Transfer Operator Reading Group Notes

Prof. Tiago Pereira's Group

December 2020 – June 2021

These notes accompany the online reading group on Transfer Operators for Professor Tiago Pereira's group at ICMC-USP. The main reference will be Omri Sarig's 2012 lecture notes [Sar12] *Introduction to the transfer operator method* for Second Brazilian School on Dynamical Systems in São Carlos, SP, Brazil, October 2012.

Zheng will record in this document the contents of each meeting, but every group member is strongly encouraged to read and edit.

Contents

0 2020.12.16 Meeting 0: Introduction by Tiago **4**
- 0.1 Invariant Measures . . . 4
- 0.2 Absolutely Continuous Invariant Measures . . . 5

1 2021.1.6 Meeting 1: Climenhaga Notes **9**
- 1.1 Invariance and Identical Distribution . . . 9
- 1.2 Ergodicity, Birkhoff and SLLN . . . 10
- 1.3 Mixing and Decay of Correlations . . . 10
- 1.4 Examples of Piecewise Expanding Interval Maps . . . 11
- 1.5 Transfer Operator . . . 13
- 1.6 Decay of Correlations for Doubling Map . . . 15
- 1.7 Spetral Gap . . . 16

2 2021.1.13 Meeting 2: Sarig L1, Transfer Operator; Definition, Basic Properties & Examples **18**
- 2.1 Transfer Operator, Definition . . . 18
- 2.2 Transfer Operator, Examples . . . 20
- 2.3 Dynamical Interpretations of Behaviors of $\widehat{T}$ . . . 21

3 2021.1.20 Meeting 3: Sarig L2, Quasi-compactness & Spectral Gap **24**
- 3.1 Spectral Gap . . . 25
- 3.2 Quasi-compactness . . . 28

4 2021.1.27 Meeting 4: Sarig L2 continued, Application of Hennion's Theorem in Continued Fractions **31**
- 4.1 Sufficient Conditions for Quasi-compactness; Hennion's Theorem . . . 31
- 4.2 Application to Continued Fractions . . . 31

5 2021.2.3 – 3.3 Meetings 5 – 8: Sarig A3, Hennion's Theorem **37**
- 5.1 Reduction to $k = 1$ Case. . . . 37
- 5.2 Conditional Closure & Riesz Lemmas . . . 38
- 5.3 Step I . . . 39
- 5.4 Step II . . . 42
- 5.5 Step III . . . 43
- 5.6 Step IV . . . 43
- 5.7 Step V . . . 44

5.8 Step VI . . . 45
5.9 Step VII . . . 46

6 Gabriel: Exercise 1.5.4 **47**
6.0.1 Characterizations of Weak Mixing . . . 47
6.0.2 The Exercise . . . 51

7 Edmilson: space of Lipschitz functions **53**

8 2020.3.10 – 24 Meetings 9–11: Sarig L3, Analytic Perturbation Theory **57**
8.1 Calculus in Banach Spaces . . . 57
8.1.1 Riemann Integral and Riemann Sums . . . 57
8.1.2 Line Integral for $\mathbb{C}$-valued Functions . . . 58
8.1.3 Line Integral for B-valued Functions . . . 61
8.1.4 Differentiation . . . 61
8.2 The Resolvent and Eigenprojections . . . 64
8.3 Analytic Perturbation Theorem . . . 69

9 Herbert: Conditional Expectation **72**

10 Herbert: Kac's Lemma **74**

11 Sarig L4, Application to Central Limit Theorem **77**
11.1 Some Probability Theory . . . 77
11.1.1 Berry-Esseen Smoothing Inequality . . . 78
11.2 Nagaev's Method . . . 78
11.3 First Derivative λ_z' at $z = 0$ . . . 79
11.4 Second Derivative λ_z'' at $z = 0$ . . . 80
11.5 Limit of Characteristic Functions $\varphi_{\frac{\psi_n}{\sqrt{n}}}(t)$ . . . 81
11.6 Positivity of σ . . . 82

12 Gabriel – Sarig A3 Mixing and exactness for the Gauss map **84**

13 Future Exercises **89**

A Lebesgue-Radon-Nikodym Theorem; a Special Case **90**

B Hahn-Banach Theorem and Consequences **93**

C Open Mapping Theorem **96**

Provisional Schedule for Exercise Presentations

17 Feb 2021	**Gabriel**	Sarig pp6 Exercise 1.5.4.
24 Feb & 3 Mar 2021	**Edmilson**	Check $(\mathrm{Lip}, \|\cdot\|_{\mathrm{Lip}})$ is a Banach space; Prove $\|\hat{\psi}\|_\infty \leqslant \lvert\hat{\psi}\rvert_{\mathrm{Lip}}$ for complex Lipschitz observables $\hat{\psi} : [0,1] \to \mathbb{C}$ with $\int \hat{\psi} \mathrm{dLeb} = 0$; Discuss how norm dominance in the subspace is related to the ideas of compact embeddings
10 Mar 2021	**Herbert**	Sarig pp25–26 Appendix A.1 Conditional Expectations; Prove Kac's Lemma; Mention ergodic theory in terms of conditional expectation with respect to the σ-algebra of sets invariant under the transformation.
17 Mar 2021	**Hans**	Sarig pp10 Exercise 2.5.
30 June 2021	**Gabriel**	Sarig pp27–28 Appendix A.2 Mixing and Exactness for the Gauss Map
TBD	**Edson**	BV: the space of functions with bounded variation; [discuss with Tiago], cf. Viana's *Stochastic Dynamics for Deterministic Systems* Chapter 3

0 2020.12.16 Meeting 0: Introduction by Tiago

0.1 Invariant Measures

Given measurable transformation $T : X \to X$ on measurable space $(X, \mathcal{B})$, our interest will be on invariant measures μ for T, that is, $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.

Why the interest in invariant measures?

1. Invariant measures are weighted averages of ergodic invariant measures (ergodic decomposition), which capture the long-term behavior of a "typical" trajectory (Birkhoff: time average limits to space average a.e.)

   ergodic μ may not be informative if $\mu = \delta_a$, where $T(a) = a$, because $\{a, a, \cdots\}$ is the only μ-typical trajectory. To say its time average limits to space average is to say

   $$\phi(a) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \phi \circ T^n(a) = \int_X \phi \mathrm{d}\delta_a = \phi(a).$$

   But we knew this long-term behavior already, as a is always fixed.

   other times, e.g., when μ is equivalent to Lebesgue measure on X, then we may interpret ergodicity to mean that space average coincides with time average for practically all initial conditions.

2. more reasons?

Why is invariant measure defined via pre-images $\mu \circ T^{-1} = \mu$, as opposed to forward-images $\mu \circ T = \mu$?

1. Trivially, when T is measurably invertible, the two candidate notions of invariance are equivalent.

2. Less trivially, $T_*\mu = \mu \circ T^{-1}$ is the natural way for T to evolve μ, whereas $\nu = \mu \circ T$ is not. Indeed, invariance means that μ is fixed by the dynamics on the space of measures on $(X, \mathcal{B})$ induced by measurable transformation T. To define invariance, we first need to specify this induced dynamics by T.

   Let μ be a given measure on $(X, \mathcal{B})$; this is interpreted as an initial distribution of masses on X. We explore the natural/possible ways for T to "evolve" μ, that is, what becomes of the distribution μ after each point $x \in X$ has evolved into $T(x)$.

   First, $T_*\mu = \mu \circ T^{-1}$ is a natural way to evolve μ. To measure any event $A \in \mathcal{B}$ one time unit into the future, we take

   $$T_*\mu(A) := \mu(T^{-1}(A)),$$

   where $T^{-1}(A)$ is measurable by measurability of T. The forward direction of time is naturally understood as forward/positive iterates of T, so that, relative to the present A, $T^{-1}(A)$ speaks of the past while $T(A)$ of the future. In this way, the past $\mu(T^{-1}(A))$ informs/determines the present $T_*\mu(A)$.

   Now let us try to evolve μ a different way, namely, into a measure ν on $(X, \mathcal{B})$ defined by

   $$\nu(A) := \mu(T(A)), \qquad \forall A \in \mathcal{B}.$$

   (a) problem of **direction of time**: the future $\mu(T(A))$ informs/determines the present $\nu(A)$; this is unnatural in that it violates the flow of time.

   (b) problem of **measurability**: the measurability of T does not guarantee that $T(A)$ is measurable for all measurable sets A.

   (c) problem of **well-definedness**: if there are two disjoint sets $A_1, A_2 \in \mathcal{B}$ with $T(A_1) = T(A_2) = A$, then

   $$\nu(A_1) + \mu(A_2) = \nu(A_1 \cup A_2) = \mu(T(A_1 \cup A_2)) = \mu(A) = \nu(A_i), \qquad i = 1, 2,$$

   which implies $\mu(A) = \nu(A_1) = \nu(A_2) = 0$. In order that this implication does not lead to contradictions, we must require the pair (T, μ) be such that

   $$A_1, A_2 \in \mathcal{B} \text{ with } A_1 \cap A_2 = \varnothing, T(A_1) = T(A_2) \quad \Rightarrow \quad \mu(T(A_i)) = 0.$$

This happens when T is injective. If T is merely injective but not surjective, then T^{-1} is not defined on the entire space X. However, when T is bijective, by taking $S = T^{-1}$, we reduce to $\nu = S_*\mu$, so this way of evolution simply runs time backwards.

What about when T is not injective, but (T, μ) just happens to satisfy the above requirement?

A question related to the previous one is why a measurable transformation is defined via pre-images $T^{-1}(\mathcal{B}) \subseteq \mathcal{B}$, as opposed to forward-images $T(\mathcal{B}) \subseteq \mathcal{B}$?

An easy (circular) answer is that we need this notion of measurability to properly define $T_*\mu$.

Proposition 0.1 (Characterization of Invariance via Observables)**.** *Let $(X, \mathcal{B}, \mu)$ be a measure space and $T : X \to X$ a measurable transformation. Then, μ is T-invariant if and only if*

$$\int_X \phi \circ T \mathrm{d}\mu = \int_X \phi \mathrm{d}\mu, \qquad \forall \phi \in L^1(\mu).$$

Proof. $(\Rightarrow)$ By invariance, we have

$$\int_X \mathbb{1}_A \circ T \mathrm{d}\mu = \int_X \mathbb{1}_{T^{-1}(A)} \mathrm{d}\mu = \mu(T^{-1}(A)) = \mu(A) = \int_X \mathbb{1}_A \mathrm{d}\mu, \qquad \forall A \in \mathcal{B}.$$

Linearity generalizes this to all simple functions ϕ, and Dominated Convergence generalizes to all $\phi \in L^1(\mu)$.

$(\Leftarrow)$ By taking $\phi = \mathbb{1}_A \in L^1(\mu)$, we have

$$\mu(T^{-1}(A)) = \int_X \mathbb{1}_{T^{-1}(A)} \mathrm{d}\mu = \int_X \mathbb{1}_A \circ T \mathrm{d}\mu = \int_X \mathbb{1}_A \mathrm{d}\mu = \mu(A), \qquad \forall A \in \mathcal{B}.$$

This completes the proof. □

0.2 Absolutely Continuous Invariant Measures

Not all invariant measures are informative. Suppose T has a fixed point $x_0 = T(x_0) \in X$. Then, the point measure δ_{x_0} is T-invariant

$$T_*\delta_{x_0} = \delta_{T(x_0)} = \delta_{x_0}.$$

This only repeats the given information that x_0 is fixed by T, only in the language of the induced dynamics of T_* on the space of measures.

More generally, for any periodic orbit $\mathbf{x} = \{x_0 = T^p(x_0), \cdots, x_{p-1}\}$ of T of period $p \geqslant 1$, the average of point measures on each point in the periodic orbit

$$\delta_{\mathbf{x}} := \frac{\delta_{x_0} + \cdots + \delta_{x_{p-1}}}{p}$$

is an (uninformative) T-invariant probability measure.

An interesting dynamical system (e.g. Bernoulli maps on the circle, Anosov diffeomorphisms, the Horseshoe) tends to have many periodic orbits, which give rise to many uninformative invariant measures. It is therefore natural to restrict our attention to the "meaningful" or "informative" invariant measures.

When X is an open set in $\mathbb{R}^d$ (or more generally, a Riemannian manifold), invariant measures which are absolutely continuous with respect to the Lebsgue (Riemannian volume) measure on X tend to be meaningful, in the following sense.

A pointwise property P_B (e.g. long-term statistics of a trajectory) is understood to be observable (in physical and numerically simulated experiments) if it holds for every point in a set B of positive volume $\mathrm{Leb}(B) > 0$. Absolute continuity of invariant measure $\mu \ll \mathrm{Leb}$ implies

$$\mathrm{Leb}(B) > 0 \quad \text{for any } B \text{ with } \mu(B) > 0;$$

in other words, positivity of $\mu(B)$ guarantees observability of property P_B. In this sense, absolutely continuous invariant measures are meaningful and informative. A famous example is the Liouville measure for Hamiltonian systems.

Moreover, with the additional restriction of absolute continuity, one may hope to, in some cases, establish existence and uniqueness of an acip (absolutely continuous invariant probability measure), and then arrive at further dynamical insights.

There is a natural way to construct absolutely continuous measures. Let $f \in L^1(\text{Leb})$. Then, the finite measure μ_f given by

$$\mu_f(A) := \int_A f \mathrm{d}\text{Leb}.$$

is absolutely continuous with respect to Leb. This construction can also be reversed.

Theorem 0.2 (Radon-Nikodym, [LM94] Theorem 2.2.1)**.** *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space and ν a finite measure on $(X, \mathcal{A})$ with $\nu \ll \mu$. Then, there is a μ-essentially unique $f \in L^1(\mu)$ with $f \geqslant 0$ such that*

$$\nu(A) = \int_A f \mathrm{d}\mu, \qquad \forall A \in \mathcal{A}.$$

We write $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and call it the Radon-Nikodym derivative of ν with respect to μ.

We now turn the discussion on invariance of absolutely continuous measures into the context of Radon-Nikodym derivatives. Consider $f \in L^1(\text{Leb})$ and μ_f its induced a.c. measure. By Characterization of Invariance via Observables Proposition 0.1, T-invariance of measure μ_f reduces to

$$\int_X \phi \circ T \mathrm{d}\mu_f = \int_X \phi \mathrm{d}\mu_f, \qquad \forall \phi \in L^1(\mu_f).$$

By change of variables formula,

$$\int_X (\phi \circ T) f \mathrm{d}\text{Leb} = \int_X \phi \circ T \mathrm{d}\mu_f = \int_X \phi \mathrm{d}T_* \mu_f = \int_X \phi \frac{\mathrm{d}T_*\mu_f}{\mathrm{d}\text{Leb}} \mathrm{d}\text{Leb},$$

where

$$\widehat{T} f := \frac{\mathrm{d}T_*\mu_f}{\mathrm{d}\text{Leb}}$$

denotes the Radon-Nikodym derivative. For its existence, we need to ensure $T_*\mu_f \ll \text{Leb}$; this will be guaranteed by requiring the transformation T to be "nonsingular". More details will be given in the next lecture. Under appropriate hypotheses, the *Transfer Operator* $\widehat{T}$ is defined for all $f \in L^1(\text{Leb})$ and returns $\widehat{T}(f) \in L^1(\text{Leb})$.

Briefly recapped, the nonsingular transformation $T : X \to X$ induces a new dynamics on the space of (finite a.c.) measures on X given by the push-forward T_*, which in turn induces the dynamics $\widehat{T}$ on the space $L^1(\text{Leb})$ of Radon-Nikodym derivatives (for signed real measures).

We illustrate how $f \mapsto \widehat{T} f$ is related to the dynamics of T.

Example 0.3 (Bernoulli Map)**.** Consider the circle $S^1 = \mathbb{R}/\mathbb{Z}$ and the Bernoulli map

$$T : S^1 \to S^1, \qquad x \mapsto 2x \mod 1.$$

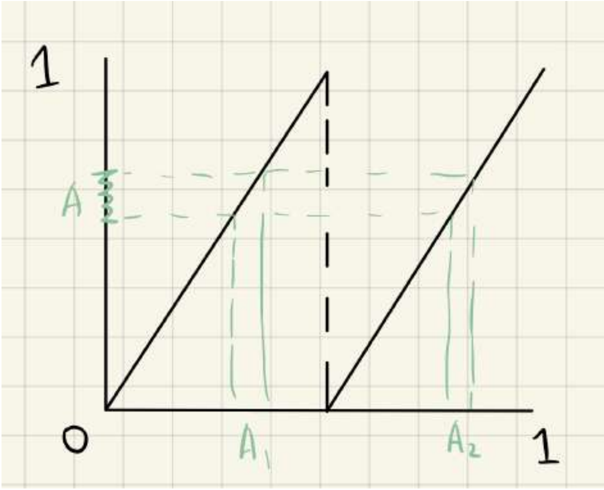


Figure 1: The inverse image of an interval A in the circle under the Bernoulli map is the union $A_1 \cup A_2$ of two intervals of half length.

We claim Leb is invariant for T. It is easy to check invariance for intervals: if A is an interval on the circle, then

$$\mathrm{Leb}(T^{-1}(A)) = \mathrm{Leb}(A_1 + A_2) = \frac{1}{2}\mathrm{Leb}(A) + \frac{1}{2}\mathrm{Leb}(A) = \mathrm{Leb}(A).$$

Since the intervals generate the Borel σ-algebra on the circle, T-invariance of Leb follows.

Note that the invariance of Leb is equivalent to

$$\widehat{T}1 = 1.$$

Hence, any constant function is invariant under $\widehat{T}$.

To calculate $\widehat{T}f$ for a general $f \in L^1(\mathrm{Leb})$, take observable $\phi : S^1 \to \mathbb{R}$.

$$\begin{aligned}\int_{S^1} (\phi \circ T) f \mathrm{dLeb} &= \int_0^1 \phi(2x \mod 1) f(x) \mathrm{dLeb}(x) = \int_0^{1/2} \phi(2x) f(x) \mathrm{d}(x) + \int_{1/2}^1 \phi(2x-1) f(x) \mathrm{d}(x) \\ &= \int_0^1 \phi(y) f\left(\frac{y}{2}\right) \frac{1}{2} \mathrm{d}y + \int_0^1 \phi(y) f\left(\frac{y+1}{2}\right) \frac{1}{2} \mathrm{d}y = \int_0^1 \phi(y) \frac{f\left(\frac{y}{2}\right) + f\left(\frac{y+1}{2}\right)}{2} \mathrm{d}y.\end{aligned}$$

where we change variables $y = 2x$ in the first integral and $y = 2x - 1$ in the second integral. This shows that

$$\widehat{T}f(y) = \frac{f\left(\frac{y}{2}\right) + f\left(\frac{y+1}{2}\right)}{2}.$$

When $f \in L^1(\mathrm{Leb})$ is a Lipschitz function with Lipschitz constant $K > 0$, we investigate the Lipschitz constant of $\widehat{T}f$. For any $x, y \in S^1$, we have

$$\begin{aligned}|\widehat{T}f(x) - \widehat{T}f(y)| &= \left| \frac{f\left(\frac{x}{2}\right) + f\left(\frac{x+1}{2}\right)}{2} - \frac{f\left(\frac{y}{2}\right) + f\left(\frac{y+1}{2}\right)}{2} \right| = \left| \frac{f\left(\frac{x}{2}\right) - f\left(\frac{y}{2}\right)}{2} + \frac{f\left(\frac{x+1}{2}\right) - f\left(\frac{y+1}{2}\right)}{2} \right| \\ &\leqslant \left| \frac{f\left(\frac{x}{2}\right) - f\left(\frac{y}{2}\right)}{2} \right| + \left| \frac{f\left(\frac{x+1}{2}\right) - f\left(\frac{y+1}{2}\right)}{2} \right| \leqslant \frac{K}{4} d(x,y) + \frac{K}{4} d(x+1, y+1) = \frac{K}{2} d(x,y).\end{aligned}$$

This shows $\widehat{T}f \in \mathrm{Lip}_{K/2}$, and iterates into

$$\widehat{T}^n f \in \mathrm{Lip}_{K/2^n}, \quad \forall n \in \mathbb{N}.$$

So $\widehat{T}$ takes a K-Lipschitz f into Lipschitz functions $\widehat{T}^n f$ with Lipschitz constants shrinking exponentially fast to 0. This is a good indication that the operator $\widehat{T}$ "squashes" all Lipschitz functions into constant functions, which are invariant under $\widehat{T}$.

In the next lecture, we will give a precise argument by showing that $\widehat{T}$ contracts on a subspace of the space of Lipschitz functions, appropriately normed.

Tiago remarks that this contraction argument scheme, together with Banach Fixed Point Theorem, is a very general one in Transfer Operator methods for constructing invariant Radon-Nikodym derivatives, and thereby invariant a.c. probability measures. The magic (underrated component) of this scheme is to find the appropriate subspace H of Radon-Nikodym derivatives for $\widehat{T}$ to act on, and to norm it appropriately so that $\widehat{T}$ contracts thereon, while ensuring **a certain domination of norms**, in this case:

$$\|f\|_\infty \leqslant \mathrm{Lip}(f),$$

where $\mathrm{Lip}(f)$ is the best Lipschitz constant of f.

1 2021.1.6 Meeting 1: Climenhaga Notes

In this lecture, we hope to define the transfer operator for a nonsingular transformation, finish Climenhaga's lecture notes, making rigorous the contraction argument on the space of Lipschitz functions and understanding the magic-ness of the domination of norms. For a continuation of Climenhaga's notes, follow the posts here: `https://vaughnclimenhaga.wordpress.com/2013/01/30/spectral-methods-in-dynamics/`.

In this lecture, we follow Vaughn Climenhaga's brief lecture notes [Cli13] on *Spectral methods in dynamics*, in order to deepen and sharpen the intuition of the subject which Tiago introduced last time.

Let X be a compact metric space; in this lecture, it is okay to assume X is the compact unit interval $[0, 1]$. And let $T : X \to X$ be a continuous (or at least piecewise continuous) transformation. Here, T specifies the dynamical system with states in X, and is taken to be "chaotic", in the sense that two nearby states will rapidly be driven far apart by the dynamics, e.g., doubling map on the circle $x \mapsto 2x \mod 1$.

A measurable function $\varphi : X \to \mathbb{C}$ is called an *observable*, and represents an observation or measurement of the dynamical system (T, X) made at time 0. Measurements made at future times $k \geqslant 0$ are given by the time series

$$\{\varphi \circ T^k\}_{k \geqslant 0}.$$

When X is also a probability space, then these are random variables and characterize the statistical properties of dynamical system (T, X). Our central interest is to investigate whether or not these random variables are independent or uncorrelated, so as to conclude statistical results.

For independent and identically distributed (iid) random variables, e.g., a fair coin flip, there are various statistical results, including the Strong Law of Large Numbers (SLLN) and Central Limit Theorem (CLT).

Theorem 1.1 (Strong Law of Large Numbers; [Dur13] Theorem 2.4.1)**.** *Let $X_1, X_2, \cdots$ be pairwise independent and identically distributed random variables with $\mathbb{E}[|X_i|] < +\infty$. Then,*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n \to +\infty]{a.s.} \mu := \mathbb{E}[X_i].$$

Theorem 1.2 (Central Limit Theorem; [Dur13] Theorem 3.4.1)**.** *Let $X_1, X_2, \cdots$ be a sequence of iid random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 \in (0, +\infty)$. Then,*

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow[n \to +\infty]{\text{in distribution}} N(0, 1).$$

In many cases, it turns out that both can hold for our random variables $\{\varphi \circ T^k\}$ of interest, even though they are not really iid because of the strong correlation between φ and $\varphi \circ T^k$ for small times k. When this correlation decays as $k \to +\infty$, it is reasonable to ask if SLLN and CLT hold.

1.1 Invariance and Identical Distribution

Let μ be a Borel probability measure on X. We say μ is T*-invariant* if

$$\mu \circ T^{-1}(A) = \mu(A), \qquad \forall \text{ Borel set } A \subseteq X.$$

If we interpret $\mu(A)$ as the probability of event $x \in A$ occurring at time 0, then $\mu(T^{-1}(A))$ denotes the probability of event $x \in T^{-1}(A)$ occurring at time 0, or equivalently, $T(x) \in A$, that is, x lands in A after 1 unit of time. Therefore, invariance amounts to the condition that the probability of event A at time 0 ($x \in A$) is the same as it is at any future time $k \geqslant 0$ ($x \in T^{-k}(A)$, or equivalently, $T^k(x) \in A$).

Invariance of the probability measure μ on state space X ensures that the sequence of random variables

$$\varphi \circ T^k : X \to \mathbb{C}$$

are identically distributed, sharing the common distribution

$$(\varphi \circ T^k)_* \mu = \mu \circ (\varphi \circ T^k)^{-1} = \mu \circ T^{-k} \circ \varphi^{-1} = \mu \circ \varphi^{-1} = \varphi_* \mu.$$

Independence, however, as mentioned before, still fails.

1.2 Ergodicity, Birkhoff and SLLN

Recall Proposition 0.1 which characterizes invariance defined in terms of Borel subsets via L^1 observables:

$$\mu \circ T^{-1} = \mu \qquad \Longleftrightarrow \qquad \int \varphi \mathrm{d}\mu = \int \varphi \circ T \mathrm{d}\mu, \quad \forall \varphi \in L^1(X, \mu).$$

In other words, invariance means that the expected value $\mathbb{E}_\mu[\varphi] = \int \varphi \mathrm{d}\mu$ of any observation φ made at time 0 is the same as if it is $\mathbb{E}[\varphi \circ T^k] = \int \varphi \circ T^k \mathrm{d}\mu$ to be made at any future time $k \geqslant 0$.

Recall also *ergodicity* of invariant probability measure μ with respect to transformation T, defined equivalently by any one of the three conditions:

1. if event A is invariant, that is, $A = T^{-1}(A)$, then A is either null or full, that is, $\mu(A) \in \{0, 1\}$.
2. if observable $\varphi \in L^1(X, \mu)$ is invariant, that is, $\varphi = \varphi \circ T$ μ–a.e., then $\varphi \equiv \text{const}$ μ–a.e.
3. μ cannot be written as a convex combination of two other invariant measures.

The ergodic thesis is that "time average equals space average".

Theorem 1.3 (Birkhoff; [Dur13] Theorem 7.2.1). *If μ is an invariant probability measure for $T : X \to X$, then for any observable $\varphi \in L^1(X, \mu)$, we have*

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ T^k(x) \xrightarrow[n \to +\infty]{\text{a.s. \& } L^1} \mathbb{E}[\varphi | \mathcal{I}],$$

where $\mathcal{I}$ denotes the σ-subalgebra consisting of almost invariant events.

The ergodic thesis is a direct consequence of this theorem applied to an ergodic invariant probability measure.

Corollary 1.4. *If μ is an ergodic invariant probability measure for $T : X \to X$, then $\mathcal{I}$ is trivial, i.e., consists of either full or null events, and hence*

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ T^k(x) \xrightarrow[n \to +\infty]{\text{a.s. \& } L^1} \mathbb{E}[\varphi | \mathcal{I}] \equiv \mathbb{E}[\mathbb{E}[\varphi | \mathcal{I}]] = \mathbb{E}[\varphi].$$

This is precisely the SLLN for our sequence of random variables $\{\varphi \circ T^k\}$.

1.3 Mixing and Decay of Correlations

To obtain CLT, we really need to worry about independence. Let us recall some definitions of independence in probability space $(X, \mathcal{B}, \mu)$.

1. two events A, B are *independent* if $\mu(A \cap B) = \mu(A)\mu(B)$;
2. two σ-subalgebras $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{B}$ are *independent* if any pair of events $B_i \in \mathcal{B}_i$, $i = 1, 2$, are independent;
3. two random variables $\varphi, \psi : X \to Y$ are *independent* if their induced σ-subalgebras $\varphi^{-1}(\mathcal{B}_Y)$ and $\psi^{-1}(\mathcal{B}_Y)$ are independent.

If the state of our invariant system (T, μ) at time k were (completely) independent of the state at time 0, then $\mathcal{B}$ and $T^{-k}(\mathcal{B})$ would be two independent σ-subalgebras, that is,

$$\mu(A \cap T^{-k}(B)) = \mu(A)\mu(T^{-k}(B)) = \mu(A)\mu(B), \quad \forall A, B \in \mathcal{B}.$$

Example 1.5 (Bernoulli Shift/ Coin Tosses). Even though it is highly unlikely that a nontrivial dynamical system at time 0 is truly independent of itself at time $k \geqslant 0$, there are cases where the observation φ at time 0 is truly independent of the same observation $\varphi \circ T^k$ at time $k \geqslant 1$. Consider the experiment of tossing a fair coin, modelled by Bernoulli shift on two symbols

$$\sigma : \Sigma_2^+ \to \Sigma_2^+, \quad \mathbb{P} = p^{\mathbb{N}}.$$

If we simply observe the outcome of the experiment at time $k \geqslant 0$, then we have iid random variables (projections down to the k-th coordinate)

$$\pi_k = \pi_0 \circ \sigma^k.$$

As we consider deterministic systems with short-term correlations, this complete independence generally fails, but we can still ask for it to hold asymptotically:

$$\lim_{k \to +\infty} \mu(A \cap T^{-k}(B)) = \mu(A)\mu(B), \quad \forall A, B \in \mathcal{B}.$$

This is the defining condition for a *mixing* measure.

In order to study the statistical behavior of system (T, μ), we need to understand the "rate of mixing", that is, the rate at which correlations decay to 0.

In the same spirit as in Proposition 0.1, the mixing property defined in terms of Borel subsets can be characterized via observables:

$$\lim_{k \to +\infty} \int (\varphi \circ T^k)\psi \mathrm{d}\mu = \int \varphi \mathrm{d}\mu \int \psi \mathrm{d}\mu, \quad \forall \varphi, \psi \in L^2(X, \mu).$$

A priori this convergence can happen arbitrarily slowly for a mixing measure μ, and it generally does. However, on a "reasonable nice" subspace of $L^2(X, \mu)$ – where the transfer operator has nice "spectral properties", this convergence happens exponentially fast. To remind ourselves of the goals now:

- find invariant measures for system $T : X \to X$, some of which are uninformative;
  - for the ergodic invariant measures, Birkhoff yields SLLN;
    - within the ergodic invariant measures, we now want to find the ones for which
      - correlation decays exponentially fast;
      - CLT holds;
      - maybe more statistical laws hold.

1.4 Examples of Piecewise Expanding Interval Maps

Example 1.6 (Doubling Map)**.** On the compact unit interval $X = [0, 1]$, consider doubling map

$$T : X \to X, \quad x \mapsto 2x \mod 1.$$

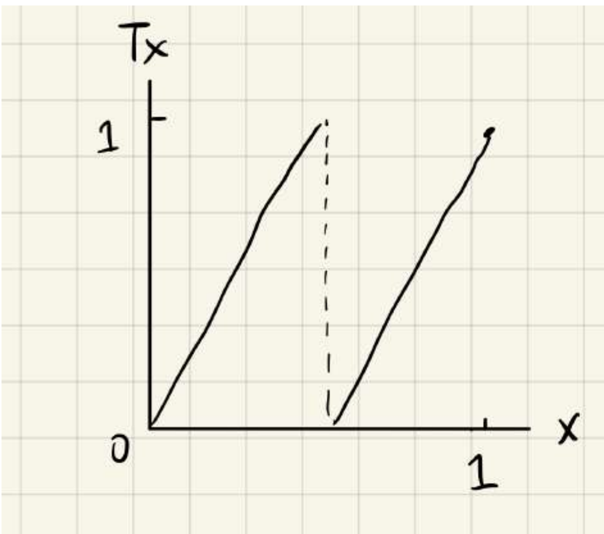


Figure 2: The doubling map $T : x \mapsto 2x \mod 1$.

1. Leb $= T_*$Leb is invariant and, in fact, ergodic;
2. the derivative $T'(x) = 2$ for all $x \in [0, 1/2) \cup (1/2, 1]$. (If we consider the doubling map on the circle instead, then $x = 1/2$ will no longer be a discontinuity and so $T' \equiv 2$.)

Example 1.7 (Piecewise Expanding Interval Map)**.** Partition the compact unit inverval $[0, 1]$ into finitely many subitervals $I_1, \cdots, I_d$. Let $T : [0, 1] \to [0, 1]$ be a map whose restriction to the interior of each subinterval I_i is C^2 and

$$|T'| \geqslant \lambda > 1, \quad \text{where differentiable.}$$

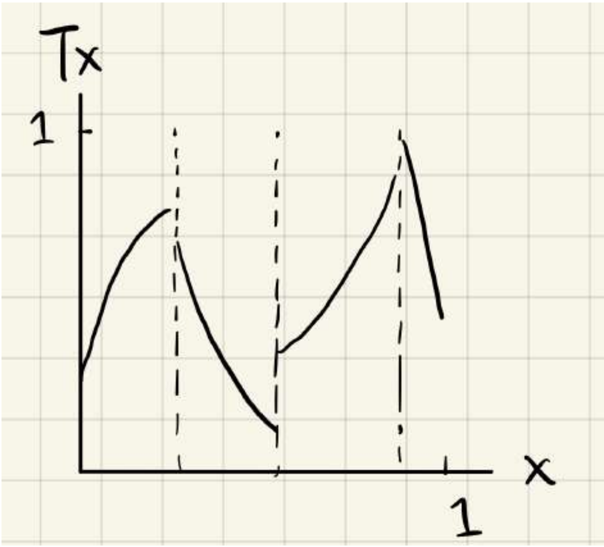


Figure 3: A piecewise expanding interval map.

The Lebesgue measure serves as a reference of "observability" in the sense that a property that holds for a set of points of positive Lebesgue measure is understood as observable in physical and simulated experiments.

As Leb is already invariant and ergodic for the doubling map, it is therefore natural to talk about statistical properties of the doubling map system with respect to Leb.

The more general piecewise expanding interval map T is chaotic in the sense that nearby points are driven far apart exponentially quickly by the expanding condition on T. But Leb is generally not invariant for T, so which invariant measure μ should we use to study its statistical properties?

A good restriction is absolute continuity ($\mu \ll$ Leb), because it ensures that a μ-a.e. result will hold for Leb-a.e. point and hence will be observable.

Let $\mathcal{M}$ denote the space of absolutely continuous probability measures on X, and $D(X, \text{Leb})$ the space of densities

$$D(X, \text{Leb}) := \{\psi \in L^1(X, \text{Leb}) : \psi \geqslant 0, \|\psi\|_{L^1} = 1\}.$$

Radon-Nikodym Theorem[1] provides a correspondence between $\mathcal{M}$ and D.

$$\mathcal{M} \to D(X, \mathrm{Leb}), \quad \mu \mapsto \frac{\mathrm{d}\mu}{\mathrm{dLeb}}$$
$$D(X, \mathrm{Leb}) \to \mathcal{M}, \quad \psi \mapsto \mu_\psi, \quad \mu_\psi(A) := \int_A \psi \mathrm{dLeb}$$

1.5 Transfer Operator

The transformation $T : X \to X$ induces push-forward dynamics on the space of (a.c. probability) measures

$$T_* : \mathcal{M} \to \mathcal{M}, \quad \mu \mapsto T_*\mu,$$

which, via the Radon-Nikodym correspondence, give rise to dynamics on the space of densities

$$D \to D, \quad \psi \mapsto \widehat{T}\psi,$$

where $\widehat{T}\psi$ is the density of a.c. probability measure $T_*\mu_\psi$. To guarantee the existence and uniqueness of $\widehat{T}\psi$, we use the Radon-Nikodym Theorem. For this, we require the nonsingularity condition on transformation T.

Definition 1.9 (Nonsingular Transformation)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. A measurable transformation $T : X \to X$ is called *nonsingular* with respect to μ if

$$\mu(A) = 0 \quad \Rightarrow \quad \mu(T^{-1}(A)) = 0, \quad \forall A \in \mathcal{A}.$$

Definition 1.10 (Transfer Operator & Koopman Operator)**.** When the transformation $T : X \to X$ is nonsingular, and $\psi \in L^1(X, \mathrm{Leb})$, then the push-forward T_* takes the a.c. signed finite measure μ_ψ into another a.c. signed finite measure $T_*\mu_\psi$, which necessarily has an essentially unique Radon-Nikodym derivative[2]

$$\widehat{T}\psi := \frac{\mathrm{d}T_*\mu_\psi}{\mathrm{dLeb}} \in L^1(X, \mathrm{Leb}).$$

Equivalently, the transfer operator may be defined as the adjoint of the *Koopman operator*

$$L^\infty(X, \mathrm{Leb}) \to L^\infty(X, \mathrm{Leb}), \quad \varphi \mapsto \varphi \circ T.$$

We remark that (i) $\varphi \circ T \in L^\infty$ for any $\varphi \in L^\infty$, provided T is nonsingular; (ii) L^∞ is the dual of L^1; and (iii) the Koopman operator is bounded and linear. Linearity is clear. To check boundedness, note, by nonsingularity of T, we have

$$\|\varphi \circ T\|_{L^\infty} = \operatorname{ess\,sup}|\varphi \circ T| \leqslant \operatorname{ess\,sup}|\varphi| = \|\varphi\|_{L^\infty}$$

[1]

Theorem 1.8 (Radon-Nikodym, [LM94] Theorem 2.2.1)**.** *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space and ν a finite measure on $(X, \mathcal{A})$ with $\nu \ll \mu$. Then, there is a μ-essentially unique $f \in L^1(\mu)$ with $f \geqslant 0$ such that*

$$\nu(A) = \int_A f \mathrm{d}\mu, \quad \forall A \in \mathcal{A}.$$

We write $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and call it the Radon-Nikodym derivative of ν with respect to μ.

[2]For the Radon-Nikodym derivative of an a.c. signed finite measure, we need a more general version of Radon-Nikodym Theorem.
Theorem 1.11 (Lebesgue-Radon-Nikodym; [Rud87] Theorem 6.10)**.** *Let μ be a real positive σ-finite measure on a measurable space $(X, \mathcal{F})$, and λ another complex measure on $(X, \mathcal{F})$.*

(a) There is then a unique pair of complex measures λ_a and λ_s on $(X, \mathcal{F})$ such that

$$\lambda = \lambda_a + \lambda_s, \quad \lambda_a \ll \mu, \quad \lambda_s \perp \mu$$

Moreover, if λ is real positive and finite, then so are λ_a and λ_s.

(b) There is a unique $h \in L^1(\mu)$ such that

$$\lambda_a(E) = \int_E h \mathrm{d}\mu, \quad \forall E \in \mathcal{F}.$$

The transfer operator

$$\widehat{T} : L^1(X, \mathrm{Leb}) \to L^1(X, \mathrm{Leb}), \qquad \psi \mapsto \widehat{T}\psi,$$

being the dual of the Koopman operator, is then defined by duality equation

$$\int (\varphi \circ T)\psi \mathrm{dLeb} = \int \varphi(\widehat{T}\psi)\mathrm{dLeb}, \qquad \forall \varphi \in L^\infty(X, \mathrm{Leb}).$$

The two definitions are equivalent by essential uniqueness.

When $T : X \mapsto X$ is a piecewise expanding map on the compact unit interval $X = [0, 1]$, then T is nonsingular, and we may write out explicitly the action of the transfer operator:

$$\widehat{T}\psi(x) = \frac{\mathrm{d}T_*\mu_\psi}{\mathrm{dLeb}} = \sum_{y \in \{T^{-1}x\}} \frac{\psi(y)}{|T'(y)|}. \tag{1}$$

Indeed, denote the restrictions of T on each subinterval by $T_i := T|_{I_i}$ and observe each T_i is monotone because $|T_i'| \geqslant \lambda > 1$. So each T_i is a diffeomorphism from I_i onto its image $T(I_i)$. By the Inverse Function Theorem, we have

$$(T_i^{-1})'(x) = \frac{1}{T'(T_i^{-1}(x))}, \qquad x \in I_i.$$

To verify the formula (1), it suffices to check

$$\int_A \sum_{y \in \{T^{-1}x\}} \frac{\psi(y)}{|T'(y)|} \mathrm{dLeb}(x) = T_*\mu_\psi(A), \qquad \forall A \in \mathcal{B}.$$

We first split the preimage $T^{-1}(A)$ into the finite disjoint union

$$T^{-1}(A) = \bigcup_i T_i^{-1}(A),$$

where $T_i^{-1} \subseteq I_i$ are disjoint, and then compute

$$\begin{aligned}
T_*\mu_\psi(A) =& \mu_\psi(T^{-1}(A)) = \int_{T^{-1}(A)} \psi(y)\mathrm{d}y = \int_{\bigcup_i T_i^{-1}(A)} \psi(y)\mathrm{d}y \\
=& \sum_i \int_{T_i^{-1}(A)} \psi(y)\mathrm{d}y \quad \text{because the finite union is disjoint} \\
=& \sum_i \int_A \psi(T_i^{-1}(x)) \left| \det \mathrm{D}T_i^{-1}|_x \right| \mathrm{d}x \quad \text{change variables } y = T_i^{-1}(x) \\
=& \int_A \sum_i \psi(T_i^{-1}(x)) \left| \frac{1}{T'(T_i^{-1}(x))} \right| \mathrm{d}x \\
=& \int_A \sum_{y \in \{T^{-1}x\}} \frac{\psi(y)}{|T'(y)|} \mathrm{d}x.
\end{aligned}$$

The boundaries of the subintervals can be ignored here because they form a null set. In fact, the formula (1) holds in general for C^1 expanding maps on a compact Riemannian manifold, cf. Chapter 11 of [VO16], and can be further extended to piecewise invertible expanding maps on a compact subset in Euclidean space, cf. [Sau98].

The transfer operator $\widehat{T}$ is useful, firstly, because it reduces the quest for an acip to that of a fixed density, i.e., a fixed point of $\widehat{T}$ in D, that is, an eigenfunction of $\widehat{T}$ with eigenvalue 1; secondly, because by iterating the duality equation, we otain

$$\int (\varphi \circ T^k)\psi \mathrm{dLeb} = \int \varphi(\widehat{T}^k\psi)\mathrm{dLeb}, \qquad \varphi \in L^\infty(X, \mathrm{Leb}), \quad \psi \in L^1(X, \mathrm{Leb}), \quad k \geqslant 0,$$

and thus the decay of correlations condition

$$\lim_{k\to+\infty}\int(\varphi\circ T^k)\psi d\mu=\lim_{k\to+\infty}\int\varphi(\widehat{T}^k\psi)\mathrm{dLeb}=\int\varphi d\mu\int\psi d\mu,\qquad\forall\varphi,\psi\in L^2(X,\mu)$$

can be understood in terms of the spectral properties of the transfer operator $\widehat{T}$ apart from its eigenvalue 1.

The (normalized) eigenfunction corresponding to the largest eigenvalue is the density of the acip, and the presence of a "spectral gap" between this eigenvalue and smaller eigenvalues leads to exponential decay of $\widehat{T}^k\psi$ when $\int\psi\mathrm{dLeb}=0$. This is how spectrum of $\widehat{T}$ informs decay of correlation and other statistical properties.

Viana in *Stochastic Dynamics for Deterministic Systems* proves the Central Limit Theorem for smooth expanding maps on manifolds and for piecewise expanding maps. His results imply that CLT holds for our case of piecewise expanding interval maps, but his proof method relies on an abstract CLT based on the martingale central limit theorem, and therefore may not be very relevant to our discussion of the Transfer Operator method.
Though he does use the Transfer Operator method to obtain decay of correlations for a different class of observables; Viana's Transfer Operator method employs the projective metric on cones of densities, as opposed to Climenhaga's Lipschitz norm.

1.6 Decay of Correlations for Doubling Map

From Tiago's introductory lecture, we know the transfer operator $\widehat{T}$ for the doubling map $T:[0,1]\to[0,1]$, $x\mapsto 2x \mod 1$ has the form

$$\widehat{T}\psi(x)=\frac{\psi\left(\frac{x}{2}\right)+\psi\left(\frac{x+1}{2}\right)}{2},\qquad\psi\in L^1([0,1],\mathrm{Leb}).$$

(Reality check: this is consistent with formula 1.)

Note

$$\widehat{T}1=1,$$

which is equivalent to the Lebesgue measure itself being invariant for T.

To prove exponential decay of correlations, we need to find a suitable Banach space $\subseteq L^1$, where $\widehat{T}$ acts with a "spectral gap".

On the space Lip of Lipschitz continuous functions $\psi:[0,1]\to\mathbb{C}$, the best Lipschitz constant

$$|\psi|_{\mathrm{Lip}}:=\sup_{x\neq y\in[0,1]}\frac{|\psi(x)-\psi(y)|}{|x-y|}$$

is a semi-norm, which fails to be a true norm only because it vanishes on all constant functions.

We improve upon $|\cdot|_{\mathrm{Lip}}$ and define a true norm $\|\cdot\|_{\mathrm{Lip}}$ on Lip by

$$\|\psi\|_{\mathrm{Lip}}:=\|\psi\|_{L^\infty}+|\psi|_{\mathrm{Lip}}.$$

The key reason for considering $|\cdot|_{\mathrm{Lip}}$ and $\|\cdot\|_{\mathrm{Lip}}$ is that $\widehat{T}$ shrinks the semi-norm $|\cdot|_{\mathrm{Lip}}$ by half:

$$|\widehat{T}\psi|_{\mathrm{Lip}}\leqslant\frac{1}{2}|\psi|_{\mathrm{Lip}},\qquad\forall\psi\in\mathrm{Lip},$$

as was proved in the last lecture.

