Exact functional quantization of weighted random series

Scott Schubr
Luminary Oddities

November 2026

Abstract

We determine the leading constant for optimal quantization of weighted random series $\sum_j \sqrt{\lambda_j} X_j e_j$ with regularly varying squared weights $\lambda_j \sim \phi(j)$ of index $-b < -1$ and non-Gaussian or dependent coefficients. We build on the theorem for compact stationary coefficients proved in the second edition of *Wasserstein transport for Kloosterman paths* (Luminary Research Archive, DOI 10.5281/zenodo.24445508), where the squared constant is $(b-1)^{b-1}J^b/b^b$ and J is an integral of the Lagrangian of the process rate-distortion function. The constant persists under a finite second moment for excess-distortion probabilities and loss powers below two, and at the sharp log log tail scale for higher powers. For a stationary Gaussian sequence it equals $\frac{2^{1-b}b^b}{b-1}\big(\int f^{1/b}\,d\omega/2\pi\big)^b$, where f is the absolutely continuous spectral density, so the singular part of the spectrum contributes nothing; a Weyl law for weighted Toeplitz operators on $\mathbb{Z}^q$ with matrix spectral measures and direction-dependent regularly varying weights gives the corresponding small-deviation constants and mean-square quantization constants for Gaussian fields. For compact Hilbert alphabets the critical metric entropy power is $2/(b-1)$, with a sharp logarithmic refinement and, for pure power weights, an exact integral test for general entropy envelopes. For independent Hilbert letters with finite second moment, and weights whose slowly varying part is almost nondecreasing, a finite rate-distortion integral gives the exact constant for every loss moment of order below $2/b$; for power and log-power weights $j^{-b}(\log ej)^\nu$ with $\nu \geq 0$ the order $2/b$ is sharp. For telegraph coefficients with weights j^{-b}, when the switching probability tends to zero together with the code size, the joint limit holds for every excess-distortion level and for moments below $2/(b-1)$, with the critical switching scale identified.

Contents

1 Introduction **2**

2 Stationary coefficients: the results we build on **3**
2.1 The constant . . . 3
2.2 Rate and distortion profiles . . . 4
2.3 The converse at every scale . . . 4
2.4 Arbitrary stationary laws . . . 5
2.5 Dependence and the role of the hypotheses . . . 5

3 Exact weighted functional quantization beyond bounded coefficients **6**
3.1 The limit and its hypotheses . . . 6
3.2 Finite codebooks for unbounded prefixes . . . 7
3.3 The approximation argument . . . 8
3.4 Sharpness of the marginal moment weight . . . 9

3.5 Compact Hilbert coefficients . . . 10
3.6 Gaussian time series and spectral density . . . 10

4 The sharp metric entropy threshold for weighted stationary quantization 12
4.1 Setting and results . . . 12
4.2 Optimized nets on the initial coordinates . . . 13
4.3 A compact mixing counterexample at the boundary . . . 15
4.4 General entropy envelopes . . . 16
4.5 The excess-distortion constant for independent letters . . . 17

5 The sharp loss-moment threshold for independent Hilbert coefficients 18
5.1 The moment threshold . . . 18
5.2 A random covering bound with no additive information term . . . 19
5.3 Encoding residuals at several distortion levels . . . 20
5.4 Rare letters and the critical moment . . . 21

6 Weighted Toeplitz eigenvalues and Gaussian small deviations 21
6.1 The spectral law . . . 22
6.2 Finite sections and trace cutoffs . . . 23
6.3 Gaussian consequences . . . 25

7 Joint slow-switching limits for weighted telegraph quantization 26
7.1 Joint limit . . . 27
7.2 Coding an initial time interval . . . 28
7.3 Transferring the middle interval . . . 28
7.4 Moment convergence below the early-flip threshold . . . 30
7.5 All fixed moments for subpolynomial switching . . . 30
7.6 Failure at and above the critical moment . . . 31
7.7 The critical switching scale . . . 31

1 Introduction

Functional quantization asks for the best N-point approximation of a random element of a Hilbert space. For Gaussian processes Luschgy and Pagès [13] identified the sharp leading constant from the covariance eigenvalues. Dereich, Fehringer, Matoussi and Scheutzow [2] linked quantization to small-ball probabilities, Dereich and Scheutzow [3] established the high-resolution quantization and entropy-coding constant of fractional Brownian motion, Dereich [5, 4] gave explicit asymptotic formulas for the coding complexity of diffusion processes in terms of the diffusion coefficient, and Aurzada, Dereich, Scheutzow and Vormoor [23] treated jump processes. For weighted series with non-Gaussian or dependent coefficients, sharp constants were first obtained for compact stationary coefficients in [1, Section K]. This article extends that theorem to unbounded, Hilbert-valued and Gaussian coefficients, and to telegraph coefficients whose switching rate varies with the code size.

Section 2 states, without proof, the results of [1, Section K] that we use: the leading constant is an integral of the Lagrangian of the process rate-distortion function, a universal codebook sequence attains the whole distribution of ergodic-component constants, and a slowly switching binary Markov chain has a strictly smaller constant than independent signs with the same marginal, with slow-switching limit given by an explicit functional of the rate-distortion Lagrangian of the

continuous-time telegraph process. For independent bounded coefficients this contains the finite-support constant of the Kloosterman limit [1].

The new results of this article are in Sections 3–7. Section 3 removes boundedness: a finite second moment suffices for excess-distortion probabilities and loss powers below two, and a vanishing weighted tail at the sharp $\log\log$ scale suffices for higher powers, with finite-entropy examples showing that a bounded tail does not. Its Gaussian corollary shows that only the absolutely continuous part of the spectral measure enters the constant. Section 4 finds the sharp metric entropy boundary for compact Hilbert alphabets, where the critical covering power is $2/(b-1)$, and, for pure power weights, an exact integral test for entropy envelopes. The counterexamples are single stationary laws with mixing shifts on multiscale alphabets, whereas Lindenstrauss and Tsukamoto [17] use multiscale alphabets to separate the supremum of rate-distortion dimensions from metric mean dimension. Section 5 shows that for independent Hilbert letters with finite second moment, and weights whose slowly varying part is almost nondecreasing, a finite rate-distortion integral gives the exact constant for every moment of order below $2/b$, and that $2/b$ is sharp for the weights $j^{-b}(\log ej)^{\nu}$, $\nu \geq 0$. Section 6 proves a Weyl law for weighted Toeplitz covariance operators on $\mathbb{Z}^q$ with matrix spectral measures and regularly varying, direction-dependent weights. It is the lattice–Toeplitz counterpart of the classical supercritical theorem of Birman and Solomyak [6], in which singular parts of the measure do not enter the leading term, and it comes with an explicit atomless singular example showing that the lower-order contribution can decay arbitrarily slowly. It extends the scalar small-deviation theorem of Lifshits and Nazarov [20] and gives exact small-deviation constants and mean-square quantization constants for Gaussian fields. Section 7, for weights j^{-b}, lets the switching probability of the telegraph example vanish together with the code size and identifies the critical scale at which the moment limit changes.

2 Stationary coefficients: the results we build on

This section states, without proof, the results on compact stationary coefficients that the rest of the article uses. All of them are proved in the part "Exact functional quantization for weighted stationary sequences" of the second edition of the article [1, Section K], whose notation we keep. The leading finite-support approximation constant for Hilbert-valued series with regularly varying squared weights and compact stationary ergodic coefficients is an integral of the Lagrangian of the process rate-distortion function. It holds for every fixed distortion moment and every fixed excess-distortion probability, with a strong converse. For arbitrary stationary laws one universal sequence of codebooks attains the distribution of the ergodic-component constants, and dependence can strictly lower the constant while the one-coordinate law stays fixed.

2.1 The constant

Let $X = (X_j)_{j\geq 1}$ be a stationary ergodic sequence in a fixed compact subset K of $\mathbb{R}^d$. Let C denote its closed convex hull. Let $\lambda_j > 0$ with $\lambda_j \sim \phi(j)$, where ϕ is positive and regularly varying of index $-b$, $b > 1$. The weighted vector $Y = (\sqrt{\lambda_j} X_j)_{j\geq 1}$ belongs to $\ell^2(\mathbb{R}^d)$. For $s \in [1, \infty)$ let $e_{N,s}(Y)$ be the infimum over codebooks of size at most N of the L^s norm of distance to the codebook. Let $d_N(\eta; Y)$ be the optimal squared-distance threshold exceeded with probability at most $\eta \in (0, 1)$.

With logarithms natural, put

$$F_m(a) = \frac{1}{m} \inf \left\{ I(X_1^m; V_1^m) + a\, \mathbb{E} \sum_{j=1}^{m} \|X_j - V_j\|^2 \right\}, \qquad a > 0,$$

where $X_1^m = (X_1, \dots, X_m)$, and the infimum is over channels with V_1^m in C^m. Write $F(a) = \lim_{m\to\infty} F_m(a)$ and

$$J = \int_0^\infty F(u^{-b})\,du, \qquad \kappa_b = \frac{(b-1)^{b-1}}{b^b} J^b.$$

Thus $F(a) = \inf_D\{R(D) + aD\}$ is the Lagrangian of the process rate-distortion function R for per-coordinate squared error, and J integrates this Lagrangian along the weights. When the dependence on b is clear we write κ for κ_b; for a nonergodic law, κ_θ denotes the constant of the ergodic component θ.

Theorem 2.1 ([1, Section K])**.** *The displayed limit $F(a)$ exists for every $a > 0$, and $J < \infty$. For every fixed $s \in [1, \infty)$ and $\eta \in (0, 1)$,*

$$\lim_{N\to\infty} \frac{e_{N,s}(Y)^2}{(\log N)\phi(\log N)} = \lim_{N\to\infty} \frac{d_N(\eta; Y)}{(\log N)\phi(\log N)} = \kappa_b.$$

If $J > 0$, then for any $c < \kappa_b$ the largest mass covered by N balls of squared radius $c(\log N)\phi(\log N)$ tends to zero.

When the coefficients are independent and identically distributed, $F_m = F_1$, and the theorem contains the exact bounded-coordinate formula for the Kloosterman limit proved in the part "The finite-support approximation constant of the Kloosterman limit" of [1]. Luschgy and Pagès [13] proved the corresponding sharp Gaussian Hilbert-space formula from the covariance eigenvalues. The coding input in [1] is the stationary block coding theorem [9, Section 11.4, Theorem 11.4.1 and the random-coding bound (11.14)], and the converse uses the block-entropy method underlying sample converses for lossy coding [10, 12], with one code budget across unequal coordinate penalties.

2.2 Rate and distortion profiles

The following facts are established in [1, Section K]. The sequence $mF_m(a)$ is subadditive, and uniformly in m,

$$0 \le F_m(a) \le C_0 \min\{a, 1 + \log(1 + a)\}.$$

The limit F is increasing and concave, and locally absolutely continuous on $(0, \infty)$. Where it is differentiable, set $D(a) = F'(a)$ and $I(a) = F(a) - aD(a)$; these are the limiting normalized distortion and information of near-optimal channels at penalty a. Then

$$\int_0^\infty u^{-b} D(u^{-b})\,du = J/b, \qquad \int_0^\infty I(u^{-b})\,du = (1 - 1/b)J =: B_0. \tag{1}$$

2.3 The converse at every scale

The converse of [1, Section K] controls one global codebook by conditional block entropies across several frequency bands. In the setting of Theorem 2.1, suppose a sequence of codebooks of sizes N has probability bounded below by a positive constant of squared error at most t_L, along a scale $L \to \infty$. Then

$$\liminf_{L\to\infty} \frac{\log N + t_L/\phi(L)}{L} \ge J. \tag{2}$$

Optimizing over L, with the choice $L = (\log N)/B_0$, gives the strong converse in Theorem 2.1. The matching codes of [1, Section K] use rate $B_0 L$ and weighted error $(J/b)L\phi(L)$ on the central bands, by (1), together with Euclidean grids on the low coordinates and a constant tail.

2.4 Arbitrary stationary laws

Let the compact coefficient law now be stationary, with ergodic decomposition $\mathsf{P} = \int \mathsf{P}_\theta \, d\vartheta(\theta)$. Write $\kappa_b(\theta)$ for the constant of Theorem 2.1 computed on P_θ, and G for its distribution function under ϑ. These constants are bounded uniformly and measurable.

Theorem 2.2 ([1, Section K])**.** *Under any stationary law on K, for every fixed finite $s \geq 1$,*

$$\lim_{N\to\infty} \frac{e_{N,s}(Y)^2}{(\log N)\phi(\log N)} = \left(\int \kappa_b(\theta)^{s/2} \, d\vartheta(\theta) \right)^{2/s}.$$

At each continuity point x of G, the greatest mass that N balls of squared radius $x(\log N)\phi(\log N)$ can cover converges to $G(x)$. One sequence of codebooks independent of the stationary law attains both conclusions. The lower and upper limits of the normalized fixed-level quantile are bounded respectively by

$$\inf\{x : G(x) \geq 1 - \eta\}, \qquad \inf\{x : G(x) > 1 - \eta\}.$$

If these bounds coincide, the quantile converges to their common value.

The proof in [1, Section K] applies the ergodic strong converse conditionally on θ: for any one sequence of N-point codebooks and any component with $\kappa_b(\theta) > 0$, the conditional probability of normalized squared error below any fixed smaller threshold tends to zero. We use this conditional converse repeatedly below. Gray [9, Theorem 11.3.1] gives the corresponding ergodic decomposition of the fixed-rate distortion function for ordinary additive distortion, and universal stationary lossy coding goes back to Yang and Kieffer [11]. The common codes are built from the following finite-alphabet lemma.

Lemma 2.3 ([1, Section K])**.** *On a finite alphabet with bounded one-letter distortion, at each fixed rate $r > 0$ there are common length-n codebooks of size at most e^{nr} with the following property. For any stationary ergodic source and any fixed m-block channel whose information per letter is strictly less than r, the code's average distortion is at most the channel distortion per letter plus any fixed slack with probability tending to one.*

2.5 Dependence and the role of the hypotheses

The following statements are proved in [1, Section K].

When the constant vanishes. For a compact finite-dimensional alphabet, $J = 0$ exactly when the shift has zero Kolmogorov–Sinai entropy.

Binary sources and Fourier series. Let F_{iid} and F_{Markov} be the objectives of independent uniform signs and of the stationary symmetric binary Markov source with the same marginal and flip probability $\rho \in (0, 1)$, $\rho \neq 1/2$. Then

$$F_{\mathrm{iid}}(a) = \min_{0 \leq z \leq 1} \left[\log 2 - H_2((1+z)/2) + a(1 - z^2) \right],$$

where H_2 is binary entropy, and $F_{\mathrm{Markov}} \leq F_{\mathrm{iid}}$ everywhere, with strict inequality for $a > 1/2$. Hence the Markov quantization constant is strictly smaller than the independent one, although the one-coordinate laws agree. For $0 < \rho \leq 1/4$,

$$c \min(a, \rho) \leq F_{\mathrm{Markov}}(a) \leq C \min\{a, \rho[1 + \log_+(a/\rho)]\}, \tag{3}$$

so that $J \asymp \rho^{1-1/b}$ and $\kappa_b \asymp \rho^{b-1}$. By Parseval's identity, the same constants hold for the random Fourier series $\sum_{j \geq 1} \sqrt{\lambda_j} X_j e^{2\pi i j t}$ in $L^2[0, 1]$.

