Ergodic Theory in Lean 4

Marcel Morgenstern

July 6, 2026

Chapter 1

Introduction

This blueprint accompanies a *sorry-free* Lean 4 + Mathlib formalization of the **Oseledets multiplicative ergodic theorem** (MET) and a broad layer of companion results. It documents the mathematical content of the development and the dependency structure of its proof: every node carries a `\lean` annotation recording the fully qualified name of the corresponding Lean declaration, and a green checkmark records that the statement (and, where shown, its proof) is fully formalized. The continuous integration deliberately does not build a `doc-gen4` API reference (doing so would regenerate documentation for the entire Mathlib import closure), so the annotations are names to be looked up in the Lean source repository linked from this blueprint.

1.1 Setting

Throughout, (X, μ) is a probability space and $T : X \to X$ is an ergodic measure-preserving transformation. A *linear cocycle* of dimension d is generated by a measurable map

$$A : X \to \mathrm{Mat}_d(\mathbb{R}), \qquad \det(A\,x) \neq 0 \text{ for all } x,$$

whose n-step product along the orbit of x is

$$A^{(n)}(x) \;=\; A(T^{n-1}x) \cdots A(Tx)\,A(x), \qquad A^{(0)}(x) = I.$$

We impose the one-sided integrability condition

$$\log^+ \|A\|\,,\; \log^+ \left\|A^{-1}\right\| \in L^1(\mu),$$

where matrices act on EuclideanSpace $\mathbb{R}$ (Fin d) so that $\|\cdot\|$ is the L^2 operator norm (submultiplicative). For $v \neq 0$ the *Lyapunov exponent* in the direction v is the growth rate

$$\bar{\lambda}(x, v) \;=\; \limsup_{n\to\infty} \tfrac{1}{n} \log\left\|A^{(n)}(x)\,v\right\|.$$

1.2 The three headline theorems of the multiplicative core

The multiplicative core of the development proves three principal theorems, each formalized sorry-free and verified (by a guarded axiom audit) to depend only on the standard axioms $\{$`propext`, `Classical.choice`, `Quot.sound`$\}$.

- **One-sided MET (filtration form),** : there are finitely many distinct Lyapunov exponents $\lambda_1 > \cdots > \lambda_k$ and, for μ-a.e. x, a strictly decreasing, A-equivariant, measurable filtration
$$\text{EuclideanSpace } \mathbb{R} \text{ (Fin } d) = V_x^0 \supsetneq V_x^1 \supsetneq \cdots \supsetneq V_x^k = \{0\}$$
along which $\frac{1}{n} \log \left\| A^{(n)}(x)\, v \right\| \to \lambda_i$ for every $v \in V_x^i \setminus V_x^{i+1}$. This is the central result; it is proved in Chapter 5.

- **Two-sided splitting,** : when both the forward and backward filtrations are available, they are transverse and the space splits a.e. into an A-equivariant direct sum of *Oseledets subspaces* $\mathbb{R}^d = \bigoplus_i E_x^i$ on which the cocycle grows at the exact two-sided rate λ_i. Chapter 7.

- **Continuous-flow MET,** : the analogue for a continuous-time linear cocycle over a measurable flow, obtained by reduction to the time-one map together with a between-times sandwich estimate. Chapter 8.

1.3 Structure of the proof

The proof proceeds in layers, mirrored by the chapters of this blueprint. Chapter 2 sets up the cocycle, the operator norm and its measurability, and the Furstenberg–Kesten extremal exponents. Chapter 3 develops the ergodic-theoretic engine: the maximal ergodic inequality, the pointwise Birkhoff theorem, and—crucially—Kingman's subadditive ergodic theorem, which converts the subadditivity of the log-norms into almost-everywhere limits. Chapter 4 introduces the Lyapunov exponent as a measurable function, the Lyapunov spectrum, and the limsup filtration together with the measurability of its subspaces. Chapter 5 assembles the one-sided theorem through the Oseledets limit, the spectral upper bound and determinant squeeze, the spectral identification of the limsup filtration, and a top-gap envelope induction. The companion results (Chapter 6), the two-sided splitting (Chapter 7), and the continuous-flow theorem (Chapter 8) complete the multiplicative core.

1.4 Beyond the multiplicative theory

The library does not stop at the multiplicative theorems. Four further layers, each with its own chapter, extend the development across classical entropy theory, symbolic dynamics, dimension theory, and smooth examples.

- **Kolmogorov–Sinai entropy** (Chapter 9): the Shannon entropy of a finite measurable partition, conditional entropy, the Kolmogorov–Sinai entropy as a Fekete limit over iterated joins, the generator theorems and the Abramov–Rokhlin addition formula, closing where the additive and multiplicative theories meet—the Margulis–Ruelle inequality $h(T) \le \sum_{\lambda_i > 0} \lambda_i$ and Rokhlin's volume-distortion identity.

- **Generators** (Chapter 10): the pointwise Shannon–McMillan–Breiman theorem (entropy equipartition), the Rokhlin–Kakutani tower lemma, a symbolic coding stack, and **Krieger's finite generator theorem**: an ergodic aperiodic automorphism of entropy below $\log k$ admits a two-sided generating partition with at most k cells.

- **Multifractal analysis** (Chapter 11): the coarse-grained formalism—partition function Z_q, mass exponent $\tau(q)$, Rényi dimensions D_q, singularity spectrum $f(\alpha)$—together with the pointwise local dimension, the Frostman/Billingsley bridge to Hausdorff dimension, and a Bernoulli-suspension witness whose Rényi spectrum is provably q-dependent.

- **Smooth maps and worked examples** (Chapter 12): the derivative cocycle connecting the MET to smooth dynamics, the foliation-free expanding-case identity of the Pesin and Rokhlin right-hand sides, and the classical examples—the doubling map with its instantiated Rokhlin equality at rate $\log 2$, and the ergodic Arnold cat map with its explicit Lyapunov spectrum.

Like the headline theorems, every result in these layers is formalized sorry-free and passes the same guarded axiom audit. The blueprint documents these principal chains rather than every module of the library; some supporting material (for instance the singular-filtration measurability layer) is not given a chapter of its own.

1.5 A finite-dimensional quantum-information layer

Built on the same matrix and continuous-functional-calculus infrastructure, the development also formalizes a self-contained *finite-dimensional quantum-information* layer, documented in the final two chapters. Chapter 13 develops the quantum relative entropy and its monotonicity: density matrices and the von Neumann entropy, the Umegaki relative entropy and Klein's inequality, **Lieb's joint-convexity theorem**, and the **data-processing inequality** for completely positive trace-preserving maps. Chapter 14 treats the **Petz recovery map** and **both directions of Petz's equality theorem**—recovery implies saturation of the data-processing inequality and, fully generally, saturation implies recoverability—together with the **CNT quantum dynamical entropy**, whose abelian corner recovers the classical Kolmogorov–Sinai entropy of the ergodic-theory core. Like the rest of the development, these results are formalized sorry-free and pass the same guarded axiom audit.

Chapter 2

The linear cocycle and Furstenberg–Kesten

2.1 The measure-preserving system and the linear cocycle

The Oseledets theorem studies the long-term growth of products of matrices driven by a dynamical system. The data are a measure-preserving self-map $T : X \to X$ of a probability space (X, μ) and a matrix-valued generator $A : X \to \mathrm{Matrix}(\mathrm{Fin}\, d)(\mathrm{Fin}\, d)\, \mathbb{R}$. The fundamental object is the *iterated linear cocycle*: the product of the generator evaluated along the orbit of x. We use the convention that the newest factor sits on the left.

Definition 2.1 (The entrywise measurable structure on matrices)**.** A matrix is a function $m \to n \to \alpha$, and we equip Matrix $m\, n\, \alpha$ with the Pi (product) σ-algebra induced from the entry type α. For finitely many entries over a Borel α this agrees with the Borel σ-algebra. Mathlib does not register this automatically because Matrix is a `def` rather than reducibly the underlying Pi type, so the ambient Pi instance does not transfer; we install it explicitly.

Definition 2.2 (The iterated linear cocycle)**.** Given $A : X \to \mathrm{Matrix}(\mathrm{Fin}\, d)(\mathrm{Fin}\, d)\, \mathbb{R}$ and $T : X \to X$, define $A^{(n)} =$ `cocycle` $A\, T\, n$ by recursion on n:

$$A^{(0)}(x) = 1, \qquad A^{(n+1)}(x) = A^{(n)}(Tx) \cdot A(x).$$

Unfolding, $A^{(n)}(x) = A(T^{n-1}x) \cdots A(Tx)\, A(x)$, with the newest factor on the left. The matrix norm throughout is the scoped L2 operator norm, which is submultiplicative; vectors live in EuclideanSpace $\mathbb{R}\,(\mathrm{Fin}\, d)$ and the action is via `toEuclideanCLM`.

Theorem 2.3 (The cocycle identity)**.** *For all $m, n \in \mathbb{N}$ and $x \in X$,*

$$A^{(m+n)}(x) = A^{(m)}(T^n x) \cdot A^{(n)}(x).$$

Proof. Induction on n with x generalized. For $n = 0$ both sides equal $A^{(m)}(x)$. For the step, reassociate $m + (n + 1)$ and apply the defining recursion $A^{(k+1)}(x) = A^{(k)}(Tx) \cdot A(x)$ on both sides, then use the inductive hypothesis at Tx together with $T^{n+1}x = T^n(Tx)$ and associativity of matrix multiplication. □

Definition 2.4 (One-sided log-integrability of the generator)**.** The hypothesis `IntegrableLogNorm` $A\, \mu$ asserts that the positive part of the log-norm of the generator is integrable: $\log^+ \|A(\cdot)\| \in L^1(\mu)$, where $\log^+ t = \max(\log t, 0)$. This is the standard integrability assumption of the Furstenberg–Kesten and Oseledets theorems; combined with the same hypothesis for the inverse generator A^{-1} it pins both extremal Lyapunov exponents in $\mathbb{R}$.

Theorem 2.5 (Measurability of the cocycle iterates)**.** *If A and T are measurable then for each n the iterate $x \mapsto A^{(n)}(x)$ is measurable for the entrywise structure 2.1.*

Proof. Induction on n. The base case is a constant map. For the step, the recursion writes $A^{(n+1)}$ as the product $(A^{(n)} \circ T) \cdot A$; matrix multiplication is jointly measurable on the Pi structure because each entry of a product is the finite sum $\sum_k M_{ik} N_{kj}$ of products of measurable coordinate projections, so the result follows by composing the inductive hypothesis with T and multiplying by A. □

2.2 Measurability of the operator norm and the inverse

To feed the cocycle into the ergodic theory we must know that the operator norm and the matrix inverse are measurable on the entrywise structure. The subtlety is that Mathlib's `Measurable.norm` is stated for a `BorelSpace`, whereas our matrix σ-algebra is the Pi structure; the two coincide here because the L2 operator-norm topology is definitionally the Pi product topology.

Lemma 2.6 (The Pi structure is an opens-measurable space)**.** *The entrywise (Pi) measurable structure on* $\mathrm{Matrix}(\mathrm{Fin}\, d)(\mathrm{Fin}\, d)\, \mathbb{R}$ *is an `OpensMeasurableSpace` for the L2 operator-norm topology, since that topology is installed (via `replaceTopology`) to be definitionally the Pi product topology, of which the Pi σ-algebra is exactly the Borel structure.*

Theorem 2.7 (Measurability of the L2 operator norm)**.** *The map $M \mapsto \|M\|$ on* $\mathrm{Matrix}(\mathrm{Fin}\, d)(\mathrm{Fin}\, d)\, \mathbb{R}$ *is measurable.*

Proof. The norm is continuous for the operator-norm topology, and by 2.6 that topology's Borel structure is the entrywise Pi structure; continuous maps into a Borel codomain are measurable. □

Theorem 2.8 (Measurability of the determinant)**.** *The determinant $M \mapsto \det M$ is measurable.*

Proof. By the Leibniz formula $\det M = \sum_\sigma \mathrm{sgn}(\sigma) \prod_i M_{i,\sigma(i)}$, the determinant is a finite sum of finite products of measurable coordinate projections, hence measurable. □

Theorem 2.9 (Measurability of the matrix inverse)**.** *The inverse $M \mapsto M^{-1}$ is measurable on the entrywise structure.*

Proof. Writing $M^{-1} = (\det M)^{-1} \cdot \mathrm{adj}(M)$, each entry is a ratio of polynomials in the entries. The adjugate is measurable (each entry is a determinant of a row update, again a polynomial in the entries), and $(\det M)^{-1}$ is measurable by 2.8 and measurability of inversion on $\mathbb{R}$; the product is therefore measurable entrywise. □

2.3 Positivity and submultiplicativity of the log-norm

To take logarithms cleanly we need the iterate norms to be strictly positive, which forces the generator to be everywhere invertible ($\det A \neq 0$) and the dimension to be nonzero. These hypotheses also make the log of a product split as a genuine sum, which is what yields subadditivity rather than a mere inequality with junk values.

Lemma 2.10 (Invertibility of the iterates)**.** *If* $\det A(x) \neq 0$ *for every* x*, then* $\det A^{(n)}(x) \neq 0$ *for all* n, x*.*

Proof. Induction on n. The base case is $\det 1 = 1$. For the step, $\det(A^{(n)}(Tx) \cdot A(x)) = \det A^{(n)}(Tx) \cdot \det A(x)$, and both factors are nonzero by the inductive hypothesis and the assumption on A. ☐

Lemma 2.11 (The unit matrix has norm one)**.** *When* $d \neq 0$*,* $\|(1 : \mathrm{Matrix}(\mathrm{Fin}\, d)(\mathrm{Fin}\, d)\, \mathbb{R})\| = 1$ *for the L2 operator norm.*

Proof. For $d \neq 0$ the space EuclideanSpace $\mathbb{R}\,(\mathrm{Fin}\, d)$ is nontrivial. The star-algebra equivalence `toEuclideanCLM` sends the identity matrix to the identity operator and preserves the norm, and the operator norm of the identity on a nontrivial space is 1. There is no `NormOneClass` instance for the matrix operator norm, so this must be proved by hand; note that at $d = 0$ the statement is false ($\|1\| = 0$), which is why $d \neq 0$ is required. ☐

Lemma 2.12 (Positivity of the iterate norms)**.** *Assume* $\det A(x) \neq 0$ *for all* x *and* $d \neq 0$*. Then* $0 < \|A^{(n)}(x)\|$ *for every* n, x*. The analogous statement* `norm_inv_cocycle_pos` *holds for the inverse iterates* $\|(A^{(n)}(x))^{-1}\|$*.*

Proof. If $\|A^{(n)}(x)\| = 0$ then $A^{(n)}(x)$ is the zero matrix, whose determinant vanishes (as $d \neq 0$), contradicting 2.10. The inverse case is identical using $\det((A^{(n)}(x))^{-1}) = (\det A^{(n)}(x))^{-1} \neq 0$. ☐

Theorem 2.13 (Subadditivity of the log-norm cocycle)**.** *Assume* $\det A \neq 0$ *everywhere and* $d \neq 0$*. Then* $g_n(x) = \log \|A^{(n)}(x)\|$ *is a subadditive cocycle over* T*:*

$$g_{m+n}(x) \leq g_m(x) + g_n(T^m x).$$

Proof. Rewrite $m+n$ as $n+m$ and apply the cocycle identity 2.3 to get $A^{(m+n)}(x) = A^{(n)}(T^m x) \cdot A^{(m)}(x)$. Submultiplicativity of the operator norm and monotonicity of log give $\log \|A^{(m+n)}(x)\| \leq \log(\|A^{(n)}(T^m x)\| \cdot \|A^{(m)}(x)\|)$; both factors are strictly positive by 2.12, so log of the product splits as the sum $\log \|A^{(n)}(T^m x)\| + \log \|A^{(m)}(x)\|$, which is the required bound after commuting the summands. ☐

Theorem 2.14 (Subadditivity of the inverse log-norm cocycle)**.** *Under the same hypotheses,* $g_n(x) = \log \|(A^{(n)}(x))^{-1}\|$ *is a subadditive cocycle over* T*.*

Proof. As above, after 2.3 the inverse of the product reverses the order: $(A^{(m+n)}(x))^{-1} = (A^{(m)}(x))^{-1} \cdot (A^{(n)}(T^m x))^{-1}$. Submultiplicativity, monotone log, strict positivity of each inverse norm (2.12), and log of a product splitting as a sum give the subadditive inequality. ☐

2.4 Birkhoff-sum sandwich bounds

The two-sided integrability hypothesis controls the log-norm by Birkhoff sums of $\log^+ \|A\|$ and $\log^+ \|A^{-1}\|$ on both sides. These bounds drive both the integrability of each level and the bounded-below proviso of Kingman's theorem.

Lemma 2.15 (Upper Fekete bound)**.** *For* $\det A \neq 0$ *everywhere and* $d \neq 0$*,*

$$\log \|A^{(n)}(x)\| \leq \sum_{k<n} \log^+ \|A(T^k x)\|.$$

Proof. Induction on n. The base case is $0 \le 0$. For the step, the recursion gives $A^{(n+1)}(x) = A^{(n)}(Tx) \cdot A(x)$; submultiplicativity and log-splitting bound $\log\|A^{(n+1)}(x)\|$ by $\log\|A(x)\| + \log\|A^{(n)}(Tx)\|$, and then $\log\|A(x)\| \le \log^+\|A(x)\|$ together with the inductive hypothesis at Tx peels off one Birkhoff term. □

Lemma 2.16 (Lower bound via the inverse Birkhoff sum)**.** *Under the same hypotheses,*

$$-\sum_{k<n} \log^+\|(A(T^k x))^{-1}\| \le \log\|A^{(n)}(x)\|.$$

Proof. From $A^{(n)}(x) \cdot (A^{(n)}(x))^{-1} = 1$ and submultiplicativity, $1 = \|1\| \le \|A^{(n)}(x)\| \cdot \|(A^{(n)}(x))^{-1}\|$ (using 2.11 and 2.10), so taking logs gives $0 \le \log\|A^{(n)}(x)\| + \log\|(A^{(n)}(x))^{-1}\|$. Bounding the inverse log-norm above by its Birkhoff sum 2.17 and rearranging yields the claim. □

Lemma 2.17 (Upper Fekete bound for the inverse cocycle)**.** *Under the same hypotheses,*

$$\log\|(A^{(n)}(x))^{-1}\| \le \sum_{k<n} \log^+\|(A(T^k x))^{-1}\|.$$

Proof. Induction on n, exactly as for the forward bound but with the inverse generator: the recursion and `mul_inv_rev` give $(A^{(n+1)}(x))^{-1} = (A(x))^{-1} \cdot (A^{(n)}(Tx))^{-1}$, and submultiplicativity plus log-splitting and $\log\|(A(x))^{-1}\| \le \log^+\|(A(x))^{-1}\|$ peel off one Birkhoff term. □

2.5 Integrability of each level

Lemma 2.18 (Integral of a Birkhoff sum)**.** *For measure-preserving T and integrable f, $\int \sum_{k<n} f(T^k x)\, d\mu = n \int f\, d\mu$.*

Proof. Each composition $f \circ T^k$ is integrable and integral-preserving because T^k is measure-preserving, so $\int f \circ T^k\, d\mu = \int f\, d\mu$. Summing the n equal terms over the finite range gives $n \int f\, d\mu$. □

Theorem 2.19 (Integrability of the log-norm levels)**.** *Let T be measure-preserving for a finite measure μ, A measurable and everywhere invertible, $d \ne 0$, with both $\log^+\|A\|$ and $\log^+\|A^{-1}\|$ integrable. Then each $x \mapsto \log\|A^{(n)}(x)\|$ is integrable. The companion `integrable_logNorm_inv_cocycle` gives the same for the inverse iterates.*

Proof. The level g_n is sandwiched between $-B_n^-$ and B_n^+, where $B_n^\pm$ are the Birkhoff sums of $\log^+\|A\|$ and $\log^+\|A^{-1}\|$ (Lemmas 2.15, 2.16). Both $B_n^\pm$ are nonnegative and integrable (finite sums of integrable, m.p. precompositions), so $|g_n| \le B_n^+ + B_n^-$ pointwise. Since g_n is measurable (2.7 composed with the cocycle), domination by the integrable $B_n^+ + B_n^-$ gives integrability. □

2.6 The Furstenberg–Kesten theorems

Feeding the subadditive cocycles into Kingman's ergodic theorem produces the two extremal Lyapunov exponents as a.e.-constant limits. The bounded-below proviso of Kingman is supplied by the cross integrability hypothesis: for the top exponent the lower bound comes from $\log^+\|A^{-1}\|$, and vice versa. This is precisely where the second integrability hypothesis keeps the limit finite in $\mathbb{R}$ rather than $-\infty$.

Theorem 2.20 (Furstenberg–Kesten, top exponent)**.** *Let T be ergodic for a probability measure μ, let A be measurable and everywhere invertible with $\log^+\|A\|, \log^+\|A^{-1}\| \in L^1(\mu)$. Then there is a constant $\lambda_1 \in \mathbb{R}$ (the top Lyapunov exponent) with*

$$\frac{1}{n}\log\left\|A^{(n)}(x)\right\| \xrightarrow{n\to\infty} \lambda_1 \qquad \textit{for } \mu\textit{-a.e. } x.$$

Proof. If $d = 0$ the matrix algebra is trivial, every norm is 0, and the limit is the constant 0. Otherwise set $d \neq 0$ and let $g_n = \log\left\|A^{(n)}\right\|$. It is a subadditive cocycle (2.13) with each level integrable (2.19). For the bounded-below proviso, the lower bound 2.16 and the Birkhoff integral identity 2.18 give $(\int g_{n+1})/(n+1) \geq -\int \log^+\left\|A^{-1}\right\| d\mu$, a constant lower bound. The ergodic Kingman theorem 3.26 then returns the a.e. constant limit, which is the top exponent λ_1 (see 4.8). □

Theorem 2.21 (Furstenberg–Kesten, bottom exponent)**.** *Under the same hypotheses there is a constant $\lambda \in \mathbb{R}$ with*

$$\frac{1}{n}\log\left\|(A^{(n)}(x))^{-1}\right\| \xrightarrow{n\to\infty} \lambda \qquad \textit{for } \mu\textit{-a.e. } x,$$

so the bottom Lyapunov exponent $\lambda_k = -\lambda$ exists and is finite.

Proof. Identical to the top case with the inverse subadditive cocycle $g_n = \log\left\|(A^{(n)})^{-1}\right\|$ (2.14). The roles of the two integrability hypotheses swap: the bounded-below proviso now uses the lower bound $-\sum_{k<n}\log^+\left\|A(T^k x)\right\| \leq \log\left\|(A^{(n)}(x))^{-1}\right\|$, obtained from $\|1\| \leq \left\|A^{(n)}\right\| \cdot \left\|(A^{(n)})^{-1}\right\|$ (2.11) and the forward Fekete bound 2.15, together with the integral identity 2.18. Ergodic Kingman 3.26 delivers the a.e. constant limit. □

Chapter 3

Ergodic theorems

This chapter develops the three classical ergodic theorems that drive the multiplicative ergodic theorem: the *maximal ergodic inequality* of Hopf and Garsia, the *pointwise (Birkhoff) ergodic theorem*, and *Kingman's subadditive ergodic theorem*. The first is the analytic gateway; the second turns the gateway into almost-everywhere convergence of Birkhoff averages; the third is the genuine engine of Oseledets, turning a subadditive sequence of log-norms into an almost-everywhere limit.

Throughout, $T : X \to X$ is a map on a measurable space carrying a measure μ, $\texttt{birkhoffSum}\,T\,g\,n\,x = \sum_{k<n} g(T^{[k]}x)$ is the n-th Birkhoff partial sum, and $\texttt{birkhoffAverage}\,\mathbb{R}\,T\,g\,n\,x = n^{-1}\,\texttt{birkhoffSum}\,T\,g\,n\,x$ is the Birkhoff average. The arguments are due to Garsia, Katznelson–Weiss, and Karlsson; the formalization keeps every quantity in $\mathbb{R}$ (or, where $-\infty$ cannot be excluded a priori, in `EReal`) to avoid junk values.

3.1 Maximal ergodic inequality

Definition 3.1 (Garsia's maximal function)**.** For a measurable T, a function $g : X \to \mathbb{R}$, $N : \mathbb{N}$ and x, the *maximal function* is

$$\texttt{maxBirkhoff}\,T\,g\,N\,x = \max_{0\le k\le N} \texttt{birkhoffSum}\,T\,g\,k\,x,$$

the nonempty `Finset.sup'` of the Birkhoff partial sums over $k \in \texttt{range}\,(N+1)$. Since the $k = 0$ term equals $\texttt{birkhoffSum}\,T\,g\,0\,x = 0$, the maximal function is its own positive part.

Lemma 3.2 (Nonnegativity of the maximal function)**.** *For all g, N, x one has $0 \le \texttt{maxBirkhoff}\,T\,g\,N\,x$.*

Proof. The index $k = 0$ lies in $\texttt{range}\,(N+1)$, and the corresponding term is $\texttt{birkhoffSum}\,T\,g\,0\,x = 0$. The supremum over a nonempty finite set dominates each of its terms, so the maximal function is at least 0. ☐

Lemma 3.3 (Recursion for the maximal function)**.** *For all g, N, x,*

$$\texttt{maxBirkhoff}\,T\,g\,(N+1)\,x = \texttt{birkhoffSum}\,T\,g\,(N+1)\,x \sqcup \texttt{maxBirkhoff}\,T\,g\,N\,x.$$

Proof. This is the identity $\texttt{range}\,(N+2) = \texttt{insert}\,(N+1)\,(\texttt{range}\,(N+1))$ applied to the supremum: splitting off the top index gives the join of the new term with the previous maximum. Both inequalities follow from the universal/existential characterizations of `Finset.sup'`. The recursion is what makes integrability provable by induction on N. ☐

Lemma 3.4 (Garsia's pointwise inequality)**.** *On the set where the maximal function is positive, for every x with $0 < \mathit{maxBirkhoff}\, T\, g\, N\, x$,*

$$\mathit{maxBirkhoff}\, T\, g\, N\, x \le g\, x + \mathit{maxBirkhoff}\, T\, g\, N\, (T x).$$

Proof. Pull the constant $g\,x$ through the supremum: by the additive shift of a nonempty `sup'` and the recursion $\texttt{birkhoffSum}\, T\, g\, (k+1)\, x = g\, x + \texttt{birkhoffSum}\, T\, g\, k\, (T x)$, the right-hand side equals $\max_{0 \le k \le N} \texttt{birkhoffSum}\, T\, g\, (k+1)\, x$, the maximum of the *shifted* partial sums. The maximum defining the left-hand side is attained at some index k_0; positivity forces $k_0 \ge 1$ (the index 0 gives the value 0), so $\texttt{birkhoffSum}\, T\, g\, k_0\, x$ is one of the shifted sums and is therefore dominated by their maximum. □

Lemma 3.5 (Measurability and integrability of the maximal function)**.** *If T is measure-preserving and g is integrable, then each Birkhoff partial sum and the maximal function $\mathit{maxBirkhoff}\, T\, g\, N$ are integrable; if moreover g and T are measurable, then $\mathit{maxBirkhoff}\, T\, g\, N$ is measurable.*

Proof. Each summand $g \circ T^{[j]}$ is integrable because $T^{[j]}$ is measure-preserving, so the finite Birkhoff sum is integrable; measurability is likewise a finite sum of measurable compositions. Integrability of the maximal function follows by induction on N using the recursion theorem 3.3 and the fact that a join of two integrable functions is integrable; measurability is the measurability of a finite `sup'`. □

Proposition 3.6 (Garsia's inequality at a fixed level)**.** *For a measure-preserving T and a measurable integrable g, and every $N : \mathbb{N}$,*

$$0 \le \int_{\{x \,:\, 0 < \mathit{maxBirkhoff}\, T\, g\, N\, x\}} g\, d\mu.$$

Proof. Write E for the level set and $M = \texttt{maxBirkhoff}\, T\, g\, N$. Integrating the pointwise inequality theorem 3.4 over E gives $\int_E (M - M \circ T) \le \int_E g$. Now $\int_E M \circ T \le \int_X M \circ T = \int_X M$ by nonnegativity and measure-preservation (via `integral_map` and $\texttt{map}\, T\, \mu = \mu$, avoiding any embedding hypothesis), and $\int_X M = \int_E M$ since M vanishes off E. Chaining these cancels the two maximal-function integrals and leaves $0 \le \int_E (M - M \circ T) \le \int_E g$. □

Lemma 3.7 (Level sets exhaust the target set)**.** *The level sets $\{x : 0 < \mathit{maxBirkhoff}\, T\, g\, N\, x\}$ are monotone in N, and their union over all N is exactly $\{x : \exists n,\ 0 < \mathit{birkhoffSum}\, T\, g\, (n{+}1)\, x\}$.*

Proof. Monotonicity is `Finset.sup'` monotonicity over the nested ranges. For the union: $0 < \texttt{maxBirkhoff}\, T\, g\, N\, x$ means some $\texttt{birkhoffSum}\, T\, g\, k\, x > 0$ with $k \le N$; the index 0 gives 0 and is excluded, so $k = n + 1$ for some n. Conversely a positive partial sum at index $n + 1$ makes the maximal function positive at level $N = n + 1$. □

Theorem 3.8 (Maximal ergodic inequality (Hopf–Garsia))**.** *For a measure-preserving T and an integrable $f : X \to \mathbb{R}$,*

$$0 \le \int_{\{x \,:\, \exists n,\ 0 < \mathit{birkhoffSum}\, T\, f\, (n+1)\, x\}} f\, d\mu.$$

Proof. First treat a measurable integrable g: by theorem 3.6 each level-set integral is nonnegative, and these integrals converge to the integral over the union theorem 3.7 by monotone convergence of set integrals, so the limit is nonnegative. For general integrable f, pass to a measurable representative $g =^{\text{a.e.}} f$; the Birkhoff sums of f and g agree a.e., so the two target sets agree a.e. (the one for g being measurable, the one for f null-measurable), and the integrands agree a.e., which transfers the inequality from g to f. □

3.2 Birkhoff

Theorem 3.9 (Conditional expectation commutes with the dynamics)**.** *For a finite measure, a measure-preserving measurable T and integrable g, the conditional expectation onto the σ-algebra* $\texttt{invariants}\, T$ *of T-invariant sets is a.e. T-invariant:*

$$(\mu[g \mid \texttt{invariants}\, T]) \circ T =^{a.e.} \mu[g \mid \texttt{invariants}\, T].$$

Proof. Set-integral invariance of $h \circ T$ over a measurable invariant set reduces, via $\texttt{map}\, T\, \mu = \mu$, to the integral of h. Hence $\mu[g \circ T \mid \mathcal{I}]$ and $\mu[g \mid \mathcal{I}]$ have equal integrals over every invariant set, so by uniqueness of the conditional expectation they agree a.e.; combining this with the commutation identity $\mu[g \circ T \mid \mathcal{I}] =^{\text{a.e.}} (\mu[g \mid \mathcal{I}]) \circ T$ (again by uniqueness, using that T is $(\mathcal{I}, \mathcal{I})$-measurable) yields the invariance. □

Lemma 3.10 (Subexponential orbital tail)**.** *For a measure-preserving T and integrable g, the orbital tail $n^{-1}\, g(T^{[n]} x)$ tends to 0 for μ-a.e. x.*

Proof. A Borel–Cantelli argument. For each threshold $\delta = 1/(k+1)$ the series $\sum_n \mu\{x : (n+1)\delta \le |g\, x|\}$ is finite: the pointwise count of crossed thresholds is at most $|g\, x| / \delta$, and integrating (Tonelli) bounds the series by $\delta^{-1} \int |g|$. Measure-preservation transfers finiteness to the shifted orbit $g \circ T^{[n]}$, so a.e. only finitely many n cross any fixed threshold; choosing k large makes the tail eventually smaller than any ε. □

Lemma 3.11 (A.e. boundedness of Birkhoff averages)**.** *For a finite measure, a measure-preserving T and integrable g, the Birkhoff averages $n \mapsto \texttt{birkhoffAverage}\, \mathbb{R}\, T\, g\, (n+1)\, x$ are a.e. bounded above (and, applied to $-g$, a.e. bounded below).*

Proof. The maximal ergodic inequality applied to $g - c$ yields, for the maximal set $B_c = \{x : \exists n,\ c < \texttt{birkhoffAverage}\, \mathbb{R}\, T\, g\, (n+1)\, x\}$, the estimate $c\, \mu B_c \le \int_{B_c} g \le \int |g|$. Taking $c = k \in \mathbb{N}$ gives $\mu B_k \le \int |g| / k \to 0$, so the intersection $\bigcap_k B_k$ is null. Off this null set the range of Birkhoff averages is bounded above. □

Lemma 3.12 (Limsup invariance and vanishing perturbations)**.** *The pointwise limsup $x \mapsto \limsup_n \texttt{birkhoffAverage}\, \mathbb{R}\, T\, g\, n\, x$ is a.e. T-invariant.*

Proof. The difference $A_n(g)(Tx) - A_n(g)(x) = n^{-1}(g(T^{[n]} x) - g\, x)$ tends to 0 a.e. by the tail estimate theorem 3.10. Two bounded sequences differing by a null sequence have equal limsup (proved via `limsup_add_const` and an ε-argument), and boundedness holds a.e. at both x and Tx by theorem 3.11; hence the limsup is a.e. T-invariant. □

Proposition 3.13 (The core maximal-inequality step)**.** *For a finite measure, measure-preserving T, integrable g and $\varepsilon > 0$, the superlevel set where the limsup of the Birkhoff averages exceeds $\mu[g \mid \mathcal{I}] + \varepsilon$ is null.*

Proof. Write $L = \mu[g \mid \mathcal{I}]$ and $Ls = \limsup A_\bullet(g)$. The set $E = \{L + \varepsilon < Ls\}$ is a.e. T-invariant (theorem 3.9, theorem 3.12), hence a.e. equal to a genuinely invariant measurable set E'. Feed $\varphi = \mathbf{1}_{E'}(g - L - \varepsilon)$ to the maximal ergodic inequality theorem 3.8; on E' the partial sums of φ telescope (using orbit-constancy of L), and the maximal set equals E'. Thus $0 \le \int_{E'} (g - L - \varepsilon) = -\varepsilon\, \mu E'$ (using $\int_{E'} g = \int_{E'} L$), forcing $\mu E' = 0$. □

Theorem 3.14 (Pointwise (Birkhoff) ergodic theorem)**.** *For a finite measure, a measure-preserving T and an integrable g, the Birkhoff averages converge μ-a.e. to the conditional expectation of g onto the invariant σ-algebra:*

$$\texttt{birkhoffAverage}\, \mathbb{R}\, T\, g\, n\, x \to (\mu[g \mid \texttt{invariants}\, T])(x).$$

Proof. Unioning the null superlevel sets of theorem 3.13 over $\varepsilon = 1/(k+1)$ gives $\limsup A_\bullet(g) \le \mu[g \mid \mathcal{I}]$ a.e.; applying the same to $-g$ (and using $\limsup(-a) = -\liminf a$) gives $\mu[g \mid \mathcal{I}] \le \liminf A_\bullet(g)$ a.e. With a.e. boundedness theorem 3.11, the sandwich $\limsup \le \mu[g \mid \mathcal{I}] \le \liminf$ forces convergence to $\mu[g \mid \mathcal{I}]$. ☐

Corollary 3.15 (Ergodic case of Birkhoff)**.** *For an ergodic T on a probability space and integrable g, the Birkhoff averages converge μ-a.e. to the space average $\int g\, d\mu$.*

Proof. By theorem 3.14 the limit is $\mu[g \mid \mathcal{I}]$, which is a.e. T-invariant theorem 3.9; ergodicity forces it a.e. constant. Its integral equals $\int g$ (conditional expectation preserves the integral) and equals the constant times $\mu(X) = 1$, so the constant is $\int g$. ☐

3.3 Kingman

Definition 3.16 (Subadditive cocycle)**.** A sequence $g : \mathbb{N} \to X \to \mathbb{R}$ is a *subadditive cocycle* over T when

$$g\,(m+n)\,x \le g\,m\,x + g\,n\,(T^{[m]}x) \qquad \text{for all } m, n, x.$$

For $g_n = \log \|A^{(n)}\|$ this is submultiplicativity of the operator norm composed with the cocycle identity.

Lemma 3.17 (Singleton and block subadditivity)**.** *For a subadditive cocycle and $n : \mathbb{N}$, one has $g\,(n+1)\,x \le$ `birkhoffSum` $T\,(g\,1)\,(n+1)\,x$; more generally, for any decomposition of $[0, N)$ into $k+1$ consecutive blocks of lengths $\ell_0, \dots, \ell_k$, the cocycle is dominated by the sum of the block values along the orbit at the frontiers $T^{[\sum_{j<i} \ell_j]}x$.*

Proof. Both are inductions that peel off the last block and apply the defining subadditivity at the split point. The statement is restricted to nonempty decompositions (and to index $n+1$) because subadditivity at $(0,0)$ only forces $0 \le g\,0\,x$, the wrong sign for a one-sided bound at 0. ☐

Definition 3.18 (Normalized cocycle)**.** The *normalized cocycle* is `cdiv` $g\,n\,x = g\,(n+1)\,x/(n+1)$, the sequence whose a.e. limit is the content of Kingman's theorem; its `EReal` coercion is `ecdiv` $g\,n\,x$, used where $-\infty$ cannot be excluded a priori.