Since every $\psi\in\mathrm{Lip}$ can be written into

$$\psi=c_\psi 1+\hat{\psi},$$

where $c_\psi=\int\psi\mathrm{dLeb}$ is the average of ψ and $\int\hat{\psi}\mathrm{dLeb}=0$, it follows that the space Lip decomposes into

$$\mathrm{Lip}=\mathbb{C}1\oplus H,\qquad H:=\{\hat{\psi}\in\mathrm{Lip}:\int\hat{\psi}\mathrm{dLeb}=0\}.$$

Note this decomposition is forward-invariant under $\widehat{T}$. Indeed, $\widehat{T}1 = 1$ implies $\widehat{T}(\mathbb{C}1) = \mathbb{C}1$; on the other hand, if $\int \hat{\psi}\text{dLeb} = 0$, then

$$\int \widehat{T}\hat{\psi}\text{dLeb} = \int (1 \circ T)\hat{\psi}\text{dLeb} = \int \hat{\psi}\text{dLeb} = 0,$$

and so $\widehat{T}(H) \subseteq H$.

Also, if $\hat{\psi} \in H$, then $\|\hat{\psi}\|_{L^\infty} \leqslant |\hat{\psi}|_{\text{Lip}}$. Indeed, the range $\hat{\psi}([0,1])$ of observable $\hat{\psi}$ has diameter

$$\text{diam}\hat{\psi}([0,1]) \leqslant |\hat{\psi}|_{\text{Lip}}\text{diam}([0,1]) = |\hat{\psi}|_{\text{Lip}},$$

and contains 0 in its closed convex hull[3], because $\int \hat{\psi}\text{dLeb} = 0$. We thus conclude

$$\|\hat{\psi}\|_{L^\infty} = \text{ess sup}|\hat{\psi}| \leqslant \text{diam}\hat{\psi}([0,1]) \leqslant |\hat{\psi}|_{\text{Lip}}, \quad \forall \hat{\psi} \in H.$$

To estimate the decay of correlations

$$\begin{aligned} C_k(\varphi, \psi) &= \int (\varphi \circ T^k)\psi\text{dLeb} - (\int \varphi\text{dLeb})(\int \psi\text{dLeb}) = \int \varphi(\widehat{T}^k\psi)\text{dLeb} - (\int \varphi\text{dLeb})c_\psi \\ &= \int \varphi\widehat{T}^k(c_\psi 1 + \hat{\psi})\text{dLeb} - c_\psi \int \varphi\text{dLeb} = \int \varphi(\widehat{T}^k\hat{\psi})\text{dLeb}, \end{aligned}$$

note

$$\|\widehat{T}^k\hat{\psi}\|_{L^\infty} \leqslant |\widehat{T}^k\hat{\psi}|_{\text{Lip}} \leqslant 2^{-k}|\hat{\psi}|_{\text{Lip}} = 2^{-k}|\psi|_{\text{Lip}}, \quad \forall \hat{\psi} \in H,$$

and hence we conclude, for $\varphi \in L^1([0,1], \text{Leb})$ and $\psi \in \text{Lip}$, the correlations decay exponentially fast

$$|C_k(\varphi, \psi)| = \left| \int \varphi(\widehat{T}^k\hat{\psi})\text{dLeb} \right| \leqslant \|\varphi\|_{L^1}\|\widehat{T}^k\hat{\psi}\|_{L^\infty} \leqslant 2^{-k}\|\varphi\|_{L^1}|\psi|_{\text{Lip}}.$$

1.7 Spetral Gap

Definition 1.12 (Spectrum). The *spectrum* of a bounded linear operator $A : B \to B$ on a Banach space B is defined to be

$$\sigma(A) := \{\lambda \in \mathbb{C} : A - \lambda\text{id is not an invertible operator on } B\}.$$

Remark 1.13.
1. the point spectrum consisting of all eigenvalues of A is contained in the spectrum $\sigma(A)$, but they are not always equal.
2. the spectrum $\sigma(A)$ is always compact and nonempty.

From previous discussion on doubling map T, we know $\widehat{T}1 = 1$ and so 1 is an eigenfunction corresponding to eigenvalue 1 of operator $\widehat{T}$.

By invariant decomposition $\text{Lip} = \mathbb{C}1 \oplus H$, we deduce

$$\sigma(\widehat{T}) = \{1\} \cup \sigma(\widehat{T}|_H).$$

In other words, apart from the eigenvalue 1, the spectrum $\sigma(\widehat{T})$ is determined by its action on H.

[3] To see this geometrically, one may view the integral of $\hat{\psi}$ against the probability measure Leb on $[0,1]$ as an average for all values $\hat{\psi}(x), x \in [0,1]$. This average is approximated by a finite mixture (convex combination) evenly spaced out on the interval, that is,

$$\mu_n := \frac{1}{n+1}\sum_{i=0}^{n} \delta_{i/n} \xrightarrow{*} \text{Leb}.$$

Hence,

$$\frac{1}{n+1}\sum_{i=0}^{n} \hat{\psi}(i/n) = \int \hat{\psi}\text{d}\mu_n \to \int \hat{\psi}\text{dLeb} = 0.$$

This gives a direct proof that $0 = \int \hat{\psi}\text{dLeb}$ is in the closed convex hull of $\hat{\psi}([0,1])$. For a more general discussion of how the integral against a probability measure compares with a convex combination, see `https://mathoverflow.net/questions/164836/is-an-integral-against-a-probability-measure-in-the-convex-hull-of-the-range`.

Definition 1.14 (Spectral Radius)**.** The *spectral radius* $\rho(A)$ of a bounded linear operator $A : B \to B$ on a Banach space B is defined to be

$$\rho(A) := \sup\{|\lambda| : \lambda \in \sigma(A)\}.$$

From functional analysis (see [Con85] Proposition 3.8), we have

$$\rho(A) = \lim_{n\to+\infty} \|A^n\|^{1/n} \leqslant \|A\|,$$

where $\|\cdot\|$ is any norm on B.

On $H \subseteq \mathrm{Lip}$, the semi-norm $|\cdot|_{\mathrm{Lip}}$ becomes a true norm, because the only constant function in H is the zero function. And norm $|\cdot|_{\mathrm{Lip}}$ is equivalent to $\|\cdot\|_{\mathrm{Lip}}$

$$|\hat{\psi}|_{\mathrm{Lip}} \leqslant \|\hat{\psi}\|_{\mathrm{Lip}} = \|\hat{\psi}\|_{L^\infty} + |\hat{\psi}|_{\mathrm{Lip}} \leqslant 2|\hat{\psi}|_{\mathrm{Lip}}, \quad \forall \hat{\psi} \in H.$$

Since the operator $\widehat{T}$ shrinks the best Lipschitz constant by half, it follows that

$$\rho(\widehat{T}|_H) \leqslant |\widehat{T}|_H|_{\mathrm{Lip}} \leqslant 1/2.$$

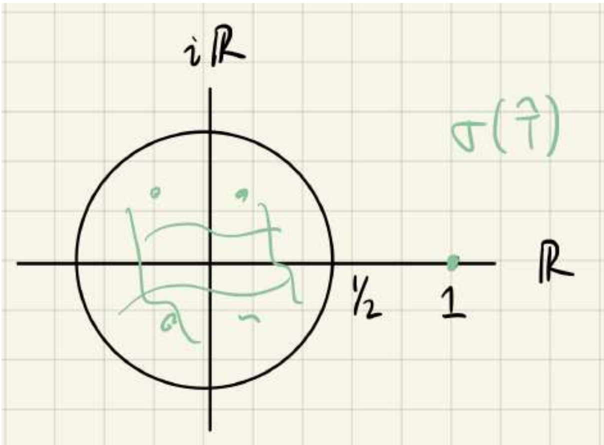


Figure 4: The spectrum $\sigma(\widehat{T})$ of the transfer operator $\widehat{T}$ for the doubling map $T : x \mapsto 2x \mod 1$ has a gap.

Definition 1.15 (Spectral Gap)**.** A bounded linear operator $A : B \to B$ on a Banach space B is said to have a *spectral gap* if

1. A has at most finitely many eigenvalues on the unit circle;
2. the rest of the spectrum $\sigma(A)$ is contained in a disk centered at 0 of radius $\rho < 1$.

In order to generalize our argument to a piecewise expanding interval map T, we need to find a suitable Banach space $B \subseteq L^1([0,1], \mathrm{Leb})$, where the transfer operator $\widehat{T}$ acts with a spectral gap. It then follows that

1. the eigenfunction(s) corresponding to eigenvalue 1 are precisely the densities for acim; there will not be any eigenvalue outside the unit disk, because $\widehat{T}$ is a Markov operator and so, in particular, $\|\widehat{T}\psi\|_{L^1} \leqslant \|\psi\|_{L^1}$ for all $\psi \in L^1$.
2. fix any $r \in (\rho, 1)$, where $D(0, \rho)$ contains the rest of the spectrum $\sigma(\widehat{T}|_B)$. Then, there is $C_r > 0$ such that

$$\|\widehat{T}^k\|_B \leqslant C_r r^k.$$

So the correlations decay exponentially fast at rate r, for any pair of observables φ, ψ chosen from suitable function spaces.

Furthermore, it will also be interesting to consider a more general class of transfer operators associated to "potential functions", for which the largest eigenvalue may not be 1.

2 2021.1.13 Meeting 2: Sarig L1, Transfer Operator; Definition, Basic Properties & Examples

For a chaotic dynamical system $T : S \to S$, such as ink diffusion in water, it is usually difficult to predict the evolution of an individual trajectory, because "chaotic" means that two nearby points are driven apart rapidly by the dynamics. However, it is often easier to study the evolution of mass densities for such systems. In the ink example, even though the trajectory of any ink particle in water is intractable, the densities of all ink particles in water will tend to become uniform.

The evolution of mass densities is given by the transfer operator.

2.1 Transfer Operator, Definition

Definition 2.1 (Nonsingular Transformation). Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space, and $T : X \to X$ a *non-singular* transformation, that is,

$$\mu(T^{-1}E) = 0 \iff \mu(E) = 0, \quad \forall E \in \mathcal{B}.$$

Remark 2.2. Note Sarig's definition of nonsingularity is stronger than that in Lasota-Mackey [LM94], where only one direction of implication $\mu(E) = 0 \Rightarrow \mu(T^{-1}E) = 0$ is required. Even the weaker version suffices to guarantee existence of transfer operator. The σ-finiteness of μ is required for Radon-Nikodym Theorem.

One starts by distributing the ink particles in water according to density $f\mathrm{d}\mu$, where $f \in L^1(\mu)$, $f \geqslant 0$, and poses the question what becomes of this density, after applying transformation T to each point $x \in X$?

The mass of points landing in E is given by

$$\begin{aligned}\int \mathbb{1}_E(Tx)f(x)\mathrm{d}\mu(x) &= \int \mathbb{1}_{T^{-1}E}\mathrm{d}\mu_f(x) \quad \text{where } \mu_f \text{ is defined by } \mu_f(E) = \int_E f\mathrm{d}\mu \\ &= \int \mathbb{1}_E \mathrm{d}\mu_f \circ T^{-1} \\ &= \int \mathbb{1}_E \frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu}\mathrm{d}\mu,\end{aligned}$$

where the Radon-Nikodym derivative $\frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu}$ exists and is unique because $\mu_f \circ T^{-1} \ll \mu$ by nonsingularity of T. Indeed, if $\mu(E) = 0$, then $\mu(T^{-1}E) = 0$ by nonsingularity of T, and hence $\mu_f \circ T^{-1}(E) = \int_{T^{-1}E} f\mathrm{d}\mu = 0$.

We may extend this procedure to all $f \in L^1(\mu)$ and obtain the definition of the transfer operator $\widehat{T}$.

Definition 2.3 (Transfer Operator). The *transfer operator* $\widehat{T}$ of a nonsingular transformation $(T, X, \mathcal{B}, \mu)$ is defined to be

$$\widehat{T} : L^1(\mu) \to L^1(\mu), \quad f \mapsto \frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu},$$

where $\frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu}$ is the Radon-Nikodym derivative of absolutely continuous signed measure $\mu_f \circ T^{-1}$ with respect to μ.

We conveniently characterize this abstract definition of the transfer operator via observables.

Proposition 2.4 (Characterization of Transfer Operator via Observables). *For any $f \in L^1(\mu)$, $\widehat{T}f$ is the unique element in $L^1(\mu)$ such that*

$$\int \varphi(\widehat{T}f)\mathrm{d}\mu = \int (\varphi \circ T) f\mathrm{d}\mu, \quad \forall \varphi \in L^\infty(\mu).$$

Proof. First we show the above duality equation holds. Let $\varphi \in L^\infty(\mu)$.

$$\begin{aligned}
\int \varphi(\widehat{T} f)\mathrm{d}\mu &= \int \varphi \frac{\mathrm{d}\mu_f \circ T^{-1}}{\mathrm{d}\mu}\mathrm{d}\mu \quad \text{by definition of } \widehat{T} \\
&= \int \varphi \mathrm{d}\mu_f \circ T^{-1} \quad \text{by definition of Radon-Nikodym derivative} \\
&= \int (\varphi \circ T)\mathrm{d}\mu_f \quad \text{change variables} \\
&= \int (\varphi \circ T) f \mathrm{d}\mu \quad \text{by definition of } \mu_f.
\end{aligned}$$

Next, we show that the duality equation characterizes $\widehat{T} f$. Suppose $h_1, h_2 \in L^1(\mu)$ both satisfy the duality equation. Let $\varphi := \mathrm{sign}(h_1 - h_2)$, which is bounded and hence L^∞. Then, for any $\varphi \in L^\infty(\mu)$, according to the Characterization of $\widehat{T}$ via Observables Proposition 2.4, we have

$$\int |h_1 - h_2|\mathrm{d}\mu = \int \varphi(h_1 - h_2)\mathrm{d}\mu = \int \varphi h_2 \mathrm{d}\mu - \int \varphi h_2 \mathrm{d}\mu = \int (\varphi \circ T) f \mathrm{d}\mu - \int (\varphi \circ T) f \mathrm{d}\mu = 0.$$

It then follows that $h_1 = h_2$ μ-a.e. This shows the duality equation uniquely determines $\widehat{T} f$ and completes the proof. ☐

Proposition 2.5 (Basic Properties of Transfer Operator)**.** *1. The transfer operator $\widehat{T}$ is a positive bounded linear operator on L^1 with induced operator norm $\|\widehat{T}\|_{L^1} = 1$.*

2. For any $f \in L^1(\mu)$ and $g \in L^\infty(\mu)$, we have $\widehat{T}((g \circ T) f) = g(\widehat{T} f)$ μ-a.e.

3. If T preserves μ, then for any $f \in L^1(\mu)$, we have $(\widehat{T} f) \circ T = \mathbb{E}_\mu[f | T^{-1}\mathcal{B}]$ μ-a.e.

Proof. **1.** "**positivity**" means if $f \in L^1$ has $f \geqslant 0$ a.e., then $\widehat{T} f \geqslant 0$ a.e. To see this, fix any $f \in L^1$ with $f \geqslant 0$ and let $\varphi := \mathbb{1}_{\{\widehat{T} f < 0\}}$. Note φ is bounded and hence L^∞. Then, we have

$$0 \geqslant \int_{\{\widehat{T} f < 0\}} \widehat{T} f \mathrm{d}\mu = \int \varphi(\widehat{T} f)\mathrm{d}\mu = \int (\varphi \circ T) f \mathrm{d}\mu \geqslant 0.$$

This shows $\int_{\{\widehat{T} f < 0\}} \widehat{T} f \mathrm{d}\mu = 0$, and hence $\mu\{\widehat{T} f < 0\} = 0$; in other words, $\widehat{T} f \geqslant 0$ a.e., as desired.

For **boundedness**, fix any $f \in L^1$ and let $\varphi := \mathrm{sign}(\widehat{T} f)$. Again, φ is bounded and hence L^∞. Then, by Hölder Inequality, we have

$$\|\widehat{T} f\|_{L^1} = \int \varphi(\widehat{T} f)\mathrm{d}\mu = \int (\varphi \circ T) f \mathrm{d}\mu \leqslant \|\varphi \circ T\|_{L^\infty} \|f\|_{L^1} = \|f\|_{L^1}.$$

It follows by definition of operator norm that $\|\widehat{T}\|_{L^1} \leqslant 1$.

For **linearity**, let $f, g \in L^1$ and $a, b \in \mathbb{C}$. Then, $af + bg \in L^1$. By Characterization of $\widehat{T}$ via Observables Proposition 2.4, for any $\varphi \in L^\infty$, we have

$$\begin{aligned}
\int \varphi(\widehat{T}(af + bg))\mathrm{d}\mu &= \int (\varphi \circ T)(af + bg)\mathrm{d}\mu = a \int (\varphi \circ T) f \mathrm{d}\mu + b \int (\varphi \circ T) g \mathrm{d}\mu \\
&= a \int \varphi(\widehat{T} f)\mathrm{d}\mu + b \int \varphi(\widehat{T} g)\mathrm{d}\mu = \int \varphi(a\widehat{T} f + b\widehat{T} g)\mathrm{d}\mu.
\end{aligned}$$

By unique determination of the duality equation from Proposition 2.4, we conclude $\widehat{T}(af + bg) = a\widehat{T} f + b\widehat{T} g$, as desired.

To see $\|\widehat{T}\|_{L^1} = 1$, it remains to check $\|\widehat{T}\|_{L^1} \geqslant 1$. For this, take any $f \in L^1$ with $f > 0$. Then,

$$\|\widehat{T} f\|_{L^1} = \int |\widehat{T} f|\mathrm{d}\mu = \int \widehat{T} f \mathrm{d}\mu = \int (1 \circ T) f \mathrm{d}\mu = \|f\|_{L^1}.$$

It follows again from the definition of operator norm (as a supremum) that $\|\widehat{T}\|_{L^1} \geqslant 1$, and hence $\|\widehat{T}\|_{L^1} = 1$, as desired. This completes the proof of **1**.

For **2**, fix any $f \in L^1$ and $g \in L^\infty$. Note $g \circ T \in L^\infty$ by nonsingularity of T, and thus $(g \circ T)f \in L^1$. Now, for any $\varphi \in L^\infty$, we have

$$\int \varphi \widehat{T}((g \circ T)f) \mathrm{d}\mu = \int (\varphi \circ T)(g \circ T) f \mathrm{d}\mu = \int ((\varphi g) \circ T) f \mathrm{d}\mu = \int (\varphi g)(\widehat{T} f) \mathrm{d}\mu.$$

It follows again from unique determination of the duality equation from Proposition 2.4 that $\widehat{T}((g \circ T)f) = g(\widehat{T} f)$ a.e.

For **3**, assume $\mu \circ T^{-1} = \mu$ and take $f \in L^1$, $T^{-1}E \in T^{-1}\mathcal{B}$. Then,

$$\begin{aligned}
\int_{T^{-1}E} (\widehat{T} f) \circ T \mathrm{d}\mu &= \int \mathbb{1}_{T^{-1}E} (\widehat{T} f) \circ T \mathrm{d}\mu = \int (\mathbb{1}_E \circ T)((\widehat{T} f) \circ T) \mathrm{d}\mu \\
&= \int (\mathbb{1}_E (\widehat{T} f)) \circ T \mathrm{d}\mu = \int \mathbb{1}_E (\widehat{T} f) \mathrm{d}T_* \mu \quad \text{change variable} \\
&= \int \mathbb{1}_E (\widehat{T} f) \mathrm{d}\mu \quad \text{by invariance } T_*\mu = \mu \\
&= \int (\mathbb{1}_E \circ T) f \mathrm{d}\mu \quad \text{by duality equation applied to } \varphi = \mathbb{1}_E \in L^\infty \\
&= \int_{T^{-1}E} f \mathrm{d}\mu.
\end{aligned}$$

We have thus verified, by definition of conditional expectation, that $(\widehat{T} f) \circ T = \mathbb{E}_\mu[f | T^{-1}\mathcal{B}]$ μ–a.e. This completes the proof of **3**. □

2.2 Transfer Operator, Examples

Example 2.6 (Doubling Map)**.** The doubling map $T : [0, 1] \to [0, 1]$, $x \mapsto 2x \mod 1$, has transfer operator given by

$$\widehat{T} f(x) = \frac{f(\frac{x}{2}) + f(\frac{x+1}{2})}{2}.$$

For details, see Lectures 0 and 1.

Example 2.7 (Gauss Map)**.** The Gauss map $T : [0, 1] \to [0, 1]$, $x \mapsto \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function, has transfer operator given by

$$\widehat{T} f(x) = \sum_{n=1}^{\infty} \frac{1}{(x+n)^2} f(\frac{1}{x+n}).$$

Indeed, fix any $\varphi \in L^\infty$ and note

$$\begin{aligned}
\int_0^1 (\varphi \circ T)(x) f(x) \mathrm{d}x &= \sum_{n=1}^{\infty} \int_{\frac{1}{n+1}}^{\frac{1}{n}} \varphi(\frac{1}{x} - n) f(x) \mathrm{d}x \\
&= \sum_{n=1}^{\infty} \int_{\frac{1}{n+1}}^{\frac{1}{n}} \varphi(y) f(\frac{1}{y+n}) \left| \frac{-1}{(y+n)^2} \right| \mathrm{d}y \quad \text{change variables } y = \frac{1}{x} - n \\
&= \int_0^1 \varphi(y) \sum_{n=1}^{\infty} \frac{1}{(y+n)^2} f(\frac{1}{y+n}) \mathrm{d}y,
\end{aligned}$$

where we have used Monotone Convergence to interchange the limit and integral. The assertion then follows by unique determination of the duality equation from Proposition 2.4.

Example 2.8 (Piecewise Monotone Interval Map)**.** Partition $[0, 1]$ into finitely many subintervals $I_1, \cdots, I_N$. Suppose $T : [0, 1] \to [0, 1]$ is such that each $T|_{I_k}$, $k = 1, \cdots, N$, is injective and has a C^1 extension with non-zero derivative to an ϵ-neighborhood of I_k. Denote by $v_k : T(I_k) \to I_k$ the inverse branch of T on subinterval

I_k, i.e., $v_k = (T|_{I_k})^{-1}$. Then, the *piecewise monotone interval map* T has transfer operator given by

$$\widehat{T}f = \sum_{k=1}^{N} \mathbb{1}_{T(I_k)} |v_k'| f \circ v_k.$$

This is only a slight generalization from piecewise expanding interval maps; another way of writing this transfer operator is simply (1).

2.3 Dynamical Interpretations of Behaviors of $\widehat{T}$

Definition 2.9 (Weak Convergence)**.** A sequence $\{f_n\} \subseteq L^1$ is said to *converge weakly* to $f \in L^1$ if

$$\int \varphi f_n \mathrm{d}\mu \to \int \varphi f \mathrm{d}\mu, \qquad \forall \varphi \in L^\infty.$$

Note weak convergence is weaker than convergence in L^1, in that convergence in L^1 implies weak convergence by Hölder Inequality.

Proposition 2.10 (Dynamical Interpretations of Convergence of $\widehat{T}^n f$)**.** *1. If $\widehat{T}^n f$ converges weakly in L^1 to $h \int f \mathrm{d}\mu$ for* ***some*** *nonzero nonnegative $f \in L^1$, then T has an acip with h being its density.*

2. If $\widehat{T}^n f$ converges weakly in L^1 to $\int f \mathrm{d}\mu$ for ***all*** *$f \in L^1$, then μ is a mixing invariant probability.*

3. If $\widehat{T}^n f$ converges (strongly) in L^1 to $\int f \mathrm{d}\mu$ for some $f \in L^1$, then for this particular f, we have

$$|\mathrm{Cov}(f, \varphi \circ T^n)| := \left| \int f(\varphi \circ T^n)\mathrm{d}\mu - \int f \mathrm{d}\mu \int \varphi \mathrm{d}\mu \right| \leqslant \|\widehat{T}^n f - \int f \mathrm{d}\mu\|_{L^1} \|\varphi\|_{L^\infty}, \qquad \forall \varphi \in L^\infty.$$

Proof. **1.** Assume without loss of generality that $\int f \mathrm{d}\mu = 1$; otherwise, take $\tilde{f} := \frac{f}{\int f \mathrm{d}\mu}$, where $\int f \mathrm{d}\mu = \|f\|_{L^1} \neq 0$ because f is nonzero and nonnegative. Now the assumption becomes $\widehat{T}^n f$ converges weakly in L^1 to h. For any $\varphi \in L^\infty$, we have

$$\begin{aligned} \int \varphi h \mathrm{d}\mu &= \lim_{n\to+\infty} \int \varphi \widehat{T}^{n+1} f \mathrm{d}\mu = \lim_{n\to+\infty} \int (\varphi \circ T)(\widehat{T}^n f)\mathrm{d}\mu \quad \text{by definition of weak convergence} \\ &= \int (\varphi \circ T) h \mathrm{d}\mu = \int \varphi (\widehat{T} h)\mathrm{d}\mu \end{aligned}$$

It then follows that $h = \widehat{T}h$ a.e. and hence

$$\mu_h \circ T^{-1}(E) = \int \mathbb{1}_{T^{-1}E} h \mathrm{d}\mu = \int (\mathbb{1}_E \circ T) h \mathrm{d}\mu = \int \mathbb{1}_E (\widehat{T}h)\mathrm{d}\mu = \int \mathbb{1}_E h \mathrm{d}\mu = \mu_h(E), \qquad \forall E \in \mathcal{B};$$

in other words, $\mu_h = T_* \mu_h$ is an a.c. invariant measure. To see it is a probability, note

$$\|\widehat{T}^n f\|_{L^1} = \int |\widehat{T}^n f| \mathrm{d}\mu = \int \widehat{T}^n f \mathrm{d}\mu = \int (1 \circ T^n) f \mathrm{d}\mu = \int f \mathrm{d}\mu = 1.$$

Hence, the weak limit h has the same norm 1; in other words, μ_h is a probability.

2. Fix any $f \in L^1$. By 1, $\mu = \mu_1$ is an invariant probability. For mixing, note for $E, F \in \mathcal{B}$, we have

$$\begin{aligned} \mu(E \cap T^{-n}F) &= \int \mathbb{1}_E \mathbb{1}_{T^{-n}F} \mathrm{d}\mu = \int \mathbb{1}_E (\mathbb{1}_F \circ T^n)\mathrm{d}\mu = \int (\widehat{T}^n \mathbb{1}_E)\mathbb{1}_F \mathrm{d}\mu \\ &\to \int (\int \mathbb{1}_E \mathrm{d}\mu)\mathbb{1}_F \mathrm{d}\mu = \mu(E)\mu(F) \quad \text{by definition of weak convergence.} \end{aligned}$$

3. Strong convergence in L^1 means $\|\widehat{T}^n f - \int f \mathrm{d}\mu\|_{L^1} \to 0$. Fix $\varphi \in L^\infty$ and note

$$\begin{aligned}\left|\int f(\varphi \circ T^n)\mathrm{d}\mu - \int f\mathrm{d}\mu \int \varphi \mathrm{d}\mu\right| &= \left|\int (\widehat{T}^n f)\varphi \mathrm{d}\mu - \int (\int f\mathrm{d}\mu)\varphi \mathrm{d}\mu\right| \\ &= \left|\int (\widehat{T}^n f - \int f\mathrm{d}\mu)\varphi \mathrm{d}\mu\right| \leqslant \|\widehat{T}^n f - \int f\mathrm{d}\mu\|_{L^1}\|\varphi\|_{L^\infty} \quad \text{by Hölder Inequality} \\ &\to 0.\end{aligned}$$

□

Exercise 1.5.1. Show all eigenvalues of the transfer operator $\widehat{T}$ have modulus less than or equal to 1.

Proof. Suppose $\widehat{T}f = \lambda f$ for some $\lambda \in \mathbb{C}$ and nonzero $f \in L^1$. Since $\|\widehat{T}f\|_{L^1} \leqslant \|f\|_{L^1}$, it follows that

$$|\lambda|\|f\|_{L^1} = \|\lambda f\|_{L^1} = \|\widehat{T}f\|_{L^1} \leqslant \|f\|_{L^1},$$

and thus $|\lambda| \leqslant 1$ because $\|f\|_{L^1} \neq 0$. □

Exercise 1.5.2. Show that

$$\{\textbf{acip densities of } T\} = \{h \in L^1(\mu) : h \geqslant 0, \quad \widehat{T}h = h, \quad \int h\mathrm{d}\mu = 1\}.$$

Proof. ($\subseteq$) If h is an acip density of T, then $h \in L^1(\mu)$, $h \geqslant 0$ and $\int h\mathrm{d}\mu = 1$. Also, from invariance $\mu_h = T_*\mu_h$, we deduce, for any $\varphi \in L^\infty$,

$$\int (\widehat{T}h)\varphi \mathrm{d}\mu = \int h(\varphi \circ T)\mathrm{d}\mu = \int \varphi \circ T \mathrm{d}\mu_h = \int \varphi \mathrm{d}T_*\mu_h = \int \varphi \mathrm{d}\mu_h = \int h\varphi \mathrm{d}\mu,$$

and therefore $\widehat{T}h = h$ a.e.

($\supseteq$) If $h \in L^1$ has $h \geqslant 0$, $\widehat{T}h = h$ and $\int h\mathrm{d}\mu = 1$, then

$$\mu_h \circ T^{-1}(E) = \int \mathbb{1}_{T^{-1}E} h\mathrm{d}\mu = \int (\mathbb{1}_E \circ T)h\mathrm{d}\mu = \int \mathbb{1}_E(\widehat{T}h)\mathrm{d}\mu = \int \mathbb{1}_E h\mathrm{d}\mu = \mu_h(E), \quad \forall E \in \mathcal{B}.$$

We thus conclude μ_h is an acip. □

Exercise 1.5.3. If $\widehat{T}$ has an acip, say μ_h, and 1 is a *simple* eigenvalue, i.e., $\dim\{g \in L^1(\mu) : \widehat{T}g = g\} = 1$, **then the acip μ_h is unique and ergodic.**

Proof. If μ_g is another acip, then by Exercise 1.5.2, we have $\widehat{T}g = g$. Since 1 is a simple eigenvalue, it follows that $g = h$ and hence μ_h is the unique acip.

To see ergodicity, suppose the contrary. Then, μ_h can be written as a nontrivial convex combination

$$\mu_h = t\nu_1 + (1-t)\nu_2,$$

where $t \in (0,1)$ and ν_1, ν_2 are two distinct invariant probabilities. Since μ_h is a.c., it follows that both ν_1, ν_2 are a.c. as well. Then, their Radon-Nikodym derivatives $\frac{\mathrm{d}\nu_1}{\mathrm{d}\mu}$ and $\frac{\mathrm{d}\nu_2}{\mathrm{d}\mu}$ are two linearly independent vectors in $\{g \in L^1(\mu) : \widehat{T}g = g\}$, contradicting the simplicity of eigenvalue 1. □

Exercise 1.5.4. If $\widehat{T}$ has an acip, say μ_h, and 1 is a *simple* eigenvalue, and all other eigenvalues have modulus strictly less than 1, then the acip μ_h is *weak mixing*, i.e.,

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}E \cap F) - \mu(E)\mu(F)| = 0, \quad \forall E, F \in \mathcal{B}.$$

Exercise 1.5.5. If μ is a mixing invariant probability, then $\widehat{T}$ has exactly one eigenvalue on the unit circle, equal to 1, and this eigenvalue is simple.

Proof. Suppose $\lambda \in \mathbb{C}$ is an eigenvalue of $\widehat{T}$. We know already that $|\lambda| \leqslant 1$, and now want to show that either $\lambda = 1$ or $|\lambda| < 1$.

Since λ is an eigenvalue, there is some nonzero $f_0 \in L^1$ with $\widehat{T} f_0 = \lambda f_0$.

It follows from the definition of mixing that

$$\int \varphi(\widehat{T}^k f)\mathrm{d}\mu = \int (\varphi \circ T^k) f \mathrm{d}\mu \xrightarrow{k\to+\infty} \int \varphi \mathrm{d}\mu \int f \mathrm{d}\mu = \int \varphi(\int f \mathrm{d}\mu)\mathrm{d}\mu, \qquad \forall \varphi \in L^\infty, f \in L^1;$$

in other words, $\widehat{T}^k f$ converges weakly in L^1 to $\int f \mathrm{d}\mu$ for any $f \in L^1$. In particular, by taking $f = f_0 \in L^1$ and $\varphi = \operatorname{sign}(f_0) \in L^\infty$, we have

$$\lambda^k \|f_0\|_{L^1} = \lambda^k \int |f_0| \mathrm{d}\mu = \int \varphi(\lambda^k f_0)\mathrm{d}\mu = \int \varphi(\widehat{T}^k f_0)\mathrm{d}\mu \xrightarrow{k\to+\infty} \int \varphi \mathrm{d}\mu \int f_0 \mathrm{d}\mu.$$

Since f_0 is nonzero, it follows that $\|f_0\|_{L^1} > 0$, and hence the convergence of sequence $\{\lambda^k \|f_0\|_{L^1}\}$ implies either $\lambda = 1$ or $|\lambda| < 1$. □

3 2021.1.20 Meeting 3: Sarig L2, Quasi-compactness & Spectral Gap

To analyze the asymptotics of $\widehat{T}^n f$ for "nice" f, we study in this section some spectral properties of operator $\widehat{T}$.

Let $(\mathcal{L}, \|\cdot\|)$ be a Banach space, i.e., $\mathcal{L}$ is a linear space with a norm $\|\cdot\|$ that induces a complete topology on $\mathcal{L}$. Let $L : \mathcal{L} \to \mathcal{L}$ a bounded linear operator; "bounded" means that the operator norm of L is finite

$$\|L\| := \sup_{0 \neq v \in \mathcal{L}} \frac{\|Lv\|}{\|v\|} < +\infty.$$

Lemma 3.1 (Bounded/Continuous Linear Operators). *A linear operator $L : \mathcal{L} \to \mathcal{L}$ on normed linear space $\mathcal{L}$ is bounded if and only if it is continuous.*

Proof. ($\Rightarrow$) Let $\{v_n\}_n \subseteq \mathcal{L}$ be a sequence converging to some $v \in \mathcal{L}$. Then,

$$\|Av_n - Av\| = \|A(v_n - v)\| \leqslant \|A\|\|v_n - v\| \to 0.$$

($\Leftarrow$) By definition of continuity, for $\epsilon = 1$ and at point $0 \in \mathcal{L}$, there is some $\delta > 0$ for which

$$\|v\| = \|v - 0\| \leqslant \delta \quad \Rightarrow \quad \|Av\| = \|Av - A0\| \leqslant 1.$$

Take any $w \in \mathcal{L}$. Then, $\left\|\frac{\delta}{\|w\|} w\right\| = \delta$, and hence $1 \geqslant \left\|A\left(\frac{\delta}{\|w\|} w\right)\right\| = \frac{\delta}{\|w\|}\|Aw\|$, which implies

$$\|Aw\| \leqslant \frac{1}{\delta}\|w\|, \quad \forall w \in \mathcal{L}.$$

We conclude $\|A\| \leqslant \frac{1}{\delta}$. This completes the proof. □

Definition 3.2 (Eigenvalue, Eigenvector, Spectrum and Spectral Radius). We say $\lambda \in \mathbb{C}$ is an *eigenvalue* of L if

$$\exists \text{ nonzero } v \in \mathcal{L} : \quad Lv = \lambda v.$$

In this case, we say the nonzero vector v is an *eigenvector* corresponding to eigenvalue λ.

We define the *spectrum* of L to be

$$\operatorname{spec}(L) := \{\lambda \in \mathbb{C} : (\lambda I - L) \text{ has no bounded inverse}\},$$

and the *spectral radius* of L to be

$$\rho(L) := \sup\{|z| : z \in \operatorname{spec}(L)\}.$$

Note any eigenvalue of L is necessarily an element of the spectrum $\operatorname{spec}(L)$, and the converse holds when the Banach space $\mathcal{L}$ is finite dimensional. If $\dim(\mathcal{L}) = +\infty$, however, there may be points in the spectrum $\operatorname{spec}(L)$ which are not eigenvalues of L. For an explicit example, see `http://www-users.math.umn.edu/~garrett/m/fun/notes_2012-13/06b_examples_spectra.pdf`.

The definition of the spectrum $\operatorname{spec}(L)$ seems to depend not only on the linear space $\mathcal{L}$ but also on the norm $\|\cdot\|$. Under a different norm $\|\cdot\|'$ on $\mathcal{L}$, would the operator L have a different spectrum? That is, could there be a value $\lambda \in \mathbb{C}$ for which $(\lambda I - L)$ has a bounded inverse with respect to norm $\|\cdot\|$ but $(\lambda I - L)$ has no bounded inverse with respect to norm $\|\cdot\|'$?

The answer is no[4], as long as the new norm $\|\cdot\|'$ still induces a complete topology on $\mathcal{L}$ and makes L remain a bounded operator, i.e., $\|L\|' < +\infty$. This is the case when $\|\cdot\|'$ is equivalent to $\|\cdot\|$. Here is why.

[4] Is there an example where changing the norm on a Banach space $\mathcal{L}$ makes the space no longer complete and/or the linear operator L no longer bounded, to the effect that the spectrum $\operatorname{spec}(L)$ also changes??? This is of course a rather pathological situation and not the kind of things people usually do to obtain a spectral gap. Usually one tries to find a smaller invariant and closed subspace $\mathcal{L}_0 \subseteq \mathcal{L}$ for L to act on. To prove spectral gap on $\mathcal{L}_0$, changing of norm is allowed (convenient or necessary) but the new norm must satisfy some domination property (e.g. equivalence to the old norm) in order to preserve the spectrum under the two norms.

Proposition 3.3. *For a bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ on a Banach space $\mathcal{L}$, we have*

$$\operatorname{spec}(L) = \{\lambda \in \mathbb{C} : (\lambda I - L) \text{ has no inverse}\}.$$

This is a consequence of the Open Mapping Theorem.

Theorem 3.4 (Open Mapping Theorem; [Con85] Chapter III Theorem 12.1)**.** *If X, Y are Banach spaces and $L : X \to Y$ is a bounded linear surjection, then L maps open sets into open sets.*

Corollary 3.5. *If $L : \mathcal{L} \to \mathcal{L}$ is an invertible bounded linear operator on Banach space $\mathcal{L}$, then its inverse L^{-1} is also a bounded linear operator.*

Proof of Proposition 3.3. Note I and L are both bounded linear operators, and therefore so is $(\lambda I - L)$. If $(\lambda I - L)$ has an inverse, then by the corollary to Open Mapping Theorem, its inverse is necessarily bounded. □

From Functional Analysis, we know

$$\rho(L) = \lim_{n \to +\infty} \sqrt[n]{\|L^n\|} = \inf_n \sqrt[n]{\|L^n\|} \leqslant \|L\|,$$

for any equivalent norm $\|\cdot\|$ on Banach space $\mathcal{L}$; cf. [Con85] Chapter VII Section 3.

In particular, $\frac{1}{n} \log \|L^n\| = \log \sqrt[n]{\|L^n\|} \downarrow \log \rho(L)$, that is, for any $\epsilon > 0$, there is $N(\epsilon)$ for which $n \geqslant N(\epsilon)$ implies $\frac{1}{n} \log \|L^n\| \leqslant \epsilon + \log \rho(L)$, or equivalently,

$$\|L^n\| \leqslant e^{n\epsilon} \rho(L)^n.$$

In other words, for any $\epsilon > 0$, we have

$$\frac{\|L^n v\|}{\|v\|} = O(e^{n\epsilon} \rho(L)^n) \quad \text{uniformly on } \mathcal{L} \backslash \{0\}. \tag{2}$$

Proposition 3.6. *If Banach space $\mathcal{L}$ has a direct sum decomposition $\mathcal{L} = \mathcal{L}_1 \oplus \mathcal{L}_2$ into two closed L-invariant linear subspaces $\mathcal{L}_1, \mathcal{L}_2$, then the spectrum $\operatorname{spec}(L)$ of a bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ can be written as*

$$\operatorname{spec}(L) = \operatorname{spec}(L|_{\mathcal{L}_1}) \cup \operatorname{spec}(L|_{\mathcal{L}_2}).$$

Proof. Since the linear subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ are closed, they are also Banach spaces, and so the spectra of restrictions $L|_{\mathcal{L}_1}$ and $L|_{\mathcal{L}_2}$ are defined. We claim any bounded linear operator $A : \mathcal{L} \to \mathcal{L}$ on a Banach space $\mathcal{L}$ has no (bounded) inverse if and only if at least one of $A|_{\mathcal{L}_i} : \mathcal{L}_i \to \mathcal{L}_i, i = 1, 2$ has no (bounded) inverse. The assertion then is a consequence of the claim applied to $A = \lambda I - L$.

($\Rightarrow$) If $A|_{\mathcal{L}_i} : \mathcal{L}_i \to \mathcal{L}_i, i = 1, 2$ both have bounded inverses $(A|_{\mathcal{L}_i})^{-1} : \mathcal{L}_i \to \mathcal{L}_i, i = 1, 2$, then the inverse of $A : \mathcal{L} \to \mathcal{L}$ is given by

$$A^{-1} v := (A|_{\mathcal{L}_1})^{-1} v_1 + (A|_{\mathcal{L}_2})^{-1} v_2, \quad v = v_1 + v_2, \quad v_i \in \mathcal{L}_i.$$

($\Leftarrow$) WOLOG suppose $A|_{\mathcal{L}_1} : \mathcal{L}_1 \to \mathcal{L}_1$ has no (bounded) inverse. Then, $Av_1 = A|_{\mathcal{L}_1} v_1 = 0$ for some $v_1 \in \mathcal{L}_1 \backslash \{0\} \subseteq \mathcal{L} \backslash \{0\}$, and hence A is not invertible, or equivalently, A has no (bounded) inverse. This proves the claim and hence the proposition. □

3.1 Spectral Gap

Definition 3.7 (Spectral Gap)**.** We say bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ on Banach space $\mathcal{L}$ has a *spectral gap* if

$$L = \lambda P + N,$$

where

1. P is a projection, i.e., idempotent $P^2 = P$. Also, $\dim(\operatorname{Im}(P)) = 1$;
2. N is a bounded linear operator on $\mathcal{L}$ with $\rho(N) < |\lambda|$;

3. $PN = NP = 0$.

Note commutativity condition 3 implies that

$$L^2 = (\lambda P + N)(\lambda P + N) = \lambda^2 P^2 + \lambda NP + \lambda PN + N^2 = \lambda^2 P + N^2,$$

and by induction,

$$L^n = \lambda^n + N^n.$$

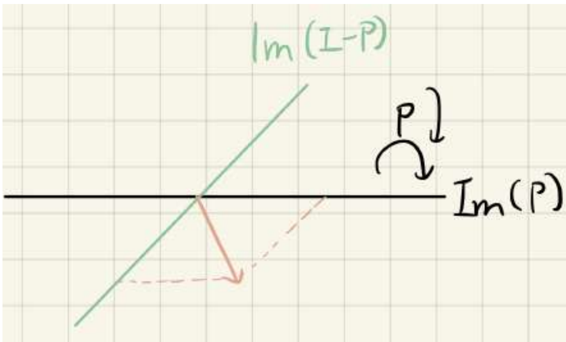
Condition 2 then yields

$$\|L^n v - \lambda^n P v\| = \|N^n v\| = o(|\lambda|^n), \quad \forall v \in \mathcal{L},$$

where we have used (2) with $\epsilon > 0$ so small that $\rho(N) + \epsilon < |\lambda|$.

Hence, if L has a spectral gap, then

$$\lambda^{-n} L^n v \xrightarrow[\text{exp.}]{n\to+\infty} Pv.$$



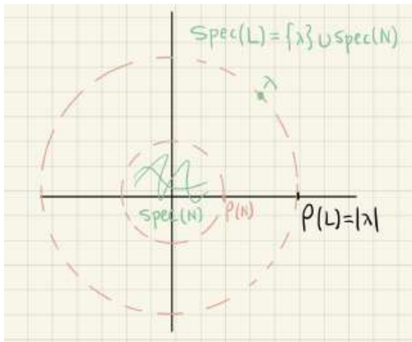


Figure 5: Left shows the one-dimensional $\mathrm{Im}(P)$ in Banach space $\mathcal{L}$. Right shows the spectral gap in the complex plane between the dominant eigenvalue λ and the rest $\mathrm{spec}(L)\backslash\{\lambda\} = \mathrm{spec}(N)$ inside a strictly smaller disk.