The slow-switching coefficient. Let Z be the stationary continuous-time sign process that flips at rate one. Define $F_{\mathrm{tel}}(a)$ to be the large-T limit of

$$\frac{1}{T}\inf_{V}\left[I(Z_{[0,T]};V)+a\,\mathbb{E}\int_0^T|Z(t)-V(t)|^2\,dt\right]$$

over reproductions in $L^2[0,T]$ taking values in $[-1,1]$ almost everywhere; this is the Lagrangian of the telegraph-process rate-distortion function, and the limit exists by subadditivity. Writing $\kappa_b(\rho)$ for the constant of the binary Markov source with flip probability ρ, the slow-switching limit exists and is an explicit functional of this Lagrangian:

$$\lim_{\rho\downarrow 0}\frac{\kappa_b(\rho)}{\rho^{b-1}}=\frac{(b-1)^{b-1}}{b^b}\left(\int_0^\infty F_{\mathrm{tel}}(v^{-b})\,dv\right)^b, \tag{4}$$

and the integral is positive and finite by (3). The proof compares the Δ-skeleton of Z, with $\Delta=-\frac{1}{2}\log(1-2\rho)$, with the Markov chain, through $|F_{\mathrm{Markov}}(\Delta a)/\Delta-F_{\mathrm{tel}}(a)|\le Ca\Delta$. The convergence in (4) concerns the weighted quantization constant after the code-size limit at each fixed ρ; Section 7 treats the two limits jointly.

The rate of the strong converse. For the indicator sequence of a stationary integer renewal process with gap probabilities proportional to m^{-a_0}, $a_0>2$, and $\lambda_j=j^{-b}$, one has $J>0$, and the strong converse holds at a rate polynomial in $\log N$: with probability bounded below by a negative power of $\log N$, the zero codeword already has squared error below any fixed positive fraction of the asymptotic constant. Thus the rate of the strong converse depends on the source.

Why finite-dimensional coefficient space matters. For iid coefficients in a Hilbert space of infinite dimension, compactness and $0<J<\infty$ do not by themselves determine the error. With $\lambda_j=j^{-b}$, a compact law placing mass $p_i=2^{-i}L_i^{-(b-1)/b}$ uniformly on M_i orthogonal points of norm 2^{-i}, with $L_i=\log M_i\asymp 2^{i^2}$, and the remaining mass at zero, has $0<J<\infty$. For the code sizes $N_i=\lfloor M_i/2\rfloor$, however, the normalized error $e^2_{N_i,2}/(\log N_i)^{1-b}$ tends to infinity. Sections 3 and 4 identify the metric entropy conditions under which Hilbert alphabets behave like finite-dimensional ones.

3 Exact weighted functional quantization beyond bounded coefficients

For weighted stationary sequences in a fixed Euclidean coefficient space, the compact-alphabet functional quantization constant remains exact under a finite second moment, both for excess-distortion probabilities and for loss powers $1\le s<2$. For $s\ge 2$ a vanishing weighted marginal tail suffices at the sharp scale of powers of $\log\log\|X_1\|$, with the regular variation factor included. A bounded weighted tail can fail even for independent scalar coefficients of finite entropy. The codes can be chosen independently of the stationary law. We also compute the Gaussian time-series constant directly from the absolutely continuous part of the spectrum.

3.1 The limit and its hypotheses

Let $X=(X_j)_{j\ge1}$ be a stationary sequence in $\mathbb{R}^d$ with $\mathbb{E}\|X_1\|^2<\infty$. Let $b>1$, and let $\lambda_j>0$ satisfy $\lambda_j\sim\phi(j)$, where ϕ is positive and regularly varying of index $-b$. Then $Y=(\sqrt{\lambda_j}X_j)_{j\ge1}$ belongs almost surely to the weighted series space $\ell^2(\mathbb{R}^d)$. Write $h_R=\sqrt{R\phi(R)}$ for $R>0$. For

$s \ge 2$, fix a positive nondecreasing function w_s on $[1,\infty)$ comparable at infinity to h_v^{-s}. Such a function exists by regular variation of positive index. Put $\Lambda(x) = 1 + \log\log(e^e + x)$, and define the tail

$$A_s(v) = \mathbb{E}\big[\|X_1\|^s \mathbf{1}_{\{\Lambda(\|X_1\|)>v\}}\big].$$

For each law of the process, put

$$F_m(a) = \frac{1}{m} \inf \left\{ I(X_1^m; V_1^m) + a\,\mathbb{E} \sum_{j=1}^{m} \|X_j - V_j\|^2 \right\}, \qquad a > 0,$$

with the infimum over reproduction channels in the coefficient space. Logs and information are in nats. For an ergodic component θ let

$$F_\theta(a) = \lim_{m\to\infty} F_{m,\theta}(a), \qquad J_\theta = \int_0^\infty F_\theta(u^{-b})\,du, \qquad \kappa_\theta = \frac{(b-1)^{b-1}}{b^b} J_\theta^b.$$

Write ϑ for the ergodic decomposition measure, and G for the distribution function of κ_θ.

Theorem 3.1. *The limits defining F_θ exist and J_θ is finite for almost all θ. At every continuity point x of G, the maximum probability covered by N balls of squared radius $x h_{\log N}^2$ converges to $G(x)$. For $1 \le s < 2$, and also for $s \ge 2$ when $w_s(v)A_s(v) \to 0$ as $v \to \infty$,*

$$\lim_{N\to\infty} \frac{\inf_{|\mathcal{C}|\le N} \big(\mathbb{E}\,\mathrm{dist}(Y,\mathcal{C})^s\big)^{1/s}}{h_{\log N}} = \left(\int \kappa_\theta^{s/2}\, d\vartheta(\theta) \right)^{1/s}.$$

A single sequence of codebooks, depending only on the dimension and weights, attains the probability and moment limits for all the indicated stationary laws and powers. The strong moment condition $\mathbb{E}[\|X_1\|^s w_s(\Lambda(\|X_1\|))] < \infty$ implies the tail condition. For pure powers $\lambda_j = j^{-b}$, the corresponding moment weight is $\Lambda(\|X_1\|)^{(b-1)s/2}$.

The compact stationary result, including conditional converses and a universal code, is stated in Section 2 and proved in [1, Section K]. For independent bounded coefficients it specializes to the finite-support constant of the Kloosterman limit [1]. Here a deterministic finite code controls rare large coefficient prefixes. This allows us to split the series into a bounded part and a residual with small marginal moment. The classical Gaussian theorem of Luschgy and Pagès [13] starts from the covariance eigenvalues of the weighted vector. Our stationary formula also works without this spectral representation.

3.2 Finite codebooks for unbounded prefixes

Lemma 3.2. *For all sufficiently large B, there are codebooks $\mathcal{B}_B$ of at most e^B points depending only on d, λ, with*

$$\|\,\mathrm{dist}(Y, \mathcal{B}_B)\|_{L^s} \le C_s h_B$$

whenever $1 \le s < 2$ and $\mathbb{E}\|X_1\|^2 \le 1$, or $s \ge 2$ and $\mathbb{E}\|X_1\|^s + \sup_{v\ge1} w_s(v)A_s(v) \le 1$. The constants are uniform over dependence and over the indicated normalized moment class, with the choice of comparable weight w_s fixed.

Proof. Take $L = \lfloor cB \rfloor$ for a sufficiently small fixed $c > 0$, and set $\Delta_{i,L} = \min\{1, \sqrt{\phi(L)/\lambda_i}\}$ for $i \le L$. For lattice index $k \in \mathbb{Z}^d$ use length $\ell(k) = c_d + (d+2)\log(1+\|k\|)$, where $\sum_k e^{-\ell(k)} \le 1$. For a nonnegative integer scale index j use length $s_j = c_0 + 2\log(1+j)$ with $\sum_j e^{-s_j} \le 1/2$. Include

every reconstruction whose first L coordinates, before weighting, lie on the respective lattices of meshes $2^j\Delta_{i,L}$, with tail zero, when the index cost $s_j + \sum_{i=1}^{L} \ell(k_i)$ is at most B. Also include zero. The exponential sum of these lengths is at most $1/2$, so there are at most $e^B/2 + 1 \le e^B$ points for large B.

For a realization choose 2^j to be the least dyadic scale at least $\rho = \max(1, L^{-1} \sum_{i \le L} \|X_i\|)$. After dividing by 2^j, round each head coordinate. The total lattice cost is bounded by

$$C_1 L + C_2 \sum_{i \le L} \log(1 + \|X_i\|/2^j) \le C_3 L.$$

Here the mesh costs sum to $O(L)$, by the Potter bound $\log(1/\Delta_{i,L}) \le C(1 + \log(L/i))$ and summation. The last inequality uses concavity of the logarithm and the definition of ρ. Choose c so that the lattice cost is at most $B/3$. The selected scale then fits whenever $j \le \lfloor e^{c_1 B} \rfloor$ for some fixed $c_1 > 0$. On this event the rounding error in the head is at most $C 2^j h_L$. Its L^s norm is $O(h_B)$: by Jensen and stationarity, $\mathbb{E}\rho^s \le 1 + \mathbb{E}\|X_1\|^s$. The tail after L has L^s norm at most $C h_B$ by Minkowski in $L^{\max(2,s)/2}$ and regular variation.

On the remaining event E, the maximum norm in the first L coefficients exceeds $T_B = 2^{\lfloor e^{c_1 B} \rfloor - 1}$. In the normalized moment classes a union bound and Markov's inequality give $\mathbb{P}(E) \le CB/T_B^2$. For $s < 2$, Hölder's inequality with the bounded L^2 norm of $\|Y\|$ gives $\| \|Y\| \mathbf{1}_E \|_{L^s} = o(h_B)$. For $s \ge 2$, split the marginal s-moment on E at $T_B^{1/(s+1)}$. On the part below the cutoff, use $\mathbb{P}(E) \le CB/T_B^2$, giving a bound $CB/T_B^{2-s/(s+1)} = o(h_B^s)$. Above the cutoff, the weighted tail bound gives $1/w_s(\Lambda(T_B^{1/(s+1)}))$. The argument of w_s there is asymptotic to $c_1 B$, so this bound is $O(h_B^s)$ by regular variation. Thus for every i, $\mathbb{E}[\|X_i\|^s \mathbf{1}_E] \le C h_B^s$. Minkowski in $L^{s/2}$ and $\sum_i \lambda_i < \infty$ bound the L^s norm of the zero-reproduction error on E by $C h_B$. This proves the lemma. □

3.3 The approximation argument

For radius $M > 0$, let $X^{(M)}$ be X projected coordinatewise onto the radius-M ball. The associated weighted vectors are related by nonexpansive projection onto the closed convex product of the weighted balls. For one stationary law with second moment $\sigma^2 = \mathbb{E}\|X_1\|^2$, let $\delta_M = \|X_1 - X_1^{(M)}\|_2$. The block objectives satisfy

$$0 \le F_m(a) - F_m^{(M)}(a) \le a(2\sigma\delta_M + \delta_M^2). \tag{5}$$

The first inequality projects input and reproduction and applies data processing. For the second, use an approximately optimal projected channel after projection of the original source. The information is unchanged from the projected channel by its conditional independence. The square root of per-coordinate distortion increases by at most δ_M. The projected distortion of a near minimizer is at most $\sigma^2 + o(1)$ by comparison with the zero channel. Let the optimization tolerance tend to zero. The same arguments show monotonicity of $F_m^{(M)}$ in M.

For each fixed a, $mF_m(a)$ is subadditive by using two adjacent block channels conditionally independently given the source. Their distortions add by stationarity and their joint mutual information is at most the sum of the two block informations. One way to see the last claim is to bound the joint relative entropy against the product of the two reproduction marginals. Thus the limit $F(a)$ exists. The zero channel gives $F(a) \le a\sigma^2$. A one-coordinate lattice quantizer of mesh $(1 + a)^{-1/2}$, with the lengths used in Lemma 3.2, gives

$$F(a) \le C_d(1 + \log(1 + a) + \mathbb{E}\log(1 + \|X_1\|)).$$

Here the entropy of the lattice index is at most the expected length, by the relative entropy inequality, and a times the squared mesh is bounded. These two estimates prove that J is finite. By (5) and monotone convergence, $J^{(M)} \uparrow J$. This applies on almost every ergodic component by the conditional second moment assumption. Measurability follows from the compact case and the increasing limit. If the marginal second moment is zero, the constant is zero. Otherwise after normalizing the marginal second moment to one, the preceding bounds give $J \le C_{d,b}$; rescaling the coefficients by σ gives $J \le C_{d,b}\sigma^{2/b}$. In particular $\sqrt{\kappa_\theta} \le C\sigma_\theta$, and its s-th moment is finite in the range of the theorem.

For any sequence of competitors, fix an ergodic component and a squared threshold $x < \kappa_\theta$. For large enough M the compact constant on the projected component exceeds x. Projection of the codebooks and the compact ergodic converse prove that the conditional mass below $xh^2_{\log N}$ tends to zero. Equivalently, under each component the normalized distance cannot fall below $\sqrt{\kappa_\theta}$ by any fixed positive amount with positive limiting probability. Integrating gives the coverage upper bound at continuity points of G, and conditionally Fatou's lemma (or integration of the conditional tail bound) gives the moment lower bound. This applies to near-optimal sequences of competitors.

For the upper bound choose $0 < \varsigma < 1$, a cutoff radius M, and a scale $\rho > 0$. With total logarithmic budget B, add a compact universal codebook for $X^{(M)}$ of log size at most $(1-\varsigma)B$ to $\rho\mathcal{B}_{\varsigma B}$, where the latter is used to encode $(X - X^{(M)})/\rho$. Each sum of two points is a candidate codeword. For any fixed $\rho > 0$ choose M large enough for the scaled residual to satisfy the normalized hypothesis of Lemma 3.2 at the power s under consideration. For $s < 2$ this follows from the vanishing residual second moment. For $s \ge 2$ put $U = \|X_1\|$ and $V_M = \|X_1 - X_1^{(M)}\|/\rho$. Then $V_M \le \rho^{-1}U\mathbf{1}_{\{U>M\}}$ and $\Lambda(U/\rho) \le \Lambda(U) + C_\rho$. For all sufficiently large v, the scaled residual's weighted tail at v is therefore at most $Cw_s(v-C_\rho)A_s(v-C_\rho)$, uniformly in M, since bounded shifts change w_s by at most a constant factor. Its supremum for v above a large fixed cutoff can be made arbitrarily small. Below that cutoff, the supremum is at most a constant times $\mathbb{E}V_M^s$, which tends to zero. The lemma then bounds the residual error in L^s by $C_s\rho h_{\varsigma B}$. The compact code has normalized error converging in probability to $\sqrt{\kappa_\theta^{(M)}}$ and has the corresponding L^s norm limit for every finite s. Finally, $h_{\gamma B}/h_B \to \gamma^{(1-b)/2}$ for $\gamma > 0$.

These choices can be made universally. For $R = \log N$, use $B = R - \sqrt{R}$ and include the codebooks for the first $\lfloor R \rfloor$ triples in a fixed enumeration of positive integer M, positive rational ρ and rational $0 < \varsigma < 1$. The union has at most e^R points for large R. For any fixed profile the resulting error is no larger than the profile error. Take $\rho \downarrow 0$ with a suitable M, and then $\varsigma \downarrow 0$. For convergence in probability to $\sqrt{\kappa_\theta}$ from above, use the bound at $s = 1$ and Markov's inequality on the residual. The compact component constants satisfy $\kappa_\theta^{(M)} \le \kappa_\theta$, and the factor $(1-\varsigma)^{(1-b)/2}$ tends to one. The converse gives convergence in probability from below. For the L^s upper bound, use Minkowski and the compact component moment limit. This proves Theorem 3.1.