Lemma 3.19 (Fekete limit of the normalized integrals)**.** *For a measure-preserving T, an integrable subadditive cocycle g whose normalized integrals are bounded below, the sequence $(\int g\,(n+1)\,d\mu)/(n+1)$ converges to the Fekete constant γ.*

Proof. Integrating the cocycle inequality and using measure-preservation ($\int g\,n\,(T^{[m]}\cdot) = \int g\,n$) shows the integral sequence $a_n = \int g\,n$ is subadditive in Fekete's sense. The $(n+1)$-indexed lower bound is bridged by hand to a lower bound on a_n/n (the $n = 0$ term is 0), and Fekete's lemma delivers convergence of a_n/n, hence of the shifted sequence, to $\gamma = \inf_n a_n/n$. ☐

Lemma 3.20 (Invariance from a one-sided orbital bound)**.** *For a finite measure, measure-preserving T and an a.e. measurable F with $F\,x \le F\,(Tx)$ for a.e. x, one has $F \circ T =^{a.e.} F$.*

Proof. For each rational c the upper level set $\{c \le F\}$ is null-measurable (via a measurable representative) and a.e. contained in its preimage $T^{-1}\{c \le F\}$, which has equal finite measure; an a.e. subset of equal finite measure is a.e. equal. Ranging over all rational c and collecting the a.e. statements gives the invariance. ☐

Lemma 3.21 (Envelopes are a.e. measurable and T-invariant)**.** *For a finite measure, measure-preserving T, an integrable subadditive cocycle with normalized integrals bounded below, the limsup envelope $f_+(x) = \limsup_n$ `cdiv` $g\,n\,x$ (and likewise the liminf envelope f_-) is a.e. T-invariant.*

Proof. The subadditive bound `cdiv` $g\,n\,x \le g\,1\,x/(n+1) + g\,n\,(Tx)/(n+1)$ differs from `cdiv` $g\,n\,(Tx)$ by a null sequence, so the vanishing-perturbation lemma for limsup gives the pointwise comparison $f_+(x) \le f_+(Tx)$ wherever the cocycle is a.e. bounded (above and below) at x and Tx. Feeding this into theorem 3.20 yields a.e. invariance. ☐

Lemma 3.22 (Integrability of the limsup envelope)**.** *Under the same hypotheses, the limsup envelope f_+ is integrable.*

Proof. The nonnegative Fatou defect $d_n(x) =$ `birkhoffAverage` $\mathbb{R}\,T\,(g\,1)\,(n+1)\,x -$ `cdiv` $g\,n\,x \ge 0$ (singleton subadditivity) controls the envelope. The `ENNReal` Fatou inequality $\int^{-} \liminf u_n \le \liminf \int^{-} u_n$ applied to $u_n =$ `ofReal` d_n, together with $\int d_n = \int g\,1 - a_{n+1}/(n+1) \to \int g\,1 - \gamma < \infty$ (theorem 3.19) and Birkhoff convergence of $A_{n+1}(g\,1)$ to a loose envelope B (theorem 3.14), shows $B - f_+ \ge 0$ has finite lower integral, hence is integrable; therefore $f_+ = B - (B - f_+)$ is integrable. ☐

Proposition 3.23 (Hard direction: $\limsup \le \liminf$ a.e.)**.** *Under the same hypotheses, for a.e. x the `EReal` limsup of the normalized cocycle is dominated by its liminf.*

Proof. This is the stopping-time / greedy block argument of Katznelson–Weiss and Karlsson. After the WLOG shift to the nonpositive process $\tilde{g}\,(n+1) \le 0$, fix $\varepsilon, M > 0$ and set $h = \mu[\max(f_-, -M) \mid \mathcal{I}]$. A greedy two-type partition of $[0, n)$ into "good" blocks (where the stopping time $\tau \le L$ realizing $\tilde{g}\,\tau \le \tau(h + \varepsilon)$ exists) and short singletons (bad/overrun) bounds, via block subadditivity theorem 3.17, $\tilde{g}\,n\,x/n \le (h\,x + \varepsilon)(1 - (L-1)/n -$ `birkhoffAverage` $\mathbf{1}_{B_L})$. Letting $n \to \infty$ (Birkhoff, theorem 3.14), then $L \to \infty$ (the bad sets B_L shrink to a null set), $M \to \infty$ and $\varepsilon \to 0$ yields $f_+ \le f_-$ a.e. The whole argument is carried in `EReal` to keep the bookkeeping clean near $-\infty$. ☐

Theorem 3.24 (Kingman core: a.e. existence of an integrable limit)**.** *For a finite measure, measure-preserving measurable T, an integrable subadditive cocycle with normalized integrals bounded below, there is an integrable G with `cdiv` $g\,n\,x \to G\,x$ for μ-a.e. x.*

Proof. Take $G = f_+$, integrable by theorem 3.22. On the a.e. good set the `EReal` limsup e satisfies $\bot < e \le B < \top$ (finiteness from the loose envelope and the Fatou step), and $\liminf = \limsup = e$ theorem 3.23; in a complete linear order equal liminf and limsup force convergence to e. Transferring to $\mathbb{R}$ gives `cdiv` $g\,n\,x \to e^{\text{toReal}} = f_+\,x$. ☐

Theorem 3.25 (Kingman's subadditive ergodic theorem)**.** *For a finite measure, a measure-preserving T, an integrable subadditive cocycle g whose normalized integrals are bounded below, there is a T-invariant integrable G with*

$$n^{-1}\,g\,n\,x \to G\,x \qquad \text{for } \mu\text{-a.e. } x.$$

Proof. Take $G = f_-$. On the a.e. set where the normalized cocycle is bounded, $\liminf \le \limsup$ is trivial and $\limsup \le \liminf$ is the hard direction inside theorem 3.24, so $f_- =^{\text{a.e.}} f_+$; the sandwich $\limsup \le f_- \le \liminf$ yields pointwise convergence to f_-, and reindexing removes the $n = 0$ term. Invariance is theorem 3.21 (liminf variant), and integrability follows from $f_- =^{\text{a.e.}} f_+$ with theorem 3.22. ☐

Corollary 3.26 (Kingman, ergodic case)**.** *For an ergodic T on a probability space, an integrable subadditive cocycle with normalized integrals bounded below, there is a constant c with $n^{-1}\, g\, n\, x \to c$ for μ-a.e. x.*

Proof. Kingman's theorem theorem 3.25 gives a T-invariant integrable limit G; ergodicity forces an a.e. T-invariant integrable function to be a.e. constant, so $G =^{\text{a.e.}} c$ and the limit is c. (That constant is the Fekete infimum γ; only a.e.-constancy is asserted, as this is what the multiplicative ergodic theorem consumes.) ☐

Chapter 4

Lyapunov exponents and the limsup filtration

This chapter constructs, from the linear cocycle A over the measure-preserving system T, the geometric scaffolding of the Oseledets theorem: a per-vector growth rate, the finite Lyapunov spectrum it produces, and the decreasing flag of subspaces along which the cocycle grows at exactly the prescribed rates. Throughout, $A\colon X \to \mathrm{Mat}_{d\times d}(\mathbb{R})$ is invertible ($\det(Ax) \neq 0$) and acts on $v \in \mathbb{R}^d$ through the Euclidean representation $\texttt{toEuclideanCLM}(M)\colon v \mapsto M \cdot v$. Two abstract ingredients are imported: the Furstenberg–Kesten extremal exponents (the a.e. limits of $\frac{1}{n}\log\|A^{(n)}(x)\|$ and $\frac{1}{n}\log\|A^{(n)}(x)^{-1}\|$) and the subadditive ergodic theorem of Kingman (). The headline statement assembled downstream from this material is .

4.1 Ultrametric growth functions

The combinatorial heart of the construction is purely linear-algebraic: a scaling-invariant, non-Archimedean real function on the nonzero vectors of a real vector space has only finitely many values, and its sublevel sets are subspaces. We isolate this with no reference to dynamics.

Definition 4.1 (Ultrametric growth function)**.** Let E be a real vector space. A function $g\colon E \to \mathbb{R}$ is an *ultrametric growth function* when it is scaling-invariant,

$$g(c \cdot v) = g(v) \qquad (c \neq 0),$$

and non-Archimedean (strong triangle inequality),

$$g(v + w) \leq \max(g(v),\, g(w)) \qquad (v, w,\, v + w \neq 0).$$

The value $g(0)$ is never used; the side conditions $v \neq 0$ are carried explicitly to avoid extended-real arithmetic.

Lemma 4.2 (Strict ultrametric equality)**.** *If g is an ultrametric growth function and $g(v) \neq g(w)$ (with $v, w, v + w$ nonzero), then $g(v + w) = \max(g(v), g(w))$.*

Proof. By symmetry ($v+w = w+v$) assume $g(v) < g(w)$, so $\max = g(w)$. The non-Archimedean inequality gives $g(v + w) \leq g(w)$. Conversely $w = (v + w) + (-v)$ and $g(-v) = g(v)$ (scaling by -1), so $g(w) \leq \max(g(v + w), g(v))$. Were $g(v + w) < g(w)$, then both arguments of this max would be $< g(w)$, contradicting the bound. Hence $g(w) \leq g(v + w)$, giving equality. ☐

Lemma 4.3 (Sum of distinct-value vectors)**.** *Let g be an ultrametric growth function, s a nonempty finite index set, and $v\colon s \to E$ a family of nonzero vectors with $g \circ v$ injective on s. Then $\sum_{i\in s} v_i \neq 0$ and $g(\sum_{i\in s} v_i) = \sup_{i\in s} g(v_i)$.*

Proof. Strong induction on s, peeling off one element a. For the inductive step, the tail sum is nonzero with g-value $g(v_b)$ for some b in the tail; since $g(v_a) \neq g(v_b)$, 4.2 applies to $v_a + \sum_{\text{tail}} v_i$, yielding both nonvanishing and the equality of the value with the new maximum. The two conclusions are proved jointly because each step needs the tail subsum to be nonzero. □

Lemma 4.4 (Distinct values are independent)**.** *If g is an ultrametric growth function and $v\colon \iota \to E$ is a family of nonzero vectors with $g \circ v$ injective, then v is linearly independent over $\mathbb{R}$.*

Proof. Suppose $\sum_j c_j v_j = 0$ with some $c_i \neq 0$. Restrict to the support $t = \{j : c_j \neq 0\} \ni i$, nonempty. The scaled vectors $c_j v_j$ are nonzero, and $g(c_j v_j) = g(v_j)$ by scaling-invariance, so $g \circ (c \cdot v)$ is still injective on t. By 4.3 the support sum is nonzero, contradicting $\sum_{j\in t} c_j v_j = 0$. □

Lemma 4.5 (Finiteness of the value set)**.** *If E is finite-dimensional and g is an ultrametric growth function, then $\{g(v) : v \neq 0\}$ is finite, with at most $\dim_{\mathbb{R}} E$ elements.*

Proof. If the value set were infinite, pick $\dim_{\mathbb{R}} E + 1$ distinct values and witnessing nonzero vectors. By 4.4 these are linearly independent, exceeding $\dim_{\mathbb{R}} E$, a contradiction. □

Definition 4.6 (Sublevel submodule)**.** For an ultrametric growth function g and threshold $t \in \mathbb{R}$, the *sublevel set*

$$\mathrm{sublevel}(g, t) = \{\, v \mid v = 0 \ \lor\ g(v) \le t \,\}$$

is a submodule of E: it contains 0; closure under addition is the non-Archimedean inequality together with $\max(gv, gw) \le t$; closure under scaling is scaling-invariance. These submodules are monotone in t (`sublevel_mono`).

4.2 The upper Lyapunov growth function

We now instantiate the abstract machinery at the per-vector logarithmic growth rate of the cocycle.

Definition 4.7 (Defining sequence)**.** For $x \in X$ and $v \in \mathbb{R}^d$ the *growth sequence* is

$$\mathrm{growthSeq}(A, T, x, v)(n) = \frac{1}{n} \log \left\| A^{(n)}(x) \cdot v \right\|,$$

where $A^{(n)}(x) = \texttt{cocycle}\, A\, T\, n\, x$ is the n-step cocycle iterate acting on v via `toEuclideanCLM`.

Definition 4.8 (Upper Lyapunov growth function)**.** The *upper Lyapunov growth function* is the lim sup of the growth sequence,

$$\bar{\lambda}(v) \;=\; \mathrm{lambdaBar}(A, T, x, v) \;=\; \limsup_{n\to\infty} \frac{1}{n} \log \left\| A^{(n)}(x) \cdot v \right\|.$$

The basic per-n sandwich is the submultiplicativity of the operator norm: $\left\|A^{(n)}(x)^{-1}\right\|^{-1} \|v\| \le \left\|A^{(n)}(x) \cdot v\right\| \le \left\|A^{(n)}(x)\right\| \|v\|$, which after taking $\frac{1}{n}\log$ pins growthSeq between two sequences differing from the Furstenberg–Kesten data by a term tending to 0. This yields both boundedness and finiteness.

Lemma 4.9 (Scaling invariance)**.** *For $c \neq 0$ and $v \neq 0$, $\bar\lambda(c \cdot v) = \bar\lambda(v)$.*

Proof. By linearity $\|A^{(n)}(x)(c \cdot v)\| = |c|\, \|A^{(n)}(x)v\|$, so the two growth sequences differ by $\frac{1}{n}\log|c|$, which $\to 0$. A lim sup is unchanged under a perturbation tending to zero (a robust helper proved directly on the defining sets $\{a : \forall^\infty n,\ u_n \leq a\}$, with no boundedness hypothesis). The bound is unconditional. □

Lemma 4.10 (Finiteness sandwich)**.** *Assume T ergodic on a probability space, A measurable and invertible, with $\log\|A\|$ and $\log\|A^{-1}\|$ integrable. Then there exist $\lambda_{\mathrm{bot}} \leq \lambda_{\mathrm{top}}$ such that for a.e. x and every $v \neq 0$, $\bar\lambda(v) \in [\lambda_{\mathrm{bot}}, \lambda_{\mathrm{top}}]$.*

Proof. Take λ_{top} and λ'_k from the Furstenberg–Kesten limits of $\frac{1}{n}\log\|A^{(n)}\|$ and $\frac{1}{n}\log\|(A^{(n)})^{-1}\|$, and set $\lambda_{\mathrm{bot}} = -\lambda'_k$. The ordering follows from $\|A^{(n)}\|\|(A^{(n)})^{-1}\| \geq 1$. On the intersection of the two full-measure convergence sets, the upper sandwich bounds $\bar\lambda(v) = \limsup \mathrm{growthSeq}$ above by λ_{top}, while the lower sandwich bounds the lim inf below by $-\lambda'_k$; since $\liminf \leq \limsup$ the value lies in the interval. □

Lemma 4.11 (Non-Archimedean inequality)**.** *For nonzero $v, w, v + w$, if the three growth sequences are bounded, then $\bar\lambda(v + w) \leq \max(\bar\lambda(v), \bar\lambda(w))$.*

Proof. From the triangle inequality $\|A^{(n)}(v + w)\| \leq \|A^{(n)}v\| + \|A^{(n)}w\| \leq 2\max(\|A^{(n)}v\|, \|A^{(n)}w\|)$, taking $\frac{1}{n}\log$ gives $\mathrm{growthSeq}(v+w)(n) \leq \frac{1}{n}\log 2 + \max(\mathrm{growthSeq}(v)(n), \mathrm{growthSeq}(w)(n))$. The term $\frac{1}{n}\log 2 \to 0$, and $\limsup\max = \max\limsup$ for bounded sequences, giving the claim. □

Theorem 4.12 ($\bar\lambda$ is an ultrametric growth function, a.e.)**.** *Under the hypotheses of 4.10, for a.e. x the function $v \mapsto \bar\lambda(v)$ is an ultrametric growth function.*

Proof. Scaling-invariance is 4.9 (trivial on $v = 0$). The non-Archimedean axiom is 4.11, whose boundedness hypotheses are discharged on the full-measure Furstenberg–Kesten convergence set, where the growth sequence of every nonzero vector is bounded above and below by the two FK sandwich sequences. □

Theorem 4.13 (A-equivariance, a.e.)**.** *Under the same hypotheses, for a.e. x and every $v \neq 0$,*

$$\bar\lambda_x(v) = \bar\lambda_{Tx}(Ax \cdot v).$$

Proof. The cocycle identity $A^{(n+1)}(x) = A^{(n)}(Tx)\,A(x)$ gives $\mathrm{growthSeq}_x(v)(n+1) = \frac{1}{n+1}\log\|A^{(n)}(Tx)(Ax \cdot v)\|$. Reindexing the lim sup by one, the two scalings differ by $(\frac{1}{n+1} - \frac{1}{n})\log\|\cdot\| = -\frac{1}{n+1} \cdot (\frac{1}{n}\log\|\cdot\|)$, which tends to 0 precisely because $\frac{1}{n}\log\|\cdot\|$ is bounded. The boundedness is needed at the image point Tx; it holds a.e. in x by pulling back the a.e. boundedness at a generic point through the measure-preserving T. □

4.3 The Lyapunov spectrum and the descending exponent list

Definition 4.14 (Lyapunov spectrum)**.** The *Lyapunov spectrum* at x is the finite set of realized values

$$\mathrm{lyapunovSpectrum}(A, T, x) = \{\, \bar\lambda_x(v) : v \neq 0 \,\},$$

defined as a `Finset` via 4.5 on the good set where $\bar\lambda_x$ is an ultrametric growth function, and as $\emptyset$ off it. A value lies in it iff it is realized by some nonzero vector.

Definition 4.15 (Multiplicity count and descending list)**.** Write $k = \mathrm{specCard}(A, T, x)$ for the number of distinct exponents (the cardinality of the spectrum). The *exponent list* $\mathrm{specList}\colon \mathrm{Fin}\, k \to \mathbb{R}$ enumerates the spectrum in *strictly descending* order, $\lambda_0 > \lambda_1 > \cdots > \lambda_{k-1}$, obtained from the order embedding of the finset composed with index reversal. It is strictly antitone, every $\mathrm{specList}(i)$ lies in the spectrum, and every spectrum value is $\mathrm{specList}(i)$ for a unique i.

4.4 The limsup filtration

Definition 4.16 (Sublevel subspace)**.** The *sublevel subspace* at threshold t,

$$\mathrm{lambdaSublevel}(A, T, x, t) = \{\, v \mid v = 0 \ \vee\ \bar{\lambda}_x(v) \le t \,\},$$

is the submodule 4.6 of $\bar{\lambda}_x$ at t on the good set, and $\perp$ off it.

Definition 4.17 (Oseledets filtration / limsup flag)**.** The *limsup flag* at x is the family $\mathrm{vflag}(A, T, x)\colon \mathrm{Fin}\,(k+1) \to \mathrm{Submodule}$ with

$$\mathrm{vflag}(A, T, x)(j) = \begin{cases} \mathrm{lambdaSublevel}(A, T, x,\ \mathrm{specList}(j)) & j < k, \\ \perp & j = k. \end{cases}$$

With the descending enumeration, level j is the sublevel set at λ_j, so the flag decreases from the whole space down to $\perp$.

Lemma 4.18 (Extremal levels)**.** *On the good set,* $\mathrm{vflag}(A, T, x)(0) = \top$*; and unconditionally* $\mathrm{vflag}(A, T, x)(\mathrm{last}) = \perp$ *(`vflag_last`).*

Proof. For any $v \ne 0$, $\bar{\lambda}_x(v)$ lies in the spectrum, so $k > 0$ and $\mathrm{specList}(0)$ is the maximum of the spectrum; thus $\bar{\lambda}_x(v) \le \mathrm{specList}(0)$ and v lies in level 0. Level k is $\perp$ by definition. ☐

Theorem 4.19 (Strict decrease)**.** *On the good set,* $\mathrm{vflag}(A, T, x)(i{+}1) \subsetneq \mathrm{vflag}(A, T, x)(i)$ *for each interior index i.*

Proof. Inclusion: since specList is strictly antitone, $\mathrm{specList}(i{+}1) < \mathrm{specList}(i)$, so the sublevel at the smaller threshold is contained in that at the larger. Strictness: a witness w with $\bar{\lambda}_x(w) = \mathrm{specList}(i)$ exists (the value is realized); it lies in level i but its value exceeds $\mathrm{specList}(i{+}1)$, so it is not in level $i{+}1$. ☐

Lemma 4.20 (Stratum exactness)**.** *On the good set, if* $v \in \mathrm{vflag}(A, T, x)(i)$ *but* $v \notin \mathrm{vflag}(A, T, x)(i{+}1)$, *then* $\bar{\lambda}_x(v) = \mathrm{specList}(i) = \lambda_i$.

Proof. Membership in level i gives $\bar{\lambda}_x(v) \le \lambda_i$. Since $\bar{\lambda}_x(v)$ is a spectrum value, $\bar{\lambda}_x(v) = \lambda_j$ for some $j \ge i$. Non-membership in level $i{+}1$ rules out $\bar{\lambda}_x(v) \le \lambda_{i+1}$, forcing $j \le i$. Hence $j = i$ by injectivity of the strictly antitone list. ☐

Theorem 4.21 (*A*-equivariance of spectrum and flag, a.e.)**.** *Under the standing hypotheses, for a.e. x the spectrum is invariant,* $\mathrm{lyapunovSpectrum}(A, T, x) = \mathrm{lyapunovSpectrum}(A, T, Tx)$ *(`lyapunovSpectrum_equivariant_ae`), and the action of Ax maps each flag level (each sublevel subspace) at x onto the corresponding level at Tx:*

$$(Ax)_* \,\mathrm{lambdaSublevel}(A, T, x, t) = \mathrm{lambdaSublevel}(A, T, Tx, t).$$

Proof. The bijection $v \mapsto Ax{\cdot}v$ preserves $\bar{\lambda}$ by 4.13 (a.e.), hence carries witnesses at x to witnesses at Tx and conversely (using $(Ax)^{-1}$), giving the spectrum identity. The same value-preserving bijection sends $\{v = 0 \vee \bar{\lambda}_x(v) \le t\}$ onto $\{w = 0 \vee \bar{\lambda}_{Tx}(w) \le t\}$, which is the claimed image of sublevel sets. ☐

4.5 Measurability of the filtration

The Oseledets theorem requires the flag to vary measurably in x. Mathlib has no measurable structure on submodules, so we encode a subspace by its orthogonal-projection matrix.

Definition 4.22 (Projection matrix encoding)**.** For $K \le \mathbb{R}^d$, $\mathrm{orthProjMatrix}(K)$ is the matrix of the orthogonal projection onto K, namely the preimage of K.`starProjection` under the star-algebra isomorphism `toEuclideanCLM`. A subspace is determined by this matrix, which lives in a space carrying the Borel/Pi measurable structure. Its (i, j) entry equals the i-th coordinate of the projection applied to the standard basis vector e_j (`orthProjMatrix_apply`).

Definition 4.23 (Measurable family of subspaces)**.** A subspace-valued map $V : X \to \mathrm{Submodule}\,\mathbb{R}\,\mathbb{R}^d$ is a *measurable family of subspaces* when $x \mapsto \mathrm{orthProjMatrix}(Vx)$ is measurable. Equivalently (`measurable_orthProjMatrix_iff`), for each standard basis index j the $\mathbb{R}^d$-valued map $x \mapsto (Vx)$.`starProjection`(e_j) is measurable.

Lemma 4.24 (Scalar growth is measurable)**.** *For fixed v, the map $x \mapsto \bar{\lambda}_x(v)$ is measurable.*

Proof. It is the lim sup of the sequence $x \mapsto \frac{1}{n}\log\left\|A^{(n)}(x)\cdot v\right\|$. Each term is measurable: $x \mapsto A^{(n)}(x)$ is measurable (measurability of the cocycle), and $M \mapsto \|M \cdot v\|$ is continuous (a fixed-vector linear map of M, on a finite-dimensional space, post-composed with the norm), so the composite with log is measurable, and a lim sup of measurable functions is measurable. □

Lemma 4.25 (Polynomial in a measurable matrix)**.** *For a fixed real polynomial q, the map $a \mapsto q(a)$ on $\mathrm{Mat}_{d\times d}(\mathbb{R})$ is measurable.*

Proof. Induction on q over the constant/sum/monomial generators, using that matrix addition and multiplication are measurable in each argument (`instMeasurableAdd Matrix` and the matrix `MeasurableMul` instance), whence $a \mapsto a^n$ is measurable (`measurable_matrix_pow`). □

Theorem 4.26 (CFC measurability via interpolating polynomial)**.** *Let $M : X \to \mathrm{Mat}_{d\times d}(\mathbb{R})$ be measurable with each Mx self-adjoint, and let $g : \mathbb{R} \to \mathbb{R}$. If a* fixed *polynomial q agrees with g on the spectrum of every Mx, then $x \mapsto \mathrm{cfc}\, g\,(Mx)$ is measurable.*

Proof. On the spectrum of Mx the continuous functional calculus of g coincides with that of q, and for a polynomial $\mathrm{cfc}\, q\,(Mx) = q(Mx)$. Thus pointwise $\mathrm{cfc}\, g\,(Mx) = q(Mx)$, which is measurable in x by 4.25. This uses only the bare Hermitian CFC instance, avoiding the isometric CFC (absent for real matrices) and any measurable selection. □

Theorem 4.27 (CFC measurability for continuous functions)**.** *Let M be measurable with each Mx self-adjoint, and $f : \mathbb{R} \to \mathbb{R}$ continuous. Then $x \mapsto \mathrm{cfc}\, f\,(Mx)$ is measurable.*

Proof. A single polynomial need not agree with f on the unbounded family of spectra, so approximate per point: by Weierstrass choose, for each k, a polynomial q_k with $|q_k - f| \le 1/(k{+}1)$ on $[-k, k]$. Each spectrum is finite, hence in some $[-R, R]$, so $q_k \to f$ uniformly on $\mathrm{spectrum}(Mx)$ and $\mathrm{cfc}\, q_k\,(Mx) \to \mathrm{cfc}\, f\,(Mx)$. Each $x \mapsto \mathrm{cfc}\, q_k\,(Mx) = q_k(Mx)$ is measurable, and matrix-entrywise the metrizable limit upgrades to measurability of $x \mapsto \mathrm{cfc}\, f\,(Mx)$. □

These CFC tools deliver MeasurableSubspace for the concrete Oseledets flag once the Oseledets limit operator $\Lambda x = \lim_n \left((A^{(n)})^\top A^{(n)}\right)^{1/(2n)}$ is available: each flag projection is realized as a spectral band projector $P_i x = \mathrm{cfc}\, g_i\,(\Lambda x)$ of Λx for a continuous gap function g_i, so the projection matrix of its range equals $P_i x$ definitionally, and measurability of $x \mapsto P_i x$ follows from 4.27 (or, on the gapped good set where the spectrum is the fixed Lyapunov set, from 4.26). This is the measurability input feeding .

Chapter 5

The one-sided multiplicative ergodic theorem

This is the crux of the development. We work over a probability space (X, μ) with an ergodic measure-preserving transformation T and a measurable cocycle generator $A : X \to \mathrm{Mat}_{d\times d}(\mathbb{R})$ with $\det(Ax) \neq 0$, subject to the one-sided integrability $\log^+ \|A\|, \log^+ \|A^{-1}\| \in L^1(\mu)$. The matrices act on $\mathbb{R}^d$ (the Euclidean space) through the operator $v \mapsto Av$, so that the relevant norm is the L^2 operator norm, which is submultiplicative; we write $A^{(n)}(x)$ for the cocycle $A(T^{n-1}x) \cdots A(x)$. The whole chapter is organized around a single positive semidefinite limiting matrix, the *Oseledets limit*

$$\Lambda(x) \;=\; \lim_{n\to\infty} \big((A^{(n)}(x))^\top A^{(n)}(x)\big)^{1/2n},$$

whose eigenspaces, once their growth rates are matched to the deterministic singular-value exponents, yield the filtration of the target theorem.

5.1 The Oseledets limit

The candidate approximants are the symmetric positive roots $q_n(x) := \big((A^{(n)}(x))^\top A^{(n)}(x)\big)^{1/2n}$, realized through the continuous functional calculus on the Gram matrix. The eigenvalues of $q_n(x)$ are the $1/n$-th roots of the singular values of $A^{(n)}(x)$; the heart of this section is that they converge a.e. to the exponentials of the deterministic Lyapunov exponents.

Theorem 5.1 (Deterministic singular-value exponents)**.** *There is an antitone sequence $\lambda^0 : \mathbb{N} \to \mathbb{R}$ (antitone on $[0, d)$) such that, for every $i < d$ and μ-a.e. x,*

$$\tfrac{1}{n} \log(\sigma_i(A^{(n)}(x))) \longrightarrow \lambda_i^0,$$

where σ_i is the i-th sorted singular value.

Proof. Package the ergodic limits $\Gamma_k = \lim \frac{1}{n} \log s_k(A^{(n)}(x))$ of the products of the top k singular values, $0 \le k \le d$, obtained from Kingman's subadditive ergodic theorem applied to the subadditive cocycle $\log\left\|\bigwedge^k A^{(n)}\right\|$ (the exterior-power functor turns submultiplicativity of the $\bigwedge^k$ operator norms into subadditivity). Then $\lambda_i^0 := \Gamma_{i+1} - \Gamma_i$; the difference of the two a.e. Γ-limits gives the i-th singular-value exponent. Antitonicity of consecutive λ_i^0 descends from the antitone ordering of the singular values inside each n, and chaining yields full antitonicity on $[0, d)$. $\square$

Lemma 5.2 (Block-value step function reproduces the spectrum)**.** *Let* stepVal λ^0 D *be the step function* $e^{\lambda^0_{D-1}} + \sum_{k=1}^{D-1}(e^{\lambda^0_{k-1}} - e^{\lambda^0_k})\,\mathbf{1}_{(c_k,\infty)}$ *with thresholds* $c_k = e^{(\lambda^0_k+\lambda^0_{k-1})/2}$ *strictly inside the* k*-th gap. If* λ^0 *is antitone on* $[0, D)$ *and* $j < D$*, then* stepVal λ^0 $D\,(e^{\lambda^0_j}) = e^{\lambda^0_j}$*.*

Proof. At the argument $e^{\lambda^0_j}$ the threshold indicator $\mathbf{1}_{(c_k,\infty)}$ is 1 exactly when $k > j$ (since λ^0 is antitone and c_k lies strictly between λ^0_k and λ^0_{k-1}), so only the increments above index j survive. Those increments $e^{\lambda^0_{k-1}} - e^{\lambda^0_k}$ telescope to $e^{\lambda^0_j} - e^{\lambda^0_{D-1}}$, which added to the constant base $e^{\lambda^0_{D-1}}$ returns $e^{\lambda^0_j}$. ☐

Lemma 5.3 (Spectral deviation bound)**.** *For a self-adjoint matrix* M *and any function* g*,*

$$\|M - g(M)\| \;\le\; \sum_j \left|\mu_j - g(\mu_j)\right|,$$

where μ_j *ranges over the sorted eigenvalues of* M*.*

Proof. Writing $M = \mathrm{id}(M)$ gives $M - g(M) = (\mathrm{id} - g)(M)$ by linearity of the continuous functional calculus, and the operator norm of a self-adjoint matrix functional calculus is the largest absolute eigenvalue $\max_j \left|\mu_j - g(\mu_j)\right|$, which is bounded by the full nonnegative sum. ☐

Lemma 5.4 (Per-term band-projector convergence)**.** *For* μ*-a.e.* x *and every threshold index* $k \in [1, d)$*, the block term* $(e^{\lambda^0_{k-1}} - e^{\lambda^0_k}) \cdot P_n^{c_k}(x)$ *converges, where* P_n^c *is the band projector* $\mathbf{1}_{(c,\infty)}(q_n(x))$*.*

Proof. At a genuine gap $\lambda^0_k < \lambda^0_{k-1}$ the threshold c_k is strictly separated from the two limiting eigenvalue clusters, so once the sorted eigenvalues of $q_n(x)$ have converged the count of eigenvalues above c_k stabilizes and the corresponding spectral projector is Cauchy in operator norm. At a non-gap ($\lambda^0_{k-1} = \lambda^0_k$) the coefficient $e^{\lambda^0_{k-1}} - e^{\lambda^0_k}$ vanishes, so the term is constantly 0. ☐

Theorem 5.5 (Existence of the Oseledets limit)**.** *For* μ*-a.e.* x *the approximants* $q_n(x)$ *converge in the matrix metric to a single matrix* $\Lambda(x)$*.*

Proof. The eigenvalues $\mu_{j,n} = \sigma_j^{1/n}$ of $q_n(x)$ converge a.e. to the exponentials $e^{\lambda^0_j}$ (Theorem 5.1). Form the block approximant $\Lambda_n(x) :=$ stepVal $\lambda^0\,d\,(q_n(x))$, a finite linear combination of band projectors. By Lemma 5.3, $\|q_n(x) - \Lambda_n(x)\| \le \sum_j \left|\mu_{j,n} - \mathrm{stepVal}(\mu_{j,n})\right|$; each summand is eventually $\left|\mu_{j,n} - e^{\lambda^0_j}\right| \to 0$ because the step function reproduces the exponentials on the spectrum (Lemma 5.2). Meanwhile $\Lambda_n(x)$ converges as a finite sum of convergent band-projector terms (Lemma 5.4). Adding the two convergences gives $q_n(x) \to \Lambda(x)$, and the limit is selected pointwise. ☐

Definition 5.6 (The named Oseledets limit)**.** $\Lambda(x) := (\,\lim_n (q_n(x))_{ij}\,)_{ij}$ is the entrywise real lim sup/limit of the matrix entries of $q_n(x)$; it is a total, measurable function of x.

Theorem 5.7 (The limit is the a.e. limit of the approximants)**.** *For* μ*-a.e.* x*,* $q_n(x) \to \Lambda(x)$ *in the matrix metric, and* Λ *is measurable.*

Proof. On the a.e. full convergence set of Theorem 5.5 the entrywise limit recovers the matrix limit (matrix convergence in finite dimensions is entrywise), so the entrywise lim inf defining Λ equals the genuine limit. Measurability is entrywise: each entry is a limit of measurable functions of x, and a lim inf of measurable $\mathbb{R}$-valued functions is measurable. ☐

Proposition 5.8 (Structure of the limit)**.** *For μ-a.e. x, $\Lambda(x)$ is self-adjoint and positive semidefinite.*

Proof. Self-adjointness $M^\top = M$ is an entrywise closed condition preserved under the matrix limit of the self-adjoint approximants $q_n(x)$. For positive semidefiniteness, the quadratic form $M \mapsto v^\top M v$ is continuous, so $v^\top \Lambda(x) v = \lim_n v^\top q_n(x) v \geq 0$ as a limit of nonnegatives. $\square$

5.2 The per-vector lower bound

The lower half of the exact growth law isolates one band of the spectrum of $q_n(x)$ and shows that a vector with nonzero projection onto the band grows at least at the band rate.

Lemma 5.9 (Gram quadratic-form band bound)**.** *For self-adjoint Q, a band indicator $\chi = \mathbf{1}_{(c,\infty)}$, and a continuous $f \geq 0$ on $\operatorname{spec}(Q)$ with $a \leq f(t)$ whenever $c < t$,*

$$a \, \|\chi(Q)\, v\|^2 \;\leq\; \langle f(Q)\, v,\, v\rangle.$$

Proof. The band projector $\chi(Q)$ is a self-adjoint idempotent, so $\|\chi(Q)v\|^2 = \langle \chi(Q)v, v\rangle$. The gap operator $(f - a\chi)(Q)$ is positive semidefinite because $f - a\chi \geq 0$ on the spectrum (above c, $f \geq a = a\chi$; below, $\chi = 0$ and $f \geq 0$). Expanding $\langle (f - a\chi)(Q)v, v\rangle \geq 0$ gives the claim. $\square$

Lemma 5.10 (Band lower bound for the cocycle)**.** *For $c \geq 0$ and $n \geq 1$,*

$$c^{2n} \, \|P_n^c(x)\, v\|^2 \;\leq\; \left\|A^{(n)}(x)\, v\right\|^2.$$

Proof. Raising $q_n(x) = (\mathrm{gram}_n)^{1/2n}$ to the $2n$-th power via the functional calculus recovers the Gram matrix $\mathrm{gram}_n = (A^{(n)})^\top A^{(n)}$ (the composed powers compose to the identity on the nonnegative spectrum). Apply Lemma 5.9 with $f(t) = t^{2n}$ and $a = c^{2n}$: above c one has $t^{2n} \geq c^{2n}$. The right-hand inner product is then $\langle \mathrm{gram}_n v, v\rangle = \left\|A^{(n)}(x)v\right\|^2$. $\square$

Lemma 5.11 (The band correction vanishes)**.** *If $P_n^c(x) \to P$ with $Pv \neq 0$, then $\frac{1}{n}\log\|P_n^c(x)\, v\| \to 0$.*

Proof. The evaluation $M \mapsto Mv$ is continuous in finite dimensions, so $P_n^c(x)v \to Pv \neq 0$ and $\|P_n^c(x)v\| \to \|Pv\| > 0$. Hence the log converges to the finite number $\log\|Pv\|$, and dividing by $n \to \infty$ sends it to 0. $\square$

Proposition 5.12 (Per-vector liminf lower bound)**.** *If $P_n^c(x) \to P$ with $c > 0$ and $Pv \neq 0$, and the cocycle growth sequence is cobounded, then*

$$\log c \;\leq\; \liminf_n \; \tfrac{1}{n}\log\left\|A^{(n)}(x)\, v\right\|.$$

Proof. Taking logs in Lemma 5.10 and dividing by $2n$ gives, eventually,

$$\log c + \tfrac{1}{n}\log\|P_n^c(x)v\| \;\leq\; \tfrac{1}{n}\log\left\|A^{(n)}(x)v\right\|.$$

The left side converges to $\log c$ since the band-correction term vanishes (Lemma 5.11). Passing to the $\liminf$ along the inequality, using its boundedness and the coboundedness of the right side (supplied by the Furstenberg–Kesten integrability of the top exponent), yields the bound. $\square$

Lemma 5.13 (Band-projector nesting, kernel propagation)**.** *For thresholds $c \leq c'$ with limit band projectors P, P', if $Pv = 0$ then $P'v = 0$.*

Proof. The finite-n bands are nested: $\mathbf{1}_{(c,\infty)} \cdot \mathbf{1}_{(c',\infty)} = \mathbf{1}_{(c',\infty)}$ on the spectrum since $(c', \infty) \subseteq (c, \infty)$, so $P_n^c P_n^{c'} = P_n^{c'}$; passing to the limit gives $PP' = P'$. Both limit projectors are symmetric (limits of self-adjoint matrices), so transposing gives $P'P = P'$, whence $P'v = P'(Pv) = 0$. $\square$

5.3 The spectral upper bound

The upper half is the genuinely non-elementary step. On the ultrametric-growth good set the per-vector upper growth exponent $\overline{\lambda}(x, v)$ of a vector v in the i-th stratum equals its spectral value specList i exactly, so in particular $\limsup_n \frac{1}{n} \log \|A^{(n)}(x)v\| \leq \text{specList}\, i$. This replaces an earlier determinant/volume-squeeze argument — the two *superseded-route* lemmas below, whose restricted-operator-norm files were removed from the development as dead code; their labels are retained only so cross-references resolve.

Lemma 5.14 (Tempering)**.** *If T is measure-preserving and $g \in L^1(\mu)$, then for μ-a.e. x, $\frac{1}{n}\, g(T^n x) \to 0$.*

Proof. The series $\sum_n g(T^n x)/n^2$ is a.e. finite by integrability and invariance of μ, so its terms $g(T^n x)/n^2 \to 0$; the Borel–Cantelli/sublinear-growth argument then gives $g(T^n x) = o(n)$, i.e. $\frac{1}{n} g(T^n x) \to 0$. ☐

Lemma 5.15 (Slow-volume exponent squeeze (superseded route))**.** This lemma belonged to a superseded determinant/volume-squeeze route and has no counterpart in the current Lean development; the statement is kept for exposition and the label only so that cross-references resolve. *Given a volume cocycle whose top, slow and remaining log-exponent sequences satisfy a sum law* vol = slow + rest *with the appropriate limits, the slow restricted-operator-norm exponent obeys* $\limsup_n \frac{1}{n} \log(\text{slow}_n) \leq \lambda_i$.