Proposition 3.8 (Explanation of the name "spectral gap"; Ex 2.1). *If L has a spectral gap, then λ is a simple eigenvalue and there is a "gap" $\gamma_0 > 0$ such that*

$$\mathrm{spec}(L)\backslash\{\lambda\} \subseteq \{|z| \leqslant e^{-\gamma_0}|\lambda|\}.$$

Proof. To see λ is an eigenvalue, for which every nonzero $v \in \mathrm{Im}(P)$ is a corresponding eigenvector, note

$$L(Pv) = (\lambda P + N)(Pv) = \lambda P^2 v + NPv = \lambda(Pv), \quad \forall v \in \mathcal{L}.$$

Since $\dim(\mathrm{Im}(P)) = 1$ by assumption, it follows that there are such $v \in \mathrm{Im}(P)$ and all of them are eigenvectors of L corresponding to λ.

To see λ is simple, we show every eigenvector corresponding to λ belongs to the one-dimensional subspace $\mathrm{Im}(P)$. Indeed, if $Lv = \lambda v$, then

$$\lambda^{-n} L^n v = v \xrightarrow[n\to+\infty]{} Pv,$$

which implies that $v = Pv \in \mathrm{Im}(P)$. Together with the above, we conclude simplicity of λ via

$$\{v \in \mathcal{L} : Lv = \lambda v\} = \mathrm{Im}(P).$$

Now suppose $z \in \mathbb{C}$ has $|z| > \rho(N)$ and $z \neq \lambda$.

(a) The equation $(zI - L)v = w$ has a solution $v \in \mathrm{Im}(P)$ if and only if $w \in \mathrm{Im}(P)$; in this case, $v = (z-\lambda)^{-1}w$ is the unique solution in $\mathrm{Im}(P)$.

Indeed, write $v \in \mathrm{Im}(P)$ as $v = Pv'$ for some $v' \in \mathcal{L}$, and the equation becomes

$$w = (zI - L)v = (zI - L)Pv' = zPv' - (\lambda P + N)Pv' = zPv' - \lambda Pv' = (z - \lambda)Pv' = (z - \lambda)v,$$

which has a solution if and only if $w \in \mathrm{Im}(P)$; in this case, $v = (z - \lambda)^{-1}w$ is the unique solution in $\mathrm{Im}(P)$, as desired.

(b) The same equation $(zI - L)v = w$ always has a unique solution in $\mathrm{Ker}(P)$, given by $v = (zI - N)^{-1}w$.

Indeed, for $v \in \mathrm{Ker}(P)$, the equation becomes

$$w = (zI - L)v = zv - (\lambda P + N)v = zv - \lambda Pv - Nv = (zI - N)v.$$

But $|z| > \rho(N)$ implies that $(zI - N)$ has a bounded inverse, and so $v = (zI - N)^{-1}w$ is the unique solution in $\mathrm{Ker}(P)$.

(c) For any $v \in \mathcal{L}$, we have $Pv \in \mathrm{Im}(P)$ and $(I - P)v \in \mathrm{Ker}(P)$.

Indeed, $P(I - P)v = P(v - Pv) = Pv - P^2v = 0$.

(d) $(zI - L)$ has a bounded inverse on $\mathcal{L}$, given by

$$(zI - L)^{-1} = (z - \lambda)^{-1}P + (zI - N)^{-1}(I - P).$$

Note $P = \lambda^{-1}(L - N)$ is a bounded linear operator, and $(I - P)$ is as well because

$$\|(I - P)v\| = \|v - Pv\| \leqslant \|v\| + \|Pv\| = (1 + \|P\|)\|v\|.$$

It follows that the expression above indeed defines a bounded[5] linear operator on $\mathcal{L}$.

To see it is the inverse of $(zI - L)$, we show

$$v = (z - \lambda)^{-1}Pw + (zI - N)^{-1}(I - P)w$$

is the unique solution in $\mathcal{L}$ to equation $(zI - L)v = w$.

Write $w = Pw + (I - P)w$. On one hand, $v_1 = (z - \lambda)^{-1}Pw$ is the unique solution in $\mathrm{Im}(P)$ to equation $(zI - L)v_1 = Pw$, by (a). On the other hand, $v_2 = (zI - N)^{-1}(I - P)w$ is the unique solution in $\mathcal{L}$ to equation $(zI - L)v_2 = (I - P)w$, by (b). Therefore, $v = v_1 + v_2$ solves equation

$$(zI - L)v = (zI - L)v_1 + (zI - L)v_2 = Pw + (I - P)w = w.$$

For uniqueness, suppose $v' \in \mathcal{L}$ is another solution, that is,

$$(zI - L)v' = w = (zI - L)v.$$

Applying operator P to both sides of the equality yields

$$zPv' - P(\lambda P + N)v' = P(zI - L)v' = Pw = P(zI - L)v = zPv - P(\lambda P + N)v$$

and so

$$(z - \lambda)Pv' = (z - \lambda)Pv,$$

which implies

$$Pv' = Pv.$$

Now $(zI - L)v' = (zI - L)v$ becomes $zv - \lambda Pv' - Nv' = zv - \lambda Pv - Nv$, and hence $(zI - N)v' = (zI - N)v$. By invertibility of $(zI - N)$, we conclude uniqueness $v' = v$. This completes the proof of (d).

[5]Another way to prove boundness is by **Open Mapping Theorem** [Con85] Chapter III Theorem 12.1: If X and Y are Banach spaces and $A : X \to Y$ is a continuous linear surjection, then A is an open map.
Indeed, once we prove $(zI - L)$ is invertible on $\mathcal{L}$ (it is also bounded/continuous because I and L are), it then follows from the Open Mapping Theorem that $(zI - L)$ is an open map, and hence its inverse $(zI - L)^{-1}$ is continuous/bounded.

To find γ_0, note (3d) implies that any $z \in \mathbb{C}$ with $|z| > \rho(N)$ and $z \neq \lambda$ cannot belong to the spectrum $\mathrm{spec}(L)$. In particular,

$$\mathrm{spec}(L)\backslash\{\lambda\} \subseteq \{|z| \leqslant \rho(N)\}.$$

Hence, by taking

$$0 < \gamma_0 \leqslant \log|\lambda| - \log\rho(N),$$

we have found the required gap. □

Remark 3.9. Sarig's definition of spectral gap is less intuitive than Climenhagas's, but allows any dominant eigenvalue λ, and requires that it be the unique one on the circle passing through it, and that it be simple.

In the example of the doubling map $T : x \mapsto 2x \mod 1$, we considered the action of its transfer operator $\widehat{T}$ on the space Lip of Lipschitz functions. We saw pictorially that $\widehat{T} : \mathrm{Lip} \to \mathrm{Lip}$ has a spectral gap. Now we can verify it under Sarig's definition by writing

$$\widehat{T} = 1P + N,$$

where

$$P\psi := \int \psi \mathrm{d}x, \qquad N\psi := \widehat{T}\hat{\psi} = \widehat{T}(\psi - \int \psi \mathrm{d}x).$$

1. Clearly, $P = P^2$ is idempotent and $\mathrm{Im}(P) = \mathbb{C}1$ is the one-dimensional (invariant) eigenspace of $\widehat{T} : \mathrm{Lip} \to \mathrm{Lip}$ corresponding to the dominant eigenvalue $\lambda = 1$. Also, H is invariant because

$$\int \hat{\psi} = 0 \quad \Rightarrow \quad \int \widehat{T}\hat{\psi}\mathrm{dLeb} = \int (1 \circ T)\hat{\psi}\mathrm{dLeb} = \int \hat{\psi}\mathrm{dLeb} = 0,$$

and so $\widehat{T}(H) \subseteq H$. Again, $\mathbb{C}1$ is closed because it has finite dimension 1, and $H = \mathrm{Im}(N)$ is closed for being the image of a bounded linear operator with finite codimension.

2. We showed that $|N\psi|_{\mathrm{Lip}} = |\widehat{T}\hat{\psi}|_{\mathrm{Lip}} \leqslant \frac{1}{2}|\hat{\psi}|_{\mathrm{Lip}} = \frac{1}{2}|\psi|_{\mathrm{Lip}}$. Since $|\cdot|_{\mathrm{Lip}}$ and $\|\cdot\|_{\mathrm{Lip}} = |\cdot|_{\mathrm{Lip}} + \|\cdot\|_{L^\infty}$ are equivalent norms on H, it follows that N is a bounded linear operator on H, with spectral radius

$$\rho(N) \leqslant |N|_{\mathrm{Lip}} \leqslant \frac{1}{2} < 1 = |1|.$$

3. It is easy to verify that

$$PN\psi = \int (\widehat{T}(\psi - \int \psi \mathrm{d}x))\mathrm{d}x = \int \widehat{T}\psi \mathrm{d}x - \int \psi \mathrm{d}x = \int (1 \circ T)\psi \mathrm{d}x - \int \psi \mathrm{d}x = 0, \quad \forall \psi \in \mathrm{Lip},$$

and

$$NP\psi = \widehat{T}(\int \psi \mathrm{d}x - \int (\int \psi \mathrm{d}x)\mathrm{d}x) = \widehat{T}0 = 0, \quad \forall \psi \in \mathrm{Lip}.$$

3.2 Quasi-compactness

Definition 3.10 (Quasi-compactness)**.** A bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ on a Banach space $\mathcal{L}$ is called *quasi-compact* if there are a direct sum decomposition $\mathcal{L} = F \oplus H$ and a constant $\rho \in (0, \rho(L))$ such that

1. F and H are closed and L-invariant, i.e., $L(F) \subseteq F$ and $L(H) \subseteq H$;
2. $\dim(F) < +\infty$ and every eigenvalue λ of $L|_F : F \to F$ has modulus $|\lambda| > \rho$;
3. $\rho(L|_H) < \rho$.

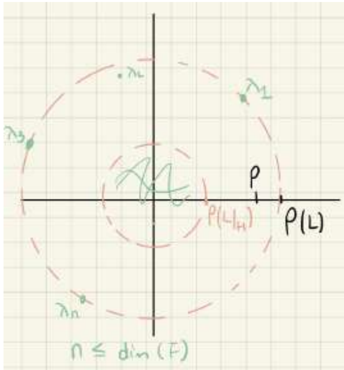


Figure 6: Quasi-compactness is weaker than the spectral gap in that it allows multiple (finitely many) eigenvalues of the same largest modulus in the complex plane.

Remark 3.11 (Quasi-compactness is weaker than spectral gap)**.** If L has a spectral gap, then L is quasi-compact. Indeed, suppose L has a spectral gap. Then, decomposition

$$\mathcal{L} = \mathrm{Im}(P) \oplus \mathrm{Im}(I - P)$$

and any constant $\rho \in (\rho(N), |\lambda|)$ satisfy the requirement in the definition of quasi-compactness.

Note any $v \in \mathcal{L}$ can be written as $v = Pv + v - Pv = Pv + (I - P)v$, where $Pv \in \mathrm{Im}(P)$ and $(I - P)v \in \mathrm{Im}(I - P)$. To see the sum is direct, suppose $v \in \mathrm{Im}(P) \cap \mathrm{Im}(I - P)$. Then, $v = Pv_1 = (I - P)v_2$ for some $v_1, v_2 \in \mathcal{L}$, and hence

$$v = Pv_1 = P^2 v_1 = P(Pv_1) = Pv = P(I - P)v_2 = Pv_2 - P^2 v_2 = 0.$$

1. $\mathrm{Im}(P)$ is closed because $\dim(\mathrm{Im}(P)) = 1$ and every finite-dimensional linear subspace is closed. It is also L-invariant because for any $Pv \in \mathrm{Im}(P)$, we have

   $$L(Pv) = (\lambda P + N)Pv = \lambda P^2 v + NPv = \lambda Pv \in \mathrm{Im}(P).$$

   $\mathrm{Im}(I - P)$ is closed because $(I - P) = (I - \lambda^{-1}(L - N))$ is a bounded linear operator and $\mathrm{Im}(I - P)$ has codimension $\mathrm{codim}(\mathrm{Im}(I - P)) = \dim(\mathrm{I} - \mathrm{P}) = 1$.[6] It is also L-invariant because for any $v - Pv \in \mathrm{Im}(I - P)$, we have

   $$L(v - Pv) = Lv - LPv = Lv - PLv = (I - P)(Lv) \in \mathrm{Im}(I - P),$$

   where we have used commutativity

   $$LP = (\lambda P + N)P = \lambda P^2 + NP = \lambda P^2 = \lambda P^2 + PN = P(\lambda P + N) = PL.$$

2. $\dim(\mathrm{Im}(P)) = 1 < +\infty$ and $L|_{\mathrm{Im}(P)}$ has exactly one eigenvalue λ, with $|\lambda| > \rho$ by construction.
3. By (3d) in Proposition 3.8, we have

   $$\mathrm{spec}(L|_{\mathrm{Im}(I-P)}) \subseteq \mathrm{spec}(L) \backslash \{\lambda\} \subseteq \{|z| \leqslant \rho(N)\},$$

   and hence

   $$\rho(L|_{\mathrm{Im}(I-P)}) \leqslant \rho(N) < \rho.$$

Hence, spectral gap is a special case of the weaker notion of quasi-compactness.

Proposition 3.12 (Quasi-compactness and Spectral Gap; Ex 2.2)**.** *If $L : \mathcal{L} \to \mathcal{L}$ is a quasi-compact linear operator on Banach space $\mathcal{L}$, L has a unique eigenvalue λ on the circle $\{z \in \mathbb{C} : |z| = \rho(L)\}$, and λ is simple, then L has a spectral gap.*

[6] According to Pietro Majer's answer `https://mathoverflow.net/q/30881`, it is a consequence of the Open Mapping Theorem that a linear subspace in a Banach space, of finite codimension, and which is the image of a Banach space via a linear bounded operator, is closed.

Proof. Let nonzero $v_1 \in \mathcal{L}$ be an eigenvector corresponding to λ. Then, $v_1 \in F$ and decompose F into

$$F = \operatorname{span}(v_1) \oplus H'',$$

where H'' is a finite-dimensional linear subspace with $L(H'') \subseteq H''$ and $\rho(L|_{H''}) < |\lambda|$ because λ is the only eigenvalue on the circle $\{z \in \mathbb{C} : |z| = \rho(L)\}$ and because λ is simple.

Put

$$H' := H'' \oplus H.$$

Then, H' is a closed linear subspace with $L(H') \subseteq H'$ and $\rho(L|_{H'}) < |\lambda|$. Define

$$P := \pi_1, \qquad N := L \circ \pi_2,$$

where $\pi_1 : \mathcal{L} \to \operatorname{span}(v_1)$ is the projection to $\operatorname{span}(v_1)$ and $\pi_2 : \mathcal{L} \to H'$ is the projection to H'.

Note any $v \in \mathcal{L}$ can be uniquely written as $v = \pi_1(v) + \pi_2(v) = av_1 + v_2$ for some scalar a and vector $v_2 \in H'$. So

$$Lv = L(av_1 + v_2) = \lambda a v_1 + Lv_2 = \lambda P v + (L \circ \pi_2)v, \qquad \forall v \in \mathcal{L}.$$

This shows

$$L = \lambda P + N.$$

Now let us check this decomposition satisfies the three conditions of spectral gap:

1. $P = \pi_1$ is a projection, that is, $P^2 = \pi_1^2 = \pi_1 = P$ and $\operatorname{Im}(P) = \operatorname{Im}(\pi_1) = \operatorname{span}(v_1)$, so $\dim(\operatorname{Im}(P)) = 1$;
2. $N = L \circ \pi_2$ is a bounded linear operator with $\|N\| = \|L \circ \pi_2\| \leqslant \|L\|$ and spectral radius

$$\rho(N) = \rho(L \circ \pi_2) = \rho(L|_{H'}) < |\lambda|.$$

3. $PNv = \pi_1(L(\pi_2 v)) = 0$ for any $v \in \mathcal{L}$ because $\pi_2 v \in H'$ implies $L(\pi_2 v) \in H'$ by invariance and hence $PNv = 0$ because the sum is direct. Similarly, $NPv = L(\pi_2(\pi_1 v)) = 0$ for any $v \in \mathcal{L}$. This shows $PN = NP = 0$.

We have shown that L has a spectral gap. □

Proposition 3.13 (Ex 2.3). *Let $\widehat{T}$ be the transfer operator of a nonsingular map $(T, X, \mathcal{B}, \mu)$. Suppose there is a linear subspace $\mathcal{L} \subseteq L^1(\mu)$ with norm $\|\cdot\|_{\mathcal{L}} \geqslant \|\cdot\|_{L^1}$ such that*

1. *$(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$ is a Banach space;*
2. *$\widehat{T}(\mathcal{L}) \subseteq \mathcal{L}$;*
3. *$\widehat{T} : \mathcal{L} \to \mathcal{L}$ is quasi-compact.*

If T has mixing acip density $h \in \mathcal{L}$, then $\widehat{T}$ has a spectral gap on $\mathcal{L}$ with $\lambda = 1$ and $Pf = h \int f \mathrm{d}\mu$.

Proof. Since T has mixing acip, it follows from Exercise 1.5.5 that $\widehat{T}$ has exactly one eigenvalue on the unit circle, equal to 1, and 1 is simple. Together with quasi-compactness of $\widehat{T}$ on $\mathcal{L}$, it follows from Proposition 3.12 that $\widehat{T}$ has a spectral gap on $\mathcal{L}$ with $\lambda = 1$.

To verify the action of projection $Pf = h \int f \mathrm{d}\mu$, we note $P^2 f = P(h \int f \mathrm{d}\mu) = h \int (h \int f \mathrm{d}\mu) \mathrm{d}\mu = h \int f \mathrm{d}\mu = Pf$ because h is a density. Also, $\operatorname{Im}(P) = \operatorname{span}(h)$ has one dimension. But this only verifies that this action of P is consistent with the requirements.

To determine the action of projection P in decomposition $\widehat{T} = \lambda P + N$, we know from requirement $\operatorname{Im}(P) = \{f \in L^1 : \widehat{T} f = f\} = \operatorname{span}(h)$ that $Pf = a(f)h$ for some scalar $a(f)$. This functional $a : L^1 \to \mathbb{C}$ must satisfy $a(h) = 1$ and must be a bounded linear functional. The integral functional $a(f) = \int f \mathrm{d}\mu$ of course satisfies this requirement, but there may be others that are also admissible. **Question: Is P unique? Or is the representation $L = \lambda P + N$ unique??**

why is it important that $\|\cdot\|_{\mathcal{L}} \geqslant \|\cdot\|_{L^1}$??? □

4 2021.1.27 Meeting 4: Sarig L2 continued, Application of Hennion's Theorem in Continued Fractions

4.1 Sufficient Conditions for Quasi-compactness; Hennion's Theorem

The transfer operator $\widehat{T}$ generally does not have a spectral gap on L^1. So we need to find a smaller $\widehat{T}$-invariant Banach subspace $\mathcal{L} \subseteq L^1$ with norm $\|\cdot\|_{\mathcal{L}} \geqslant \|\cdot\|_{L^1}$ such that $\widehat{T}|_{\mathcal{L}} : \mathcal{L} \to \mathcal{L}$ has a spectral gap. This will give information on $\widehat{T}^n f$ for $f \in \mathcal{L}$.

Theorem 4.1 (Doeblin-Fortet; Ionescu-Tulcea-Marinescu; Hennion). *Suppose $(\mathcal{L}, \|\cdot\|)$ is a Banach space and $L : \mathcal{L} \to \mathcal{L}$ is a bounded linear operator with spectral radius $\rho(L)$. Assume there exists a semi-norm $\|\cdot\|'$ on $\mathcal{L}$ such that*

1. ***Continuity:*** *$\mathcal{L} \to \mathbb{R}, v \mapsto \|v\|'$ is a continuous function;*
2. ***Precompactness:*** *for any sequence $\{f_n\} \subseteq \mathcal{L}$, if $\sup \|f_n\| \leqslant +\infty$, then there is a subsequence n_k and $g \in \mathcal{L}$ such that*
$$\|Lf_{n_k} - g\|' \xrightarrow{k \to +\infty} 0;$$
3. ***Boundedness:***
$$\exists M > 0, \forall f \in \mathcal{L} : \quad \|Lf\|' \leqslant M\|f\|';$$
4. ***Doeblin-Fortet Inequality:*** *there are $k \geqslant 1$, $r \in (0, \rho(L)), R > 0$ such that*
$$\|L^k f\| \leqslant r^k \|f\| + R\|f\|', \quad \forall f \in \mathcal{L}.$$

   Then, $L : \mathcal{L} \to \mathcal{L}$ is quasi-compact.

We will prove this theorem in the next lecture. In this lecture, we present an application to continued fractions.

4.2 Application to Continued Fractions

Every irrational number $x \in [0, 1] \backslash \mathbb{Q}$ has a unique *continued fraction* representation as

$$x = \cfrac{1}{a_1(x) + \frac{1}{a_2(x) + \cdots}}, \quad a_i(x) \in \mathbb{N}.$$

We will be interested in the asymptotic distribution of $a_n(x)$ for large n.

Theorem 4.2 (Gauss; Kuzmin; Lévy). *For every natural number N,*

$$\mathrm{Leb}\{x \in [0, 1] : a_{n+1}(x) = N\} \xrightarrow[n \to +\infty]{\text{exp.}} \frac{1}{\ln 2} \frac{\ln\left(1 + \frac{1}{N}\right)}{\ln\left(1 + \frac{1}{N+1}\right)}.$$

Proof. We will use the Gauss map

$$T : [0, 1] \to [0, 1], \quad x \mapsto \frac{1}{x} \mod 1.$$

Note for any $x \in [0, 1] \backslash \mathbb{Q}$, we have

$$Tx = \frac{1}{x} \mod 1 = a_1(x) + \frac{1}{a_2(x) + \cdots} \mod 1 = \frac{1}{a_2(x) + \cdots}$$

and induction gives

$$T^n x = \cfrac{1}{a_{n+1}(x) + \frac{1}{a_{n+2}(x) + \cdots}}.$$

It then follows that

$$a_{n+1}(x) = N \iff T^n x \in \left(\frac{1}{N+1}, \frac{1}{N}\right),$$

and therefore

$$\text{Leb}\{x \in [0,1]\backslash\mathbb{Q} : a_{n+1}(x) = N\} = \int \mathbb{1}_{(\frac{1}{N+1},\frac{1}{N})} \circ T^n(x)\mathrm{d}x = \int (\widehat{T}^n 1)\mathbb{1}_{(\frac{1}{N+1},\frac{1}{N})} = \int_{\frac{1}{N+1}}^{\frac{1}{N}} (\widehat{T}^n 1)(x)\mathrm{d}x.$$

GOAL: Find a Banach space $\mathcal{L} \subseteq L^1$ on which $\widehat{T}$ is quasi-compact and $\|\cdot\|_{\mathcal{L}} \geqslant \|\cdot\|_{L^1}$

Then, the fact that Gauss map T has a mixing acip density

$$h(x) = \frac{1}{\ln 2}\frac{1}{1+x}$$

(we will prove this fact in the next lecture), together with Exercise 2.3, yields that $\widehat{T}$ has a spectral gap on $\mathcal{L}$ with $\lambda = 1$ and $Pf = h\int f\mathrm{d}\mu$ for $f \in \mathcal{L}$. This implies

$$\widehat{T}^n 1 = \lambda^{-n}\widehat{T}^n 1 \xrightarrow[n\to+\infty]{\text{exp. in } \mathcal{L}} P1 = h.$$

But $\|\cdot\|_{\mathcal{L}} \geqslant \|\cdot\|_{L^1}$ and hence

$$\|\widehat{T}^n 1 - h\|_{L^1} \leqslant \|\widehat{T}^n 1 - h\|_{\mathcal{L}} \xrightarrow[n\to+\infty]{\text{exp.}} 0.$$

We thus conclude

$$\text{Leb}\{x \in [0,1]\backslash\mathbb{Q} : a_{n+1}(x) = N\} = \int_{\frac{1}{N+1}}^{\frac{1}{N}} (\widehat{T}^n 1)(x)\mathrm{d}x \xrightarrow[n\to+\infty]{\text{exp.}} \int_{\frac{1}{N+1}}^{\frac{1}{N}} h(x)\mathrm{d}x = \int_{\frac{1}{N+1}}^{\frac{1}{N}} \frac{1}{\ln 2}\frac{1}{1+x}\mathrm{d}x = \frac{1}{\ln 2}\frac{\ln\left(1+\frac{1}{N}\right)}{\ln\left(1+\frac{1}{N+1}\right)},$$

as required.

To accomplish the goal, we follow two steps.

Step I: Find the Banach space $\mathcal{L}$. Take linear space

$$\mathcal{L} = \{\text{Lipschitz functions } f : [0,1] \to \mathbb{C}\} \subseteq L^1,$$

normed by

$$\|f\| := \|f\|_{L^\infty} + \text{Lip}(f), \qquad \text{Lip}(f) := \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|}.$$

Then, $(\mathcal{L}, \|\cdot\|)$ is a Banach space. For this, take a Cauchy sequence $\{f_n\}_n \subseteq \mathcal{L}$, and we show it converges to some $f \in \mathcal{L}$. Note Cauchy in $\mathcal{L}$ implies Cauchy in L^∞ and hence Cauchy in L^1. Since L^1 is complete, the sequence $\{f_n\}_n$ converges in L^1-norm to some $f \in L^1$. Check completeness, namely convergence in $\|\cdot\|$-norm.

Lemma 4.3 (Exercise 2.4). *1. If $f.g \in \mathcal{L}$, then $\|f \cdot g\| \leqslant \|f\|\|g\|$;*

2. If $a \geqslant 1$, then $\left\|\frac{1}{(a+x)^2}\right\| \leqslant \frac{3}{a^2}$;

3. If $f \in \mathcal{L}$ and $a \geqslant 1$, then $\left\|f\left(\frac{1}{a+x}\right)\right\| \leqslant \|f\|$.

Proof. **1.**

$$\begin{aligned}\|f\cdot g\| =&\|f\cdot g\|_{L^\infty}+\mathrm{Lip}(f\cdot g)=\max_x|f(x)g(x)|+\sup_{x\neq y}\frac{|f(x)g(x)-f(y)g(y)|}{|x-y|}\\ \leqslant&\max_x|f(x)|\max_x|g(x)|+\sup_{x\neq y}\frac{|f(x)g(x)-f(x)g(y)|+|f(x)g(y)-f(y)g(y)|}{|x-y|}\\ \leqslant&\|f\|_{L^\infty}\|g\|_{L^\infty}+\sup_{x\neq y}\frac{|f(x)||g(x)-g(y)|}{|x-y|}+\sup_{x\neq y}\frac{|g(y)||f(x)-f(y)|}{|x-y|}\\ \leqslant&\|f\|_{L^\infty}\|g\|_{L^\infty}+\max_x|f(x)|\sup_{x\neq y}\frac{|g(x)-g(y)|}{|x-y|}+\max_y|g(y)|\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}\\ =&\|f\|_{L^\infty}\|g\|_{L^\infty}+\|f\|_{L^\infty}\mathrm{Lip}(g)+\|g\|_{L^\infty}\mathrm{Lip}(f)\\ \leqslant&\|f\|_{L^\infty}(\|g\|_{L^\infty}+\mathrm{Lip}(g))+(\|g\|_{L^\infty}+\mathrm{Lip}(g))\mathrm{Lip}(f)=\|f\|\|g\|\end{aligned}$$

2. Note $x\mapsto\frac{1}{(a+x)^2}$ is decreasing and hence realizes maximum value at $x=0$ on $[0,1]$, so $\max_x\frac{1}{(a+x)^2}=\frac{1}{a^2}$. Also, $(x,y)\mapsto\frac{|x+y+2a|}{(a+x)^2(a+y)^2}$ has no critical points on $[0,1]^2$ and hence achieves maximum at boundaries, namely, when $x=y=0$, the maximal value is $\frac{2}{a^3}$. Now we compute

$$\begin{aligned}\left\|\frac{1}{(a+x)^2}\right\|=&\max_x\frac{1}{(a+x)^2}+\sup_{x\neq y}\frac{\left|\frac{1}{(a+x)^2}-\frac{1}{(a+y)^2}\right|}{|x-y|}=\frac{1}{a^2}+\sup_{x\neq y}\frac{1}{|x-y|}\frac{|(a+y)^2-(a+x)^2|}{(a+x)^2(a+y)^2}\\ =&\frac{1}{a^2}+\sup_{x\neq y}\frac{|x+y+2a|}{(a+x)^2(a+y)^2}=\frac{1}{a^2}+\frac{2}{a^3}\leqslant\frac{3}{a^2}.\end{aligned}$$

3. Note $\mathrm{Lip}(\frac{1}{a+x})=\sup_{x\neq y}\frac{\left|\frac{1}{a+x}-\frac{1}{a+y}\right|}{|x-y|}=\sup_{x\neq y}\frac{1}{|a+x||a+y|}=\frac{1}{a^2}$ because $a\geqslant 1>0$. Now we compute

$$\begin{aligned}\left\|f\left(\frac{1}{a+x}\right)\right\|=&\sup_x\left|f\left(\frac{1}{a+x}\right)\right|+\mathrm{Lip}(f(\frac{1}{a+x}))\leqslant\sup_x|f(x)|+\mathrm{Lip}(f)\mathrm{Lip}(\frac{1}{a+x})\\ =&\|f\|_{L^\infty}+\mathrm{Lip}(f)\frac{1}{a^2}\leqslant\|f\|_{L^\infty}+\mathrm{Lip}(f)\quad\text{because } a\geqslant 1.\end{aligned}$$

This completes the proof of the three estimates in Lemma 4.3. □

Recall from Exercise 1.3 that the transfer operator $\widehat{T}$ of the Gauss map T is given by

$$\widehat{T}f(x)=\sum_{a=1}^{\infty}\frac{1}{(a+x)^2}f\left(\frac{1}{a+x}\right).$$

By Lemma 4.3, we have

$$\|\widehat{T}f\|=\left\|\sum_{a=1}^{\infty}\frac{1}{(a+x)^2}f\left(\frac{1}{a+x}\right)\right\|=\sum_{a=1}^{\infty}\left\|\frac{1}{(a+x)^2}\right\|\left\|f\left(\frac{1}{a+x}\right)\right\|\leqslant\sum_{a=1}^{\infty}\frac{3}{a^2}\|f\|.$$

This shows $\widehat{T}(\mathcal{L})\subseteq\mathcal{L}$ and $\|\widehat{T}\|\leqslant\sum_{a=1}^{\infty}\frac{3}{a^2}$; hence $\widehat{T}$ is a bounded linear operator on Banach space $(\mathcal{L},\|\cdot\|)$.

Step II: Verify conditions of Hennion's Theorem 4.1 for (semi-)norm $\|\cdot\|'=\|\cdot\|_{L^1}$ on Banach space $(\mathcal{L},\|\cdot\|)$.

1. **Continuity.** Let sequence $\{f_n\}_n\subseteq\mathcal{L}$ converge in $\|\cdot\|$-norm to some $f\in\mathcal{L}$. Then, $\|f_n-f\|_{L^\infty}\leqslant\|f_n-f\|\to 0$ and hence the sequence $\{f_n\}_n$ converges to f in L^1-norm. Therefore,

$$\|f_n\|'\equiv\|f_n\|_{L^1}\to\|f\|_{L^1}\equiv\|f\|'.$$

2. **Precompactness.** Let sequence $\{f_n\}_n\subseteq\mathcal{L}$ be such that $\sup_n\|f_n\|<+\infty$. Then, on one hand, we have $\sup_n\|f_n\|_{L^\infty}<+\infty$ and hence $\{f_n\}_n$ is uniformly bounded; on the other hand, we also have

$\sup_n \mathrm{Lip}(f_n) < +\infty$ and hence $\{f_n\}_n$ is uniformly equicontinuous. According to Arzela-Ascoli Theorem, there is a subsequence $\{f_{n_k}\}_k$ uniformly convergent to some function $f : [0, 1] \to \mathbb{C}$ with

$$\mathrm{Lip}(f) \leqslant \sup_n \mathrm{Lip}(f_n) < +\infty.$$

This shows $f \in \mathcal{L}$. Uniform convergence is equivalent to convergence in L^∞, which implies convergence in L^1, and hence

$$\|f_{n_k} - f\|_{L^1} \xrightarrow{k\to+\infty} 0.$$

Since $\widehat{T}$ is a bounded operator on L^1, it follows that

$$\|\widehat{T} f_{n_k} - \widehat{T} f\|' \equiv \|\widehat{T} f_{n_k} - \widehat{T} f\|_{L^1} \xrightarrow{k\to+\infty} 0.$$

Note $\widehat{T} f \in \mathcal{L}$ because $f \in \mathcal{L}$ and $\widehat{T}(\mathcal{L}) \subseteq \mathcal{L}$.

3. **Boundedness.**

$$\|\widehat{T} f\|' \equiv \|\widehat{T} f\|_{L^1} \leqslant \|f\|_{L^1} \equiv \|f\|'.$$

4. **Doeblin-Fortet Inequality**. We first prove a lemma.

Lemma 4.4 (Exercise 2.5). *For each $a \in \mathbb{N}$, define*

$$v_a(x) := \frac{1}{a + x}, \qquad x \in [0, 1],$$

their compositions

$$v_{a_1,\cdots,a_n} := v_{a_n} \circ \cdots \circ v_{a_1}.$$

(a) For any $f \in L^1$,

$$\widehat{T}^n f = \sum_{a_1,\cdots,a_n=1}^{\infty} |v'_{a_1,\cdots,a_n}| f \circ v_{a_1,\cdots,a_n}.$$

(b) There are constants $C > 0$ and $\theta \in (0, 1)$ such that for any string $\underline{a} := a_1 \cdots a_n$ of any length $n \geqslant 1$, we have

$$|v_{\underline{a}}(x) - v_{\underline{a}}(y)| < C\theta^n |x - y|.$$

(c) There is a constant $H > 1$ such that for any $x, y \in [0, 1]$ and any string $\underline{a} := a_1 \cdots a_n$ of any length $n \geqslant 1$, we have

$$\left| \frac{v'_{\underline{a}}(x)}{v'_{\underline{a}}(y)} - 1 \right| \leqslant H|x - y|.$$

(d) There is another constant $G > 1$ such that for any $x \in [0, 1]$ and any string $\underline{a} := a_1 \cdots a_n$ of any length $n \geqslant 1$, we have

$$G^{-1}\mathrm{Leb}(v_{\underline{a}}[0, 1)) \leqslant |v'_{\underline{a}}(x)| \leqslant G\mathrm{Leb}(v_{\underline{a}}[0, 1)).$$

(e) $v_{\underline{a}}[0, 1)$ are non-overlapping sub-intervals of $[0, 1)$.

Proof of (a). **Base Case** $n + 1$.

$$\sum_{a_1=1}^{\infty} |v'_{a_1}(x)| f(v_{a_1}(x)) = \sum_{a_1=1}^{\infty} \left| \frac{-1}{(a_1 + x)^2} \right| f\left(\frac{1}{a_1 + x} \right) = \widehat{T} f(x).$$

Assume for n and show for $n+1$.

$$\widehat{T}^{n+1}f(x) = \widehat{T}^n(\widehat{T}f)(x) = \sum_{a_1,\cdots,a_n=1}^{\infty} |v'_{a_1,\cdots,a_n}(x)|(\widehat{T}f)\circ v_{a_1,\cdots,a_n}(x)$$

$$= \sum_{a_1,\cdots,a_n=1}^{\infty} |v'_{a_n}(v_{a_1,\cdots,a_{n-1}}(x))||v'_{a_{n-1}}(v_{a_1,\cdots,a_{n-2}}(x))|\cdots|v'_{a_1}(x)|$$

$$\sum_{a_{n+1}=1}^{\infty} \frac{1}{(a_{n+1}+v_{a_1,\cdots,a_n}(x))^2} f\left(\frac{1}{a_{n+1}+v_{a_1,\cdots,a_n}(x)}\right)$$

$$= \sum_{a_1,\cdots,a_{n+1}=1}^{\infty} |v'_{a_n}(v_{a_1,\cdots,a_{n-1}}(x))||v'_{a_{n-1}}(v_{a_1,\cdots,a_{n-2}}(x))|\cdots|v'_{a_1}(x)|$$

$$|v'_{a_{n+1}}(v_{a_1,\cdots,a_n}(x))|f(v_{a_1,\cdots,a_n}(x))$$

$$= \sum_{a_1,\cdots,a_{n+1}=1}^{\infty} |v'_{a_1,\cdots,a_{n+1}}(x)|f\circ v_{a_1,\cdots,a_{n+1}}(x)$$

This completes the induction and proves point (a). □

Proof of (b). and the rest...

Now we use the lemma to verify Doeblin-Fortet Inequality. First, we estimate $\mathrm{Lip}(\widehat{T}^n f)$.

$$|(\widehat{T}^n f)(x) - (\widehat{T}^n f)(y)| = \sum_{\underline{a}} |v'_{\underline{a}}(x)|f(v_{\underline{a}}(x)) - |v'_{\underline{a}}(y)|f(v_{\underline{a}}(y))$$

$$\leqslant \sum_{\underline{a}} \big||v'_{\underline{a}}(x)|f(v_{\underline{a}}(x)) - |v'_{\underline{a}}(y)|f(v_{\underline{a}}(x))\big| + \big||v'_{\underline{a}}(y)|f(v_{\underline{a}}(x)) - |v'_{\underline{a}}(y)|f(v_{\underline{a}}(y))\big|$$

$$= \sum_{\underline{a}} \big||v'_{\underline{a}}(x)| - |v'_{\underline{a}}(y)|\big| \cdot |f(v_{\underline{a}}(x))| + |v'_{\underline{a}}(y)| \cdot |f(v_{\underline{a}}(x)) - f(v_{\underline{a}}(y))|$$

$$\leqslant \sum_{\underline{a}} |v'_{\underline{a}}(x) - v'_{\underline{a}}(y)| \cdot |f(v_{\underline{a}}(x))| + |v'_{\underline{a}}(y)| \cdot |f(v_{\underline{a}}(x)) - f(v_{\underline{a}}(y))|$$

$$\leqslant \sum_{\underline{a}} |v'_{\underline{a}}(y)| \cdot \left|\frac{v'_{\underline{a}}(x)}{v'_{\underline{a}}(y)} - 1\right| \cdot |f(v_{\underline{a}}(x))| + \|v'_{\underline{a}}\|_{L^\infty} \cdot \mathrm{Lip}(f)|v_{\underline{a}}(x) - v_{\underline{a}}(y)|$$

For any Lipschitz function $f : J \to \mathbb{C}$ on interval J, we have the diameter of its range bounded by

$$\mathrm{diam}(f(J)) \leqslant \mathrm{Lip}(f)\mathrm{Leb}(J),$$

and the mean value is realized at some point $x_0 \in J$

$$\frac{1}{\mathrm{Leb}(J)} \int_J |f(t)|\mathrm{d}t = f(x_0)\mathrm{d}t \in f(J).$$

It then follows that

$$|f(x)| \leqslant \frac{1}{\mathrm{Leb}(J)} \int_J |f(t)|\mathrm{d}t + \mathrm{Lip}(f)\mathrm{Leb}(J), \qquad \forall x \in J.$$

We continue the estimates for $\mathrm{Lip}(\widehat{T}^n f)$.

$$
\begin{aligned}
|(\widehat{T}^n f)(x) - (\widehat{T}^n f)(y)| &\leqslant \sum_{\underline{a}} |v'_{\underline{a}}(y)| \cdot \left| \frac{v'_{\underline{a}}(x)}{v'_{\underline{a}}(y)} - 1 \right| \cdot |f(v_{\underline{a}}(x))| + \|v'_{\underline{a}}\|_{L^\infty} \cdot \mathrm{Lip}(f) |v_{\underline{a}}(x) - v_{\underline{a}}(y)| \\
&\leqslant \sum_{\underline{a}} G\mathrm{Leb}(v_{\underline{a}}[0,1)) \cdot H|x-y| \cdot \left(\frac{1}{\mathrm{Leb}(v_{\underline{a}}[0,1))} \int_{v_{\underline{a}}[0,1)} |f(t)| \mathrm{d}t + \mathrm{Lip}(f) \mathrm{Leb}(v_{\underline{a}}[0,1)) \right) \\
&\qquad + G\mathrm{Leb}(v_{\underline{a}}[0,1)) \cdot \mathrm{Lip}(f) C\theta^n |x-y| \\
&= \sum_{\underline{a}} GH|x-y| \int_{v_{\underline{a}}[0,1)} |f(t)| \mathrm{d}t + \sum_{\underline{a}} G\mathrm{Leb}(v_{\underline{a}}[0,1)) \cdot H|x-y| \mathrm{Lip}(f) \mathrm{Leb}(v_{\underline{a}}[0,1)) \\
&\qquad + \sum_{\underline{a}} G\mathrm{Leb}(v_{\underline{a}}[0,1)) \cdot \mathrm{Lip}(f) C\theta^n |x-y| \\
&\leqslant \sum_{\underline{a}} GH|x-y| \int_{v_{\underline{a}}[0,1)} |f(t)| \mathrm{d}t + \sum_{\underline{a}} G\mathrm{Leb}(v_{\underline{a}}[0,1)) \cdot H|x-y| \mathrm{Lip}(f) C\theta^n \\
&\qquad + \sum_{\underline{a}} G\mathrm{Leb}(v_{\underline{a}}[0,1)) \cdot \mathrm{Lip}(f) C\theta^n |x-y|,
\end{aligned}
$$

where in the last inequality we have used

$$\mathrm{Leb}(v_{\underline{a}}[0,1)) = |v_{\underline{a}}(0) - v_{\underline{a}}(1)| < C\theta^n |0-1| = C\theta^n.$$

But $v_{\underline{a}}[0,1)$ are non-overlapping subintervals of $[0,1)$, and so $\sum_{\underline{a}} \mathrm{Leb}(v_{\underline{a}}[0,1)) \leqslant 1$. We continue.

$$
\begin{aligned}
|(\widehat{T}^n f)(x) - (\widehat{T}^n f)(y)| \leqslant& GH|x-y| \int_{\bigcup_{\underline{a}} v_{\underline{a}}[0,1)} |f(t)| \mathrm{d}t + GH|x-y| \mathrm{Lip}(f) C\theta^n + G\mathrm{Lip}(f) C\theta^n |x-y| \\
\leqslant& |x-y| \left(GH\|f\|_{L^1} + (H+1) G\mathrm{Lip}(f) C\theta^n \right).
\end{aligned}
$$

It follows that

$$\mathrm{Lip}(\widehat{T}^n f) \leqslant GH\|f\|_{L^1} + (H+1)GC\theta^n \mathrm{Lip}(f).$$

Second, we estimate $\|\widehat{T}^n f\|_{L^\infty}$. Since $\widehat{T}^n f$ is Lipschitz on interval $[0,1]$, it follows that, for any $x \in [0,1]$,

$$|\widehat{T}^n f(x)| \leqslant \frac{1}{\mathrm{Leb}[0,1]} \int_0^1 |\widehat{T}^n f(t)| \mathrm{d}t + \mathrm{Lip}(\widehat{T}^n f) \mathrm{Leb}[0,1] = \int_0^1 |\widehat{T}^n f(t)| \mathrm{d}t + \mathrm{Lip}(\widehat{T}^n f) = \|\widehat{T}^n f\|_{L^1} + \mathrm{Lip}(\widehat{T}^n f),$$

and hence

$$\|\widehat{T}^n f\|_{L^\infty} \leqslant \|\widehat{T}^n f\|_{L^1} + \mathrm{Lip}(\widehat{T}^n f) \leqslant \|f\|_{L^1} + \mathrm{Lip}(\widehat{T}^n f) \leqslant (GH+1)\|f\|_{L^1} + (H+1)GC\theta^n \mathrm{Lip}(f).$$

Putting the two estimates for $\mathrm{Lip}(\widehat{T}^n f)$ and $\|\widehat{T}^n f\|_{L^\infty}$ together, we obtain

$$\|\widehat{T}^n f\| \equiv \|\widehat{T}^n f\|_{L^\infty} + \mathrm{Lip}(\widehat{T}^n f) \leqslant (2GH+1)\|f\|' + 2(H+1)GC\theta^n \mathrm{Lip}(f).$$

By slightly increasing θ to $r \in (\theta, 1)$, we absorb the multiplicative constant

$$2(H+1)GC\theta^k \leqslant r^k$$

for any sufficiently large k.

We have verified the Doeblin-Fortet Inequality, and hence all conditions for Hennion's Theorem 4.1 are met. The proof of Theorem 4.2 is complete. □

5 2021.2.3 – 3.3 Meetings 5 – 8: Sarig A3, Hennion's Theorem

In this meeting we will start the proof of Hennion's Theorem.

Theorem 5.1 (Hennion). *Suppose $(\mathcal{L}, \|\cdot\|)$ is a Banach space and $L : \mathcal{L} \to \mathcal{L}$ is a bounded linear operator with spectral radius $\rho(L)$. Assume there exists a semi-norm $\|\cdot\|'$ on $\mathcal{L}$ such that*

1. ***Continuity:*** *$\mathcal{L} \to \mathbb{R}, v \mapsto \|v\|'$ is a continuous function;*
2. ***Precompactness:*** *for any sequence $\{f_n\} \subseteq \mathcal{L}$, if $\sup \|f_n\| \leqslant +\infty$, then there is a subsequence n_k and $g \in \mathcal{L}$ such that*
$$\|Lf_{n_k} - g\|' \xrightarrow{k \to +\infty} 0;$$
3. ***Boundedness:***
$$\exists M > 0, \forall f \in \mathcal{L} : \quad \|Lf\|' \leqslant M\|f\|';$$
4. ***Doeblin-Fortet Inequality:*** *there are $k \geqslant 1$, $r \in (0, \rho(L)), R > 0$ such that*
$$\|L^k f\| \leqslant r^k \|f\| + R\|f\|', \quad \forall f \in \mathcal{L}.$$

Then, $L : \mathcal{L} \to \mathcal{L}$ is quasi-compact.