3.4 Sharpness of the marginal moment weight

Theorem 3.3. *For every $b > 1$ and every $s \ge 2$, there are iid real coefficients with finite s-th moment and finite entropy, and weights $\lambda_j = j^{-b}$, for which $0 < J < \infty$ but $e_{N,s}(Y)/h_{\log N} \to \infty$. They can be chosen to have finite $\mathbb{E}|X_1|^s\Lambda(|X_1|)^\alpha$ for any fixed $0 \le \alpha < (b-1)s/2$. For general regularly varying weights, let $W \ge 1$ be nondecreasing and suppose $\liminf_{v\to\infty} W(v)h_v^s = 0$. Then a finite $\mathbb{E}|X_1|^s W(\Lambda(|X_1|))$ does not suffice for the formula, even for iid scalar letters with finite entropy and $0 < J < \infty$. Even $\sup_v w_s(v)A_s(v) < \infty$ can fail to give the exact constant. Conversely, the vanishing tail condition is strictly weaker than the strong weighted moment condition.*

Proof. Choose $\alpha < \gamma < (b-1)s/2$, take atoms $x_i = 3^i$ for $i \geq 3$, and put respective probabilities

$$p_i = \frac{c}{3^{is} i (\log i)^{1+\gamma}}.$$

Place the remaining mass at zero with $c > 0$ small. The weighted moment and entropy are finite since $\Lambda(3^i) \asymp \log i$. Independence gives $F_m(a) = F_1(a)$ by tensorization. Sending the exact letter or reproducing zero gives $F(a) \leq \min\{H(X_1), a\mathbb{E}|X_1|^2\}$, which also proves $J < \infty$. The objective is positive for all $a > 0$: achieving arbitrarily small distortion allows classification of one isolated atom of intermediate probability with error tending to zero, forcing positive information by the binary entropy inequality.

Each first-coordinate codepoint lies within $3^i/4$ of at most one x_i, by geometric separation. Any N-point code therefore misses all but at most N atoms in these respective intervals. The resulting first-coordinate lower bound on s-th power distortion is a constant times the sum of $1/[i(\log i)^{1+\gamma}]$ over the missed indices. This decreasing sequence gives a lower bound of order $(\log N)^{-\gamma}$. Since $h^s_{\log N} = (\log N)^{-(b-1)s/2}$, the ratio diverges.

For the general claim, choose $v_l \to \infty$ sparse enough that $W(v_l) h^s_{v_l} \leq 2^{-2l}$. For a sufficiently large constant C, the disjoint index blocks $[M_l, 2M_l]$ with $M_l = \lfloor e^{v_l - C} \rfloor$ satisfy $\Lambda(3^i) \leq v_l$ throughout block l. Give each block a total s-moment mass proportional to $2^{-l}/W(v_l)$, shared equally among its indices. The probability at $x_i = 3^i$ is the assigned moment mass divided by 3^{is}, and the remaining probability is at zero. This gives the claimed weighted moment by monotonicity. The entropy is finite, for example because $p_i \leq C 3^{-is}$, which also makes $\sum_i \sqrt{p_i}$ finite. The same information argument gives $0 < J < \infty$. At $N_l = \lfloor M_l/2 \rfloor$, at least half of the l-th block is missed in the first coordinate. Its normalized distortion is bounded below by a constant times $2^{-l}/[W(v_l) h^s_{\log N_l}]$, which tends to infinity since $\log N_l = v_l - O(1)$ and h is regularly varying.

For the bounded-tail claim choose v_l with $h^s_{v_{l+1}} \leq h^s_{v_l}/2$, use the same blocks, and give block l mass $h^s_{v_l}$ in s-moment units. For a threshold between v_{l-1} and v_l, earlier blocks have no mass above it and the remaining mass is at most $2h^s_{v_l}$. Multiplication by $w_s(v_l)$ is bounded. The first-coordinate distortion divided by $h^s_{\log N_l}$ is bounded below by a fixed positive constant, as before. Taking v_1 large makes $\mathbb{E}|X_1|^s$ arbitrarily small and hence makes $\sqrt{K} \leq C\|X_1\|_2$ arbitrarily small, so the exact constant fails. All nonzero atoms exceed one, and the distribution again has finite entropy and positive J.

For strictness, with the same choice of v_l, put s-moment mass $h^s_{v_l}/l$ at a single positive atom with $\Lambda(x_l) = v_l$. The preceding tail bound after weighting is $O(1/l)$ between consecutive thresholds, while the strong weighted moment diverges by comparison with the harmonic series. □

3.5 Compact Hilbert coefficients

Remark 3.4. The projections used in (5) require only Hilbert norms and nonexpansive projection. Hence the approximation argument extends to coefficients in a fixed compact set K of a separable Hilbert space once the initial coordinates are coded by nets of controlled size. Theorem 4.1 carries this out with an optimized covering radius in each coordinate: if $\log N(K, \varepsilon) = O(\varepsilon^{-\beta})$ with $0 < \beta < 2/(b-1)$, the probability and all finite power moment limits of Theorem 3.1 hold for every stationary law on K, with one universal sequence of codebooks depending on K and the weights. Section 4 also shows that the power $2/(b-1)$ is sharp.

3.6 Gaussian time series and spectral density

Corollary 3.5. *Let X be a centered stationary real Gaussian sequence with finite spectral measure μ. Let f be the density of its absolutely continuous part relative to $d\omega/(2\pi)$ on $[-\pi, \pi]$. Then the*

normalized squared quantization constant for every finite moment and every fixed excess-distortion level in $(0,1)$ *is*

$$\frac{2^{1-b}b^b}{b-1}\left(\int_{-\pi}^{\pi} f(\omega)^{1/b}\frac{d\omega}{2\pi}\right)^b.$$

The singular part of the spectral measure contributes zero. At a fixed positive variance the constant is strictly smaller than that of the iid Gaussian sequence unless the spectral measure is a constant multiple of $d\omega/(2\pi)$.

Proof. Let $T_m(\mu)$ be the Toeplitz block covariance. Diagonalizing it and using independent Gaussian rate distortion gives $F_m(a)$ as the mean over its eigenvalues of

$$H_a(t) = \begin{cases} at, & 2at \le 1, \\ \frac{1}{2}(1+\log(2at)), & 2at > 1. \end{cases}$$

Indeed, for a scalar Gaussian of variance t, the rate at distortion $0 < D \le t$ is $\frac{1}{2}\log(t/D)$. Minimize after adding aD. The scalar formula follows by the Gaussian maximum-entropy bound on the reconstruction error and the independent Gaussian error construction. Information tensorization then gives the stated within-block sum.

The Toeplitz eigenvalue distribution theorem for Radon measures [15, Theorem 1.1] identifies the limiting mean of a continuous, compactly supported eigenvalue test function with its integral applied to f. It also applies to H_a by truncation: $H_a(t) = o(t)$ at infinity, and $\operatorname{tr}(T_m(\mu))/m$ is the fixed spectral mass. Consequently the process objective for the full Gaussian law is $F(a) = \int H_a(f(\omega))\, d\omega/(2\pi)$.

When μ is absolutely continuous the covariance at lag j tends to zero by the Riemann–Lebesgue lemma. Gaussian cylinder variables then mix by convergence of joint Gaussian laws, which implies ergodicity. Thus in this case F is the objective of the sole component.

Tonelli's theorem and the substitution $u = t^{1/b}v$ give

$$\int_0^\infty H_{u^{-b}}(t)\, du = t^{1/b}\frac{2^{1/b}b^2}{2(b-1)}.$$

The integral at $t = 1$ splits at $u = 2^{1/b}$. The definition of κ_θ yields the displayed constant in the absolutely continuous case. Constant f recovers the independent Gaussian formula of Luschgy and Pagès [13, Theorem 2.2].

For a purely singular law $F(a) = 0$. Every block channel for the full law satisfies the Markov relation $\theta \to X_1^m \to V_1^m$ with the ergodic label θ. Its mutual information therefore bounds conditional mutual information averaged over θ. Taking the infimum gives $\int F_{m,\theta}(a)\, d\vartheta(\theta) \le F_m(a)$. Fatou's lemma, first at positive rational a and then monotonicity in a, gives $F_\theta(a) = 0$ almost surely. Theorem 3.1 now gives quantization error $o(h_{\log N})$ for all finite powers, and convergence of normalized distance to zero in probability.

For general μ, write the process as an independent sum of Gaussian processes with its absolutely continuous and singular spectra. Give a fixed positive fraction of the logarithmic code budget to the singular part, add the codebooks, and send that fraction to zero. The singular part has negligible error on this scale, giving the claimed upper limits. For the lower limits, condition on the realized singular vector. The translated codebook is a competitor for the independent absolutely continuous part, so it cannot improve its optimal coverage or moment risk. This proves the same constant without an ergodicity assumption on the full law. Finally, $\int f \le \mu([-\pi,\pi])$, and strict Jensen for $t^{1/b}$ gives the last assertion. □

4 The sharp metric entropy threshold for weighted stationary quantization

We identify the sharp metric entropy growth that guarantees the process rate-distortion constant for quantization of weighted stationary sequences in a compact Hilbert alphabet. For squared weights $j^{-b}[\log(ej)]^{\nu}$, $b > 1$, the critical entropy power is $2/(b-1)$. At that power an entropy saving of order $(\log(1/r))^{-\gamma}$ suffices precisely when $\gamma(b-1) > \max\{b, -\nu\}$, uniformly over stationary laws on each such alphabet. At and below the boundary there are compact alphabets and laws with a mixing shift and a positive finite process constant for which the fixed-rate formula fails. Under pure power weights we also give a sharp integral test for general entropy envelopes. For independent square-integrable Hilbert letters under pure power weights, a finite process integral already suffices for the exact excess-distortion constant. For general regularly varying weights, we characterize exactly when this independent-letter conclusion follows from a finite process integral alone.

4.1 Setting and results

Let K be a fixed compact subset of a separable real Hilbert space, and write $H_K(r) = \log N(K, r)$ for its covering entropy in the Hilbert norm. Let $X = (X_j)_{j\ge1}$ be stationary in K, and $Y = (\sqrt{\lambda_j}X_j)_{j\ge1}$ its weighted vector. The squared weights $\lambda_j > 0$ satisfy $\lambda_j \sim \phi(j)$, with $\phi > 0$ regularly varying of index $-b$, $b > 1$. Set $h_R = \sqrt{R\phi(R)}$. For a law of X the normalized block objective is

$$F_m(a) = \frac{1}{m} \inf \left\{ I(X_1^m; V_1^m) + a\,\mathbb{E} \sum_{j=1}^{m} \|X_j - V_j\|^2 \right\}, \qquad a > 0.$$

The infimum is over channels in the Hilbert space. Write $F(a) = \lim_m F_m(a)$, $J = \int_0^\infty F(u^{-b})\,du$, and $\kappa = (b-1)^{b-1}J^b/b^b$. For a stationary nonergodic law let κ_θ denote this constant on its ergodic components. An N-point codebook is any finite subset of the weighted Hilbert space of size at most N.

Say that the exact formula holds if, at continuity points x of the distribution of κ_θ, the maximum mass covered by N balls of squared radius $xh^2_{\log N}$ tends to $\mathbb{P}(\kappa_\theta \le x)$, and the minimum L^s distance to such a codebook divided by $h_{\log N}$ tends to $(\mathbb{E}\kappa_\theta^{s/2})^{1/s}$ for every finite $s \ge 1$. In the ergodic case this identifies every fixed nontrivial excess-distortion quantile.

Theorem 4.1. *The exact formula holds for every stationary law on K if $H_K(r) = O(r^{-\beta})$ for $0 < \beta < 2/(b-1)$. When $\lambda_j \sim j^{-b}[\log(ej)]^{\nu}$, $\nu \in \mathbb{R}$, it also holds under*

$$H_K(r) = O\Big(r^{-2/(b-1)}[\log(e/r)]^{-\gamma}\Big), \qquad \gamma(b-1) > \max\{b, -\nu\}.$$

In both cases a single sequence of codebooks, depending on K and the weights, attains the limits simultaneously for all stationary laws and all indicated powers.

Theorem 4.2. *For $b > 1$, $\nu \in \mathbb{R}$, and $0 \le \gamma(b-1) \le \max\{b, -\nu\}$, there is a compact Hilbert alphabet with*

$$H_K(r) = O\Big(r^{-2/(b-1)}[\log(e/r)]^{-\gamma}\Big)$$

and a stationary law on it with a mixing shift and $0 < J < \infty$, for which the exact formula fails for $\lambda_j \sim j^{-b}[\log(ej)]^{\nu}$. The formula fails at every fixed excess-distortion level. If $\gamma(b-1) \le b$, the maximal mass covered at every fixed multiple of $h_{\log N}$ tends to zero along all N. If $b < \gamma(b-1) < -\nu$, the same holds along a subsequence. In every case κ can be made arbitrarily small and positive.

Theorem 4.3. *Let $E : (0,1] \to [1,\infty)$ be finite and nonincreasing, and suppose $\lambda_j \sim j^{-b}$. The condition $H_K(r) = O(E(r))$ guarantees the exact formula for all compact K and every stationary law on K if and only if*

$$\int_0^1 r^{2/b-1} E(r)^{1-1/b}\, dr < \infty.$$

If the integral diverges there is a compact counterexample with a mixing shift, $0 < J < \infty$, and zero limiting coverage at each fixed multiple of $h_{\log N}$.

Theorem 4.4. *Write $\lambda_j \sim j^{-b}\ell(j)$, where $\ell(x) = x^b\phi(x) > 0$ is slowly varying. For iid Hilbert letters with finite second moment and $J < \infty$, the optimal squared threshold at every fixed excess-distortion probability in $(0,1)$, divided by $h^2_{\log N}$, converges to κ for every such law if and only if*

$$\sup_{1\le j\le R,\ j \text{ integer}} \ell(j)/\ell(R) = O(1) \qquad (R \to \infty).$$

If this condition fails, a compact iid alphabet and a law with $0 < J < \infty$ give zero coverage at each fixed multiple of the predicted scale along a subsequence. For logarithmic modifiers $\ell(x) = [\log(ex)]^\nu$, the condition is equivalent to $\nu \ge 0$.

The bounded Euclidean theorem with a universal code is stated in Section 2 and proved in [1, Section K]. The upper bound here optimizes the covering radius in each coordinate, and then applies the same finite-dimensional approximation. The lower construction uses layers with a long initial stretch of independent symbols at a small entropy rate. Classical functional product quantizers and metric entropy bounds [13, 16] provide the related independent-coordinate and deterministic approximation setting, including the entropy of the Gaussian Strassen ball. Lindenstrauss and Tsukamoto [17, Proposition 1.7 and Section 5] use multiscale compact Hilbert alphabets to separate metric mean dimension from rate distortion for additive block metrics, comparing the supremum of rate-distortion dimensions over all invariant measures with metric mean dimension. Our counterexamples also live on multiscale alphabets, but the metric is one-sided and weighted, and each counterexample is a single stationary law with a mixing shift.

4.2 Optimized nets on the initial coordinates

Set

$$B_K(a) = \inf_{0<r\le 1}\{H_K(r) + ar^2\}.$$

It is finite, nondecreasing and concave on $(0,\infty)$. The zero reproduction and a net-valued deterministic channel give $F(a) \le \min\{Ca, B_K(a)\}$. The limit defining F exists by subadditivity of $mF_m(a)$, using independent channels on adjacent blocks conditional on the source and stationarity. The same observations apply to every ergodic component.