Proof. The total volume exponent is the Furstenberg–Kesten determinant limit $\sum_j \lambda_j^0$; the fast-block volume is the exterior-power Kingman limit; their difference forces the slow-block volume exponent. The squeeze converts a sum identity of limits into an upper bound for the slow factor once the fast and remaining factors are pinned, the angle/tilt between fast and slow blocks tempering to zero so no cross term inflates the slow volume. ☐

Theorem 5.16 (Spectral upper bound on each stratum)**.** *For μ-a.e. x, every index i and every vector v in the i-th stratum* vflag i.castSucc $\setminus$ vflag i.succ *one has*

$$\limsup_n \ \tfrac{1}{n} \log \|A^{(n)}(x)\, v\| \ \leq \ \text{specList}\, i.$$

Proof. On the IsUltrametricGrowth good set the per-vector upper growth function $\overline{\lambda}(x, v)$ equals the exact stratum exponent specList i; since this $\overline{\lambda}$ is by definition the lim sup of $\frac{1}{n} \log \|A^{(n)} v\|$, the bound follows (indeed with equality). The good set itself is built from the subexponential one-step log-norm factor supplied by the tempering estimate (Lemma 5.14), so the argument is non-circular. ☐

Theorem 5.17 (Upper bound on the slow space)**.** *On the ultrametric-growth good set, every vector v of the Λ-slow band* vslow(e^t) *with* $\limsup_n \frac{1}{n} \log \|A^{(n)} v\| \leq t$ *lies in the growth sublevel* $\{v : \overline{\lambda}(x, v) \leq t\}$.

Proof. The lim sup of $\frac{1}{n} \log \|A^{(n)} v\|$ is by definition the upper growth function $\overline{\lambda}(x, v)$; the hypothesis says it is $\leq t$, which is exactly membership in the sublevel lambdaSublevel t. The zero vector lies in every submodule. ☐

5.4 Spectral identification of the filtration

The two bounds match the spectral filtration of Λ with the analytic limsup filtration. The bridge is that the finite-n band projectors converge a.e. to the functional calculus indicator of Λ.

Theorem 5.18 (Band projectors converge to the CFC indicator)**.** *For μ-a.e. x, every $c > 0$ that is not one of the limiting eigenvalues $e^{\lambda_{\text{sing}}(x,i)}$ of $\Lambda(x)$ satisfies*

$$P_n^c(x) \longrightarrow \mathbf{1}_{(c,\infty)}(\Lambda(x)).$$

Proof. Since c avoids the spectrum of $\Lambda(x)$, there is a gap $\delta > 0$ between c and every eigenvalue. Replace the discontinuous indicator by a continuous $\delta/2$-clamp surrogate χ that is Lipschitz and agrees with $\mathbf{1}_{(c,\infty)}$ at distance $\geq \delta/2$ from c. Once the sorted eigenvalues of $q_n(x)$ are within $\delta/2$ of their limits, χ and $\mathbf{1}_{(c,\infty)}$ agree on both spectra, so $P_n^c(x) = \chi(q_n(x))$ eventually; by Lipschitz continuity of the functional calculus and $q_n(x) \to \Lambda(x)$ (Theorem 5.7), $\chi(q_n(x)) \to \chi(\Lambda(x)) = \mathbf{1}_{(c,\infty)}(\Lambda(x))$. □

Theorem 5.19 (Reverse slow-flag inclusion)**.** *For μ-a.e. x and every t,* $\text{lambdaSublevel}(x,t) \leq \text{vslow}(x, e^t)$.

Proof. Contrapositively, a vector $v \notin \text{vslow}(e^t)$ has nonzero component in the band of $\Lambda(x)$ above e^t; by Theorem 5.18 the finite band projectors converge to the corresponding CFC indicator with $Pv \neq 0$, so the per-vector lower bound (Proposition 5.12) forces $\liminf \frac{1}{n} \log \|A^{(n)} v\| > t$, hence $\overline{\lambda}(x,v) > t$ and $v \notin \text{lambdaSublevel}\, t$. Kernel propagation across nested thresholds (Lemma 5.13) makes this consistent for all t. □

Theorem 5.20 (The slow flag equals the limsup sublevel)**.** *Under the spectral upper bound and the reverse inclusion, for μ-a.e. x and every t,*

$$\text{vslow}(x, e^t) \;=\; \text{lambdaSublevel}(x,t).$$

Proof. Two inclusions: the forward one (Theorem 5.17, from the upper bound) shows slow vectors grow slowly, hence lie in the sublevel; the reverse one (Theorem 5.19, from the lower bound) shows slowly-growing vectors lie in the slow space. Antisymmetry gives the identity, simultaneously for all t on the a.e. ultrametric-growth good set. □

5.5 Ruelle's reverse cofactor bound and the top-gap envelope

The upper bound on an individual vector reduces, after diagonalizing the limit, to controlling the overlap matrix between the sorted Gram eigenbasis at level n and the limiting eigenbasis. The graded overlap is controlled by a leakage induction; Ruelle's cofactor estimate then converts a one-sided forward decay into the full pairwise rate.

Lemma 5.21 (Ruelle reverse cofactor bound)**.** *Let S be orthogonal ($SS^\top = 1$) with the graded forward decay $|S_{ab}| \leq c \cdot e^{-\max(g_b - g_a, 0)}$. Then every entry obeys the reverse bound at the full pairwise rate:*

$$|S_{ij}| \;\leq\; (d-1)!\, c^{d-1}\, e^{-(g_i - g_j)}.$$

Proof. Since $S^{-1} = S^\top$, the entry S_{ij} is $(\det S)^{-1}$ times the cofactor $\text{adj}(S)_{ji}$, and $|\det S| = 1$ because $SS^\top = 1$. Expanding the minor by the Leibniz formula, every surviving permutation term collects the level imbalance $g_i - g_j$ by telescoping the forward factors, each forward factor contributing at most $c\, e^{-\max(\cdot,0)}$; there are at most $(d-1)!$ such terms. □

Definition 5.22 (Top-gap fast-band-mass envelope)**.** TopGapMassEnvelope $A\,T\,\lambda^0\,x$ asserts the uniform geometric leakage of fast-band mass across each genuine gap: for every cut tolerance δ there is a constant C controlling, eventually and uniformly, the band mass that crosses the top gap of each stratum.

Lemma 5.23 (Multi-source geometric envelope)**.** *If a nonnegative sequence a obeys a one-step recursion fed by finitely many source sequences each decaying geometrically with ratio $\rho < 1$, then a_n is bounded by a fixed multiple of the summed source envelopes for all n.*

Proof. Each single source contributes a geometric partial sum bounded by $K/(1-\rho)$; summing the finitely many per-source envelopes and folding them through the linear one-step recursion gives a uniform bound on the chained quantity a_n. □

Lemma 5.24 (Per-stratum envelope step)**.** *At a fixed cut strictly inside a gap of width $\geq G$, the one-step band-mass increment is the current mass damped by e^{-G} plus a tempered source term; iterating produces the per-stratum leakage envelope.*

Proof. In the sorted-Gram-eigenbasis block decomposition, the band mass above the cut at step $n+1$ is the mass at step n attenuated by the singular-value ratio across the gap (bounded by e^{-nG}-type damping after n iterates) plus the contribution injected by the one-step generator $A(T^n x)$, whose log-norm is tempered to $o(n)$. This is exactly the geometric one-step recursion fed by tempered sources of Lemma 5.23. □

Theorem 5.25 (The top-gap envelope, a.e.)**.** *For μ-a.e. x, the top-gap fast-band-mass envelope* TopGapMassEnvelope $A\,T\,\lambda^0\,x$ *holds.*

Proof. Fix the deterministic distinct gap $G > 0$ separating distinct exponents. On the a.e. set where every singular-value exponent converges (Theorem 5.1) and the one-step generator log-norm is tempered (Lemma 5.14), build the per-stratum leakage envelope for each gap pair (Lemma 5.24) and assemble them into the top-gap envelope: each stratum's fast-band mass crossing its top gap is geometrically damped, uniformly in the cut, by a single per-pair constant maximized over the finitely many pairs. □

5.6 Constancy of the spectrum

The deterministic exponent set is constant in x by construction, so once the per-point spectrum is identified with it, ergodic constancy is automatic.

Lemma 5.26 (Spectrum identity from two inclusions)**.** *If at x every realized exponent is a deterministic one and every deterministic exponent is attained, then* $\mathrm{lyapunovSpectrum}(x) = \mathrm{distinctExp}\,\lambda^0\,d$.

Proof. Both directions are finite-set inclusions; antisymmetry of $\subseteq$ gives the equality of the two finite subsets of $\mathbb{R}$. □

Theorem 5.27 (Ergodic constancy of the spectrum)**.** *Given a.e. that every realized value of the upper growth function is a deterministic exponent (upper inclusion) and every deterministic exponent is attained (lower inclusion), for μ-a.e. x,*

$$\mathrm{lyapunovSpectrum}(x) \;=\; \mathrm{distinctExp}\,\lambda^0\,d.$$

Proof. On the ultrametric-growth good set the two Finset inclusions are equivalent to native statements about the upper growth function $\overline{\lambda}(x, \cdot)$: the upper inclusion is the spectral-upper-bound output (each stratum value is a deterministic exponent), the lower inclusion is attainment from the lower bound. Lemma 5.26 then gives the identity with the deterministic constant set; since that set does not depend on x, the identification is T-invariant and the spectrum is a.e. constant — ergodicity adds nothing beyond the deterministic value already in hand. ☐

5.7 Assembling the target theorem

The forward graded-overlap bound (built from the top-gap envelope and Ruelle's reverse bound) gives the spectral upper bound hupper; combined with the reverse inclusion it yields the slow flag, and the lower bound supplies attainment. Together they discharge the three a.e. interfaces (spectrum, slow flag, exact growth) of the filtration assembly.

Theorem 5.28 (Filtration from the spectral upper bound)**.** *Assume the per-vector spectral upper bound on the slow flag, the reverse slow-flag inclusion, the two spectrum inclusions, and the band-projector convergence datum. Then there exist k, strictly decreasing $\lambda : \mathrm{Fin}\, k \to \mathbb{R}$, and a measurable family V forming a.e. a strictly decreasing A-equivariant flag along which $\frac{1}{n} \log \|A^{(n)}(x)v\| \to \lambda_i$ on each stratum.*

Proof. The deterministic exponents λ^0 come from Theorem 5.1. The spectrum interface is discharged by the two inclusions (constancy, Theorem 5.27); the slow-flag interface by vslow = lambdaSublevel (Theorem 5.20); the exact-growth interface by combining the unconditional upper half (the stratum value $\overline{\lambda} = \lambda_i$ on vflag), the lower half (Proposition 5.12 fed the band datum), and Furstenberg–Kesten boundedness into a two-sided limit. These three interfaces feed the generic slow-flag assembly. ☐

Theorem 5.29 (Filtration from the top-gap envelope)**.** *Under the standing ergodic, invertible, log-integrable hypotheses, and assuming the top-gap envelope* TopGapMassEnvelope *quantified over λ^0, the full Oseledets filtration conclusion holds.*

Proof. Diagonalize $\Lambda(x)$ by its limit eigenbasis with eigenvalues $e^{\lambda_{\text{sing}}}$ and slow-orthogonality. The forward graded-overlap bound, consuming the envelope (Definition 5.22), yields the one-sided forward decay of the overlap matrix between the sorted Gram eigenbasis and the limit eigenbasis; Ruelle's reverse cofactor estimate (Lemma 5.21) upgrades this to the full pairwise rate, which is exactly the slow-restriction bound feeding the spectral upper bound hupper on the limit slow space. The band-projector convergence (Theorem 5.18) supplies the reverse slow-flag inclusion and the lower-bound datum, and the spectrum inclusions follow from the slow-flag identity. These discharge all hypotheses of Theorem 5.28. ☐

Theorem 5.30 (One-sided Oseledets multiplicative ergodic theorem)**.** *Let μ be a probability measure, $T : X \to X$ ergodic measure-preserving, and $A : X \to \mathrm{Mat}_{d\times d}(\mathbb{R})$ measurable with $\det(Ax) \neq 0$ and $\log^+ \|A\|, \log^+ \|A^{-1}\| \in L^1(\mu)$. Then there are k distinct Lyapunov exponents $\lambda : \mathrm{Fin}\, k \to \mathbb{R}$, strictly decreasing, and a measurable family of subspaces*

$$V : \mathrm{Fin}(k+1) \to X \to \mathrm{Submodule}_{\mathbb{R}}(\mathbb{R}^d)$$

with each $x \mapsto V_i\, x$ measurable, such that for μ-a.e. x: $V_0\, x = \top$, $V_k\, x = \bot$; the flag is strictly decreasing, $V_{i+1}\, x < V_i\, x$; it is A-equivariant, $A(x)\, V_i\, x = V_i\, (Tx)$; and along it the cocycle grows at the exact rate λ_i:

$$\frac{1}{n} \log \|A^{(n)}(x)\, v\| \longrightarrow \lambda_i \qquad \text{for every } v \in V_i\, x \setminus V_{i+1}\, x.$$

Proof. If $d = 0$ the trivial flag $\top = \bot$ with no exponents discharges the statement. For $d > 0$, the top-gap envelope holds a.e. (Theorem 5.25), so the conditional assembly Theorem 5.29 applies directly and produces the exponents, the measurable equivariant flag, and the exact per-stratum growth limits. □

Chapter 6

Companion results and extensions

The Oseledets multiplicative ergodic theorem `oseledets_filtration` delivers, μ-a.e., a strictly decreasing A-equivariant flag $\mathbb{R}^d = V_0(x) \supsetneq \cdots \supsetneq V_k(x) = 0$ together with a strictly decreasing exponent list $\lambda_0 > \cdots > \lambda_{k-1}$ governing the growth $\frac{1}{n} \log \|A^{(n)}(x)v\| \to \lambda_i$ on each stratum $V_i \setminus V_{i+1}$. This chapter collects its companion results. Many are surprisingly cheap: they follow from the *statement* of the main theorem, quantified over an arbitrary witness of its conclusion, with no access to the construction. To make this precise we bundle the conclusion as a predicate.

6.1 The bundled predicate and uniqueness

Definition 6.1 (Bundled Oseledets filtration)**.** For a measure μ, map T, generator A, count k, exponent list $\lambda : \mathrm{Fin}\, k \to \mathbb{R}$ and flag $V : \mathrm{Fin}(k+1) \to X \to \mathrm{Submodule}$, the predicate $\mathrm{IsOseledetsFiltration}\ \mu\ T\ A\ k\ \lambda\ V$ asserts: λ is strictly antitone; each level V_i is a measurable subspace family; and μ-a.e. x carries the strictly decreasing A-equivariant flag $\mathbb{R}^d = V_0(x) \supsetneq \cdots \supsetneq V_k(x) = 0$ with exact growth rate λ_i on the stratum $V_i \setminus V_{i+1}$. This is byte-identical to the conclusion of the main theorem.

Theorem 6.2 (Repackaged existence)**.** *Under the standing hypotheses (ergodic T, invertible measurable A with $\log^+ \|A\|, \log^+ \|A^{-1}\| \in L^1$), there exist k, λ, V with* $\mathrm{IsOseledetsFiltration}\ \mu\ T\ A\ k\ \lambda\ V$.

Proof. Destructure the conclusion of `oseledets_filtration` and repackage its three conjuncts as the bundled predicate; the data match definitionally. ☐

Theorem 6.3 (Canonical sublevel characterization)**.** *If* $\mathrm{IsOseledetsFiltration}\ \mu\ T\ A\ k\ \lambda\ V$ *holds, then μ-a.e. each interior flag level is exactly a growth-sublevel set: for every i and vector v,*

$$v \in V_i(x) \iff v = 0 \ \vee\ \limsup_n \tfrac{1}{n} \log \|A^{(n)}(x)v\| \le \lambda_i.$$

Proof. At a good point pick the stratum index j of $v \neq 0$; the per-stratum convergence gives $\limsup = \lambda_j$. Membership in V_i forces $i \le j$, so by antitonicity $\lambda_j \le \lambda_i$; conversely if $v \notin V_i$ then $j < i$ and $\lambda_j > \lambda_i$ strictly, contradicting the bound. The disjunct $v = 0$ handles $\log 0$. No machinery beyond the a.e. block is used. ☐

Theorem 6.4 (Uniqueness of the spectrum and filtration)**.** *On a probability space, any two Oseledets filtration data (k, λ, V) and (k_2, λ_2, V_2) for the same cocycle agree: $k = k_2$, the exponents coincide under the index cast, and μ-a.e. the flags agree level by level.*

Proof. At a single good point the set of realized growth limits equals range λ and range λ_2; two strictly antitone enumerations of one finite set coincide, giving $k = k_2$ and $\lambda = \lambda_2$. Levelwise identity then follows from the sublevel characterization 6.3 applied to both data. □

6.2 The top exponent as operator-norm growth

Lemma 6.5 (Nontriviality). *On a probability space with $0 < d$, any Oseledets filtration has $0 < k$.*

Proof. If $k = 0$ then at a good point $\mathbb{R}^d = V_0(x) = V_{\text{last}}(x) = 0$, forcing finrank $= 0$, contradicting $d > 0$. □

Theorem 6.6 (Top exponent = norm growth). *On a probability space, with A invertible and $0 < k$, μ-a.e. the operator-norm growth rate of the cocycle converges to the top exponent:*

$$\tfrac{1}{n}\log\|A^{(n)}(x)\| \to \lambda_0.$$

Proof. Two-sided squeeze from the flag block. *Lower:* a vector v in the top stratum has $\frac{1}{n}\log\|A^{(n)}v\| \to \lambda_0$ and $\|A^{(n)}v\| \le \|A^{(n)}\|\,\|v\|$. *Upper:* the column-sum bound $\|M\| \le \sum_j \|Me_j\|$ on the L^2 operator norm, each basis vector being nonzero, gives eventually $\frac{1}{n}\log\|A^{(n)}\| \le \lambda_0 + \varepsilon$. Neither Furstenberg–Kesten nor singular values are needed. □

Corollary 6.7 (Identification of the Furstenberg–Kesten constant). *Any constant c to which $\frac{1}{n}\log\|A^{(n)}(x)\|$ converges μ-a.e. (e.g. the Furstenberg–Kesten constant) equals λ_0.*

Proof. Both 6.6 and the hypothesis hold at a common good point; uniqueness of limits gives $\lambda_0 = c$. □

6.3 A.e.-constant multiplicities

Theorem 6.8 (Deterministic dimension profile). *For ergodic T and invertible A, every Oseledets filtration has a deterministic dimension profile: there is a strictly decreasing $m : \mathrm{Fin}(k+1) \to \mathbb{N}$ with $m_0 = d$, $m_k = 0$ and, μ-a.e.,* finrank $V_i(x) = m_i$.

Proof. The dimension $x \mapsto$ finrank $V_i(x)$ is measurable via the trace of the orthogonal projector, and T-invariant a.e. because equivariance through the invertible $A(x)$ preserves dimension. Ergodicity makes each invariant $\mathbb{N}$-valued function a.e. constant. The profile structure (StrictAnti, endpoints) is read off one good point. □

Corollary 6.9 (Per-exponent multiplicities). *For ergodic T and invertible A, each exponent λ_i carries a positive deterministic multiplicity $m_i = \dim V_i - \dim V_{i+1}$ with $\sum_i m_i = d$.*

Proof. Set m_i to the consecutive dimension drops of 6.8; positivity is strict antitonicity, and the telescoping sum equals $m_0 - m_k = d$. □

Theorem 6.10 (MET with multiplicities). *Under the standing hypotheses there exist k, λ, V and a strictly decreasing m with $m_0 = d$, $m_k = 0$, such that* IsOseledetsFiltration $\mu\ T\ A\ k\ \lambda\ V$ *holds and μ-a.e.* finrank $V_i(x) = m_i$.

Proof. Obtain a witness from 6.2 and apply 6.8 to it. □

6.4 The Lyapunov spectrum

Definition 6.11 (Sorted spectrum)**.** The full Lyapunov spectrum with multiplicity is the total function exponents : $\mathrm{Fin}\, d \to \mathbb{R}$, whose i-th entry is the deterministic limit of $\frac{1}{n} \log \sigma_i(A^{(n)})$, sorted non-increasingly. The top entry is topExponent.

Theorem 6.12 (Defining σ-limit and order)**.** *For each sorted index i and μ-a.e. x, $\frac{1}{n} \log \sigma_i(A^{(n)}(x)) \to \mathrm{exponents}_i$; moreover* exponents *is antitone.*

Proof. The deterministic exponent sequence is extracted by Classical.choose from the singular-value convergence theorem underlying the MET; antitonicity and the a.e. limit are its defining specification. ☐

Theorem 6.13 (Eigenvalue tie)**.** *μ-a.e., $\exp(\mathrm{exponents}_i)$ is the i-th sorted eigenvalue of the Oseledets limit matrix $\Lambda(x)$.*

Proof. The σ-limit identifies $\mathrm{lamSing}(x, i) = \mathrm{exponents}_i$ a.e.; combine with the eigenvalues of Λ being e^{lamSing}. ☐

6.5 Exponent sums

Theorem 6.14 (Sign characterizations of exponent sums)**.** *The sum* sumPosExp *of the strictly positive exponents is nonnegative, vanishes iff all exponents are ≤ 0, and is strictly positive iff some exponent is positive (and dually* $\mathrm{sumNegExp} \leq 0$ *with the mirror characterizations).*

Proof. Each summand of sumPosExp is strictly positive on the filter, so the sum is ≥ 0; a sum of nonnegatives vanishes iff the filter is empty, i.e. no exponent is positive. ☐

Theorem 6.15 (Telescoping identity for partial sums)**.** *For $k \leq d$, the ergodic growth rate Γ_k of the product of the top-k singular values equals the sum of the top-k exponents: $\Gamma_k = \sum_{i<k} \mathrm{exponents}_i$.*

Proof. Since $\mathrm{sprod}_k = \prod_{i<k} \sigma_i$, the normalized $\log \mathrm{sprod}_k$ is the finite sum of the per-index $\frac{1}{n} \log \sigma_i$, each converging to $\mathrm{exponents}_i$. The sum of convergents converges to the sum of limits; uniqueness against the defining limit of Γ_k closes it. ☐

6.6 Exterior (wedge) growth

Definition 6.16 (Exterior cocycle generator)**.** The k-th exterior generator $\mathrm{extGen}\, k\, A$ sends x to the k-th compound matrix $C_k(Ax)$ of $k \times k$ minors. Its iterated cocycle is the compound of the iterate: $\mathrm{cocycle}(\mathrm{extGen}\, k\, A)\, T\, n\, x = C_k(A^{(n)}(x))$.

Theorem 6.17 (k-volume growth rate)**.** *For $k \leq d$ and μ-a.e. x, the operator-norm growth of the compound cocycle converges to Γ_k: $\frac{1}{n} \log \left\| C_k(A^{(n)}(x)) \right\| \to \Gamma_k$, the k-dimensional volume growth rate.*

Proof. The operator norm of the compound matrix is sprod_k, the product of the top-k singular values; rewrite and apply the defining a.e. limit of Γ_k. ☐

Corollary 6.18 (Positive sum as a maximal partial sum)**.** *Writing $k_+ = \#\{i : 0 < \mathrm{exponents}_i\}$, the positive-exponent sum equals the partial sum $\Gamma_{k_+} = \sum_{i<k_+} \mathrm{exponents}_i$.*

Proof. By antitonicity the strictly positive entries are exactly the top k_+ indices, so the filtered positive sum coincides with the top-k_+ prefix sum, which is Γ_{k_+} via 6.15. ☐

6.7 The trace–determinant identity

Lemma 6.19 (Product of singular values is the absolute determinant)**.** *For every* n*,* x*:* $\operatorname{sprod} A\, T\, d\, n\, x = \left|\det A^{(n)}(x)\right|$.

Proof. Squaring, $\operatorname{sprod}_d^2 = \prod_i \sigma_i^2 = \det(M^\top M) = (\det M)^2$ for the symmetric Gram operator; take the nonnegative square root. No invertibility is needed. □

Theorem 6.20 (Determinant identity)**.** *The sum of all Lyapunov exponents equals the integral of* $\log|\det|$ *of the generator:*

$$\sum_i \mathrm{exponents}_i = \int_X \log|\det A(x)|\; d\mu.$$

Proof. Two a.e. limits of $\frac{1}{n}\log\left|\det A^{(n)}\right|$: it equals $\frac{1}{n}\log \operatorname{sprod}_d \to \Gamma_d = \sum_i \mathrm{exponents}_i$ by 6.19 and 6.15; and $\log\left|\det A^{(n)}\right|$ is the additive Birkhoff sum of $\log|\det A|$, whose ergodic average tends to $\int \log|\det A|$. Uniqueness of limits closes the identity. □

Corollary 6.21 (Volume contraction)**.** *If* $\sum_i \mathrm{exponents}_i < 0$ *then* μ*-a.e.* $\left|\det A^{(n)}(x)\right| \to 0$.

Proof. Since $\frac{1}{n}\log\left|\det A^{(n)}\right|$ tends to a negative constant, $\log\left|\det A^{(n)}\right| \to -\infty$, so its exponential tends to 0. □

6.8 The inverse / time-reversed spectrum

Theorem 6.22 (Inverse cocycle exponents)**.** *For each sorted index* i *and* μ*-a.e.* x*, the singular-value exponents of the inverse-matrix cocycle are the negated, reversed exponents of* A*:*

$$\tfrac{1}{n}\log\sigma_i\big((A^{(n)}(x))^{-1}\big) \to -\mathrm{exponents}_{\mathrm{rev}\, i}.$$

Proof. Singular-value reciprocity $\sigma_i(M^{-1}) = \sigma_{\mathrm{rev}\, i}(M)^{-1}$ for invertible M, applied to the iterate, turns the forward limit $\mathrm{exponents}_{\mathrm{rev}\, i}$ into its negative. □

Corollary 6.23 (Top of the reversed spectrum is minus the bottom)**.** μ*-a.e. the largest exponent of the inverse cocycle is* $-\mathrm{exponents}_{d-1}$*, the negative of the smallest forward exponent.*

Proof. Specialize 6.22 at $i = 0$, where $\mathrm{rev}\, 0 = d - 1$. □

6.9 Restriction to invariant subbundles

Definition 6.24 (Invariant subbundle)**.** An invariant subbundle is a measurable family of fibre subspaces $W(x) \le \mathbb{R}^d$ that is A-equivariant a.e.: $A(x)\, W(x) = W(Tx)$.

Lemma 6.25 (Dimension interlacing)**.** *At each ambient flag level, the dimension captured by the subbundle is bounded by the ambient dimension:* $\dim\big(W(x) \cap V_i(x)\big) \le \dim V_i(x)$*, so the restricted multiplicities are a sub-multiset of the ambient ones.*

Proof. Monotonicity of finrank under $W \cap V_i \le V_i$. □

Theorem 6.26 (Restricted strict Oseledets filtration). *For ergodic T, invertible A, and an invariant subbundle W, collapsing the constant-dimension levels of $i \mapsto W \cap V_i$ yields a genuine* strict *Oseledets filtration inside W: a strictly antitone λ' and a measurable family V' with, μ-a.e., $V'_0(x) = W(x)$, $V'_{k'}(x) = 0$, strictly descending and A-equivariant, with exact growth rate λ'_i per stratum and all levels $\leq W(x)$.*

Proof. Obtain a forward witness via 6.2, restrict the flag to W and read its a.e.-constant dimension profile (the restricted analogue of 6.8); enumerate the first-occurrence surviving indices via an order embedding to collapse repeats into a strict flag, inheriting equivariance and per-stratum growth. The top level is W, not $\top$, so the result is stated directly rather than through 6.1. □

6.10 The non-ergodic spectrum

Theorem 6.27 (Non-ergodic exponents). *For merely measure-preserving T (no ergodicity) and invertible measurable A with the standing integrability, there is a family of T-invariant integrable functions $\lambda_i : X \to \mathbb{R}$ such that for each $i < d$ and μ-a.e. x, $\frac{1}{n} \log \sigma_i(A^{(n)}(x)) \to \lambda_i(x)$. The exponents become invariant functions rather than constants.*

Proof. The non-ergodic Kingman theorem applied to the subadditive cocycle $\log \operatorname{sprod}_k$ produces invariant integrable partial-sum limits G_k; set $\lambda_i = G_{i+1} - G_i$, with the per-$\sigma$ telescoping as in the ergodic case. □

Corollary 6.28 (Non-ergodic positive-exponent sum). *Summing the positive parts* $\max(\lambda_i(x), 0)$ *over $i < d$ yields a single T-invariant integrable function $G_+ : X \to \mathbb{R}$, the non-ergodic analogue of the positive-exponent sum.*

Proof. A finite sum of positive parts of the invariant integrable functions of 6.27 is invariant and integrable. □

6.11 Regularity of the exponents

Theorem 6.29 (Fekete infimum representation). *The partial-sum growth rate is the infimum over n of the normalized integrals:*

$$\Gamma_k = \inf_n \frac{1}{n+1} \int_X \log \operatorname{sprod}_k(n+1, x)\, d\mu.$$

Proof. The integral sequence is subadditive (Fekete), so it converges to its infimum; a Fatou estimate against the dominating Birkhoff averages of $\log^+ \|A^{\pm 1}\|$ identifies that limit with the a.e. growth rate Γ_k in both directions. □

Theorem 6.30 (Upper semicontinuity of partial sums and top exponent). *Along a filter of generators $B_i \to A$ for which each fixed-n integral is continuous, the partial-sum rate is upper semicontinuous:* $\limsup_i \Gamma_k(B_i) \leq \Gamma_k(A)$*; specializing $k = 1$ gives the same for the top exponent. This is USC, not continuity — equality can fail when a spectral gap closes.*

Proof. Γ_k is an infimum of the per-n continuous normalized integrals (6.29); for each n, $\Gamma_k(B_i)$ is below the n-th integral, whose limit along the filter is the n-th integral of A, so $\limsup_i \Gamma_k(B_i)$ is at most every term of the infimum. □

Theorem 6.31 (Lower semicontinuity of the bottom exponent). *The bottom exponent $\lambda_d = \Gamma_d - \Gamma_{d-1}$ is lower semicontinuous in the generator: $\lambda_d(A) \leq \liminf_i \lambda_d(B_i)$.*

Proof. Writing $\Gamma_d = \int \log|\det|$ (the determinant identity 6.20), which is continuous in the generator under the hypothesis, and Γ_{d-1} upper semicontinuous (6.30), the difference $\Gamma_d - \Gamma_{d-1}$ is lower semicontinuous. □

6.12 Singular one-sided bounds

Theorem 6.32 (Forward top value)**.** *For ergodic T and a possibly-singular measurable generator with only $\log^+\|A\| \in L^1$ (no invertibility, no inverse integrability), the normalized positive-part log-norms $\frac{1}{n}\log^+\|A^{(n)}(x)\|$ converge μ-a.e. to a constant λ_1^+.*

Proof. Apply the ergodic Kingman theorem to the subadditive, bounded-below, integrable cocycle $\log^+\|A^{(n)}\|$; only the forward integrability is used. □

Theorem 6.33 (Upper bound on the singular top exponent)**.** *Under the same singular hypotheses there is λ_1^+ with, μ-a.e.,*

$$\limsup_n \left(\tfrac{1}{n}\log\|A^{(n)}(x)\| : \mathrm{EReal}\right) \le \lambda_1^+.$$

The lim sup *is taken in* EReal *so the bound is unconditional even when the growth tends to $-\infty$; this is one-sided only.*

Proof. Termwise $\log \le \log^+$, then pass to the EReal lim sup; since the $\log^+$ sequence converges to λ_1^+ (6.32), its lim sup is λ_1^+. □

Theorem 6.34 (Sharp lim sup in the expanding case)**.** *There is a forward top value λ_1^+ (the a.e. limit of $\frac{1}{n}\log^+\|A^{(n)}\|$) such that, whenever $\lambda_1^+ > 0$, μ-a.e. the genuine log-norm* lim sup *is exactly λ_1^+:*

$$\limsup_n \left(\tfrac{1}{n}\log\|A^{(n)}(x)\| : \mathrm{EReal}\right) = \lambda_1^+.$$

The positivity hypothesis is essential; in the contracting case $\lambda_1^+ = 0$ the genuine growth may tend to $-\infty$ and equality fails.

Proof. The $\le$ half is 6.33. For $\ge$: where $\frac{1}{n}\log^+\|A^{(n)}\| \to \lambda_1^+ > 0$ the sequence is eventually positive, forcing $\log\|A^{(n)}\| > 0$, so $\log^+ = \log$ eventually; the two EReal sequences agree eventually, hence so do their lim sups, which equal λ_1^+. □

Theorem 6.35 (Singular top-k volume upper bound)**.** *For ergodic T and a possibly-singular generator with $\log^+\|A\| \in L^1$, there is a constant Γ_k^+ with, μ-a.e.,*

$$\limsup_n \left(\tfrac{1}{n}\log \mathrm{sprod}_k(x) : \mathrm{EReal}\right) \le \Gamma_k^+.$$

The top-k volume growth is bounded above unconditionally (the EReal lim sup *allows volume collapse), again one-sided only.*

Proof. Since $\mathrm{sprod}_k \ge 0$ is submultiplicative with no invertibility, the same $\log^+$-of-a-nonnegative-subadditive-quantity Kingman construction gives the a.e.-constant value Γ_k^+ and the termwise $\log \le \log^+$ domination transfers to the EReal lim sup. □

Chapter 7

The two-sided splitting

The one-sided multiplicative ergodic theorem produces, for a measurable matrix cocycle over an ergodic system, a decreasing measurable equivariant *filtration* $\mathbb{R}^d = V_0(x) \supsetneq V_1(x) \supsetneq \cdots \supsetneq V_k(x) = 0$ whose successive quotients carry the distinct Lyapunov exponents $\lambda_0 > \cdots > \lambda_{k-1}$. When the base dynamics is *invertible* and the generator is everywhere invertible with $\log^+ \|A\|, \log^+ \|A^{-1}\| \in L^1$, the filtration upgrades to a genuine direct-sum *splitting* $\mathbb{R}^d = E_0(x) \oplus \cdots \oplus E_{k-1}(x)$, with each E_i equivariant and the cocycle growing at the exact rate $+\lambda_i$ forward and $-\lambda_i$ backward on $E_i \setminus \{0\}$.

This chapter develops that upgrade. The strategy applies the one-sided theorem twice — once forward to (T, A), once to a reflected *backward generator* over T^{-1} — and intersects the two filtrations. Two genuinely new analytic nodes carry the argument: an identification of the Kingman subadditive constant with the limit of integral means, and a backward-orbit growth envelope for restricted cocycles. Their consequence is the *transversality crux*: forward and backward exponents of any common nonzero vector add to a nonnegative number, so opposite-side sublevels meet in 0. A pure multiset reflection lemma then aligns the backward spectrum to the negated forward one, after which a telescoping-flag lattice argument assembles the splitting.

7.1 The invertible setup and the backward generator

The base dynamics is an invertible measure-preserving system $T : X \simeq_m X$ (a `MeasurableEquiv`) with $T^{-1} = T.\mathrm{symm}$. The backward iterates of the cocycle $A^{(n)}(x) = \mathrm{cocycle}\, A\, T\, n\, x$ are governed by the *backward generator*, a generator over T^{-1} whose cocycle reflects the forward one through inverse and orbit reversal.

Definition 7.1 (Backward generator)**.** For $A : X \to \mathrm{Matrix}\,(\mathrm{Fin}\, d)\,(\mathrm{Fin}\, d)\,\mathbb{R}$ and $T : X \simeq_m X$, the backward generator is

$$\mathrm{backwardGen}\, A\, T\, x \;=\; \left(A(T^{-1}x)\right)^{-1}.$$

Its cocycle is taken over T^{-1}.

Lemma 7.2 (Cocycle recursion, newest factor on the right)**.** *For any generator A and map T, $A^{(n+1)}(x) = A(T^{[n]}x) \cdot A^{(n)}(x)$.*

Proof. This is the companion of the standard recursion $A^{(n+1)}(x) = A^{(n)}(Tx) \cdot A(x)$: apply the additive law $A^{(m+n)} = A^{(m)}(T^{[n]}\cdot) \cdot A^{(n)}$ with $m = 1$ after commuting the summands. ☐

Lemma 7.3 (Backward cocycle identity)**.** *Writing $B = \mathrm{backwardGen}\, A\, T$, the cocycle of B over T^{-1} satisfies*

$$B^{(n)}(x) \;=\; \left(A^{(n)}(T^{-n}x)\right)^{-1}, \qquad B^{(n)}(T^n y) \;=\; \left(A^{(n)}(y)\right)^{-1}.$$

Proof. Induct on n. The successor step expands $B^{(n+1)}$ by the left recursion, applies the inductive hypothesis at $T^{-1}x$, rewrites $B = (A \circ T^{-1})^{-1}$ by theorem 7.2, and collapses the product of inverses by Matrix.mul_inv_rev; a left-inverse iterate identity $(T \circ T^{-1})^{[n]} = \mathrm{id}$ reindexes the orbit point. The dual form follows by substituting $x = T^n y$ and cancelling $(T^{-1} \circ T)^{[n]} = \mathrm{id}$. This identity is exactly what makes the second (backward) growth limit in the splitting theorem the cocycle $(A^{(n)}(T^{-n}x))^{-1}$, avoiding any moving-point singular-value bookkeeping. □

Proposition 7.4 (Backward standing hypotheses)**.** *If $\det A(x) \neq 0$ for all x, A is measurable, T is measure-preserving, and $\log^+ \|A\|, \log^+ \|A^{-1}\| \in L^1(\mu)$, then the backward system (T^{-1}, B) satisfies the same four hypotheses: $\det B \neq 0$ everywhere, B measurable, and $\log^+ \|B\|, \log^+ \|B^{-1}\| \in L^1(\mu)$.*

Proof. Invertibility of $B = (A \circ T^{-1})^{-1}$ is $\det(M^{-1}) = (\det M)^{-1} \neq 0$; measurability is composition of matrix inversion with the measurable T^{-1}. For integrability, T^{-1} is measure-preserving (Mathlib's MeasurePreserving.symm), so $\log^+ \|B\| = \log^+ \|A^{-1}\| \circ T^{-1}$ and $\log^+ \|B^{-1}\| = \log^+ \|A\| \circ T^{-1}$ (using $(M^{-1})^{-1} = M$) transfer by composition. Crucially T ergodic implies T^{-1} ergodic (Mathlib's Ergodic.symm), so the one-sided theorem applies verbatim to the backward system. □

Lemma 7.5 (Biinvariant conull set)**.** *For T measure-preserving on a probability space and a conull measurable set S, there is a conull measurable $S' \subseteq S$ invariant under both T and T^{-1}.*

Proof. Take $S' = (\bigcap_n (T^{[n]})^{-1} S) \cap (\bigcap_n (T^{-[n]})^{-1} S)$. Each iterate is measure-preserving, so every preimage of the conull S is conull and the countable intersection is conull. Membership is "all forward and backward iterates land in S"; applying T or T^{-1} merely shifts the index, the cross term collapsing through $T^{-1} \circ T = \mathrm{id}$. □

7.2 The strong one-sided export with dimensions

The crux argument needs the dimensions of the filtration spaces, which the headline one-sided statement quantifies away. We re-run the one-sided composition with concrete witnesses, exposing the spectrum λ and the filtration V together with a dimension formula.