Proof of Hennion's Theorem.

5.1 Reduction to $k = 1$ Case.

It suffices to prove the case $k = 1$. Indeed, assume $k = 1$ case holds. Take L for which the Doeblin-Fortet Inequality holds for some $k \geqslant 2$. We show this implies L is quasi-compact.

Note for bounded linear operator $\tilde{L} := L^k$ and semi-norm $\|\cdot\|'$, we easily verify the four conditions for Hennion's Theorem and so the $k = 1$ case yields that $\tilde{L}$ is quasi-compact.

Since $\tilde{L} = L^k$ is quasi-compact, it follows that (i) $\operatorname{spec}(L^k)$ contains only finitely many points $z \in \mathbb{C}$ with $|z| = \rho(L^k)$, (ii) every $z \in \operatorname{spec}(L^k)$ with $|z| = \rho(L^k)$ is an eigenvalue of L^k with finite multiplicity, and (iii) points in $\operatorname{spec}(L^k)$ do not accumulate to the circle $\{|z| = \rho(L^k)\}$.

By Spectral Mapping Theorem[7], we have

$$\operatorname{spec}(L^k) = (\operatorname{spec}(L))^k.$$

Now, (i) implies that $\operatorname{spec}(L)$ also contains finitely many points $z \in \mathbb{C}$ with $|z| = \rho(L)$. (ii) implies that every $z \in \operatorname{spec}(L)$ with $|z| = \rho(L)$ is an eigenvalue of L with finite multiplicity; indeed, take any such z. Then, $z^k \in (\operatorname{spec}(L))^k = \operatorname{spec}(L^k)$ with $|z^k| = |z|^k = (\rho(L))^k = \rho(L^k)$. According to (ii), z^k is an eigenvalue of L^k with finite multiplicity; in other words, there is some $v \in \mathcal{L}\backslash\{0\}$ such that $L^k v = z^k v$.

Fix $\rho \in (r, \rho(L))$ and define closed annulus in the complex plane

$$A(\rho, \rho(L)) := \{z \in \mathbb{C} : \rho \leqslant |z| \leqslant \rho(L)\}.$$

Lemma 5.2. *Under the hypotheses of Hennion's Theorem, for any $z \in A(\rho, \rho(L))$, we have*

(i) $K(z) := \bigcup_{\ell \geqslant 1} \ker(zI - L)^\ell$ is a finite-dimensional linear subspace, and $I(z) := \bigcap_{\ell \geqslant 1} \operatorname{Im}(zI - L)^\ell$ is a closed linear subspace.

(ii) $K(z), I(z)$ are both L-invariant and $B = K(z) \bigoplus I(z)$.

(iii) $(zI - L) : I(z) \to I(z)$ is a bijection with bounded inverse.

[7] **Spectral Mapping Theorem [Con85] Theorem VII.4.10.** If $a \in \mathcal{A}$ and $f \in \operatorname{Hol}(a)$, then
$$\operatorname{spec}(f(a)) = f(\operatorname{spec}(a)).$$
Here, $\mathcal{A}$ is the Banach algebra $B(\mathcal{L})$ and $f : z \mapsto z^k$.

(iv) $\{\lambda \in A(\rho, \rho(L)) : K(\lambda) \neq \{0\}\}$ *is finite and non-empty.*

Note (i) and (ii) imply, according to Proposition 3.6, that for any $z \in A(\rho, \rho(L))$, we have

$$\mathrm{spec}(L) = \mathrm{spec}(L|_{K(z)}) \cup \mathrm{spec}(L|_{I(z)}).$$

But (iii) implies $z \notin \mathrm{spec}(L|_{I(z)})$. If $z \in \mathrm{spec}(L) \cap A(\rho, \rho(L))$, then z must belong to $\mathrm{spec}(L|_{K(z)})$ and hence $K(z) \neq \{0\}$. This implies

$$\mathrm{spec}(L) \cap A(\rho, \rho(L)) \subseteq \{\lambda \in A(\rho, \rho(L)) : K(\lambda) \neq \{0\}\}.$$

By (iv), we conclude the intersection $\mathrm{spec}(L) \cap A(\rho, \rho(L))$ is finite, and it must be nonempty too by definition of spectral radius $\rho(L)$. Write

$$\mathrm{spec}(L) \cap A(\rho, \rho(L)) = \{\lambda_1, \cdots, \lambda_t\}.$$

If z is not an eigenvalue, then $(zI - L)$ is invertible, and so are its positive powers $(zI - L)^\ell$, $\ell \geqslant 1$. Therefore, $K(z) = \{0\}$, and so $B = I(z)$. But (iii) then implies $(zI - L)$ has a bounded inverse on $I(z) = B$, and hence $z \notin \mathrm{spec}(L) \cap A(\rho, \rho(L))$. We conclude each element λ_i in the finite nonempty intersection $\mathrm{spec}(L) \cap A(\rho, \rho(L))$ is an eigenvalue of L; moreover, each λ_i has finite geometric multiplicity by (i).

By forming

$$F := \bigoplus_{i=1}^{t} K(\lambda_i), \qquad H := \bigcap_{i=1}^{t} I(\lambda_i),$$

we will show (v) F is a direct sum, $\dim(F) < +\infty$, $L(F) \subseteq F$, and the eigenvalues of $L|_F$ are exactly $\lambda_1, \cdots, \lambda_t$; (vi) H is closed, $L(H) \subseteq H$, and $B = F \oplus H$; (vii) $\rho(L|_H) \leqslant \rho$. It will then follow by definition that L is quasi-compact.

5.2 Conditional Closure & Riesz Lemmas

To prove Lemma 5.2, we first prove the following result, which will be our main technical tool to utilize the Doeblin-Fortet Inequality.

Lemma 5.3 (Conditional Closure Lemma). *Under the hypotheses of Hennion's Theorem, fix $z \in \mathbb{C}$ with $|z| > r$, and let $\{g_n\}_n \subseteq B$ be a sequence such that each equation*

$$g_n = (zI - L)f_n \tag{3}$$

has a solution $f_n \in B$. If $\|g_n - g\| \to 0$ as $n \to +\infty$ and $\sup_n \|f_n\| < +\infty$, then the sequence $\{f_n\}_n$ has a subsequence in B converging to a solution $f \in B$ to the limiting equation

$$g = (zI - L)f.$$

Proof of Lemma 5.3. From equation (3), we have

$$g_n - g_m = (zI - L)f_n - (zI - L)f_m = z(f_n - f_m) - L(f_n - f_m), \tag{4}$$

and hence

$$|z| \cdot \|f_n - f_m\| = \|g_n - g_m + L(f_n - f_m)\| \leqslant \|g_n - g_m\| + r\|f_n - f_m\| + R\|f_n - f_m\|',$$

by Doeblin-Fortet Inequality ($k = 1$). Rearranging terms yields

$$\|f_n - f_m\| \leqslant \frac{\|g_n - g_m\| + \|f_n - f_m\|'}{|z| - r}. \tag{5}$$

First note

$$\|g_n - g_m\| \leqslant \|g_n - g\| + \|g - g_m\| \xrightarrow{m,n \to +\infty} 0.$$

To deal with $\|f_n - f_m\|'$, we start again from (4) and deduce

$$|z| \cdot \|f_n - f_m\|' \leqslant \|g_n - g_m\|' + \|Lf_n - Lf_m\|'$$

by Triangle Inequality for the semi-norm $\|\cdot\|'$. Since $\sup_n \|f_n\| < +\infty$, there is a subsequence $\{Lf_{n_k}\}_k$ such that $\|Lf_{n_k} - h\|' \xrightarrow{k\to+\infty} 0$ for some $h \in B$ by Precompactness. Thus,

$$\|f_{n_k} - f_{m_l}\|' \leqslant \frac{\|g_{n_k} - g_{m_l}\|' + \|Lf_{n_k} - h\|' + \|h - Lf_{m_l}\|'}{|z|} \xrightarrow{k,l\to+\infty} 0.$$

Plugging this into estimate (5), we obtain Cauchyness

$$\|f_{n_k} - f_{m_l}\| \xrightarrow{k,l\to+\infty} 0,$$

and so, by completeness of B, there is some $f \in B$ with

$$\|f_{n_k} - f\| \xrightarrow{k\to+\infty} 0.$$

Since $(zI - L)$ is continuous, it follows that

$$g = \lim_{k\to+\infty} g_{n_k} = \lim_{k\to+\infty} (zI - L)f_{n_k} = (zI - L) \lim_{k\to+\infty} f_{n_k} = (zI - L)f.$$

This completes the proof of Conditional Closure Lemma. ☐

Before proving Lemma 5.2, we prove another separation result for general normed vector spaces.

Lemma 5.4 (Riesz Lemma)**.** *Let $(V, \|\cdot\|)$ be a normed vector space and $U \subseteq V$ a linear subspace with $\overline{U} \neq V$. Then, for any $r \in (0, 1)$, there is $v \in V$ with $\|v\| = 1$ and $\operatorname{dist}(v, U) \geqslant t$.*

Proof of Riesz Lemma. Fix any $v_0 \in V \backslash \overline{U}$. By definition of $\operatorname{dist}(v_0, U) = \inf_{u\in U} \|v_0 - u\|$, there is some $u_0 \in U$ with

$$\operatorname{dist}(v_0, U) \leqslant \|v_0 - u_0\| \leqslant \frac{1}{t}\operatorname{dist}(v_0, U).$$

Note that for any $u \in U$, we have

$$\left\| \frac{v_0 - u_0}{\|v_0 - u_0\|} - \frac{u}{\|v_0 - u_0\|} \right\| = \frac{\|v_0 - (u_0 + u)\|}{\|v_0 - u_0\|} \geqslant \frac{\operatorname{dist}(v_0, U)}{\frac{1}{t}\operatorname{dist}(v_0, U)} = t.$$

Hence, $v := \frac{v_0 - u_0}{\|v_0 - u_0\|}$ is the desired vector. This completes the proof fo Riesz Lemma. ☐

Remark 5.5. As Tiago remarked on 2021.2.3, Riesz Lemma is commonly used to produce a sequence of vectors on the unit sphere that has no convergent subsequence, leading to non-compactness of the unit ball in an infinite-dimensional Banach space, for instance. We will see it in action many times in the proof of Hennion's Theorem.

We are now ready to prove Lemma 5.2.

5.3 Step I

Proof of Lemma 5.2. **Step I.** If $|z| > r$, then

1. $\ker(zI - L)^\ell$ is finite-dimensional for all $\ell \geqslant 1$.
2. $\operatorname{Im}(zI - L)^\ell$ is closed for all $\ell \geqslant 0$.
3. There exists $\ell \geqslant 1$ such that $K(z) = \ker(zI - L)^\ell$ and $I(z) = \operatorname{Im}(zI - L)^\ell$.

Proof of Step I. Fix $z \in \mathbb{C}$ with $|z| > r$. Set $K_\ell := \ker(zI - L)^\ell$. We induct on $\ell \geqslant 1$ to show $\dim(K_\ell) < +\infty$ for all $\ell \geqslant 1$.

Base Case $\ell = 1$**.** For a contradiction, suppose $\dim(K_1) = +\infty$. Take any $f_1 \in K_1$ with $\|f_1\| = 1$. Since $S_1 := \operatorname{span}(f_1) \subseteq K_1$ is one-dimensional, it is closed and is not the entire space K_1, Riesz Lemma applied with $t = \frac{1}{2}$ yields $f_2 \in K_1$ with $\|f_2\| = 1$ and $\operatorname{dist}(f_2, S_1) \geqslant \frac{1}{2}$. Continuing this way, we obtain a sequence $\{f_n\}_n \subseteq K_1$ such that

$$\|f_n\| = 1 \quad \text{and} \quad \|f_n - f_m\| \geqslant \frac{1}{2}, \quad \forall n \neq m.$$

Then, $\sup_n \|f_n\| = 1 < +\infty$ and $(zI - L)f_n = 0$ because $f_n \in K_1 = \ker(zI - L)$. Conditional Closure Lemma then yields a convergent subsequence of $\{f_n\}_n$, contradicting the fact that $\|f_n - f_m\| \geqslant \frac{1}{2}$ for all $n \neq m$. We thus conclude the base case that $\dim(K_1) < +\infty$.

Assume ℓ, show $\ell + 1$. Again, for a contradiction, suppose $\dim(K_{\ell+1}) = +\infty$. Then, by Riesz Lemma, we can construct a sequence $\{f_n\}_n \subseteq K_{\ell+1}$ such that

$$\|f_n\| = 1 \quad \text{and} \quad \|f_n - f_m\| \geqslant \frac{1}{2}, \quad \forall n \neq m.$$

But $f_n \in K_{\ell+1} = \ker(zI - L)^{\ell+1}$ implies that

$$g_n := (zI - L)f_n \in \ker(zI - L)^{\ell} = K_\ell,$$

and

$$\|g_n\| \leqslant |z| \cdot \|f_n\| + \|L\| \cdot \|f_n\| = |z| + \|L\|.$$

By Induction Hypothesis, $\dim(K_\ell) < +\infty$, and hence the unit ball in K_ℓ is compact; in particular, the sequence $\{g_n\}_n$ has a convergent subsequence $\{g_{n_k}\}_k$. Conditional Closure Lemma then yields a convergent subsequence $\{f_{n_{k_l}}\}_l$, contradicting the fact that $\|f_n - f_m\| \geqslant \frac{1}{2}$ for all $n \neq m$. We thus conclude the induction step that $\dim(K_{\ell+1}) < +\infty$. This completes the induction and shows $\dim(K_\ell) < +\infty$ for all $\ell \geqslant 1$.

To see $I_\ell := \mathrm{Im}(zI - L)^\ell$ is closed for all $\ell \geqslant 1$. We induct on $\ell \geqslant 0$, where $(zI - L)^0 = I$.

Base Case $\ell = 0$. Clearly $I_0 = B$ is closed.

Assume ℓ, show $\ell + 1$. Take a sequence $\{g_n\}_n \subseteq I_{\ell+1}$, and assume $g_n \xrightarrow[n\to+\infty]{B} g$ for some $g \in B$. We need to show $g \in I_{\ell+1}$.

Since $g_n \in I_{\ell+1} = (zI - L)I_\ell$, we may write

$$g_n = (zI - L)f'_n, \quad \text{for some } f'_n \in I_\ell.$$

Since K_1 is finite dimensional and I_ℓ is closed by Induction Hypothesis, it follows that the intersection $K_1 \cap I_\ell$ is also closed and finite-dimensional[8], and hence there is some $h \in K_1 \cap I_\ell$ such that

$$\|f'_n - h\| = \mathrm{dist}(f'_n, K_1 \cap I_\ell) = \min_{h' \in K_1 \cap I_\ell} \|f'_n - h'\|.$$

Take $f_n := f'_n - h$ so that

$$g_n = (zI - L)f_n \quad \text{and} \quad \|f_n\| = \mathrm{dist}(f_n, K_1 \cap I_\ell).$$

CLAIM 1: $\sup_n \|f_n\| < +\infty$. Otherwise, there exists a subsequence $\{f_{n_k}\}_k$ with $\|f_{n_k}\| \xrightarrow{k\to+\infty} +\infty$. Then we have

$$\frac{g_{n_k}}{\|f_{n_k}\|} \xrightarrow{k\to+\infty} 0,$$

because $g_{n_k} \xrightarrow{k\to+\infty} g$. Conditional Closure Lemma then yields a subsequence $\{f_{n_{k_l}}\}_l$ with

$$\frac{f_{n_{k_l}}}{\|f_{n_{k_l}}\|} \xrightarrow[l\to+\infty]{B} h, \quad \text{for some } h \in B,$$

and h solves the limiting equation

$$0 = (zI - L)h.$$

This shows $h \in K_1$. Since $f_{n_{k_l}} \in I_\ell$ and I_ℓ is closed by Induction Hypothesis, it follows that $h \in I_\ell$ and hence $h \in K_1 \cap I_\ell$. In particular, $\mathrm{dist}(h, K_1 \cap I_\ell) = 0$, contradicting $\mathrm{dist}(f_{n_k}, K_1 \cap I_\ell) = \|f_{n_k}\| \to +\infty$. This proves CLAIM 1.

---

[8] As Matheus pointed out, finite-dimensionality is important: mere closedness of a linear subspace S in a Banach space is not sufficient to guarantee that for any vector v outside S, there is a vector $s \in S$ with $\|v - s\| = \mathrm{dist}(v, S)$. For a concrete example, see **https://math.stackexchange.com/questions/296354/given-a-point-x-and-a-closed-subspace-y-of-a-normed-space-must-the-distance**.

Now that $\sup_n \|f_n\| < +\infty$, Conditional Closure Lemma yields a subsequence $\{f_{n_k}\}_k$ converging to some f in B, where f solves the limiting equation

$$g = (zI - L)f.$$

Since $f_{n_k} \in I_\ell$ and I_ℓ is closed by Induction Hypothesis, it follows that the limit $f \in I_\ell$ as well, and thus $g \in (zI - L)I_\ell \equiv I_{\ell+1}$. We conclude the induction step that $I_{\ell+1}$ is closed. This completes the induction and shows I_ℓ is closed for all $\ell \geqslant 0$.

To show $K(z) = K_\ell$ for some $\ell \geqslant 1$, we will prove the ascending sequence $K_1 \subseteq K_2 \subseteq \cdots$ eventually stabilizes. For a contradiction, suppose the opposite. Then, there are infinitely many n for which $K_{n-1} \subsetneq K_n$; collect these indices to form a strictly ascending subsequence $K_{n_{k-1}} \subsetneq K_{n_k}$ of linear subspaces in B. Riesz Lemma applied again with $t = \frac{1}{2}$ yields a sequence $f_{n_k} \in K_{n_k}$ with

$$\|f_{n_k}\| = 1 \quad \text{and} \quad \operatorname{dist}(f_{n_k}, K_{n_{k-1}}) \geqslant \frac{1}{2}.$$

In particular, the sequence $\{f_{n_k}\}_k$ is $\frac{1}{2}$-separated.

CLAIM 2: the sequence $\{L^m f_{n_k}\}_k$ is $\frac{|z|^m}{2}$-separated for all $m \geqslant 1$. Write

$$z^{-m}L^m f_{n_{k+l}} - z^{-m}L^m f_{n_k} = f_{n_{k+l}} - \left[(I - z^{-m}L^m)f_{n_{k+l}} + z^{-m}L^m f_{n_k}\right].$$

We will show the term in the square bracket belongs to $K_{n_{k+l-1}}$, which will then imply

$$\|L^m f_{n_{k+l}} - L^m f_{n_k}\| \geqslant |z|^m \cdot \operatorname{dist}(f_{n_{k+l}}, K_{n_{k+l-1}}) \geqslant \frac{|z|^m}{2}, \quad \forall k, l \geqslant 1;$$

in other words, we will then have proven CLAIM 2.

First observe $L(K_\ell) \subset K_\ell$; indeed, if $f \in K_\ell$, then

$$(zI - L)^\ell Lf = L(zI - L)^\ell f = 0.$$

This implies $L^m f_{n_k} \in K_{n_k}$ because $f_{n_k} \in K_{n_k}$.

Second observe $(zI - L)K_\ell \subseteq K_{\ell-1}$; indeed, if $(zI - L)^\ell f = 0$, then

$$(zI - L)^{\ell-1}(zI - L)f = 0.$$

Thus, together with the first observation, we obtain

$$(I - z^{-m}L^m)f_{n_{k+l}} = \sum_{j=0}^{m-1} z^{-j}L^j(I - z^{-1}L)f_{n_{k+l}} \in \sum_{j=0}^{m-1} L^j K_{n_{k+l-1}} \subseteq K_{n_{k+l-1}}.$$

It follows that

$$\left[(I - z^{-m}L^m)f_{n_{k+l}} + z^{-m}L^m f_{n_k}\right] \in K_{n_{k+l-1}},$$

as desired. This proves CLAIM 2.

To derive a contradiction in order to conclude $K(z) = K_\ell$, recall the Doeblin-Fortet Inequality, which we have assumed to hold for $k = 1$. Iterating the inequality yields

$$\|L^m f\| \leqslant r^m \|f\| + R \sum_{j=1}^{m} r^j \|L^{m-k} f\|'.$$

Taking $f = Lf_{n_k} - Lf_{n_l}$, we obtain

$$\begin{aligned}\|L^{m+1}f_{n_k} - L^{m+1}f_{n_l}\| &\leqslant r^m \|Lf_{n_k} - Lf_{n_l}\| + R \sum_{j=1}^{m} r^j \|L^{m-j}Lf_{n_k} - L^{m-j}Lf_{n_l}\|' \\ &\leqslant r^m \|L\| 2 + R \sum_{j=1}^{m} r^j M^{m-j} \|Lf_{n_k} - Lf_{n_l}\|',\end{aligned}$$

where we have used the fact that $\|f_{n_k}\| = 1$ and the Boundedness assumption.

Since $\sup_k \|f_{n_k}\| = 1 < +\infty$, Precompactness yields a subsequence $\{f_{n_{k_i}}\}_i$ such that

$$\|Lf_{n_{k_i}} - h\|' \xrightarrow{i\to+\infty} 0, \qquad \text{for some } h \in B.$$

Hence, for any $\epsilon > 0$, there are $l \neq k$ so large that

$$\|L^{m+1} f_{n_k} - L^{m+1} f_{n_l}\| \leqslant 2r^m \|L\| + \epsilon.$$

Choose m so large that $2r^m\|L\| \leqslant \frac{|z|^m}{4}$ and choose $\epsilon < \frac{|z|^m}{4}$. We thus obtain $k_i \neq k_j$ such that

$$\|L^{m+1} f_{n_{k_i}} - L^{m+1} f_{n_{k_j}}\| < \frac{|z|^m}{2},$$

contradicting the $\frac{|z|^m}{2}$-separation of sequence $\{L^m f_{n_k}\}_k$ for all $m \geqslant 1$ from CLAIM 2. We conclude that the sequence $K_1 \subseteq K_2 \subseteq \cdots$ eventually stabilizes and hence $K(z) = K_\ell$ for some $\ell \geqslant 1$, as required.

A similar argument shows that the descending sequence $I_1 \supseteq I_2 \supseteq \cdots$ also eventually stabilizes. Step I is complete.

5.4 Step II

Step II. $LK(z) \subseteq K(z)$, $LI(z) \subseteq I(z)$ and $B = K(z) \oplus I(z)$.

Proof of Step II. If $f \in K(z) \equiv \bigcup_{\ell \geqslant 0} \ker(zI - L)^\ell$, then $(zI-L)^\ell f = 0$ for some $\ell \geqslant 0$. Since

$$(zI - L)^\ell Lf = L(zI - L)^\ell f = L0 = 0,$$

it follows that $Lf \in K(z)$. This shows $LK(z) \subseteq K(z)$.

If $f \in I(z) \equiv \bigcap_{\ell \geqslant 0} \mathrm{Im}(zI - L)^\ell$, then for each $\ell \geqslant 0$, there is some $g_\ell \in B$ such that

$$(zI - L)^\ell g_\ell = f.$$

Now

$$Lf = L(zI - L)^\ell g_\ell = (zI - L)^\ell L g_\ell \in \mathrm{Im}(zI - L)^\ell, \qquad \forall \ell \geqslant 0.$$

This shows $Lf \in I(z)$ and hence $LI(z) \subseteq I(z)$.

Since both sequences K_ℓ and I_ℓ eventually stabilize $K(z) = K_m$ and $I(z) = I_m$ for some $m \geqslant 0$, it suffices to show that $B = K_m \oplus I_m$.

First we show $B = K_m + I_m$. If $f \in B$, then

$$(zI - L)^m f \in I_m = I_{2m}.$$

So there is some $g \in B$ with $(zI - L)^m f = (zI - L)^{2m} g$, and hence

$$(zI - L)^m [f - (zI - L)^m g] = 0.$$

Now

$$f = [f - (zI - L)^m g] + (zI - L)^m g \in K_m + I_m,$$

and hence $B = K_m + I_m$, as desired.

To see the sum is direct, we show $K_m \cap I_m = \{0\}$. Take $f \in K_m \cap I_m$. Since $f \in I_m$, we have $f = (zI - L)^m g$ for some $g \in B$. But also $f \in K_m$, and so

$$(zI - L)^{2m} g = (zI - g)^m f = 0.$$

Therefore, $g \in K_{2m} = K_m$, and hence $f = (zI - L)^m g = 0$. This shows $K_m \cap I_m = \{0\}$ and completes the proof of Step II. $\square$

5.5 Step III

Step III. We show $(zI - L) : I(z) \to I(z)$ is a bijection with bounded inverse.

Proof fof Step III. Let $m \geqslant 0$ be so large that $K(z) = K_m$ and $I(z) = I_m$. Note

$$\ker(zI - L) \cap I(z) = K_1 \cap I_m \subseteq K_m \cap I_m = \{0\}$$

because the sum $K_m \oplus I_m = B$ is direct from Step II. So $(zI - L)$ is injective on $I(z)$.

Also,

$$(zI - L)I(z) = (zI - L)I_m = I_{m+1} = I_m = I(z),$$

and so $(zI - L)$ is surjective on $I(z)$. We have shown that $(zI - L)$ is a bijection on $I(z)$.

Since $I(z)$ is a closed linear subspace from Step I, it follows that $I(z)$ is a Banach space (under the same norm as B), and hence the linear bijection $(zI - L)$ is an open mapping on $I(z)$. We conclude it has a bounded inverse on $I(z)$. This completes the proof of Step III. □

5.6 Step IV

Step IV. $K(z) = \{0\}$ for all but finitely many $z \in A(\rho, \rho(L))$, and $K(z) \neq \{0\}$ for at least one $z \in \mathbb{C}$ with $|z| = \rho(L)$.

Proof of Step IV. First, for a contradiction, suppose that $K(z_i) \neq \{0\}$ for infinitely many $\{z_i\}_{i \geqslant 1} \subseteq A(\rho, \rho(L))$. By compactness of the closed annulus $A(\rho, \rho(L))$, we obtain a subsequence $z_n \xrightarrow{n \to +\infty} z \in A(\rho, \rho(L))$.

On the one hand, since $K(z_n) \neq \{0\}$, we have $\ker(z_n I - L) \neq \{0\}$. Indeed, if $\ker(zI - L) = \{0\}$, then

$$(zI - L)f = 0 \quad \Rightarrow \quad f = 0,$$

and so

$$(zI - L)^\ell f = (zI - L)(zI - L)^{\ell-1} f = 0 \quad \Rightarrow \quad (zI - L)^{n-1} f = 0 \quad \Rightarrow \cdots \Rightarrow \quad f = 0,$$

that is, $\ker(zI - L)^\ell = \{0\}$, and hence, $K(z) = \{0\}$. On the other hand, if $w \neq z$, then

$$wf \neq zf, \quad \forall f \in B \backslash \{0\},$$

and hence $(wI - L)f \neq (zI - L)f$, therefore, $\ker(zI - L) \cap \ker(wI - L) = \{0\}$.

The above two observations allows us to form the direct sums

$$F_n := \ker(z_1 I - L) \oplus \cdots \oplus \ker(z_n I - L).$$

Note $F_1 \subsetneq F_2 \subsetneq \cdots$. By Riesz Lemma, we construct a sequence $f_n \in \ker(z_n I - L) \subseteq F_n$ with

$$\|f_n\| = 1, \quad \operatorname{dist}(f_n, F_{n-1}) \geqslant \frac{1}{2}.$$

Since $f_n \in \ker(z_n I - L)$, we have $(z_n I - L)f_n = 0$, that is, $z_n f_n = Lf_n$. Now for any $n, k, m \geqslant 1$, we have

$$\|L^m f_{n+k} - L^m f_n\| = \|z_{n+k}^m f_{n+k} - z_n^m f_n\| \geqslant \operatorname{dist}(z_{n+k}^m f_{n+k}, F_n) \geqslant \frac{1}{2}|z_{n+k}|^m \geqslant \frac{1}{2}\rho^m,$$

because $z \in A(\rho, \rho(L))$.

We derive a contradiction in a similar way as we did in Step I. By iterating the Doeblin-Fortet Inequality ($k = 1$) for m times, we obtain

$$\|L^m f\| \leqslant r^m \|f\| + R \sum_{j=1}^{n} r^j \|L^{m-j} f\|', \quad \forall f \in B, \forall m \geqslant 1.$$

Apply this to vector $(Lf_{n+k} - Lf_n)$ to get

$$\|L^{m+1}f_{n+k} - L^{m+1}f_n\| \leqslant r^m\|Lf_{n+k} - Lf_n\| + R\sum_{j=1}^{m} r^j\|L^{m-j}Lf_{n+k} - L^{m-j}Lf_n\|'$$
$$\leqslant r^m\|L\|2 + R\sum_{j=1}^{m} r^j M^{m-j}\|Lf_{n+k} - Lf_n\|', \quad \forall n, k, m \geqslant 1.$$

Since $\sup_n \|f_n\| = 1 < +\infty$, it follows from Precompactness that there is a subsequence n_l and some $h \in B$ with

$$\|Lf_{n_l} - h\|' \xrightarrow{l\to+\infty} 0.$$

Fix $\epsilon > 0$. We have

$$\|L^{m+1}f_{n_{l+1}} - L^{m+1}f_{n_l}\| \leqslant r^m\|L\|2 + \epsilon, \quad \forall m \geqslant 1, \quad \text{for large } l.$$

Since $r < \rho$, by taking a large m, we will have

$$\|L^{m+1}f_{n_{l+1}} - L^{m+1}f_{n_l}\| < \frac{1}{2}\rho^m,$$

a contradiction. We conclude there are at most finitely many $z \in A(\rho, \rho(L))$ for which $K(z) = \{0\}$.

To see there is at least one $z \in \mathbb{C}$ with $|z| = \rho(L)$ and $K(z) \neq \{0\}$, suppose the contrary. Then, there is some $\rho' < \rho(L)$ such that $K(z) = 0$ for all $z \in \mathbb{C}$ with $|z| \geqslant \rho'$. So $I(z) = B$ for all $z \in \mathbb{C}$ with $|z| \geqslant \rho'$; it follows from Step III that $(zI - L)$ has a bounded inverse on $I(z) = B$. This implies $\rho(L) \leqslant \rho' < \rho(L)$, a contradiction. We conclude there is at least one such z and complete the proof of Step IV. □

We have proven all four statements of Lemma 5.2. □

As discussed after the statement of Lemma 5.2, we now have the complete list of eigenvalues of L in $A(\rho, \rho(L))$

$$\{\lambda_1, \cdots, \lambda_t\} = \operatorname{spec}(L) \cap A(\rho, \rho(L)).$$

5.7 Step V

Step V. We show that the sum forming $F := \bigoplus_{i=1}^{t} K(\lambda_i)$ is direct $\dim(F) < +\infty$, $LF \subseteq F$, and that the eigenvalues of $L|_F$ are exactly $\lambda_1, \cdots, \lambda_t$.

Proof of Step V. To see the sum is direct, take $v_i \in K(\lambda_i)\backslash\{0\}$ with $\sum_{i=1}^{t} \alpha_i v_i = 0$ for some scalars $\alpha_i \in \mathbb{C}$. We need to show $\alpha_i = 0$ for all $i = 1, \cdots, t$.

For a contradiction, suppose $\alpha_j \neq 0$ for some j. By Step I, there is some $m \geqslant 1$ such that $K(\lambda_i) = \ker(\lambda_i I - L)^m$ for all $i = 1, \cdots, t$. Define formal polynomials

$$p_i(Z) := (\lambda_i - Z)^m, \quad q_j(Z) := \prod_{i\neq j} p_i(Z).$$

Note $q_j(L)v_i = 0$ for all $i \neq j$, and hence

$$0 = q_j(L)\left(\sum_{i=1}^{t} \alpha_i v_i\right) = \alpha_j q_j(L)v_j.$$

But $\alpha_j = 0$, and so $q_j(L)v_j = 0$.

Since polynomials $p_j(Z)$ and $q_j(Z)$ have no zeros in common, it follows that they are relatively prime and hence there are two other formal polynomials $a(Z), b(Z)$ such that

$$a(Z)p_j(Z) + b(Z)q_j(Z) = 1.$$

Evaluating this equation at $Z = L$ and applying it to v_j, we obtain

$$v_j = (a(L)p_j(L) + b(L)q_j(L))v_j = a(L)p_j(L)v_j + b(L)q_j(L)v_j = 0,$$

a contradiction. We thus conclude the sum is direct.

Since $\dim(K(\lambda_i)) < +\infty$ for each $i = 1, \cdots, t$ from Step I, it follows that the finite direct sum $F \equiv \bigoplus_{i=1}^t K(\lambda_i)$ also has finite dimension.

Again, from Step I, we have $LK(\lambda_i) \subseteq K(\lambda_i)$ for each $i = 1, \cdots t$; it then follows that $LF \subseteq F$.

From the eigenequation $\lambda f = L|_F f$, we derive

$$\sum_{i=1}^t \alpha_i \lambda f_i = \lambda f = L|_F f = L|_F \sum_{i=1}^t \alpha_i f_i = \sum_{i=1}^t \alpha_i \lambda_i f_i,$$

and hence

$$\alpha_i \lambda = \alpha_i \lambda_i, \quad \forall i = 1, \cdots, t.$$

Since $f \neq 0$, at least one of α_i is nonzero. Also, $\lambda_i \in A(\rho, \rho(L)) \subseteq \mathbb{C} \backslash \{0\}$. It then follows that $\alpha_i \neq 0$ for exactly one i, and for this i, we have $\lambda = \lambda_i$. This shows the eigenvalues of $L|_F$ all belong to $\{\lambda_1, \cdots, \lambda_t\}$.

Conversely, for any $i = 1, \cdots, t$, we have $K(\lambda_i) \neq \{0\}$, and hence $\ker(\lambda_i I - L) \neq \{0\}$. It follows that there is some nonzero $f \in \ker(\lambda_i I - L) K(\lambda_i) \subseteq F$ for which

$$L|_F f = Lf = \lambda_i f.$$

This shows $f \in F \backslash \{0\}$ is an eigenvector for eigenvalue λ_i of $L|_F$, and hence the eigenvalues of $L|_F$ are precisely $\{\lambda_1, \cdots, \lambda_t\}$. The proof of Step V is complete. □

5.8 Step VI

Step VI. We show $H \equiv \bigcap_{i=1}^t I(\lambda_i)$ is closed and L-invariant, and $B = F \oplus H$.

Proof of Step VI. From Step I, we have already each $I(\lambda_i)$ is closed and L-invariant, and hence the intersection H is also closed and L-invariant.

For each $i = 1, \cdots, t$, we have from Step II that $B = K(\lambda_i) \oplus I(\lambda_i)$, and thus a projection operator

$$\pi_i : B \to K(\lambda_i), \quad \pi_i(f) \in K(\lambda_i), \quad (I - \pi_i) f \in I(\lambda_i).$$

Firstly, note

$$\pi_i L = L \pi_i.$$

Indeed, since $LK(\lambda_i) \subseteq K(\lambda_i)$ and $LI(\lambda_i) \subseteq I(\lambda_i)$ from Step II, it follows that

$$\pi_i L f = \pi_i L(\pi_i f + (I - \pi_i) f) = \pi_i (L \pi_i f) + \pi_i (L(I - \pi_i) f) = L \pi_i f, \quad \forall f \in B.$$

Secondly, note $\pi_i \pi_j = 0$ for all $i \neq j$. Indeed, take $u \in B$ and let

$$v := \pi_j(u) \in K(\lambda_j) = \ker(\lambda_j I - L)^m.$$

By binomial expansion, we have

$$\begin{aligned} 0 = (\lambda_j I - L)^m v =& (\lambda_j I - \lambda_i I + \lambda_i I - L)^m v = \sum_{k=0}^m \binom{m}{k} (\lambda_j - \lambda_i)^{m-k} (\lambda_i I - L)^k v \\ =& (\lambda_j - \lambda_i)^m v + \sum_{k=1}^m \binom{m}{k} (\lambda_j - \lambda_i)^{m-k} (\lambda_i I - L)^k v \end{aligned}$$

By reorganizing terms, we obtain

$$v = -(\lambda_j - \lambda_i)^{-m} \sum_{k=1}^m \binom{m}{k} (\lambda_j - \lambda_i)^{m-k} (\lambda_i I - L)^k v.$$

Iterating this equation m times, we have

$$v = \left[-(\lambda_j - \lambda_i)^{-m} \sum_{k=1}^{m} \binom{m}{k} (\lambda_j - \lambda_i)^{m-k} (\lambda_i I - L)^k \right]^m v \in \mathrm{Im}(\lambda_i I - L)^m.$$

It follows that $v \in I(\lambda_i) = \ker \pi_i$ and hence $(\pi_i \pi_j)u = \pi_i(v) = 0$ for any $u \in B$, as desired.

Now we are ready to prove $B = F \oplus H$. Decompose any $f \in B$ into

$$f = \sum_{i=1}^{t} \pi_i(f) + \left(f - \sum_{i=1}^{t} \pi_i(f) \right).$$

Clearly $\sum_{i=1}^{t} \pi_i(f) \in F$. Also,

$$\left(f - \sum_{i=1}^{t} \pi_i(f) \right) \in \bigcap_{i=1}^{t} \ker \pi_i = \bigcap_{i=1}^{t} I(\lambda_i) = H.$$

This shows $B = F + H$.

To see the sum is direct, take $f \in F \cap H$. Then, $\pi_i(f) = 0$ for each i because $f \in H = \bigcap_{i=1}^{t} I(\lambda_i) = \bigcap_{i=1}^{t} \ker \pi_i$. But $f \in F$ implies $f = \sum_{i=1}^{t} \pi_i(f) = 0$. This shows $F \cap H = \{0\}$ and hence the sum is direct, as desired. The proof of Step VI is complete. □

5.9 Step VII

Step VII. We show $\rho(L|_H) \leqslant \rho$.

Proof of Step VII. It suffices to show that $(zI - L) : H \to H$ has a bounded inverse for any $z \in \mathbb{C}$ with $|z| \geqslant \rho$. Fix such a z and an $h \in H$.

If $|z| > \rho(L)$, then clearly $(zI - L)$ has a bounded inverse on B, and hence on H. Now suppose $z \in A(\rho, \rho(L))$.

If $z \notin \{\lambda_1, \cdots, \lambda_t\}$, then $K(z) = \{0\}$. So $I(z) = B$ and hence $(zI - L)$ has a bounded inverse on $I(z) = B$, according to Step III. Now suppose $z = \lambda_i$ for some i.

Recall from Step III that $(\lambda_i I - L)$ is an isomorphism on $I(\lambda_i)$. So there is a unique $f \in I(\lambda_i)$ for which $h = (\lambda_i I - L)f$. We show $f \in H$.

According to Steps V and VI, it suffices to check $\pi_j(f) = 0$ for all $j = 1, \cdots, t$.

If $j = i$, then $f \in I(\lambda_i) = \ker \pi_i$ and so $\pi_i(f) = 0$.

If $j \neq i$, then, by the first observation that $\pi_j L = L \pi_j$, we have

$$0 = \pi_j(h) = \pi_j(\lambda_i I - L)f = (\lambda_i I - L)\pi_j f.$$

This shows $\pi_j f \in \ker(\lambda_i I - L) \subseteq K(\lambda_i)$ and hence $\pi_j f \in K(\lambda_i) \cap K(\lambda_j) = \{0\}$. We conclude

$$f \in \bigcap_{j=1}^{t} \ker \pi_j \equiv H,$$

and hence for each $h \in H$, there is a unique $f \in H$ for which $h = (zI - L)f$. In other words, $(zI - H)$ is invertible on H. But H is a Banach space for being closed in Banach space B, and hence Open Mapping Theorem yields that $(zI - H)$ has a bounded inverse on H. This proves Step VII. □

In summary, we have proven that

$$B = F \oplus H$$

is an L-invariant decomposition with F finite-dimensional, H closed; all eigenvalues of $L|_F$ have modulus $\geqslant \rho$ because $\{\lambda_1, \cdots, \lambda_t\} \subseteq A(\rho, \rho(L))$, and $\rho(L|_H) \leqslant \rho$. We conclude that L is quasi-compact and close the proof of Hennion's Theorem. □

6 Gabriel: Exercise 1.5.4

Exercise 1.5.4. If $\widehat{T}$ has an acip, say μ_h, and 1 is a *simple* eigenvalue, and all other eigenvalues have modulus strictly less than 1, then the acip μ_h is *weak mixing*, i.e.,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}E \cap F) - \mu(E)\mu(F)| = 0, \quad \forall E, F \in \mathcal{B}.$$

6.0.1 Characterizations of Weak Mixing

Let $(X, \mathcal{B}, \mu, T)$ be a Lebesgue measure-preserving space. The space $L^2(\mu)$ is Hilbert with the following inner product:

$$\langle f, g\rangle = \int f\overline{g}\mathrm{d}\mu.$$

Definition 6.1. A complex number λ is an *eigenvalue* of T if it is an eigenvalue of the Koopman operator $U_T : L^2(\mu) \to L^2(\mu)$, i.e., if there exists $f \in L^2(\mu)$ such that $f \neq 0$ and $f \circ T = \lambda f$. Such an f is called an *eigenfunction* corresponding to λ.

Lemma 6.2. *If λ is an eigenvalue of T, then $|\lambda| = 1$.*

Proof. Suppose $U_T(f) = \lambda f$, where $f \neq 0$. Then:

$$\|f\|^2 = \|U_T(f)\|^2 = \|\lambda f\|^2 = |\lambda|^2\|f\|^2.$$

Therefore $|\lambda| = 1$. □

Definition 6.3. The probability invariant measure μ is *weak mixing* if, for every $E, F \in \mathcal{B}$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\mu(T^{-i}E \cap F) - \mu(E)\mu(F)| = 0. \tag{6}$$

Lemma 6.4. *(Koopman-von Newmann Lemma) If $(a_n)_{n\in\mathbb{N}}$ is a bounded sequence of real numbers then the following are equivalent:*

(a)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |a_i| = 0.$$

(b) There exists a subset $\mathcal{N}$ of $\mathbb{N}$ of density zero such that $\lim_{n\to\infty} a_n = 0$ provided $n \notin \mathcal{N}$. Density zero means that:

$$\left(\frac{\mathit{cardinality}(\mathcal{N} \cap \{0, \dots, n-1\})}{n}\right) \longrightarrow 0.$$

(c)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |a_i|^2 = 0.$$

Proof. If $M \subset \mathbb{N}$, denote by $\alpha_M(n)$ the cardinality of $M \cap \{0, \dots, n-1\}$.

(a)⇒(b) For each $k > 0$, define $J_k := \{n \in \mathbb{N}; |a_n| \geqslant 1/k\}$. Then $J_1 \subset J_2 \subset \cdots$. Each J_k has density zero since:

$$\frac{1}{n}\sum_{i=0}^{n-1} |a_i| \geqslant \frac{1}{n}\frac{1}{k}\alpha_{J_k}(n).$$

Therefore there exist integers $0 = m_0 < m_1 < m_2 < \cdots$ such that, for $n \geqslant m_k$:

$$\frac{1}{n}\alpha_{J_{k+1}}(n) < \frac{1}{k+1}.$$

Define:

$$\mathcal{N} := \bigcup_{k=0}^{\infty} [J_{k+1} \cap [m_k, m_{k+1})] .$$

We now show that $\mathcal{N}$ has density zero. Since $J_1 \subset J_2 \subset \cdots$, if $m_k \leqslant n < m_{k+1}$, then:

$$\mathcal{N} \cap [0, n) = [\mathcal{N} \cap [0, m_k)] \cup [\mathcal{N} \cap [m_k, n)] \subset [J_k \cap [0, m_k)] \cup [J_{k+1} \cap [0, n)] .$$

Therefore:

$$\frac{1}{n}\alpha_{\mathcal{N}}(n) \leqslant \frac{1}{n} \leqslant \frac{1}{n}\left(\alpha_{J_k}(m_k) + \alpha_{J_{k+1}}(n)\right) \leqslant \frac{1}{n}\left(\alpha_{J_k}(n) + \alpha_{J_{k+1}}(n)\right) \leqslant \frac{1}{k} + \frac{1}{k+1}.$$

Hence $(1/n)\alpha_{\mathcal{N}}(n) \to 0$ as $n \to \infty$, and so $\mathcal{N}$ has density zero. Now, if $n > m_k$ and $n \notin \mathcal{N}$, then $n \notin J_{k+1}$ and, therefore, $|a_n| < 1/(k+1)$. Hence:

$$\lim_{\mathcal{N} \not\ni n \to \infty} |a_n| = 0.$$

(b)⇒(a) Suppose $|a_n| \leqslant K$ for all $n \in \mathbb{N}$, and fix $\varepsilon > 0$. There exists N_ε such that $n \geqslant N_\varepsilon$ implies:

$$\frac{1}{n}\sum_{i=0}^{n-1}|a_i| = \frac{1}{n}\left[\sum_{i \in \mathcal{N} \cap \{0,\ldots,n-1\}} |a_i| + \sum_{i \notin \mathcal{N} \cap \{0,\ldots,n-1\}} |a_i|\right] < \frac{K}{n}\alpha_{\mathcal{N}}(n) + \varepsilon < (K+1)\varepsilon.$$

(a)⇔(c) By the above it suffices to note that $\lim_{\mathcal{N} \not\ni n \to \infty} |a_n| = 0$ iff $\lim_{\mathcal{N} \not\ni n \to \infty} |a_n|^2 = 0$. □

Theorem 6.5. *Let $(X, \mathcal{B}, \mu, T)$ be a Lebesgue probability invariant space. The following are equivalent:*

(a) μ is weak mixing;

(b) for all $f \in L^2(\mu)$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i(f), f\rangle - \langle f, \mathbb{1}\rangle\langle \mathbb{1}, f\rangle| = 0;$$

(c) for all $f, g \in L^2(\mu)$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i(f), g\rangle - \langle f, \mathbb{1}\rangle\langle \mathbb{1}, g\rangle| = 0;$$

(d) if $f \in L^2(\mu)$ is such that $U_T(f) = \lambda f$ for some $\lambda \in \mathbb{C}$, then f is constant almost everywhere.