We first prove sufficiency of the bounds

$$\int_0^1 B_K(u^{-b})\, du < \infty, \qquad \lim_{\alpha\downarrow 0}\limsup_{R\to\infty} \frac{1}{R}\sum_{j\le \alpha R}\left[1 + B_K(\lambda_j/\phi(R))\right] = 0. \tag{6}$$

Choose nested finite-dimensional orthogonal projections P_l such that $\sup_{x\in K}\|(I-P_l)x\| \le \delta_l \to 0$. They can be obtained by spanning finite nets. Projection of both source and reproduction locally decreases distortion and mutual information. Conversely a channel for the projected source can be used on the unprojected source after projection, with the same information and root mean-square error per letter increased by at most δ_l. Comparison with the zero channel bounds the distortion

of a near minimizer uniformly. Hence, for fixed a, the projected block objective differs from the full one by at most $Ca(\delta_l+\delta_l^2)$, uniformly in block length and in the law on K. The projected objectives increase with l, by nestedness, and the integrability assumption gives convergence of their constants on every ergodic component by monotone convergence. Projection of any competitor codebook and the compact Euclidean converse of Section 2 give the conditional lower bound on distance at every threshold below $\sqrt{\kappa_\theta}$. Integrating over components gives the moment lower bound and the coverage upper bound at continuity points.

For the upper bound consider the residual alphabet $(I-P_l)K$. Its covering entropy is at most H_K because orthogonal projection is a contraction. At budget scale R, for $j \le \alpha R$ choose a radius whose objective is at most $B_K(a_j)+1$, where $a_j = \lambda_j/\phi(R)$. The log-size of the product of these nets and the squared head error divided by $\phi(R)$ are both bounded by the sum in (6). Put the tail to zero. Its error is at most $C\delta_l\alpha^{(1-b)/2}h_R$ for each fixed $\alpha > 0$, by summation of the weights. Thus for any reserved fraction $\varsigma > 0$ and error tolerance, first choose α small so that the head budget is at most ςR and the deterministic head error is within the tolerance times h_R. Then choose l large for the tail bound. Add the residual codebook to the compact universal codebook for P_lX at logarithmic budget $(1-\varsigma)R$. The two codebook sizes multiply and their errors add. The projected compact normalized error converges in probability to $\sqrt{\kappa_\theta^{(l)}}$ and in every finite moment by the compact theorem. The scale factor from the compact rate budget tends to $(1-\varsigma)^{(1-b)/2}$.

To get a single code sequence, enumerate the profiles with integer l and positive rational α, ς. At total log budget B, include the first $\lfloor B \rfloor$ profiles evaluated at $R = B - \sqrt{B}$, each using at most e^R points. Include only those satisfying the residual budget. Every fixed profile with strict slack in the residual budget is included for large B and the union has at most e^B points. The error of the union is bounded by each fixed profile's error. Sending the tolerances and then ς to zero proves the coverage lower bound and the moment upper bound. The integral condition bounds the component constants uniformly. This proves both exact limits.

For $H_K(r) \le Cr^{-\beta}$, choose $r = a^{-1/(2+\beta)}$ for $a \ge 1$. It gives $B_K(a) \le Ca^{\beta/(2+\beta)}$. Since $b\beta/(2+\beta) < 1$, the integral converges. Potter bounds give, for sufficiently small $\xi > 0$, an upper bound $C(R/j)^v$ on the summands of (6) when $j \le \alpha R$, where $v = (b+\xi)\beta/(2+\beta) < 1$. The normalized sum is $O(\alpha^{1-v})$.

At the critical power $\beta = 2/(b-1)$, choose

$$r = a^{-1/(2+\beta)}[\log(e+a)]^{-\gamma/(2+\beta)} \qquad (a \ge 1).$$

Then $B_K(a) \le Ca^{1/b}[\log(e+a)]^{-p}$, where $p = \gamma(b-1)/b > \max(1, -\nu/b)$. This is integrable at $a = u^{-b}$ near $u = 0$. For the logarithmic weights write $v_j = \log(ej)$ and $T = \log(eR)$. The normalized head sum is bounded by $C\alpha$ plus

$$C \sum_{j \le \alpha R} \frac{1}{j[1+\log(R/j)]^p} \left(\frac{v_j}{T}\right)^{\nu/b}.$$

Indeed $\log(e+a_j) \asymp 1 + \log(R/j)$ there: the extra term in $\log a_j$ is $\nu \log(v_j/T)$, with magnitude bounded by $|\nu| \log(1+\log(R/j))$ up to a constant. For $\sqrt{R} < j \le \alpha R$, the ratio v_j/T is bounded away from zero. Integral comparison with the substitution $u = \log(R/j)$ gives the limsup bound $C[\log(1/\alpha)]^{1-p}$. For $j \le \sqrt{R}$, the contribution is at most

$$CT^{-p-\nu/b}\left(1 + \int_1^{T/2+1} v^{\nu/b}\, dv\right).$$

This tends to zero under the two strict inequalities on p, including $\nu/b = -1$. This proves Theorem 4.1.

4.3 A compact mixing counterexample at the boundary

Take $r_i = 2^{-i}$ and $\beta = 2/(b-1)$. In mutually orthogonal finite-dimensional spaces choose the alphabet $K_i = \{r_i e_{i,k} : 1 \le k \le M_i\}$ with orthonormal directions, where

$$H_i = \log M_i \asymp 2^{i\beta} i^{-\gamma}.$$

The full alphabet consists of all sums of one member from each K_i, with the series converging uniformly in norm. The product is compact: the finite initial layers are finite and squared tail norms sum to $O(2^{-2i})$. At scale $r \asymp 2^{-i}$, specifying enough initial layers to make the squared tail $O(r^2)$ costs $\sum_{k \le i+O(1)} H_k = O(2^{i\beta} i^{-\gamma})$. This gives the stated covering bound. We may start with i large enough that every $H_i \ge 10$.

First consider $p = \gamma(b-1)/b \le 1$. We specify the required independent stretches before defining the process. Write

$$a_i = (r_i^2 H_i^{b-1})^{1/b} \asymp i^{-\gamma(b-1)/b}.$$

The sum diverges in the stated range. Choose disjoint finite consecutive blocks I_l of layers with $A_l = \sum_{i \in I_l} a_i \to \infty$. Choose increasing rate boundaries $R_l \to \infty$ large enough that $R_l a_i/(A_l H_i) \ge 1$ for $i \in I_l$, and the uniform regular variation bounds used below hold on the corresponding finite collection of scaled intervals for every $R \ge R_l$. For $R_l \le R < R_{l+1}$, use $n_i(R) = \lceil 8Ra_i/(A_l H_i) \rceil$, $i \in I_l$. Fix $T_i \ge 2n_i(R_{l+1})$ in these layers, using the formula with block I_l at the endpoint. Any remaining layers may have template length one.

For layer i, the state is a word of T_i letters in $\{1, \dots, M_i\}$ and a cyclic position. Keep the word and advance the position by one at each step except that, with probability $\rho_i > 0$, both the word (uniform independently in its letters) and the position (uniform) are replaced independently. Use the stationary product distribution of uniform word and uniform position. The finite-state chain is mixing by the independent resets. Its output is the current letter represented as a point of K_i. Within any stretch of length at most T_i, conditional on the resets and positions, all output letters are independent and uniform: each uninterrupted segment visits distinct slots, and successive words are independent. Use independent chains for different layers, and sum their output vectors. The countable product is mixing in the measure-preserving sense, since cylinder events depending on finitely many layers and times mix and approximate every event. The shift on the Hilbert-valued output is therefore stationary and mixing.

The entropy rate of layer i is at most

$$\hbar_i = H_{\mathrm{bin}}(\rho_i) + \rho_i(T_i H_i + \log T_i),$$

by describing the initial state, reset indicators and new states. For each $a > 0$, encode finitely many layers using either the exact sequence of their output letters or zero, and ignore the remaining layers. Distortions add by orthogonality, and independent layer channels allow the information costs to add. Taking block length to infinity and then increasing the number of considered layers gives

$$F(a) \le \sum_i \min(\hbar_i, a r_i^2).$$

Indeed the never-encoded tail costs at most $a \sum_{i>m} r_i^2$, which tends to zero. Integrating the minima gives

$$J \le \frac{b}{b-1} \sum_i r_i^{2/b} \hbar_i^{1-1/b}.$$

Choose each ρ_i small enough that the sum is arbitrarily small. Add one independent binary iid layer of arbitrarily small positive amplitude, in a new orthogonal dimension. It preserves the entropy

bound and yields $J > 0$: projection onto this layer reduces the objective, and the binary iid objective is positive at every positive penalty by information tensorization and the binary decoding bound. Its contribution to the upper bound on J can be made arbitrarily small.

Fix $R \in [R_l, R_{l+1})$, suppress the argument of $n_i(R)$, and consider the features in layers $i \in I_l$ at times $n_i \le j < 2n_i$. They are jointly independent uniforms on their layer alphabets. For any codeword, each feature is called matched if its projected weighted coordinate is within $\sqrt{\lambda_j} r_i/3$ of the correct letter. The two-letter separation is $\sqrt{2\lambda_j} r_i$, so the probability of matching a particular feature is at most e^{-H_i}. A missed feature contributes at least $\lambda_j r_i^2/9$ to squared distance. By the choice of n_i,

$$\frac{\lambda_j r_i^2}{9} \ge c_b \phi(R) A_l^b H_i, \qquad n_i \le j < 2n_i.$$

In fact $n_i \le 9Ra_i/(A_l H_i)$, $a_i^b = r_i^2 H_i^{b-1}$, and uniform regular variation gives $\lambda_j \ge \frac{1}{2}\phi(R)(j/R)^{-b}$ for the relevant intervals by the choice of R_l. The total entropy weight of all features is $\sum_{i \in I_l} n_i H_i \in [8R, 9R]$. Since $H_i \ge 10$, the number of features is at most R.

A ball of squared radius $c_b R \phi(R) A_l^b$ can therefore meet only samples having match weight at least $7R$. For each subset with that weight, independence bounds its probability by e^{-7R}. There are at most 2^R subsets, so the coverage of at most e^R codewords tends to zero uniformly. As $A_l \to \infty$, the squared radius divided by h_R^2 tends to infinity along all $R = \log N$. This proves the first case of Theorem 4.2.

For $p > 1$ but $p \le -\nu/b$, use the same alphabet and resets with template lengths one. Choose the reset probabilities and binary layer as before to make J positive and as small as desired. At time one, the i-th component is uniform on M_i separated letters. With $N_i = \lfloor e^{H_i/2} \rfloor$, a ball of radius $\sqrt{\lambda_1} r_i/3$ matches at most one such letter. The N_i-ball coverage at that radius is at most N_i/M_i, tending to zero. Since $H_i \asymp 2^{i\beta} i^{-\gamma}$ and $\log H_i \asymp i$,

$$\frac{r_i^2}{h_{\log N_i}^2} \asymp i^{-[\gamma(b-1)+\nu]}.$$

It diverges for $p < -\nu/b$, and is bounded below by a fixed positive number at equality. At equality choose κ smaller than the resulting positive lower bound on the normalized squared radius. Every fixed excess-distortion quantile then exceeds the predicted constant along this sequence. The same coverage bound gives failure under every fixed power loss. This completes the lower bound.

4.4 General entropy envelopes

Replace E by a constant multiple if needed so that it bounds H_K on $(0, 1]$. The quantity $B_K(a)$ is at most twice

$$C_E(a) = \inf_{0 < r \le 1} \max(E(r), ar^2).$$

For $t > E(1)$ let $r_t = \inf\{r : E(r) \le t\}$. If $a < t/r_t^2$, choose an admissible radius arbitrarily close to r_t to see that $C_E(a) \le t$. In the case $r_t = 0$ this conclusion holds for every finite a. By the layer-cake representation and Tonelli,

$$\begin{aligned}\int_0^1 C_E(u^{-b})\, du &\le E(1) + \int_{E(1)}^{\infty} t^{-1/b} r_t^{2/b}\, dt \\ &\le E(1) + C_b \int_0^1 r^{2/b-1} E(r)^{1-1/b}\, dr.\end{aligned}$$

The second inequality writes $r_t^{2/b}$ as the integral of $(2/b)r^{2/b-1}$ on $(0, r_t)$, where $E(r) > t$. For pure power weights $a_j \le C(R/j)^b$. The function $u \mapsto B_K(Cu^{-b})$ is nonincreasing, so its integral on $(0, \alpha)$ dominates the normalized right-endpoint sum for $j \le \alpha R$. This integral tends to zero with α by the preceding bound. Thus (6) applies.

For the converse, put $E_i = E(2^{-i})$ and $p_0 = 1 - 1/b$. Monotonicity shows that divergence of the integral is equivalent to divergence of $\sum_i 2^{-2i/b} E_i^{p_0}$. Write $\Delta_i = E_i - E_{i-1}$. Subadditivity of the power $p_0 \in (0, 1)$ and summation of the geometric factors show that divergence forces $\sum_i 2^{-2i/b} \Delta_i^{p_0} = \infty$. Discarding all increments smaller than 11 preserves divergence. For each retained layer take $\log M_i = \log \lfloor e^{\Delta_i} \rfloor$ and $r_i = c_0 2^{-i}$ with a sufficiently small fixed $c_0 > 0$. For $2^{-(i+1)} \le r < 2^{-i}$, specifying all layers through i leaves a tail of diameter at most r. Its log covering cost is at most $\sum_{j \le i} \Delta_j \le E_i \le E(r)$. An independent binary layer adds only a constant to this budget. The sum $\sum_i (r_i^2 (\log M_i)^{b-1})^{1/b}$ diverges. This is exactly the condition used to choose the blocks and template lengths for the preceding mixing counterexample. That construction therefore gives zero limiting coverage at every fixed multiple of the predicted scale, with an arbitrarily small positive J. This proves Theorem 4.3.

4.5 The excess-distortion constant for independent letters

For iid letters the block objective tensorizes to the one-letter objective, since coordinatewise channels attain the upper bound and mutual information dominates the sum of coordinate informations when the input coordinates are independent. In particular the same applies to any deterministic projection or residual of the letters.

We use the random coding method of Kostina and Verdú [18, Theorem 10] in the following form. For any coupled source and reproduction (U, V) with mutual information $I < \infty$, let $\iota(U, V) = \log(dP_{UV}/d(P_U P_V))$. Under the joint law the negative tail satisfies $\mathbb{P}(\iota < -x) \le e^{-x}$ for $x > 0$ by change of measure, and thus $\mathbb{E}(\iota_+) \le I + 1$. Given a desired distortion threshold, call a pair good if its distortion is within the threshold and $\iota \le B/2$. The set of source values with conditional good-pair probability less than $1/2$ has probability at most twice the joint bad-pair probability. For every other source value, an independent reproduction draw from P_V has success probability at least $\frac{1}{2} e^{-B/2}$ by change of measure. Drawing $\lfloor e^B \rfloor$ independent codepoints therefore gives some deterministic codebook with failure probability at most twice the original joint bad-pair probability plus $o(1)$ as $B \to \infty$. The term from information is bounded by $2(I+1)/B$ by Markov's inequality.

Take nested finite-dimensional orthogonal projections P_l tending strongly to the identity and write $W_l = (I - P_l)X_1$ for the residual letter. By the projection comparisons, the projected process constants increase to κ, using the vanishing mean-square residual in the block comparison. The residual objective $F_l^{\mathrm{res}}(a)$ is bounded by $F(a)$ by projection, and also by $a\mathbb{E}\|W_l\|^2$. Thus $J_l^{\mathrm{res}} \to 0$ by dominated convergence. The projected stationary laws may have unbounded coefficients, but the finite-dimensional second-moment excess-distortion theorem of Section 3 applies to them.