Proposition 7.6 (Strong one-sided export)**.** *Under the one-sided hypotheses (with* [NeZero d]*) there exist a decreasing exponent sequence* $\mathrm{lam0} : \mathbb{N} \to \mathbb{R}$ *realizing the a.e. per-index singular-value limits, and a measurable filtration* $V : \mathrm{Fin}(k+1) \to X \to \mathrm{Submodule}\,\mathbb{R}\,(\mathbb{R}^d)$ *where* $k = \mathrm{numExp}\,\mathrm{lam0}\,d$*, such that for a.e. x: $V_0 x = \top$, $V_k x = \bot$, V is strictly decreasing and A-equivariant, every $v \in V_i x \setminus V_{i+1} x$ grows at rate* $\mathrm{expEnum}\,\mathrm{lam0}\,d\,i$*, and*

$$\mathrm{finrank}\,(V_i x) \;=\; \#\{\, j < d : \mathrm{lam0}\, j \le \mathrm{expEnum}\,\mathrm{lam0}\,d\,i \,\}.$$

Proof. This is the committed one-sided assembly, re-run with the concrete witness $V = V'\,A\,T\,\mathrm{lam0}$: obtain lam0 from the singular-value exponents, discharge the top-gap mass envelope, build the spectral, slow-flag and growth interfaces exactly as the headline proof does, and read the structural block off the same conull set. The dimension formula is supplied by theorem 7.7 combined with the slow-flag identification $V'_i x = \mathrm{vslow}\,(\exp \lambda_i)\, x$. □

Proposition 7.7 (Rank of the slow space)**.** *For a.e. x and all $t \in \mathbb{R}$,*

$$\operatorname{finrank}\left(\operatorname{vslow} A\, T\, (\exp t)\, x\right) \;=\; \#\{\, j < d : \operatorname{lam0} j \le t \,\}.$$

Proof. The sanitized limit operator $\widehat{\Lambda}$ has an orthonormal eigenbasis with eigenvalues $\exp(\operatorname{lamSing} x\, e)$, and a.e. $\operatorname{lamSing} x\, j = \operatorname{lam0} j$ (a countable conjunction over $j < d$). The slow space is the range of a spectral sublevel projector $\operatorname{cfc} f\, \widehat{\Lambda}$; theorem 7.8 shows it acts diagonally on the eigenbasis with eigenvalues in $\{0, 1\}$, and the rank of such a self-adjoint idempotent equals the number of 1-eigenvectors, which by exp-monotonicity of the threshold is $\#\{j : \operatorname{lam0} j \le t\}$. ☐

Lemma 7.8 (Functional calculus on an eigenvector)**.** *If M is self-adjoint and $Mv = c\,v$ for $v \ne 0$, then $(\operatorname{cfc} f\, M)\, v = f(c) \cdot v$.*

Proof. The eigenvalue c lies in the (finite) spectrum of M. Pick a Lagrange interpolating polynomial q with $q = f$ on the spectrum; then $\operatorname{cfc} f\, M = \operatorname{cfc} q\, M = q(M)$, and $q(M)v = q(c)v = f(c)v$ because v is an eigenvector. This is pointwise, so no measurability is incurred. ☐

7.3 The Kingman means identification

The repository's ergodic Kingman theorem exhibits a constant c with $(1/n)g_n \to c$ a.e., but defers identifying c with the Fekete infimum of the integral means. The two-sided theorem is the first consumer that needs the means form, in order to equate the Kingman constants of a subadditive cocycle over T and of its orbit-reversed version over T^{-1}.

Proposition 7.9 (Kingman constant as the limit of integral means)**.** *Under the hypotheses of the ergodic Kingman theorem, there is a constant c with*

$$\frac{1}{n+1}\int_X g_{n+1}\, d\mu \longrightarrow c \qquad \text{and} \qquad \tfrac{1}{n}\, g_n(x) \longrightarrow c \;\; \text{for a.e. } x.$$

Proof. Let L be the Fekete limit of the means (existing by subadditivity). For $c \le L$: iterate subadditivity $g_{mn}(x) \le \sum_{j<m} g_n(T^{[jn]}x)$, divide by mn and let $m \to \infty$; the left side converges to c along a subsequence, the right to the Birkhoff average of $(1/n)g_n$ for the measure-preserving $T^{[n]}$ (no ergodicity needed), whose integral is $(1/n)\int g_n$; integrate the inequality. For $c \ge L$: apply Fatou to the nonnegative sequence $A_n - \operatorname{cdiv}_n$, where A_n is the Birkhoff average of g_1 (nonnegative by single-step subadditivity) and cdiv_n tends to L; since $A_n - \operatorname{cdiv}_n \to B - c$ a.e. and $\int A_n = \int g_1$, Fatou yields $\int g_1 - L \ge \int g_1 - c$. ☐

7.4 Restricted cocycles and their exponent

To measure the growth of $A^{(n)}$ restricted to a filtration level V_i, we form a restricted cocycle by post-composing the orthogonal projector onto V_i. A *floor* term repairs the failure of everywhere-subadditivity off the good set, so that Kingman applies without an a.e. caveat.

Definition 7.10 (Floored restricted log-cocycle)**.** For a measurable family $V : X \to \operatorname{Submodule} \mathbb{R}\, (\mathbb{R}^d)$, set $\operatorname{restGen} A\, V\, x = A(x) \cdot P_{V(x)}$ (with P_K the orthogonal projector onto K), $\operatorname{sFloor} A\, T\, n\, x = \prod_{j<n} \left\|(A(T^{[j]}x))^{-1}\right\|^{-1}$, and

$$\operatorname{restLog} A\, V\, T\, n\, x \;=\; \log\Bigl(\left\|\operatorname{cocycle}(\operatorname{restGen} A\, V)\, T\, n\, x\right\| \,\sqcup\, \operatorname{sFloor} A\, T\, n\, x\Bigr).$$

Lemma 7.11 (Everywhere subadditivity)**.** restLog $A\,V\,T$ *is an everywhere subadditive cocycle: for all* m, n, x*,* $\mathrm{restLog}\,(m+n)\,x \le \mathrm{restLog}\,m\,x + \mathrm{restLog}\,n\,(T^{[m]}x)$*.*

Proof. The floor is multiplicative, $\mathrm{sFloor}\,(m+n) = \mathrm{sFloor}\,m \cdot (\mathrm{sFloor}\,n) \circ T^{[m]}$, the norm of the restricted product is submultiplicative, and $\max(ab, cd) \le \max(a, c)\max(b, d)$ for nonnegatives; taking log (monotone) gives subadditivity with no exceptional set. The floor is the only mechanism that keeps the everywhere signature Kingman requires; on the good set it is dominated and disappears. ☐

Lemma 7.12 (Restricted Kingman, both directions of time)**.** *Kingman applied to* restLog *over* T *yields a constant* χ_V*; the orbit-reversed cocycle* $h_n(x) = \mathrm{restLog}_n(T^{-n}x)$ *is subadditive over* T^{-1} *with the same integral means, hence converges a.e. to the same* χ_V*.*

Proof. Over T, theorem 7.11 together with integrability of the endpoints gives a Kingman constant χ_V. The reversed family $h_n = \mathrm{restLog}_n \circ T^{-n}$ is subadditive over T^{-1} by an index juggle through the additive law, and $\int h_n = \int \mathrm{restLog}_n$ by measure-preservation of T^{-n}. Since T^{-1} is ergodic, theorem 7.9 forces the two Kingman constants — both being the common limit of the identical integral means — to coincide. ☐

Proposition 7.13 (Restricted exponent equals λ_i)**.** *For the forward level* V_i*, the restricted Kingman constant equals the* i*-th exponent:* $\chi_{V_i} = \mathrm{expEnum}\,\mathrm{lam0}\,d\,i$*.*

Proof. Lower bound $\ge$: on the good set, for $v \in V_i x$ the restricted cocycle equals $A^{(n)}(x)v$; a stratum witness from strictness of the filtration realizes the rate λ_i. Upper bound $\le$: pointwise, expand any $P_{V_i(y)}w$ in a classical orthonormal basis of $V_i(y)$; then $\|A^{(n)}Pw\| \le \sum_j \|A^{(n)}e_j\|$ with each e_j in a stratum of index $\ge i$, so its growth is $\le \lambda_i$; the log-of-a-finite-sum lemma passes the maximum through. Both bounds match the a.e.-constant Kingman limit. ☐

Proposition 7.14 (Backward-orbit growth envelope)**.** *For a.e.* x*,*

$$\limsup_n \tfrac{1}{n} \log\bigl\|A^{(n)}(T^{-n}x) \cdot P_{V_i(T^{-n}x)}\bigr\| \;\le\; \mathrm{expEnum}\,\mathrm{lam0}\,d\,i.$$

Proof. This is the analytic heart. By theorem 7.12 the reversed restricted log converges a.e. to the same constant χ_{V_i}, which theorem 7.13 identifies as λ_i. On the good set the floor is absorbed, so the lim sup of the genuine restricted norm along the backward orbit is bounded by λ_i. Only the $\le$ direction is consumed downstream. ☐

7.5 The transversality crux

The envelope feeds the central nonnegativity fact: a vector cannot decay backward strictly faster than $-\lambda_i$ while lying in V_i. The crux is pointwise — ergodicity is not used here — and directly kills nonzero intersection vectors of opposite-side sublevels.

Proposition 7.15 (Sublevels of opposite sign are transverse)**.** *Fix* x*. If the forward envelope for* V_x *holds at rate* a*,* U_x *decays backward at rate* $\le b$*, and* $a + b < 0$*, then* $V_x \sqcap U_x = \bot$*.*

Proof. Suppose $0 \ne v \in V_x \sqcap U_x$. Put $v_n = B^{(n)}(x)v$; by theorem 7.3 $v = A^{(n)}(T^{-n}x)\,v_n$ and, by forward equivariance along the backward orbit, $v_n \in V_i(T^{-n}x)$. Hence $\log\|v\| \le \text{restLog-envelope}_n(x) + \log\|v_n\|$, whose right side behaves like $n(a + b + 2\varepsilon)$, tending to $-\infty$. This contradicts $\|v\|$ being a positive constant. ☐

Proposition 7.16 (The a.e. crux)**.** *For a.e. x, for every forward level i and backward level s with* $\mathrm{expEnum}\,\mathrm{lam0}\,d\,i + \mathrm{expEnum}\,\mathrm{mu0}\,d\,s < 0$,

$$V_i x \;\sqcap\; W_s x \;=\; \bot.$$

Proof. Bundle all the a.e. facts — the forward envelope (one application of theorem 7.14 per forward level), the backward growth (flag descent in W) and equivariance along the orbit — onto a single biinvariant conull set via theorem 7.5. On it, every pair (i, s) with negative exponent sum satisfies the hypotheses of theorem 7.15, so the intersection is $\bot$. The quantifiers over the finite $\mathrm{Fin}\,k \times \mathrm{Fin}\,l$ are discharged simultaneously. □

Corollary 7.17 (Counting bound)**.** *For all $a, b \in \mathbb{R}$ with $a + b < 0$,*

$$\#\{\, j < d : \mathrm{lam0}\,j \le a \,\} \;+\; \#\{\, j < d : \mathrm{mu0}\,j \le b \,\} \;\le\; d.$$

Proof. Convert thresholds to filtration levels and apply theorem 7.16 at one good point: the two sublevel spaces meet in $\bot$, so by the Grassmann formula their dimensions sum to at most $\dim \mathbb{R}^d = d$. The dimension formula of theorem 7.6 (forward and backward) turns the dimensions into the stated counts. The bound is deterministic, so it holds outright. □

7.6 Spectral reflection

The backward spectrum must be aligned to the forward one. The determinant identity $\sum_j \mathrm{lam0}\,j = \int \log|\det A|$ — and its backward negation — together with the counting bound pin the backward exponents to the negated, reversed forward ones, by a purely combinatorial multiset argument that avoids any exterior-power calculus for $\|\bigwedge^q M^{-1}\|$.

Proposition 7.18 (Forward determinant sum)**.** *Any exponent sequence* lam0 *realizing the a.e. singular-value limits satisfies* $\sum_{j<d} \mathrm{lam0}\,j = \int_X \log|\det A(x)|\; d\mu$.

Proof. By uniqueness of a.e. limits at a common conull point, $\mathrm{lam0}\,j$ equals the chosen spectrum $\mathrm{exponents}\,j$ for each $j < d$; the sum of those is the integral of $\log|\det A| = \log \prod_j \sigma_j$ via the established singular-value/determinant identity. □

Proposition 7.19 (Backward sum is the negated forward sum)**.** *For a backward exponent sequence* mu0, $\sum_{j<d} \mathrm{mu0}\,j = -\sum_{j<d} \mathrm{lam0}\,j$.

Proof. Apply theorem 7.18 to the backward system (legitimate by theorem 7.4): $\sum \mathrm{mu0}\,j = \int \log|\det B|$. The pointwise identity $\log|\det B| = -\log\left|\det(A \circ T^{-1})\right|$ and the change of variables along the measure-preserving T^{-1} give $\int \log|\det B| = -\int \log|\det A| = -\sum \mathrm{lam0}\,j$. □

Proposition 7.20 (Reflection lemma)**.** *Let $p, q : \mathbb{N} \to \mathbb{R}$ be antitone on $[0, d)$. If $\#\{p \le a\} + \#\{q \le b\} \le d$ whenever $a + b < 0$, and $\sum_{j<d} q\,j = -\sum_{j<d} p\,j$, then $q\,j = -p\,(d-1-j)$ for all $j < d$.*

Proof. Apply the counting bound at b and $a = -q(j) - \varepsilon$: since antitone tuples are their own sorted enumerations, this gives the sorted domination $q\,j \ge -p\,(d-1-j)$ for each j. A pointwise $\ge$ whose total sum is an equality forces pointwise equality, so $q\,j = -p\,(d-1-j)$. This is pure finite combinatorics with no analytic content. □

Corollary 7.21 (Aligned backward index)**.** *Under the reflection* $\mathrm{mu0}\,j = -\mathrm{lam0}\,(d-1-j)$, *the backward count of distinct exponents equals the forward one, and the index* $\mathrm{sidx}\,i = \mathrm{cast}\,(\mathrm{Fin.rev}\,i)$ *satisfies* $\mathrm{expEnum}\,\mathrm{mu0}\,d\,(\mathrm{sidx}\,i) = -\mathrm{expEnum}\,\mathrm{lam0}\,d\,i$.

Proof. From theorem 7.20 the multisets of distinct values agree up to negation and reversal, so $\text{numExp}\,\text{mu0}\,d = \text{numExp}\,\text{lam0}\,d$ and the enumerations satisfy $\text{expEnum}\,\text{mu0}\,d\,a = -\text{expEnum}\,\text{lam0}\,d\,(\text{Fin.rev}\,a)$. Substituting $a = \text{sidx}\,i = \text{Fin.rev}\,i$ and computing the index by omega gives the negated forward exponent. □

7.7 Measurability of the intersection bundle

The split bundle is $x \mapsto V_i x \sqcap W_{\text{sidx}\,i}\,x$. Measurable subspaces do not close under $\sqcap$ for free, and Mathlib has no alternating-projection theorem; instead a single von Neumann-style power lemma supplies measurability.

Proposition 7.22 (Powers of the projector triple converge to the intersection projector)**.** *For subspaces K, L of $\mathbb{R}^d$, with P_K, P_L the orthogonal projectors,*

$$(P_K\, P_L\, P_K)^n \longrightarrow P_{K \sqcap L} \qquad (n \to \infty).$$

Proof. The matrix $S = P_K P_L P_K$ is self-adjoint, PSD, and a contraction, so its eigenvalues lie in $[0, 1]$; by theorem 7.23 its 1-eigenspace is exactly $K \sqcap L$. Diagonalizing by the spectral theorem, $c^n \to 1$ if $c = 1$ and $\to 0$ otherwise, so S^n converges to the orthogonal projection onto the 1-eigenspace, namely $P_{K \sqcap L}$. □

Lemma 7.23 (The intersection fixes exactly $K \sqcap L$)**.** $(P_K\, P_L\, P_K)\, v = v$ *if and only if* $v \in K \sqcap L$.

Proof. The "if" direction is immediate since both projectors fix vectors in $K \sqcap L$. For "only if", $\langle Sv, v\rangle = \|P_L P_K v\|^2$; the Cauchy–Schwarz equality chain $\|v\|^2 = \langle Sv, v\rangle \le \|v\|^2$ forces $v \in K$ and $P_K v \in L$, whence $v \in K \sqcap L$. □

Proposition 7.24 (Measurable intersection of measurable subspaces)**.** *If $x \mapsto V(x)$ and $x \mapsto W(x)$ are measurable subspace families, then so is $x \mapsto V(x) \sqcap W(x)$.*

Proof. Each $x \mapsto (P_{V(x)} P_{W(x)} P_{V(x)})^n$ is measurable (matrix powers of measurable matrices). By theorem 7.22 these converge pointwise to $P_{V(x)\sqcap W(x)}$, so the limiting projector is measurable entrywise as a limit of measurable functions; this exhibits the intersection family as measurable. □

7.8 Assembling the splitting

With the aligned reflection, the crux disjointness, and the dimension formulas in hand, the splitting is assembled by intersecting each forward level with the matching backward level and telescoping a flag into a direct sum.

Lemma 7.25 (Telescoping-flag lattice lemma)**.** *A descending flag $V_0 \supseteq \cdots \supseteq V_k = \bot$ in a modular lattice, with complements E_i satisfying $V_i = E_i \sqcup V_{i+1}$ and $E_i \sqcap V_{i+1} = \bot$, telescopes into an independent family (E_i) with $\bigsqcup_i E_i = V_0$.*

Proof. Induct on k. The cons step shows that prepending a head E_0 disjoint from the supremum of an already-independent tail preserves independence, using modularity to peel E_0 off the join $E_0 \sqcup C$ intersected with the tail bound. The supremum identity unfolds $\bigsqcup_i E_i = E_0 \sqcup \bigsqcup_i E_{i+1} = E_0 \sqcup V_1 = V_0$ via the first telescoping identity. □

Proposition 7.26 (Splitting at a point)**.** *Define* $E_i = V_i \sqcap W_{\mathrm{sidx}\, i}$*. Then* $\operatorname{finrank} E_i \geq 1$*, the telescoping identities* $V_i = E_i \sqcup V_{i+1}$ *and* $E_i \sqcap V_{i+1} = \bot$ *hold, and consequently* (E_i) *is independent with* $\bigsqcup_i E_i = V_0 = \top$*.*

Proof. The crux theorem 7.16 at the negated-exponent pair gives $V_{i+1} \sqcap W_{\mathrm{sidx}\, i} = \bot$, so by Grassmann $V_{i+1} \sqcup W_{\mathrm{sidx}\, i} = \top$, and the reflection theorem 7.21 fixes $\operatorname{finrank} W_{\mathrm{sidx}\, i} = d - \#\{\mathrm{lam0} < \lambda_i\}$. Combining with the dimension formula of theorem 7.6 yields $\operatorname{finrank} E_i = \#\{\mathrm{lam0} \leq \lambda_i\} - \#\{\mathrm{lam0} < \lambda_i\} \geq 1$, the telescoping totality $V_i = E_i \sqcup V_{i+1}$ by equal finrank, and disjointness from a second crux instance. theorem 7.25 then telescopes to independence and total supremum. □

Theorem 7.27 (The two-sided Oseledets splitting)**.** *Let* $T : X \simeq_m X$ *be an invertible ergodic measure-preserving transformation of a probability space, and let* $A : X \to \mathrm{Matrix}\,(\mathrm{Fin}\, d)\,(\mathrm{Fin}\, d)\,\mathbb{R}$ *be measurable with* $\det A(x) \neq 0$ *for all* x *and* $\log^+ \|A\|, \log^+ \|A^{-1}\| \in L^1(\mu)$*. Then there exist* $k \in \mathbb{N}$*, a strictly decreasing* $\lambda : \mathrm{Fin}\, k \to \mathbb{R}$*, and measurable subspace families* $E : \mathrm{Fin}\, k \to X \to \mathrm{Submodule}\,\mathbb{R}\,(\mathbb{R}^d)$ *such that for* μ*-a.e.* x*:*

- $\mathbb{R}^d = \bigoplus_i E_i(x)$ *is an internal direct sum with every* $E_i(x) \neq \bot$*;*
- *each* E_i *is* A*-equivariant:* $(A(x))(E_i(x)) = E_i(Tx)$*;*
- *for every nonzero* $v \in E_i(x)$*,*

$$\tfrac{1}{n} \log\big\|A^{(n)}(x)\, v\big\| \to \lambda_i \qquad \textit{and} \qquad \tfrac{1}{n} \log\Big\|\big(A^{(n)}(T^{-n}x)\big)^{-1} v\Big\| \to -\lambda_i.$$

Proof. For $d = 0$ take $k = 0$ (the empty internal sum of the zero module). For $d > 0$, apply theorem 7.6 forward to obtain $\mathrm{lam0}, V$ and, via theorem 7.4, to the backward system (T^{-1}, B) to obtain $\mathrm{mu0}, W$, each with its dimension formula. Run the crux theorem 7.16 and counting bound theorem 7.17; combine with theorem 7.19 through theorem 7.20 to get the reflection $\mathrm{mu0}\, j = -\mathrm{lam0}\,(d - 1 - j)$, hence $l = k$ and the index alignment theorem 7.21. Set $\lambda = \mathrm{expEnum}\, \mathrm{lam0}\, d$ and $E_i(x) = V_i x \sqcap W_{\mathrm{sidx}\, i}\, x$; measurability is theorem 7.24. On a conull set, theorem 7.26 provides the internal direct sum and nonzeroness. Equivariance is Submodule.map over $\sqcap$ (injective $A(x)$): the forward factor is the one-sided equivariance, and the backward factor is transported through $\mathrm{backwardGen}\, A\, T\, (Tx) = (A(x))^{-1}$ by theorem 7.3. For growth, a nonzero $v \in E_i(x)$ lies in the i-th forward stratum and the $\mathrm{rev}\, i$-th backward stratum (two crux instances exclude deeper levels), so the two one-sided growth limits apply; the backward limit is rewritten as $-\lambda_i$ through theorem 7.3 and theorem 7.21. □

Chapter 8

The continuous-flow multiplicative ergodic theorem

8.1 Overview

The discrete multiplicative ergodic theorem governs a single measure-preserving map $T : X \to X$ and the iterates of a matrix cocycle generated one step at a time. This chapter lifts that theorem to *continuous time*: the map T is replaced by a one-parameter measure-preserving flow $\varphi : \mathbb{R} \to X \to X$, and the iterated cocycle by a continuous-time linear cocycle $A : \mathbb{R} \to X \to \mathrm{Mat}_d(\mathbb{R})$. The conclusion is a finite Lyapunov spectrum $\lambda_1 > \cdots > \lambda_k$, a measurable filtration that is *flow-equivariant* at every real time, and the exact continuous-parameter growth $t^{-1} \log \|A(t, x)v\| \to \lambda_i$ as $t \to \infty$ over $\mathbb{R}$.

The strategy is a reduction, not a redevelopment of the ergodic machinery for $\mathbb{R}$. We set $T := \varphi(1)$ and read the discrete cocycle off the sampled flow cocycle; the proved discrete theorem 5.30 delivers the integer-time conclusion, and two analytic devices lift it to the continuous parameter: a *between-times sandwich* that controls growth on each interval $[n, n+1)$, and a *shift-invariance of the growth* $\limsup$ that promotes the discrete (time-one) equivariance to equivariance at every real time. No continuous-time Kingman theorem is needed; the integer clock appears only as a technical reduction device. Throughout, X carries no topology: the flow is a measurable measure-preserving action of $(\mathbb{R}, +)$, and the cocycle is measurable in the state at each fixed time.

8.2 The continuous-time data

Definition 8.1 (Measure-preserving one-parameter flow)**.** A *measure-preserving one-parameter flow* on a measurable space X for a measure μ is a family $\varphi : \mathbb{R} \to X \to X$ together with the data $\varphi(0) = \mathrm{id}$, $\varphi(s + t) = \varphi(s) \circ \varphi(t)$ for all $s, t \in \mathbb{R}$, and a proof that each time-t map $\varphi(t)$ preserves μ. No topology on X is assumed; in particular no joint continuity in (t, x) is required.

Lemma 8.2 (Integer times are iterates of the time-one map)**.** *For a measure-preserving flow φ and $n \in \mathbb{N}$, the integer-time map of the flow is the n-fold iterate of its time-one map: $\varphi(n) = (\varphi(1))^{[n]}$.*

Proof. Induction on n. The base case $\varphi(0) = \mathrm{id} = (\varphi(1))^{[0]}$ is the time-zero law. For the step, write $n + 1 = n + 1$ as reals and apply additivity $\varphi(n + 1) = \varphi(n) \circ \varphi(1)$, then the inductive hypothesis and $(\varphi(1))^{[n+1]} = (\varphi(1))^{[n]} \circ \varphi(1)$. □

Definition 8.3 (Continuous-time linear cocycle over a flow)**.** A *continuous-time linear cocycle* over a measure-preserving flow φ, valued in invertible $d \times d$ real matrices, is a family $A : \mathbb{R} \to X \to \mathrm{Mat}_d(\mathbb{R})$ with $A(0, x) = 1$, the cocycle identity (newest factor on the left) $A(t + s, x) = A(t, \varphi(s)x)\, A(s, x)$, a proof that $\det A(t, x) \neq 0$ for all t, x, and measurability of each time-t map $A(t, \cdot)$.

Proposition 8.4 (Reduction identity at integer times)**.** *For a flow cocycle A over φ, every $n \in \mathbb{N}$ and every x,*

$$A(n, x) = \mathrm{cocycle}\,(A(1, \cdot))\,(\varphi(1))\, n\, x,$$

i.e. at integer times the continuous-time cocycle equals the discrete iterated cocycle generated by its time-one map $A(1, \cdot)$ over the time-one dynamics $\varphi(1)$.

Proof. Induction on n. At $n = 0$ both sides are the identity. For the step, split $A((n + 1), x) = A(n, \varphi(1)x)\, A(1, x)$ by the cocycle identity at $t = n$, $s = 1$, match the recursion cocycle $(n+1)\, x = \mathrm{cocycle}\, n\, (\varphi(1)x) \cdot A(1, x)$, and apply the inductive hypothesis at the point $\varphi(1)x$. □

8.3 Reduction to the discrete theorem

The discrete theorem requires the time-one generator to have integrable positive log-norm, both forward and inverse. These follow by evaluating the uniform dominating hypotheses at $s = 1$.

Lemma 8.5 (Integrability of the time-one log-norm)**.** *If $g \in L^1(\mu)$ dominates $\log^+ \|A(s, x)\|$ for all $s \in [0, 1]$ and all x, then the time-one map $A(1, \cdot)$ has integrable positive log-norm.*

Proof. The map $x \mapsto \log^+ \|A(1, x)\|$ is measurable (composition of the measurable $\log^+$, the operator norm, and the measurable time-one cocycle map) and is dominated pointwise by g, taking $s = 1 \in [0, 1]$ in the hypothesis. Dominated by an integrable function, it is integrable. □

Lemma 8.6 (Integrability of the inverse time-one log-norm)**.** *If $g' \in L^1(\mu)$ dominates $\log^+ \left\|(A(s, x))^{-1}\right\|$ for all $s \in [0, 1]$ and all x, then the inverse $(A(1, \cdot))^{-1}$ has integrable positive log-norm.*

Proof. Identical to 8.5, inserting the measurable matrix-inversion map and evaluating the dominating hypothesis at $s = 1$. □

Proposition 8.7 (Discrete filtration for the time-one data)**.** *Let μ be a probability measure, φ a measure-preserving flow with $\varphi(1)$ ergodic, and A a flow cocycle with the two uniform integrable dominators g, g'. Then there exist k, exponents $\lambda : \mathrm{Fin}\, k \to \mathbb{R}$, and a subspace family V forming an Oseledets filtration for the generator $A(1, \cdot)$ over the dynamics $\varphi(1)$.*

Proof. Apply the discrete theorem 5.30 to the ergodic map $\varphi(1)$ and generator $A(1, \cdot)$, feeding it invertibility ($\det A(1, \cdot) \neq 0$), measurability, and the two integrability inputs from 8.5 and 8.6. The output is the desired discrete Oseledets datum. □

8.4 Between integer times

The discrete theorem controls growth only along integer times; the next results bridge the gap to a continuous parameter. Both rest on the orbital sublinearity of an integrable function along the flow's integer orbit.

Lemma 8.8 (Error sublinearity along the integer orbit)**.** *For integrable g, g' and a measure-preserving flow φ, for almost every x the combined fluctuation along the integer orbit vanishes:*

$$n^{-1}(g(\varphi(n)x) + g'(\varphi(n)x)) \to 0 \quad (n \to \infty).$$

Proof. Apply the Birkhoff orbital-tail estimate (a.e. $n^{-1}h(T^{[n]}x) \to 0$ for integrable h) to $h = g + g'$ and the time-one map $T = \varphi(1)$, then rewrite the iterate orbit $(\varphi(1))^{[n]}x$ as $\varphi(n)x$ using 8.2. □

Theorem 8.9 (Between-times sandwich: continuous growth equals integer-time growth)**.** *Fix a flow φ, a flow cocycle A with uniform forward/inverse one-step controls g, g' on $[0, 1]$, a point x, and a nonzero vector v. Suppose the integer fluctuation error vanishes and the integer-time average converges, $n^{-1}\log\|A(n,x)v\| \to L$. Then the continuous-time average converges to the same limit:*

$$t^{-1}\log\|A(t,x)v\| \to L \quad (t \to \infty).$$

Proof. Write $t = r + n$ with $n = \lfloor t \rfloor \geq 1$ and $r \in [0, 1)$. The cocycle identity splits $A(t,x) = A(r,\varphi(n)x)\,A(n,x)$, so with $w = A(n,x)v$ one sandwiches $\log\|A(t,x)v\|$ between $\log\|w\| - \log\|(A(r,\varphi(n)x))^{-1}\|$ and $\log\|w\| + \log\|A(r,\varphi(n)x)\|$. Both correction terms are bounded in absolute value by $g(\varphi(n)x) + g'(\varphi(n)x)$ via the $[0, 1]$ controls (using $\|M\|\|M^{-1}\| \geq 1$). Dividing by t, the error term vanishes by hypothesis, the discrete average converges to L along $\lfloor t \rfloor$, and the floor ratio $\lfloor t \rfloor/t \to 1$; a squeeze over the floor delivers the continuous-time limit. □

8.5 Equivariance at every real time

The discrete theorem gives equivariance one integer step at a time. To obtain equivariance at every real time t_0 we use the intrinsic growth characterization: a vector lies in level V_i iff it is zero or its discrete growth lim sup is $\leq \lambda_i$. The fixed matrix $A(t_0, x)$ is a bounded bijection, so it perturbs the per-step log-norm by $o(n)$, hence leaves the growth lim sup unchanged.

Lemma 8.10 (Fixed-time log-norm is sublinear)**.** *Fix a real time t_0. For almost every x, both $n^{-1}\log\|A(t_0,\varphi(n)x)\|$ and $n^{-1}\log\|(A(t_0,\varphi(n)x))^{-1}\|$ tend to 0 as $n \to \infty$.*

Proof. One builds an integrable function H dominating both $\log^+\|A(t_0,\cdot)\|$ and $\log^+\|(A(t_0,\cdot))^{-1}\|$: by induction on $\lfloor t_0 \rfloor$ one splits $A((\rho+n)+1,\cdot) = A(\rho+n,\varphi(1)\cdot)\,A(1,\cdot)$, bounds $\log^+$ of a product by the sum, and uses measure-preservation of $\varphi(1)$; negative times reduce to positive ones via $A(t_0, y) = (A(-t_0,\varphi(t_0)y))^{-1}$. The two-sided bound $|\log\|M\|| \leq \log^+\|M\| + \log^+\|M^{-1}\|$ turns H into an integrable dominator for the absolute log-norm; the Birkhoff orbital tail of H along the integer orbit and a squeeze finish the proof. □

Theorem 8.11 (Shift-invariance of the growth limsup)**.** *Fix a real time t_0. For almost every x and every test vector u, the discrete-time growth* lim sup *of the cocycle applied to u at x equals that at $\varphi(t_0)x$ applied to the pushed-forward vector $A(t_0,x)\,u$.*

Proof. First, a.e. the discrete growth average $n^{-1}\log\|\mathrm{cocycle}\,n\,x\,u\|$ has bounded range, with upper and lower bounds from the Furstenberg–Kesten Fekete inequalities $\log\|\mathrm{cocycle}\| \leq \mathrm{birkhoffSum}(\log^+\|A(1,\cdot)\|)$ (and the inverse version), whose Birkhoff averages converge by the ergodic theorem. The cocycle identity and 8.4 give the shift relation $\mathrm{cocycle}\,n\,(\varphi(t_0)x) \cdot A(t_0,x) = A(t_0,\varphi(n)x) \cdot \mathrm{cocycle}\,n\,x$, so the two growth averages differ by $n^{-1}(\log\|A(t_0,\varphi(n)x)(\cdots)\| - \log\|\cdots\|)$, which is squeezed to 0 by 8.10. A difference tending to 0 between two range-bounded sequences leaves the lim sup unchanged. □

Theorem 8.12 (Flow-equivariance of the filtration at every real time)**.** *Let V be the Oseledets filtration of the time-one data, and fix $t_0 \in \mathbb{R}$. Then for almost every x and every level i,*

$$\operatorname{map}(A(t_0, x))(V_i\, x) = V_i\,(\varphi(t_0)x).$$

Proof. Use the growth characterization $v \in V_i\, x \iff v = 0 \vee \limsup \le \lambda_i$ at x and, pulled back along the measure-preserving $\varphi(t_0)$, at $\varphi(t_0)x$. The map $P = A(t_0, x)$ is a bijection with inverse $A(t_0, x)^{-1}$. For a non-bottom level, prove both inclusions by le_antisymm: membership of v (resp. $P^{-1}v$) is transported through the equivalence of growth lim sups supplied by 8.11. The bottom level $V_k = \bot$ maps to $\bot$. □

8.6 The continuous-flow theorem

Theorem 8.13 (Continuous-flow multiplicative ergodic theorem)**.** *Let μ be a probability measure on X, let φ be a measure-preserving one-parameter flow whose time-one map $\varphi(1)$ is μ-ergodic, and let A be a continuous-time linear cocycle over φ valued in invertible $d \times d$ real matrices. Suppose $g, g' \in L^1(\mu)$ satisfy, for all $s \in [0, 1]$ and all x,*

$$\log^+ \|A(s, x)\| \le g(x), \qquad \log^+ \left\|(A(s, x))^{-1}\right\| \le g'(x).$$

Then there exist $k \in \mathbb{N}$, a strictly decreasing sequence of exponents $\lambda : \operatorname{Fin} k \to \mathbb{R}$, and a measurable family of subspaces $V : \operatorname{Fin}(k + 1) \to X \to \operatorname{Submodule} \mathbb{R}\,(\operatorname{EuclideanSpace} \mathbb{R}\,(\operatorname{Fin} d))$ such that:

- *λ is strictly decreasing and each V_i is a measurable subspace family;*
- ***(full flow equivariance)** for every $t \in \mathbb{R}$, almost every x has $A(t, x) \cdot V_i\, x = V_i\,(\varphi(t)x)$ for all i;*
- *almost every x carries the strict flag $\top = V_0\, x \supsetneq \cdots \supsetneq V_k\, x = \bot$, and on each stratum $v \in V_i\, x \setminus V_{i+1}\, x$ the continuous-time growth rate is exactly λ_i:*

$$t^{-1} \log \|A(t, x)v\| \to \lambda_i \quad (t \to \infty).$$

Proof. Take (k, λ, V) from the reduction 8.7; this supplies strict antitonicity, measurability, the strict flag, and the integer-time growth rates. Full flow equivariance at each t is 8.12. For the continuous-time growth, work a.e. on the discrete-conclusion set intersected with the error-sublinearity set of 8.8. Fix a stratum vector $v \neq 0$; rewriting via the reduction identity 8.4 turns the discrete stratum growth $n^{-1} \log \|\operatorname{cocycle} n\, x\, v\| \to \lambda_i$ into $n^{-1} \log \|A(n, x)v\| \to \lambda_i$. The between-times sandwich 8.9 then upgrades this integer-time limit to the continuous-parameter limit $t^{-1} \log \|A(t, x)v\| \to \lambda_i$. □

Chapter 9

Kolmogorov–Sinai entropy

Alongside the multiplicative theory of the preceding chapters, the library develops the *classical entropy theory* of a measure-preserving system: the Shannon entropy of a finite measurable partition, its conditional refinement, the Kolmogorov–Sinai entropy $h(T)$ as a Fekete limit over iterated joins, and the structural theorems that make it computable — the one- and two-sided generator theorems and the Abramov–Rokhlin addition formula. The chapter closes where the additive and multiplicative theories meet: the Margulis–Ruelle inequality $h(T) \le \sum_{\lambda_i>0} \lambda_i$ bounding the entropy by the positive Lyapunov exponents of the derivative cocycle, and Rokhlin's volume-distortion identity $h(T, \xi) = \int \log|\det DT|\ d\mu$.

Throughout, (α, μ) is a measure space with μ a probability measure, T a measure-preserving transformation, and partitions are finite, indexed by a `Fintype`. Where regular conditional probabilities are needed (conditional entropy, the generator theorems, Abramov–Rokhlin), the space is additionally assumed standard Borel. We write $\eta(t) = -t \log t$ (Mathlib's negMulLog, with $\eta(0) = 0$ built in).

9.1 Partitions and Shannon entropy

Definition 9.1 (Shannon entropy of a cell family)**.** The *Shannon entropy* of a finite family of cells $s : \iota \to \mathcal{P}(\alpha)$ with respect to a measure μ is

$$H_\mu(s) \;=\; \sum_i \eta(\mu(s_i)) \;=\; -\sum_i \mu(s_i) \log \mu(s_i),$$

the average information gained by learning which cell a μ-random point lies in. The definition is on *loose data* — an arbitrary `Fintype`-indexed family of sets — so it applies both to genuine partitions and to intermediate non-partition families.

Definition 9.2 (Finite measurable partition)**.** A *finite measurable partition* of (α, μ) is a structure bundling a `Fintype`-indexed family of cells $A_i \subseteq \alpha$ together with proofs that each cell is measurable, that the cells are pairwise almost-everywhere disjoint, and that they cover the whole space, $\bigcup_i A_i = \alpha$.