Proof.

(a)⇒(b) By weak mixing, we have that, for $A, B \in \mathcal{B}$:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i(\chi_A), \chi_B\rangle - \langle \chi_A, \mathbb{1}\rangle\langle \mathbb{1}, \chi_B\rangle| = 0.$$

Fixing B and picking h to be a simple function, the bi-linearity of inner product and triangle inequality imply that:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\langle U_T^i(h), \chi_B\rangle - \langle h, \mathbb{1}\rangle\langle \mathbb{1}, \chi_B\rangle| = 0.$$

Now, fixing h, we conclude that:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(h),h\rangle-\langle h,\mathbb{1}\rangle\langle\mathbb{1},h\rangle|=0.$$

So (b) is valid for simple functions. Suppose $f\in L^2(\mu)$ and let $\varepsilon>0$. Choose a simple function h such that $\|f-h\|_2<\varepsilon$, and choose $N(\varepsilon)$ so that $n\geqslant N(\varepsilon)$ implies:

$$\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(h),h\rangle-\langle h,\mathbb{1}\rangle\langle\mathbb{1},h\rangle|<\varepsilon.$$

Then, if $n\geqslant N(\varepsilon)$:

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(f),f\rangle-\langle f,\mathbb{1}\rangle\langle\mathbb{1},f\rangle| &\leqslant \frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(f),f\rangle-\langle U_T^i(h),f\rangle|+\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(h),f\rangle-\langle U_T^i(h),h\rangle|\\ &\quad+\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(h),h\rangle-\langle h,\mathbb{1}\rangle\langle\mathbb{1},h\rangle|\\ &\quad+\frac{1}{n}\sum_{i=0}^{n-1}|\langle h,\mathbb{1}\rangle\langle\mathbb{1},h\rangle-\langle f,\mathbb{1}\rangle\langle\mathbb{1},h\rangle|\\ &\quad+\frac{1}{n}\sum_{i=0}^{n-1}|\langle f,\mathbb{1}\rangle\langle\mathbb{1},h\rangle-\langle f,\mathbb{1}\rangle\langle\mathbb{1},f\rangle|\\ &\leqslant \frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(f-h),f\rangle|+\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(h),f-h\rangle|+\varepsilon\\ &\quad+|\langle\mathbb{1},h\rangle||\langle h-f,\mathbb{1}\rangle|+|\langle f,\mathbb{1}\rangle||\langle\mathbb{1},h-f\rangle|\\ &\leqslant \|f-h\|_2\|f\|_2+\|f-h\|_2\|h\|_2+\varepsilon+\|h\|_2\|f-h\|_2=\|f\|_2\|h-f\|_2\\ &\leqslant \varepsilon\|f\|_2+\varepsilon(\|f\|_2+\varepsilon)+\varepsilon+(\|f\|_2+\varepsilon)\varepsilon+\varepsilon\|f\|_2.\end{aligned}$$

Therefore, $\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(f),f\rangle-\langle f,\mathbb{1}\rangle\langle\mathbb{1},f\rangle|=0$.

(b)⇒(c) Let $f\in L^2(\mu)$ and let $\mathcal{H}_f$ denote the smallest (closed) subspace of $L^2(\mu)$ containing f and the constant functions and satisfying $U_T\mathcal{H}_f\subset\mathcal{H}_f$. Define:

$$\mathcal{F}_f:=\left\{g\in L^2(\mu);\ \frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(f),g\rangle-\langle f,\mathbb{1}\rangle\langle\mathbb{1},g\rangle|=0\right\}.$$

By hypothesis, $\mathcal{F}_f$ is a closed subspace of $L^2(\mu)$ containing f and the constant functions. Since this set is U_T invariant, it contains $\mathcal{H}_f$. Now, if $g\in\mathcal{H}_f^\perp$, then $\langle U_T^n(f),g\rangle=0$ for $n\geqslant 0$ and $\langle\mathbb{1},g\rangle=0$. Therefore, $\mathcal{H}_f^\perp\subset\mathcal{F}_f$, i.e., $\mathcal{F}_f=L^2(\mu)$.

(c)⇒(a) Just choose characteristic function.

(b)⇒(d) Suppose $U_T(f)=\lambda f$ for some $f\in L^2(\mu)$. If $\lambda=1$, then f is constant a.e. by ergodicity of μ (weak mixing implies ergodicity). If $\lambda\neq 1$, then:

$$\lambda\int f\,\mathrm{d}\mu=\int f\circ T\,\mathrm{d}\mu=\int f\,\mathrm{d}T_*\mu=\int f\,\mathrm{d}\mu.$$

Hence $\langle f,\mathbb{1}\rangle=0$. Then, by hypothesis:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|\langle\lambda^i f,f\rangle|=\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_T^i(f),f\rangle|=0.$$

Since $|\lambda|=1$ this gives $\langle f,f\rangle=0$ and therefore $f=0$ a.e.

(d)⇒(a) This step requires some additional results as follows.

Definition 6.6. If $n < 0$, we define $U_T^n := (U_T^*)^{|n|}$, where U_T^* is the unique operator such that $\langle U_T^*(f), g\rangle = \langle f, U_T^*(g)\rangle$ for every $f, g \in L^2(\mu)$. Note that, if T is invertible, then $U_{T}^{-1} = (U_T)^{-1}$.

Definition 6.7. If $f \in L^2\backslash\{0\}$, we define the *spectral measure* of f as being the unique measure ν_f on $\mathbb{S}^1$ such that $\langle U_T^n(f), f\rangle = \int_{S^1} z^n \mathrm{d}\nu_f$ for every $n \in \mathbb{Z}$. (existence and uniqueness of ν_f?)

Proposition 6.8 ([?], Proposition 3.3)**.** *If T satisfies (d) on Theorem 6.5, then all the spectral measures of $f \in L^2$ such that $\int f d\mu = 0$ are non-atomic (i.e. unitary sets have zero measure).*

Proof. Suppose $f \in L^2(\mu)$ has measure zero and that ν_f has an atom $\lambda \in \mathbb{S}^1$. We will construct an eigenfunction with eigenvalue λ. Consider the sequence:

$$\frac{1}{n}\sum_{i=0}^{n-1} \lambda^{-i} U_T^i(f).$$

This sequence is bounded in norm, therefore has a weakly convergent subsequence (why?) (here we use the that $L^2(\mu)$ is separable – a consequence of the fact that $(X, \mathcal{B}, \mu)$ is a Lebesgue space):

$$\frac{1}{n_k}\sum_{i=0}^{n_k-1} \lambda^{-i} U_T^i(f) \xrightarrow{w} g.$$

The limit g satisfies $\langle U_T(g), h\rangle = \langle \lambda g, h\rangle$ for all $h \in L^2(\mu)$ (check!). Therefore g is an eigenfunction with eigenvalue λ. Now we show g is non-constant. We have:

$$\begin{aligned}
\langle g, f\rangle &= \lim_{k\to\infty} \frac{1}{n_k}\sum_{i=0}^{n_k-1} \lambda^{-i}\langle U_T^i(f), f\rangle = \lim_{k\to\infty} \frac{1}{n_k}\sum_{i=0}^{n_k-1} \int_{\mathbb{S}^1} \lambda^{-i} z^i \mathrm{d}\nu_f(z) \\
&= \nu_f(\{\lambda\}) + \lim_{k\to\infty} \frac{1}{n_k}\sum_{i=0}^{n_k-1} \int_{\mathbb{S}^1\backslash\{\lambda\}} \lambda^{-i} z^i \mathrm{d}\nu_f(z) \\
&= \nu_f(\{\lambda\}) + \lim_{k\to\infty} \int_{\mathbb{S}^1\backslash\{\lambda\}} \frac{1}{n_k}\cdot\frac{1-\lambda^{n_k} z^{n_k}}{1-\lambda^{-1}z} \mathrm{d}\nu_f(z).
\end{aligned}$$

The limit tends to zero, because the integrand tends to zero and is uniformly bounded by 1. Thus $\langle g, f\rangle = \nu_f(\{\lambda\}) \neq 0$, whence g is non-constant. This contradicts the hypothesis on T. □

Lemma 6.9. *If T satisfies (d) on Theorem 6.5, then for every real-valued $f \in L^2(\mu)$:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\langle U_T^i(f), f\rangle - \langle f, \mathbb{1}\rangle^2\right| = 0.$$

Proof. It is enough to consider the case when $\int f \mathrm{d}\mu = 0$ (when $\int f \mathrm{d}\mu \neq 0$, applie the result for $F := f - \int f \mathrm{d}\mu$). Let ν_f be the spectral measure of f. Then, for each $n \in \mathbb{N}$:

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} \left|\langle U_T^i(f), f\rangle\right|^2 &= \frac{1}{n}\sum_{i=0}^{n-1} \left|\int_{\mathbb{S}^1} z^n \mathrm{d}\nu_f(z)\right|^2 \\
&= \frac{1}{n}\sum_{i=0}^{n-1} \left(\int_{\mathbb{S}^1} z^n \mathrm{d}\nu_f(z)\right)\overline{\left(\int_{\mathbb{S}^1} z^n \mathrm{d}\nu_f(z)\right)} \\
&= \frac{1}{n}\sum_{i=0}^{n-1} \int_{\mathbb{S}^1}\int_{\mathbb{S}^1} z^n \overline{w}^n \mathrm{d}\nu_f(z)\mathrm{d}\nu_f(w) \\
&= \int_{\mathbb{S}^1}\int_{\mathbb{S}^1} \frac{1}{n}\left(\sum_{i=0}^{n-1} z^n\overline{w}^n\right) \mathrm{d}\nu_f(z)\mathrm{d}\nu_f(w).
\end{aligned}$$

The integrand tends to zero and is bounded outside $\triangle := \{(z,w);\ z = w\}$ (why?). If we can show that $(\nu_f \times \nu_f)(\triangle) = 0$, then it will follow that:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \langle U_T^i(f), f \rangle \right|^2 = 0.$$

This is indeed the case: T satisfies (d) on Theorem6.5, so, by the previous proposition, ν_f is non-atomic. By Fubini-Tonelli, it follows that $(\nu_f \times \nu_f)(\triangle) = \int_{\mathbb{S}^1} \nu_f(\{w\}) \mathrm{d}\nu_f(w) = 0$.

Now, by Koopman-von Newmann Lemma, for every bounded non-negative sequence $(a_n)_{n\in\mathbb{N}}$, $\frac{1}{n}\sum_{i=0}^{n-1} a_n \to 0$ iff $\frac{1}{n}\sum_{i=0}^{n-1} a_n^2 \to 0$, and this completes the proof. ☐

Proposition 6.10. *If T satisfies (d) on Theorem6.5, then for every real-valued $f, g \in L^2(\mu)$:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \langle U_T^i(f), g \rangle - \langle f, \mathbb{1} \rangle \langle g, \mathbb{1} \rangle \right| = 0. \tag{7}$$

Proof. Again it is enough to consider the case when $\int f \mathrm{d}\mu = 0$. Define:

$$S(f) := \overline{\mathrm{span}\,\{U_T^n(f);\ n \geqslant 0\} \cup \{1\}}.$$

Then $L^2(\mu) = S(f) \oplus S(f)^\perp$.

- Every $g \in S(f)$ satisfies (7). In fact, note that if $g_1, \ldots, g_m$ satisfy (7), then so does $g := \sum \alpha_i g_i$ for any $\alpha_i \in \mathbb{R}$. Therefore it is enough to check (7) for $g := U_T^n(f)$ and $g \equiv$ constant. Constant functions satisfy (7) because, for such functions, $\int g \cdot f \circ T^n \mathrm{d}\mu = 0$ for all $n \geqslant 0$ since $\int f \mathrm{d}\mu = 0$. Set $g := U_T^n(f)$ for some $n \geqslant 0$. Then, for all $m \geqslant n$:

$$\int g \cdot f \circ T^m \mathrm{d}\mu = \int (f \circ T^n)(f \circ T^m) \mathrm{d}\mu = \int f \cdot f \circ T^{m-n} \mathrm{d}\mu \overset{\mathcal{N} \not\ni n \to \infty}{\longrightarrow} 0,$$

  for some $\mathcal{N} \subset \mathbb{N}$ of density zero, by Lemma6.9. The claim follows by Koopman-von Newmann Lemma.
- Every $g \perp S(f) \oplus \{\text{constants}\}$ satisfies (7), because $\langle g,\ f \circ T^n \rangle$ is eventually zero and $\int g \mathrm{d}\mu = \langle g,\ \mathbb{1} \rangle = 0$. ☐

Remark 6.11. Choosing characterisc functions in (7), we conclude the implication (d)⇒(a) on Theorem6.5. ☐

6.0.2 The Exercise

Exercise 1.5.3. If $\widehat{T}$ has an acip, say μ_h, and 1 is a *simple* eigenvalue, i.e., $\dim\{g \in L^1(\mu) : \widehat{T} g = g\} = 1$, **then the acip μ_h is unique and ergodic.**

Remark 6.12. Being μ_h the unique acip, we conclude that $\mu_h = \mu$.

Lemma 6.13. *[[?], Proposition 1.1] The following are equivalent:*

- *μ is ergodic;*
- *if $f : X \to \mathbb{R}$ is a measurable function such that $f \circ T = f$ a.e., then f is constant a.e.*

Lemma 6.14. *If $f \in L^1(\mu)$ and $g \in L^\infty(\mu)$, the following equality holds:*

$$\widehat{T}\left((f \circ T) \cdot g\right) = f \cdot \widehat{T}(g).$$

Proof. Fix $\varphi \in L^\infty(\mu)$. Then:

$$\int \varphi \cdot \widehat{T}\left((f \circ T) \cdot g\right) \mathrm{d}\mu = \int (\varphi \circ T) \cdot (f \circ T) \cdot g \mathrm{d}\mu = \int [(\varphi \cdot f) \circ T] \cdot g \mathrm{d}\mu \overset{9}{=} \int (\varphi \cdot f) \cdot \widehat{T}(g) \mathrm{d}\mu = \int \varphi \cdot \left[f \cdot \widehat{T}(g) \right] \mathrm{d}\mu.$$

Therefore $\widehat{T}\left((f \circ T) \cdot g\right) = f \cdot \widehat{T}(g)$. ☐

[9] Repeat the proof of Characterization of the Transfer Operator interchanging the roles of $L^1(\mu)$ and $L^\infty(\mu)$; in fact it is valid for every $p, q \in [1, \infty]$ such that $p^{-1} + q^{-1} = 1$.

Exercise 1.5.4. If $\widehat{T}$ has an acip, say μ_h, and 1 is a *simple* eigenvalue, and all other eigenvalues have modulus strictly less than 1, then the acip μ_h is *weak mixing*.

Proof. By Exercise 1.5.3, we know that $\mu_h = \mu$ and μ is ergodic.

Let $f \in L^2(\mu)$ and $\lambda \in \mathbb{C}$ be such that $f \circ T = \lambda f$. We need to show that f is constante a.e. By Lemma6.2, $|\lambda| = 1$. Now, by the previous lemma, we have:

$$\widehat{T}(\lambda f) = \widehat{T}\left((f \circ T) \cdot 1\right) = f \cdot \widehat{T}(\mathbb{1}) = f.$$

This implies that $\widehat{T}(f) = \frac{1}{\lambda} f$. But $|\frac{1}{\lambda}| = |\lambda| = 1$ and, by hypothesis, we conclude that $\lambda = 1$. Being μ ergodic, it follows from Lemma6.13 that f is constant a.e. □

Next Exercises: Show that Exercises 1.5.3 and 1.5.4 are "iff".

7 Edmilson: space of Lipschitz functions

Check $(\mathrm{Lip}, \|\cdot\|_{\mathrm{Lip}})$ is a Banach space;

Proof. **Disclaimer.** We denote the space of Lipschitz functions as $(\mathcal{L}, \|\cdot\|)$ where

$$\|f\| := \|f\|_\infty + \mathrm{Lip}(f),$$

$$\mathrm{Lip}(f) = \sup_{x \neq y} \left\{ \frac{|f(x) - f(y)|}{|x - y|} \right\}.$$

To prove that $(\mathcal{L}, \|\cdot\|)$ is a Banach space, take a Cauchy sequence $\{f_n\}_n \subset \mathcal{L}$. Hence, for a fixed $\varepsilon > 0$ there exists $N_0 > 0$ such that for $n, m > N_0$ we have

$$\|f_n - f_m\| < \varepsilon. \tag{8}$$

By the definition of $\|\cdot\|$ we conclude

$$\begin{aligned} \varepsilon &> Lip(f_n - f_m) \\ &> \frac{|f_n(x) - f_n(y) - (f_m(x) - f_m(y))|}{|x - y|}. \end{aligned}$$

Note that $f_n(x)$ is clearly a Cauchy sequence of complex numbers for any fixed $x \in [0, 1]$. In particular, by completeness of $\mathbb{C}$ there is a limit $f(x)$ for each x. Thus, we get a limiting function $f(t)$. Moreover, letting $m \to \infty$ in

$$\frac{|f_n(x) - f_n(y) - (f_m(x) - f_m(y))|}{|x - y|} < \varepsilon, \quad \forall n, m > N_0, x \neq y \in [0, 1]$$

we see

$$\frac{|f_n(x) - f_n(y) - (f(x) - f(y))|}{|x - y|} < \varepsilon.$$

If we define $g_n := f_n - f$, rearranging the above expression, we obtain

$$|g_n(x) - g_n(y)| < \varepsilon |x - y|,$$

which means, g_n is a Lipschitz function, i.e., $f_n - f \in \mathcal{L}$. By assumption, $f_n \in \mathcal{L}$, so we conclude $f = f_n - g_n \in \mathcal{L}$, since the difference of two Lipschitz functions is Lipschitz.

Since $f \in \mathcal{L}$, we proved that any Cauchy sequence with respect to $\|\cdot\|$ converged to a point inside the space. Hence, the claim follows. ☐

Prove $\|f\|_\infty \leqslant \mathrm{Lip}(f)$ for complex Lipschitz observables $f : [0, 1] \to \mathbb{C}$ with $\int f \mathrm{dLeb} = 0$.

Definition 7.1. A subset $S \subset \mathbb{R}^n$ is said to be *convex* if $(1 - \lambda)x + \lambda y \in S$ for all $x, y \in S$ and $0 < \lambda < 1$.

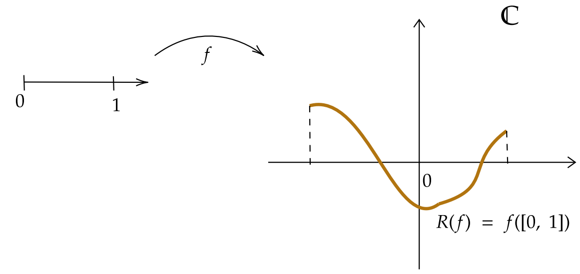
Definition 7.2. The intersection of all the convex sets containing a given subset $S \subset \mathbb{R}^n$ is called the *convex hull* of S and is denoted by $conv\ S$.

Let $x_1, \ldots, x_m \in \mathbb{R}^n$. A vector sum

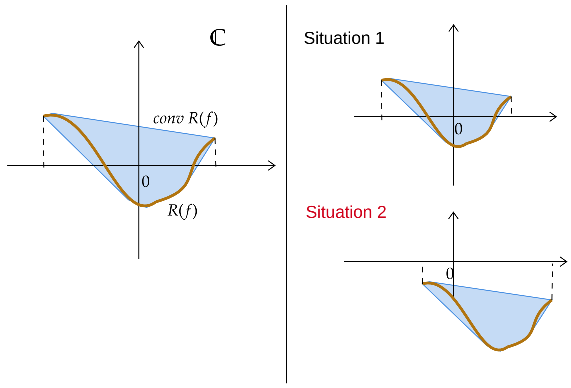
$$\lambda_1 x_1 + \cdots + \lambda_m x_m$$

is called a *convex combination* of $x_1, \ldots, x_m$ if the coefficients λ_i are all non-negative and $\lambda_1 + \cdots + \lambda_m = 1$.

Theorem 7.3 ([Roc70]). *For any $S \subset \mathbb{R}^n$, $conv\ S$ consists of all the convex combinations of the elements of S.*



(a) Image under f is denoted $R(f) \subset \mathbb{C}$.



(b) Left panel shows the convex hull of $R(f)$ in shaded blue. Note the origin is not in the $R(f)$, even though we there are points of the type $(x, 0)$ and $(0, y)$. Right panel shows two distinct situations: Situation 1 is the case in which the origin is inside the convex hull of $R(f)$, *conv* $R(f)$, and situation 2 is the opposite case where the origin is in the complement of *conv* $R(f)$.

Figure 7: The reason for the origin to be inside the closed convex hull $\overline{R(f)}$ is captured by the geometric picture depicted in Situation 1 and 2. Note both figures differ in where the argument of replacing sup $|f(x)|$ by the diameter of $R(f)$ is correct.

Proof. Having Figure 7 in mind the proof goes as follow. Assume 0 is in the closed convex hull of $f([0,1])$ (0 is inside the blue region). Then,

$$\begin{aligned}\|f\|_\infty &= ess \sup_{x\in[0,1]} |f(x)| \\ &\leqslant \operatorname{diam}(f([0,1])) \\ &\leqslant \operatorname{Lip}(f)\operatorname{diam}([0,1]) = \operatorname{Lip}(f).\end{aligned}$$

We prove the following claim: $0 \in \overline{conv\ f} \subset \mathbb{C}$.

We can construct a probability measure (an average of delta measures over equally spaced points in the interval $[0,1]$) such that in the limit this converges to the Leb measure on $[0,1]$. This average is a convex combination on the interval, and in the limit, on $\mathbb{C}$. That is,

$$\mu_n := \frac{1}{n+1}\sum_{i=0}^{n} \delta_{i/n} \xrightarrow{*} \text{Leb}.$$

Hence, in particular, since f is continuous, then

$$\frac{1}{n+1}\sum_{i=0}^{n} f(i/n) = \int f d\mu_n \to \int f d\text{Leb} = 0.$$

So, we conclude 0 could be written as a limit of a sequence $(y_n)_n$ of the form $y_n = \sum_{i=0}^n \lambda_i f(i/n)$, $\sum_{i=0}^n \lambda_i = 1$ and for each i: $\lambda_i \geqslant 0$. ☐

Discuss how norm dominance in the subspace is related to the ideas of compact embedding.

Finite-dimensional example. Pick the real line $\mathbb{R}$ and think as an object into the plane, $\mathbb{R}^2$, where both space are given by the Euclidean norm. So, we could think as a map:

$$\begin{aligned} i : (\mathbb{R}, \|\cdot\|_2) &\to (\mathbb{R}^2, \|\cdot\|_2) \\ x &\mapsto (x, 0). \end{aligned}$$

In this case, $\|x\|_2 = \|(x,0)\|_2$ for every $x \in \mathbb{R}$.

This motivates us to introduce the notion of *continuous embedding*.

Definition 7.4. (Continuous embedding) Let $\mathcal{B}$ and $\mathcal{B}'$ be two normed vector spaces, with norms $\|\cdot\|_{\mathcal{B}}$ and $\|\cdot\|_{\mathcal{B}'}$ respectively, such that $\mathcal{B} \subset \mathcal{B}'$. If the inclusion map (identity function)

$$\begin{aligned} i : &\mathcal{B} \to \mathcal{B}' \\ &x \mapsto x \end{aligned}$$

is continuous, i.e. if there exists a constant $C \geqslant 0$ such that

$$\|x\|_{\mathcal{B}'} \leqslant C\|x\|_{\mathcal{B}}$$

for every $x \in \mathcal{B}$, then $\mathcal{B}$ is said to be continuously embedded in $\mathcal{B}'$.

From this definition, the *norm dominance* we observed so far are:

- From the previous exercise, we conclude $\mathcal{L}$ is continuously embedded into L^∞.
- And more importantly, since $X = [0,1]$ has finite measure, we conclude $\mathcal{L}$ is continuously embedded into L^1 ($\|f\|_1 \leqslant \|f\|_\infty$).

Norm dominance per se is not enough to prove Hennion's theorem. We need in addition that the embedding of $\mathcal{B}$ into $\mathcal{B}'$ is a compact operator.

Definition 7.5. (Compact embedding) Let $\mathcal{B}$ and $\mathcal{B}'$ be two normed vector spaces, and suppose that $\mathcal{B} \subset \mathcal{B}'$. We say that $\mathcal{B}$ is compactly embedded in $\mathcal{B}'$ ($\mathcal{B} \subset\subset \mathcal{B}'$) if

- $\mathcal{B}$ is continuously embedded in $\mathcal{B}'$; and
- the embedding of $\mathcal{B}$ into $\mathcal{B}'$ is a compact operator, i.e., any bounded sequence in $\mathcal{B}$ admits a subsequence which is Cauchy in the norm of $\mathcal{B}'$.

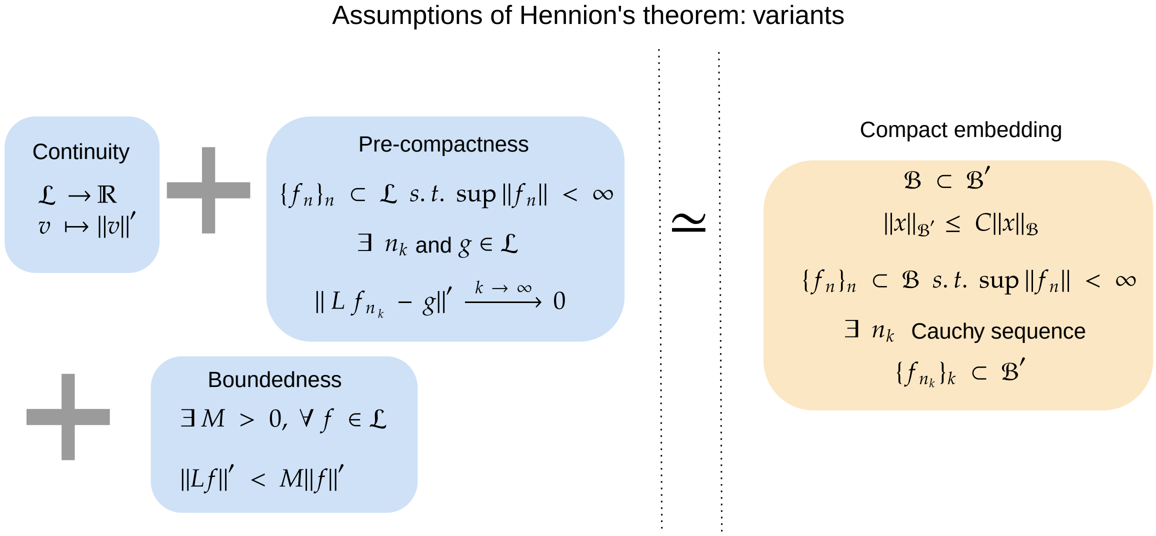


Figure 8: Left hand side shows Sarig's statement of Hennion's theorem. Right hand side shows a variant formulation in terms of compact embedding.

Theorem 7.6 ([Bal18]). *Let $L : \mathcal{B} \to \mathcal{B}$ be a bounded operator on a Banach space $(\mathcal{B}, \|\cdot\|)$, and let $(\mathcal{B}', \|\cdot\|')$ be a Banach space containing $\mathcal{B}$ such that the inclusion $\mathcal{B} \subset \mathcal{B}'$ is compact. Assume that there exist two sequences of real numbers r_n and R_n such that for any $n \geqslant 1$ and any $\varphi \in \mathcal{B}$*

$$\|L^n \varphi\| \leqslant r_n \|\varphi\| + R_n \|\varphi\|'.$$

Then the essential spectral radius *of L on $\mathcal{B}$ is not larger than*

$$\liminf_{n\to\infty} (r_n)^{1/n}.$$

For a definition of *essential spectral radius* and an alternative definition of quasi-compactness see Hennion and Herve's book [Hen01] (Sarig cite this book in the end of Appendix A3.

8 2020.3.10 – 24 Meetings 9–11: Sarig L3, Analytic Perturbation Theory

This section studies the analytic perturbation of a bounded linear operator and culminates in Theorem 8.19, which says that a small analytic perturbation does not destroy the spectral gap. We will later use this result to prove the Central Limit Theorem.

Let $(\mathcal{L}, \|\cdot\|)$ be a Banach space over $\mathbb{C}$, $B = B(\mathcal{L})$ denote the space of all bounded linear operators $L : \mathcal{L} \to \mathcal{L}$ with norm

$$\|L\| := \sup_{x \in \mathcal{L} \setminus \{0\}} \frac{\|Lx\|}{\|x\|}.$$

Denote by $\mathcal{L}^*$ the space of all bounded linear functionals $\varphi : \mathcal{L} \to \mathbb{C}$ with norm

$$\|\varphi\| := \sup_{x \in \mathcal{L} \setminus \{0\}} \frac{|\varphi(x)|}{\|x\|},$$

and similarly denote by B^* the space of all bounded linear functionals $\varphi : B \to \mathbb{C}$.

We will be interested in one-parameter (complex) family $L_z, z \in U \subseteq \mathbb{C}$ of bounded linear operators $L_z : \mathcal{L} \to \mathcal{L}$. More precisely, let $U \subseteq \mathbb{C}$ be open and nonempty, and consider

$$L : U \to B, \quad z \mapsto L_z.$$

This dependence of $L_z \in B$ on $z \in U$ will be "analytic" (a notion to be specified below). Hence, we may think of the family L_z as an analytic perturbation of some fixed operator L_{z_0}.

8.1 Calculus in Banach Spaces

8.1.1 Riemann Integral and Riemann Sums

Recall the area under a continuous curve $f : [a, b] \to \mathbb{R}$ is calculated by Riemann integral $\int_a^b f \mathrm{d}x$, defined as the limit of Riemann sums

$$\sum_{i=1}^{n} f(\xi_{[t_i, t_{i+1}]})[t_{i+1} - t_i],$$

where $a = t_1 < t_2 < \cdots < t_{n+1} = b$, $\xi_{[t_i, t_{i+1}]} \in [t_i, t_{i+1}]$, and the limit is taken as the mesh size $\max_{i=1,\cdots,n} |t_{i+1} - t_i|$ tends to 0.

To see this limit exists, take two meshes $M' = \{t'_1, \cdots, t'_{n'+1}\}$ of size δ' and $M'' = \{t''_1, \cdots, t''_{n''+1}\}$ of size δ''. Amalgamate the two together to obtain a finer mesh $M = \{t_1, \cdots, t_{n+1}\}$ of size $\delta := \min\{\delta', \delta''\}$. Denote by $R(M)$ the Riemann sum for mesh M.

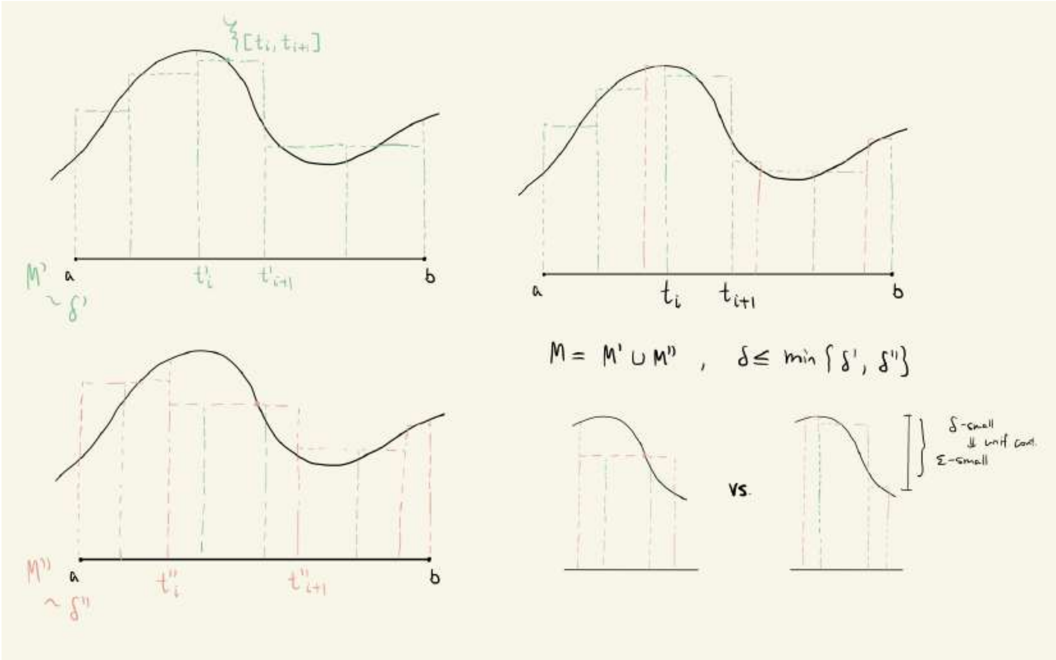


Figure 9: Two meshes M' of size δ' and M'' of size δ'' amalgamate to a refinement of both meshes, namely, M of size δ. The Riemann sums $R(M')$ and $R(M'')$ corresponding to the two meshes can be compared against $R(M)$, establishing Cauchy-ness of the Riemann sums.

Fix $\epsilon > 0$. By uniform continuity of continuous f on compact interval $[a, b]$, there is some $\delta(\epsilon) > 0$ for which $|x - y| \leqslant \delta(\epsilon)$ implies $|f(x) - f(y)| \leqslant \epsilon$.

When $\delta', \delta'' \leqslant \delta(\epsilon)$, we have

$$\begin{aligned}
&|R(M') - R(M'')| \leqslant |R(M') - R(M)| + |R(M) - R(M'')| \\
=&\left|\sum_{i=1}^{n'} f(\xi'_{[t'_i,t'_{i+1}]})[t'_{i+1} - t'_i] - \sum_{i=1}^{n} f(\xi_{[t_i,t_{i+1}]})[t_{i+1} - t_i]\right| + \left|\sum_{i=1}^{n} f(\xi_{[t_i,t_{i+1}]})[t_{i+1} - t_i] - \sum_{i=1}^{n''} f(\xi''_{[t''_i,t''_{i+1}]})[t''_{i+1} - t''_i]\right| \\
=&\left|\sum_{i=1}^{n} f(\xi'_{[t'_{j(i)},t'_{j(i)+1}]})[t_{i+1} - t_i] - \sum_{i=1}^{n} f(\xi_{[t_i,t_{i+1}]})[t_{i+1} - t_i]\right| + \left|\sum_{i=1}^{n} f(\xi_{[t_i,t_{i+1}]})[t_{i+1} - t_i] - \sum_{i=1}^{n} f(\xi''_{[t''_{j(i)},t''_{j(i)+1}]})[t_{i+1} - t_i]\right|,
\end{aligned}$$

where $[t'_{j(i)}, t'_{j(i)+1}]$ is the unique interval from mesh M' which contains the finer interval $[t_i, t_{i+1}]$ from mesh M. We continue

$$\begin{aligned}
|R(M') - R(M'')| &\leqslant \sum_{i=1}^{n} \max_{s,t\in[t'_{j(i)},t'_{j(i)+1}]} |f(t) - f(s)|[t_{i+1} - t_i] + \sum_{i=1}^{n} \max_{s,t\in[t''_{j(i)},t''_{j(i)+1}]} |f(t) - f(s)|[t_{i+1} - t_i] \\
&\leqslant \sum_{i=1}^{n} \epsilon[t_{i+1} - t_i] + \sum_{i=1}^{n} \epsilon[t_{i+1} - t_i] = 2\epsilon(b - a)
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, it follows that the Riemann sums are Cauchy, and hence converge by completeness of $\mathbb{R}$.

8.1.2 Line Integral for $\mathbb{C}$-valued Functions

Let $U \subseteq \mathbb{C}$ be open and nonempty, $\gamma \subseteq U$ a curve with smooth parametrization $z(t)$, $t \in [a, b]$.

Suppose $f : \gamma \to \mathbb{C}$ is continuous. Then the *line integral* $\int_\gamma f(z)\mathrm{d}z$ of f on curve γ is defined as the limit of Riemann sums

$$\sum_{i=1}^{n} f(z(\xi_i))[z(t_{i+1}) - z(t_i)],$$

where $a = t_1 < t_2 < \cdots < t_{n+1} = b$, $\xi_i \in [t_i, t_{i+1}]$, and the limit is taken as the mesh size $\max_{i=1,\cdots,n} |t_{i+1} - t_i|$ tends to 0.

To show the limit exists[10], we make use of the C^1-smoothness of z by writing

$$\sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))[z(t_{i+1}) - z(t_i)] = \sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))\frac{z(t_{i+1}) - z(t_i)}{t_{i+1} - t_i}[t_{i+1} - t_i]. \tag{9}$$

Again, take two meshes $M' = \{t'_1, \cdots, t'_{n'+1}\}$ and $M'' = \{t''_1, \cdots, t''_{n''+1}\}$ of sizes δ' and δ'', and amalgamate the two together to obtain a finer mesh $M = \{t_1, \cdots, t_{n+1}\}$ of size $\delta := \min\{\delta', \delta''\}$. Denote by $R(M)$ the Riemann sum (9) for mesh M.

Fix $\epsilon > 0$. By differentiability of z at $t \in [a, b]$, there is some $\delta_1(\epsilon, t) > 0$ for which

$$|s - t| \leqslant \delta_1(\epsilon, t) \quad \Rightarrow \quad \left|\frac{|z(t) - z(s)|}{t - s} - z'(t)\right| \leqslant \epsilon.$$

By compactness of $[a, b]$, we obtain a uniform size $\delta_1(\epsilon) > 0$ for which the above implication holds. So when $\delta', \delta'' \leqslant \delta_1(\epsilon)$, for mesh M, we have

$$\left|\frac{|z(t_{i+1}) - z(t_i)|}{t_{i+1} - t_i} - z'(t_i)\right| \leqslant \epsilon.$$

And similarly for meshes M' and M''.

Since z is C^1, its first derivative z' is C^0, and hence uniformly continuous on compact interval $[a, b]$. So there is some $\delta_2(\epsilon) > 0$ for which

$$|s - t| \leqslant \delta_2(\epsilon) \quad \Rightarrow \quad |z'(t) - z'(s)| \leqslant \epsilon.$$

By uniform continuity of continuous function $f \circ z$ on compact interval $[a, b]$, there is some $\delta_0(\epsilon) > 0$ for which

$$|s - t| \leqslant \delta_0(\epsilon) \quad \Rightarrow \quad |f(z(t)) - f(z(s))| \leqslant \epsilon.$$

[10] Another, perhaps simpler, way to show convergence of Riemann sums is to make the Cauchy-ness argument directly for $\sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))[z(t_{i+1}) - z(t_i)]$. We will use C^1-smoothness of the parametrization z for γ to establish $\sum_{i=1}^{n} |z(t_{i+1}) - z(t_i)| \to \mathrm{length}(\gamma) < +\infty$, which is not generally true for continuous curves.

When $\delta', \delta'' \leqslant \min\{\delta_0(\epsilon), \delta_1(\epsilon), \delta_2(\epsilon)\}$, we have

$$
\begin{aligned}
&|R(M') - R(M'')| \leqslant |R(M') - R(M)| + |R(M) - R(M'')| \\
=&\left|\sum_{i=1}^{n'} f(z(\xi'_{[t'_i,t'_{i+1}]}))\frac{z(t'_{i+1}) - z(t'_i)}{t'_{i+1} - t'_i}[t'_{i+1} - t'_i] - \sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))\frac{z(t_{i+1}) - z(t_i)}{t_{i+1} - t_i}[t_{i+1} - t_i]\right| \\
&+ \left|\sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))\frac{z(t_{i+1}) - z(t_i)}{t_{i+1} - t_i}[t_{i+1} - t_i] - \sum_{i=1}^{n''} f(z(\xi''_{[t''_i,t''_{i+1}]}))\frac{z(t''_{i+1}) - z(t''_i)}{t''_{i+1} - t''_i}[t''_{i+1} - t''_i]\right| \\
\leqslant&\left|\sum_{i=1}^{n'} f(z(\xi'_{[t'_i,t'_{i+1}]}))\frac{z(t'_{i+1}) - z(t'_i)}{t'_{i+1} - t'_i}[t'_{i+1} - t'_i] - \sum_{i=1}^{n'} f(z(\xi'_{[t'_i,t'_{i+1}]}))z'(t'_i)[t'_{i+1} - t'_i]\right| \\
&+ \left|\sum_{i=1}^{n'} f(z(\xi'_{[t'_i,t'_{i+1}]}))z'(t'_i)[t'_{i+1} - t'_i] - \sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))z'(t_i)[t_{i+1} - t_i]\right| \\
&+ 2\left|\sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))z'(t_i)[t_{i+1} - t_i] - \sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))\frac{z(t_{i+1}) - z(t_i)}{t_{i+1} - t_i}[t_{i+1} - t_i]\right| \\
&+ \left|\sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))z'(t_i)[t_{i+1} - t_i] - \sum_{i=1}^{n''} f(z(\xi''_{[t''_i,t''_{i+1}]}))z'(t''_i)[t''_{i+1} - t''_i]\right| \\
&+ \left|\sum_{i=1}^{n''} f(z(\xi''_{[t''_i,t''_{i+1}]}))z'(t''_i)[t''_{i+1} - t''_i] - \sum_{i=1}^{n''} f(z(\xi''_{[t''_i,t''_{i+1}]}))\frac{z(t''_{i+1}) - z(t''_i)}{t''_{i+1} - t''_i}[t''_{i+1} - t''_i]\right| \\
\leqslant&\sum_{i=1}^{n'} |f(z(\xi'_{[t'_i,t'_{i+1}]}))|\epsilon[t'_{i+1} - t'_i] \\
&+ \left|\sum_{i=1}^{n} f(z(\xi'_{[t'_{j(i)},t'_{j(i)+1}]}))z'(t'_{j(i)})[t_{i+1} - t_i] - \sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))z'(t_i)[t_{i+1} - t_i]\right| \\
&+ 2\sum_{i=1}^{n} |f(z(\xi_{[t_i,t_{i+1}]}))|\epsilon[t_{i+1} - t_i] \\
&+ \left|\sum_{i=1}^{n} f(z(\xi_{[t_i,t_{i+1}]}))z'(t_i)[t_{i+1} - t_i] - \sum_{i=1}^{n} f(z(\xi''_{[t''_{j(i)},t''_{j(i)+1}]}))z'(t''_{j(i)})[t_{i+1} - t_i]\right| \\
&+ \sum_{i=1}^{n''} |f(z(\xi''_{[t''_i,t''_{i+1}]}))|\epsilon[t''_{i+1} - t''_i].
\end{aligned}
$$

By denoting $C_1 := \max_{t\in[a,b]} |f(z(t))|$, we continue

$$
|R(M') - R(M'')| \leqslant 4C_1\epsilon(b - a) + 2\sum_{i=1}^{n} \max_{s,t\in[t'_{j(i)},t'_{j(i)+1}]} |f(z(s))z'(s) - f(z(t))z'(t)|[t_{i+1} - t_i].
$$

By denoting $C_2 := \max_{t\in[a,b]} |z'(t)|$, we have

$$
\begin{aligned}
&\max_{s,t\in[t'_{j(i)},t'_{j(i)+1}]} |f(z(s))z'(s) - f(z(t))z'(t)| \\
\leqslant&\max_{s,t\in[t'_{j(i)},t'_{j(i)+1}]} |f(z(s))z'(s) - f(z(s))z'(t)| + |f(z(s))z'(t) - f(z(t))z'(t)| \\
\leqslant&\max_{s,t\in[t'_{j(i)},t'_{j(i)+1}]} C_1|z'(s) - z'(t)| + C_2|f(z(s)) - f(z(t))| \\
\leqslant&C_1\epsilon + C_2\epsilon.
\end{aligned}
$$

Now we have

$$
\begin{aligned}
|R(M') - R(M'')| \leqslant&4C_1\epsilon(b - a) + 2\sum_{i=1}^{n}(C_1\epsilon + C_2\epsilon)[t_{i+1} - t_i] \\
=&4C_1\epsilon(b - a) + 2(C_1\epsilon + C_2\epsilon)(b - a) = \epsilon(6C_1 + 2C_2)(b - a).
\end{aligned}
$$

Since $\epsilon > 0$ was arbitrary, it follows that the Riemann sums are Cauchy and hence converge by completeness of $\mathbb{C}$.