To encode the residual at log budget B, use independent near-optimal letter channels for $j \le \lfloor B \rfloor$ with penalties $a_j = \lambda_j / \phi(B)$. Use zero for the rest. The sum of the information and the mean squared head distortion divided by $\phi(B)$ is at most $1 + \sum_{j \le B} F_l^{\mathrm{res}}(a_j)$. Under the weight hypothesis of Theorem 4.4, $a_j \le C(B/j)^b$. The same right-endpoint integral bound gives a total of at most $1 + CBJ_l^{\mathrm{res}}$. The squared tail distortion has expectation at most $Ch_B^2 \mathbb{E}\|W_l\|^2$. Apply the covering bound to these coupled weighted vectors. For every $\delta > 0$, the residual admits a codebook of log size at most B with

$$\limsup_{B \to \infty} \mathbb{P}\{\text{residual squared error} > \delta h_B^2\} \le C J_l^{\mathrm{res}} + C\delta^{-1}(J_l^{\mathrm{res}} + \mathbb{E}\|W_l\|^2).$$

For fixed δ this tends to zero as $l \to \infty$. Combine with the projected codebook by adding codepoints, reserving a fraction $\varsigma > 0$ of the logarithmic budget for the residual. With ς fixed, first make the desired residual threshold small and then choose l large to make its failure probability small. The projected constant is at most κ, and its budget contributes the factor $(1-\varsigma)^{(1-b)/2}$ in normalized root error. For fixed ς the corresponding factor for the residual scale is $\varsigma^{(1-b)/2}$. Sending $\varsigma \downarrow 0$ proves coverage tending to one at every squared threshold strictly above κ. Projection of any competitor followed by the finite-dimensional strong converse gives coverage tending to zero at thresholds below κ, using $l \to \infty$. This proves the positive assertion.

If the weight bound fails, choose $R_i \to \infty$ and integers $1 \le m_i \le R_i$ with $\ell(m_i)/\ell(R_i) \ge 2^{2bi}$. Uniform slow variation gives $R_i/m_i \to \infty$. Use an orthogonal product of alphabets of size M_i, with $H_i = \log M_i = \log\lceil \exp(8R_i/m_i)\rceil$ and $r_i^2 = 2^{-ib} H_i^{1-b}$. The squared radii are summable and the resulting alphabet is compact. Independently in each layer and at each time, take a uniform letter from that layer. The exact-or-zero channels on finitely many layers give

$$J \le \frac{b}{b-1} \sum_i r_i^{2/b} H_i^{1-1/b} = \frac{b}{b-1} \sum_i 2^{-i} < \infty,$$

and one nontrivial layer gives positivity by projection. At $N_i = \lfloor e^{R_i} \rfloor$, test the m_i independent letters of layer i at times $m_i \le j < 2m_i$. On that interval λ_j is comparable to λ_{m_i} by regular variation and positivity of the finitely many initial weights. A ball of squared radius $c\, m_i \lambda_{m_i} r_i^2$, with a small fixed $c > 0$, must match at least half of those letters in the sense used in the mixing construction. The union over subsets and centers bounds the coverage by $\exp(R_i) 2^{m_i} \exp(-m_i H_i/2)$, tending to zero. The radius squared divided by $h^2_{\log N_i}$ has order $2^{-ib}\ell(m_i)/\ell(R_i)$, which diverges. This completes Theorem 4.4.

5 The sharp loss-moment threshold for independent Hilbert coefficients

We determine which loss moments are controlled by a finite one-letter rate-distortion integral for regularly weighted independent Hilbert coefficients. When the slowly varying part of the squared weights is uniformly almost nondecreasing, every moment of order $0 < s < 2/b$ has the same exact constant as the excess-distortion threshold, where $-b < -1$ is the weight index. For the weights $j^{-b}[\log(ej)]^{\nu}$ with $\nu \ge 0$ the boundary $2/b$ is sharp. One compact independent-letter example with positive finite rate-distortion integral has diverging normalized errors along a subsequence at every moment order $s \ge 2/b$.

5.1 The moment threshold

Let the coefficients X_j be iid in a separable real Hilbert space with $\mathbb{E}\|X_1\|^2 < \infty$. Assume $\lambda_j > 0$ and $\lambda_j \sim \phi(j) = j^{-b}\ell(j)$, with $b > 1$ and ℓ positive and slowly varying. Put $Y = (\sqrt{\lambda_j} X_j)_{j \ge 1}$ and $h_R = \sqrt{R\phi(R)}$. For positive finite s, let $e_{N,s}$ be the infimum of $(\mathbb{E}\,\mathrm{dist}(Y, \mathcal{C})^s)^{1/s}$ over all codebooks $|\mathcal{C}| \le N$. Set

$$F(a) = \inf_V \{ I(X_1; V) + a\, \mathbb{E}\|X_1 - V\|^2 \}, \qquad J = \int_0^\infty F(u^{-b})\, du, \qquad \kappa = \frac{(b-1)^{b-1}}{b^b} J^b.$$

Information is measured in nats. Independence makes $F(a)$ the corresponding per-letter objective for every block length.

Theorem 5.1. *Suppose $J < \infty$ and*

$$\sup_{1 \le j \le R,\ j \text{ integer}} \ell(j)/\ell(R) = O(1). \tag{7}$$

Then for every $0 < s < 2/b$,

$$\frac{e_{N,s}(Y)^2}{h^2_{\log N}} \longrightarrow \kappa.$$

If (7) fails, there is a compact iid alphabet with $0 < J < \infty$ for which this limit fails for every $s > 0$.

Theorem 5.2. *For every $b > 1$ and every $\nu \in \mathbb{R}$, with $\lambda_j \sim j^{-b}[\log(ej)]^{\nu}$, there is a compact iid Hilbert alphabet with $0 < J < \infty$ such that $e_{N,s}/h_{\log N}$ diverges along a common subsequence for every $s \ge 2/b$.*

The theorem stated in Section 2 gives the weighted rate-distortion formula over all finite loss moments for compact finite-dimensional alphabets, and Section 3 permits finite second moments for loss powers below two in Euclidean alphabets. In a general Hilbert alphabet, a finite J is enough for the exact excess-distortion threshold precisely under condition (7), uniformly over iid laws, by Theorem 4.4. Theorems 5.1 and 5.2 distinguish this probability limit from mean convergence without a metric entropy bound. Kostina and Verdú [18, Theorems 1 and 10] provide the general fixed-rate random covering framework. Here the one-sided weighting allows a normalized tail of order $t^{-1/b}$ in squared distance, and the endpoint construction uses first occurrences at many time scales.

5.2 A random covering bound with no additive information term

Let P be the product of the source and reproduction marginals and Q their joint law. Assume $I = D(Q\|P) < \infty$, and write $\iota = \log(dQ/dP)$. For $u > 1$, let $A = \{\iota > u\}$ and $q = Q(A)$. Then $P(A) \le e^{-u}q$. The binary relative entropy bound gives

$$I \ge qu + (1-q)\log(1-q) \ge q(u-1), \qquad \text{so} \qquad Q\{\iota > u\} \le I/(u-1).$$

The endpoints $q = 0, 1$ have the same bound by continuity. This keeps the information-tail estimate proportional to I even at very small I.

For a specified distortion threshold, call a coupled pair good if it meets the threshold and $\iota \le B/2$. Source points with conditional good-pair probability less than $1/2$ have probability at most twice the joint bad-pair probability. At every other source point, the reproduction marginal gives the good set mass at least $\frac{1}{2}e^{-B/2}$ by change of measure. For an iid codebook of $\lfloor e^B \rfloor$ marginal draws, the expected uncovered mass is therefore at most

$$2Q\{\text{coupled distortion exceeds the threshold}\} + \frac{2I}{B/2-1} + \exp(-ce^{B/2})$$

for $B > 2$ sufficiently large. Consequently some deterministic codebook of this size attains the bound.

5.3 Encoding residuals at several distortion levels

Let W be any iid residual letter, with second moment $M = \mathbb{E}\|W\|^2$ and one-letter integral $J_W < \infty$. At target squared radius th_B^2 for $t > 0$, use product channels through $j \le \lfloor B \rfloor$ with penalties $a_j = \lambda_j/[t\phi(B)]$, and use zero afterwards. By (7), $a_j \le C(B/j)^b/t$. The normalized right Riemann sum of the objectives is at most its monotone integral:

$$\sum_{j \le B} F_W(a_j) \le B \int_0^1 F_W(Cu^{-b}/t)\, du$$
$$\le CBt^{-1/b} J_W.$$

The last step substitutes a rescaled u and integrates on the whole positive half-line. Choose the total near-minimization tolerance at most e^{-B}. The total information and the expected squared head distortion divided by $t\phi(B)$ are jointly bounded by the last display plus this tolerance. The zero tail adds mean squared distortion at most CMh_B^2 by regular variation. The covering bound, applied to the weighted vectors, gives a code with log size at most B and probability of squared error above th_B^2 at most

$$C(t^{-1/b} J_W + t^{-1} M) + O(e^{-B}/B) + \exp(-ce^{B/2}). \tag{8}$$

The constants are uniform in t and the residual law.

For fixed $0 < s < 2/b$, use a code union over the dyadic thresholds $1 \le t \le B^L$, with $L > 0$ large and fixed, and also at a fixed $0 < \delta < 1$. Include zero. Give each profile logarithmic budget $B - O(\log\log B)$, so the union has at most e^B points. This changes h_B by a factor tending to one. By taking a fixed factor in the thresholds to account for dyadic rounding and budget rescaling, (8) bounds its squared-error tail through B^L; the two exponentially small terms remain negligible on integration over this range. Above B^L, zero reproduction bounds the normalized squared first moment by CM/h_B^2. For $p = s/2 < 1/b < 1$, that part of the p-moment is at most

$$CM(B^L)^{p-1}/h_B^2 = o(1)$$

by choosing L large enough, using the slow variation of ℓ. Between one and B^L, the tail bound integrates against $p\, t^{p-1}$, because $p < 1/b$. Between a fixed constant multiple of δ and one, use the bound from the δ profile. Below that level use probability at most one. We obtain a residual code sequence with normalized s-th error moment of limit superior at most

$$C_s\big(\delta^{s/2} + (\delta^{-1} + \delta^{-1/b})(J_W + M)\big). \tag{9}$$

Let P_l be nested finite-dimensional orthogonal projections tending strongly to the identity. The projected constants increase to κ by the projection comparison in Section 4. For $W_l = (I - P_l)X_1$, the residual second moment M_l tends to zero. The residual objective is at most $F(a)$, by projection, and at most aM_l by zero reproduction. Dominated convergence gives $J_{W_l} \to 0$. Thus (9) can be made arbitrarily small for the projected residual, by first choosing δ small and then l large.

Reserve a fraction $\varsigma > 0$ of the log budget for the residual, and add its codepoints to the codebook for $P_l X$ at fraction $1 - \varsigma$. At each fixed l, the projected law is finite-dimensional with finite second moment. Its exact moment limit for $1 \le s < 2$ follows from Section 3; for $0 < s < 1$, Jensen gives the upper limit from $s = 1$ and the excess-distortion converse gives the lower limit. For $s \ge 1$ use Minkowski for the two reconstruction errors; for $0 < s < 1$ use subadditivity of the s-th power. With ς fixed, the residual scale factor $\varsigma^{(1-b)/2}$ is fixed. Choose l and δ to make the residual negligible and send $\varsigma \downarrow 0$. This gives the upper bound in Theorem 5.1. Its moment lower

bound follows from the excess-distortion strong converse of Section 4. If (7) fails, the compact iid counterexample in Section 4 has vanishing coverage at every fixed multiple of the predicted scale along a subsequence. Hence all positive moments diverge on that scale there.

5.4 Rare letters and the critical moment

For the counterexample fix $r_i = 2^{-i}$ and $p_i = \exp(-e^{2i})$, $i \geq 1$. In mutually orthogonal finite-dimensional spaces take M_i orthogonal letters of norm r_i. Choose $H_i = \log M_i$ so that

$$C_i = r_i^2 H_i^{b-1} \asymp p_i^{-b} e^{-bi}.$$

Such integer M_i exist by rounding the exponential of the prescribed H_i. The i-th layer is zero with probability $1 - p_i$, and otherwise uniform on its M_i letters. All times and layers are independent. The alphabet of all allowed sums is compact by summability of r_i^2.

The one-letter objective is bounded by

$$F(a) \leq \sum_i \min(H_{\mathrm{bin}}(p_i) + p_i H_i,\, a p_i r_i^2).$$

This follows by exact-or-zero encoding of finitely many layers and taking the deterministic zero tail. The integrated bound is at most $C_b \sum_i p_i (r_i^2 H_i^{b-1})^{1/b} < \infty$: indeed $H_{\mathrm{bin}}(p_i) \leq C p_i H_i$ for large i, and integrating each minimum gives $b/(b-1)$ times the corresponding weighted geometric mean. The sum is finite because $p_i C_i^{1/b} \asymp e^{-i}$. One nontrivial finite layer gives $J > 0$ by projecting onto it and applying the binary decoding bound at sufficiently small distortion.

Put $N_i = \lfloor e^{H_i/2} \rfloor$. Let T_i be the time of the first nonzero letter of layer i. Conditional on $T_i = j$, the active letter is uniform among M_i separated values. Any codebook of size at most N_i misses this coordinate at radius $\sqrt{\lambda_j} r_i/3$ with conditional probability at least $1 - N_i/M_i$. The events $T_i = j$ are disjoint and have probability at least $c p_i$ for $j \leq 1/p_i$. With $s = 2/b$, summing those lower bounds for $p_i^{-1/2} \leq j \leq p_i^{-1}$ gives

$$e_{N_i, 2/b}^{2/b} \geq c p_i r_i^{2/b} [\log(1/p_i)]^{1+\nu/b}.$$

Indeed $\lambda_j^{1/b} \asymp j^{-1} [\log(ej)]^{\nu/b}$ on that interval, and the sum of $1/j$ there has order $\log(1/p_i)$. Also $\log H_i \asymp \log(1/p_i)$ by its definition. After division by $h_{\log N_i}^{2/b}$ the bound is of order at least

$$p_i C_i^{1/b} \log(1/p_i) \asymp e^i,$$

which diverges. For every $s \geq 2/b$ the same holds for the normalized L^s distance by moment monotonicity. This proves Theorem 5.2.

6 Weighted Toeplitz eigenvalues and Gaussian small deviations

We prove a Weyl law on $\mathbb{Z}^q$ for Toeplitz covariance operators with regularly varying, direction-dependent diagonal weights and arbitrary finite positive matrix spectral measures. The constant uses only the absolutely continuous matrix density, which may be unbounded. This is the lattice–Toeplitz counterpart of the classical supercritical theorem of Birman and Solomyak, in which singular parts of the measure do not enter the leading term. We give an explicit atomless singular example in which the lower-order contribution decays arbitrarily slowly. Finite Toeplitz sections, a low-rank cutoff and trace estimates give the result. Applications include exact leading constants for weighted Hilbert small deviations and mean-square functional quantization of Gaussian fields.

6.1 The spectral law

Let $\mathfrak{m}(d\omega) = d\omega/(2\pi)^q$ be normalized Haar measure on the q-torus, $q \geq 1$. Let μ be a finite positive $d \times d$ matrix-valued measure with absolutely continuous part $F\, d\mathfrak{m}$. Here positive means that $\mu(E)$ is positive semidefinite for every Borel set E. Let $\phi > 0$ be regularly varying of index $-b$, $b > q$, and let the squared weights $\lambda_j \geq 0$ on $\mathbb{Z}^q$ satisfy

$$\frac{\lambda_j}{\phi(|j|)} - c(j/|j|) \longrightarrow 0 \qquad (|j| \to \infty),$$

where $c \geq 0$ is continuous on the unit sphere, and λ_0 is arbitrary and finite. The norm is Euclidean. This includes one-sided weights for $q = 1$ by setting them to zero on the unused side. Put $\beta = b/q > 1$, and set

$$A_{jk} = \sqrt{\lambda_j \lambda_k} \int e^{i(k-j)\cdot\omega}\, d\mu(\omega), \qquad S = \frac{1}{q} \int_{\mathbb{S}^{q-1}} c(\xi)^{1/\beta}\, d\xi \int_{\mathbb{T}^q} \operatorname{tr} F(\omega)^{1/\beta}\, d\mathfrak{m}.$$

For $q = 1$, sphere measure is counting measure on $\{-1, 1\}$. The operator A on $\ell^2(\mathbb{Z}^q; \mathbb{C}^d)$ is positive trace class: its finite compressions are positive Gram matrices and their traces are bounded by $(\sum_j \lambda_j) \operatorname{tr} \mu(\mathbb{T}^q)$. Write $N_A(x)$ for the number of eigenvalues strictly above $x > 0$.