Lemma 9.3 (Entropy is at most $\log k$)**.** *A finite measurable partition of a probability space into k cells (with $k \ge 1$) has Shannon entropy at most $\log k$.*

Proof. The cell measures $p_i = \mu(A_i)$ sum to 1 by finite additivity over the a.e.-disjoint cover. Jensen's inequality for the concave η with equal weights $1/k$ gives $k^{-1}H \le \eta(\sum_i k^{-1}p_i) = \eta(1/k) = k^{-1}\log k$, and the positive factor k^{-1} cancels. □

Theorem 9.4 (Subadditivity under joins)**.** *For two finite measurable partitions $\alpha = (A_i)$ and $\beta = (B_j)$ of a probability space, the entropy of the* join *(common refinement) $\alpha \vee \beta$, whose cell family at (i, j) is the intersection $A_i \cap B_j$, satisfies*

$$H(\alpha \vee \beta) \ \leq \ H(\alpha) + H(\beta).$$

Proof. The discrete Gibbs inequality $\sum_x p_x \log p_x \geq \sum_x p_x \log q_x$ (proved termwise from $\log(q/p) \leq q/p - 1$) is applied to the joint distribution $p_{(i,j)} = \mu(A_i \cap B_j)$ and the product distribution $q_{(i,j)} = \mu(A_i)\mu(B_j)$. The marginal identities $\sum_j p_{(i,j)} = \mu(A_i)$ and $\sum_i p_{(i,j)} = \mu(B_j)$ make both vectors probability vectors and expand $-\sum p_{(i,j)} \log q_{(i,j)}$ into exactly $H(\alpha) + H(\beta)$. □

Lemma 9.5 (Invariance under a measure-preserving pullback)**.** *For a measure-preserving T and a finite measurable partition β, the* pullback partition *$T^{-1}\beta$ with cells $T^{-1}B_j$ has the same Shannon entropy as β: $H(T^{-1}\beta) = H(\beta)$.*

Proof. Each cell measure is preserved, $\mu(T^{-1}B_j) = \mu(B_j)$, so the corresponding η-terms agree summand by summand. □

9.2 Conditional entropy

Definition 9.6 (Conditional Shannon entropy)**.** On a standard Borel probability space, the *conditional Shannon entropy* of a finite family of cells s given a sub-σ-algebra $\mathcal{A}$ is

$$H(s \mid \mathcal{A}) \ = \ \int_\alpha \sum_i \eta(\kappa_\omega(s_i))\, d\mu(\omega),$$

the μ-average of the pointwise entropy computed against the regular conditional probability kernel $\kappa = \mathrm{condExpKernel}\, \mu\, \mathcal{A}$. It measures the information about the partition remaining after conditioning on $\mathcal{A}$; conditioning on the trivial σ-algebra recovers the ordinary entropy.

Theorem 9.7 (Conditioning does not increase entropy)**.** *For any finite measurable partition P of a standard Borel probability space and any sub-σ-algebra $\mathcal{A} \leq m_\alpha$,*

$$H(P \mid \mathcal{A}) \ \leq \ H(P).$$

Proof. Jensen's inequality for the concave η, applied cell by cell: the μ-average of $\eta(\kappa_\omega(P_i))$ is at most η of the average $\int \kappa_\omega(P_i)\, d\mu(\omega)$, which the disintegration identity $\int \kappa_\omega(P_i)\, d\mu = \mu(P_i)$ turns into $\eta(\mu(P_i))$. Summing over i gives the bound. □

9.3 Kolmogorov–Sinai entropy as a Fekete limit

Definition 9.8 (Flat iterated join)**.** For a measure-preserving T and a finite measurable partition α, the *iterated join* $\bigvee_{k=0}^{n-1} T^{-k}\alpha$ is realized as the partition indexed by the *flat* function type $\mathrm{Fin}\, n \to \iota$, whose cell at f is the intersection $\bigcap_{k<n} T^{-k}(\alpha_{f(k)})$. Cells are measurable as finite intersections of preimages under the iterates T^k, pairwise a.e. disjoint (two distinct indices disagree at some coordinate k, where measure-preservation of T^k transports the disjointness of α), and cover the space. For $n = 0$ this is the trivial one-cell partition.

Definition 9.9 (Iterated-join entropy sequence)**.** The *iterated-join entropy sequence* of (T, α) is

$$n \longmapsto H\Big(\bigvee_{k=0}^{n-1} T^{-k}\alpha\Big),$$

the Shannon entropy of the flat n-fold join. Its term at 0 is 0, and every term is nonnegative.

Theorem 9.10 (Subadditivity of the entropy sequence)**.** *For a measure-preserving T on a probability space and all n, m,*

$$H\Big(\bigvee_{k=0}^{n+m-1} T^{-k}\alpha\Big) \;\le\; H\Big(\bigvee_{k=0}^{n-1} T^{-k}\alpha\Big) + H\Big(\bigvee_{k=0}^{m-1} T^{-k}\alpha\Big).$$

Proof. Splitting the index type along $\mathrm{Fin}\, n \oplus \mathrm{Fin}\, m \simeq \mathrm{Fin}(n+m)$ exhibits, cell by cell, the $(n+m)$-fold join as the ordinary join of the n-fold join with the T^n-pullback of the m-fold join. Join subadditivity (Theorem 9.4) bounds the entropy by the sum of the two join entropies, and pullback invariance (Theorem 9.5) — T^n is measure-preserving — identifies the second summand with the m-fold join entropy. □

Definition 9.11 (Per-partition Kolmogorov–Sinai entropy)**.** The *Kolmogorov–Sinai entropy of T relative to a finite partition α* is the Fekete limit of the subadditive iterated-join entropy sequence,

$$h(T, \alpha) \;=\; \lim_{n\to\infty} \frac{1}{n}\, H\Big(\bigvee_{k=0}^{n-1} T^{-k}\alpha\Big),$$

packaged through Mathlib's `Subadditive.lim` (an infimum of the averaged terms, so the definition is total).

Theorem 9.12 (Fekete convergence)**.** *The averaged iterated-join entropies converge: $\frac{1}{n}H(\bigvee_{k<n} T^{-k}\alpha) \to h(T, \alpha)$ as $n \to \infty$.*

Proof. Fekete's subadditivity lemma (`Subadditive.tendsto_lim`) applied to the sequence of Theorem 9.10; its boundedness-below hypothesis is discharged from the nonnegativity of the entropies, each averaged term being at least 0. □

Definition 9.13 (Kolmogorov–Sinai entropy of the system)**.** The *Kolmogorov–Sinai entropy of the system* is the supremum of the partition-relative entropies over all finite measurable partitions,

$$h(T) \;=\; \sup_{\alpha} h(T, \alpha),$$

taken in the complete lattice $\overline{\mathbb{R}} = [-\infty, +\infty]$ (`EReal`) so that the definition is total even when the entropy is infinite. The supremum ranges over $\mathrm{Fin}\, n$-indexed partitions for every $n \in \mathbb{N}$, which by reindexing invariance realizes the full partition supremum; each $h(T, \alpha)$ is below $h(T)$ (), and $h(T) \ge 0$, witnessed by the trivial one-cell partition.

9.4 The generator theorems

Computing $h(T)$ from the definition requires exhausting all finite partitions. The Kolmogorov–Sinai generator theorem collapses the supremum to a single partition, provided that partition generates the full σ-algebra under the dynamics.

Definition 9.14 (One-sided generating partition)**.** A finite measurable partition P is *(one-sided) generating* for (α, T, μ) if the forward-saturated σ-algebra it generates is the ambient one:

$$\bigvee_{n\in\mathbb{N}} T^{-n}\sigma(P) \;=\; m_\alpha,$$

where $\sigma(P)$ is the σ-algebra generated by the cells of P. This is the standard generator condition for a (possibly non-invertible) endomorphism: the whole measurable structure is recovered from the observable P and its forward time-translates.

Theorem 9.15 (Kolmogorov–Sinai generator theorem)**.** *Let T be a measure-preserving transformation of a standard Borel probability space and P a one-sided generating finite partition. Then the entropy of the system is attained on P:*

$$h(T) \;=\; h(T, P).$$

No ergodicity is required; standard Borelness and the probability hypothesis feed the kernel disintegration, and the saturation supplied by the generating hypothesis is load-bearing.

Proof. By $\leq$-antisymmetry; $h(T, P) \leq h(T)$ is free from the defining supremum. For the converse it suffices to bound $h(T, Q) \leq h(T, P)$ for every finite partition Q. Apply the unconditional Abramov–Rokhlin partition identity (Theorem 9.20) to the refinement $R = Q \vee P$ over the factor P, conditioning on the *full* σ-algebra: the conditional entropy against m_α vanishes (each kernel cell mass is a.e. 0 or 1, and η kills both), the generating hypothesis supplies the saturation $\bigvee_n \sigma(P\text{-join}) = m_\alpha$ that the moving-index limit needs, and each $(Q \vee P)$-join cell lies inside a single P-join cell. Hence $h(T, Q) \leq h(T, Q \vee P) = h(T, P) + 0$. □

Definition 9.16 (Two-sided generating partition)**.** For a measure-preserving *automorphism* $e : \alpha \simeq \alpha$, a finite partition P is *two-sided generating* if the σ-algebra saturated under both time directions is the ambient one:

$$\bigvee_{n\in\mathbb{Z}} e^{-n}\sigma(P) \;=\; m_\alpha.$$

This is the correct condition for invertible systems: forward one-sided generation is strictly stronger and provably fails for genuine automorphisms such as the two-sided Bernoulli shift.

Theorem 9.17 (Two-sided generator theorem)**.** *Let e be a measure-preserving automorphism of a standard Borel probability space and P a two-sided generating finite partition. Then*

$$h(e) \;=\; h(e, P).$$

Proof. Again by reduction to $h(e, Q) \leq h(e, P)$ for arbitrary Q, but along the *symmetric windows* $\bigvee_{j=-N}^{N} e^j P$, realized as the e^N-pullback of the forward $(2N{+}1)$-fold join. Invertibility enters exactly once, in the window-entropy invariance $h(e, \bigvee_{-N}^{N} e^j P) = h(e, P)$, assembled from the iterate invariance $h(e, e^{-c}R) = h(e, R)$ ($c \in \mathbb{Z}$, using the inverse e^{-1}) and the block identity $h(e, \bigvee_0^{k-1} e^{-i}P) = h(e, P)$ via a sequence-level telescoping and a Cesàro shift. Conditioning Q on the increasing window σ-algebras and letting $N \to \infty$, the relative term vanishes by Lévy-upward continuity since the windows saturate m_α by the two-sided generating hypothesis, leaving $h(e, Q) \leq h(e, P)$. □

9.5 The Abramov–Rokhlin addition formula

Definition 9.18 (Relative Kolmogorov–Sinai entropy of a partition)**.** For a sub-σ-algebra $\mathcal{A} \le m_\alpha$ with $T^{-1}\mathcal{A} \le \mathcal{A}$ (forward invariance), the *relative entropy of T on a partition α given $\mathcal{A}$* is the Fekete limit of the conditional iterated-join entropies,

$$h(T, \alpha \mid \mathcal{A}) \;=\; \lim_{n\to\infty} \frac{1}{n} H\Big(\bigvee_{k=0}^{n-1} T^{-k}\alpha \;\Big|\; \mathcal{A}\Big),$$

the conditional sequence being subadditive by the conditional analogues of Theorem 9.4 and Theorem 9.5.

Definition 9.19 (Relative Kolmogorov–Sinai entropy of the system)**.** The *relative entropy of the system given $\mathcal{A}$* is the $\overline{\mathbb{R}}$-valued supremum $h(T \mid \mathcal{A}) = \sup_\alpha h(T, \alpha \mid \mathcal{A})$ over all finite measurable partitions, mirroring Theorem 9.13.

Theorem 9.20 (Abramov–Rokhlin partition identity)**.** *Let T preserve μ on a standard Borel probability space, let $\mathcal{A} \le m_\alpha$ be forward invariant, and let P, Q be finite partitions such that (i) each cell of the n-fold P-join lies μ-a.e. inside a single cell of the n-fold Q-join (a refinement witness g), and (ii) the join filtration of Q saturates $\mathcal{A}$: $\bigvee_n \sigma(\bigvee_{k<n} T^{-k}Q) = \mathcal{A}$. Then*

$$h(T, P) \;=\; h(T, Q) + h(T, P \mid \mathcal{A}).$$

The saturation hypothesis is essential: the identity is false without it.

Proof. Per n, the refinement collapses the join $H(A_n \vee B_n)$ to $H(B_n)$ plus the conditional term, and the finite conditional chain rule expresses $H(A_n) = H(B_n) + H(A_n \mid \sigma(B_n))$ for the join partitions $A_n = \bigvee_{k<n} T^{-k}P$, $B_n = \bigvee_{k<n} T^{-k}Q$. Dividing by n, the first two averaged terms converge to $h(T, P)$ and $h(T, Q)$ by Fekete. For the conditional term the conditioning σ-algebra *moves* with n; a martingale (Lévy-upward) σ-convergence argument, valid under the saturation (ii), shows the Cesàro defect between conditioning on $\sigma(B_n)$ and on the fixed limit $\mathcal{A}$ vanishes, so the moving conditional average converges to $h(T, P \mid \mathcal{A})$. □

Theorem 9.21 (Abramov–Rokhlin addition formula)**.** *Let $\pi : (\alpha, T, \mu) \to (\beta, S, \nu)$ be a factor map of measure-preserving systems on probability spaces (α standard Borel), and let P, R be finite partitions of the total and factor systems, supplied with the generator reductions $h(T) = h(T, P)$, $h(S) = h(S, R)$, the relative-generator reduction $h(T \mid \pi^{-1}\mathcal{B}_\beta) = h(T, P \mid \pi^{-1}\mathcal{B}_\beta)$, and the partition-level identity of Theorem 9.20 for P over $\pi^{-1}R$. Then the entropy of the total system splits as*

$$h(T) \;=\; h(S) + h(T \mid \pi^{-1}\mathcal{B}_\beta),$$

the factor entropy plus the fibrewise entropy relative to the pulled-back factor σ-algebra $\pi^{-1}\mathcal{B}_\beta =$ comap $\pi\, m_\beta$.

Proof. Pure $\overline{\mathbb{R}}$-algebra threading three proved ingredients: rewrite $h(T) = h(T, P)$ by the generator reduction, expand by the partition-level identity into $h(T, \pi^{-1}R) + h(T, P \mid \pi^{-1}\mathcal{B}_\beta)$, identify $h(T, \pi^{-1}R) = h(S, R)$ by the factor-relative invariance (entropies of pulled-back joins agree cell-measure by cell-measure since π intertwines the dynamics and pushes μ to ν), and fold back with the two remaining reductions. In the sharpened form `abramov_rokhlin_of_W3` the partition identity is discharged down to the single martingale σ-convergence input, and when R generates the base (`IsGenerating` ν S R) that too is proved, making the formula unconditional. □

9.6 The Margulis–Ruelle inequality and Rokhlin's formula

The bridge back to the multiplicative theory of Chapters 2–6: for a smooth ergodic self-map T of $\mathbb{R}^d$ (as `EuclideanSpace`), the derivative cocycle $x \mapsto D_x T$ satisfies the standing hypotheses of the multiplicative ergodic theorem, and its Lyapunov exponents $\lambda_1 \geq \ldots \geq \lambda_d$ (Theorem 6.11) control the Kolmogorov–Sinai entropy from above.

Theorem 9.22 (Margulis–Ruelle inequality, abstract reduction)**.** *Let T be an ergodic self-map of $\mathbb{R}^d$ with everywhere-nonsingular derivative cocycle and one-sided log-integrable derivative data (Theorem 2.4 for DT and DT^{-1}), and suppose the* geometric atom-counting bound *holds: every finite partition α satisfies $h(T,\alpha) \leq \sum_{\lambda_i>0} \lambda_i$, the sum of the strictly positive Lyapunov exponents of the derivative cocycle (a deterministic constant, the spectrum being a.e. constant by ergodicity). Then the system entropy obeys*

$$h(T) \;\leq\; \sum_{\lambda_i>0} \lambda_i.$$

Proof. The pure lattice step: unfold the defining double supremum of $h(T)$ over arities n and Fin n-indexed partitions, and apply `iSup_le` twice in $\overline{\mathbb{R}}$ to the uniform per-partition hypothesis. The content of this node is the exact interface: everything around the genuinely geometric per-partition estimate is proved. □

Theorem 9.23 (Sharp Margulis–Ruelle inequality)**.** *Same setting (ergodic T on $\mathbb{R}^d$, nonsingular log-integrable derivative cocycle). Assume only the honest non-compactness input — the Mañé/Katok geometric atom count: for every finite partition P there are $\varepsilon > 0$ and $C \geq 0$ such that, μ-a.e., the number of atoms of the n-fold refinement meeting an orbit eventually satisfies*

$$\#\{atoms\} \;\leq\; C \cdot \mathcal{N}_\varepsilon(D_x(T^n)\,\overline{B}(0,\varepsilon)),$$

the ε-covering number of the image ellipsoid. Then

$$h(T) \;\leq\; \sum_{\lambda_i>0} \lambda_i.$$

Some such distortion hypothesis is unavoidable on the noncompact phase space: the bare inequality is false in general (Riquelme 2017).

Proof. The one-step *sharp anisotropic covering count* $\mathcal{N}_\varepsilon(L\,\overline{B}(0,\varepsilon)) \;\leq\; 6^d \prod_i \max(1, \sigma_i(L))$ is proved in-tree by a constructive SVD ellipsoid domination, and is composed with the assumed atom count to bound the atom growth by the positive singular-value product $\prod_i \max(1, \sigma_i(D_x T^n))$. Along a.e. orbit, $\frac{1}{n}\log$ of that product tends to $\sum \lambda_i^+$ by the multiplicative ergodic theorem (Theorem 5.30); the log-cardinality bound passes through the Fekete limit to give $h(T,P) \leq \sum_{\lambda_i>0} \lambda_i$ per partition, and Theorem 9.22 folds the supremum. □

Theorem 9.24 (Rokhlin's volume-distortion formula)**.** *Let T be a measure-preserving, differentiable self-map of $\mathbb{R}^d$ with $\mu \ll \mathrm{vol}$, everywhere-nonsingular derivative, a one-sided generating finite partition ξ on whose cells T is injective, and μ-integrable $\log\rho$ (ρ the density $d\mu/d\mathrm{vol}$) and $\log|\det DT|$. Then the per-partition entropy is the integrated volume distortion:*

$$h(T,\xi) \;=\; \int \log\,|\det D_x T|\; d\mu(x).$$

Together with the determinant identity of Theorem 6.20, the right-hand side is the sum $\sum_i \lambda_i$ of all *Lyapunov exponents of the derivative cocycle — so for expanding maps (all exponents positive) this saturates the Margulis–Ruelle bound, giving Pesin's entropy formula in that regime.*

Proof. Three identities composed. First, the sharp-rate form of the Fekete limit expresses $h(T, \xi)$ as the conditional entropy $H(\xi \mid \bigvee_k \sigma(T^{-1}\xi\text{-joins}))$ of the partition against its own strict future. Second, the generating hypothesis glues that future σ-algebra to comap $T\, m_\alpha$. Third, Coudène's conditional-expectation computation evaluates $H(\xi \mid T^{-1} m_\alpha)$ by the change-of-variables formula on each injectivity cell, producing $\int \log|\det(D_x T)|\, d\mu$; a final bridge identifies the Fréchet-derivative determinant with the determinant of the matrix derivative cocycle. An instantiated equality (the doubling map on the circle) is exhibited elsewhere in the library. □

Chapter 10

Generators: Shannon–McMillan–Breiman and Krieger's theorem

The Kolmogorov–Sinai entropy of the classical-entropy development measures the average information produced by a transformation per unit time. This chapter documents the *generator* side of that theory (the 29 modules of `ErgodicTheory/Krieger/`): the pointwise *Shannon–McMillan–Breiman theorem* (entropy equipartition), the *Rokhlin–Kakutani tower lemma*, the symbolic coding stack built on them (name counting and sentinel prefix codes), and the two generator headliners — the countable finite-entropy two-sided generator of Rokhlin, Keane–Serafin and Downarowicz, and *Krieger's finite generator theorem*: an ergodic aperiodic automorphism of entropy below $\log k$ admits a two-sided generating partition with at most k cells.

Throughout, (α, μ) is a probability space, T a measure-preserving transformation, and P a finite measurable partition indexed by a finite type ι (a `MeasurePartition`: measurable cells, pairwise a.e.-disjoint, covering α). For the coding and generator layers, e is a measure-preserving *automorphism* (a measurable equivalence $\alpha \simeq \alpha$) of a standard Borel probability space and iterates range over $\mathbb{Z}$. We write $\bigvee_{k=0}^{n-1} T^{-k}P$ for the n-fold iterated join, $H(\cdot)$ for Shannon entropy, and $h(P, T)$ for the Kolmogorov–Sinai entropy of the pair. Every result below is formalized sorry-free; the two headline generator theorems are stated relative to explicitly disclosed hypothesis bundles (their honest formalization boundary, described in place).

10.1 The information function and its chain rules

Definition 10.1 (Information function of the iterated join)**.** Every point x has an n-step *itinerary* $f : \mathrm{Fin}\, n \to \iota$, recording for each $k < n$ which cell of P contains $T^k x$; among the admissible codes the *least* one in a fixed enumeration is chosen, which makes the selector genuinely measurable (the cells may overlap on a null set, so a bare choice function would not be). The *atom of* x is the cell $\bigcap_{k<n} T^{-k}(P_{f(k)})$ of the iterated join containing x, and the *information function* is

$$i_n(x) \;=\; -\log \mu(\mathrm{atom}_n(x)),$$

the surprise of learning the n-name of x. It is nonnegative and measurable (): in its simple-function form it is a finite sum of indicators of the itinerary fibers, each the difference of a join cell and the finitely many strictly smaller cells.

Theorem 10.2 (The information function integrates to the join entropy)**.** *For every n,*

$$\int_\alpha i_n \, d\mu \;=\; H\Big(\bigvee_{k=0}^{n-1} T^{-k}P\Big),$$

the Shannon entropy of the n-fold iterated join (the n-th term of the Kolmogorov–Sinai entropy sequence).

Proof. Write i_n as the finite sum of indicators of the itinerary fibers weighted by $-\log \mu(\text{cell})$ and integrate termwise. Each fiber differs from its join cell only inside the union of the *other* cells — a null set by pairwise a.e.-disjointness — so fiber and cell have the same measure, and each summand becomes $\mu(\text{cell}) \cdot (-\log \mu(\text{cell})) = \text{negMulLog}\, \mu(\text{cell})$, which summed over the cells is the join entropy. □

Theorem 10.3 (Telescoped Breiman chain rule, entropy level)**.** *On a standard Borel space, the n-step join entropy telescopes into conditional entropies:*

$$H\Big(\bigvee_{k=0}^{n-1} T^{-k}P\Big) \;=\; \sum_{k=0}^{n-1} H\Big(P \,\Big|\, \sigma(T^{-1} \textstyle\bigvee_{j=0}^{k-1} T^{-j}P)\Big),$$

where the k-th conditioning σ-algebra is generated by the cells of the T-pullback of the k-step join.

Proof. Induction on n from the one-step chain rule: reindexing along $\text{Fin}\, n \times \iota \simeq \text{Fin}(n{+}1)$ (peeling the first symbol) identifies the $(n{+}1)$-fold join with the join $B \vee P$, $B := T^{-1}(\bigvee_0^{n-1} T^{-k}P)$; the absolute chain rule $H(B \vee P) = H(B) + H(P \mid B)$ and T-invariance of join entropy give $H(\bigvee_0^{n}) = H(\bigvee_0^{n-1}) + H(P \mid \sigma(B))$, the conditional entropy bridged from its additive-over-cells form to the σ-algebra form. □

Theorem 10.4 (The sharp KS rate as a conditional entropy)**.** *On a standard Borel space (with ι nonempty), the Kolmogorov–Sinai entropy of the pair equals the conditional entropy of P given the σ-algebra of the strict future:*

$$h(P,T) \;=\; H\Big(P \,\Big|\, \bigvee_k \sigma(T^{-1} \textstyle\bigvee_{j=0}^{k-1} T^{-j}P)\Big) \;=\; H\Big(P \,\Big|\, \sigma\Big(\textstyle\bigvee_{j\ge 1} T^{-j}P\Big)\Big).$$

This is the integral-level statement of the sharp SMB rate: the precise constant the pointwise theorem converges to. No ergodicity is needed.

Proof. The conditioning σ-algebras increase in k, so the fixed-partition Lévy theorem (martingale convergence of conditional entropies) sends $H(P \mid \sigma_k) \to H(P \mid \bigvee_k \sigma_k)$. By Theorem 10.3, $\frac{1}{n} H(\bigvee_0^{n-1} T^{-k}P)$ is exactly the Cesàro average of that convergent sequence, hence tends to the same limit; matching it against the Fekete limit defining $h(P,T)$ gives the identity. □

10.2 The pointwise Shannon–McMillan–Breiman theorem

Theorem 10.5 (Crude name-count bound, Birkhoff-free)**.** *For a measure-preserving T and a finite partition indexed by a nonempty ι,*

$$\limsup_{n\to\infty} \frac{i_n(x)}{n} \;\le\; \log \#\iota \qquad \textit{for } \mu\textit{-a.e. } x.$$

Neither ergodicity nor any ergodic theorem enters: this is the Algoet–Cover engine fed the uniform competing measure, whose partition-function bound $\int \exp(i_n - n\log\#\iota)\,d\mu \le 1$ is the lemma .

Proof. On each itinerary fiber the integrand $\exp(i_n - n\log\#\iota)$ is the constant $\mu(\text{cell})^{-1}\cdot(\#\iota)^{-n}$ (or vanishes with the cell), and the join has at most $(\#\iota)^n$ non-null cells, giving the partition-function bound. Markov's inequality plus Borel–Cantelli then make $\frac{1}{n}i_n \le \log\#\iota + \varepsilon$ eventually a.e. for every $\varepsilon > 0$. ☐

Definition 10.6 (Conditional information function)**.** For a sub-σ-algebra $\mathcal{A}$ of a standard Borel space, the *conditional information function* of P given $\mathcal{A}$ is

$$g_{\mathcal{A}}(x) \;=\; \sum_i \mathbf{1}_{P_i}(x)\cdot(-\log\mu[P_i \mid \mathcal{A}](x)),$$

the conditional probability realized by the regular conditional probability kernel; exactly one indicator survives at each point. Its integral is the conditional entropy: for $\mathcal{A} \le m_\alpha$, $\int g_{\mathcal{A}}\,d\mu = H(P \mid \mathcal{A})$ (); in particular $g_{\mathcal{A}}$ is integrable.

Theorem 10.7 (Chung's maximal inequality)**.** *Let g_k be the conditional information function of P given the k-th Breiman conditioning σ-algebra $\mathcal{C}_k = \sigma(T^{-1}\bigvee_{j=0}^{k-1} T^{-j}P)$, and let $g^* = \sup_k g_k$ be the Chung maximal information function (, valued in $[0,\infty]$). Then for every cell P_{i_0} and every $\lambda > 0$,*

$$\mu(\{x \in P_{i_0} \mid g^*(x) > \lambda\}) \;\le\; e^{-\lambda}.$$

Consequently $g^* \in L^1(\mu)$, with $\int g^*\,d\mu \le H(P) + 1$ () — Chung's L^1 domination, the one genuinely analytic leaf of the SMB proof.

Proof. A stopping-time (first-passage) argument: stratify $\{g^* > \lambda\} \cap P_{i_0}$ by the first level k at which the conditional probability of P_{i_0} drops below $e^{-\lambda}$. Each stratum is $\mathcal{C}_k$-measurable, so its intersection with P_{i_0} has measure at most $e^{-\lambda}$ times the stratum's measure by the defining property of conditional probability; summing the disjoint strata gives the tail bound. Integrability follows by the layer-cake formula: summing the per-cell tails yields $\int g^* \le H(P) + 1$. ☐

Theorem 10.8 (Pointwise SMB, from the Breiman telescoping)**.** *Let T be ergodic and let (i_n) be any sequence of functions satisfying, for μ-a.e. x and all n, the Breiman telescoping identity*

$$i_n(x) \;=\; \sum_{j=0}^{n-1} g_{n-j}(T^j x),$$

where g_k is the conditional information function of P given $\mathcal{C}_k$. Then

$$\frac{i_n(x)}{n} \;\longrightarrow\; h(P,T) \qquad \text{for } \mu\text{-a.e. } x.$$

Proof. Split $\frac{1}{n}\sum_{j<n} g_{n-j}(T^j x)$ into the Birkhoff main term $\frac{1}{n}\sum_{j<n} g_\infty(T^j x)$ and the Cesàro tail $\frac{1}{n}\sum_{j<n}(g_{n-j}-g_\infty)(T^j x)$, where g_∞ is the conditional information function given the strict-future σ-algebra $\bigvee_k \mathcal{C}_k$. The main term converges a.e. to $\int g_\infty\,d\mu = h(P,T)$ by the pointwise ergodic theorem and the sharp rate identity (Theorem 10.4). The tail vanishes a.e. by the Maker/Breiman dominated-Cesàro argument (): $|g_{n-j} - g_\infty|$ is dominated by the integrable maximal function $g^* + g_\infty$ of Theorem 10.7, and Lévy downward convergence $g_k \to g_\infty$ a.e. makes the dominated Birkhoff averages of the sup-tails vanish. ☐

Theorem 10.9 (Pointwise Shannon–McMillan–Breiman theorem). *Let T be an ergodic measure-preserving transformation of a standard Borel probability space and P a finite measurable partition (nonempty index). Then the concrete information functions of Theorem 10.1 satisfy*

$$\frac{i_n(x)}{n} \longrightarrow h(P,T) \qquad \text{for } \mu\text{-a.e. } x:$$

the measure of the atom of x in the n-fold join decays like $e^{-n\,h(P,T)}$ (entropy equipartition, the AEP of ergodic theory).

Proof. The one-step factorization of the information weight (peeling the first symbol of the itinerary) telescopes, a.e., the concrete $i_{n+1}(x)$ into the Breiman partial sum $\sum_{j<n} g_{n-j}(T^j x)$ plus a single edge term $g_0(T^n x)$. The partial sum converges after division by n via Theorem 10.8; the edge term dies since $\frac{1}{n} g_0(T^n x) \to 0$ a.e. for the integrable g_0; an index shift finishes. □

Theorem 10.10 (In-measure upper equipartition). *For ergodic T, the in-measure SMB upper bound holds: for every $\varepsilon > 0$,*

$$\mu\Big(\Big\{x \;\Big|\; h(P,T) + \varepsilon < \tfrac{1}{N}\, i_N(x)\Big\}\Big) \longrightarrow 0 \qquad (N \to \infty).$$

This McMillan-type form is exactly what the name-count covering bound below consumes.

Proof. Almost-everywhere convergence (Theorem 10.9) makes the deviation sets eventually empty along a.e. orbit; convergence in measure of their indicators to the empty set follows on the finite measure space. □

10.3 The Rokhlin tower lemma

Theorem 10.11 (Rokhlin–Kakutani tower lemma). *Let e be an ergodic measure-preserving automorphism of a standard Borel probability space with non-atomic μ. For every height $N \geq 1$ and every $\varepsilon > 0$ there is a measurable base B whose first N iterates $B, eB, \dots, e^{N-1}B$ are pairwise disjoint and whose union covers all but ε of the space:*

$$\mu\Big(\bigcup_{i=0}^{N-1} e^i B\Big) > 1 - \varepsilon.$$

(Ergodicity plus non-atomicity supplies the aperiodicity the classical statement assumes.)

Proof. The Kakutani skyscraper construction. Pick a positive measurable set A with $\mu(A) < \varepsilon/N$; by ergodicity a.e. point returns to A, so the return-time levels of A tile the space mod 0. Grouping each return column into complete blocks of height N yields the base B; the levels are disjoint by construction, and the uncovered part lies in the top incomplete floors of the columns, of total measure at most $N\mu(A) < \varepsilon$. □

10.4 The coding stack

The passage from equipartition to a finite generator is combinatorial: count the names that carry most of the mass, encode them into a slightly larger alphabet by a prefix-free (sentinel) code painted along Rokhlin towers, and phrase "generator" as a mod-0 statement about the two-sided itinerary σ-algebra. We record one representative node for each layer.

Theorem 10.12 (AEP covering bound: few names carry the mass)**.** *Let T be measure preserving and assume the in-measure SMB upper bound of Theorem 10.10. Write $h = h(P,T)$. For every $\varepsilon > 0$ and all sufficiently large N, there is a finite set S of rank-N names $g : \mathrm{Fin}\,N \to \iota$ with*

$$\mu\Big(\bigcup_{g\in S}\ \bigcap_{k<N} T^{-k}P_{g(k)}\Big) \ \geq\ 1-\varepsilon \qquad \text{and} \qquad \#S\ \leq\ \lfloor e^{N(h+\varepsilon)}\rfloor.$$

The set S is the set of good names *() — names whose join cell has measure at least $e^{-N(h+\varepsilon)}$; the cardinality bound is the unconditional pigeonhole .*

Proof. Pigeonhole: cells of measure $\geq e^{-N(h+\varepsilon)}$ number at most $e^{N(h+\varepsilon)}$, since they are essentially disjoint in a probability space. Coverage: the good cells cover the set $\{\frac{1}{N}i_N \leq h+\varepsilon\}$, whose complement has measure $< \varepsilon$ eventually by the in-measure SMB upper bound. □

Theorem 10.13 (Sentinel prefix code)**.** *Over an alphabet* $\mathrm{Fin}\,l$ *with a reserved sentinel letter s, any finite name set with* $\#\mathrm{Name} \leq (l-1)^m$ *admits an injection* $\mathrm{enc} : \mathrm{Name} \hookrightarrow \mathrm{List}(\mathrm{Fin}\,l)$ *into blocks of length $m+1$, each containing the sentinel exactly once (at its end). The blocks form a prefix-free family, so arbitrary concatenations are uniquely decodable (); when $e^{N(h+\varepsilon)} < k^N$, the good names of Theorem 10.12 embed into* $\mathrm{Fin}\,k$*-blocks of length $O(N)$ — the symbolic code from which the coding partition is read off along a Rokhlin tower.*

Proof. The name set embeds into the fixed-length data words $\mathrm{Fin}\,m \to \mathrm{Fin}(l-1)$ over the non-sentinel letters by the cardinality bound; appending the sentinel produces blocks in which the sentinel marks exactly the block boundaries, so a straightforward take/drop induction recovers the block decomposition of any concatenation, giving injectivity. □

Definition 10.14 (Two-sided generation mod 0)**.** The *two-sided saturation* of a finite partition P under an automorphism e is the σ-algebra $\bigvee_{n\in\mathbb{Z}} e^{-n}\sigma(P)$ generated by the two-sided P-itinerary. The partition *two-sidedly generates* (α, e, μ) *mod* 0 when the ambient σ-algebra is contained in the μ-completion of the saturation — every measurable set is recovered from the itinerary up to a null set. This mod-0 form (not the literal equality of σ-algebras) is the faithful conclusion of Krieger's theorem. The Countable-indexed analogue for a family of cells $Q : \kappa \to \mathrm{Set}\,\alpha$ is .

Definition 10.15 (Countable Shannon entropy)**.** For a countable family of cells $s : \iota \to \mathrm{Set}\,\alpha$, the *countable Shannon entropy* is $cH_\mu(s) = \sum_i \mathrm{negMulLog}\,\mu(s_i) = -\sum_i \mu(s_i)\log\mu(s_i)$, an unconditionally convergent `tsum`. Downarowicz's Fact 1.1.4 supplies the finiteness criterion (): if the index-weighted masses satisfy $\sum_i i\cdot\mu(s_i) < \infty$, the entropy terms are summable, so $cH_\mu(s)$ is a genuine finite sum. This is the form in which the Keane–Serafin construction certifies finite entropy of its limit partition.

10.5 The generator theorems

Theorem 10.16 (Countable finite-entropy two-sided generator (Rokhlin; Keane–Serafin))**.** *Let (α,μ) be standard Borel with automorphism e, and let $Q : \mathbb{N} \to \mathrm{Set}\,\alpha$ be a countable family of measurable cells such that (i) the two-sided Q-itinerary recovers each set of a fixed standard-Borel generating sequence up to a μ-null set, and (ii) the entropy terms $i \mapsto \mathrm{negMulLog}\,\mu(Q_i)$ are summable. Then there is a countable family that two-sidedly generates (α, e, μ) mod 0 and has finite static Shannon entropy — sub-problem A of Krieger's theorem (Downarowicz, Thm. 4.2.3, first half). The structural reduction from the inductive Keane–Serafin data — an increasing sequence of finite partitions with per-level recovery and a summable geometric mass envelope —*

is ; the construction of that data from the raw dynamical hypotheses (the SMB-driven Rokhlin-tower marker painting of Keane–Serafin §2) is the disclosed residual, isolated as an explicit hypothesis bundle rather than formalized.

Proof. Recovering every set of a generating sequence mod 0 pushes the ambient σ-algebra into the μ-completion of the two-sided saturation of Q (monotone-class/completion bookkeeping on the countable saturation), which is mod-0 two-sided generation; the entropy summability is carried verbatim. In the levels form, the per-level families are enumerated through the pairing bijection $\mathbb{N} \simeq \mathbb{N} \times \mathbb{N}$, the per-level recovery lifts to the union family, and the geometric envelope certifies summability via Theorem 10.15. ☐

Definition 10.17 (Krieger coding data)**.** For an automorphism e and $k \in \mathbb{N}$, a *Krieger coding datum* bundles: the measure-preservation of e; a countable index type κ with measurable cells $Q : \kappa \to \mathrm{Set}\,\alpha$ two-sidedly generating mod 0; and a $\mathrm{Fin}\,k$-valued partition P that *codes* Q two-sidedly mod 0 (): the two-sided P-itinerary recovers each Q-cell up to a μ-null set. This is exactly the object the Krieger construction produces when $\log k$ exceeds the entropy: Q from Theorem 10.16, and P from the column coding of the $\leq k^N$ good names (Theorem 10.12 and Theorem 10.13) along refining Rokhlin towers (Theorem 10.11).

Theorem 10.18 (Recovery assembly of Krieger's theorem)**.** *Given a Krieger coding datum D for (e, μ, k), there exists a partition P into at most k cells that two-sidedly generates (α, e, μ) mod 0. No entropy, ergodicity or aperiodicity hypotheses enter: those are consumed upstream, in producing D. This is the pure recovery content of Krieger's theorem.*

Proof. Cross-layer recovery. Mod-0 generation by Q places the ambient σ-algebra inside the completion of the two-sided Q-saturation; since the P-itinerary recovers each Q-cell mod 0, that saturation lies inside the completion of the two-sided P-saturation; completion-monotonicity and idempotence chain the inclusions, so D's own coding partition is the generator. ☐

Theorem 10.19 (Krieger's finite generator theorem)**.** *Let e be an ergodic, aperiodic (: every set of n-periodic points, $n \neq 0$, is μ-null), measure-preserving automorphism of a probability space, and let $k \in \mathbb{N}$ satisfy*

$$h(e) \ < \ \log k,$$

where $h(e)$ is the Kolmogorov–Sinai entropy of the system. If the Krieger coding construction supplies a coding datum (Theorem 10.17), then e admits a finite two-sided generator of size $\leq k$, mod 0*: a partition P indexed by* $\mathrm{Fin}\,k$ *with* IsGeneratingTwoSidedMod0 e P*. The entropy threshold is pinned to the genuine $h(e)$, so the hypothesis is a real entropy constraint.* Interface disclosure: *the coding datum has no in-repo constructor — the tower-painting combinatorics that would discharge it from the dynamical hypotheses alone are not formalized — so this is the honest conditional form of Krieger's theorem, with the dynamical hypotheses retained as the faithful classical interface that the upstream layers (Theorem 10.11, Theorem 10.12, Theorem 10.13 and Theorem 10.16) are designed to feed.*

Proof. Immediate from the recovery assembly Theorem 10.18 applied to the supplied coding datum, which already carries the measure-preservation the cross-layer recovery needs. ☐

Chapter 11

Multifractal analysis

This chapter documents the *coarse-grained (finite-resolution) multifractal analysis* of an invariant probability measure, formalized in `ErgodicTheory/Multifractal/` (26 modules under the `ErgodicTheory.Multifractal` aggregator). The theory quantifies how unevenly a measure μ distributes mass over the cells of a finite partition at scale ε: the partition function Z_q, the mass exponent $\tau(q)$, the Rényi (generalized) dimensions D_q, and the singularity spectrum $f(\alpha)$, the Legendre transform of τ.