8.1.3 Line Integral for B-valued Functions

Now suppose $L : \gamma \to B$ is continuous. Then the *line integral* $\int \gamma L(z)\mathrm{d}z$ on curve γ is defined as the limit of the Riemann sums

$$\sum_{i=1}^{n} L(z(\xi_i))[z(t_{i+1}) - z(t_i)],$$

where $a = t_1 < t_2 < \cdots < t_{n+1} = b$, $\xi_i \in [t_i, t_{i+1}]$, and the limit in B is taken as $\max_{i=1,\cdots,n} |t_{i+1} - t_i| \to 0$.

To see the Riemann sums are Cauchy, we proceed exactly the same way as we did for line integral for $\mathbb{C}$-valued functions, replacing $|f(z(t))|$ by $\|L(z(t))\|$. Then, completeness of B gives convergence.

If $\rho : [a, b] \to [a, b]$, $\rho' > 0$, is a reparametrization, then the line integral

$$\int_\gamma L(\tilde{z})\mathrm{d}\tilde{z} = \int_a^b L(\tilde{z}(t))\tilde{z}'(t)\mathrm{d}t$$

of L on curve γ reparametrized by $\tilde{z} = z \circ \rho$ coincides with

$$\int_\gamma L(z)\mathrm{d}z = \int_a^b L(z(t))z'(t)\mathrm{d}t$$

by change of variables $s = \rho(t)$. Indeed,

$$\int_\gamma L(\tilde{z})\mathrm{d}\tilde{z} = \int_a^b L(\tilde{z}(t))\tilde{z}'(t)\mathrm{d}t = \int_a^b L(z \circ \rho(t))z'(\rho(t))\rho'(t)\mathrm{d}t = \int_a^b L(z(s))z'(s)\mathrm{d}s = \int_\gamma L(z)\mathrm{d}z.$$

This shows that the line integral $\int_\gamma L(z)\mathrm{d}z$ of L on curve γ is independent of the smooth parametrization z.

Lemma 8.1 (Ex 3.1). *Suppose $L : \gamma \to B$ is continuous. For any $\varphi \in \mathcal{L}^*$, we have*

$$\varphi\left[\int_\gamma L(z)\mathrm{d}z\right] = \int_\gamma \varphi[L(z)]\mathrm{d}z.$$

For any $T \in B$, we have

$$T\left[\int_\gamma L(z)\mathrm{d}z\right] = \int_\gamma T[L(z)]\mathrm{d}z.$$

Proof. As φ and T are both continuous, it is easy to see that they commute with the limit of the Riemann sums. □

8.1.4 Differentiation

Theorem 8.2 (Analyticity Theorem). *For any complex family $L : U \to B$ of bounded linear operators $L_z : \mathcal{L} \to \mathcal{L}$, $z \in U \subseteq \mathbb{C}$, on a complex Banach space $\mathcal{L}$, the following two notions of analyticity are equivalent.*

1. ***Weak Analyticity.*** *For any $\varphi \in B^*$, the function $\varphi[L(\cdot)] : U \to \mathbb{C}$ is holomorphic.*
2. ***Strong Analyticity.*** *For every $z \in U$, there is some $L'(z) \in B$, called the "derivative of L at z", such that*

$$\left\| \frac{L(z+h) - L(z)}{h} - L'(z) \right\| \xrightarrow{|h| \to 0} 0$$

A proof is given in Sarig's Appendix, and we will prove it later.

Lemma 8.3 (Rules of Differentiation, Ex 3.2). *Suppose $L, L_1, L_2 : U \to B$ are analytic.*

1. $(L_1(z) + L_2(z))' = L_1'(z) + L_2'(z)$;

2. $(L_1(z) \circ L_2(z))' = L_1'(z) \circ L_2(z) + L_1(z) \circ L_2'(z)$;

3. *when L_z is invertible for every $z \in U$, then* $(L_z^{-1})' = -L_z^{-1} \circ L_z' \circ L_z^{-1}$;

4. *if $\varphi \in B^*$, then* $(\varphi \circ L_z)' = \varphi \circ L_z'$.

Proof. **1.** Note

$$\begin{aligned}
&\left\| \frac{(L_1 + L_2)(z+h) - (L_1 + L_2)(z)}{h} - (L_1'(z) + L_2'(z)) \right\| \\
=& \left\| \frac{L_1(z+h) - L_1(z)}{h} - L_1'(z) + \frac{L_2(z+h) - L_2(z)}{h} - L_2'(z) \right\| \\
\leqslant& \left\| \frac{L_1(z+h) - L_1(z)}{h} - L_1'(z) \right\| + \left\| \frac{L_2(z+h) - L_2(z)}{h} - L_2'(z) \right\| \xrightarrow{|h| \to 0} 0.
\end{aligned}$$

2. Note

$$\begin{aligned}
&\left\| \frac{L_1(z+h) \circ L_2(z+h) - L_1(z) \circ L_2(z)}{h} - (L_1'(z) \circ L_2(z) + L_1(z) \circ L_2'(z)) \right\| \\
\leqslant& \left\| \frac{L_1(z+h) \circ L_2(z+h) - L_1(z+h) \circ L_2(z)}{h} - L_1(z+h) \circ L_2'(z) \right\| + \| L_1(z+h) \circ L_2'(z) - L_1(z) \circ L_2'(z) \| \\
&+ \left\| \frac{L_1(z+h) \circ L_2(z) - L_1(z) \circ L_2(z)}{h} - L_1'(z) \circ L_2(z) \right\| \\
\leqslant& \| L_1(z+h) \| \left\| \frac{L_2(z+h) - L_2(z)}{h} - L_2'(z) \right\| + \| L_1(z+h) - L_1(z) \| \| L_2'(z) \| \\
&+ \left\| \frac{L_1(z+h) - L_1(z)}{h} - L_1'(z) \right\| \| L_2(z) \| \xrightarrow{|h| \to 0} 0.
\end{aligned}$$

3. Note

$$\begin{aligned}
&\left\| \frac{L^{-1}(z+h) - L^{-1}(z)}{h} + L^{-1}(z) \circ L'(z) \circ L^{-1}(z) \right\| \\
\leqslant& \left\| \frac{L^{-1}(z+h) - L^{-1}(z)}{h} + L^{-1}(z+h) \circ L'(z) \circ L^{-1}(z) \right\| \\
&+ \| L^{-1}(z) \circ L'(z) \circ L^{-1}(z) - L^{-1}(z+h) \circ L'(z) \circ L^{-1}(z) \| \\
=& \left\| L^{-1}(z+h) \circ \left[\frac{L(z) - L(z+h)}{h} + L'(z) \right] \circ L^{-1}(z) \right\| + \| [L^{-1}(z) - L^{-1}(z+h)] \circ L'(z) \circ L^{-1}(z) \| \\
\leqslant& \| L^{-1}(z+h) \| \left\| \frac{L(z+h) - L(z)}{h} - L'(z) \right\| \| L^{-1}(z) \| + \| L^{-1}(z) - L^{-1}(z+h) \| \| L'(z) \| \| L^{-1}(z) \|.
\end{aligned}$$

Since every L_z, $z \in U$ is invertible, so is every $L(z+h)$ for small h by openness of U. It follows by Open Mapping Theorem that these inverses are also bounded linear operators on $\mathcal{L}$. Hence, the above estimates tend to 0 as $|h| \to 0$, by definition of $L'(z)$. This proves the formula.

4. Note

$$\left\| \frac{\varphi \circ L(z+h) - \varphi \circ L(z)}{h} - \varphi \circ L'(z) \right\| \leqslant \| \varphi \| \left\| \frac{L(z+h) - L(z)}{h} - L'(z) \right\| \xrightarrow{|h| \to 0} 0.$$

□

Theorem 8.4 (Cauchy Integral Formula). *If $L : U \to B$ is analytic on U, then L is differentiable infinitely many times on U. Moreover, for any $z \in U$ and any simple closed smooth curve $\gamma \subseteq U$ around z, we have*

$$L(z) = \frac{1}{2\pi i} \int_\gamma \frac{L(\xi)}{\xi - z} \mathrm{d}\xi, \quad \textit{and} \quad L^{(n)}(z) = \frac{n!}{2\pi i} \int_\gamma \frac{L(\xi)}{(\xi - z)^{n+1}} \mathrm{d}\xi.$$

Proof. By weak analyticity, $\varphi[L(\cdot)] : U \to \mathbb{C}$ is holomorphic for any $\varphi \in B^*$. By Exercise 3.2.4 and strong analyticity, we have

$$\frac{\mathrm{d}}{\mathrm{d}z}\varphi[L(z)] = \varphi[L'(z)]$$

is holomorphic too. In other words, $L'(z)$ is (weak) analytic. Induction gives that $L(z)$ is differentiable infinitely many times on U.

For any $\varphi \in B^*$, note

$$\varphi\left[\frac{1}{2\pi i}\int_\gamma \frac{L(\xi)}{\xi - z}\mathrm{d}\xi\right] = \frac{1}{2\pi i}\int_\gamma \frac{\varphi[L(\xi)]}{\xi - z}\mathrm{d}\xi = \varphi[L(z)],$$

by the usual Cauchy Integral Formula for the holomorphic function $\varphi[L(z)] : U \to \mathbb{C}$. Since this equality holds for all $\varphi \in B^*$, and bounded linear functionals separate points, see Proposition B.8, it follows that

$$\frac{1}{2\pi i}\int_\gamma \frac{L(\xi)}{\xi - z}\mathrm{d}\xi = L(z).$$

The formula for higher derivatives is proved exactly the same way using the usual Cauchy Integral Formula for holomorphic function $\varphi[L(z)] : U \to \mathbb{C}$. ☐

Proposition 8.5 (Ex 3.3). *If $L : U \to B$ is analytic and γ is a simple closed smooth contractible curve in U, then*

$$\int_\gamma L(z)\mathrm{d}z = 0.$$

Proof. Fix any $\varphi \in B^*$. By weak analyticity, the function $\varphi[L(\cdot)] : U \to \mathbb{C}$ is holomorphic. It then follows from Exercise 3.1 and a basic property of holomorphic functions that

$$\varphi[\int_\gamma L(z)\mathrm{d}z] = \int_\gamma \varphi[L(z)]\mathrm{d}z = 0 = \varphi[0].$$

Since this holds for any $\varphi \in B^*$ and bounded linear functionals separate points, the asserting follows. ☐

Proposition 8.6 (Ex 3.4). *If a sequence $\{T_n\}_{n\geqslant 0} \subseteq B$ satisfies $\|T_n\| = O(r^n)$ for some $r > 0$, then the power series*

$$\sum_{n\geqslant 0}(z-a)^n T_n$$

converges and is analytic on $\{z \in U : |z-a| < 1/r\}$.

Proof. The big-O notation means that there are $C, N > 0$ such that

$$\|T_n\| \leqslant Cr^n, \quad \forall n \geqslant N.$$

Fix any $z \in U$ with $|z-a| < 1/r$. Then,

$$\sum_{n\geqslant 0}\|(z-a)^n T_n\| = \sum_{n=0}^{N-1}|z-a|^n\|T_n\| + \sum_{n\geqslant N}|z-a|^n\|T_n\| \leqslant \sum_{n=0}^{N-1}|z-a|^n\|T_n\| + \sum_{n\geqslant N}|z-a|^n Cr^n.$$

Since $|z-a|r < 1$ by choice of z, we conclude the power series converges absolutely on $\{z : |z-a| < 1/r\}$. Write $T(z) := \sum_{n\geqslant 0}(z-a)^n T_n$ for $z \in U$ with $|z-a| < 1/r$. Fix $\varphi \in B^*$. By continuity and linearity of φ, we have

$$\varphi[T(z)] = \sum_{n\geqslant 0}(z-a)^n\varphi[T_n].$$

Since the power series for $\varphi[T(\cdot)]$ converges absolutely

$$\sum_{n\geqslant 0}|(z-a)^n\varphi[T_n]| \leqslant \sum_{n\geqslant 0}|z-a|^n\|\varphi\|\|T_n\| = \|\varphi\|\sum_{n\geqslant 0}\|(z-a)^n T_n\| < +\infty$$

on $\{z \in U : |z-a| < 1/r\}$, it follows from properties of holomorphic functions and power series that $\varphi[T(\cdot)]$ is holomorphic on $\{z \in U : |z-a| < 1/r\}$. This implies T is (weak) analytic on $\{z \in U : |z-a| < 1/r\}$. ☐

Proposition 8.7 (Ex 3.5). *A family $L : U \to B$ is analytic on open set $U \subseteq \mathbb{C}$ if and only if for each $a \in U$, there is a sequence $L_n(a) \in B$, $n \geqslant 0$, and a radius $r(a) > 0$ such that*

$$\|L_n(a)\| = O(r(a)^{-n}) \quad \text{and} \quad L(z) = \sum_{n \geqslant 0} (z-a)^n L_n(a) \ \text{ on } \ \{z \in U : |z-a| < r(a)\}.$$

Proof. ($\Leftarrow$) By Exercise 3.4, we have L analytic on $\{z \in U : |z-a| \leqslant 1/r(a)\}$ for each $a \in U$. This implies that L is analytic on the entire open set U.

($\Rightarrow$) Fix $a \in U$ and take $R(a) > 0$ with $\overline{\mathbb{D}(a, R(a))} \subseteq U$. Define

$$F(z) := L(z+a), \qquad z \in \mathbb{D}(0, R(a)).$$

Note F is analytic on $\mathbb{D}(0, R(a))$. Fix $z \in \mathbb{D}(0, R(a))$ and let $r := \frac{|z| + R(a)}{2}$. For any $\xi \in \partial\mathbb{D}(0, r)$, we have $\left|\frac{z}{\xi}\right| < 1$. So

$$\frac{1}{\xi - z} = \frac{1}{\xi} \frac{1}{1 - \frac{z}{\xi}} = \frac{1}{\xi} \sum_{n \geqslant 0} \left(\frac{z}{\xi}\right)^n.$$

By Cauchy Integral Formula, we have

$$\begin{aligned} F(z) = & \frac{1}{2\pi i} \int_{\partial\mathbb{D}(0,r)} \frac{F(\xi)}{\xi - z} \mathrm{d}\xi = \frac{1}{2\pi i} \int_{\partial\mathbb{D}(0,r)} F(\xi) \frac{1}{\xi} \sum_{n \geqslant 0} \left(\frac{z}{\xi}\right)^n \mathrm{d}\xi \\ = & \int_{\partial\mathbb{D}(0,r)} \sum_{n \geqslant 0} \frac{1}{2\pi i} F(\xi) \frac{1}{\xi} \left(\frac{z}{\xi}\right)^n \mathrm{d}\xi. \end{aligned}$$

It is easy to see that $\sum_{n \geqslant 0} \frac{1}{2\pi i} F(\xi) \frac{1}{\xi} \left(\frac{z}{\xi}\right)^n$ converges absolutely and uniformly on $\partial\mathbb{D}(0, r)$ and hence we may interchange the limit and integral. We continue

$$F(z) = \sum_{n \geqslant 0} z^n \int_{\partial\mathbb{D}(0,r)} \frac{1}{2\pi i} \frac{F(\xi)}{\xi^{n+1}} \mathrm{d}\xi.$$

Now returning to L, we have

$$L(z) = F(z-a) = \sum_{n \geqslant 0} (z-a)^n \int_{\partial\mathbb{D}(a,r)} \frac{1}{2\pi i} \frac{L(\xi)}{(\xi - a)^{n+1}} \mathrm{d}\xi \ \text{ on } \ \{z \in U : |z-a| < R(a)\}.$$

Take $L_n(a) := \int_{\partial\mathbb{D}(a,r)} \frac{1}{2\pi i} \frac{L(\xi)}{(\xi - a)^{n+1}} \mathrm{d}\xi$. Note $\frac{R(a)}{2} \leqslant r < R(a)$ and hence

$$\|L_n(a)\| \leqslant 2\pi r \frac{1}{2\pi} r^{-(n+1)} \max_{\xi \in \partial\mathbb{D}(a,r)} \|L(\xi)\| \leqslant \left(\frac{R(a)}{2}\right)^{-n} \max_{\xi \in \overline{\mathbb{D}(a,R(a))}} \|L(\xi)\|.$$

This shows $\|L_n(a)\| = O(r(a)^{-n})$ where $r(a) = R(a)/2$. The proof is complete. □

8.2 The Resolvent and Eigenprojections

Recall the spectrum of $L \in B$ is defined as

$$\operatorname{spec}(L) := \{\lambda \in \mathbb{C} : (\lambda I - L) \text{ has no (bounded) inverse}\}.$$

Proposition 8.8 (Ex 3.6). *The spectrum* $\operatorname{spec}(L)$ *of any $L \in B$ is a compact set in $\mathbb{C}$.*

Proof. **1. If $\|L\| < 1$, then $(I - L)$ has a bounded inverse given by**

$$(I - L)^{-1} = I + L + L^2 + \cdots.$$

Indeed, note $\sum_{i=0}^{\infty} \|L^i\| \leqslant \sum_{i=0}^{\infty} \|L\|^i < \frac{1}{1-\|L\|} < +\infty$, and hence the partial sums

$$L_n := \sum_{i=0}^{n} L^i$$

are Cauchy. It follows from the completeness of Banach space B that L_n converges to some $L_\infty \in B$, and in particular, $L^n \to 0$. On the other hand,

$$\begin{aligned}(I-L)L_\infty &= \lim_{n\to+\infty}(I-L)L_n = \lim_{n\to+\infty} L_n - LL_n = \lim_{n\to+\infty}(I+L+L^2+\cdots+L^n)-(L+L^2+\cdots+L^n+L^{n+1}) \\ &= I - \lim_{n\to+\infty} L^{n+1} = I\end{aligned}$$

A similar argument shows that $L_\infty(I-L) = I$. We conclude $L_\infty = (I-L)^{-1}$.

2. $(zI - L)$ **has a bounded inverse provided** $|z|$ **sufficiently large.**

For $|z| > 0$, write $zI - L = z(I - z^{-1}L)$. When $|z| > \|L\|$, we have $\|z^{-1}L\| = |z|^{-1}\|L\| < 1$ and hence, by item 1, $(I - z^{-1}L)$ has a bounded inverse; in this case, so does $(zI - L)$.

It follows that $\mathrm{spec}(L) \subseteq \overline{\mathbb{D}(0, \|L\|)}$. In other words, $\mathrm{spec}(L)$ is bounded[11] in $\mathbb{C}$. It remains to show that $\mathrm{spec}(L)$ is closed in $\mathbb{C}$.

3. Show that if $(I - L)$ has a bounded inverse, then so does any $(I - L_1)$ with $\|L_1 - L\|$ sufficiently small.

We show that for any invertible $F \in B(\mathcal{L})$, we have $F + G$ also invertible, provided $\|G\|$ sufficiently small. Indeed, note

$$F + G = F(I + F^{-1}G).$$

When $\|G\| < \|F^{-1}\|^{-1}$ (note $\|F^{-1}\| < +\infty$ by Open Mapping Theorem), we have

$$\|F^{-1}G\| \leqslant \|F^{-1}\|\|G\| < 1,$$

and hence $(I + F^{-1}G)$ is invertible by item 1. This implies $F + G$ is invertible too.

In fact, if f is any Lipeomorphism on a Banach space X and g is a Lipschitz function with sufficiently small Lipschitz constant, then $f + g$ is another Lipeomorphism. Indeed, since

$$f + g = f \circ (\mathrm{id} + f^{-1} \circ g),$$

it follows that when

$$\mathrm{Lip}(g) < (\mathrm{Lip}(f^{-1}))^{-1},$$

we have $\mathrm{Lip}(f^{-1} \circ g) \leqslant \mathrm{Lip}(f^{-1})\mathrm{Lip}(g) < 1$, and hence $(\mathrm{id} + f^{-1} \circ g)$ is a Lipeomorphism by Lipschitz Inverse Function Theorem (a consequence of the Banach Fixed Point Theorem).

4. Take a sequence $\{z_n\}_n \subseteq \mathrm{spec}(L)$ with $z_n \to z$ for some $z \in \mathbb{C}$, and we show $z \in \mathrm{spec}(L)$.

If $z = 0$, then we need to show that $(0I - L) = -L$ has no bounded inverse, or equivalently, that L has no bounded inverse. Suppose the contrary. Since $(z_nI - L)$ has no bounded inverse, it follows that

$$(z_nI - L)L^{-1} = z_nL^{-1} - I$$

has no bounded inverse; otherwise, $L^{-1}((z_nI - L)L^{-1})^{-1}$ would be a bounded inverse for $(z_nI - L)$. Equivalently, $I - z_nL^{-1}$ has no bounded inverse. Since

$$\|z_nL^{-1} - 0\| = |z_n|\|L^{-1}\| \xrightarrow{n\to+\infty} |z|\|L^{-1}\| = 0,$$

it follows from item 3 that $I = I - 0$ has no bounded inverse, a contradiction. We conclude in this case L has no bounded inverse, as desired.

[11] It is worth noting that Sarig purposefully proved boundedness of the spectrum from scratch, rather than make use of estimates $\rho(L) = \lim_{n\to+\infty} \sqrt[n]{\|L^n\|} = \inf_{n\geqslant 1} \sqrt[n]{\|L^n\|}$, in order to avoid circular arguments. More specifically, the proof of this limiting equality hinges on the analycity of the resolvent, which in turn assumes compactness of the spectrum.

When $z \neq 0$, we assume without loss of generality that $z_n \neq 0$ for all n. If $\left(I - \frac{L}{z_n}\right)$ were to have a bounded inverse, then $z_n^{-1}\left(I - \frac{L}{z_n}\right)^{-1}$ would be a bounded inverse for $(z_n I - L)$, contradicting $z_n \in \text{spec}(L)$; we thus conclude that each $\left(I - \frac{L}{z_n}\right)$ has no bounded inverse. But since

$$\left\| \frac{L}{z_n} - \frac{L}{z} \right\| = \left| \frac{1}{z_n} - \frac{1}{z} \right| \|L\| \xrightarrow{n \to +\infty} 0,$$

it follows from item 3 that $\left(I - \frac{L}{z}\right)$ has no bounded inverse. In other words, $z \in \text{spec}(L)$. This shows $\text{spec}(L)$ is closed and completes the proof. ☐

Definition 8.9 (Resolvent)**.** Given a bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ on Banach space $\mathcal{L}$, on the complement of its spectrum, we define the resolvent

$$R : \mathbb{C} \backslash \text{spec}(L) \to B, \quad z \mapsto (zI - L)^{-1}.$$

Proposition 8.10 (Properties of the Resolvent, Ex 3.7)**.** *1.* ***Commutation.*** *$R(z)L = LR(z)$.*

2. ***Resolvent Identity.*** *$R(w) - R(z) = (z - w)R(z)R(w)$.*
3. ***Neumann's Expansion.*** *$R(z) = \sum_{n=0}^{\infty} (-1)^n (z - z_0)^n R(z_0)^{n+1}$, for any $z_0 \in \mathbb{C} \backslash \text{spec}(L)$ and z sufficiently close to z_0.*
4. ***Analyticity.*** *The resolvent $R : \mathbb{C} \backslash \text{spec}(L) \to B$ is an analytic function.*

Remark 8.11. It makes sense to speak of the analyticity of the resolvent, because we know $\mathbb{C} \backslash \text{spec}(L)$ is a nonempty open set from compactness of $\text{spec}(L)$.

Proof. **1.** Note

$$R(z)L = (zI - L)^{-1}L = (zI - L)^{-1}L(zI - L)(zI - L)^{-1} = (zI - L)^{-1}(zI - L)L(zI - L)^{-1} = LR(z).$$

2. Note

$$\begin{aligned} R(w) - R(z) =& (wI - L)^{-1} - (zI - L)^{-1} = (zI - L)^{-1}[(zI - L) - (wI - L)](wI - L)^{-1} \\ =& (zI - L)^{-1}[z - w](wI - L)^{-1} = (z - w)R(z)R(w). \end{aligned}$$

3. Fix any $z_0 \in \mathbb{C} \backslash \text{spec}(L)$. Since $\text{spec}(L)$ is closed in $\mathbb{C}$, it follows that its complement $\mathbb{C} \backslash \text{spec}(L)$ is open. So take $z \in \mathbb{C} \backslash \text{spec}(L)$ so close to z_0 that

$$|z - z_0| \|R(z_0)\| < 1.$$

Now we start from the Resolvent Identity

$$\begin{aligned} R(z) - R(z_0) =& (z_0 - z)R(z_0)R(z) \\ R(z) - (z_0 - z)R(z_0)R(z) =& R(z_0) \\ [I - (z_0 - z)R(z_0)]R(z) =& R(z_0) \end{aligned}$$

By choice of z, we have

$$\|(z_0 - z)R(z_0)\| = |z_0 - z| \|R(z_0)\| < 1,$$

and hence, by Ex 3.6.1, the operator $[I - (z_0 - z)R(z_0)]$ has a bounded inverse given by the fundamental series expansion.

$$\begin{aligned} R(z) =& [I - (z_0 - z)R(z_0)]^{-1}R(z_0) \\ =& \left[\sum_{n=0}^{\infty} (z_0 - z)^n R(z_0)^n \right] R(z_0) \\ =& \sum_{n=0}^{\infty} (-1)^n (z - z_0)^n R(z_0)^{n+1}. \end{aligned}$$

4. By choice of z from item 3, we have $T_n := (-1)^n R(z_0)^{n+1}$ satisfying

$$\|T_n\| = \left\|(-1)^n R(z_0)^{n+1}\right\| \leqslant \|R(z_0)\|^{n+1} < \|R(z_0)\||z - z_0|^{-n}.$$

By Ex 3.5, we conclude R analytic on disks $\{z \in \mathbb{C}\backslash\mathrm{spec}(L) : |z - z_0| < \|R(z_0)\|^{-1}\}$ for any $z_0 \in \mathbb{C}\backslash\mathrm{spec}(L)$. In other words, R is analytic on $\mathbb{C}\backslash\mathrm{spec}(L)$, as required. □

Proposition 8.12 (Spectrum is Nonempty). *The spectrum* $\mathrm{spec}(L)$ *of any bounded linear operator* $L : \mathcal{L} \to \mathcal{L}$ *on a Banach space* $\mathcal{L}$ *is nonempty.*

Proof. Suppose the contrary. Then, the resolvent R is defined on the entire $\mathbb{C}$ and is thus is an entire function. Also, for any $|z| > \|L\|$, we have

$$R(z) = (zI - L)^{-1} = [z(I - z^{-1}L)]^{-1} = z^{-1}(I - z^{-1}L)^{-1}.$$

As $z \to \infty$, note $z^{-1} \to 0$, $z^{-1}L \to 0$, $(I - z^{-1}L)^{-1} \to I$ and hence $R(z) \to 0$. This shows R vanishes at ∞, and in particular, must be bounded. Liouville's Theorem thus implies R is a constant function.

However,

$$\frac{\mathrm{d}}{\mathrm{d}z}R(z) = -(zI - L)^{-1} \circ I \circ (zI - L)^{-1} = -(zI - L)^{-2} \not\equiv 0,$$

a contradiction. We conclude $\mathrm{spec}(L) \neq \varnothing$. □

Proposition 8.13 ([Con85] VII.3.8). *For any bounded linear operator* $L : \mathcal{L} \to \mathcal{L}$ *on a Banach space* $\mathcal{L}$*, the limit* $\lim_{n\to+\infty} \|L^n\|^{1/n}$ *exists and equals* $\rho(L)$.

Proof. Let $G := \{z \in \mathbb{C} : z^{-1} \in \mathbb{C}\backslash\mathrm{spec}(L)\} \cup \{0\}$, and define

$$f : G\backslash\{0\} \to B(\mathcal{L}), \qquad z \mapsto (z^{-1}I - L)^{-1}.$$

As $\lambda \to \infty$, note $(I - \lambda^{-1}L) \to I$, $\lambda^{-1} \to 0$ and hence $(\lambda I - L)^{-1} = \lambda^{-1}(I - \lambda^{-1}L)^{-1} \to 0$. This shows $f(z) \to 0$ as $z \to 0$, and hence 0 is a removable singularity. By defining $f(0) = 0$, we have f analytic on G.

Take its power series expansion for $|z| < \|L\|^{-1}$

$$f(z) = (z^{-1} - L)^{-1} = [z^{-1}(I - zL)]^{-1} = z(I - zL)^{-1} = z\sum_{n=0}^{\infty} z^n L^n.$$

The largest radius of convergence for this power series is

$$\begin{aligned} R =& \mathrm{dist}(0, \partial G) = \mathrm{dist}(0, \mathrm{spec}(L)^{-1}), \qquad \mathrm{spec}(L)^{-1} := \{z^{-1} : z \in \mathrm{spec}(L)\}. \\ =& \inf\{|z| : z^{-1} \in \mathrm{spec}(L)\} = \rho(L)^{-1}. \end{aligned}$$

On the other hand, by Sarig Exercise 3.5, we have $\|L^n\| = O(R^{-n})$, that is,

$$\|L^n\| \leqslant CR^{-n}, \qquad \text{for some } C > 0.$$

This yields $R^{-1} \geqslant C^{-1}\|L^n\|^{1/n}$ for every $n \geqslant 1$, and hence

$$\rho(L) = R^{-1} = \limsup_{n\to+\infty} \|L^n\|^{1/n}.$$

Now if $z \in \mathbb{C}$ and $n \geqslant 1$, then

$$z^n I - L^n = (zI - L)(z^{n-1}I + z^{n-2}L + \cdots + L^{n-1}) = (z^{n-1}I + z^{n-2}L + \cdots + L^{n-1})(zI - L).$$

If $(z^n I - L^n)$ is invertible, then $(z^n I - L^n)^{-1}(z^{n-1}I + z^{n-2}L + \cdots + L^{n-1})$ is the inverse of $(zI - L)$. For any $z \in \mathrm{spec}(L)$, we have $(zI - L)$ non-invertible, and hence $z^n I - L^n = z^n(I - z^{-n}L^n)$ is also non-invertible for all $n \geqslant 1$. This implies

$$\|z^{-n}L^n\| = |z|^{-n}\|L^n\| < 1, \qquad \forall z \in \mathrm{spec}(L),$$

and hence $|z| \leqslant \|L^n\|^{1/n}$. We conclude

$$\rho(L) \leqslant \liminf_{n\to\infty} \|L^n\|^{1/n} = \limsup_{n\to+\infty} \|L^n\|^{1/n} = \rho(L),$$

and hence the limit $\lim_{n\to+\infty} \|L^n\|^{1/n}$ exists and equals $\rho(L)$. □

Lemma 8.14 (Fekete's Subadditive Lemma). *If a sequence $\{a_n\}_{n\geqslant 1} \subseteq \mathbb{R}$ is subadditive, i.e.,*

$$a_{n+m} \leqslant a_n + a_m, \qquad \forall m, n \geqslant 1,$$

then the limit $\lim_{n\to+\infty} a^n/n$ exists and equals $\inf_{n\geqslant 1} a^n/n$.

Proof. Let $L := \inf_{n\geqslant 1} a_n/n$ and fix $L' > L$. Choose $k \geqslant 1$ with $a_k/k < L'$. For $n > k$, by Division Algorithm, there are integers $p_n \geqslant 1$ and $q_n \in [0, k-1]$ such that $n = p_n k + q_n$. By subadditivity, we have

$$a_n = a_{p_n k + q_n} \leqslant p_n a_k + a_{q_n}.$$

Dividing both sides by n yields

$$\frac{a_n}{n} \leqslant \frac{p_n k}{n}\frac{a_k}{k} + \frac{a_{q_n}}{n}.$$

As $n \to +\infty$, we have $\frac{p_n k}{n} \to 1$, $\frac{a_{q_n}}{n} \leqslant \frac{\max\{a_i : i=0,\cdots,k-1\}}{n} \to 0$, and hence

$$L \leqslant \lim_{n\to+\infty} \frac{a_n}{n} \leqslant \frac{a_k}{k} < L', \qquad \forall L' > L.$$

The assertion follows. □

Corollary 8.15. *For any bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ on a Banach space $\mathcal{L}$,*

$$\rho(L) = \lim_{n\to+\infty} \|L^n\|^{1/n} = \inf_{n\geqslant 1} \|L^n\|^{1/n}.$$

Proof. Since $\|L^{n+m}\| = \|L^n L^m\| \leqslant \|L^n\|\|L^m\|$, it follows that $\log\|L^n\|$ is subadditive. By Fekete's Subadditive Lemma, we have

$$\lim_{n\to+\infty} \frac{\log\|L^n\|}{n} = \inf_{n\geqslant 1} \frac{\log\|L^n\|}{n}.$$

Exponentiating both sides yields $\lim_{n\to+\infty} \|L^n\|^{1/n} = \inf_{n\geqslant 1} \|L^n\|^{1/n}$, as required. □

Theorem 8.16 (Separation of Spectrum). *Suppose the spectrum* $\mathrm{spec}(L)$ *of a bounded linear operator $L : \mathcal{L} \to \mathcal{L}$ on Banach space $\mathcal{L}$ admits a decomposition into the disjoint union*

$$\mathrm{spec}(L) = \Sigma_{\mathrm{in}} \dot{\cup} \Sigma_{\mathrm{out}}$$

of two compact pieces Σ_{in} and Σ_{out}, and γ is a smooth closed curve in $\mathbb{C}\backslash\mathrm{spec}(L)$ with Σ_{in} inside and Σ_{out} outside. Then,

1. *The operator defined by line integral*

$$P := \frac{1}{2\pi i}\int_\gamma (zI - L)^{-1} \mathrm{d}z$$

   is a projection, i.e., $P^2 = P$. Hence, $\mathcal{L} = \ker(P) \oplus \mathrm{Im}(P)$.

2. *$PL = LP$. So,* $\ker(P)$ *and* $\mathrm{Im}(P)$ *are both L-invariant.*

3. $\mathrm{spec}(L|_{\mathrm{Im}(P)}) = \Sigma_{\mathrm{in}}$ *and* $\mathrm{spec}(L|_{\ker(P)}) = \Sigma_{\mathrm{out}}$.

A proof is given in Sarig's Appendix, and we will prove it later.

Definition 8.17 (Eigenprojection)**.** The projection operator $P = \frac{1}{2\pi i}\int_\gamma (zI - L)^{-1}\mathrm{d}z$ is called the *eigenprojection of* Σ_{in}.

Corollary 8.18. *When L has a spectral gap with representation $L = \lambda P + N$, the eigenprojection P_λ of $\Sigma_{\text{in}} = \{\lambda\}$ equals P.*

Proof. Note λ is a simple eigenvalue of $L : \mathcal{L} \to \mathcal{L}$ with one-dimensional eigenspace $E(\lambda) = \mathrm{Im}(P)$, and $\mathcal{L} = \ker(P) \oplus \mathrm{Im}(P)$. We claim $\mathrm{Im}(P_\lambda) \subseteq \mathrm{Im}(P)$.

Suppose not. Then, $(I-P)\mathrm{Im}(P_\lambda) \neq \{0\}$ is a closed nontrivial L-invariant linear subspace. Now $\rho(N) < |\lambda|$ implies $\lambda \notin \mathrm{spec}(N|_{\ker(P)}) = \mathrm{spec}(L|_{\ker(P)})$; it then follows from $(I-P)\mathrm{Im}(P_\lambda) \subseteq \ker(P)$ that $\mathrm{spec}(L|_{(I-P)\mathrm{Im}(P_\lambda)}) \subseteq \mathrm{spec}(L|_{\ker(P)})$ and hence

$$\lambda \notin \mathrm{spec}(L|_{(I-P)\mathrm{Im}(P_\lambda)}).$$

On the other hand, note $\mathrm{spec}(L|_{(I-P)\mathrm{Im}(P_\lambda)}) \subseteq \mathrm{spec}(L|_{\mathrm{Im}(P_\lambda)})$. But since[12]

$$\varnothing \neq \mathrm{spec}(L|_{(I-P)\mathrm{Im}(P_\lambda)}) \subseteq \mathrm{spec}(L|_{\mathrm{Im}(P_\lambda)}) = \Sigma_{\text{in}} = \{\lambda\},$$

we must have $\mathrm{spec}(L|_{(I-P)\mathrm{Im}(P_\lambda)}) = \{\lambda\}$, a contradiction. We conclude $\mathrm{Im}(P_\lambda) \subseteq \mathrm{Im}(P)$. But since $\mathrm{spec}(L|_{\mathrm{Im}(P_\lambda)}) = \Sigma_{\text{in}} = \{\lambda\}$, it follows that $\mathrm{Im}(P_\lambda)$ is nontrivial and hence must coincide with the one-dimensional space $\mathrm{Im}(P)$.

Now that P and P_λ are both projections to the same space $E(\lambda)$, we conclude $P = P_\lambda$, as desired. □

More generally, if $P^2 = P$, $Q^2 = Q$ and $\mathrm{Im}(P) = \mathrm{Im}(Q)$, then $P = Q$.
Indeed, any $v \in \mathcal{L}$ can be written as

$$v = Pv + (I-P)v,$$

where $Pv \in \mathrm{Im}(P)$ and $(I-P)v \in \ker(P)$. This shows $\mathcal{L} = \ker(P) + \mathrm{Im}(P)$. To see the sum is direct, take any $v \in \ker(P) \cap \mathrm{Im}(P)$. So $Pv = 0$ and $v = Pu$ for some $u \in \mathcal{L}$, hence $v = Pu = P^2u = P(Pu) = Pv = 0$, We conclude $\mathcal{L} = \ker(P) \oplus \mathrm{Im}(P)$. But the same argument for Q yields

$$v = Qv + (I-Q)v,$$

where $Qv \in \mathrm{Im}(Q) = \mathrm{Im}(P)$ and so we must have $Pv = Qv$ because the sum is direct (the representation is unique). This shows $Pv = Qv$ for any $v \in \mathcal{L}$ and so $P = Q$.

8.3 Analytic Perturbation Theorem

Theorem 8.19 (Analytic Perturbation Theorem)**.** *Let $U \ni 0$ be an open subset of $\mathbb{C}$ and $L : U \to B(\mathcal{L}), z \mapsto L_z$ be an analytic family of bounded linear operators $L_z : \mathcal{L} \to \mathcal{L}$ on a Banach space $\mathcal{L}$. If L_0 has spectral gap, then there is some $\epsilon > 0$ such that L_z has spectral gap whenever $|z| < \epsilon$. Moreover, there are λ_z, P_z, N_z analytic on $\{z \in \mathbb{C} : |z| < \epsilon\}$ such that*

1. *$L_z = \lambda_z P_z + N_z$;*
2. *$P_z \in B(\mathcal{L})$ with $P_z^2 = P_z$ and $\dim(\mathrm{Im}(P_z)) = 1$;*
3. *$P_z N_z = N_z P_z = 0$;*
4. *$\rho(N_z) < |\lambda_z| - \kappa$ for some $\kappa > 0$ independent of $z \in \{z \in \mathbb{C} : |z| < \epsilon\}$.*

Proof. Since L_0 has spectral gap, we have

$$\mathrm{spec}(L_0) = \{\lambda_0\} \dot\cup \Sigma, \qquad \Sigma \subseteq \{z : |z| < \rho(L_0) = |\lambda_0|\}.$$

Take a small circle $\gamma \subseteq \mathbb{C}\backslash\mathrm{spec}(L_0)$ with λ_0 inside γ and Σ outside γ.

Step 1. There is some $\epsilon_1 > 0$ such that $\gamma \subseteq \mathbb{C}\backslash\mathrm{spec}(L_z)$ for all z with $|z| < \epsilon_1$.

[12] The spectrum $\mathrm{spec}(L)$ must be nonempty, because otherwise the resolvent $R : z \mapsto (zI - L)^{-1}$ would be defined on $\mathbb{C}$, and so is an entire function. It is not difficult to show that R vanishes at ∞ and so is bounded; Liouville's Theorem implies R must be constant, contradicting the fact that R has nonzero first derivative $R' \neq 0$. For more details, see [Con85] Theorem VII.3.6.

Indeed, by choice of γ, we have that $\xi I - L_0$ has a bounded inverse for all $\xi \in \gamma$. Ex 3.6.3 says that having a bounded inverse is an open property in $B(\mathcal{L})$, that is, if F has a bounded inverse, then so does $F + G$, provided $\|G\|$ sufficiently small. This implies that

$$\Lambda := \{(\xi, z) \in \mathbb{C} \times \mathbb{C} : (\xi I - L_z) \text{ has a bounded inverse}\}$$

is an open neighborhood of compact set $\gamma \times \{0\}$. By compactness, we produce a uniform size $\epsilon_1 > 0$ for which $\gamma \times \{z : |z| < \epsilon_1\} \subseteq \Lambda$, as desired.

Step 2. For any $|z| < \epsilon_1$,

$$P_z := \frac{1}{2\pi i} \int_\gamma (\xi I - L_z)^{-1} \mathrm{d}\xi$$

is a projection and $P_z L_z = L_z P_z$. Moreover, there is $\epsilon_2 \in (0, \epsilon_1)$ such that P_z is analytic on $\{z : |z| < \epsilon_2\}$.

Indeed, fix z with $|z| < \epsilon_1$. Step 1 gives $\gamma \subseteq \mathbb{C} \backslash \mathrm{spec}(L_z)$.

If $\mathrm{spec}(L_z)$ does not intersect the region inside γ, then $(\xi I - L_z)^{-1}$ is well-defined and hence analytic on a region in which γ is contractible; we thus conclude $P_z = 0$ by Ex 3.3. Then, trivially we have $P_z^2 = 0^2 = 0 = P_z$ and $P_z L_z = 0 L_z = 0 = L_z 0 = L_z P_z$.

If $\mathrm{spec}(L_z)$ intersects the region inside γ, then Separation of Spectrum Theorem yields $P_z^2 = P_z$ and $P_z L_z = L_z P_z$.

On $\{z : |z| < \epsilon_1\}$, the family P_z is well-defined and hence analytic.

Step 3. There is $\epsilon_3 \in (0, \epsilon_2)$ such that $\dim(\mathrm{Im}(P_z)) = 1$ for all $|z| < \epsilon_3$.

We say $P, Q \in B(\mathcal{L})$ are *similar* if there is a linear isomorphism π of $\mathcal{L}$ such that $P = \pi^{-1} Q \pi$.

Kato Lemma: If projections $P, Q \in B(\mathcal{L})$ have $\|P - Q\| < 1$, then they are similar to each other.

By continuity (from analyticity) of P_z, there is some $\epsilon_3 \in (0, \epsilon_2)$ such that $\|P_z - P_0\| < 1$ whenever $|z| < \epsilon_3$. Kato's Lemma then yields linear isomorphism π_z of $\mathcal{L}$ with $P_z = \pi_z^{-1} P_0 \pi_z$ and hence $\dim(\mathrm{Im}(P_z)) = \dim(\mathrm{Im}(P_0)) = 1$.

Step 4. Definition of λ_z.

Take $|z| < \epsilon_3$. Since $P_z L_z = L_z P_z$ from Step 2, we have $L(\mathrm{Im}(P_z)) \subseteq \mathrm{Im}(P_z)$. Since $\dim(\mathrm{Im}(P_z)) = 1$, it follows that linear map

$$L_z : \mathrm{Im}(P_z) \to \mathrm{Im}(P_z)$$

takes the form $f \mapsto \lambda_z f$ for some $\lambda_z \in \mathbb{C}$. This shows $L_z P_z = \lambda_z P_z$.

To see λ_z depends analytically on z, take any $f \in \mathcal{L} \backslash \ker(P_0)$. By Hahn-Banach Theorem, there is some $\varphi \in \mathcal{L}^*$ such that $\varphi(P_0 f) > 0$. By continuity (from analyticity) of P_z, there is $\epsilon_4 \in (0, \epsilon_3)$ such that

$$\varphi(P_z f) > 0, \qquad \forall |z| < \epsilon_4.$$

Now the expression

$$\lambda_z = \frac{\varphi(L_z P_z f)}{\varphi(P_z f)}$$

shows that λ_z is analytic on $\{z : |z| < \epsilon_4\}$.

Step 5. Definition of N_z.

Define $N_z := L_z(I - P_z)$. Note it is analytic on $\{z : |z| < \epsilon_4\}$ because both L_z and P_z are analytic there.