Theorem 6.1. *For every $t > 0$,*

$$\lim_{L \to \infty} \frac{N_A(t\phi(L))}{L^q} = S\, t^{-1/\beta}.$$

Consequently, if the eigenvalues (σ_n) are in decreasing order with zeros appended when needed, then

$$\frac{\sigma_n}{\phi(n^{1/q})} \longrightarrow S^\beta.$$

In particular, the singular part of the matrix measure has zero leading coefficient.

That the singular part of the measure does not enter the leading term is a classical phenomenon for high-order spectral problems. M. G. Krein discovered it in 1951 for the singular string, and Birman and Solomyak [6] and Borzov [7] proved it for a wide class of high-order spectral problems, including $(-\Delta)^l u = \lambda P u$ in a domain of $\mathbb{R}^n$ with $2l > n$ and a measure P; see the account in the introduction of Rozenblum and Shargorodsky [8]. The nonzero spectrum of A coincides with that of the operator on the q-torus formed from the measure μ and the Fourier multiplier with symbol $\sqrt{\lambda_k}$, a multiplier of order $-b/2$ up to slowly varying factors, with $b > q$. Thus Theorem 6.1 is the lattice–Toeplitz counterpart of the supercritical Birman–Solomyak theorem. Its proof is by finite Toeplitz sections and trace approximation, and it covers matrix-valued spectral measures on $\mathbb{Z}^q$, regularly varying and direction-dependent diagonal weights, and unbounded matrix densities. The sharpness paragraph in Section 6.2 gives an explicit atomless singular measure for which the lower-order contribution decays arbitrarily slowly.

For scalar pure power weights in one dimension, Hong, Lifshits and Nazarov [19, Theorem 1.1] proved a small-deviation asymptotic, with the covariance eigenvalue law in equation (2.7) of the proof of their Theorem 1.1, under an extra filter summability assumption for amplitude powers $1/2 < p < 1$; whether this assumption is needed is asked in Remark 1.3 of the arXiv version of [19], cited as Remark 1.2 in [20]. Lifshits and Nazarov [20, Theorem 3.1 and the following discussion] removed that assumption using their spectral lemma for integrable scalar symbols. Our proof recovers their scalar asymptotic and extends it to multidimensional matrix symbols, finite

matrix spectral measures with singular parts, and slowly varying modifiers of the diagonal weights. On compact rectangles the absolutely continuous finite-section distribution also follows from the multilevel block generalized locally Toeplitz algebra of Barbarino, Garoni and Serra-Capizzano, using properties GLT 1, GLT 3 and GLT 4 in the summary of [22, Section 5]. For unweighted scalar finite sections the Radon-measure distribution theorem is due to Tyrtyshnikov and Zamarashkin [15, Theorem 1.1]. Ponge and Tian [21] develop a general perturbation theory for spectral asymptotics with regularly varying scales.

6.2 Finite sections and trace cutoffs

Let K_μ send a finitely supported vector x to $\sum_j \sqrt{\lambda_j} e^{ij\cdot\omega} x_j$ in the Hilbert space with squared norm $\int p^*\, d\mu\, p$. After quotienting zero-norm functions and completing, K_μ is Hilbert–Schmidt and $A = K_\mu^* K_\mu$. Its columns have squared norms summing over coordinate j to $\lambda_j \operatorname{tr}\mu(\mathbb{T}^q)$.

We use the singular-value counting inequality

$$n(s_1 + s_2; B + C) \le n(s_1; B) + n(s_2; C), \qquad s_1, s_2 > 0,$$

where $n(s; B)$ counts singular values strictly above s. It follows from the minimax characterization by intersecting subspaces on which each operator has norm bounded by its threshold. A finite-rank perturbation changes this count by at most its rank at the same threshold, by intersecting with its kernel and applying the argument in both directions. The analogous inequality for eigenvalues of a sum of positive compact operators follows from the quadratic-form minimax principle.

For fixed $0 < \alpha < T$, keep only columns with $\alpha L < |j| \le TL$. The omitted head has rank $O(d\alpha^q L^q)$ asymptotically. The squared Hilbert–Schmidt norm of the omitted tail is bounded by $C \operatorname{tr}\mu(\mathbb{T}^q) T^{q-b} L^q \phi(L)$ as $L \to \infty$. Indeed, lattice annuli have $O(r^q)$ points at scale r, and uniform regular variation and summation on dyadic shells bound $\sum_{|j|>TL} \lambda_j$ at this order [14]. The tail estimate and counting inequality give opposite one-sided bounds with thresholds $(1-\epsilon)\sqrt{t\phi(L)}$ and $(1+\epsilon)\sqrt{t\phi(L)}$ on the middle operator. Their errors in the count divided by L^q are at most $C[\alpha^q + \epsilon^{-2} T^{q-b} \operatorname{tr}\mu(\mathbb{T}^q)/t]$. This allows $\alpha \downarrow 0$, $T \uparrow \infty$ and then $\epsilon \downarrow 0$ when the limiting count is continuous in t.

Trigonometric polynomial matrix density. Suppose first that F is a positive semidefinite trigonometric polynomial with Fourier support $|k|_\infty \le m_0$. The weighted matrix is banded. In the spatial variable $u = j/L$, approximate the fixed annulus $\alpha < |u| \le T$ by a finite grid of axis-parallel cubes wholly inside it. The uncovered boundary cells have total volume tending to zero with the mesh, so deleting all their lattice indices changes the middle count divided by L^q by an amount tending to zero. There are $O(L^{q-1})$ indices within distance m_0 of the boundaries of each fixed grid cube. Erasing cross-cube entries therefore changes the middle count by $o(L^q)$ for every fixed grid.

On the retained annulus the weights divided by $\phi(L)$ converge uniformly to the continuous profile $g(u) = |u|^{-b} c(u/|u|)$. In a cube their lower and upper bounds converge to the infimum and supremum of g there, respectively. For its unweighted covariance matrix T_0 and weight diagonal D, the eigenvalues of DT_0D are those of $T_0^{1/2} D^2 T_0^{1/2}$; order bounds on D^2 compare them to constant-weight multiples of T_0. The unweighted section on a lattice cube of n indices, divided in count by n, has eigenvalue distribution $\int \sum_{a=1}^{d} \delta_{f_a(\omega)}\, d\mathfrak{m}$, where f_a are the eigenvalues of F. For each fixed positive integer power k, all rows farther than km_0 from the boundary have the same diagonal block as the whole-lattice convolution power. Thus the corresponding normalized trace tends to $\int \operatorname{tr} F^k\, d\mathfrak{m}$. The zeroth trace is d. Polynomial approximation of continuous test functions on the bounded spectral interval then proves the distribution claim.

These facts give a Riemann integral for the normalized count in the middle. To justify this even for atoms of the unweighted eigenvalue distribution, first sandwich the indicator above t between continuous functions with transition widths $\eta > 0$. For each fixed η, the oscillation of g tends to zero uniformly over the cubes, and the unweighted eigenvalues are bounded. The upper and lower cube comparisons therefore converge to the integrals of the corresponding continuous tests of $g(u)f_a(\omega)$. At $t > 0$ the level set $g(u)f_a(\omega) = t$ has zero product measure: for fixed direction and ω, at most one radius has this value. Letting $\eta \downarrow 0$ gives the limit, continuous at every $t > 0$,

$$\int_{\alpha<|u|\le T} \int_{\mathbb{T}^q} \#\{a : \ g(u)f_a(\omega) > t\}\, d\mathfrak{m}\, du.$$

Apply the head and tail estimates and use polar coordinates and Tonelli. For each direction and eigenvalue the allowed radii range from zero to $(c(\xi)f_a(\omega)/t)^{1/b}$. This proves the claimed formula for polynomial densities.

Integrable matrix density. For a nonnegative integrable scalar g, apply the factorization to $gI_d\, d\mathfrak{m}$. Remove only columns with $|j| \le \alpha L$. The finite rank bound and the Hilbert–Schmidt estimate for the rest give

$$\limsup_{L\to\infty} \frac{N_{A_{gI_d}}(t\phi(L))}{L^q} \le Cd\alpha^q + Cd\, t^{-1}\alpha^{q-b}\|g\|_{L^1}. \tag{10}$$

In particular this bound tends to zero as $\|g\|_{L^1}$ tends to zero, by choosing α proportional to $\|g\|_{L^1}^{1/b}$ for fixed t. For a general integrable positive matrix density, take its Fejér means F_m, positive matrix trigonometric polynomials converging in L^1. With $g_m = \|F - F_m\|_{\mathrm{op}}$ pointwise, both quadratic form inequalities $A_F \le A_{F_m} + A_{g_m I_d}$ and $A_{F_m} \le A_F + A_{g_m I_d}$ hold. Use the eigenvalue counting inequality at a fixed relative threshold slack, let $m \to \infty$ using (10), and then send the slack to zero. The polynomial constants converge because Weyl's finite-dimensional eigenvalue inequality gives $|f_a - f_{m,a}| \le g_m$ and $|x^{1/\beta} - y^{1/\beta}| \le |x - y|^{1/\beta}$ for $x, y \ge 0$. This proves the limit for all integrable densities.

Singular spectral mass. For a singular scalar spectral measure ν, we first show that the unweighted section on a lattice cube of side length m has $o(m^q)$ eigenvalues above any fixed positive threshold. Fix $\eta > 0$. Choose a compact Haar-null set E with $\nu(E^c) \le \eta$, and cover it by a finite union O of open boxes of Haar measure at most η, with positive distance from E to O^c. Project onto the discrete Fourier grid modes whose frequencies lie in O. This projection P has rank at most $(\eta + o(1))m^q$.

The unweighted Toeplitz covariance $T_m(\nu)$ is the integral of the rank-one forms from the exponential vectors $v_m(\omega) = (e^{-ij\cdot\omega})_j$, using the same sign convention for the discrete grid. For $\omega \in E$, any frequency outside O differs in some coordinate by at least a fixed $\delta > 0$ on the circle. In that coordinate the squared coefficient of the vector in the normalized one-dimensional discrete Fourier basis is at most C_δ/m, by the geometric sum formula. Summing over at most m such indices gives C_δ, while summing along each other coordinate gives m, by Parseval. A union over the q possible separated coordinates gives $\|(I - P)v_m(\omega)\|^2 \le C_\delta m^{q-1}$. Outside E use $\|v_m(\omega)\|^2 = m^q$. Hence $\mathrm{tr}((I - P)T_m(\nu)(I - P))/m^q \le \eta + O_\delta(m^{-1})$. Interlacing and the trace bound, followed by $\eta \downarrow 0$, prove the asserted unweighted count. For $q = 1$ it also follows from the Radon-measure Toeplitz theorem [15].

The unweighted covariance on the retained middle indices is a principal compression of a single unweighted cube of side $O(L)$, so the same $o(L^q)$ count holds there. The middle weights are at

most $C_{\alpha,T}\phi(L)$. The weighted eigenvalue comparison above gives a zero normalized count at every scaled threshold. Remove the head and tail as before. For a positive singular matrix measure, domination by its scalar trace measure times I_d gives the same conclusion. Add the absolutely continuous part and use positivity for the lower bound and the counting inequality with threshold slack for the upper bound.

Finally, apply the count limit with $L = n^{1/q}$ and t above and below S^β. This gives the eigenvalue limit, using only the upper comparison if $S = 0$, and proves Theorem 6.1.

Sharpness for singular measures. Assume that c is positive somewhere. Given any sequence $\epsilon_n > 0$ tending to zero, there is an atomless singular scalar spectral probability ν whose covariance eigenvalues satisfy

$$\limsup_{n\to\infty} \frac{\sigma_n}{\epsilon_n \phi(n^{1/q})} = \infty.$$

Choose an increasing sequence m_l with $\epsilon_{m_l^q} \le 2^{-2l}$. Let ρ_m be the uniform probability on the m^q-point torus grid and ν_0 any fixed symmetric atomless singular probability supported on a compact Haar-null set. Take $\nu = \sum_{l\ge 1} 2^{-l}(\rho_{m_l} * \nu_0)$. This is symmetric, atomless, and singular since its mass lies in a countable union of translates of that null set.

For $m = m_l$, the Fourier coefficients of $\rho_m * \nu_0$ vanish at every nonzero k with $|k|_\infty < m$. Its unweighted covariance on any axis-parallel lattice cube of side m is therefore the identity. Choose such a cube at distance comparable to m and within an open cone where c is bounded below by a positive constant. This is possible by placing the cube center at a sufficiently large fixed multiple of m in a direction of positivity. Uniform regular variation gives $\lambda_j \ge C\phi(m)$ on the cube for all sufficiently large m. Positivity of the other terms and the minimax principle then give $\sigma_{m_l^q} \ge C2^{-l}\phi(m_l)$. This proves the claim. In particular, the zero leading coefficient gives no uniform rate of improvement over the absolutely continuous scale even for atomless spectra of real Gaussian fields.

6.3 Gaussian consequences

Let the matrix measure be that of a real centered stationary Gaussian vector field (U_j). The conjugate symmetry of the spectral measure makes the covariance matrix of the weighted sequence real; its real and complex eigenvalue counts agree. Put $Q = \sum_j \lambda_j \|U_j\|^2$, $R = L^q$, and $\widetilde{\phi}(R) = \phi(R^{1/q})$. The latter is regularly varying of index $-\beta$. Write

$$A_\beta = \frac{2^{1/\beta - 1}\pi S}{\sin(\pi/\beta)}.$$

The following limit includes the case $S = 0$.

Corollary 6.2. *For every $v > 0$,*

$$\lim_{R\to\infty} \frac{-\log \mathbb{P}\{Q \le vR\widetilde{\phi}(R)\}}{R} = (\beta - 1)\left(\frac{A_\beta}{\beta}\right)^{\beta/(\beta-1)} v^{-1/(\beta-1)}.$$

In particular, for $\phi(L) = L^{-b}$ and $S > 0$,

$$-\log \mathbb{P}\{Q \le a^2\} \sim (\beta - 1)\left(\frac{A_\beta}{\beta}\right)^{\beta/(\beta-1)} a^{-2/(\beta-1)} \qquad (a \downarrow 0).$$

Proof. By the Gaussian spectral decomposition, Q has the law of $\sum_{j\ge1}\sigma_j Z_j^2$ with independent standard real Gaussians Z_j. Write $\Psi(r)=-\log \mathbb{E}e^{-rQ}$. For each $t>0$, Theorem 6.1 and regular variation give

$$\frac{\Psi(t/\widetilde{\phi}(R))}{R}=\frac{1}{2R}\sum_{j\ge1}\log(1+2t\sigma_j/\widetilde{\phi}(R))\longrightarrow A_\beta t^{1/\beta}.$$

Indeed the sum on a fixed middle $[\alpha R,TR]$ is a Riemann sum with integrand $\frac{1}{2}\log(1+2tS^\beta u^{-\beta})$. Potter bounds give head contribution $O(\alpha(1+|\log\alpha|))$ after division by R and an $o(1)$ contribution from finitely many initial indices. For the tail, $\log(1+x)\le x$ and regularly varying summation give $O(T^{1-\beta})$. The integral over the entire half line is $A_\beta t^{1/\beta}$, using integration by parts and $\int_0^\infty(1+x^\beta)^{-1}dx=\pi/[\beta\sin(\pi/\beta)]$.