The development has four layers. First, an *abstract, measure-free core*: all quantities are defined on a bare finite weight family $p : \iota \to \mathbb{R}$ (think $p_i = \mu(\mathrm{cell}_i)$), with the probability hypotheses carried on the lemmas, never baked into the definitions. The structural heart is the Hölder / cumulant-convexity argument: $q \mapsto \log Z_q$ is *convex*, hence the mass exponent τ is *concave* (for a scale $0 < \varepsilon < 1$, since $\log \varepsilon < 0$ flips the sign), and D_q is *non-increasing* in q. Second, the *measure/flow layer* discharges the abstract hypotheses from an actual invariant probability measure and identifies the $q = 1$ branch with the Shannon entropy of the partition. Third, the *fine-scale (pointwise) theory*: the local dimension $d_\mu(x) = \lim_{r\to 0^+} \log \mu(B(x,r))/\log r$, proved to exist and equal the ambient dimension in the absolutely-continuous case, together with the Frostman/Billingsley bridge from pointwise dimension to *Hausdorff* dimension, and the symbolic entropy = dimension identity $\dim_H = h_\mu(\sigma)/\log 2$ on the full shift, made unconditional for Bernoulli measures. Fourth, a genuinely *multifractal witness*: the constant-roof suspension flow of a biased two-sided Bernoulli shift, an ergodic flow of positive entropy whose Rényi spectrum is provably q-dependent.

Everything below is formalized sorry-free and verified by the guarded axiom audit to rest only on $\{$`propext`, `Classical.choice`, `Quot.sound`$\}$. Throughout, ι is a finite index type and the exponent x^q is the real-base real-exponent power (`Real.rpow`).

11.1 The coarse-grained formalism

Definition 11.1 (Generalized partition function)**.** For a finite weight family $p : \iota \to \mathbb{R}$ and $q \in \mathbb{R}$, the *generalized partition function* is

$$Z_q = \sum_{i\,:\,p_i>0} p_i^q,$$

the sum over the *occupied* cells only. The positivity guard is load-bearing at $q = 0$: it forces empty cells ($p_i = 0$) to contribute 0 rather than $0^0 = 1$, so that Z_0 counts the occupied cells.

For $q \neq 0$ the guard is removable ($0^q = 0$), and for a probability family ($p_i \geq 0$, $\sum_i p_i = 1$) one has $Z_1 = 1$.

Definition 11.2 (Mass exponent)**.** The *mass exponent* of the family p at scale ε is

$$\tau(q) \;=\; \frac{\log Z_q}{\log \varepsilon},$$

defined for every q with no case split. For a probability family $\tau(1) = 0$, since $Z_1 = 1$.

Definition 11.3 (Rényi / generalized dimension)**.** The *Rényi (generalized) dimension* of p at scale ε is

$$D_q \;=\; \frac{\tau(q)}{q-1} \quad (q \neq 1), \qquad D_1 \;=\; \frac{\sum_i p_i \log p_i}{\log \varepsilon}.$$

At $q = 1$ the general formula is the indeterminate $0/0$, so the L'Hôpital value — the *information dimension* — is supplied directly as a separate branch. (By the Mathlib convention $\log 0 = 0$, the $q = 1$ numerator needs no positivity guard.)

Definition 11.4 (Singularity spectrum)**.** The *singularity spectrum* of p at scale ε is the Legendre transform of the mass exponent,

$$f(\alpha) \;=\; \inf_{q \in \mathbb{R}} \big(q\alpha - \tau(q)\big).$$

It is an *infimum*: since τ is concave (Theorem 11.6), the supremum of $q\alpha - \tau(q)$ would be $+\infty$.

11.2 Structure of the spectrum: convexity and monotonicity

Theorem 11.5 (Log-convexity of the partition function)**.** *Let $p : \iota \to \mathbb{R}$ satisfy $p_i \geq 0$ for all i and $p_i > 0$ for some i. Then $q \mapsto \log Z_q$ is* convex *on all of $\mathbb{R}$. This is the cumulant-convexity / Hölder property, the mathematical core of the multifractal theory.*

Proof. Derivative-free. The midpoint inequality $Z_{aq_1+bq_2} \leq Z_{q_1}^{a}\, Z_{q_2}^{b}$ (for $a, b > 0$, $a + b = 1$) is exactly the two-term Hölder inequality with conjugate exponents $1/a, 1/b$ applied on the support $\{i : p_i > 0\}$ (the lemma `partitionFunction_holder`). The positivity hypothesis gives $Z_q > 0$ at every q, so taking logarithms and using monotonicity of log turns the multiplicative bound into the convexity inequality for $\log \circ Z$. ☐

Theorem 11.6 (Concavity of the mass exponent)**.** *Under the same hypotheses on p, for a scale $0 < \varepsilon < 1$ the mass exponent $q \mapsto \tau(q) = \log Z_q / \log \varepsilon$ is* concave *on $\mathbb{R}$.*

Proof. Since $0 < \varepsilon < 1$, the denominator $\log \varepsilon$ is negative; multiplying the convex $\log Z_q$ by the nonpositive constant $1/\log \varepsilon$ flips convexity to concavity. Formally, $c \cdot \log Z$ with $c = -(\log \varepsilon)^{-1} \geq 0$ is convex, and τ is its negation. ☐

Theorem 11.7 (Antitonicity of the Rényi dimension)**.** *Let p be a probability weight family ($p_i \geq 0$, $\sum_i p_i = 1$, at least one $p_i > 0$) and $0 < \varepsilon < 1$. Then $q \mapsto D_q$ is* non-increasing *(`Antitone`) on all of $\mathbb{R}$ — including across the information-dimension branch point $q = 1$.*

Proof. The classical secant-slope argument. Write $h(q) = \log Z_q$; it is convex and $h(1) = 0$ for a probability family, so the secant slope $g(q) = h(q)/(q-1)$ anchored at 1 is non-decreasing on $\{q \neq 1\}$ (`ConvexOn.secant_mono`). For $q \neq 1$, $D_q = g(q)/\log \varepsilon$, and $\log \varepsilon < 0$ flips monotone to antitone. The subtle point is gluing in $q = 1$: the information-dimension numerator $\sum_i p_i \log p_i$ is exactly the derivative $h'(1)$ (each occupied summand $q \mapsto p_i^q$ is a real exponential), and the convex supporting-line inequalities give $g(q) \leq h'(1) \leq g(q')$ for $q < 1 < q'$; dividing by $\log \varepsilon < 0$ inserts D_1 into the antitone family. □

11.3 The measure and flow layer

The abstract core specializes to a genuine invariant probability measure μ together with a finite measurable partition P (a `MeasurePartition`), by taking the weight family $p_i = \mu(\mathrm{cell}_i)$. The probability hypotheses are now discharged *from the measure*: nonnegativity, the normalization $\sum_i p_i = 1$, and the existence of a cell of positive mass all follow from μ being a probability measure.

Definition 11.8 (Rényi dimension of a measure)**.** For a measure μ on α, a finite measurable partition P of μ indexed by ι, and $\varepsilon, q \in \mathbb{R}$, the *Rényi dimension of* μ at partition scale ε is the abstract D_q of the cell-mass family $i \mapsto \mu(P_i)$ (real-valued via `toReal`). The companion definitions `partitionFunctionMeasure` and `massExponentMeasure` specialize Z_q and τ in the same way.

Theorem 11.9 (Antitonicity for a probability measure)**.** *For a probability measure μ, a finite measurable partition P, and a scale $0 < \varepsilon < 1$, the Rényi dimension $q \mapsto D_q(\mu, P, \varepsilon)$ is non-increasing in q.*

Proof. Apply Theorem 11.7 to the cell-mass family: nonnegativity is `ENNReal.toReal` ≥ 0, the normalization is the partition identity $\sum_i \mu(P_i) = 1$, and at least one cell has positive mass because the total mass is 1. □

Theorem 11.10 (Information dimension is entropy over $-\log \varepsilon$)**.** *For a probability measure μ, a partition P, and any ε,*

$$D_1(\mu, P, \varepsilon) \;=\; \frac{-H(P)}{\log \varepsilon},$$

where $H(P) = \sum_i \mathrm{negMulLog}(\mu(P_i)) = -\sum_i \mu(P_i) \log \mu(P_i)$ is the Shannon entropy of the partition. For $0 < \varepsilon < 1$ this is the familiar $D_1 = H(P)/\log(1/\varepsilon)$.

Proof. Unfold the $q = 1$ branch of D_q: its numerator $\sum_i \mu(P_i) \log \mu(P_i)$ is term-by-term the negation of the negMulLog sum defining $H(P)$. □

Definition 11.11 (Rényi dimension of a flow's invariant measure)**.** For a measure-preserving flow φ with invariant probability measure μ (Definition 8.1), a partition P, and ε, q, the *Rényi dimension of the flow's invariant measure* is $D_q(\varphi, P, \varepsilon) = D_q(\mu, P, \varepsilon)$. The flow is an explicit (unused) argument whose *type* documents that μ is flow-invariant; the multifractal API consumes any invariant probability measure.

Corollary 11.12 (Flow-level antitonicity)**.** *For a measure-preserving flow φ of a probability measure μ, a partition P, and $0 < \varepsilon < 1$, the flow Rényi dimension $q \mapsto D_q(\varphi, P, \varepsilon)$ is non-increasing in q.*

Proof. `renyiDimFlow` unfolds to `renyiDimMeasure`, so this is Theorem 11.9 verbatim. □

11.4 Local dimension and Hausdorff dimension

Definition 11.13 (Upper local dimension)**.** For a measure μ on a (pseudo-)metric measurable space E and a point x, the *upper local (pointwise) dimension* is

$$\bar{d}_\mu(x) \;=\; \limsup_{r\to 0^+} \frac{\log \mu(\bar{B}(x,r))}{\log r},$$

the lim sup along the filter $r \to 0^+$ of the closed-ball mass quotient. Where the genuine limit exists (as in the absolutely-continuous case below), this lim sup is the honest local dimension.

Theorem 11.14 (Local dimension in the absolutely-continuous case)**.** *Let E be a finite-dimensional real inner-product space (Borel-measurable) and let μ be a probability measure on E absolutely continuous with respect to an additive Haar measure ν (e.g. Lebesgue). Then for μ-almost every x the local-dimension quotient* $\log \mu(\bar{B}(x,r))/\log r$ converges*, as $r \to 0^+$, to the ambient dimension* $\dim_{\mathbb{R}} E = \operatorname{finrank}_{\mathbb{R}} E$*.*

Proof. Pure measure differentiation, no dynamics. Besicovitch differentiation gives $\mu(\bar{B}(x,r))/\nu(\bar{B}(x,r)) \to (d\mu/d\nu)(x)$ as $r \to 0^+$, μ-a.e., with the Radon–Nikodym density finite and positive μ-a.e. (from $\mu \ll \nu$). The Haar ball-volume scaling $\nu(\bar{B}(x,r)) = r^d\,\nu(\bar{B}(0,1))$ with $d = \operatorname{finrank}_{\mathbb{R}} E$ factorizes the ball mass as $\operatorname{ratio}(r)\cdot r^d \cdot C$ with $\operatorname{ratio}(r) \to L > 0$ and $C > 0$; a logarithm-limit lemma then shows the quotient is $(\log \operatorname{ratio}(r) + \log C)/\log r + d \to d$, since $\log r \to -\infty$ kills the bounded numerator. □

Corollary 11.15 (A.e. value of the local dimension)**.** *Under the same hypotheses, $\bar{d}_\mu(x) =$* $\operatorname{finrank}_{\mathbb{R}} E$ *for μ-almost every x: the measure is exact-dimensional with dimension equal to the ambient dimension.*

Proof. Where the genuine limit of Theorem 11.14 exists, the lim sup defining $\bar{d}_\mu$ returns that limit. □

Theorem 11.16 (Local-to-Hausdorff dimension bridge)**.** *Let μ be a probability measure on a Borel second-countable metric space E, let $\alpha > 0$, and let s be a set of full μ-measure ($\mu(s^{\mathsf{c}}) = 0$) such that* for every $x \in s$ *the local-dimension quotient* $\log \mu(\bar{B}(x,r))/\log r$ *tends to α as $r \to 0^+$. Then* $\dim_H s = \alpha$*.*

Proof. Two mass-distribution arguments over a bare metric space. *Lower bound (Frostman):* for each $a < \alpha$, the pointwise limit yields on a positive-measure measurable piece of s a uniform upper ball bound $\mu(\bar{B}(x,r)) \le r^a$ at small radii; then $\mu\restriction_A \le \mu_H^a$ (any small-diameter set meeting A sits in a controlled ball), so $\mu_H^a(s) > 0$ and $a \le \dim_H s$; let $a \uparrow \alpha$. *Upper bound (Billingsley):* for $a > \alpha$, the limit produces at arbitrarily small radii the *lower* bound $\mu(\bar{B}(x,r)) \ge r^a$ (the positive limit forces positive ball masses); a Vitali enlargement of a disjoint subfamily of such balls covers s with $\sum(\operatorname{diam})^a \le (2\tau)^a \mu(E) < \infty$, so $\mu_H^a(s) < \infty$ and $\dim_H s \le a$; let $a \downarrow \alpha$. The pointwise (not merely a.e.) hypothesis is essential for the upper bound, since a μ-null subset can carry extra Hausdorff dimension. □

Theorem 11.17 (Hausdorff dimension of full-measure sets, a.c. case)**.** *Let μ be a probability measure on a finite-dimensional real inner-product space E, absolutely continuous with respect to a Haar measure. Then* every *set s of full μ-measure has Hausdorff dimension equal to the ambient dimension:* $\dim_H s = \operatorname{finrank}_{\mathbb{R}} E$*.*

Proof. The upper bound is monotonicity: $\dim_H s \le \dim_H E = \operatorname{finrank}_{\mathbb{R}} E$. The lower bound is the Frostman direction of the bridge, fed by the a.e. local-dimension limit of Theorem 11.14. □

11.5 The symbolic entropy–dimension identity

On the one-sided full shift $\Sigma = \alpha_0^{\mathbb{N}}$ over a finite alphabet, equipped with Mathlib's `PiNat` ultrametric $d(x, y) = (1/2)^{\mathrm{firstDiff}(x,y)}$, closed balls of radius $(1/2)^n$ *are* the n-step join atoms of the time-0 coordinate partition. The Shannon–McMillan–Breiman theorem therefore turns the ball-mass quotient into an entropy, and the bridge of Theorem 11.16 converts it into a Hausdorff dimension. The base $\log 2$ is fixed by the ultrametric, never a free parameter.

Theorem 11.18 (Entropy = Hausdorff dimension on the full shift)**.** *Let μ be a shift-invariant probability measure on the full shift Σ such that the left shift σ is ergodic for μ, and suppose the Kolmogorov–Sinai entropy of the coordinate partition is positive. Then there is a full-measure carrier set $s \subseteq \Sigma$ ($\mu(s^{\mathsf{c}}) = 0$) with*

$$\dim_H s \;=\; \frac{h_\mu(\sigma)}{\log 2},$$

where $h_\mu(\sigma)$ is the partition-independent *system entropy (*`ksEntropy`*, Definition 9.13).*

Proof. Atoms are cylinders are dyadic closed balls, so the unconditional pointwise Shannon–McMillan–Breiman theorem gives the dyadic mass quotient $\log \mu(\bar{B}(x, (1/2)^n))/\log((1/2)^n) \to h/\log 2$ μ-a.e., with h the coordinate-partition entropy (`ksEntropyPartition`, Definition 9.11). An ultrametric sandwich (balls are constant on each dyadic gap, and $\log r \to -\infty$) upgrades the dyadic limit to the continuum limit $r \to 0^+$. Feeding the conull carrier of that pointwise limit into Theorem 11.16 with $\alpha = h/\log 2 > 0$ gives $\dim_H s = h/\log 2$; finally the coordinate partition is a generator, so the Kolmogorov–Sinai generator theorem identifies h with the system entropy $h_\mu(\sigma)$. ☐

Theorem 11.19 (Unconditional Bernoulli witness)**.** *Let* $\mathrm{bern}\,\nu$ *be the Bernoulli (i.i.d. product) measure on the full shift with single-symbol law ν, and suppose ν charges two* distinct *symbols $i \neq j$ with positive mass. Then there is a* $\mathrm{bern}\,\nu$*-conull set s with*

$$\dim_H s \;=\; \frac{H(\nu)}{\log 2}, \qquad H(\nu) = \sum_a \mathrm{negMulLog}(\nu\{a\})$$

the single-symbol Shannon entropy (`Hnu` *in the sources). No ergodicity or positive-entropy hypothesis remains: both standing conditionals of Theorem 11.18 are discharged for the Bernoulli case.*

Proof. Ergodicity of the shift for $\mathrm{bern}\,\nu$ is Kolmogorov's 0–1 law applied to the tail-measurable invariant sets (`ergodic_shiftMap_bern`). The coordinate-partition entropy equals $H(\nu)$, which is strictly positive because the two charged symbols force $\nu\{i\} \in (0, 1)$ (`Hnu_pos`). Theorem 11.18 then applies, and the system entropy identity $h_{\mathrm{bern}\,\nu}(\sigma) = H(\nu)$ (`ksEntropy_bern_eq`, via the generator theorem) rewrites the dimension. ☐

11.6 The Bernoulli-suspension flow: a genuinely multifractal witness

The remaining question is non-vacuity of the flow-level formalism: is there an ergodic measure-preserving *flow* of positive entropy whose Rényi spectrum genuinely depends on q? The witness is the constant-roof ($\tau \equiv 1$) suspension of the *two-sided* Bernoulli shift $T =$ `biShiftEquiv` over the i.i.d. product measure $\mathrm{bernZ}\,\nu$ on $\alpha_0^{\mathbb{Z}}$, with a *biased* two-symbol law ν.

Definition 11.20 (The constant-roof Bernoulli suspension flow)**.** The *Bernoulli suspension flow* is the time-translation flow on the suspension (mapping-torus) of the two-sided Bernoulli shift T over $\mathrm{bernZ}\,\nu$ with constant roof $\tau \equiv 1$: points are orbits $[x, s]$ of the identification $(x, s) \sim (Tx, s - 1)$, the flow is $\zeta_t[x, s] = [x, s + t]$, and it preserves the normalized suspension measure $\hat{\mu}$ (the product of $\mathrm{bernZ}\,\nu$ with Lebesgue on the fibres; the constant roof makes the normalizing constant 1). It is a `MeasurePreservingFlow` of $\hat{\mu}$ (Definition 8.1).

Theorem 11.21 (Ergodicity of the suspension flow)**.** *Assume the base shift T is ergodic for* $\mathrm{bernZ}\,\nu$*. Then every measurable set A invariant under* all *time-t maps of the suspension flow ($\zeta_t^{-1}(A) = A$ for every $t \in \mathbb{R}$) is null or conull: $\hat{\mu}(A) = 0$ or $\hat{\mu}(A) = 1$. The base hypothesis is discharged unconditionally by the two-sided Bernoulli ergodicity theorem (*`ergodic_biShiftEquiv_bernZ`*); the companion* `ergodic_bernSuspensionFlow_uncond` *records the unconditional statement. By contrast, the* time-1 map alone is never ergodic *(*`not_ergodic_bernSuspensionFlow_one`*): the saturated section set $\{[x, s] : \mathrm{fract}\, s < 1/2\}$ is invariant of mass $1/2$ — the constant-roof special-flow dichotomy of Cornfeld–Fomin–Sinai.*

Proof. Lift A through the quotient map $\pi(x, s) = [x, s]$. Invariance under all vertical translations shows membership of $[x, s]$ depends only on the base point, so the lift is a cylinder $B \times \mathbb{R}$ with $B = \{x : [x, 0] \in A\}$. The identification generator $(x, s) \mapsto (Tx, s - 1)$ fixes π, so B is shift-invariant; it is measurable, and the constant-roof box computation gives $\hat{\mu}(A) = \mathrm{bernZ}\,\nu(B)$. Base ergodicity's zero–one law finishes. □

Theorem 11.22 (Entropy of the suspension flow)**.** *The Kolmogorov–Sinai entropy (Definition 9.13) of the* time-1 map *of the Bernoulli suspension flow equals the single-symbol Shannon entropy:*

$$h_{\hat{\mu}}(\zeta_1) \;=\; H(\nu) \qquad (\textit{as an extended real}).$$

In particular the flow's metric entropy (defined as the entropy of its time-1 map) is $H(\nu)$, strictly positive for a genuinely biased ν.

Proof. The fundamental-domain equivalence onto $\alpha_0^{\mathbb{Z}} \times [0, 1)$ conjugates ζ_1 to the *frozen* product $T \times \mathrm{id}$ and carries $\hat{\mu}$ to $\mathrm{bernZ}\,\nu \otimes \mathrm{Leb}\restriction_{[0,1)}$. Conjugacy invariance of h, the frozen-factor product identity $h(T \times \mathrm{id}) = h(T)$, and the two-sided Bernoulli system-entropy identity $h(T) = H(\nu)$ (generator theorem on the two-sided coordinate partition) chain together. □

Definition 11.23 (The witness partition)**.** The *witness partition* of the suspension measure $\hat{\mu}$ is the base time-0 coordinate partition of $\mathrm{bernZ}\,\nu$ pulled back along the base projection (factor map) $\pi : [x, s] \mapsto T^{\lfloor s \rfloor} x$, which is measure-preserving onto $\mathrm{bernZ}\,\nu$. Its cells are indexed by $\mathrm{Fin}(\mathrm{card}\,\alpha_0)$; the crux mass identity is that pulling back does not change cell masses, so the j-th cell carries the single-symbol mass $\nu\{a_j\}$ of the corresponding symbol.

Theorem 11.24 (Heterogeneity of the witness)**.** *If ν charges two distinct symbols $i \neq j$ with* different *masses ($\nu\{i\} \neq \nu\{j\}$), then the witness partition is* heterogeneous*: two of its cells carry distinct $\hat{\mu}$-mass (*`IsHeterogeneous`*, the honest non-uniformity predicate whose negation is exactly the equal-measure hypothesis of the monofractal degeneracy $D_q \equiv \log N/(-\log \varepsilon)$).*

Proof. By the mass identity, the cells indexed by i and j carry masses $\nu\{i\}$ and $\nu\{j\}$, which differ by hypothesis (the `toReal` coercion is injective on finite masses). □

Theorem 11.25 (q-dependence of the flow's Rényi spectrum)**.** *Let α_0 consist of exactly two symbols $i \neq j$, let ν charge both with positive but* different *masses ($\nu\{i\} \neq \nu\{j\}$ after* `toReal`*),* *and let* $0 < \varepsilon < 1$*. Then the Rényi dimension of the suspension flow's invariant measure on the witness partition takes different values at two exponents:*

$$\exists\, q_1, q_2, \quad D_{q_1}(\varphi, P, \varepsilon) \;\neq\; D_{q_2}(\varphi, P, \varepsilon).$$

The exhibited exponents are the explicit $q_1 = 0$, $q_2 = 1$ *(`renyiDimFlow_bernSuspension_zero_ne_one`): concretely* $D_0 = \log 2/(-\log \varepsilon)$ *(both cells occupied) while* $D_1 = H(\nu)/(-\log \varepsilon)$ *(the information dimension), and these differ precisely because the bias forces the strict inequality* $H(\nu) < \log 2$*. The witness is therefore non-vacuous: the q-dependence is driven by the genuine bias of* ν*, not satisfied trivially.*

Proof. A transfer argument. The flow witness's cell masses agree, up to the $\alpha_0 \simeq \mathrm{Fin}(\mathrm{card}\, \alpha_0)$ reindex, with those of the *one-sided* base coordinate partition under $\mathrm{bern}\, \nu$; since the Rényi dimension depends only on the cell-mass family, the flow spectrum equals the base spectrum at every q (`renyiDimFlow_bernSuspension_eq_base`). On the base, D_0 is the box-counting value $\log 2/(-\log \varepsilon)$ (two occupied cells) and D_1 is the information dimension $H(\nu)/(-\log \varepsilon)$ (Theorem 11.10); the strict two-point entropy bound $H(\nu) < \log 2$ for a biased law separates them. □

Chapter 12

Smooth maps and worked examples

The multiplicative ergodic theorem of Chapter 5 is a statement about abstract measurable matrix cocycles. This chapter connects it to the setting it was invented for — *smooth dynamics* — and then instantiates the whole chain on classical worked examples.

The bridge is the *derivative (tangent) cocycle*: for a differentiable self-map T of $E = \mathbb{R}^d$ (formalized as EuclideanSpace $\mathbb{R}\,(\mathrm{Fin}\,d)$), the chain rule makes the family of Jacobian matrices $x \mapsto D_xT$ a linear cocycle over T, so the Oseledets theorem applies verbatim and produces the Lyapunov exponents of the smooth system. On top of the bridge we prove the expanding-map results: every Lyapunov exponent of a uniformly expanding map is at least $\log K > 0$, and consequently the sum of the *positive* exponents $\sum \lambda^+$ (the Pesin / Margulis–Ruelle right-hand side) collapses onto the full exponent sum and hence, by the trace–determinant identity, onto $\int \log |\det D_xT| \; d\mu$ (the Rokhlin right-hand side). This is the honest, *foliation-free* expanding-case identity of the two right-hand sides — no stable foliation, no SRB-density machinery, and *not* a claim of the full Pesin entropy formula. The entropy-side companion, Rokhlin's formula $h_\mu(T, \xi) = \int \log |\det D_xT| \; d\mu$, is proved via a conditional-expectation / change-of-variables argument (Coudène), and the two are chained into an expanding-map Pesin formula — a correct implication whose hypothesis bundle is disclosed to be vacuous on the non-compact space $\mathbb{R}^d$, the genuinely instantiated equality living on the compact circle.

The worked examples then exercise every layer. The *doubling map* $y \mapsto 2y$ on the unit circle has top exponent $\log 2$, positive-exponent sum $\log 2$, a per-partition Margulis–Ruelle bound at exactly that rate, and — the highlight — the genuinely instantiated Rokhlin *equality* $h(\alpha, T) = \int \log |\det DT| \; d\mu = \log 2$ for the binary partition. The *Arnold cat map*, the hyperbolic automorphism of $\mathbb{T}^2$ induced by the matrix

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

is formalized as genuine toral dynamics: it is measure-preserving and ergodic for Haar measure (a Fourier-analytic proof on the multivariate character basis), its Lyapunov spectrum is $\log((3 \pm \sqrt{5})/2)$, and the derivative cocycle of its linear lift to the universal cover has strictly positive top exponent — the formalized hyperbolicity of the cat map. Everything in this chapter is sorry-free.

12.1 The derivative cocycle of a smooth self-map

Definition 12.1 (Derivative cocycle generator)**.** For a self-map T of $E = \mathrm{EuclideanSpace}\ \mathbb{R}\,(\mathrm{Fin}\, d)$, the *derivative cocycle generator* is the matrix-valued map

$$x \;\longmapsto\; (\texttt{toEuclideanCLM})^{-1}(D_xT) \;\in\; \mathrm{Matrix}_d(\mathbb{R}),$$

the matrix representing the Fréchet derivative $D_xT = \mathrm{fderiv}\ \mathbb{R}\, T\, x$, transported along the star-algebra equivalence between $d \times d$ real matrices and continuous linear endomorphisms of E. Since that equivalence is an L^2-operator-norm isometry, the generator has the same norm as D_xT. Feeding it to the iterated cocycle construction of Theorem 2.2 yields the *tangent cocycle* of the smooth system.

Theorem 12.2 (Chain-rule cocycle identity)**.** *For a differentiable T, the n-th cocycle iterate of the derivative generator represents the derivative of the n-th iterate of the map:*

$$\textit{toEuclideanCLM}(\mathrm{cocycle}(\mathrm{derivativeCocycle}\, T)\, T\, n\, x) \;=\; D_x(T^{[n]}).$$

Proof. Induction on n. The base case is $D(\mathrm{id}) = \mathrm{id}$. For the step, peel the innermost factor T simultaneously from the cocycle recursion $A^{(n+1)}(x) = A^{(n)}(Tx) \cdot A(x)$ and from the iterate $T^{[n+1]} = T^{[n]} \circ T$; the chain rule $D_x(T^{[n]} \circ T) = D_{Tx}(T^{[n]}) \circ D_xT$ (differentiability of T gives differentiability of all iterates) matches the two factorizations, and the star-algebra equivalence turns the matrix product into the composition. □

Theorem 12.3 (Oseledets theorem for the derivative cocycle)**.** *Let μ be a probability measure on $E = \mathrm{EuclideanSpace}\ \mathbb{R}\,(\mathrm{Fin}\, d)$ and let T be ergodic for μ, differentiable, with everywhere nonvanishing Jacobian determinant* $\det(\mathrm{derivativeCocycle}\ T\, x) \neq 0$ *and with the log-integrability* $\log^+ \|D_xT\|,\ \log^+\|(D_xT)^{-1}\| \in L^1(\mu)$ *(stated for the matrix generator and its matrix inverse; by the isometry these are exactly the* fderiv *conditions). Then:*

1. *for every n and x, the cocycle iterate represents $D_x(T^{[n]})$ — the cocycle is the genuine tangent cocycle; and*
2. *there exist $k \le d$, strictly decreasing exponents $\lambda_1 > \cdots > \lambda_k$, and a measurable family of subspaces $E = V_0(x) \supsetneq V_1(x) \supsetneq \cdots \supsetneq V_k(x) = 0$, a.e. equivariant under D_xT, such that for every $v \in V_{i-1}(x) \setminus V_i(x)$,*

$$\frac{1}{n}\log\|D_x(T^{[n]})v\| \;\longrightarrow\; \lambda_i.$$

Proof. The first conjunct is Theorem 12.2. The second is the one-sided Oseledets theorem (Theorem 5.30) applied to the generator $A = \mathrm{derivativeCocycle}\ T$, whose measurability is proved entrywise: each entry is a (continuous) coordinate projection of $x \mapsto (D_xT)e_j$, measurable by Mathlib's measurability of the Fréchet derivative in the base point. □

12.2 Uniformly expanding maps: the foliation-free right-hand-side identity

A differentiable map T is *uniformly expanding* with constant $K > 1$ when $K\,\|v\| \le \|D_xT\, v\|$ for every base point x and tangent vector v. The results of this section are stated exactly as the module discloses them: they establish the *all-positive-spectrum collapse* of $\sum \lambda^+$ onto $\int \log|\det D_xT|\, d\mu$ — the honest, foliation-free expanding-case identity of the Pesin and Rokhlin right-hand sides — and no more.

Lemma 12.4 (Compounded expansion bound)**.** *If T is differentiable and uniformly expanding with constant $K > 1$, then the derivative of the n-th iterate stretches every vector by at least K^n:*

$$K^n \|v\| \le \left\|D_x(T^{[n]})v\right\| \qquad \text{for all } x, v, n.$$

Proof. Induction on n, using the chain rule $D_x(T^{[n+1]}) = D_{Tx}(T^{[n]}) \circ D_xT$ to peel off one expanding factor at each step: $K^{n+1}\|v\| = K^n(K\|v\|) \le K^n \|D_xT\,v\| \le \left\|D_{Tx}(T^{[n]})(D_xT\,v)\right\|$. □

Lemma 12.5 (Every singular value is at least K^n)**.** *Under the same hypotheses, every singular value of the cocycle iterate $A^{(n)}(x) = \mathrm{cocycle}(\mathrm{derivativeCocycle}\,T)\,T\,n\,x$ is at least K^n: $K^n \le \sigma_i(A^{(n)}(x))$ for every $i < d$.*

Proof. By Theorem 12.2 the iterate acts as $D_x(T^{[n]})$, so Theorem 12.4 gives the uniform stretching $K^n\|v\| \le \|fv\|$ for $f = A^{(n)}(x)$. A uniform lower stretching bound passes to every singular value: evaluate f on the unit-norm right singular vector u_i (an eigenvector of f^*f), where $\sigma_i(f) = \|fu_i\| \ge K^n\|u_i\| = K^n$. □

Theorem 12.6 (Every exponent is at least $\log K$)**.** *For an ergodic, log-integrable, differentiable uniformly expanding map with constant $K > 1$ (and $d \ge 1$), each Lyapunov exponent of the tangent cocycle satisfies $\log K \le \lambda_i$.*

Proof. Pick a base point where the per-index singular-value limit $\frac{1}{n}\log\sigma_i(A^{(n)}(x)) \to \lambda_i$ of Theorem 6.12 holds. By Theorem 12.5 each pre-limit term is at least $\frac{1}{n}\log K^n = \log K$, and the limit inherits the lower bound. □

Corollary 12.7 (Positivity of the whole spectrum)**.** *Every Lyapunov exponent of a uniformly expanding map is strictly positive: $0 < \log K \le \lambda_i$.*

Proof. $K > 1$ gives $\log K > 0$; chain with Theorem 12.6. □

Proposition 12.8 (All-positive-spectrum collapse)**.** *For a uniformly expanding map the positive-part exponent sum is the full exponent sum: $\sum \lambda^+ = \sum_i \lambda_i$ (with multiplicity).*

Proof. By Theorem 12.7 the filter $\{i \mid 0 < \lambda_i\}$ is all of the index set, so the two finite sums have identical terms. □

Theorem 12.9 (The expanding-case right-hand-side identity)**.** *For an ergodic, log-integrable, differentiable uniformly expanding self-map T, the sum of the strictly positive Lyapunov exponents — the Pesin / Margulis–Ruelle right-hand side — equals the integrated volume distortion — the Rokhlin right-hand side:*

$$\sum \lambda^+ = \int \log\left|\det(\mathrm{derivativeCocycle}\,T\,x)\right|\, d\mu(x).$$

This is the honest foliation-free instance of the identity between the two right-hand sides; it is not *a claim of the full Pesin entropy formula (no entropy appears in the statement).*

Proof. Since all exponents are positive, $\sum \lambda^+ = \sum \lambda$ (Theorem 12.8), and the trace–determinant identity (Theorem 6.20) rewrites the full sum as $\int \log|\det A|\, d\mu$. □

12.3 Rokhlin's entropy formula for an expanding map

The entropy-side companion identifies the Kolmogorov–Sinai entropy itself with the determinant integral, for an absolutely continuous invariant measure. The proof is Coudène's conditional-expectation argument: the conditional entropy of a partition given the σ-algebra $T^{-1}\mathcal{A}$ is computed by a per-branch change of variables.

Definition 12.10 (Injectivity partition)**.** For a self-map T of EuclideanSpace $\mathbb{R}\,(\mathrm{Fin}\,d)$ and a finite measurable partition ξ, the predicate `IsInjectivityPartition` $\mu\,T\,\xi$ packages the two hypotheses the conditional-expectation proof needs: T is injective on each cell of ξ, and each cell is measurable. (These are literally the hypotheses of Mathlib's change-of-variables lemma; no Markov condition and no generating condition are baked in.)

Theorem 12.11 (Conditional entropy equals the Jacobian integral)**.** *Let μ be an invariant probability measure with $\mu \ll$ volume, let T be differentiable with everywhere nonvanishing Jacobian, let ξ be an injectivity partition, and assume $\log\rho \in L^1(\mu)$ for the density $\rho = d\mu/d\,\mathrm{vol}$ and $\log|\det DT| \in L^1(\mu)$. Then the conditional Shannon entropy of ξ given the pulled-back σ-algebra $T^{-1}\mathcal{A}$ is*

$$H(\xi \,|\, T^{-1}\mathcal{A}) \;=\; \int \log|\det D_xT|\; d\mu(x),$$

independently of the partition *(no generating hypothesis).*

Proof. Per cell ξ_i, the change-of-variables crux (`measure_cell_inter_preimage_eq_setLIntegral_transfer`) recovers $\mu(\xi_i \cap T^{-1}B)$ as the integral over $T(\xi_i)\cap B$ of the per-branch transfer density $\rho(g_i^{-1}y)/\,|\det DT|_{g_i^{-1}y}$, where g_i^{-1} is the branch inverse of T on ξ_i. This identifies the regular-conditional kernel mass of each cell with the branch-weight candidate; the entropy integrand $\sum_i$ negMulLog of these masses pulls out, per cell, to $\int_{\xi_i}(\log|\det DT| + \log\rho\circ T - \log\rho)d\mu$, the cells partition the space, and the density bracket $\log\rho\circ T - \log\rho$ telescopes to 0 by T-invariance of μ, leaving the Jacobian integral. □

The assembly of Theorem 12.11 into the per-partition Rokhlin formula $h_\mu(T,\xi) = \int \log|\det D_xT|\; d\mu$ for a one-sided *generating* injectivity partition — via the sharp-rate identity expressing $h_\mu(T,\xi)$ as the conditional entropy of ξ against its own strict future, and the σ-algebra glue identifying that future with $T^{-1}\mathcal{A}$ — is the node Theorem 9.24 of the classical-entropy chapter. Here we chain it with the expanding-case right-hand-side identity into the Pesin formula.

Theorem 12.12 (Expanding-map Pesin formula (vacuous on $\mathbb{R}^d$, disclosed))**.** *For an ergodic, absolutely continuous ($\mu \ll$ volume), differentiable, uniformly expanding self-map of* EuclideanSpace $\mathbb{R}\,(\mathrm{Fin}\,d)$ *with everywhere-nonsingular derivative, log-integrable derivative data, a one-sided generating injectivity partition, and integrable $\log\rho$ and $\log|\det DT|$:*

$$h_\mu(T) \;=\; \sum \lambda^+.$$

Honest disclosure: *this is a correct implication whose hypothesis bundle has* no model *on the non-compact space $\mathbb{R}^d$ — a globally uniformly expanding map of $\mathbb{R}^d$ admits no ergodic absolutely continuous invariant probability measure (uniform expansion forces mass to escape to infinity; for $T = c\cdot\mathrm{id}$ the nested preimages of a ball force an atom at the fixed point). The theorem is therefore vacuously true as stated on $\mathbb{R}^d$, and is disclosed as such; the instantiated Pesin/Rokhlin equality lives on the compact circle (Theorem 12.18).*

Proof. Compose three theorems: the Kolmogorov–Sinai generator theorem $h_\mu(T) = h_\mu(T, \xi)$ for the generating ξ, Rokhlin's per-partition formula (Theorem 9.24), and the expanding-case right-hand-side identity (Theorem 12.9), aligning the two determinant hypotheses along the fderiv-to-matrix bridge. □

12.4 The doubling map

The phase space is the unit circle $\mathbb{T}$ = UnitAddCircle with its Haar probability measure (Mathlib's default volume), and the map is doublingMap : $y \mapsto 2{\cdot}y$ (), ergodic by Mathlib's `AddCircle.ergodic_nsmul` (). Its derivative is the constant 1×1 matrix (2), realized as a constant cocycle.