Since $P_z^2 = P_z$ and $L_z P_z = P_z L_z$ from Step 2, we have

$$P_z N_z = P_z L_z (I - P_z) = P_z (I - P_z) L_z = (P_z - P_z^2) L_z = 0$$

and also

$$N_z P_z = L_z (I - P_z) P_z = L_z (P_z - P_z^2) = 0.$$

From $L_z P_z = \lambda_z P_z$ in Step 4, we deduce

$$L_z = L_z - L_z P_z + L_z P_z = L_z(I - P_z) + L_z P_z = N_z + \lambda_z P_z.$$

It remains to show $\rho(N_z) < |\lambda_z|$. Recall

$$\rho(L) = \lim_{n\to+\infty} \sqrt[n]{\|L^n\|} = \inf_{n\geqslant 1} \sqrt[n]{\|L^n\|}.$$

Fix any $\delta > 0$. Applying the above estimate to N_0 yields some $n \geqslant 1$ so large that

$$\sqrt[n]{\|N_0^n\|} < e^{\delta}\rho(N_0).$$

Since $z \mapsto N_z$ is analytic on $\{z : |z| < \epsilon_4\}$ from Step 5, it follows that $z \mapsto \sqrt[n]{\|N_z^n\|}$ is continuous there. So there is $\epsilon_5 \in (0, \epsilon_4)$ such that

$$\sqrt[n]{\|N_z^n\|} < e^{\delta}\sqrt[n]{\|N_0^n\|} < e^{2\delta}\rho(N_0), \quad \forall |z| < \epsilon_5.$$

By continuity (from analyticity) of $z \mapsto \lambda_z$, there is $\epsilon_6 \in (0, \epsilon_5)$ such that

$$|\lambda_z| > e^{-\delta}|\lambda_0|, \quad \forall |z| < \epsilon_6.$$

Pick $\delta > 0$ so small that $e^{3\delta}\rho(N_0) < |\lambda_0|$. Then, we have

$$\rho(N_z) \leqslant \sqrt[n]{\|N_z^n\|} < e^{2\delta}\rho(N_0) < e^{-\delta}|\lambda_0| < |\lambda_z|, \quad \forall |z| < \epsilon_6.$$

Up to further shrinking ϵ_6, we can make $\kappa := \min_{|z|<\epsilon_6} |\lambda_z| - e^{-\delta}|\lambda_0| > 0$. ☐

9 Herbert: Conditional Expectation

Definition 9.1 (Conditional Expectation)**.** Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(\mathcal{B})$. The conditional expectation of f with respect to $\mathcal{D}$ is the function $\mathbb{E}[f|\mathcal{D}] \in L^1(\mathcal{D})$ such that for all $D \in \mathcal{D}$ we have

$$\int_D f \mathrm{d}\mu = \int_D \mathbb{E}[f|\mathcal{D}] \mathrm{d}\mu$$

Theorem 9.2 (Existence and Uniqueness Of Conditional expectation)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(\mathcal{B})$. The conditional expectation $\mathbb{E}[f|\mathcal{D}] \in L^1(\mathcal{D})$ exists and is unique.*

Proof. First we prove the existence, define the complex measure

$$\nu : \mathcal{D} \to \mathbb{C}, \quad D \mapsto \int_D f \mathrm{d}\mu.$$

Remark 9.3. For every $D \in \mathcal{D}$. It is easy to check that this is indeed a complex measure. In fact: We have that

$$\nu(\varnothing) = \int_\varnothing f \mathrm{d}\mu = 0,$$

as ν is a measure. Likewise let $\{E_k\}$ be a countable disjoint family of sets. Then

$$\nu(\cup_k E_k) = \int_{\cup E_k} f \mathrm{d}\mu = \int \sum_{k=1}^{\infty} f \ \mathbb{1}_{E_k} \ \mathrm{d}\mu.$$

Now let

$$g_n = \sum_{k=1}^{n} f \ \mathbb{1}_{E_k},$$

then we have that $g_n \to f \ \mu - a.e.$ and $|g_n| \leqslant f \in L^1(\mathcal{B})$ so we can write

$$\int \sum_{k=1}^{\infty} f \ \mathbb{1}_{E_k} \ \mathrm{d}\mu = \int \lim_n g_n \mathrm{d}\mu = \lim_n \int g_n \mathrm{d}\mu = \lim_n \int \sum_{k=1}^{n} f \ \mathbb{1}_{E_k} \mathrm{d}\mu = \lim_n \sum_{k=1}^{n} \int f \ \mathbb{1}_{E_k} \mathrm{d}\mu = \sum_{k=1}^{\infty} \int_{E_k} f \ \mathrm{d}\mu = \sum_{k=1}^{\infty} \nu(E_k).$$

This is finite and absolutely convergent as $\int |f| \mathrm{d}\mu < \infty$ and hence ν is a complex measure.

Moreover if $D \in \mathcal{D}$ is such that $\mu(D) = 0$ then $\nu(D) = 0$ as well. Then we have $\nu \ll \mu$, the measure ν is absolutely continuous with respect to measure μ. Therefore we can apply the Radon-Nikodym theorem to find a derivative $g = \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^1(\mathcal{D})$. Then for every $D \in \mathcal{D}$

$$\int_D g \mathrm{d}\mu = \nu(D) = \int_D f \mathrm{d}\mu.$$

Thus g is the conditional expectation of f with respect to σ-algebra $\mathcal{D}$.

Now, to prove uniqueness, assume that $g, h \in L^1(\mathcal{D})$ are both conditional expectations of f. Then for each $D \in \mathcal{D}$ we have

$$\int_D g \mathrm{d}\mu = \int_D f \mathrm{d}\mu = \int_D h \mathrm{d}\mu$$

this implies that

$$\int_D (h - g) \mathrm{d}\mu = 0.$$

If the following set $\{x \in X : h(x) - g(x) \neq 0\}$ has positive measure then without loss of generality the set $\{x \in X : h(x) - g(x) > 0\}$ has positive measure. Hence there is some $\epsilon > 0$ such that the set $D_\epsilon = \{x \in$

$X : h(x) - g(x) > \epsilon\}$. has positive measure. Note that both h and g are measurable in $\mathcal{D}$ we conclude that $D_\epsilon = \{x \in X : h(x) - g(x) > \epsilon\} = \{x \in X : (h-g)(x) > \epsilon\} \in \mathcal{D}$ and hence

$$0 < \epsilon\mu(D_\epsilon) \leqslant \int_{D_\epsilon} (h-g)\mathrm{d}\mu = 0$$

which is a contradiction. □

Proposition 9.4. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(\mathcal{B})$ be real valued. The conditional expectation $\mathbb{E}[f|\mathcal{D}]$ satisfies*

$$\inf_{x\in X} f(x) \leqslant \mathbb{E}[f|\mathcal{D}](y) \leqslant \sup_{x\in X} f(x) \qquad a.s.$$

Proof. Fix $\epsilon > 0$ and let $D = \{y \in X : \mathbb{E}[f|\mathcal{D}](y) < \inf_{x\in X} f(x) - \epsilon\} \in \mathcal{D}$. We have

$$\mu(D) \inf_{x\in X} f(x) \leqslant \int_D f\mathrm{d}\mu = \int_D \mathbb{E}[f|\mathcal{D}]\mathrm{d}\mu \leqslant \mu(D)(\inf_{x\in X} f(x) - \epsilon).$$

Then we have $\epsilon\mu(D) \leqslant 0$, and thus $\mu(D) = 0$. Since $\epsilon > 0$ was arbitrary we conclude that almost surely $\inf_{x\in X} f(x) \leqslant \mathbb{E}[f|\mathcal{D}](y)$.

□

Lemma 9.5. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra and let $f \in L^1(X, \mathcal{B})$. Then almost everywhere*

$$|\mathbb{E}[f \mid \mathcal{D}]| \leqslant \mathbb{E}[|f| \mid \mathcal{D}]. \tag{10}$$

Proof. Asume that f is real value. Define $A = \{x \in X : \mathbb{E}[f|\mathcal{D}] > 0\}$ and $B = \{x \in X : \mathbb{E}[f|\mathcal{D}] \leqslant 0\}$. Let $D \in \mathcal{D}$ the set of points where the inequality (10) fails. Let $D^+ = D \cap A$ and $D^- = D \cap B$. Note also that $D = D^+ \cup D^-$, then

$$\begin{aligned}
\int_D |\mathbb{E}[f \mid \mathcal{D}]|\mathrm{d}\mu &= \int_{D^+} |\mathbb{E}[f \mid \mathcal{D}]|\mathrm{d}\mu + \int_{D^-} |\mathbb{E}[f \mid \mathcal{D}]|\mathrm{d}\mu \quad \text{by } D = D^+ \cup D^- \\
&= \left|\int_{D^+} \mathbb{E}[f \mid \mathcal{D}]\mathrm{d}\mu\right| + \left|\int_{D^-} \mathbb{E}[f \mid \mathcal{D}]\mathrm{d}\mu\right| \\
&= \left|\int_{D^+} f\mathrm{d}\mu\right| + \left|\int_{D^-} f\mathrm{d}\mu\right| \qquad \text{by definition of Conditional Expectation} \\
&\leqslant \int_{D^+} |f|\mathrm{d}\mu + \int_{D^-} |f|\mathrm{d}\mu \\
&= \int_{D^+} \mathbb{E}[|f| \mid \mathcal{D}]\mathrm{d}\mu + \int_{D^-} \mathbb{E}[|f| \mid \mathcal{D}]\mathrm{d}\mu \\
&= \int_D \mathbb{E}[|f| \mid \mathcal{D}]\mathrm{d}\mu.
\end{aligned}$$

Since D is the set of points where the inequality (10) fails, we conclude that $\mu(D) = 0$ as desired. □

Proposition 9.6. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra. The operator*

$$\mathbb{E}[\ .\ |\mathcal{D}] : L^1(X, \mathcal{B}) \to L^1(X, \mathcal{D}),$$

is continuous.

Proof. The idea is to show that the norm of the operator is 1. Let $f \in L^1(X, \mathcal{B})$ and by Lemma 9.5 we have

$$||\mathbb{E}[f \mid \mathcal{D}]|| = \int_X |\mathbb{E}[f \mid \mathcal{D}]| \mathrm{d}\mu \leqslant \int_X \mathbb{E}[|f| \mid \mathcal{D}] \mathrm{d}\mu = \int_X |f| \mathrm{d}\mu = ||f||.$$

□

Now if $f \in L^2(X, \mathcal{B})$ one can use the Hilbert space structure to give a different characterization of the conditional expectation.

Theorem 9.7. *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $\mathcal{D} \subset \mathcal{B}$ be a σ-algebra. Consider $P : L^2(X, \mathcal{B}) \to L^2(X, \mathcal{D})$ be the orthogonal projection, then for every $f \in L^2(X, \mathcal{B})$ we have $\mathbb{E}[f \mid \mathcal{D}] = Pf$.*

Proof. By definiton of orthogonal projection, for any function $g \in L^2(X, \mathcal{D})$ we have $\langle f - Pf, g \rangle = 0$. Therefore

$$\begin{aligned}
\int_D Pf \mathrm{d}\mu &= \int_X \mathbb{1}_D Pf \mathrm{d}\mu \\
&= \langle \mathbb{1}_D, Pf \rangle \\
&= \langle \mathbb{1}_D, f \rangle \\
&= \int_X \mathbb{1}_D f \mathrm{d}\mu \\
&= \int_D f \mathrm{d}\mu,
\end{aligned}$$

and hence $Pf = \mathbb{E}[f \mid \mathcal{D}]$ as desired. □

10 Herbert: Kac's Lemma

Let M be the phase space of a physical system, for example let M include all possible states of molecules in a box. The σ-algebra $\mathcal{B}$ represents the collections of observable states of the system and $\mu(A)$ is the probability of observing the state A. If f gives the discrete time evolution of the system, it is reasonable to expect that if the system is in equilibrium, f preserves the measure μ, that is, the probability of observing a certain state is independent on time. Consider now an initial state in which all the particles are in half of the box (imagine of having a wall which separates the box and then removing it). By Poincaré Recurrence Theorem, almost surely, all the molecules will return at some point in the same half of the box. This seems counter-intuitive. In reality, this is not a paradox, but simply the fact that the event will happen almost surely does not say anything about the time it will take to happen again (the recurrence time). Indeed, one can show that if the transformation is ergodic, the average recurrence time is inversely proportional to the measure of the set to which one wants to return.

Let again $f : M \to M$ be a measurable transformation.and μ a f-invariant finite measure. Let $E \subset M$ be any measurable set with $\mu(E) > 0$. Consider the first-return time function $\rho_E : E \to \mathbb{N} \cup \{\infty\}$, defined by

$$\rho_E(x) = \min\{n \geqslant 1 : f^n(x) \in E\}$$

whenever the set in the right-hand side is non-empty, otherwise $\rho_E(x) = \infty$.

Now we will show that this function ρ_E is integrable. To this end, we introduce

$$E_0 = \{x \in E : f^n(x) \notin E \text{ for all } n \geqslant 1\}$$
$$E_0^* = \{x \in M : f^n(x) \notin E \text{ for all } n \geqslant 0\}.$$

So, E_0 is the set of points of E that never return to E, and E_0^* is the set of points of M that never enter in E. Note that $\mu(E_0) = 0$ by the Poincaré recurrence theorem.

Theorem 10.1 (Kac). *Let $f : M \to M$ be a measurable transformation, μ a finite f-invariant measure and $E \subset M$ a subset of positive measure. Then, the function ρ_E is integrable and*

$$\int_E \rho_E \mathrm{d}\mu = \mu(M) - \mu(E_0^*)$$

Proof. For each $n \geqslant 1$ let us define

$$E_n = \{x \in E : f(x) \notin E, ..., f^{n-1}(x) \notin E \text{ but } f^n(x) \in E\}$$

$$E_n^* = \{x \in M : x \notin E, ..., f^{n-1}(x) \notin E \text{ but } f^n(x) \in E\}$$

This means that E_n is the set of points of E that return to E for the first time precisely at the moment n.

$$E_n = \{x \in E : \rho_E(x) = n\} = E \cap f^{-n}E \setminus \bigcup_{1 \leqslant k \leqslant n-1} E_k, \qquad \rho_E^{-1}(\{n\}) = E_n$$

and E_n^* is the set of points that is not in E and enter in E for the first time at the moment n. These sets are measurable since E is measurable and so the function ρ_E is measurable.

Afirmation: for $n \geqslant 1$ the sets E_n and E_n^* are pairwise disjoint. In fact, let E_i and E_j with $i < j$ and cosnider $E_i \cap E_j \neq \emptyset$. Take $x \in E_i$ then $x \in E$, $f^k(x) \notin E$ for $1 \leqslant k \leqslant i-1$ and $f^i(x) \in E$. Observe that $f^j(x) = f^{j-i}(f^i(x)) \in E$, then $f^i(x) \in E_{j-i}$, this means that $f^i(x) \in E$, $f^l(x) \notin E$ for $i+1 \leqslant l \leqslant j-1$. But $x \in E_j$ and $i < j$ then $f^i(x) \notin E$. Therefore the sets $\{E_n\}$ are pairwise disjoint. The proof that $E_j^* \cap E_i^* = \emptyset$ and $E_j^* \cap E_i = \emptyset$ are similar

So

$$\begin{aligned}
\mu(M) =& \mu\left(\left[\bigcup_{n\geqslant 0} E_n\right] \cup \left[\bigcup_{n\geqslant 0} E_n^*\right]\right) \\
=& \mu\left(\bigcup_{n\geqslant 0} E_n\right) + \mu\left(\bigcup_{n\geqslant 0} E_n^*\right) \\
=& \sum_{n=0}^{\infty} \mu(E_n) + \sum_{n=0}^{\infty} \mu(E_n^*) \\
=& \mu(E_0^*) + \mu(E_0) + \sum_{n=1}^{\infty} \mu(E_n) + \sum_{n=1}^{\infty} \mu(E_n^*) \\
=& \mu(E_0^*) + \sum_{n=1}^{\infty} \mu(E_n) + \sum_{n=1}^{\infty} \mu(E_n^*).
\end{aligned}$$

Then

$$\mu(M) - \mu(E_0^*) = \sum_{n=1}^{\infty} (\mu(E_n) + \mu(E_n^*)), \tag{11}$$

But the measure is finite, then $\mu(E_m), \mu(E_m^*) \to 0$ when $m \to \infty$. Now observe that

$$f^{-1}(E_n^*) = E_{n+1}^* \cup E_{n+1} \qquad \text{for every } n.$$

In fact, $y \in f^{-1}(E_n^*)$ then $f(y) \in E_n^*$ this means that the first iterate of $f(y)$ that belongs to E is $f^n(f(y)) = f^{n+1}(y)$ and that occurs if and only if $y \in E_n^*$ or else $y \in E_{n+1}$.

So, given that μ is invariant, we have

$$\mu(E_n^*) = \mu(f^{-1}(E_n^*)) = \mu(E_{n+1}^*) + \mu(E_{n+1}) \qquad \text{for every } n.$$

Observe that

$$\mu(E_{n+1}^*) = \mu(E_{n+2}^*) + \mu(E_{n+2})$$

then

$$\mu(E_n^*) = \mu(E_{n+2}^*) + \mu(E_{n+2}) + \mu(E_{n+1})$$

Now,

$$\mu(E_{n+2}^*) = \mu(E_{n+3}^*) + \mu(E_{n+3})$$

so,

$$\mu(E_n^*) = \mu(E_{n+3}^*) + \mu(E_{n+3}) + \mu(E_{n+2}) + \mu(E_{n+1}).$$

Applying this relation successively, we find that

$$\mu(E_n^*) = \mu(E_m^*) + \sum_{i=n+1}^{m} \mu(E_i) \qquad \text{for every } m > n.$$

Taking the limit as $m \to \infty$ we find that

$$\mu(E_n^*) = \sum_{i=n+1}^{\infty} \mu(E_i). \tag{12}$$

Replace (12) in the equality (11). In this way we find that

$$\mu(M) - \mu(E_0^*) = \sum_{n=1}^{\infty} \left(\sum_{i=n}^{\infty} \mu(E_i) \right) = \sum_{n=1}^{\infty} n\mu(E_n) = \sum_{n=1}^{\infty} \int_{E_n} \rho_E \, d\mu = \int_E \rho_E \, d\mu.$$

□

Remark 10.2. When the system (f, μ) is ergodic, the set E_0^* has zero measure. Then the conclusion of the Kac theorem means that

$$\frac{1}{\mu(E)} \int_E \rho_E \, d\mu = \frac{\mu(M)}{\mu(E)}$$

for every measurable set E with positive measure. The left-hand side of this expression is the mean return time to E. The mean retunr time is inversely proportional to the measure of E

Remark 10.3. Consider μ probability measure on M, let E be as in the statement of Kac's Lemma. One can define, by restriction to E, an induced σ-algebra $\mathcal{B}_E$ given by

$$\mathcal{B}_E = \{A \cap E : A \in \mathcal{B}\},$$

and an induced measure μ_E on $(E, \mathcal{B}_E)$ given by restriction

$$\mu_E(A) = \frac{\mu(A \cap E)}{\mu(E)}, \quad \textbf{for all } A \in \mathcal{B}.$$

This define a probability measure (conditional measure) μ_E on E, so that $\mu_E(E) = 1$. Then Kac's Lemma says that

$$\mathbb{E}[\rho_E \mid E] = \int_E \rho_E \, d\mu_E = \frac{1}{\mu(E)} \int_E \rho_E \, d\mu = \frac{1}{\mu(E)}.$$

So

$$\mathbb{E}[\rho_E \mid E] = \frac{1}{\mu(E)}$$

Thus, the average return time to E is $1/\mu(E)$.

11 Sarig L4, Application to Central Limit Theorem

Theorem 11.1 (Central Limit Theorem). *Let $(X, \mathcal{B}, \mu, T)$ be a mixing, probability preserving dynamical system. Suppose $\widehat{T} : \mathcal{L} \to \mathcal{L}$ has a spectral gap on a Banach space $\mathcal{L} \subseteq L^1(\mu)$ containing the constant functions, with norm $\|\cdot\|$ satisfying*[13]

$$\|fg\| \leqslant \|f\|\|g\|, \quad \& \quad \|\cdot\| \geqslant \|\cdot\|_{L^1}.$$

Let $\psi \in \mathcal{L}$ be bounded with $\int \psi \mathrm{d}\mu = 0$. If the cohomological[14] *equation $\psi = v - v \circ T$ has no solution $v \in \mathcal{L}$, then there is $\sigma > 0$ such that*

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi \circ T^k \xrightarrow[n\to+\infty]{\text{dist.}} N(0, \sigma^2);$$

in other words, for any interval $[a, b]$, we have

$$\mu\left\{x \in X : \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi \circ T^k(x) \in [a, b]\right\} \xrightarrow{n\to+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-t^2/\sigma^2} \mathrm{d}t.$$

Remark 11.2.

1. It seems that the mixing condition is unnecessary. Note $T_*\mu = \mu$ implies $\widehat{T}\mathbb{1} = \mathbb{1}$ and hence 1 is an eigenvalue of $\widehat{T} : L^1 \to L^1$ with the one-dimensional space consisting of the constant functions $\{\text{Const}\} \subseteq E(1)$ contained in the eigenspace $E(1)$ corresponding to eigenvalue 1. Then, the spectral gap condition on $\widehat{T} : \mathcal{L} \to \mathcal{L}$, where $\mathcal{L} \supseteq \{\text{Const}\}$, implies that 1 is the dominant eigenvalue and it is simple. However, it is unclear whether or not the spectral gap property implies mixing?? According to Exercise 1.5.4, the condition that 1 is the only eigenvalue on the unique circle and 1 is simple (not necessarily spectral gap) only implies weak mixing.
2. The Banach space $\mathcal{L}$ must contain the constant functions because the proof requires an operator $\widehat{T}_t$ defined on $\mathcal{L}$ and apply it to $\mathbb{1} \in \mathcal{L}$.

   It also seems to make sense to require further that $\mathcal{L}$ be a Banach algebra (closed under multiplication); only then can we guarantee $fg \in \mathcal{L}$ for any $f, g \in \mathcal{L}$, and for any $\psi \in \mathcal{L}$ we have $e^{it\psi} \in \mathcal{L}$; any Banach algebra must satisfy $\|fg\| \leqslant \|f\|\|g\|$. Sarig himself says so too `https://www.youtube.com/watch?v=ApTbp8FtFJg`.
3. The observable $\psi \in \mathcal{L}$ is assumed to be $M \to \mathbb{R}$ in order for the convergence in distribution to make sense.

   ψ being real may also imply that $\widehat{T}_t : f \mapsto e^{it\psi} f$ has norm $\|\widehat{T}_t\| \leqslant 1$?? Not sure how to prove this though.

11.1 Some Probability Theory

Definition 11.3 (Distribution Function). Let X be an $\mathbb{R}$-valued random variable. Its *distribution function*

$$F_X : \mathbb{R} \to [0, 1]$$

is defined as

$$F_X(t) := \mathbb{P}[X < t].$$

Definition 11.4 (Convergence in Distribution). Let $X_n, n = 1, 2, \cdots$ and Y be $\mathbb{R}$-valued random variables, not necessarily defined on the same probability space. We say the sequence X_n of random variables *converges in distribution to* random variable Y, written $X_n \xrightarrow[n\to+\infty]{\text{dist.}} Y$, if

$$F_{X_n}(t) \equiv \mathbb{P}[X_n < t] \xrightarrow{n\to+\infty} \mathbb{P}[Y < t] \equiv F_Y(t) \quad \text{at all continuity points } t \text{ of } F_Y(t);$$

in other words, the convergence holds for all $t \in \mathbb{R}$ where $F_Y(t) = \mathbb{P}[Y < t]$ is continuous.[15]

[13] For intance, the Lipschitz functions on $[0, 1]$ with $\|\cdot\|_{\text{Lip}}$ norm.

[14] For an explanation of the name "cohomological equation", see `https://amathew.wordpress.com/2010/07/17/the-cohomological-equation-for-dynamical-systems/` and `https://terrytao.wordpress.com/2008/12/21/cohomology-for-dynamical-systems/`.

[15] We only require the convergence for continuity points to allow convergence in distribution of $X_n = 2 - \frac{1}{n}$ to $Y = 2$, where the convergence $F_{X_n}(t) \to F_Y(t)$ fails at $t = 2$, a discontinuity point of $F_Y(t)$.

Definition 11.5 (Characteristic Function)**.** The *characteristic function* $\varphi_X(t)$ of a $\mathbb{R}$-valued random variable X is defined as

$$\varphi_X(t) := \mathbb{E}[e^{itX}].$$

Theorem 11.6 (Lévy's Continuity Theorem)**.** *A sequence of $\mathbb{R}$-valued random variables X_n converges in distribution to an $\mathbb{R}$-valued random variable if and only if*

$$\varphi_{X_n}(t) \xrightarrow{n\to+\infty} \varphi_Y(t), \qquad \forall t \in \mathbb{R}.$$

11.1.1 Berry-Esseen Smoothing Inequality

Sarig states an inequality, which he then proves in the Appendix, and then uses it to prove a special case of Lévy's Continuity Theorem, where $Y = N(0, \sigma^2)$. See Exercise 4.1. This special case suffices for proving our Central Limit Theorem.

11.2 Nagaev's Method

We now prove the Central Limit Theorem, following the functional-analytic Nagaev's method. Write

$$\psi_n := \psi + \psi \circ T + \cdots + \psi \circ T^{n-1}.$$

By Lévy's Continuity Theorem, it suffices to show convergence of characteristic functions, namely,

$$\varphi_{\frac{\psi_n}{\sqrt{n}}}(t) = \mathbb{E}[e^{i\frac{t}{\sqrt{n}}\psi_n}] = \int_X e^{i\frac{t}{\sqrt{n}}\psi_n} \mathrm{d}\mu \xrightarrow{n\to+\infty} e^{-\frac{1}{2}\sigma^2t^2} = \varphi_{N(0,\sigma^2)}(t), \qquad \forall t \in \mathbb{R}. \tag{13}$$

Define operators

$$\widehat{T}_t f := \widehat{T}(e^{it\psi}f),$$

where t will be taken real for now, but later we will extend it to $z \in \mathbb{C}$ to exploit Analytic Perturbation Theory.

Proposition 11.7 (Exercise 4.2)**.** $\widehat{T}_t^{\,n} f = \widehat{T}^n(e^{it\psi_n}f)$.

Proof. Base Case $n = 1$ is clear. Now assume for $n-1$ and show for n.

$$\begin{aligned}
\widehat{T}_t^{\,n} f =& \widehat{T}_t(\widehat{T}_t^{\,n-1} f) = \widehat{T}\left(e^{it\psi}\widehat{T}^{n-1}(e^{it\psi_{n-1}}f)\right)\\
=& \widehat{T}\left(\widehat{T}[e^{it\psi\circ T}\widehat{T}^{n-2}(e^{it\psi_{n-1}}f)]\right), \qquad \text{using Ex 1.2: } \widehat{T}[(f\circ T)g] = f(\widehat{T}g)\\
&\vdots\\
=& \widehat{T}^n[e^{it\psi\circ T^{n-1}}e^{it\psi_{n-1}}f] = \widehat{T}^n[e^{it\psi_n}f].
\end{aligned}$$

□

Note

$$\int_X \widehat{T}_t^{\,n}\mathbb{1}\mathrm{d}\mu = \int_X \widehat{T}^n(e^{it\psi_n})\mathrm{d}\mu = \int_X e^{it\psi_n}\mathbb{1}\circ T^n \mathrm{d}\mu = \int_X e^{it\psi_n} = \mathbb{E}[e^{it\psi_n}] = \varphi_{\psi_n}(t).$$

To prove (13), we need to study the behavior of $\varphi_{\frac{\psi_n}{\sqrt{n}}}(t) = \mathbb{E}[e^{i\frac{t}{\sqrt{n}}\psi_n}] = \int_X \widehat{T_{\frac{t}{\sqrt{n}}}}^{\,n}\mathbb{1}\mathrm{d}\mu$ as $n \to +\infty$ via analytic perturbation theory.

Claim: $z \mapsto \widehat{T}_z$ is an analytic family.

Indeed, by continuity of $\widehat{T}$, we obtain the expansion

$$\widehat{T}_z f = \widehat{T}(e^{iz\psi}f) = \widehat{T}\left(\sum_{n=0}^{\infty}\frac{(iz\psi)^n}{n!}f\right) = \widehat{T}f + \sum_{n=1}^{\infty}\frac{(iz)^n}{n!}\widehat{T}M_\psi^n f, \qquad \text{where } M_\psi : f \mapsto \psi f.$$

Since $\|M_\psi f\| = \|\psi f\| \leqslant \|\psi\|\|f\|$, it follows that M_ψ is a bounded linear operator with $\|M_\psi\| \leqslant \|\psi\|$, and hence

$$\|\widehat{T}M_\psi^n\| \leqslant \|\widehat{T}\|\|\psi\|^n.$$

This implies the series $\sum_{n=1}^{\infty} \frac{(iz)^n}{n!} \widehat{T}M_\psi^n$ converges in $\|\cdot\|$ norm on $\mathbb{C}$, and hence $\widehat{T}_z$ is analytic on $\mathbb{C}$, according to Ex 3.4.

Proposition 11.8.

$$\begin{aligned} \widehat{T}_z{}' &= i\widehat{T}_z M_\psi; \\ \widehat{T}_z{}'' &= -\widehat{T}_z M_\psi^2; \\ (\widehat{T}_z{}^n)' &= i\widehat{T}_z{}^n M_{\psi_n}; \\ (\widehat{T}_z{}^n)'' &= -\widehat{T}_z{}^n M_{\psi_n}^2. \end{aligned}$$

Proof. It follows directly from the expansion $\widehat{T}_z = \widehat{T} + \sum_{n=1}^{\infty} \frac{(iz)^n}{n!} \widehat{T}M_\psi^n$ that

$$\widehat{T}_z{}' = \sum_{n=1}^{\infty} \frac{n(iz)^{n-1}i}{n!} \widehat{T}M_\psi^n = i\left(\sum_{n=1}^{\infty} \frac{(iz)^{n-1}}{(n-1)!} \widehat{T}M_\psi^{n-1}\right) M_\psi = i\left(\sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \widehat{T}M_\psi^n\right) M_\psi = i\widehat{T}_z M_\psi.$$

The second derivative $\widehat{T}_z{}''$ is calculated the same way. For $(\widehat{T}_z{}^n)'$, note

$$(\widehat{T}_z{}^n)' = \left(\widehat{T}^n(e^{iz\psi_n})\right)' = \widehat{T}^n((e^{iz\psi_n})') = \widehat{T}^n(e^{iz\psi_n} iM_{\psi_n}) = i\widehat{T}^n(e^{iz\psi_n} M_{\psi_n}) = i\widehat{T}_z{}^n M_{\psi_n}.$$

The second derivative $(\widehat{T}_z{}^n)''$ is again calculated the same way. □

11.3 First Derivative λ_z' at $z = 0$

By assumption of the Central Limit Theorem, $\widehat{T}_0 = \widehat{T} = \lambda_0 P_0 + N_0$ has a spectral gap with $\lambda_0 = 1$ and $P_0 f = \int f \mathrm{d}\mu$, according to Ex 2.3 and Ex 3.8.

By Analytic Perturbation Theorem, there is an $\epsilon > 0$ such that for any $|z| < \epsilon$, the analytically perturbed operator $\widehat{T}_z$ also has a spectral gap:

$$\widehat{T}_z = \lambda_z P_z + N_z,$$

where $P_z^2 = P_z$, $\dim \mathrm{Im}(P_z) = 1$, $N_z P_z = P_z N_z = 0$, and

$$\rho(N_z) < |\lambda_z| - \kappa, \quad \text{for some uniform } \kappa > 0.$$

Differentiating equation $\widehat{T}_z P_z = \lambda_z P_z$ yields

$$\widehat{T}_z{}' P_z + \widehat{T}_z P_z' = \lambda_z' P_z + \lambda_z P_z'.$$

Left multiplying by P_z yields

$$\begin{aligned} P_z \widehat{T}_z{}' P_z + P_z \widehat{T}_z P_z' &= P_z \lambda_z' P_z + P_z \lambda_z P_z' \\ P_z \widehat{T}_z{}' P_z + \lambda_z P_z P_z' &= \lambda_z' P_z + \lambda_z P_z P_z' \quad \text{using } P_z \widehat{T}_z = \lambda_z P_z \text{ and } P_z^2 = P_z \\ P_z \widehat{T}_z{}' P_z &= \lambda_z' P_z. \end{aligned}$$

Plugging in $z = 0$ and applying to $\mathbb{1}$, we have

$$\lambda_0' = \lambda_0' P_0 \mathbb{1} = P_0 \widehat{T}_0{}' P_0 \mathbb{1} = P_0(i\widehat{T}_0 M_\psi \mathbb{1}) = iP_0(\widehat{T}_0 \psi) = i\int_X \widehat{T}\psi \mathrm{d}\mu = i\int_X \psi \mathrm{d}\mu = 0.$$

11.4 Second Derivative λ_z'' at $z = 0$

Claim: $\lambda''(0) = -\lim_{n\to+\infty} \frac{1}{n}\int_X (\psi_n)^2 \mathrm{d}\mu$.

Indeed, first we differentiate equation $\widehat{T_z}^n P_z = \lambda_z^n P_z$ to obtain

$$(\widehat{T_z}^n)' P_z + \widehat{T_z}^n (P_z)' = (\lambda_z^n)' P_z + \lambda_z^n (P_z)'.$$

Differentiating again yields

$$(\widehat{T_z}^n)'' P_z + 2(\widehat{T_z}^n)'(P_z)' + \widehat{T_z}^n (P_z)'' = (\lambda_z^n)'' P_z + 2(\lambda_z^n)'(P_z)' + \lambda_z^n (P_z)''.$$

Left multiplying by P_z yields

$$\begin{aligned}
P_z(\widehat{T_z}^n)'' P_z + 2P_z(\widehat{T_z}^n)'(P_z)' + P_z\widehat{T_z}^n (P_z)'' =& P_z(\lambda_z^n)'' P_z + 2(\lambda_z^n)' P_z(P_z)' + \lambda_z^n P_z(P_z)'' \\
P_z(\widehat{T_z}^n)'' P_z + 2P_z(\widehat{T_z}^n)'(P_z)' + \lambda_z^n P_z(P_z)'' =& P_z(\lambda_z^n)'' P_z + 2(\lambda_z^n)' P_z(P_z)' + \lambda_z^n P_z(P_z)'' \quad \text{using } P_z\widehat{T_z}^n = \lambda_z^n P_z \\
P_z(\widehat{T_z}^n)'' P_z + 2P_z(\widehat{T_z}^n)'(P_z)' =& P_z(\lambda_z^n)'' P_z + 2(\lambda_z^n)' P_z(P_z)'.
\end{aligned}$$

Evaluating at $z = 0$ yields

$$\begin{aligned}
P_0(\widehat{T_0}^n)'' P_0 + 2P_0(\widehat{T_0}^n)'(P_0)' =& P_0(\lambda_0^n)'' P_0 + 2(\lambda_0^n)' P_0(P_0)' \\
P_0(-\widehat{T_0}^n M_{\psi_n}^2) P_0 + 2P_0(i\widehat{T_0}^n M_{\psi_n})(P_0)' =& P_0(\lambda_0^n)'' P_0 + 2(\lambda_0^n)' P_0(P_0)'.
\end{aligned}$$

Note

$$\begin{aligned}
(\lambda_z^n)' =& n\lambda_z^{n-1}\lambda_z' \\
(\lambda_z^n)'' =& n(\lambda_z^{n-1})'\lambda_z' + n\lambda_z^{n-1}\lambda_z'',
\end{aligned}$$

and by evaluating at $z = 0$, we have

$$\begin{aligned}
(\lambda_0^n)' =& n\lambda_0^{n-1}\lambda_0' = n\lambda_0' = 0 \\
(\lambda_0^n)'' =& n(\lambda_0^{n-1})'\lambda_0' + n\lambda_0^{n-1}\lambda_0'' = n\lambda_0''.
\end{aligned}$$

Plugging these into the previous equation, we continue

$$P_0(-\widehat{T_0}^n M_{\psi_n}^2) P_0 + 2P_0(i\widehat{T_0}^n M_{\psi_n})(P_0)' = n\lambda_0'' P_0.$$

Applying to $\mathbb{1}$, we have

$$\begin{aligned}
n\lambda_0'' P_0\mathbb{1} =& P_0(-\widehat{T_0}^n M_{\psi_n}^2) P_0\mathbb{1} + 2P_0(i\widehat{T_0}^n M_{\psi_n})(P_0)'\mathbb{1} \\
\lambda_0'' =& \frac{1}{n}\left(-P_0\widehat{T_0}^n(\psi_n)^2 + 2iP_0\widehat{T_0}^n \psi_n (P_0)'\mathbb{1}\right) \\
=& -\frac{1}{n}\int_X \widehat{T}^n[(\psi_n)^2]\mathrm{d}\mu + 2i\frac{1}{n}\int_X \widehat{T}^n[\psi_n(P_0)'\mathbb{1}]\mathrm{d}\mu \\
=& -\frac{1}{n}\int_X (\psi_n)^2\mathrm{d}\mu + 2i\frac{1}{n}\int_X \psi_n(P_0)'\mathbb{1}\mathrm{d}\mu \\
=& -\frac{1}{n}\int_X (\psi_n)^2\mathrm{d}\mu + 2i\int_X ((P_0)'\mathbb{1})\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ T^k\mathrm{d}\mu.
\end{aligned}$$

By Birkhoff Ergodic Theorem,

$$\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ T^k \xrightarrow{n\to+\infty} \int_X \psi\mathrm{d}\mu = 0, \qquad \mu\text{-a.e.}$$

Since $|\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ T^k| \leqslant |\psi|$, it follows from Dominated Convergence Theorem that

$$\begin{aligned}
\lambda_0'' &= -\lim_{n\to+\infty}\frac{1}{n}\int_X (\psi_n)^2 \mathrm{d}\mu + 2i\lim_{n\to+\infty}\int_X ((P_0)'\mathbb{1})\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ T^k \mathrm{d}\mu \\
&= -\lim_{n\to+\infty}\frac{1}{n}\int_X (\psi_n)^2 \mathrm{d}\mu + 2i\int_X ((P_0)'\mathbb{1})\lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}\psi\circ T^k \mathrm{d}\mu \\
&= -\lim_{n\to+\infty}\frac{1}{n}\int_X (\psi_n)^2 \mathrm{d}\mu,
\end{aligned}$$

as claimed.

We now know the Taylor coefficients up to order 2:

$$\lambda_z = 1 - \frac{1}{2}\sigma^2 z^2 + O(z^3), \qquad \sigma := \sqrt{\lim_{n\to+\infty}\frac{1}{n}\int_X (\psi_n)^2 \mathrm{d}\mu}.$$

Proposition 11.9 (Green-Kubo Formula, Ex 4.4)**.**

$$\sigma^2 = \int_X \psi^2 \mathrm{d}\mu + 2\sum_{n=1}^{\infty}\int_X \psi(\psi\circ T^n)\mathrm{d}\mu.$$

Proof. By definition of σ, we have

$$\begin{aligned}
\sigma^2 &= \lim_{n\to+\infty}\frac{1}{n}\int_X (\psi_n)^2 \mathrm{d}\mu \\
&= \lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}\sum_{l=0}^{n-1}\int_X (\psi\circ T^k)(\psi\circ T^l)\mathrm{d}\mu
\end{aligned}$$

Note for $d = |k-l|$, by invariance $\mu\circ T^{-1} = \mu$, we have

$$\int_X (\psi\circ T^k)(\psi\circ T^l)\mathrm{d}\mu = \int_X \psi(\psi\circ T^d)\mathrm{d}\mu.$$

This implies

$$\begin{aligned}
\sigma^2 &= \lim_{n\to+\infty}\frac{1}{n}\left(n\int_X \psi^2\mathrm{d}\mu + \sum_{d=1}^{n-1}2(n-d)\int_X \psi(\psi\circ T^d)\mathrm{d}\mu\right) \\
&= \int_X \psi^2\mathrm{d}\mu + \lim_{n\to+\infty}\sum_{d=1}^{n-1}2\frac{n-d}{n}\int_X \psi(\psi\circ T^d)\mathrm{d}\mu \\
&= \int_X \psi^2\mathrm{d}\mu + \sum_{d=1}^{\infty}2\int_X \psi(\psi\circ T^d)\mathrm{d}\mu,
\end{aligned}$$

as desired. □

11.5 Limit of Characteristic Functions $\varphi_{\frac{\psi_n}{\sqrt{n}}}(t)$

Fix $t\in\mathbb{R}$. By previous analytic perturbation arguments, for n so large that $\frac{t}{\sqrt{n}} < \epsilon$, we have

$$\begin{aligned}
\varphi_{\frac{\psi_n}{\sqrt{n}}}(t) =& \mathbb{E}[e^{i\frac{t}{\sqrt{n}}\psi_n}] = \int_X \widehat{T_{\frac{t}{\sqrt{n}}}}^n \mathbb{1}\mathrm{d}\mu = \int_X \left(\lambda^n_{\frac{t}{\sqrt{n}}} P_{\frac{t}{\sqrt{n}}}\mathbb{1} + N^n_{\frac{t}{\sqrt{n}}}\mathbb{1}\right)\mathrm{d}\mu \\
=& \lambda^n_{\frac{t}{\sqrt{n}}}\left(1 + \int_X (P_{\frac{t}{\sqrt{n}}} - P_0)\mathbb{1}\mathrm{d}\mu + \lambda^{-n}_{\frac{t}{\sqrt{n}}}\int_X N^n_{\frac{t}{\sqrt{n}}}\mathbb{1}\mathrm{d}\mu\right) \\
=& \lambda^n_{\frac{t}{\sqrt{n}}}\left(1 + O(\|P_{\frac{t}{\sqrt{n}}} - P_0\|) + O(|\lambda^{-n}_{\frac{t}{\sqrt{n}}}|\|N^n_{\frac{t}{\sqrt{n}}}\|)\right) \qquad \text{using } \|\cdot\| \geqslant \|\cdot\|_{L^1}
\end{aligned}$$

By continuity of $z \mapsto P_z$, we have $\|P_{\frac{t}{\sqrt{n}}} - P_0\| \xrightarrow{n\to+\infty} 0$. Since $\lim_{n\to+\infty} \|N_z^n\|^{1/n} = \rho(N_z) \leqslant |\lambda_z| - \kappa$, it follows that

$$|\lambda_{\frac{t}{\sqrt{n}}}^{-n}| \|N_{\frac{t}{\sqrt{n}}}^n\| \leqslant \frac{\left\|N_{\frac{t}{\sqrt{n}}}^n\right\|}{\left|\lambda_{\frac{t}{\sqrt{n}}}\right|^n} \xrightarrow{n\to+\infty} 0.$$

With spectral gap, the long term behavior of $\widehat{T}^n \mathbb{1}$ is always dominated by the dominant eigenvalue. If we relax the hypothesis to quasi-compactness, then maybe we can recover a version of the Analytic Perturbation Theorem, and need to adjust the dominating behavior accordingly by summing over the multiplicities??? cf. Gouëzel

We continue

$$\begin{aligned}
\varphi_{\frac{\psi_n}{\sqrt{n}}}(t) &= \lambda_{\frac{t}{\sqrt{n}}}^n (1 + o(1)) \\
&= \left[1 - \frac{1}{2}\sigma^2 \left(\frac{t}{\sqrt{n}}\right)^2 + O((t/\sqrt{n})^3)\right]^n (1 + o(1)) \\
&= \left[1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})\right]^n (1 + o(1)).
\end{aligned}$$

The scaling by $\frac{1}{\sqrt{n}}$ is crucial to being able to see the limiting behavior.