The Chernoff bound gives the lower limit on $-\log\mathbb{P}/R$ as the supremum over $t>0$ of $A_\beta t^{1/\beta}-tv$, the displayed value. For the reverse bound, tilt by $e^{-rQ+\Psi(r)}$ with $r=t/\widetilde{\phi}(R)$. Concavity of the normalized Laplace exponent and convergence to a differentiable limit show that its derivatives converge at $t>0$. The tilted mean of $Q/(R\widetilde{\phi}(R))$ therefore tends to $A_\beta t^{1/\beta-1}/\beta$. Its variance tends to zero, because the tilted variance of Q is $2\sum_j(\sigma_j/(1+2r\sigma_j))^2\le\Psi'(r)/r$. For $A_\beta>0$, choose t so that the limiting mean is strictly below v, and use a thin interval around that mean contained in $[0,v]$. On this interval the cost of undoing the tilt approaches, per R, $A_\beta t^{1/\beta}-t(A_\beta t^{1/\beta-1}/\beta)$ as its width tends to zero. Let the chosen limiting mean increase to v. When $A_\beta=0$, the tilted probability of $Q/(R\widetilde{\phi}(R))\le v$ tends to one, and undoing the tilt there costs at most $\Psi(t/\widetilde{\phi}(R))=o(R)$. The pure power expression follows by taking $a^2=R^{1-\beta}$, using positive real R. □

For $q=d=1$, in the notation of [19, 20], take $b=2p$ and $c_\pm=d_\pm^2$. The spectral density of the linear representation is the squared modulus of its Fourier filter. The constant S^b is then their C in equation (1.6) of [19], and the pure power limit in Corollary 6.2 equals their $B_pC^{1/(2p-1)}$. This recovers the scalar extension of [20, Theorem 3.1] under square summability of the filter.

For $S>0$, the same eigenvalue law and the Gaussian covariance theorem of Luschgy and Pagès [13, Theorem 2.2] give the exact squared mean-square functional quantization constant, where $e_{N,2}$ is the least root mean-square distance to a codebook of at most N points:

$$\lim_{N\to\infty}\frac{e_{N,2}^2}{(\log N)\widetilde{\phi}(\log N)}=\frac{2^{1-\beta}\beta^\beta}{\beta-1}S^\beta.$$

For one-sided scalar sequences this agrees with the process rate-distortion calculation in Section 3. The coefficient S thus enters the finite-support approximation and small-deviation limits through one covariance eigenvalue law.

7 Joint slow-switching limits for weighted telegraph quantization

For stationary binary Markov coefficients and squared weights j^{-b}, $b>1$, we determine the quantization constant when the number of codepoints tends to infinity and the flip probability tends to zero simultaneously. The continuous-time telegraph constant governs every fixed excess-distortion probability for every joint sequence. It also governs power loss of order $s<2/(b-1)$. At and above this threshold the moment limit can fail: we give early-flip configurations for which the normalized moment diverges, despite the same excess-distortion limit. If $\log(1/\rho_N)=o(\log N)$, the moment limit holds for every fixed positive order. For the critical moment we also identify the scale of $\log(1/\rho_N)$ separating convergence and divergence.

7.1 Joint limit

Let $X^{(\rho)}$ be the stationary sign chain with $\mathbb{P}(X_{j+1}^{(\rho)} \neq X_j^{(\rho)} \mid X_j^{(\rho)}) = \rho$, $0 < \rho < 1/2$, and put $Y^{(\rho)} = (j^{-b/2} X_j^{(\rho)})_{j \geq 1}$. Let $e_{N,s}$ denote the least L^s distance to an N-point codebook, for any $s > 0$, and let $d_N(\epsilon)$ denote the least squared radius whose N balls cover mass at least $1 - \epsilon$. Set $R = \log N$.

Let Z be the stationary continuous-time sign process that flips at rate one. For $a > 0$, let $F_{\mathrm{tel}}(a)$ be the limit as $T \to \infty$ of

$$\frac{1}{T} \inf_V \left\{ I(Z_{[0,T]}; V) + a\, \mathbb{E} \int_0^T |Z(t) - V(t)|^2 \, dt \right\}$$

over reproductions valued in $[-1, 1]$. Define

$$K_b = \frac{(b-1)^{b-1}}{b^b} \left(\int_0^\infty F_{\mathrm{tel}}(u^{-b}) \, du \right)^b .$$

The process integral is positive and finite by the slow-switching result of Section 2.

Theorem 7.1. *For every sequence $\rho_N \to 0$, and each $0 < \epsilon < 1$,*

$$\frac{d_N(\epsilon; Y^{(\rho_N)})}{\rho_N^{b-1} (\log N)^{1-b}} \longrightarrow K_b.$$

There is a strong coverage threshold: at any fixed smaller normalized squared radius, the maximal covered mass tends to zero, and at any fixed larger radius it tends to one. For every $0 < s < 2/(b-1)$,

$$\frac{e_{N,s}(Y^{(\rho_N)})^2}{\rho_N^{b-1} (\log N)^{1-b}} \longrightarrow K_b.$$

For every $s \geq 2/(b-1)$ there exists a sequence $\rho_N \to 0$ along which the last ratio diverges. If $\log(1/\rho_N) = o(\log N)$, that ratio instead converges to K_b for every fixed $s > 0$.

Theorem 7.2. *Let $\rho_N \to 0$. For the critical power $s = 2/(b-1)$, put*

$$S_N = \log(1/\rho_N), \qquad S_N^* = \frac{N^{1/b}}{(\log N)^{(b-1)/b}}.$$

If $S_N / S_N^* \to 0$, the normalized squared moment in Theorem 7.1 converges to K_b. If $S_N / S_N^* \to \infty$, it diverges. When the ratios stay bounded above and away from zero, the normalized squared moments stay bounded and strictly above K_b in limit inferior. In particular convergence to K_b occurs exactly when $S_N / S_N^* \to 0$.

At fixed switching probability the stationary quantization theorem of Section 2, proved in [1, Section K], gives a constant $\kappa_b(\rho)$ for bounded Markov coefficients with a strong coverage converse. Section 2 takes $\rho \downarrow 0$ after $N \to \infty$ and finds $\kappa_b(\rho)/\rho^{b-1} \to K_b$, by (4). Theorems 7.1 and 7.2 take the two limits simultaneously. Finite-window jump-time quantization is studied by Aurzada, Dereich, Scheutzow and Vormoor [23], among others. Here the physical window relevant to the quantizer grows with R, and the weight is singular near the time origin after rescaling.

For a mesh $\delta > 0$ sample $Z(j\delta)$, $j \geq 1$, and use the squared metric

$$\|x - y\|_{\delta,b}^2 = \delta \sum_{j \geq 1} (j\delta)^{-b} |x_j - y_j|^2.$$

The observations have flip probability $\rho(\delta) = (1 - e^{-2\delta})/2$. The metric is δ^{1-b} times the original squared metric, and $\delta \sim \rho(\delta)$ as $\delta \to 0$. We therefore work in this metric with normalization $W_R = R^{1-b}$, letting the mesh δ tend to zero along the sequence of budgets. In this proof distance is always the square root of the corresponding weighted sum.

7.2 Coding an initial time interval

The following construction separates the singular time origin from the long intervals on which a fixed sampling mesh is accurate.

Lemma 7.3. *For all sufficiently small meshes $\delta > 0$, take a head cutoff $H > 0$, $0 < \tau \le H/2$, a precision parameter $0 < \xi \le \min(1/4, H)$, and a log budget $B \ge 4$. There is a code for the grid points with $j\delta < H$, using at most e^B codepoints, such that the probability of squared head error greater than $z > 0$ is at most*

$$C_b \left[\tau + \frac{H(1 + \log(1/\xi))}{B} + \frac{\xi H^{1-b}}{z} \right].$$

If $H \le \delta$, the empty head can be encoded exactly.

Proof. Assume $H > \delta$. Split the head into dyadic intervals $[\ell, 2\ell)$, with upper endpoints starting at H, stopping at a lower endpoint $\ell_* \in [\max(\tau, \delta/2), 2\max(\tau, \delta/2))$. On the earlier segment $j\delta < \ell_*$, reproduce the first sampled sign throughout, and encode this sign when the segment is nonempty. If $\max(\tau, \delta/2) = \delta/2$, that segment is empty. Otherwise its length is at most 2τ and the probability of an observed transition there is at most 2τ, by the union bound $\rho(\delta) \le \delta$.

Within a retained interval $[\ell, 2\ell)$ group the consecutive sampled indices into cells of $m = \max(1, \lfloor w/\delta \rfloor)$ indices, where $w = \xi(\ell/H)^b$. The final cell can be shorter. The reproduction is constant on each cell with the sign at the first index. Encode the first retained sign and, for each successive pair of cells, the parity of transitions between their starting indices, including when the next cell is in a different band. There is one starting sign for all retained cells, because they form a partition of the consecutive grid indices in the retained head. A parity attached to a cell in this band has probability at most $q = m\delta$ of being odd. Use the lengths of the Bernoulli mass function with parameter q: odd costs $-\log q$ and even costs $-\log(1 - q)$. Here $q \le 1/4$ by the choice of mesh and ξ. The exponential sum of these lengths over all binary strings equals one. The two sign bits cost at most $2\log 2$. Retain only strings within budget B, so their reproductions number at most e^B. If no string fits, use a singleton fallback and count all outcomes as overflow.

The expected parity lengths in the band total at most $C(\ell + q)(1 + \log(1/q))$, by the bound on odd probability and the number of cells. Here $q \ge w/2$ by the cell-size choice, and $q \le C\ell$ since $\delta \le 2\ell$ and $\xi \le H$ implies $w \le \ell$. The expected sum over bands is therefore at most the sum of $C\ell[1 + \log(1/\xi) + b\log(H/\ell)]$, which is $O(H(1 + \log(1/\xi)))$. Markov's inequality bounds budget overflow by $CH(1 + \log(1/\xi))/B$, since the fixed sign cost is less than half the budget.

If $m = 1$, the reproduction on that band is exact. Otherwise the probability of a wrong sign at any index in a cell is at most $m\delta \le w$. The expected squared rounding error in this band is at most $C(\ell + \delta)\ell^{-b}w \le C\xi\ell H^{-b}$. The sum over the retained bands is at most $C\xi H^{1-b}$. Markov's inequality, along with the early and overflow events, proves the lemma. □

7.3 Transferring the middle interval

Fix a small coarse mesh $h > 0$. Couple all grids by the same Z. For fixed $0 < \alpha < T$, retain the complete coarse cells $[kh, (k + 1)h)$ lying inside $[\alpha R, TR]$. Their union is an interval whose endpoints differ from $\alpha R, TR$ by at most $2h$. The fine cell weight

$$M_{k,\delta} = \delta \sum_{j\delta \in [kh, (k+1)h)} (j\delta)^{-b}$$

is $(1+o(1))h(kh)^{-b}$, uniformly in these cells, as $R\to\infty$, $\delta\to 0$ at fixed h,α,T. In fact the grid count differs from h/δ by at most two, and t^{-b} varies by a relative $O(h/(\alpha R))$ on the cell. For a coarse vector, extension by holding its values through each cell accordingly preserves its middle squared norm up to $1+o(1)$. For a fine vector, its $\delta(j\delta)^{-b}$-weighted cell average gives the corresponding coarse vector, with middle norm at most $1+o(1)$ times the fine norm by Jensen's inequality. Codeword coordinates can always be clipped to $[-1,1]$.

The sample process on the fine cells differs from the held coarse signs by middle squared norm of expectation at most

$$C_b h\,\alpha^{1-b}W_R.$$

Indeed the probability of a sign discrepancy inside a cell is bounded by the probability of a jump between the two sampling times, at most h, and the summed weights on the middle are $O(\alpha^{1-b}W_R)$. Thus the same discrepancy bound controls the comparison of weighted fine source averages with the coarse signs. Markov's inequality makes these transfer errors small in probability by taking h small at fixed α.

For the head on either grid, choose H at the left endpoint of the retained interval, so $H/R\to\alpha$. Apply Lemma 7.3 with $\xi=(H/R)^b$, budget ςR, and a small fixed $\tau>0$. For any fixed $z_0>0$, the probability of squared head error above z_0W_R is bounded in limit superior by

$$C_b\left[\tau+\frac{\alpha(1+|\log\alpha|)}{\varsigma}+\frac{\alpha}{z_0}\right].$$

The zero-reproduction squared tail past the right endpoint is at most $C_bT^{1-b}W_R$ deterministically, by integral comparison. All these bounds are uniform for the fine meshes and for small fixed coarse mesh.

Coverage above the constant. Fix a normalized root radius $u>\sqrt{K_b}$. Choose $\varsigma>0$ small enough that $(1-\varsigma)^{(1-b)/2}\sqrt{K_b}<u$, with a strict margin for the other errors. Take T large, α and τ small, and then h small. The normalized root constant for mesh h in the effective metric is

$$K(h)^{1/2}=h^{(1-b)/2}\kappa_b(\rho(h))^{1/2},\qquad K(h)\longrightarrow K_b.$$

At fixed h, the theorem of Section 2 gives a coarse code with log budget $(1-\varsigma)R$ covering mass tending to one at any normalized root radius strictly above $(1-\varsigma)^{(1-b)/2}\sqrt{K(h)}$. Restrict this code to its middle cells and extend them to the fine mesh. Concatenate the fine head code of log size at most ςR and use zero for the tail. The cardinality is at most e^R. The middle transfer, head and tail estimates make the fine normalized root error at most u with probability at least $1-\varepsilon$ in limit inferior, for arbitrarily small $\varepsilon>0$: the parameters just selected can make the total probability of transfer and head failure as small as desired within the strict radius margin. This proves the upper coverage limit.

Coverage below the constant. Fix $u<\sqrt{K_b}$, and suppose fine codebooks cover probability at least $\varepsilon>0$ along a subsequence at radius $u\sqrt{W_R}$. Choose $\varsigma>0$ small enough that $(1+\varsigma)^{(1-b)/2}\sqrt{K_b}>u$, again with margin. Choose T,α,τ,h in the same order. Transfer the fine middle code to the coarse mesh by weighted cell averaging of each codeword. Append the coarse head code of log size at most ςR and its zero tail. Whenever the fine code succeeds and the head and transfer errors are small, the coarse code succeeds within any fixed radius chosen strictly between u and $(1+\varsigma)^{(1-b)/2}\sqrt{K(h)}$ after scaling by $\sqrt{W_R}$. The radius margin absorbs the $1+o(1)$ comparison and the other norm errors. The bad-event probabilities can be made less than $\varepsilon/2$. The resulting coarse code has log size at most $(1+\varsigma)R$, contradicting its strong converse. This proves the sharp coverage limit.