Theorem 12.13 (Doubling map: top exponent $\log 2$)**.** *The constant cocycle with generator $M = (2)$ over the ergodic doubling map has top Lyapunov exponent $\log 2$.*

Proof. For a constant cocycle with symmetric invertible generator M, the sorted Lyapunov spectrum is log of the sorted eigenvalues of $|M|$ (`exponents_const`); for positive semidefinite M the functional calculus collapses $|M| = M$. The single eigenvalue of (2) is its trace 2, so the unique exponent is $\log 2$. □

Corollary 12.14 (Doubling map: positive-exponent sum $\log 2$)**.** *The sum of the strictly positive Lyapunov exponents of the doubling-map cocycle is $\log 2$, with every spectrum hypothesis (invertibility, measurability, both log-integrability conditions) discharged unconditionally from the constant-cocycle API.*

Proof. The spectrum consists of the single exponent $\log 2 > 0$, so the positive-part filter is the full (one-element) index set and the sum is that one term. □

Theorem 12.15 (Per-partition Ruelle bound for the doubling map)**.** *For any finite measurable partition P of the circle whose n-fold refinement $\bigvee_{k=0}^{n-1} T^{-k}P$ under the doubling map eventually has at most $C \cdot e^{n \log 2}$ non-empty atoms ($C \geq 1$),*

$$h(P, T) \;\leq\; \sum \lambda^+ \;=\; \log 2.$$

This is the per-partition *bound $h(\alpha, T) \leq \sum \lambda^+$, not the system Margulis–Ruelle inequality; the right-hand side is the computed Lyapunov datum (Theorem 12.14), not an abstract constant. The atom-count hypothesis is automatic for the binary partition, where the bound is attained (Theorem 12.17).*

Proof. Specialize the abstract atom-count-growth entropy bound (the arithmetic backbone of the per-partition Ruelle inequality, cf. Theorem 9.23) to the doubling map at the exponential rate $R = \log 2$, then identify the rate with the positive-exponent sum via Theorem 12.14. □

Proposition 12.16 (Entropy of a uniform-join system)**.** *Let T preserve a probability measure and let P be a finite partition indexed by a type of cardinality b such that for* every *n, every cell of the n-fold join $\bigvee_{k=0}^{n-1} T^{-k}P$ has measure exactly b^{-n}. Then $h(P, T) = \log b$.*

Proof. The n-fold join is indexed by the b^n formal cell-tuples, each of measure b^{-n}, so its Shannon entropy is a sum of b^n equal terms negMulLog$(b^{-n}) = b^{-n}\, n \log b$, totalling $n \log b$. The averaged sequence H_n/n is thus eventually the constant $\log b$, and it converges to $h(P, T)$; the two limits agree. □

Theorem 12.17 (Rokhlin equality, entropy side: $h(\alpha, T) = \log 2$)**.** *For the binary partition* $\alpha = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$ *of the circle (), the partition-relative Kolmogorov–Sinai entropy under the doubling map is exactly* $\log 2$*.*

Proof. The dynamical crux (`volume_binJoinCell`) shows every cell of the n-fold join is a dyadic arc of measure exactly 2^{-n}: the doubling map restricted to a half-arc is the affine two-fold magnification onto the whole circle, so $\mathrm{vol}(\alpha_i \cap T^{-1}B) = \mathrm{vol}(B)/2$, and induction on n halves the measure at each refinement step. Feeding this uniform-join datum into Theorem 12.16 with $b = 2$ gives $h(\alpha, T) = \log 2$. □

Theorem 12.18 (Rokhlin equality on the doubling map)**.** *For the doubling map and the binary partition,*

$$h(\alpha, T) \;=\; \int_{\mathbb{T}} \log\,|\det DT|\;d\mu \;=\; \log 2,$$

the genuinely instantiated *Pesin/Rokhlin equality on a real expanding system (in contrast to the vacuous-on-*$\mathbb{R}^d$ *assembly Theorem 12.12). The integrand is the honest log-Jacobian: the generator* (2) *is proved to be the Fréchet derivative of the doubling map's linear lift* $x \mapsto 2x$ *to the universal cover (), and the covering projection* $\mathbb{R} \to \mathbb{T}$ *intertwines that lift with the doubling map () — so* (2) *genuinely is* DT*, not an arbitrary constant of the right determinant.*

Proof. The entropy side is Theorem 12.17. For the integral side, $\det(2) = 2$, so the integrand is the constant $\log 2$, which integrates against the probability measure to $\log 2$. Both sides equal $\log 2$. □

12.5 The Arnold cat map

The Arnold cat map is the automorphism of the 2-torus $\mathbb{T}^2 = \mathrm{UnitAddTorus}(\mathrm{Fin}\,2)$ induced by the unimodular hyperbolic matrix $M \in \mathrm{SL}_2(\mathbb{Z})$ displayed in the chapter introduction. Its eigenvalues are $\lambda_\pm = (3 \pm \sqrt{5})/2$, with $\lambda_+ > 1 > \lambda_- > 0$ and $\lambda_+\lambda_- = 1$. We first read off the Lyapunov spectrum of the *matrix* (as a constant cocycle), then formalize the genuine toral dynamics, and finally combine them.

Theorem 12.19 (Cat-map matrix: closed-form Lyapunov spectrum)**.** *Realized as a constant cocycle with generator the cat-map matrix* M *over an ergodic base (here the doubling map — the spectrum depends only on* M*), the two Lyapunov exponents are*

$$\lambda_1 = \log\frac{3+\sqrt{5}}{2}, \qquad \lambda_2 = \log\frac{3-\sqrt{5}}{2}.$$

Honesty caveat (as in the Lean docstring): *the cocycle here is the constant matrix, not the derivative cocycle of the genuine toral automorphism; the genuine dynamics is treated below.*

Proof. M is symmetric positive definite with trace 3 and determinant 1, so its sorted eigenvalues $a \geq b$ satisfy $a + b = 3$, $ab = 1$, whence $(a - b)^2 = 9 - 4 = 5$ and $a, b = (3 \pm \sqrt{5})/2$. The constant-cocycle spectrum theorem (`exponents_const`) evaluates the sorted exponents to log of the sorted eigenvalues of $|M| = M$. □

Corollary 12.20 (The cat-map exponents sum to zero)**.** $\lambda_1 + \lambda_2 = 0$*: the cocycle is conservative (*$\det M = 1$*).*

Proof. $\log\lambda_+ + \log\lambda_- = \log(\lambda_+\lambda_-) = \log 1 = 0$, computing $\lambda_+\lambda_- = (9 - (\sqrt{5})^2)/4 = 1$. □

Definition 12.21 (The cat-map toral automorphism)**.** The map $\mathrm{catTorus} : \mathbb{T}^2 \to \mathbb{T}^2$, $(\mathrm{catTorus}\, y)_i = \sum_j M_{ij} \cdot y_j$, with the integer cat-map matrix M acting by integer scalar multiplication on each circle coordinate. Since $\det M = 1$, the integer matrix M^{-1} (with rows $(1, -1)$ and $(-1, 2)$) induces a two-sided inverse, so catTorus is a continuous additive automorphism of the compact group $\mathbb{T}^2$. Throughout this section $\mathbb{T}^2$ carries the product Haar *probability* measure (the normalization for which Mathlib's multivariate Fourier basis is stated).

Proposition 12.22 (Measure preservation)**.** catTorus *preserves the Haar probability measure on* $\mathbb{T}^2$.

Proof. catTorus is a continuous surjective additive homomorphism of a compact group; the pushforward of Haar measure under such a map is again a translation-invariant probability measure, hence equals Haar (Mathlib's `AddMonoidHom.measurePreserving`, with equal total mass 1). □

Theorem 12.23 (Ergodicity of the Arnold cat map)**.** catTorus *is ergodic for the Haar probability measure on* $\mathbb{T}^2$.

Proof. The classical Fourier / character argument. The Koopman operator permutes the multivariate characters: `mFourier` $n \circ \mathrm{catTorus} =$ `mFourier`$(M^\top n)$, and since M is symmetric the index action is $n \mapsto Mn$. For a measurable invariant set s, the indicator $\mathbf{1}_s \in L^2$ has Fourier coefficients constant along each index orbit $p \mapsto M^p n$. The hyperbolicity input is that this orbit is *infinite* for every $n \neq 0$ (`orbit_infinite`: pairing with the two eigen-covectors of M, a period would force $\varphi(n)\lambda_+^k = \varphi(n)$ and $\psi(n)\lambda_+^{-k} = \psi(n)$ with $\lambda_+^k \neq 1$, so $n = 0$). A square-summable sequence constant on an infinite set vanishes there, so all nonzero-index coefficients of $\mathbf{1}_s$ are 0; the Fourier series collapses to the constant term, $\mathbf{1}_s$ is a.e. constant, and s is a.e. empty or full. □

Theorem 12.24 (Cat-map spectrum over the genuine ergodic base)**.** *Realized as a constant cocycle with generator M over the* genuine *ergodic Arnold cat map* catTorus*, the two Lyapunov exponents are* $\log((3+\sqrt{5})/2)$ *and* $\log((3-\sqrt{5})/2)$ *— the same spectrum as Theorem 12.19, now over the hyperbolic toral automorphism itself rather than a surrogate base.*

Proof. The constant-cocycle spectrum theorem applies over any ergodic base; instantiate it over catTorus (Theorem 12.23) and evaluate the sorted eigenvalues of $|M| = M$ by the closed form computed for Theorem 12.19. □

Theorem 12.25 (The cat map's derivative cocycle has positive top exponent)**.** *Let* $\mathrm{catLift} : \mathbb{R}^2 \to \mathbb{R}^2$ *() be the linear lift of the cat map to the universal cover — the continuous linear map with matrix M, which genuinely lifts* catTorus*: the covering projection* $\mathbb{R}^2 \to \mathbb{T}^2$ *satisfies* $\pi \circ \mathrm{catLift} = \mathrm{catTorus} \circ \pi$ *(). Its derivative cocycle in the sense of Theorem 12.1 is the constant matrix M at every point (), and the top Lyapunov exponent of this* genuine derivative cocycle*, over the genuine ergodic base* catTorus*, is*

$$0 \;<\; \log \frac{3+\sqrt{5}}{2},$$

the formalized hyperbolicity of the Arnold cat map. (Reading the derivative on the torus manifold itself via `mfderiv` is a documented gap: Mathlib has no manifold-derivative API for `AddCircle` endomorphisms; the lift and the map share the same derivative everywhere because the covering projection is a local diffeomorphism with identity derivative.)

Proof. The Fréchet derivative of a continuous linear map is the map itself, so derivativeCocycle catLift is the constant M; transporting along this equality, the top exponent is the constant-cocycle top exponent $\log((3+\sqrt{5})/2)$ of Theorem 12.24, positive because $(3+\sqrt{5})/2 > 1$. ☐

Theorem 12.26 (Per-partition Ruelle bound for the cat map)**.** *For any finite measurable partition P of $\mathbb{T}^2$ whose n-fold refinement under* catTorus *eventually has at most $C \cdot e^{n \log \lambda_+}$ non-empty atoms ($C \geq 1$, $\lambda_+ = (3+\sqrt{5})/2$),*

$$h(P, \mathrm{catTorus}) \;\leq\; \log \lambda_+.$$

As in the doubling-map case this is the per-partition *bound, not the system inequality; the rate $\log \lambda_+$ is the genuine top Lyapunov exponent of the cat map (Theorem 12.25), and the atom-count growth is the honest named geometric input. The sharp system-level equality $h_\mu = \log \lambda_+$ is a documented wall (it needs an Adler–Weiss Markov generating partition for the upper bound and Pesin/Ledrappier–Young machinery for the lower bound).*

Proof. A thin specialization of the abstract atom-count-growth entropy bound (the arithmetic backbone of the per-partition Ruelle inequality, cf. Theorem 9.23) over the measure-preserving base catTorus (Theorem 12.22) at the rate $R = \log((3+\sqrt{5})/2)$. ☐

Chapter 13

Quantum relative entropy and its monotonicity

The multiplicative ergodic machinery of the preceding chapters is built on the continuous functional calculus of self-adjoint matrices. That same finite-dimensional matrix and CFC infrastructure supports a second, logically independent development: a *finite-dimensional quantum-information layer.* This chapter documents its foundational half — the entropies of a finite quantum system and the master inequality controlling how they behave under quantum channels.

The objects are *density matrices*: positive semidefinite complex matrices of unit trace, the finite-dimensional states of a quantum system. Attached to a state ρ is its von Neumann entropy $S(\rho)$, and to a pair ρ, σ the Umegaki *relative entropy* $S(\rho\|\sigma) = \mathrm{Tr}(\rho(\log\rho - \log\sigma))$, the finite-dimensional distinguishability functional of quantum information theory. The central theorem, the *data-processing inequality*, states that no quantum channel can increase relative entropy: distinguishability can only degrade under physical processing. We reach it through Lieb's 1973 theorem on the joint convexity of relative entropy (Lieb 1973), following the modern route of Carlen's *Trace Inequalities and Quantum Entropy* via the Effros operator perspective, and close with the easy half of Petz's equality theorem (Petz 1986, 2003) — recovery implies saturation — whose hard converse (saturation implies the existence of a Petz recovery map) is treated separately.

Everything in this chapter is formalized sorry-free, on the same matrix/CFC foundations as the multiplicative ergodic theorem, and is verified by the guarded axiom audit to rest only on {`propext`, `Classical.choice`, `Quot.sound`}. Throughout, n is a finite index type, matrices are complex $n \times n$ acting on $\mathbb{C}^n$, and Tr is the matrix trace.

13.1 Density matrices and von Neumann entropy

Definition 13.1 (Density matrix)**.** For a finite index type n, a *density matrix* is a structure bundling a matrix $\rho \in \mathrm{Matrix}_n(\mathbb{C})$ together with proofs that ρ is positive semidefinite and has unit trace, $\mathrm{Tr}\,\rho = 1$. It is the finite-dimensional state of a quantum system on $\mathbb{C}^n$.

Lemma 13.2 (Eigenvalues form a probability vector)**.** *The real eigenvalues λ_i of a density matrix ρ are nonnegative, each at most 1, and sum to 1: $\sum_i \lambda_i = \mathrm{Tr}\,\rho = 1$.*

Proof. Positive semidefiniteness gives $\lambda_i \geq 0$. The Hermitian spectral theorem identifies $\mathrm{Tr}\,\rho$

with $\sum_i \lambda_i$ as a complex number; casting the hypothesis $\operatorname{Tr}\rho = 1$ back to $\mathbb{R}$ yields $\sum_i \lambda_i = 1$. Each $\lambda_i \le 1$ is then a single term bounded by the sum of the nonnegative terms. □

Definition 13.3 (Von Neumann entropy)**.** The *von Neumann entropy* of a density matrix ρ with eigenvalues λ_i is

$$S(\rho) \;=\; \sum_i \operatorname{negMulLog}(\lambda_i) \;=\; -\sum_i \lambda_i \log \lambda_i,$$

using Mathlib's $\operatorname{negMulLog}(x) = -x \log x$, so the convention $0 \log 0 = 0$ is built in.

Theorem 13.4 (Nonnegativity of entropy)**.** *$S(\rho) \ge 0$ for every density matrix ρ.*

Proof. Each eigenvalue satisfies $0 \le \lambda_i \le 1$, and on $[0,1]$ one has $\operatorname{negMulLog}(x) = -x\log x \ge 0$. The entropy is a sum of these nonnegative terms. □

We record the partial trace here, as it is the elementary operation underlying both the reduced states of a bipartite system and the data-processing inequality.

Definition 13.5 (Right partial trace)**.** For an operator M on a bipartite system $\mathbb{C}^{n_A} \otimes \mathbb{C}^{n_B}$, the *partial trace over the B factor* is

$$(\operatorname{Tr}_B M)_{i,i'} \;=\; \sum_j M_{(i,j),(i',j)},$$

an operator on $\mathbb{C}^{n_A}$. (The left partial trace Tr_A is defined symmetrically.)

Lemma 13.6 (Partial trace preserves the trace)**.** $\operatorname{Tr}(\operatorname{Tr}_B M) = \operatorname{Tr} M$.

Proof. Both sides expand to the full double sum $\sum_{i,j} M_{(i,j),(i,j)}$ over the product index set; a reindexing of the finite sum finishes. □

Lemma 13.7 (Partial trace is completely positive)**.** *If M is positive semidefinite then so is $\operatorname{Tr}_B M$.*

Proof. Write the partial trace in Kraus/compression form $\operatorname{Tr}_B M = \sum_j E_j^* M E_j$, the sum of the conjugations of M by the block-inclusion isometries $E_j : i \mapsto (i,j)$. Each compression $E_j^* M E_j$ is positive semidefinite, and a finite sum of positive semidefinite matrices is positive semidefinite. □

13.2 Umegaki relative entropy and Klein's inequality

Definition 13.8 (Umegaki relative entropy)**.** The *Umegaki relative entropy* of ρ with respect to σ is $S(\rho\|\sigma) = \operatorname{Tr}\rho(\log\rho - \log\sigma)$. It is defined in the concrete *spectral/overlap* form

$$S(\rho\|\sigma) \;=\; \sum_k p_k \log p_k \;-\; \sum_{k,m} |\langle e_k \mid f_m\rangle|^2 \, p_k \log q_m,$$

where (p_k, e_k) are the eigenvalues/eigenvectors of ρ and (q_m, f_m) those of σ. The definition is total: with the convention $\log 0 = 0$ the σ-singular columns contribute 0.

Theorem 13.9 (Trace form of relative entropy)**.** *For faithful (positive definite) σ, the spectral definition agrees with the textbook trace form*

$$S(\rho\|\sigma) \;=\; \Re\, \operatorname{Tr}(\rho\,(\log\rho - \log\sigma)),$$

where $\log\rho, \log\sigma$ are the Hermitian continuous functional calculus of the real logarithm.

Proof. Expand both eigendecompositions. The trace–spectral bridge $\Re\ \mathrm{Tr}(\rho\, f(\tau)) = \sum_{k,m} p_k\, |\langle e_k \mid g_m\rangle|^2\, f(q_m)$ (proved once, entrywise, via the spectral theorem and the identity $z\bar{z} = |z|^2$) identifies $\Re\ \mathrm{Tr}(\rho \log \rho)$ with $\sum_k p_k \log p_k$ (the overlap of ρ with itself is the identity) and $\Re\ \mathrm{Tr}(\rho \log \sigma)$ with the cross double sum. Distributing over the difference gives the claim. □

Theorem 13.10 (Klein / Gibbs nonnegativity)**.** *For faithful (positive definite) σ, $0 \le S(\rho\|\sigma)$.*

Proof. The overlap matrix $D_{km} = |\langle e_k \mid f_m\rangle|^2$ is doubly stochastic: writing $Q = \rho.\mathrm{eigVec}^*\, \sigma.\mathrm{eigVec}$, both $QQ^* = 1$ and $Q^*Q = 1$ (the eigenvector unitaries cancel), giving unit row and column sums. Faithfulness makes every $q_m > 0$, so the support side condition is vacuous. Klein's scalar inequality (Theorem 13.14) then yields $\sum_{k,m} D_{km}\, p_k \log q_m \le \sum_k p_k \log p_k$, i.e. $0 \le S(\rho\|\sigma)$. □

Theorem 13.11 (Vanishing on the diagonal)**.** *$S(\rho\|\rho) = 0$ for every density matrix ρ.*

Proof. With $\sigma = \rho$ the overlap matrix $Q = \rho.\mathrm{eigVec}^*\, \rho.\mathrm{eigVec} = 1$, so $D_{km} = \delta_{km}$. The cross double sum collapses to $\sum_k p_k \log p_k$, exactly cancelling the first term. □

Theorem 13.12 (Unitary invariance)**.** *For any unitary W, $S(W\rho W^* \,\|\, W\sigma W^*) = S(\rho\|\sigma)$, where $W\rho W^*$ denotes the density matrix obtained by conjugating ρ.*

Proof. Conjugation by W is a $*$-algebra automorphism, so it commutes with the functional calculus: $\log(W\rho W^*) = W(\log \rho)W^*$. Hence each trace $\mathrm{Tr}((W\rho W^*) \log(W\tau W^*)) = \mathrm{Tr}(W(\rho \log \tau)W^*) = \mathrm{Tr}(\rho \log \tau)$ by cyclicity of the trace and $W^*W = 1$. Applying this to the two terms of the trace form leaves $S(\rho\|\sigma)$ unchanged. □

Theorem 13.13 (Ancilla invariance)**.** *Tensoring both arguments with a common faithful ancilla α leaves the relative entropy unchanged: for faithful σ and faithful α, $S(\rho \otimes \alpha \,\|\, \sigma \otimes \alpha) = S(\rho\|\sigma)$.*

Proof. Relative entropy is additive over Kronecker products, $S(\rho \otimes \alpha \,\|\, \sigma \otimes \alpha) = S(\rho\|\sigma) + S(\alpha\|\alpha)$, and the self-term vanishes by Theorem 13.11. □

The nonnegativity above is the operator form of *Klein's inequality*; its scalar core, which we isolate next, is the combinatorial engine behind both nonnegativity and the subadditivity of the von Neumann entropy.

Theorem 13.14 (Scalar Klein / Peierls inequality)**.** *Let $D = (D_{km})$ be a doubly stochastic $K \times M$ matrix ($D_{km} \ge 0$, all row sums and all column sums equal to 1), let $p \ge 0$ on K and $s \ge 0$ on M have equal total mass $\sum_k p_k = \sum_m s_m$, and assume the support condition $s_m = 0 \Rightarrow D_{km}p_k = 0$. Writing $a_m = \sum_k D_{km}p_k$ for the column marginal,*

$$\sum_m a_m \log s_m \ \le\ \sum_k p_k \log p_k.$$

Proof. This is the finite scalar core of Klein's inequality (Carlen, *Trace Inequalities and Quantum Entropy*, Thm. 2.11; Peierls, Thm. 2.9). The termwise Peierls bound $p - s \le p \log p - p \log s$ (valid for $p \ge 0$, $s > 0$, from $\log(s/p) \le s/p - 1$) is scaled by $D_{km} \ge 0$ and summed over the double index. Double stochasticity makes the mass-balance term $\sum_{k,m} D_{km}(p_k - s_m)$ telescope to $\sum_k p_k - \sum_m s_m = 0$; the columns with $s_m = 0$ contribute nothing by the support hypothesis (there $a_m = 0$). Rearranging the summed inequality yields the claim. □

Theorem 13.15 (Subadditivity of the von Neumann entropy)**.** *For a bipartite density matrix ρ on $n_A \otimes n_B$ with reduced density matrices $\rho_A = \mathrm{Tr}_B\, \rho$ and $\rho_B = \mathrm{Tr}_A\, \rho$, the von Neumann entropy is subadditive:*

$$S(\rho) \ \le\ S(\rho_A) + S(\rho_B).$$

Proof. An elementary route through the scalar Klein inequality, with no matrix logarithm (Carlen, *Trace Inequalities and Quantum Entropy*, Thm. 2.11; Nielsen–Chuang §11.3). Diagonalize $\rho = G\,\mathrm{diag}(p)\,G^*$, $\rho_A = U\,\mathrm{diag}(\lambda)\,U^*$, $\rho_B = V\,\mathrm{diag}(\mu)\,V^*$ and set $Q = G^*(U \otimes V)$, a unitary. Then $D_{km} = |Q_{km}|^2$ is doubly stochastic, and the conjugation–partial-trace identity $\mathrm{Tr}_B((U \otimes V)^*\rho(U \otimes V)) = U^*(\mathrm{Tr}_B\,\rho)U$ (and its left analogue) identifies the marginals of D with λ and μ. Feeding D, the eigenvalue vector p, and the product vector $s_{(i,j)} = \lambda_i\mu_j$ into Theorem 13.14 gives $\sum_i \eta(\lambda_i) + \sum_j \eta(\mu_j) \geq \sum_k \eta(p_k)$ (with $\eta(t) = -t\log t$), which is the assertion after negation. □

13.3 Lieb's joint-convexity theorem

The deep content of quantum relative entropy is Lieb's 1973 theorem: the map $(\rho, \sigma) \mapsto S(\rho\|\sigma)$ is *jointly convex*. We prove it through the Effros operator perspective, whose joint convexity follows in turn from operator convexity of $-\log$ and the Hansen–Pedersen–Jensen operator-Jensen inequality, following Carlen's *Trace Inequalities and Quantum Entropy*.

Theorem 13.16 (Operator convexity of $-\log$). *The function $x \mapsto -\log x$ is operator convex on $(0, \infty)$: for every matrix dimension its continuous functional calculus is convex in the Loewner order on the self-adjoint matrices with spectrum in $(0, \infty)$.*

Proof. Transport the statement along the $\mathbb{R}$-linear star-algebra equivalence Matrix $\simeq$ `CStarMatrix` (identity on the shared carrier, order- and cfc-preserving) onto the C^*-algebra where Mathlib's operator concavity of log lives. Negating concavity of log there and pulling back the Loewner inequality gives convexity of $-\log$ on Matrix. □

Theorem 13.17 (Hansen–Pedersen–Jensen operator-Jensen inequality). *Let f be operator convex on an interval I, let A, B be a contraction pair ($A^*A + B^*B = 1$), and let X, Y be self-adjoint with spectra in I. Then*

$$f(A^*XA + B^*YB) \ \leq \ A^*f(X)A + B^*f(Y)B.$$

Proof. The Effros / Hansen–Pedersen unitary-dilation method. Work in the doubled algebra $\mathrm{Matrix}(\mathrm{Fin}\,2 \times \mathrm{Fin}\,N)$, form the block diagonal $D = \mathrm{diag}(X, Y)$, dilate the column isometry $[A; B]$ to a unitary U, and pinch $M = U^*DU$ by the involution $V = \mathrm{diag}(1, -1)$: $\frac{1}{2}M + \frac{1}{2}VMV = \mathrm{diag}(M_{00}, M_{11})$. Operator convexity of f at dimension $2N$, applied to M and VMV, yields $f(\mathrm{diag}(M_{00}, M_{11})) \leq \mathrm{diag}(f(M)_{00}, f(M)_{11})$ blockwise, whose $(0,0)$-block is exactly the asserted inequality. (This node is stated for a general operator-convex f and specialized below to $f = -\log$.) □

Definition 13.18 (Operator perspective). For $f : \mathbb{R} \to \mathbb{R}$ and matrices L, R with R positive definite, the *operator perspective* is

$$P_f(L, R) \ = \ R^{1/2}\, f(R^{-1/2}\, L\, R^{-1/2})\, R^{1/2},$$

with the functional calculus and real powers supplied by the continuous functional calculus.

Theorem 13.19 (Effros' theorem: joint convexity of the perspective). *Let f be operator convex on I, let R_1, R_2 be positive definite with the sandwiched arguments $R_i^{-1/2}L_iR_i^{-1/2}$ self-adjoint and with spectra in I, and let $c \in [0, 1]$. Then*

$$P_f(cL_1 + (1-c)L_2,\ cR_1 + (1-c)R_2) \ \leq \ c\,P_f(L_1, R_1) + (1-c)\,P_f(L_2, R_2).$$

Proof. The Effros argument, a single application of Theorem 13.17. Writing $R = cR_1 + (1-c)R_2$, the contraction pair $A = \sqrt{c}\, R_1^{1/2} R^{-1/2}$, $B = \sqrt{1-c}\, R_2^{1/2} R^{-1/2}$ satisfies $A^*A + B^*B = R^{-1/2} R R^{-1/2} = 1$, and the self-adjoint arguments are $X = R_1^{-1/2} L_1 R_1^{-1/2}$, $Y = R_2^{-1/2} L_2 R_2^{-1/2}$. Applying Hansen–Pedersen–Jensen and conjugating the resulting inequality by $R^{1/2}$ reproduces exactly the joint-convexity estimate, once the sandwich cancellations $R^{1/2}R^{-1/2} = 1$ are used on both sides. ☐

Definition 13.20 (Operator perspective at a general index)**.** The same formula $P_f(L,R) = R^{1/2} f(R^{-1/2} L R^{-1/2}) R^{1/2}$ for matrices indexed by an arbitrary finite type m (rather than $\mathrm{Fin}\, N$).

Theorem 13.21 (Effros' theorem at a general finite index)**.** *Under the same hypotheses as Theorem 13.19, joint convexity of* opPersp f *holds over an arbitrary finite index type* m*.*

Proof. Transport along the star-algebra equivalence $\mathrm{Matrix}(m) \simeq \mathrm{Matrix}(\mathrm{Fin}(|m|))$ (a reindexing), which preserves positive definiteness, self-adjointness, spectra, the continuous functional calculus, real powers, and the Loewner order. Push the pair through the equivalence, invoke Theorem 13.19 on $\mathrm{Fin}(|m|)$, and reflect the resulting Loewner inequality back. ☐

Lemma 13.22 (Logarithm of a Kronecker product)**.** *For positive-definite* A, B, $\log(A \otimes B) = \log A \otimes 1 + 1 \otimes \log B$*.*

Proof. Simultaneously diagonalize by $U_A \otimes U_B$; the Kronecker of the two eigenvalue diagonals is $\mathrm{diag}(a_i b_j)$, and $\log(a_i b_j) = \log a_i + \log b_j$ splits the diagonal cfc additively across the two factors. Conjugating back by $U_A \otimes U_B$ gives the stated Kronecker splitting of log. ☐

Lemma 13.23 (Effros realization of the relative entropy)**.** *For positive-definite* ρ, σ*, the perspective of* $-\log$ *at the commuting pair* $L = 1 \otimes \sigma^\top$*,* $R = \rho \otimes 1$ *has the closed form*

$$P_{-\log}(1 \otimes \sigma^\top,\, \rho \otimes 1) = (\rho^{1/2}(\log \rho)\rho^{1/2}) \otimes 1 - \rho \otimes (\log \sigma)^\top.$$

Proof. The sandwiched argument is $R^{-1/2} L R^{-1/2} = \rho^{-1} \otimes \sigma^\top$. Apply Theorem 13.22 to $-\log(\rho^{-1} \otimes \sigma^\top) = \log \rho \otimes 1 - 1 \otimes (\log \sigma)^\top$ (using $\log(\rho^{-1}) = -\log \rho$ and the transpose-commutes-with-cfc identity $(\log \sigma)^\top = \log(\sigma^\top)$), then conjugate by $R^{1/2} = \rho^{1/2} \otimes 1$ and cancel with the sandwich to obtain the closed form. ☐

Lemma 13.24 (Scalar Effros functional)**.** *The positive linear functional* $M \mapsto \langle \mathrm{vec}\, 1,\, M\, \mathrm{vec}\, 1 \rangle$ *recovers the trace-form relative entropy from the perspective:*

$$\mathrm{relForm}\, P_{-\log}(1 \otimes \sigma^\top,\, \rho \otimes 1) = \mathrm{Tr}(\rho\, (\log \rho - \log \sigma)).$$

Proof. Apply relForm to the closed form of Theorem 13.23. On Kronecker products $\mathrm{relForm}(A \otimes C) = \mathrm{Tr}(AC^\top)$, so the two terms become $\mathrm{Tr}(\rho^{1/2}(\log \rho)\rho^{1/2}) = \mathrm{Tr}(\rho \log \rho)$ and $\mathrm{Tr}(\rho \log \sigma)$; their difference is the trace-form relative entropy. ☐

Definition 13.25 (Trace-form relative entropy)**.** For matrices ρ, σ, the trace-form relative entropy is $\mathrm{relEntropyMat}(\rho, \sigma) = \Re\, \mathrm{Tr}(\rho\, (\log \rho - \log \sigma))$, the logarithms taken through the continuous functional calculus.

Theorem 13.26 (Lieb's theorem: joint convexity of relative entropy)**.** *For positive-definite* $\rho_1, \rho_2, \sigma_1, \sigma_2$ *and* $c \in [0,1]$*,*

$$\mathrm{relEntropyMat}(c\rho_1 + (1-c)\rho_2,\ c\sigma_1 + (1-c)\sigma_2) \le c\, \mathrm{relEntropyMat}(\rho_1, \sigma_1) + (1-c)\, \mathrm{relEntropyMat}(\rho_2, \sigma_2).$$

Proof. Lieb's theorem (Lieb 1973; Carlen, *Trace Inequalities and Quantum Entropy*, Thm. 2.12), obtained from Effros' joint convexity of the perspective. The maps $\rho \mapsto R = \rho \otimes 1$ and $\sigma \mapsto L = 1 \otimes \sigma^\top$ are $\mathbb{R}$-linear, so applying Theorem 13.21 to $f = -\log$ (operator convex on $(0, \infty)$ by Theorem 13.16) gives a Loewner joint-convexity inequality for the perspective. The functional relForm is positive and linear; by Theorem 13.24 it turns that operator inequality into the scalar convexity of $(\rho, \sigma) \mapsto \mathrm{Tr}(\rho(\log \rho - \log \sigma))$, whose real part is relEntropyMat. □

Lemma 13.27 (Finite Jensen form of Lieb's theorem)**.** *For a finite convex combination of faithful states — weights $w_i \geq 0$ with $\sum_i w_i = 1$ and positive-definite ρ_i, σ_i —,*

$$\mathrm{relEntropyMat}\Big(\sum_i w_i \rho_i,\ \sum_i w_i \sigma_i\Big) \ \leq\ \sum_i w_i \, \mathrm{relEntropyMat}(\rho_i, \sigma_i).$$

Proof. The two-point joint convexity of Theorem 13.26 says relEntropyMat is a convex function on the set of pairs of positive-definite matrices; Jensen's inequality for a convex function over a finite convex combination upgrades the two-point estimate to the finite-sum form. □

Theorem 13.28 (Bridge to the spectral relative entropy)**.** *For density matrices ρ, σ with σ faithful (positive definite),* $\mathrm{relEntropyMat}(\rho, \sigma) = S(\rho \| \sigma)$*, the Umegaki relative entropy.*

Proof. Both sides expand through the spectral theorem: the trace form $\Re\, \mathrm{Tr}(\rho(\log \rho - \log \sigma))$ equals the spectral double sum defining $S(\rho \| \sigma)$ once the Hermitian cfc is written via eigendecompositions (this is Theorem 13.9 together with the eigenbasis form of the functional calculus). Hence the trace-form joint convexity above transfers to the Umegaki relative entropy on density matrices. □

13.4 The data-processing inequality

The Umegaki relative entropy is the finite-dimensional distinguishability functional whose *monotonicity under quantum channels* is the master inequality of quantum information. We derive it from Lieb's joint convexity in three moves: a Weyl-twirl reduction of the partial-trace case, a regularization to remove faithfulness, and a Stinespring dilation lifting the partial-trace case to every mixed-ancilla channel. We close with the no-recovery obstruction and the easy $(\Leftarrow)$ half of Petz's equality theorem (Petz 1986, 2003).

Definition 13.29 (The partial-trace DPI, as a monomorphic `Prop`)**.** `RelEntropyMonotoneUnderPartialTrace` is the proposition that for all finite index types n_A, n_E and all states ρ, σ on $n_A \times n_E$ with σ faithful (positive definite), tracing out the E-factor does not increase relative entropy,

$$S(\mathrm{Tr}_E\, \rho \| \, \mathrm{Tr}_E\, \sigma) \ \leq\ S(\rho \| \sigma).$$

The index types are quantified over `Type` (universe 0): finite-dimensional quantum information lives in universe 0, and pinning the universe lets this `Prop` be used as a reusable hypothesis whose subsystems unify with any consumer's.

Theorem 13.30 (Partial-trace DPI, faithful case)**.** *For positive-definite states ρ, σ on $n_A \times n_E$,*

$$S(\mathrm{Tr}_E\, \rho \| \, \mathrm{Tr}_E\, \sigma) \ \leq\ S(\rho \| \sigma).$$

Proof. Realize the partial trace as a Weyl twirl: with $d = \# n_E$ and $\tau = \frac{1}{d}\mathbf{1}$ the maximally mixed ancilla state, averaging ρ over the d^2 Heisenberg–Weyl unitaries W_{ab} on the E-factor gives $d^{-2} \sum_{a,b} (\mathbf{1} \otimes W_{ab}) \rho (\mathbf{1} \otimes W_{ab})^* = (\mathrm{Tr}_E\, \rho) \otimes \tau$, and likewise for σ. Each twirled state

$\rho_{ab} := (\mathbf{1} \otimes W_{ab})\rho(\mathbf{1} \otimes W_{ab})^*$ has $S(\rho_{ab}\|\sigma_{ab}) = S(\rho\|\sigma)$ by unitary invariance (Theorem 13.12), and the averaged pair has relative entropy $S((\operatorname{Tr}_E \rho) \otimes \tau \,\|\, (\operatorname{Tr}_E \sigma) \otimes \tau) = S(\operatorname{Tr}_E \rho \| \operatorname{Tr}_E \sigma)$ by ancilla additivity (Theorem 13.13). The finite-sum form of Lieb's joint convexity (Theorem 13.27) applied to the weights $w_{ab} = d^{-2}$ then yields $S(\operatorname{Tr}_E \rho \| \operatorname{Tr}_E \sigma) \le \sum_{ab} w_{ab}\, S(\rho_{ab}\|\sigma_{ab}) = S(\rho\|\sigma)$. □

Theorem 13.31 (Partial-trace DPI, unconditional)**.** *The proposition* `RelEntropyMonotoneUnderPartialTrace` *holds: the faithfulness hypothesis on ρ is unnecessary, so $S(\operatorname{Tr}_E \rho \| \operatorname{Tr}_E \sigma) \le S(\rho\|\sigma)$ for every ρ and every faithful σ.*

Proof. Regularize the first argument along the affine path $\rho_\varepsilon = (1-\varepsilon)\rho + \varepsilon\tau$ towards the maximally mixed state τ. For $\varepsilon \in (0, 1]$ the state ρ_ε is positive definite, so the faithful case (Theorem 13.30) gives $S(\operatorname{Tr}_E \rho_\varepsilon \| \operatorname{Tr}_E \sigma) \le S(\rho_\varepsilon\|\sigma)$. Both sides are continuous in ε at 0 (continuity of $M \mapsto S(M\|\sigma)$ on states with σ fixed and faithful, and continuity of the partial trace), and $\rho_\varepsilon \to \rho$; passing to the limit $\varepsilon \downarrow 0$ preserves the inequality. □

Lemma 13.32 (Tracing out an adjoined ancilla is the identity)**.** *For any state ρ and any ancilla state α, $\operatorname{Tr}_E(\rho \otimes \alpha) = \rho$.*

Proof. Entrywise, $(\operatorname{Tr}_E(\rho \otimes \alpha))_{ii'} = \sum_j \rho_{ii'}\alpha_{jj} = \rho_{ii'} \cdot \operatorname{Tr}\alpha = \rho_{ii'}$, using only $\operatorname{Tr}\alpha = 1$. □

Theorem 13.33 (Isometric-embedding invariance)**.** *For states ρ, σ on n, a faithful ancilla α on e, a unitary U on $n \times e$, and faithful σ,*

$$S(U(\rho \otimes \alpha)U^* \,\|\, U(\sigma \otimes \alpha)U^*) \;=\; S(\rho\|\sigma).$$

Proof. Unitary conjugation leaves relative entropy invariant (Theorem 13.12), reducing the claim to $S(\rho \otimes \alpha\|\sigma \otimes \alpha) = S(\rho\|\sigma)$, which is ancilla additivity (Theorem 13.13) for the faithful ancilla α. □

Theorem 13.34 (Stinespring reduction of the DPI)**.** Given *the partial-trace DPI (Theorem 13.29), every Stinespring-dilated channel $\Lambda\rho = \operatorname{Tr}_E(U(\rho\otimes\alpha)U^*)$ — adjoin a faithful ancilla α, conjugate by a unitary dilation U, then trace out the ancilla — is relative-entropy monotone on faithful states: $S(\Lambda\rho\|\Lambda\sigma) \le S(\rho\|\sigma)$.*

Proof. Apply the wall (Theorem 13.29) to the dilated states $U(\rho \otimes \alpha)U^*$ and $U(\sigma \otimes \alpha)U^*$ (the latter faithful, being a unitary conjugate of a Kronecker product of faithful states): tracing out the ancilla gives $S(\operatorname{Tr}_E U(\rho \otimes \alpha)U^* \| \operatorname{Tr}_E U(\sigma \otimes \alpha)U^*) \le S(U(\rho \otimes \alpha)U^* \| U(\sigma \otimes \alpha)U^*)$. The right-hand side equals $S(\rho\|\sigma)$ by isometric-embedding invariance (Theorem 13.33). □

Theorem 13.35 (Data-processing inequality for the faithful-ancilla mixed-Stinespring family)**.** *Unconditionally, for a faithful ancilla α, a unitary dilation U, and faithful σ,*

$$S(\operatorname{Tr}_E U(\rho \otimes \alpha)U^* \,\|\, \operatorname{Tr}_E U(\sigma \otimes \alpha)U^*) \;\le\; S(\rho\|\sigma).$$

This is the data-processing inequality for the faithful-ancilla mixed-Stinespring family; that family does not *contain every CPTP channel (amplitude damping lies outside it, since its exact dilation needs a pure, non-faithful ancilla), so it is not an in-repo DPI for a general Kraus channel.*

Proof. Feed the now-discharged partial-trace DPI (Theorem 13.31) into the Stinespring reduction (Theorem 13.34), removing its hypothesis. □

Theorem 13.36 (No faithful-monotone recovery section under a strict drop)**.** *If a map R satisfies the faithful data-processing inequality and inverts a coarse-graining Λ on the pair ρ, σ ($R(\Lambda\rho) = \rho$, $R(\Lambda\sigma) = \sigma$ with $\Lambda\sigma$ faithful), then a* strict *drop $S(\Lambda\rho\|\Lambda\sigma) < S(\rho\|\sigma)$ is impossible.*

Proof. Monotonicity of R applied to $\Lambda\rho, \Lambda\sigma$ gives $S(R(\Lambda\rho)\|R(\Lambda\sigma)) \leq S(\Lambda\rho\|\Lambda\sigma)$; the section identities rewrite the left side as $S(\rho\|\sigma)$, so $S(\rho\|\sigma) \leq S(\Lambda\rho\|\Lambda\sigma)$, contradicting the assumed strict drop. □

Corollary 13.37 (No Stinespring recovery under a strict drop)**.** *Unconditionally: a strict relative-entropy drop under a coarse-graining Λ (an endomorphism of the input space) rules out* any *Stinespring recovery channel $R\,x = \mathrm{Tr}_E(U(x \otimes \alpha)U^*)$ inverting Λ on ρ, σ.*

Proof. Specialize Theorem 13.36 to the Stinespring recovery map, whose faithful data-processing monotonicity is exactly Theorem 13.35. □

The equality case of the data-processing inequality — that saturation $S(\Lambda\rho\|\Lambda\sigma) = S(\rho\|\sigma)$ is equivalent to the existence of a Petz recovery map inverting Λ on ρ, σ — is *Petz's equality theorem.* Its elementary direction (recovery $\Rightarrow$ saturation) rests only on the monotonicity established above; the converse (saturation $\Rightarrow$ recovery) is the analytic heart of the next chapter. Both are treated in full in Chapter 14.

Chapter 14

Petz recovery and quantum dynamical entropy

This chapter documents a self-contained finite-dimensional quantum-information layer (issues #22–#28) that is built on the same matrix / continuous-functional-calculus infrastructure as the multiplicative ergodic theorem, but is logically independent of it. The objects are density matrices ρ, σ on $\mathbb{C}^d$ and completely positive trace-preserving maps (quantum channels) between matrix algebras. The central quantity is the *Umegaki relative entropy*

$$S(\rho\|\sigma) \;=\; \operatorname{tr}(\rho\,(\log\rho - \log\sigma))$$

(Lean: `relEntropy`); its defining feature is the *data-processing inequality* $S(\Lambda\rho\|\Lambda\sigma) \leq S(\rho\|\sigma)$ for every channel Λ, a consequence of Lieb's joint-convexity theorem. Two threads are developed here. The first is the *Petz recovery theorem* and both directions of Petz's equality theorem: a channel saturates the data-processing inequality on a pair of faithful states *if and only if* the input is exactly reconstructed by the Petz transpose recovery map. The general saturation $\Rightarrow$ recovery direction (Petz, *Monotonicity of quantum relative entropy revisited*, Rev. Math. Phys. 15 (2003), Thm 2; via the operator-Jensen / $-\log$ Loewner route of Carlen–Vershynina, *Recovery map stability for the data processing inequality*, 2020) is proved with **no injectivity hypothesis** on the channel, so it covers information-losing channels such as the completely depolarising channel. The second thread is the Connes–Narnhofer–Thirring quantum dynamical entropy (Connes–Narnhofer–Thirring, *Dynamical entropy of C^* algebras and von Neumann algebras*, Comm. Math. Phys. 112 (1987) 691–719, in the operational-partition formulation of Alicki–Fannes 1994), whose abelian corner collapses onto the classical Kolmogorov–Sinai entropy of the underlying measure-preserving system, tying the quantum layer back to the ergodic theory of the core. Every node below is formalized sorry-free and audited to depend on exactly {`propext`, `Classical.choice`, `Quot.sound`}. Throughout, $X^\dagger$ denotes the conjugate transpose (Hilbert–Schmidt adjoint) of X.

14.1 Kraus channels and the Petz recovery map

Definition 14.1 (Kraus channel)**.** A *Kraus channel* on $\mathrm{Matrix}_n(\mathbb{C})$ is a finite family of Kraus operators $K : \iota \to \mathrm{Matrix}_n(\mathbb{C})$ satisfying the completeness relation $\sum_i K_i^\dagger K_i = 1$. Its *Schrödinger action* is $\Lambda(X) = \sum_i K_i\, X\, K_i^\dagger$ (Lean: `toMat`, restricted to states as `toDM`), and its *Heisenberg (Hilbert–Schmidt) adjoint* is $\Lambda^\dagger(X) = \sum_i K_i^\dagger X\, K_i$ (Lean: `adj`).

Lemma 14.2 (Trace preservation)**.** *A Kraus channel is trace preserving,* $\mathrm{tr}(\Lambda X) = \mathrm{tr}\, X$ *for every* X*; consequently the adjoint is unital,* $\Lambda^\dagger(1) = 1$ *(Lean:* `adj_unital`*).*

Proof. By trace cyclicity $\mathrm{tr}(K_i X K_i^\dagger) = \mathrm{tr}(K_i^\dagger K_i X)$; summing over i and applying $\sum_i K_i^\dagger K_i = 1$ collapses the sum to $\mathrm{tr}\, X$. Unitality of $\Lambda^\dagger$ is the same completeness relation read at $X = 1$. □

Definition 14.3 (Petz recovery map)**.** For a state σ and a channel Λ, the *Petz (transpose) recovery map* is

$$P_{\sigma,\Lambda}(X) \;=\; \sqrt{\sigma}\,\Lambda^\dagger\big((\Lambda\sigma)^{-1/2}\,X\,(\Lambda\sigma)^{-1/2}\big)\,\sqrt{\sigma},$$

realised through the continuous functional calculus (`CFC.conjSqrt` for the $\sqrt{\cdot}$-conjugations and the ring inverse for $(\Lambda\sigma)^{-1/2}$).

Theorem 14.4 (Petz recovery identity)**.** *If the channel output* $\Lambda\sigma$ *is positive definite, then the Petz map recovers* σ *from* $\Lambda\sigma$*:*

$$P_{\sigma,\Lambda}(\Lambda\sigma) \;=\; \sigma.$$

Proof. Feeding $X = \Lambda\sigma$ into $P_{\sigma,\Lambda}$, the inner conjugation $(\Lambda\sigma)^{-1/2}(\Lambda\sigma)(\Lambda\sigma)^{-1/2} = 1$ collapses to the identity (this is where positive definiteness of $\Lambda\sigma$ is used); unitality $\Lambda^\dagger(1) = 1$ then leaves $\sqrt{\sigma}\cdot 1\cdot\sqrt{\sigma} = \sigma$. □

14.2 Petz's equality theorem

The Petz map already recovers σ unconditionally (Theorem 14.4); the content of the equality theorem is that it recovers the *other* state ρ exactly when the channel loses no relative entropy between ρ and σ. One direction is elementary and rests only on monotonicity (the data-processing inequality); the other is the analytic heart of the chapter.

Theorem 14.5 (Recovery ⇒ saturation)**.** *Let* Λ *and* R *be maps on states, each monotone under relative entropy (each satisfies the data-processing inequality against positive-definite second arguments). If* R *is a recovery section for the pair* ρ, σ*, i.e.* $R(\Lambda\rho) = \rho$ *and* $R(\Lambda\sigma) = \sigma$ *(with* $\sigma, \Lambda\sigma$ *positive definite), then the data-processing inequality is saturated:*

$$S(\Lambda\rho\|\Lambda\sigma) \;=\; S(\rho\|\sigma).$$

Proof. Monotonicity of Λ gives $S(\Lambda\rho\|\Lambda\sigma) \le S(\rho\|\sigma)$. For the reverse inequality, apply monotonicity of R to the pair $\Lambda\rho, \Lambda\sigma$ and rewrite through the section identities $R(\Lambda\rho) = \rho$, $R(\Lambda\sigma) = \sigma$; this yields $S(\rho\|\sigma) \le S(\Lambda\rho\|\Lambda\sigma)$. Antisymmetry gives equality. □

14.2.1 The Choi $-\log$ Loewner inequality

The engine of the converse is operator convexity of $-\log$, packaged first as a Loewner inequality for a rectangular isometry and then specialised to the Kraus column.

Theorem 14.6 (Rectangular isometry $-\log$ Loewner inequality)**.** *Let* $W : \mathbb{C}^q \to \mathbb{C}^p$ *be an isometry (*$W^\dagger W = 1$*) and* X *self-adjoint with spectrum in* $(0, \infty)$*. Then*

$$(-\log)(W^\dagger X\,W) \;\le\; W^\dagger\,(-\log)(X)\,W.$$

Proof. Extend the orthonormal columns of W to a unitary U; conjugating X by U and reindexing, $W^\dagger X W$ is the upper-left corner of $A = U^\dagger X U$. Operator convexity of $-\log$ gives the corner inequality $(-\log)(\text{corner of } A) \le \text{corner of } (-\log)(A)$ (`sum_corner_loewner` applied to `operatorConvexOn_neg_log`), which is exactly the claim after identifying both corners. □

Theorem 14.7 (Choi $-\log$ operator inequality)**.** *For Kraus operators K with $\sum_i K_i^\dagger K_i = 1$ and X self-adjoint with spectrum in $(0,\infty)$,*

$$(-\log)\Big(\sum_i K_i^\dagger X K_i\Big) \;\le\; \sum_i K_i^\dagger\,(-\log)(X)\,K_i.$$

Proof. Stack the Kraus operators into the column isometry V and let X_{bd} be the block-diagonal amplification of X. Then $V^\dagger X_{\mathrm{bd}} V = \sum_i K_i^\dagger X K_i$ and, since $-\log$ commutes with block diagonals, $V^\dagger(-\log)(X_{\mathrm{bd}})V = \sum_i K_i^\dagger(-\log)(X)K_i$. Apply Theorem 14.6 to (V, X_{bd}). □

14.2.2 The modular realisation of relative entropy

Lemma 14.8 (Modular form of the relative entropy)**.** *For faithful states ρ, σ, with (vectorised) relative modular operator $\Delta = \sigma \otimes (\rho^{-1})^\top$ and cyclic vector $\xi = \mathrm{vec}(\rho^{1/2})$,*

$$S(\rho\|\sigma) \;=\; \mathrm{Re}\langle \xi,\ (-\log)(\Delta)\,\xi\rangle.$$

Proof. Compute $(-\log)(\sigma\otimes(\rho^{-1})^\top) = -(\log\sigma)\otimes 1 + 1\otimes(\log\rho)^\top$ (via log of a Kronecker product, $\log\rho^{-1} = -\log\rho$, and the transpose law). Apply this to $\mathrm{vec}(\rho^{1/2})$ through the vec/Kronecker rule $(A\otimes B^\top)\,\mathrm{vec}\,X = \mathrm{vec}(AXB)$, read off via the Hilbert–Schmidt inner product $\langle \mathrm{vec}\,X, \mathrm{vec}\,Y\rangle = \mathrm{tr}(X^\dagger Y)$, and use $\rho^{1/2}\rho^{1/2} = \rho$ with trace cyclicity to obtain $\mathrm{tr}(\rho(\log\rho - \log\sigma)) = S(\rho\|\sigma)$. □

14.2.3 The general channel: contraction rigidity (issue #28, no injectivity)

For a *general* Kraus channel the vectorised Petz map $W =$ `petzWChanVec` is only a *contraction* ($W^\dagger W \le 1$), and the output modular operator $\Delta_{\mathrm{out}} = (\Lambda\sigma)\otimes((\Lambda\rho)^{-1})^\top$ is a genuinely separate operator: the whole-space Loewner bound $W^\dagger(-\log\Delta)W \succeq -\log\Delta_{\mathrm{out}}$ *fails* for a contraction. Two adaptations carry the argument through. The $-\log$ saturation is supplied as a *scalar* equality of quadratic forms, and the compression $Y = W^\dagger\Delta W$ is *never inverted*, which is precisely what removes the injectivity hypothesis.

Theorem 14.9 (Scalar channel $-\log$ modular gap)**.** *For a Kraus channel Λ with all four states $\rho, \sigma, \Lambda\rho, \Lambda\sigma$ faithful, if data processing is saturated ($S(\rho\|\sigma) = S(\Lambda\rho\|\Lambda\sigma)$), then the two $-\log$ modular quadratic forms agree at the output cyclic vector $\xi = \mathrm{vec}((\Lambda\rho)^{1/2})$:*

$$\mathrm{Re}\langle \xi,\ W^\dagger(-\log\Delta)W\,\xi\rangle \;=\; \mathrm{Re}\langle \xi,\ (-\log\Delta_{\mathrm{out}})\,\xi\rangle,$$

with $W =$ `petzWChanVec`, $\Delta = \sigma\otimes(\rho^{-1})^\top$, $\Delta_{\mathrm{out}} = (\Lambda\sigma)\otimes((\Lambda\rho)^{-1})^\top$.

Proof. Since $W\xi = \mathrm{vec}(\rho^{1/2})$ (cyclicity of the channel contraction), the left form is $S(\rho\|\sigma)$ and the right form is $S(\Lambda\rho\|\Lambda\sigma)$ by Lemma 14.8; the entropy equality is precisely their coincidence. □

Lemma 14.10 (Injectivity-free gap decomposition)**.** *For a contraction W with defect $(1 - W^\dagger W)\xi = 0$, positive-definite shifts X, Out, writing $\eta = \mathrm{Out}^{-1}\xi$, $b = X^{-1}(W\xi) - W\eta$, $Y = W^\dagger X W$,*

$$\langle \xi,\ (W^\dagger X^{-1} W - \mathrm{Out}^{-1})\,\xi\rangle \;=\; \langle b,\ X\,b\rangle \;+\; \langle \eta,\ (\mathrm{Out} - Y)\,\eta\rangle.$$

Proof. A pure matrix identity: the isometric proof's Y^{-1} bridge is replaced by $\mathrm{Out}^{-1}Y\mathrm{Out}^{-1}$, which cancels between the two summands, so the compression Y is never inverted. Expanding $\langle b, Xb\rangle$ produces $W^\dagger X^{-1}W$, two cross terms $W^\dagger W\mathrm{Out}^{-1}$ and $\mathrm{Out}^{-1}(W^\dagger W)$, and $\mathrm{Out}^{-1}Y\mathrm{Out}^{-1}$; the contraction defect collapses each cross term (each contains $W^\dagger W\xi = \xi$) to Out^{-1}, and the second summand contributes $\mathrm{Out}^{-1} - \mathrm{Out}^{-1}Y\mathrm{Out}^{-1}$. Adding and simplifying leaves exactly $W^\dagger X^{-1}W - \mathrm{Out}^{-1}$. □

Lemma 14.11 (Per-t intertwining at gap zero)**.** *In the setting of Lemma 14.10, with the compression bound* $W^\dagger X W \le \mathrm{Out}$*, if the total resolvent gap* $\mathrm{Re}\langle \xi, (W^\dagger X^{-1} W - \mathrm{Out}^{-1})\xi\rangle$ *vanishes, then* $X^{-1}(W\xi) = W(\mathrm{Out}^{-1}\xi)$*.*

Proof. Both summands of Lemma 14.10 are nonnegative ($X \succ 0$ and $\mathrm{Out} - Y \succeq 0$), and their real parts sum to the vanishing gap; hence each is zero. In particular $\mathrm{Re}\langle b, Xb\rangle = 0$, and positive definiteness of X forces $b = 0$, i.e. $X^{-1}(W\xi) = W(\mathrm{Out}^{-1}\xi)$. ☐

Theorem 14.12 (Scalar-sourced contraction rigidity spine)**.** *Let W be a contraction ($W^\dagger W \le 1$), $\Delta, \Delta_{\mathrm{out}}$ positive definite with compression bound $W^\dagger \Delta W \le \Delta_{\mathrm{out}}$, cyclic-norm condition $\|W\xi\| = \|\xi\|$, and the* scalar $-\log$ *saturation* $\mathrm{Re}\langle \xi, W^\dagger(-\log \Delta)W\xi\rangle = \mathrm{Re}\langle \xi, (-\log \Delta_{\mathrm{out}})\xi\rangle$*. Then for every $t > 0$,*

$$(\Delta + t)^{-1}(W\xi) \;=\; W((\Delta_{\mathrm{out}} + t)^{-1}\xi).$$

Proof. Represent $-\log$ by the integral $\int_0^\infty ((1+t)^{-1} - (\cdot + t)^{-1})\,dt$ and let $F(t)$ be the real resolvent-gap quadratic form at ξ. By the shifted compression bound $W^\dagger(\Delta + t)W \le \Delta_{\mathrm{out}} + t$ and the injectivity-free decomposition, $F(t) \ge 0$ pointwise; integrating recovers the scalar $-\log$ gap, which is 0 by hypothesis. A nonnegative continuous integrand with zero integral vanishes on $(0, \infty)$, so each $F(t) = 0$; per-t saturation (Lemma 14.11) gives the resolvent intertwining. (The general finite square index follows by an `equivFin` reindexing.) ☐

Lemma 14.13 (Contraction intertwines every continuous function)**.** *Under the hypotheses of Theorem 14.12, for every continuous g,*

$$W(g(\Delta_{\mathrm{out}})\,\xi) \;=\; g(\Delta)\,(W\xi).$$

Proof. On the finite union of the two spectra, g is a real-coefficient combination of resolvents $x \mapsto (x+t)^{-1}$ (finite-spectrum resolvent readoff). Both $g(\Delta_{\mathrm{out}})$ and $g(\Delta)$ become the corresponding operator-resolvent combinations, and the per-resolvent intertwining of Theorem 14.12 propagates linearly. ☐

Theorem 14.14 (Channel unitary-power intertwining)**.** *Under entropy saturation, the channel contraction intertwines the modular unitary power on the output cyclic vector,* $W(\Delta_{\mathrm{out}}^{it}\,\xi) = \Delta^{it}\,(W\xi)$*, with* $W\xi = \mathrm{vec}(\rho^{1/2})$*.*

Proof. Apply Lemma 14.13 with $g = \cos(t \log \cdot)$ and $g = \sin(t \log \cdot)$; the scalar saturation input is Theorem 14.9. The complex power Δ^{it} is the functional-calculus combination $\cos(t\log\cdot) + i\sin(t\log\cdot)$, so the two intertwinings combine to the unitary-power intertwining. ☐

Definition 14.15 (Modular it-intertwining)**.** The channel adjoint *intertwines the modular it-flows* if, for all t,

$$\Lambda^\dagger((\Lambda\rho)^{it}(\Lambda\sigma)^{-it}) \;=\; \rho^{it}\,\sigma^{-it}.$$

Theorem 14.16 (Equality ⇒ modular intertwining (general channel))**.** *For any Kraus channel with all four states faithful, saturation $S(\rho\|\sigma) = S(\Lambda\rho\|\Lambda\sigma)$ implies* `IntertwinesIt`*.*

Proof. From Theorem 14.14, reading off the vec-action $\Delta^{it}\,\mathrm{vec}\,X = \mathrm{vec}(P^{it} X R^{-it})$ turns the vectorised statement into $\Lambda^\dagger$-form directly (the vectorised Petz map already contains $\Lambda^\dagger$, so no amplification is needed). Cancelling the $(\Lambda\rho)^{\pm 1/2}$ twist on the input column and the $\rho^{1/2}$ factor, then taking adjoints, yields $\Lambda^\dagger((\Lambda\rho)^{it}(\Lambda\sigma)^{-it}) = \rho^{it}\sigma^{-it}$. ☐

Theorem 14.17 (Modular intertwining ⇒ Petz recovery)**.** *If* `IntertwinesIt` *holds (all four states faithful), then the Petz map recovers the input: $P_{\sigma,\Lambda}(\Lambda\rho) = \rho$.*

Proof. Analytic continuation of the it-intertwining to the value $t = -i/2$ (a Kadison-type argument on the modular flow) turns the cocycle identity into $\sqrt{\sigma}\,\Lambda^\dagger((\Lambda\sigma)^{-1/2}(\Lambda\rho)(\Lambda\sigma)^{-1/2})\sqrt{\sigma} = \rho$, which is exactly $P_{\sigma,\Lambda}(\Lambda\rho) = \rho$. ☐

Theorem 14.18 (General Petz recovery from equality — issue #28 headline)**.** *Let* Λ *be* any *Kraus channel and* ρ, σ *states with all four of* $\rho, \sigma, \Lambda\rho, \Lambda\sigma$ *positive definite. If data processing is saturated,*

$$S(\rho\|\sigma) \;=\; S(\Lambda\rho\|\Lambda\sigma),$$

then the Petz recovery map reconstructs the input state,

$$P_{\sigma,\Lambda}(\Lambda\rho) \;=\; \rho.$$

No injectivity of the channel (or of the vectorised Petz map) is assumed; the result holds for information-losing channels such as the completely depolarising channel.

Proof. Compose Theorem 14.16 (entropy equality ⇒ modular it-intertwining) with Theorem 14.17 (intertwining ⇒ recovery). This is the full-generality form of Petz (2003, Thm 2). Together with Theorem 14.5 it closes the equivalence: for faithful states, saturation of the data-processing inequality holds if and only if the Petz map recovers the input. ☐

14.3 The modular-cocycle intertwining and the injectivity-free route

The general contraction argument of the previous section is a deformation of a cleaner isometric prototype, which is worth recording on its own because it is the analytic heart of the whole equality theorem. For the partial-trace channel $\Lambda = \mathrm{Tr}_B$ on a bipartite system with faithful dilated states ω, τ (and faithful marginals $\mathrm{Tr}_B\,\omega, \mathrm{Tr}_B\,\tau$), the vectorised Petz map $W =$ `petzWvec` is a genuine *isometry*. There the whole-space $-\log$ Loewner bound (Theorem 14.6) is available, so the saturation input can be taken as a full operator gap rather than a scalar one, and the compression may be inverted freely — this is exactly the injectivity that the general route of §14.2 had to dispense with, replacing the Y^{-1} bridge by the cancelling decomposition of Lemma 14.10.

Theorem 14.19 (Partial-trace modular gap)**.** *If the partial-trace relative entropy is preserved,* $S(\mathrm{Tr}_B\,\omega\|\,\mathrm{Tr}_B\,\tau) = S(\omega\|\tau)$*, then the rectangular* $-\log$ *operator-Jensen gap for the Petz isometry* W *annihilates the output cyclic vector* $\xi = \mathrm{vec}((\mathrm{Tr}_B\,\omega)^{1/2})$*:*

$$(W^\dagger(-\log)(\Delta)\,W)\,\xi \;=\; (-\log)(W^\dagger\Delta\,W)\,\xi, \qquad \Delta = \tau\otimes(\omega^{-1})^\top.$$

Proof. By Theorem 14.6 the operator $B - A \geq 0$, where $B = W^\dagger(-\log)(\Delta)W$ and $A = (-\log)(W^\dagger\Delta W)$. Using Lemma 14.8 on both sides (the compression $W^\dagger\Delta W$ is exactly the output modular operator, and $W\xi = \mathrm{vec}(\omega^{1/2})$), the two quadratic forms at ξ equal $S(\mathrm{Tr}_B\,\omega\|\,\mathrm{Tr}_B\,\tau)$ and $S(\omega\|\tau)$; the entropy equality makes $\mathrm{Re}\langle\xi,(B-A)\xi\rangle = 0$. For a positive semidefinite matrix a vanishing real expectation forces $(B-A)\xi = 0$. ☐

Theorem 14.20 (Partial-trace equality ⇒ modular intertwining)**.** *Under the same entropy-preservation hypothesis, the amplified output modular it-cocycle equals the input one, for all* $t \in \mathbb{R}$*:*

$$((\mathrm{Tr}_B\,\omega)^{it}(\mathrm{Tr}_B\,\tau)^{-it})\otimes 1 \;=\; \omega^{it}\,\tau^{-it}.$$

Proof. The gap of Theorem 14.19 is upgraded, first to an intertwining of every resolvent $(\Delta+t)^{-1}$ (the isometric rigidity tail), then — via a finite-spectrum resolvent readoff — to the intertwining of *every* continuous function of Δ on ξ. Feeding the pair $\cos(t\log\cdot), \sin(t\log\cdot)$ assembles the unitary power Δ^{it}, giving $W(\Delta_A^{it}\xi) = \Delta^{it}(W\xi)$. Reading off the vec-action $\Delta^{it}\operatorname{vec} X = \operatorname{vec}(P^{it}XR^{-it})$, cancelling the $\omega_A^{\pm 1/2}$ twist on the input column and the $\omega^{1/2}$ factor, and taking adjoints yields the stated cocycle identity — an instance of the general `IntertwinesIt` property (Definition 14.15). □

14.4 The Connes–Narnhofer–Thirring dynamical entropy and its abelian corner

This section documents the finite-dimensional quantum dynamical entropy of Connes–Narnhofer–Thirring, in the operational-partition formulation of Alicki–Fannes (see also Ohya–Petz, *Quantum Entropy and Its Use*). The dynamics is a unital $*$-endomorphism Φ of the matrix algebra $\mathrm{Matrix}_d(\mathbb{C})$, the observable is a finite operational partition of unity, and the entropy is read off from a family of correlation density matrices. The headline is that on the *abelian corner* — diagonal dynamics on a diagonal state — the whole construction collapses onto the classical Kolmogorov–Sinai entropy of the underlying measure-preserving system (Lean: `ErgodicTheory.Entropy.ksEntropy`), tying the quantum layer back to the ergodic theory of the MET core.

Definition 14.21 (Unital $*$-endomorphism)**.** A *finite quantum dynamics* is a map $\Phi : \mathrm{Matrix}_d(\mathbb{C}) \to \mathrm{Matrix}_d(\mathbb{C})$ that is additive, multiplicative, unital ($\Phi(1) = 1$) and $*$-preserving ($\Phi(x^\dagger) = \Phi(x)^\dagger$). Additivity is carried as part of the datum: it is automatic for a $*$-homomorphism of matrix algebras, and it is exactly what lets Φ commute with the finite sums appearing in the telescoping identity below.

Definition 14.22 (Operational partition of unity)**.** An *operational partition of unity* of size k is a family $(x_i)_{i<k}$ of operators in $\mathrm{Matrix}_d(\mathbb{C})$ satisfying the partition-of-unity relation $\sum_i x_i^\dagger x_i = 1$. This is the noncommutative analogue of a measurable partition of the state space.

Definition 14.23 (Time-ordered refinement)**.** Given a dynamics Φ and an operational partition $X = (x_i)$, the *time-ordered refinement* of depth n along a word $f \in (\mathrm{Fin}\, k)^{\mathrm{Fin}\, n}$ is the operator

$$\mathrm{refine}\,\Phi\, X\, n\, f \;=\; x_{f_0}\,\Phi(x_{f_1})\,\Phi^2(x_{f_2})\cdots\Phi^{n-1}(x_{f_{n-1}}),$$

defined by the telescoping recursion $\mathrm{refine}(n+1, f) = x_{f_0}\cdot\Phi\big(\mathrm{refine}(n, \mathrm{tail}\, f)\big)$ with $\mathrm{refine}(0,\cdot) = 1$. It records the observable measured along the first n steps of the orbit under Φ.

Lemma 14.24 (Telescoping identity)**.** *The refinement of an operational partition of unity is again an operational partition of unity: for every n,*

$$\sum_{f\in(\mathrm{Fin}\,k)^{\mathrm{Fin}\,n}} \big(\mathrm{refine}\,\Phi\,X\,n\,f\big)^\dagger\,\big(\mathrm{refine}\,\Phi\,X\,n\,f\big) \;=\; 1.$$

Proof. Induct on n. Splitting the word as $f = (i, g)$ with $i = f_0$, the summand factors as $\Phi(\mathrm{refine}(n,g))^\dagger\,(x_i^\dagger x_i)\,\Phi(\mathrm{refine}(n,g))$; summing over the leading letter i and applying $\sum_i x_i^\dagger x_i = 1$ removes it. What remains is $\sum_g \Phi\big(\mathrm{refine}(n,g)^\dagger \mathrm{refine}(n,g)\big)$; pulling Φ out of the finite sum (its additivity, Definition 14.21) and applying the inductive hypothesis leaves $\Phi(1) = 1$. □

Definition 14.25 (Correlation density matrix)**.** For a dynamics Φ, a state ρ (a density matrix on $\mathbb{C}^d$) and an operational partition X, the depth-n *correlation density matrix* $\mathrm{corrMatrix}\,\Phi\,\rho\,X\,n$ is the matrix on the classical index set $(\mathrm{Fin}\,k)^{\mathrm{Fin}\,n}$ with entries

$$(g, f) \;\longmapsto\; \mathrm{tr}\big(\rho\,(\mathrm{refine}\,\Phi\,X\,n\,g)^\dagger\,\mathrm{refine}\,\Phi\,X\,n\,f\big).$$

It is a genuine density matrix: its trace is 1 by the telescoping identity (Lemma 14.24) together with $\mathrm{tr}\,\rho = 1$, and it is positive semidefinite because the quadratic form $x^\dagger M x$ equals $\mathrm{tr}(T\rho T^\dagger) \geq 0$ for $T = \sum_f x_f\,\mathrm{refine}(n, f)$.

Definition 14.26 (CNT/ALF dynamical entropy)**.** The *entropy of a partition* is the infimum von Neumann entropy rate

$$h_\Phi(\rho, X) \;=\; \inf_{n\geq 1} \frac{S\big(\mathrm{corrMatrix}\,\Phi\,\rho\,X\,n\big)}{n}, \qquad S(\tau) = -\,\mathrm{tr}(\tau\log\tau),$$

and the *CNT dynamical entropy* of Φ in the state ρ is the supremum over all finite operational partitions, $h_\Phi(\rho) = \sup_{k,X} h_\Phi(\rho, X)$ (valued in $\overline{\mathbb{R}}$). The $\inf_n$ form is the honest analogue of the subadditive limit $\lim_n S(\cdot)/n$ and sidesteps an operator-Fekete argument.

Definition 14.27 (Diagonal state)**.** For a probability vector $\mu : \mathrm{Fin}\,d \to \mathbb{R}_{\geq 0}$ ($\sum_i \mu_i = 1$), the associated diagonal state is $\rho_\mu = \mathrm{diag}\,\mu$, a density matrix on $\mathbb{C}^d$.

Definition 14.28 (Projection partition)**.** For a cell map $c : \mathrm{Fin}\,d \to \mathrm{Fin}\,k$, the diagonal *projection partition* is $\{\mathrm{diag}\,\mathbf{1}_{c^{-1}(i)}\}_{i<k}$; it is an operational partition of unity encoding the measurable partition $c^{-1}(\cdot)$ of $\mathrm{Fin}\,d$.

Definition 14.29 (The permutation dynamics (abelian corner))**.** For a permutation $\sigma \in \mathfrak{S}_d$, the *permutation dynamics* $\mathrm{adPerm}\,\sigma$ is conjugation $x \mapsto P_\sigma\,x\,P_\sigma^\dagger$ by the permutation matrix, a unital $*$-endomorphism; on diagonal matrices it acts by $(\mathrm{adPerm}\,\sigma)(\mathrm{diag}\,v) = \mathrm{diag}(v \circ \sigma)$. Paired with a σ-invariant diagonal state ρ_μ ($\mu \circ \sigma = \mu$, Definition 14.27) and a projection partition (Definition 14.28), this is the abelian corner: it mirrors the classical system $(\mathrm{Fin}\,d, \mu, \sigma)$ with the measurable partition $c^{-1}(\cdot)$.

Theorem 14.30 (Per-resolution diagonal collapse)**.** *On the abelian corner the correlation matrix is diagonal, carrying the masses of the classical n-fold join on its diagonal; hence its von Neumann entropy equals the classical iterated-join Shannon entropy of the system* $(\mathrm{Fin}\,d, \mu, \sigma)$*:*

$$S\big(\mathrm{corrMatrix}\,(\mathrm{adPerm}\,\sigma)\,\rho_\mu\,(\mathrm{projPartition}\,c)\,n\big) \;=\; H\Big(\bigvee\nolimits_{l<n} \sigma^{-l}\,c^{-1}(\cdot)\Big).$$

Proof. Because $\mathrm{adPerm}\,\sigma$ preserves diagonal matrices, the refinement of a projection partition is again diagonal, and the depth-n refinement product along a word f is exactly the indicator of the classical join cell $\bigcap_{l<n} \sigma^{-l} c^{-1}(f_l)$. Feeding these into the correlation entries and using that ρ_μ is diagonal, the off-diagonal entries vanish (two distinct words disagree in some slot, forcing an orthogonal pair of indicators) and the (f, f) entry is the mass $\mu(\bigcap_{l<n} \sigma^{-l} c^{-1}(f_l))$ of the join cell. The correlation matrix is therefore diag of the join distribution, so by $S(\mathrm{diag}\,p) = -\sum_j p_j \log p_j$ its von Neumann entropy is the Shannon entropy of the join — the classical join entropy at resolution n. □

Non-vacuity. The per-resolution collapse is not the trivial $0 = 0$: for the two-level uniform state $\mu \equiv \frac{1}{2}$ on $\mathrm{Fin}\,2$, the identity dynamics, and the identity cell map, the resolution-1 correlation matrix has von Neumann entropy $\log 2 > 0$. (The library records this as an executable positivity certificate, so the equality genuinely relates a positive quantum entropy to a positive classical join entropy.)

Theorem 14.31 (Per-partition equality of entropy rates)**.** *For each projection partition* projPartition c*, the CNT partition entropy in the abelian corner equals the classical Kolmogorov–Sinai partition entropy of the corresponding measurable partition* $c^{-1}(\cdot)$*:*

$$h_{\mathrm{adPerm}\,\sigma}(\rho_\mu, \mathrm{projPartition}\,c) \;=\; h_\sigma(\mu, c^{-1}(\cdot)).$$

Proof. Both sides are the subadditive limit of the same sequence divided by n: the quantum partition rate is $\inf_n S(\mathrm{corrMatrix}\,n)/n$ and the classical partition entropy is the corresponding limit of $H(\bigvee_{l<n} \sigma^{-l} c^{-1})/n$. The per-resolution collapse (Theorem 14.30) identifies the two defining sequences term by term, so the limits agree. □

Definition 14.32 (Abelian-corner CNT entropy)**.** The *abelian-corner CNT dynamical entropy* of adPerm σ in the state ρ_μ is the supremum of the partition rate over all *projection* operational partitions, $h^{\mathrm{ab}}_{\mathrm{adPerm}\,\sigma}(\rho_\mu) = \sup_{k,c} h_{\mathrm{adPerm}\,\sigma}(\rho_\mu, \mathrm{projPartition}\,c)$.

Theorem 14.33 (Abelian corner = Kolmogorov–Sinai entropy)**.** *Suppose every state carries positive mass ($\mu_i > 0$ for all i). Then the abelian-corner CNT dynamical entropy equals the classical Kolmogorov–Sinai entropy of the permutation system:*

$$h^{\mathrm{ab}}_{\mathrm{adPerm}\,\sigma}(\rho_\mu) \;=\; h_\mu(\sigma) \;=\; \texttt{ksEntropy}.$$

Proof. Two inequalities. For $\le$, each projection partition's rate equals a classical partition entropy (Theorem 14.31), which is dominated by the classical KS entropy (a supremum over all measurable partitions); take the supremum. For $\ge$, positivity of every μ_i forces the cells of any measurable partition P to be genuinely disjoint (not merely a.e.), so P is realized by an honest cell map c with $c^{-1}(\cdot) = P$; its projection partition then has the same rate as P, and this rate is one of the terms of the abelian supremum. Antisymmetry gives the equality. □

Theorem 14.34 (Full CNT entropy dominates KS entropy)**.** *Under the same positivity hypothesis, the* full *CNT dynamical entropy of* adPerm σ *in the state* ρ_μ *— the supremum over* all *operational partitions — dominates the classical Kolmogorov–Sinai entropy:*

$$h_\mu(\sigma) \;\le\; h_{\mathrm{adPerm}\,\sigma}(\rho_\mu).$$

Proof. The abelian dynamical entropy is a supremum over the sub-family of projection partitions, hence is $\le$ the full CNT dynamical entropy taken over all operational partitions. Rewriting the left endpoint by the corner equality (Theorem 14.33) turns this into the claimed bound. □