7.4 Moment convergence below the early-flip threshold

Write $a_* = 1/(b-1)$. We construct extra codebooks whose normalized errors have uniformly bounded s'-th moments for every $s' < 2a_*$. Let $t \geq 1$ be a normalized squared threshold and put $u = t^{-a_*}$. Use head length $H = Ru$, early cutoff $\tau = u$, and precision $\xi = \min(u/4, 1/4)$. These choices meet the head lemma conditions for large R, and $\xi \leq H$. Use a log budget at least $R/2$ for the head and put the tail to zero. The squared tail cost is at most $C_b H^{1-b} = C_b t W_R$. This bound also holds if $H \leq \delta$ and the head is empty, since $\delta^{1-b} \leq H^{1-b}$. The lemma and Markov's inequality give a code with

$$\mathbb{P}\{\text{squared error} > C_b t W_R\} \leq C_b u(1 + |\log u|).$$

Here the budget overrun term is $O(u(1 + |\log u|))$ and the rounding failure at squared radius $O(H^{1-b})$ is $O(u)$.

Take the union of such codebooks for dyadic $1 \leq t \leq R^M$, for any fixed $M > 0$, and include the two full constant-sign codewords. All dyadic codes can be given log budget $R - O(\log\log R)$, still at least $R/2$, while the number of codewords in the union stays below $N/2$ for large R. For $1 \leq t \leq R^M$, adjusting constants for dyadic rounding, its squared-error tail is $O(t^{-a_*}(1 + \log t))$. The two constant-sign codewords also give a tail bound $C_b R t^{-a_*}$ for all $t > 0$. Indeed the error to the starting sign is bounded by $C_b(j\delta)^{1-b}$ when the first flip occurs at grid index j, and the probability of a flip before physical time w is at most w. This estimate applies uniformly in the mesh.

Integrating the squared-error tail to the power $s'/2 < a_*$ gives bounded normalized $L^{s'}$ error on $[0, R^M]$. The integral beyond R^M is $O(R^{1+M(s'/2-a_*)})$, bounded for sufficiently large fixed M. Join this union to a coverage-achieving codebook of at most $N/2$ points, chosen by a diagonal argument from the proved upper coverage limit. Using $N/2$ in place of N leaves the normalization unchanged asymptotically. The combined code has normalized squared error converging in probability to K_b, by the strong converse and the coverage upper bound. For every $s < s' < 2a_*$, the s'-moment bound gives uniform integrability at power s. This proves the upper moment limit. The strong coverage lower bound supplies the moment lower limit for every competing sequence.

7.5 All fixed moments for subpolynomial switching

The same head code has an exponential overflow bound when its time length is well below the log budget. The parity bits use disjoint sets of Markov transitions, so they are independent. A bit of odd probability at most q with length L_q as in Lemma 7.3 satisfies

$$\mathbb{E}e^{\theta L_q} \leq q^{1-\theta} + (1-q)^{-\theta} \leq \exp(C_\theta q^{1-\theta}) \qquad (0 < \theta < 1).$$

In a band with lower edge ℓ there are $O(\ell/q + 1)$ bits, $q \leq C\ell$, and $q \geq \xi(\ell/H)^b/2$. For $0 < \theta < 1/b$, summing the log moment bound over bands therefore gives at most $C_{b,\theta} H \xi^{-\theta}$, since $\sum_\ell \ell^{1-b\theta} = O(H^{1-b\theta})$. The fixed two sign bits only add a constant to the log moment bound.

Fix a desired moment order $s' > 0$ and take $c > \max(1, (b-1)s'/2)$ and $0 < \theta < \min(1/b, 1/c)$. At normalized squared threshold t, put $u = t^{-1/(b-1)}$, $H = Ru$, $\tau = u^c$, and $\xi = u^c/4$. For large R these satisfy the head lemma conditions. Since $H\xi^{-\theta} = O(Ru^{1-c\theta})$, at all t above a sufficiently large fixed constant the overflow probability at log budget $B \geq R/2$ is at most $Ce^{-\theta R/4}$ by the exponential bound. The early error and head rounding failure at squared threshold $C_b t W_R$ each have probability $O(u^c)$, and the zero tail is deterministic as before. The same tail conclusion is immediate when the head is empty.

Use a union of the resulting dyadic codes up to a fixed multiple of the deterministic maximum normalized squared error $C_b(R/\delta)^{b-1}$ of the zero codeword, and include zero. There are $O(1 +$

$\log(R/\delta)$) dyadic levels. Under $\log(1/\delta) = o(R)$ the per-level budget can be kept at least $R/2$ and the union size at most $N/2$. The polynomial term integrates at power $s'/2$ by the choice of c. The exponential term contributes at most a constant times $e^{-\theta R/4}(R/\delta)^{(b-1)s'/2} = o(1)$. Thus every fixed s'-th normalized moment is bounded. As in the preceding section, combine this union with a coverage-achieving codebook of size at most $N/2$. For any $s < s'$, uniform integrability gives the exact moment limit. Since s' can be arbitrarily large and $\delta \sim \rho$, this proves the additional claim.

7.6 Failure at and above the critical moment

Return to the original grid metric $\sum_j j^{-b}|x_j - y_j|^2$. Put $\alpha_* = (b-1)s/2$.

If $\alpha_* > 1$, take $\rho_N = N^{-c}$ with

$$c > \frac{bs/2 - 1}{\alpha_* - 1} > 1.$$

Consider paths with initial sign $+1$ and exactly one flip before time index $3N+1$, at one of the $2N$ consecutive transitions with indices in $[N, 3N)$. Each such event has probability bounded below by $c_1\rho_N$, since $N\rho_N \to 0$. Paths from distinct events differ in at least one coordinate of index at most $3N$, giving distance at least $c_2 N^{-b/2}$. For radius $c_3 N^{-b/2}$ with c_3 small, one codepoint can hit at most one event. Thus any N-point code incurs s-th distortion moment at least $c_4\rho_N N^{1-bs/2}$. Relative to the predicted s-th power scale $\rho_N^{\alpha_*}(\log N)^{-\alpha_*}$, this diverges under the stated choice of c.

7.7 The critical switching scale

Lower bound. Assume $S_N/\log N \to \infty$. Set $\tau_0 = e^{-\zeta R}$ with a fixed $\zeta > 0$. Use geometric shells $L = 4^i$ with $CN^2 \le L \le \tau_0/(10\rho_N)$. There are $m \asymp S_N$ such shells. For each transition index $k \in [L, 2L)$, require initial sign $+1$ and exactly one flip before index $3L+1$, at transition k. Each of these disjoint events has probability at least $c\rho_N$. For $L < L'$, paths in the earlier shell have sign -1 and paths in the later shell sign $+1$ on the index interval between $2L$ and $3L$. Thus a codepoint can lie within distance $c_0 L^{(1-b)/2}$ of paths in at most one shell when c_0 is sufficiently small. Let n_L count these codepoints for shell L. Then $\sum_L n_L \le N$.

Within a shell, paths at distinct transition indices k, k' have squared distance at least $cL^{-b}|k-k'|$. At squared radius $c_1 L^{1-b}/(1+n_L)$, one of the n_L codepoints can therefore hit at most $Cc_1 L/(1+n_L)+1$ transition indices. Choose c_1 sufficiently small, also within the larger shell radius. Since $L \gg N$, at least half the indices are missed. At $s = 2/(b-1)$, they contribute at least $c\rho_N(1+n_L)^{-1/(b-1)}$ to the s-th moment. Summing over shells and applying convexity gives

$$\frac{e_{N,s}^s}{\rho_N/R} \ge cRS_N(1 + N/S_N)^{-1/(b-1)}.$$

This diverges if $S_N/S_N^* \to \infty$. For example, it applies to $\rho_N = e^{-N^2}$, and therefore also proves the critical failure assertion of Theorem 7.1. All events used in this lower bound lie in a set of probability $O(\tau_0)$, since they require a first flip by that order of physical time. On the complement, the strong coverage converse forces the contribution to the normalized s-th moment to have limit inferior at least $K_b^{s/2}$, for every competing sequence. Thus if S_N/S_N^* is bounded away from zero on a subsequence, the earlier shell contribution gives a strict positive gap above the predicted moment on that subsequence.

Upper bound. Use the effective metric at mesh δ and write $\tau_0 = e^{-\zeta R}$ with a sufficiently small fixed $\zeta > 0$. Reserve $N/4$ codepoints for paths whose first grid transition occurs by physical time τ_0. If there are no such transition indices this part is unnecessary. Otherwise their dyadic shells number $O(1 + S_N)$. For each shell use an equally spaced quantization of the transition index with at least $cN/(1+S_N)$ representatives, or encode all indices if a shell is smaller. The budget includes both initial signs. Under $S_N = o(S_N^*)$ the number allocated per shell tends to infinity. Decode a path with that first transition and no later transitions. The index quantization contributes at most $C(\delta L)^{1-b}/k$ squared error on a shell with indices of order L and k representatives; the shell probability is at most $C\delta L$. Consequently at the critical power its normalized moment on this early-transition event is bounded by

$$CR(1 + S_N)\big(N/(1 + S_N)\big)^{-1/(b-1)} = o(1).$$

Ignoring later transitions adds error at most $CT_2^{(1-b)/2}$, with T_2 the physical time of the second grid transition. For the first transition at time $m\delta$, the next waiting time is an independent geometric number of steps. Splitting that number at $1/\delta$ gives

$$\mathbb{E}[T_2^{-1} \mid T_1 = m\delta] \le C[1 + \log(1 + 1/(m\delta))].$$

Indeed each probability through $1/\delta$ is at most $\rho(\delta) \le \delta$, and summing the reciprocals of $m+k$ gives the logarithm. The remaining terms contribute at most a constant. Summing over $m\delta \le \tau_0$ with probabilities at most δ and comparing with an integral bounds $\mathbb{E}[T_2^{-1}\mathbf{1}_{\{T_1\le\tau_0\}}]$ by $C\tau_0(1+\log(1/\tau_0))$. Multiplication by R to normalize the critical moment leaves $o(1)$. The sum of the two early errors therefore has negligible normalized critical moment.

On the event $T_1 > \tau_0$, the constant-sign code gives squared normalized error at most $C(R/\tau_0)^{b-1}$. Use another $N/4$ codepoints for the exponential-overflow construction from the preceding section, with dyadic thresholds up to a fixed multiple of this bound, and include the two constant-sign codewords. The number of levels is $O(R)$, so its individual budgets are still at least $R/2$. Take $s' > s$ and choose the precision exponent and θ as in that construction. Choose ζ small enough that $e^{-\theta R/4}(R/\tau_0)^{(b-1)s'/2} \to 0$. The same integration gives a uniform normalized s'-moment bound on the nonearly event, regardless of the mesh. Finally include the $N/2$ codepoints achieving the exact coverage threshold. The normalized distance then converges in probability to $\sqrt{K_b}$. The negligible moment on the early event and uniform integrability on its complement prove the exact critical moment limit. The same upper proof with $S_N = O(S_N^*)$ bounds the early moment by a fixed constant, while the bound on the complement is unchanged. This proves the boundedness assertion and completes Theorem 7.2.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of each part. Anthropic Claude edited, consolidated and cross-referenced the parts and checked computations. Scott Schubr directed the research and is responsible for the content.

References

[1] S. Schubr. *Wasserstein transport for Kloosterman paths: optimal rates, prime powers and transport constants*, edition 2. Luminary Research Archive, November 2026. Section K, *Exact functional quantization for weighted stationary sequences*; the bounded independent case is the part *The finite-support approximation constant of the Kloosterman limit*. https://doi.org/10.5281/zenodo.24445508.

[2] S. Dereich, F. Fehringer, A. Matoussi, and M. Scheutzow. On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *Journal of Theoretical Probability* 16(1):249–265, 2003. https://doi.org/10.1023/A:1022242924198.

[3] S. Dereich and M. Scheutzow. High resolution quantization and entropy coding for fractional Brownian motion. *Electronic Journal of Probability* 11:700–722, 2006. https://doi.org/10.1214/EJP.v11-344.

[4] S. Dereich. The coding complexity of diffusion processes under supremum norm distortion. *Stochastic Processes and their Applications* 118(6):917–937, 2008. https://doi.org/10.1016/j.spa.2007.07.003.

[5] S. Dereich. The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion. *Stochastic Processes and their Applications* 118(6):938–951, 2008. https://doi.org/10.1016/j.spa.2007.07.002.

[6] M. Sh. Birman and M. Z. Solomyak. Spectral asymptotics of nonsmooth elliptic operators. I. *Transactions of the Moscow Mathematical Society* 27, 1972; English translation 27:1–52, 1975.

[7] V. V. Borzov. Quantitative characteristics of singular measures. In *Spectral Theory and Wave Processes*, Springer US, 1971, 37–42. https://doi.org/10.1007/978-1-4684-8926-2_5.

[8] G. Rozenblum and E. Shargorodsky. Eigenvalue estimates and asymptotics for weighted pseudodifferential operators with singular measures in the critical case. In *Partial Differential Equations, Spectral Theory, and Mathematical Physics*, EMS Series of Congress Reports, EMS Press, 2021, 331–354. https://doi.org/10.4171/ECR/18-1/20; https://arxiv.org/abs/2011.14877.

[9] R. M. Gray. *Entropy and Information Theory*. Corrected first edition. https://ee.stanford.edu/~gray/it.pdf.

[10] J. C. Kieffer. Sample converses in source coding theory. *IEEE Transactions on Information Theory*, 37(2):263–268, 1991. https://doi.org/10.1109/18.75241.

[11] E.-H. Yang and J. C. Kieffer. Simple universal lossy data compression schemes derived from the Lempel–Ziv algorithm. *IEEE Transactions on Information Theory*, 42(1):239–245, 1996. https://doi.org/10.1109/18.481794.

[12] T. Weissman and E. Ordentlich. The empirical distribution of rate-constrained source codes. *IEEE Transactions on Information Theory*, 51(11):3718–3733, 2005. https://doi.org/10.1109/TIT.2005.856982.

[13] H. Luschgy and G. Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Annals of Probability* 32:1574–1599, 2004. https://doi.org/10.1214/009117904000000324.

[14] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

[15] E. E. Tyrtyshnikov and N. L. Zamarashkin. Toeplitz eigenvalues for Radon measures. *Linear Algebra and its Applications* 343–344:345–354, 2002. https://doi.org/10.1016/S0024-3795(01)00345-7.

[16] H. Luschgy and G. Pagès. Functional quantization and entropy for stochastic processes. *Dagstuhl Seminar Proceedings* 4401:1–15, 2005. https://doi.org/10.4230/DagSemProc.04401.4.

[17] E. Lindenstrauss and M. Tsukamoto. From rate distortion theory to metric mean dimension: variational principle. https://arxiv.org/abs/1702.05722.

[18] V. Kostina and S. Verdú. Fixed-length lossy compression in the finite blocklength regime. *IEEE Transactions on Information Theory* 58:3309–3338, 2012. https://doi.org/10.1109/TIT.2012.2186786.

[19] S. Y. Hong, M. Lifshits, and A. Nazarov. Small deviations in L_2-norm for Gaussian dependent sequences. *Electronic Communications in Probability* 21, paper 41, 1–9, 2016. https://doi.org/10.1214/16-ECP4708.

[20] M. Lifshits and A. Nazarov. L_2-small deviations for weighted stationary processes. *Mathematika* 64:387–405, 2018. https://doi.org/10.1112/S0025579317000572.

[21] R. Ponge and Y. Tian. Nonclassical Weyl laws and Connes' integration for weak Lorentz ideals, I. https://arxiv.org/abs/2605.25591.

[22] G. Barbarino, C. Garoni, and S. Serra-Capizzano. Block generalized locally Toeplitz sequences: theory and applications in the multidimensional case. *Electronic Transactions on Numerical Analysis* 53:113–216, 2020. https://doi.org/10.1553/ETNA_VOL53S113.

[23] F. Aurzada, S. Dereich, M. Scheutzow, and C. Vormoor. High resolution quantization and entropy coding of jump processes. *Journal of Complexity* 25:163–187, 2009. https://doi.org/10.1016/j.jco.2008.10.002.