By L'Hôpital's Rule, we have

$$\begin{aligned}
&\lim_{n\to+\infty} \log\left[1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})\right]^n = \lim_{n\to+\infty} n \log\left[1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})\right] \\
&= \lim_{n\to+\infty} \frac{\log\left[1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})\right]}{n^{-1}} = \lim_{n\to+\infty} \frac{\frac{\frac{\sigma^2 t^2}{2} n^{-2} + O(n^{-5/2})}{1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})}}{-n^{-2}} \\
&= -\lim_{n\to+\infty} \frac{\frac{\sigma^2 t^2}{2} + O(n^{-1/2})}{1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})} = -\frac{\sigma^2 t^2}{2}.
\end{aligned}$$

This shows

$$\left[1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})\right]^n \xrightarrow{n\to+\infty} e^{-\frac{\sigma^2 t^2}{2}},$$

and hence

$$\varphi_{\frac{\psi_n}{\sqrt{n}}}(t) = \lim_{n\to+\infty} \left[1 - \frac{\sigma^2 t^2}{2} n^{-1} + O(n^{-3/2})\right]^n (1 + o(1)) = e^{-\frac{\sigma^2 t^2}{2}} = \varphi_{N(0,\sigma^2)}(t), \qquad \forall t \in \mathbb{R}.$$

11.6 Positivity of σ

It remains to show $\sigma > 0$. This will come from the non-solvability of the cohomological equation

$$\psi = v - v \circ T. \tag{14}$$

For a contradiction, suppose $\sigma = 0$. We will then construct a solution $v \in \mathcal{L}$ to the cohomological equation (14). Take

$$u := \psi + \sum_{n=1}^{\infty} \widehat{T}^n \psi,$$

where the sum converges in $\|\cdot\|$ norm because $P_0 \psi = \int \psi \mathrm{d}\mu = 0$ and so

$$\|\widehat{T}^n \psi\| = \|N_0^n \psi\| \leqslant \|N_0^n\| \|\psi\| \xrightarrow[n\to+\infty]{\text{exp.fast}} 0, \qquad \text{using} \quad \lim_{n\to+\infty} \|N_0^n\|^{1/n} = \rho(N_0) < \lambda_0 = 1.$$

By construction, we have

$$\psi = u - \widehat{T}u, \qquad \& \qquad \widehat{T}u = \sum_{n=1}^{\infty} \widehat{T}^n \psi$$

From Green-Kubo Formula, we derive

$$\begin{aligned}
0 = \sigma^2 &= \int_X \psi^2 \mathrm{d}\mu + 2\sum_{n=1}^{\infty} \int_X \psi(\psi \circ T^n)\mathrm{d}\mu = \int_X \psi^2 \mathrm{d}\mu + 2\sum_{n=1}^{\infty} \int_X (\widehat{T}^n \psi)\psi \mathrm{d}\mu \\
&= \int_X \psi^2 \mathrm{d}\mu + 2\int_X \psi \sum_{n=1}^{\infty} \widehat{T}^n \psi \mathrm{d}\mu \quad \text{by Bounded Convergence using } \|\cdot\| \geqslant \|\cdot\|_{L^1} \\
&= \int_X (u - \widehat{T}u)^2 \mathrm{d}\mu + 2\int_X (u - \widehat{T}u)\widehat{T}u \mathrm{d}\mu = \int_X (u - \widehat{T}u)(u - \widehat{T}u + 2\widehat{T}u)\mathrm{d}\mu \\
&= \int_X (u - \widehat{T}u)(u + \widehat{T}u)\mathrm{d}\mu = \int_X u^2 - (\widehat{T}u)^2 \mathrm{d}\mu = \int_X u^2 \mathrm{d}\mu - \int_X (\widehat{T}u)^2 \mathrm{d}\mu \\
&= \int_X \widehat{T}(u^2)\mathrm{d}\mu - \int_X (\widehat{T}u)^2 \mathrm{d}\mu \quad \text{using } \int \widehat{T}g \mathrm{d}\mu = \int g \mathrm{d}\mu \\
&= \int_X \widehat{T}(u^2) \circ T \mathrm{d}\mu - \int_X ((\widehat{T}u) \circ T)^2 \mathrm{d}\mu \quad \text{using invariance } \int g \mathrm{d}\mu = \int g \circ T \mathrm{d}\mu \\
&= \int_X \mathbb{E}[u^2 | T^{-1}\mathcal{B}] - \mathbb{E}[u | T^{-1}\mathcal{B}]^2 \mathrm{d}\mu \quad \text{by ex 1.2 } (\widehat{T}g) \circ T = \mathbb{E}[g | T^{-1}\mathcal{B}].
\end{aligned}$$

Jensen's Inequality for Conditional Probabilities implies

$$\mathbb{E}[u^2 | T^{-1}\mathcal{B}] \geqslant \mathbb{E}[u | T^{-1}\mathcal{B}]^2.$$

Together with the above equality, we conclude

$$\mathbb{E}[u^2 | T^{-1}\mathcal{B}] = \mathbb{E}[u | T^{-1}\mathcal{B}]^2.$$

But this equality holds only when u is $T^{-1}\mathcal{B}$-measurable and hence

$$u = \mathbb{E}[u | T^{-1}\mathcal{B}] = (\widehat{T}u) \circ T,$$

by ex 1.2. Putting $v := -\widehat{T}u$, we have

$$\psi = u - \widehat{T}u = (\widehat{T}u) \circ T - \widehat{T}u = -v \circ T + v,$$

contradicting the non-solvability of the cohomological equation in $\mathcal{L}$. This shows $\sigma > 0$ and completes the proof of the Central Limit Theorem.

12 Gabriel – Sarig A3 Mixing and exactness for the Gauss map

Theorem 12.1 ((Martingale Convergence Theorem I), Bogachev Example 10.3.14.). *Suppose $(X, \mathcal{B}, \mu)$ is a probability space, and $(\mathcal{F}_n)_{n\geqslant 1}$ is an increasing sequence of sub-σ-algebras in $\mathcal{B}$. Define $\mathcal{F} := \sigma(\bigcup_{n\geqslant 1} \mathcal{F}_n)$ (the smallest σ-algebra containing the union). If $f \in L^1(\mu)$, then $\mathbb{E}(f|\mathcal{F}_n) \longrightarrow \mathbb{E}(f|\mathcal{F})$ a.e. and in L^1.*

Theorem 12.2 ((Martingale Convergence Theorem II), Bogachev Corollary 10.3.17.). *Suppose $(X, \mathcal{B}, \mu)$ is a probability space, and $(\mathcal{F}_n)_{n\leqslant 0}$ is a sequence of sub-σ-algebras in $\mathcal{B}$ such that $\mathcal{F}_{n-1} \subset \mathcal{F}_n$ for all n. Define $\mathcal{F} := \bigcap_{n\leqslant 0} \mathcal{F}_n$ (the intersection of σ-algebras is also a σ-algebra). If $f \in L^1(\mu)$, then $\mathbb{E}(f|\mathcal{F}_n) \longrightarrow \mathbb{E}(f|\mathcal{F})$ a.e. and in L^1.*

Let $(X, \mathcal{B}, \mu, T)$ be a probability preserving space.

Definition 12.3. We say that μ is *mixing* if, for every $A, B \in \mathcal{B}$:

$$\mu(A \cap T^{-n}(B)) \longrightarrow \mu(A)\mu(B).$$

Definition 12.4. If T is a (not-invertible) non-singular map, we say that μ is *exact* if $\mu(E) \in \{0, 1\}$ for every $E \in \bigcap_{n\in\mathbb{N}} T^{-n}\mathcal{B}$.

Remark 12.5. $T^{-n}\mathcal{B} = \{T^{-n}(B);\ B \in \mathcal{B}\}$.

Proposition 12.6. *If μ is exact, then it is mixing.*

Proof. By measurability of T, it follows that $(T^{-n}\mathcal{B})_{n\in\mathbb{N}}$ is a decreasing sequence of σ-algebras. By the Martingale Convergence Theorem II, for all $A \in \mathcal{B}$:

$$\mathbb{E}(\mathbb{1}_A|T^{-n}\mathcal{B}) \xrightarrow{L^1} \mathbb{E}\left(\mathbb{1}_A\Big|\bigcap_{n\in\mathbb{N}} T^{-n}\mathcal{B}\right) = \mathbb{E}(\mathbb{1}_A|\{\varnothing, X\}) = \mu(A).$$

So for all $A, B \in \mathcal{B}$:

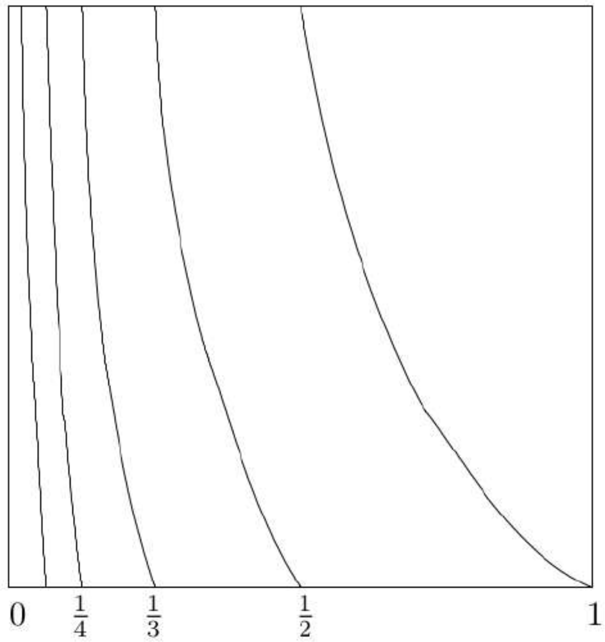
$$\begin{aligned} \mu(A \cap T^{-n}(B)) &= \int \mathbb{1}_A(\mathbb{1}_B \circ T^n)\mathrm{d}\mu = \int \mathbb{E}(\mathbb{1}_A|T^{-n}\mathcal{B})\mathbb{1}_B \circ T^n \mathrm{d}\mu \\ &= \int \mu(A)\mathbb{1}_B \circ T^n \mathrm{d}\mu + O\left(\int \Big|\mathbb{E}(\mathbb{1}_A|T^{-n}\mathcal{B}) - \mu(A)\Big|\mathrm{d}\mu\right) \longrightarrow \mu(A)\mu(B). \end{aligned}$$

□

Definition 12.7. The Gauss map $T : [0,1] \to [0,1]$ is defined as follows:

$$T(x) := \begin{cases} 0, & x = 0; \\ \dfrac{1}{x} \mod 1, & x \neq 0. \end{cases}$$

Remark 12.8. Notice that $\frac{1}{x} \mod 1$ is the fractional part of $\frac{1}{x}$, i.e., $\frac{1}{x} \mod 1 = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$. Sarig denotes $\frac{1}{x} \mod 1$ by $\left\{\frac{1}{x}\right\}$.



We will consider, on the space $[0,1]$, the Borel σ-algebra $\mathcal{B}$, and denote by m the Lesbesgue measure.

Proposition 12.9. *The Gauss map has the following invariant probability measure:*

$$d\mu = \frac{1}{\ln 2} \cdot \frac{1}{1+x} dx.$$

This means that, for every $B \in \mathcal{B}$:

$$\mu(B) = \frac{1}{\ln 2} \int_B \frac{1}{1+x} dx.$$

This measure is called Gauss measure. Moreover, Gauss and Lebesgue measures are absolutely continuous with respect to each other (i.e. they are equivalent).

Proof. Let $[a,b] \in \mathcal{B}$ be an interval, and notice that:

$$T^{-1}([a,b]) = \bigcup_{n=1}^{\infty} \left[\frac{1}{b+n}, \frac{1}{a+n} \right].$$

This union is disjoint, so:

$$\begin{aligned}
\mu(T^{-1}([a,b])) &= \frac{1}{\ln 2}\sum_{n=1}^{\infty}\int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{1}{1+x}\mathrm{d}x \\
&= \frac{1}{\ln 2}\sum_{n=1}^{\infty}\left[\ln\left(1+\frac{1}{a+n}\right)-\ln\left(1+\frac{1}{b+n}\right)\right] \\
&= \frac{1}{\ln 2}\sum_{n=1}^{\infty}\left[\ln(a+n+1)-\ln(a+n)-\ln(b+n+1)+\ln(b+n)\right] \\
&= \frac{1}{\ln 2}\lim_{N\to\infty}\sum_{n=1}^{N}\left[\ln(a+n+1)-\ln(a+n)-\ln(b+n+1)+\ln(b+n)\right] \\
&= \frac{1}{\ln 2}\lim_{N\to\infty}\left[\ln(a+N+1)-\ln(a+1)-\ln(b+N+1)+\ln(b+1)\right] \\
&= \frac{1}{\ln 2}\left[\ln(b+1)-\ln(a+1)+\lim_{N\to\infty}\ln\left(\frac{a+N+1}{b+N+1}\right)\right] \\
&= \frac{1}{\ln 2}\left[\ln(b+1)-\ln(a+1)\right] \\
&= \frac{1}{\ln 2}\int_a^b \frac{1}{1+x}\mathrm{d}x = \mu([a,b]).
\end{aligned}$$

This proves the invariance. To show that Gauss and Lebesgue measures are absolutely continuous with respect to each other, we will prove the following inequalities for every $B \in \mathcal{B}$:

$$\frac{1}{2\ln 2}m(B) \leqslant \mu(B) \leqslant \frac{1}{\ln 2}m(B).$$

In fact, for every $x \in B$, we have $\frac{1}{2} \leqslant \frac{1}{1+x} \leqslant 1$. Then:

$$\frac{1}{2\ln 2}m(B) = \frac{1}{\ln 2}\int_B \frac{1}{2}\mathrm{d}x \leqslant \frac{1}{\ln 2}\int_B \frac{1}{1+x}\mathrm{d}x = \mu(B) \leqslant \frac{1}{\ln 2}\int_B 1\mathrm{d}x = \frac{1}{\ln 2}m(B).$$ □

Remark 12.10. If $x \in (0,1)$, then x has a continued fraction expansion of the form:

$$x = \cfrac{1}{x_1 + \cfrac{1}{x_2 + \cfrac{1}{x_3 + \cdots}}},$$

where each x_j belongs to $\mathbb{N}$. Moreover, it can be shown that x is irrational if and only if it has infinite continued fraction expansion; and, in this case, the continued fraction expansion is unique. Observe that:

$$\frac{1}{x} = x_1 + \cfrac{1}{x_2 + \cfrac{1}{x_3 + \cdots}}.$$

Hence, $T(x)$ acts like a shift map, deleting the first term in the continued fraction expansion of x:

$$T(x) = \cfrac{1}{x_2 + \cfrac{1}{x_3 + \cfrac{1}{x_4 + \cdots}}}.$$

Lemma 12.11 (Exercise 2.5). *For each $a \in \mathbb{N}$, define*

$$v_a(x) := \frac{1}{a+x}, \qquad x \in [0,1],$$

and their compositions

$$v_{a_1,\cdots,a_n} := v_{a_n} \circ \cdots \circ v_{a_1}.$$

a. For any $f \in L^1$,

$$\widehat{T}^n f = \sum_{a_1,\cdots,a_n=1}^{\infty} |v'_{a_1,\cdots,a_n}| f \circ v_{a_1,\cdots,a_n}.$$

b. There are constants $C > 0$ and $\theta \in (0,1)$ such that for any string $\underline{a} := (a_1, \ldots, a_n)$ of any length $n \geqslant 1$, we have

$$|v_{\underline{a}}(x) - v_{\underline{a}}(y)| < C\theta^n |x-y|.$$

c. There is a constant $H > 1$ such that for any $x, y \in [0,1]$ and any string $\underline{a} := (a_1, \ldots, a_n)$ of any length $n \geqslant 1$, we have

$$\left|\frac{v'_{\underline{a}}(x)}{v'_{\underline{a}}(y)} - 1\right| \leqslant H|x-y|.$$

d. There is another constant $G > 1$ such that for any $x \in [0,1]$ and any string $\underline{a} := (a_1, \ldots, a_n)$ of any length $n \geqslant 1$, we have

$$G^{-1} m(v_{\underline{a}}[0,1)) \leqslant |v'_{\underline{a}}(x)| \leqslant G m(v_{\underline{a}}[0,1)).$$

e. $v_{\underline{a}}[0,1)$ are non-overlapping sub-intervals of $[0,1)$.

Theorem 12.12 (Rényi). *The Gauss map is exact with respect to Gauss measure.*

Proof. By the equivalence of Lebesgue and Gauss map, it is enough to show that T is "exact" with respect to Lebesgue measure m. For each $a \in \mathbb{N}$, let $v_a : [0,1] \to [0,1]$ denote the inverse branches $v_a(x) := \frac{1}{a+x}$, and set, for every $\underline{a} = (a_1, \ldots, a_n)$, $v_{\underline{a}} := v_{a_1} \circ \cdots \circ v_{a_n}$. Define $[\underline{a}] := v_{\underline{a}}([0,1])$. This is the set of all numbers whose continued fraction expansion starts with $\underline{a}$.

Rényi's inequality: there exists $C > 1$ such that:

$$\frac{1}{C} \cdot m[\underline{a}]m[\underline{b}] \leqslant m[\underline{a}, \underline{b}] \leqslant C \cdot m[\underline{a}]m[\underline{b}] \qquad \forall\, \underline{a}, \underline{b}.$$

Here, $[\underline{a}, \underline{b}]$ denotes the set of all numbers whose continued fraction expansion stars with $\underline{a}$ followed by $\underline{b}$. We will also denote by $|\underline{a}|$ the length of $\underline{a}$.

Proof of Rényi's inequality.

$$\begin{aligned} m[\underline{a}, \underline{b}] &= \int \mathbb{1}_{[\underline{a}]} \mathbb{1}_{[\underline{b}]} \circ T^{|\underline{a}|} dm = \int_{[\underline{b}]} \widehat{T}^{|\underline{a}|} \mathbb{1}_{[\underline{a}]} dm \\ &= \int_{[\underline{b}]} |v'_{\underline{a}}| dm \quad \textit{(by Exercise 2.5 a, because } \underline{a} \neq \underline{b} \Rightarrow \mathbb{1}_{[\underline{a}]} \circ v_{\underline{b}} = 0 \textit{ by Exercise 2.5 e)} \\ &= \int_{[\underline{b}]} G^{\pm 1} m[\underline{a}] dm = G^{\pm 1} m[\underline{a}] m[\underline{b}] \quad \textit{(by Exercise 2.5 d)}. \end{aligned}$$

Here $a = G^{\pm 1} b$ means $G^{-1} \leqslant a/b \leqslant G$. Choosing $C = G$, we have then proved that:

$$\frac{1}{C} \cdot m[\underline{a}]m[\underline{b}] \leqslant m[\underline{a}, \underline{b}] \leqslant C \cdot m[\underline{a}]m[\underline{b}].$$

□

Define, for each $n \in \mathbb{N}$, $\mathcal{F}_n := \sigma(\{[\underline{a}];\, |\underline{a}| = n\})$. Then $\mathcal{B} = \sigma(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n)$. Standard approximation arguments show that, for every $\underline{a}$ and $B \in \mathcal{B}$:

$$\frac{1}{C} \cdot m[\underline{a}]m(B) \leqslant m(\underline{a} \cap T^{-|\underline{a}|}(B)) \leqslant C \cdot m[\underline{a}]m(B).$$

We can now show exactness. Suppose $B \in \bigcap_{n\in\mathbb{N}} T^{-n}\mathcal{B}$ *and* $m(B) > 0$. *For every* $n \in \mathbb{N}$, *there exists* $B_n \in \mathcal{B}$ *such that* $B = T^{-n}(B_n)$. *Then, for every* $\underline{a}$ *with* $|\underline{a}| = n$:

$$m(B \cap [\underline{a}]) = m(T^{-n}(B_n) \cap [\underline{a}]) \geqslant \frac{1}{C} m(B_n) m[\underline{a}]. \tag{15}$$

Remember that $\frac{1}{2\ln 2} \leqslant \frac{d\mu}{dm} \leqslant \frac{1}{\ln 2}$, *where* μ *is the Gauss measure. So:*

$$m(B_n) \geqslant \ln 2 \mu(B_n) = \ln 2 \mu(B) \geqslant \frac{1}{2}\mu(B).$$

Then, by (15) *it follows that, for all* $\underline{a}$:

$$\frac{m(B \cap [\underline{a}])}{m[\underline{a}]} \geqslant \frac{m(B)}{2C}.$$

Moreover, we know that, for each $n \in \mathbb{N}$, $\mathcal{F}_n = \sigma(\{[\underline{a}];\ |\underline{a}| = n\})$. *Hence (exercise):*

$$\mathbb{E}_m(\mathbb{1}_B|\mathcal{F}_n) = \sum_{|\underline{a}|=n} \frac{m(B \cap [\underline{a}])}{m[\underline{a}]} \mathbb{1}_{[\underline{a}]}.$$

Therefore, $\mathbb{E}_m(\mathbb{1}_B|\mathcal{F}_n) \geqslant \frac{m(B)}{2C} > 0$. *But* $\mathcal{B} = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right)$, *so by the Martingale Convergence Theorem I:*

$$\lim_{n\to\infty} \mathbb{E}_m(\mathbb{1}_B|\mathcal{F}_n) = \mathbb{E}(\mathbb{1}_B|\mathcal{B}).$$

Therefore $\mathbb{1}_B > 0$ *a.e., which implies* $m(B) = 1$. □

13 Future Exercises

1. **Sarig Exercise 2.3.**

   Zheng has discussed a little of this in the 4th meeting on 27 Jan 2021, see Proposition 3.13. The sticking points are: (i) is the action of projection $Pf = h\int f\mathrm{d}\mu$ uniquely determined? if so, prove it. if not, provide an example of an alternative action. (ii) is the assumption $\|\cdot\|_{\mathcal{L}} \geqslant \|\cdot\|_{L^1}$ really necessary for this proposition?

2. **Exercise for Hennion's Theorem.** Show that it suffices to prove the case $k = 1$.

3. equivalence of weak and strong analyticity. Sargig A4

4. Separation of Spectrum Theorem. Sarig A5

5. Kato Lemma. Sarig A6

A Lebesgue-Radon-Nikodym Theorem; a Special Case

In this subsection, we prove a special case of the Lebesgue-Radon-Nikodym Theorem, [Rud87] Theorem 6.10. We follow an elegant proof due to Rudin based on Riesz Representation Theorem.

Theorem A.1 (Riesz Representation; [Rud87] Theorem 4.12)**.** *If L is a continuous linear function on Hilbert space H, then there is a unique $y \in H$ such that*

$$Lx = \langle x, y \rangle \quad \forall x \in H.$$

Another important ingredient in Rudin's proof is the fact that L^2 is a Hilbert space.

Theorem A.2 (L^2 is Hilbert)**.** *Let $(X, \mathcal{F}, \mu)$ be any measure space, and let $L^2(\mu)$ denote the space of measurable functions $f : X \to \mathbb{C}$ for which*

$$\int |f|^2 \mathrm{d}\mu < +\infty,$$

where two functions are identified when they coincide μ-a.e. By Cauchy-Schwarz Inequality, $\langle \cdot, \cdot \rangle$ given by

$$\langle f, g \rangle := \int_X f(x)\overline{g(x)} \mathrm{d}\mu x$$

defines an inner product on $L^2(\mu)$. Moreover, $(L^2(\mu), \langle \cdot, \cdot \rangle)$ is complete and hence a Hilbert space.

Next, we state and prove a special case of the Lebesgue-Radon-Nikodym Theorem.

Theorem A.3 (Lebesgue-Radon-Nikodym Theorem; Special Case)**.** *Let μ, λ be two real positive finite measures on measurable space $(X, \mathcal{F})$.*

(a) There is then a unique pair of real finite measures λ_a and λ_s on $(X, \mathcal{F})$ such that

$$\lambda = \lambda_a + \lambda_s, \quad \lambda_a \ll \mu, \quad \lambda_s \perp \mu$$

Moreover, λ_a and λ_s are finite.

(b) There is a unique $h \in L^1(\mu)$ such that

$$\lambda_a(E) = \int_E h \mathrm{d}\mu, \quad \forall E \in \mathcal{F}.$$

Proof. First we show the uniqueness of the decomposition

$$\lambda = \lambda_a + \lambda_s, \quad \lambda_a \ll \mu, \quad \lambda_s \perp \mu.$$

If $\lambda = \lambda'_a + \lambda'_s$ is another such decomposition, then

$$\mu \gg \lambda_a - \lambda'_a = \lambda'_s - \lambda_s \perp \mu.$$

This implies $\lambda_a - \lambda'_a = \lambda'_s - \lambda_s$ is the zero measure, and hence the decomposition is unique.

For existence, note

$$\varphi := \lambda + \mu$$

is another real positive finite measure on $(X, \mathcal{F})$. Define

$$\Lambda : L^2(\varphi) \to \mathbb{R}, \quad f \mapsto \int f \mathrm{d}\lambda.$$

Since the integral is linear, so is Λ. Also, Λ is bounded:

$$\begin{aligned} |\Lambda(f)| &= \left| \int f \mathrm{d}\lambda \right| \leqslant \int |f| \mathrm{d}\lambda \leqslant \int |f| \cdot 1 \\ &\leqslant \left(\int |f|^2 \mathrm{d}\varphi \right)^{1/2} \left(\int 1^2 \mathrm{d}\varphi \right)^{1/2} \quad \text{by Schwarz Inequality} \\ &= \|f\|_2 (\varphi(X))^{1/2}. \end{aligned}$$

Hence, Λ has a unique Riesz representation by some $g \in L^2(\varphi)$:

$$\int f \mathrm{d}\lambda = \int fg \mathrm{d}\varphi, \quad \forall f \in L^2(\varphi). \tag{16}$$

Part I. We show $g(x) \in [0, 1]$ for φ-a.e. $x \in X$. Taking $f = \mathbb{1}_E$ with E any measurable set in X, we have

$$0 \leqslant \lambda(E) = \int \mathbb{1}_E \mathrm{d}\lambda = \int \mathbb{1}_E g \mathrm{d}\varphi \leqslant \varphi(E).$$

This implies that

$$0 \leqslant \frac{1}{\varphi(E)} \int_E g \mathrm{d}\varphi \leqslant 1, \tag{17}$$

whenever $\varphi(E) > 0$.

For a contradiction, suppose there is some interval $[\alpha - \epsilon, \alpha + \epsilon]$ such that $[\alpha - \epsilon, \alpha + \epsilon] \cap [0, 1] = \varnothing$ and $\varphi(g^{-1}[\alpha - \epsilon, \alpha + \epsilon]) > 0$. Then, denoting $E = g^{-1}[\alpha - \epsilon, \alpha + \epsilon]$, we have

$$\left| \frac{1}{\varphi(E)} \int_E g \mathrm{d}\varphi - \alpha \right| = \left| \frac{1}{\varphi(E)} \int_E (g - \alpha) \mathrm{d}\varphi \right| \leqslant \frac{1}{\varphi(E)} \epsilon \varphi(E) = \epsilon,$$

contradicting bounds (17). We conclude any interval $[\alpha - \epsilon, \alpha + \epsilon]$ disjoint from $[0, 1]$ must have $\varphi(g^{-1}[\alpha - \epsilon, \alpha + \epsilon]) = 0$.

For every $\alpha \in \mathbb{Q} \backslash [0, 1]$, there is some $\epsilon_\alpha > 0$ for which $[\alpha - \epsilon_\alpha, \alpha + \epsilon_\alpha]$ is disjoint from $[0, 1]$, so

$$\varphi(g^{-1}[\alpha - \epsilon_\alpha, \alpha + \epsilon_\alpha]) = 0, \quad \forall \alpha \in \mathbb{Q} \backslash [0, 1].$$

This implies

$$\varphi(g^{-1}(\mathbb{R} \backslash [0, 1])) = \varphi \left(g^{-1} \left(\bigcup_{\alpha \in \mathbb{Q} \backslash [0,1]} [\alpha - \epsilon_\alpha, \alpha + \epsilon_\alpha] \right) \right) = 0,$$

that is to say, $g(x) \in [0, 1]$ for φ-a.e. $x \in X$. Up to redefining g outside this full-φ-measure set, we may assume $g(x) \in [0, 1]$ for all $x \in X$.

Part II. Define sets

$$A := \{x \in X : g(x) < 1\}, \quad B := \{x \in X : g(x) = 1\},$$

and real positive finite measures

$$\lambda_a(E) = \lambda(E \cap A), \quad \lambda_s(E) = \lambda(E \cap B).$$

Then clearly $A \cup B = X$ and thus we have decomposition

$$\lambda = \lambda_a + \lambda_s.$$

Rewrite the relation (16) into $\int f \mathrm{d}\lambda = \int fg \mathrm{d}\lambda + \int fg \mathrm{d}\mu$

$$\int f(1 - g) \mathrm{d}\lambda = \int fg \mathrm{d}\mu, \quad \forall f \in L^2(\varphi). \tag{18}$$

Take $f = \mathbb{1}_B$, and by (18) and definition of B, we see

$$0 = \int_B (1 - g) \mathrm{d}\lambda = \int_B g \mathrm{d}\mu = \mu(B).$$

This implies $\lambda_s \perp \mu$.

Part III. We show existence part of statement (b), which will imply $\lambda_a \ll \mu$. (Alternatively, one may show $\lambda_a \ll \mu$ directly.) Take any measurable set $E \subseteq X$ and define for each $n \in \mathbb{N}$

$$f_n(x) := [1 + g(x) + \cdots + (g(x))^n] \, \mathbb{1}_E.$$

Note each f_n is bounded and hence $L^2(\varphi)$. By relation (18) applied to f_n and the identity

$$(1-x)(1+x+\cdots+x^n) = 1-x^{n+1},$$

we obtain

$$\int_E (1-(g(x))^{n+1})\mathrm{d}\lambda(x) = \int f_n(1-g)\mathrm{d}\lambda = \int f_n g \mathrm{d}\mu = \int_E \left[1+g(x)+\cdots+(g(x))^n\right] g(x)\mathrm{d}\mu(x).$$

Since $g \equiv 1$ on B and $\mu(B)=0$, we have

$$\int_{E\cap A} (1-(g(x))^{n+1})\mathrm{d}\lambda(x) = \int_{E\cap A} \left[1+g(x)+\cdots+(g(x))^n\right] g(x)\mathrm{d}\mu(x).$$

For any $x \in E\cap A$, we have $g(x) \in [0,1)$ and hence $\lim\limits_{n\to+\infty}(g(x))^{n+1} = 0$. By Monotone Convergence,

$$\lambda_a(E) = \lambda(E\cap A) = \int_{E\cap A} 1\mathrm{d}\lambda = \lim_{n\to+\infty}\int_{E\cap A}(1-(g(x))^{n+1})\mathrm{d}\lambda(x).$$

On the other hand, for any $x \in E\cap A$, we have $g(x)\in[0,1)$, and hence

$$\left[1+g(x)+\cdots+(g(x))^n\right] g(x) \leqslant \frac{g(x)}{1-g(x)} < +\infty.$$

Monotone Convergence thus yields

$$\lambda_a(E) = \lim_{n\to+\infty}\int_{E\cap A}\left[1+g(x)+\cdots+(g(x))^n\right] g(x)\mathrm{d}\mu(x) = \int_{E\cap A} h(x)\mathrm{d}\mu(x) = \int_E h(x)\mathrm{d}\mu(x),$$

where $h : X \to \mathbb{R}$ (note $h \in L^1(\mu)$) is given by

$$h(x) = \begin{cases} \lim\limits_{n\to+\infty}\left[1+g(x)+\cdots+(g(x))^n\right] g(x) & x\in A \\ 0 & x \in B \end{cases}.$$

Part IV. It remains to show uniqueness part of statement (b). Suppose $h' \in L^1(\mu)$ also satisfies

$$\lambda_a(E) = \int_E h'\mathrm{d}\mu, \qquad \forall E \in \mathcal{F}.$$

Note $D := \{x \in X : h(x) > h'(x)\}$ is measurable. For a contradiction, suppose $\mu(D) > 0$. Then,

$$0 = \lambda_a(D) - \lambda_a(D) = \int_D h\mathrm{d}\mu - \int_D h'\mathrm{d}\mu = \int_D (h-h')\mathrm{d}\mu > 0,$$

a contradiction. Hence, $h \in L^1(\mu)$ is unique. The proof is complete. □

Try to generalize to $\nu \ll \mu$, where ν is finite but μ is σ-finite. Maybe even further to signed or complex measures???

B Hahn-Banach Theorem and Consequences

Definition B.1 (Closed Linear Subspace)**.** A subset E of a linear space V over $\mathbb{R}$ is called a *linear subspace* if

$$\lambda_1 f_1 + \lambda_2 f_2 \in E, \quad \forall \lambda_1, \lambda_2 \in \mathbb{R}, \forall f_1, f_2 \in E.$$

A subset E of a normed linear space V is called *closed* if

$$\lim_{n\to+\infty} f_n \in E, \quad \forall \{f_n\} \subseteq E \text{ strongly convergent in } V.$$

Definition B.2 (Sublinear Functional)**.** On a vector space X, a *sublinear functional* is a function $q : X \to \mathbb{R}$ such that

(a) $q(x+y) \leqslant q(x) + q(y)$ for all $x, y \in X$;

(b) $q(ax) = aq(x)$ for all $x \in X$ and $a \in \mathbb{R}_{\geqslant 0}$.

Note any norm is a sublinear functional. In fact, (a) is the Triangle Inequality, and (b) is Homogeneity for nonnegative scalars. So a sublinear functional q falls short of being a norm in that q can take negative values, need not satisfy Absolute Homogeneity ($\|ax\| = |a|\|x\|$ for any $a \in \mathbb{R}$ and $x \in X$) for negative scalars, and need not separate points ($\|x\| = 0$ if and only if $x = 0$).

Theorem B.3 (Hahn-Banach Theorem for Vector Spaces over $\mathbb{R}$; [Con85] Theorem 6.2)**.** *Let X be a vector space over $\mathbb{R}$ and q a sublinear functional on X. If M is a linear subspace of X and $f : M \to \mathbb{R}$ is a linear functional with $f(x) \leqslant q(x)$ for all $x \in M$, then there is a linear functional $F : X \to \mathbb{R}$ such that $F|_M = f$ and $F(x) \leqslant q(x)$ for all $x \in X$.*

The substance of the theorem is not that the extension from M to X exists, but that there is an extension that remains dominated by sublinear functional q. The proof relies on Zorn's Lemma[16].

Proof. Take $x_1 \in X \backslash M$ and define $M_1 := \text{span}\{x_1, M\}$. In order to extend f to M_1, we need to find an appropriate value $\alpha_1 \in \mathbb{R}$ for

$$f(x_1) = \alpha_1.$$

Then, by linearity, we will have $f(tx_1 + y) = tf(x_1) + f(y) = t\alpha_1 + f(y)$ for all $tx_1 + y \in \text{span}\{x_1, M\}$.

The restriction is domination by q. Finding such an appropriate α_1 reduces to meeting the following two conditions.

(i) for $t > 0$, we need $f(tx_1 + y) \leqslant q(tx_1 + y) = tq(x_1 + t^{-1}y)$ for all $y \in M$, or equivalently,

$$t^{-1} f(tx_1 + y) = \alpha_1 + f(t^{-1}y) \leqslant q(x_1 + t^{-1}y), \quad \forall y \in M.$$

In other words, we need

$$\alpha_1 \leqslant q(x_1 + y') - f(y'), \quad \forall y' \in M.$$

(ii) for $t < 0$, the condition $f(tx_1 + y) \leqslant q(tx_1 + y) = (-t)q(-x_1 + (-t)^{-1}y)$ for all $y \in M$ is equivalent to

$$(-t)^{-1} f(tx_1 + y) = -\alpha_1 + f((-t)^{-1}y) \leqslant q(-x_1 + (-t)^{-1}y), \quad \forall y \in M.$$

In other words, we need

$$\alpha_1 \geqslant f(y'') - q(-x_1 + y''), \quad \forall y'' \in M.$$

Note that for any $y', y'' \in M$, we have

$$f(y') + f(y'') = f(y' + y'') \leqslant q(y' + y'') \leqslant q(x_1 + y') + q(-x_1 + y'').$$

That is,

$$f(y'') - q(-x_1 + y'') \leqslant q(x_1 + y') - f(y'), \quad \forall y', y'' \in M.$$

[16] **Zorn's Lemma:** If every nonempty chain in a nonempty partially ordered set P has an upper bound in P, then P has at least one maximal element.

Therefore, there exists an $\alpha_1 \in \mathbb{R}$ satisfying conditions (i) and (ii), and we have extended f to M_1.

The proof is completed by Zorn's Lemma. Define the set P to be the collection of all linear subspaces N with $M \subseteq N \subseteq X$ for which there exists a linear functional $g : N \to \mathbb{R}$ with $g|_N = f$ and $g(x) \leqslant q(x)$ for all $x \in N$. Note $P \ni M$ is nonempty. Partially order P by inclusion $\subseteq$. Then, any nonempty chain $\{N_c\}_{c \in C} \subseteq P$ has an upper bound $N = \bigcup_{c \in C} N_c \in P$. Zorn's Lemma yields a maximal element $N_0 \in P$.

We show $N_0 = X$. Suppose not. Then, there is some $x_2 \in X \backslash N_0$ so that $N_1 := \text{span}\{x_2, N_0\}$ is a strictly larger linear subspace than N_0. Since we can extend g from N_0 to N_1 as we did from M to M_1, it follows that N_0 is not a maximal element of P, a contradiction. We conclude $N_0 = X$ and complete the proof. □

An easy application of Hahn-Banach allows us to extend a bounded linear functional defined on a linear subspace to the entire space, while preserving the operator norm.

Corollary B.4. *Let $(X, \|\cdot\|)$ be a normed linear space over $\mathbb{R}$, M a linear subspace, and $f : M \to \mathbb{R}$ a bounded linear functional. Then, there exists $F \in X'$ such that $F|_M = f$ and $\|F\| = \|f\|$.*

Proof. Define $q : X \to \mathbb{R}$ by

$$q(x) := \|f\|\|x\|, \qquad \forall x \in \mathbb{R}.$$

Then, q is a sublinear functional on X with $f(x) \leqslant q(x)$ for all $x \in M$. Hahn-Banach Theorem extends f to a linear functional $F : X \to \mathbb{R}$ such that $F|_M = f$ and $F(x) \leqslant q(x) = \|f\|\|x\|$ for all $x \in X$. This implies $\|F\| \leqslant \|f\|$. On the other hand, $\|F\| \geqslant \|F|_M\| = \|f\|$. So we conclude $\|F\| = \|f\|$. □

This type of norm-preserving extension of bounded linear functionals allows us to explore a certain symmetry in the norms of X and its dual X'.

Corollary B.5. *If X is a normed linear space and $x \in X$, then*

$$\|x\| = \sup\{|f(x)| : f \in X', \|f\| \leqslant 1\}.$$

Moreover, this supremum is attained.

Proof. Let $\alpha = \sup\{|f(x)| : f \in X', \|f\| \leqslant 1\}$. If $f \in X'$ with $\|f\| \leqslant 1$, then $|f(x)| \leqslant \|f\|\|x\| \leqslant \|x\|$. This shows $\alpha \leqslant \|x\|$. On the other hand, let

$$M = \{\beta x : x \in \mathbb{R}\},$$

and define

$$g : M \to \mathbb{R} : \qquad g(\beta x) = \beta\|x\|.$$

Note $g \in M'$ and $\|g\| = 1$. By the preceding corollary, we can extend g to $f \in X'$ with $f(x) = g(x) = \|x\|$, while preserving the norm $\|f\| = \|g\| = 1$. By definition of α as supremum, we have $\alpha \geqslant |f(x)| = \|x\|$, and evidently this supremum is attained by f. □

Proposition B.6 (Geometric Hahn-Banach Theorem; [Con85] Corollary 6.8)**.** *Let X be a normed linear space over $\mathbb{R}$, M a closed linear subspace, and $x_0 \in X \backslash M$ with $d = \text{dist}(x_0, M)$. Then, there exists $f \in X'$ with $\|f\| = 1/d$ such that (i) $f(x_0) = 1$ and (ii) $f(x) = 0$ for all $x \in M$.*

Remark B.7. Since M is a closed linear subspace, $x_0 \notin X \backslash M$ implies $d = \text{dist}(x_0, M) > 0$. Indeed, suppose the contrary, so $0 = \text{dist}(x_0, M) = \inf_{y \in M} \|x_0 + y\|$. Then, there is a sequence $\{y_n\} \subseteq M$ such that $\lim_{n \to +\infty} \|x_0 + y_n\| = 0$. This means $\lim_{n \to +\infty} y_n = -x_0$ and this limit belongs to M because M is closed. But M is also a linear space, so we conclude $x_0 \in M$, contradicting $x_0 \in X \backslash M$.

The Geometric Hahn-Banach Theorem is interpreted to mean that a closed linear subspace M can be separated from any outside vector $x_0 \in X \backslash M$, not only by a positive distance, but also by a vector (bounded linear functional $f \in X'$) so that f is orthogonal to M but not orthogonal to x_0.

Proof of Proposition B.6. Let $Q : X \to X/M$ be the quotient map. By definition,

$$\|x_0 + M\| = \inf\{x_0 + y : y \in M\} = \text{dist}(x_0, M) = d > 0.$$

By preceding corollary, there is a bounded linear functional $g \in (X/M)'$ such that $g(x_0 + M) = d$ and $\|g\| = 1$. Define

$$f : X \to \mathbb{R}, \qquad x \mapsto d^{-1} g \circ Q(x).$$

Clearly, f is continuous, $f(x) = 0$ for all $x \in M$, and $f(x_0) = 1$.

To verify $\|f\| = d^{-1}$. Note on the one hand,

$$|f(x)| = d^{-1}|g \circ Q(x)| \leqslant d^{-1}\|g\|\|Q(x)\| \leqslant d^{-1}\|x\|, \qquad \forall x \in X,$$

and hence $\|f\| \leqslant d^{-1}$. On the other hand, since $\|g\| = 1$, by the definition of operator norm and continuity of g, there is a sequence $\{x_n\} \subseteq X$ such that $|g(x_n + M)| \to 1$ and $\|x_n + M\| < 1$ for all n. Let $\{y_n\} \subseteq M$ be a sequence such that $\|x_n + y_n\| < 1$. Then,

$$|f(x_n + y_n)| = |d^{-1} g(x_n + M)| \to d^{-1},$$

and hence $\|f\| \geqslant d^{-1}$. We conclude $\|f\| = d^{-1}$, as required. This completes the proof of Proposition B.6. □

Proposition B.8 (Bounded Linear Functionals Separate Points). *If x, y are two distinct points in a normed linear space X, then there is some bounded linear functional $f \in X^*$ for which $f(x) \neq f(y)$.*

Proof. At least one of x, y must be nonzero, so assume $x \neq 0$ without loss of generality. Take $M = \text{span}(x)$.

If $y \in M$, then define bounded linear functional f on M by setting $f(x) = 1$ and extend f to X by Hahn-Banach; since $y \neq x$, it follows that $y = \alpha x$ for some $\alpha \neq 1$ and hence $f(y) = f(\alpha x) = \alpha f(x) = \alpha \neq 1 = f(x)$, as required.

Now suppose $y \notin M$. So $d = \text{dist}(y, M) > 0$ because M is closed for being a 1-dimensional linear subspace. Then, Proposition B.6 yields some $f \in X^*$ for which $f(y) = 1$, $f(x) = 0$. This completes the proof. □

C Open Mapping Theorem

Theorem C.1 (Open Mapping Theorem; [Con85] Theorem 12.1)**.** *If X and Y are Banach spaces and $A : X \to Y$ is a continuous linear surjection, then $A(G)$ is open in Y whenever G is open in X.*

Proof. For $r > 0$, $x \in X$ and $y \in Y$, let $B_X(x, r) := \{x' \in X : \|x' - x\| < r\}$ denote the open ball in X centered at x of radius r and $B_Y(y, r) := \{y' \in Y : \|y' - y\| < r\}$ denote the open ball in Y centered at y of radius r. When $x = 0$ (or $y = 0$ respectively), we write simply $B_X(r) = B_X(0, r)$ (and $B_Y(r) = B_Y(0, r)$).

Claim 1. $0 \in \text{int cl } A(B(r))$ **for any** $r > 0$**.**

Proof fo Claim 1. Since $\bigcup_{k=1}^{\infty} B_X(kr/2) = X$ and A is surjective, we have

$$Y = \bigcup_{k=1}^{\infty} \text{cl } A(B_X(kr/2)) = \bigcup_{k=1}^{\infty} k \cdot \text{cl } A(B_X(r/2)).$$

Baire Category Theorem 1[17]**: Every complete metric space X is a Baire space, i.e., any countable intersection $\bigcap_{n\in\mathbb{N}} U_n$ of open dense sets $U_n \subseteq X$ is dense in X.**

Immediately, Banach spaces X and Y are Baire spaces. For a contradiction, suppose each closed set $\text{cl } A(B_X(kr/2))$, $k \geqslant 1$, has empty interior. Then each complement $Y \backslash \text{cl } A(B_X(kr/2))$, $k \geqslant 1$, is open and dense in Y. It follows that the countable intersection

$$\bigcap_{k=1}^{\infty} Y \backslash \text{cl } A(B_X(kr/2)) = Y \backslash \bigcup_{k=1}^{\infty} \text{cl } A(B_X(kr/2)) = \varnothing$$

is dense in Y, a contradiction. We thus conclude that there is some $k \geqslant 1$ for which $k \cdot \text{cl } A(B_X(r/2)) = \text{cl } A(B_X(kr/2))$ has nonempty interior. Hence, $V := \text{int cl } A(B_X(r/2)) \neq \varnothing$.

Let $y_0 \in V$ and $s > 0$ be such that $B_Y(y_0, s) \subseteq V \subseteq \text{cl } A(B_X(r/2))$. Let $y \in B_Y(s)$ so that $y_0 + y \in B_Y(y_0, s) \subseteq \text{cl } A(B_X(r/2))$. Now that both points $y_0, y_0 + y \in \text{cl } A(B_X(r/2))$, there are two sequences $\{x_n\}_n$ and $\{z_n\}_n$ in $B_X(r/2)$ such that $A(x_n) \to y_0$ and $A(z_n) \to y_0 + y$. We thus obtain $\{z_n - x_n\}_n \subseteq B_X(r)$ with $A(z_n - x_n) \to y \in B_Y(s)$. Since $y \in B_Y(s)$ was arbitrary, we have shown that $B_Y(s) \subseteq \text{cl } A(B_X(r))$. From $0 \in \text{int } B_Y(s)$, we conclude $0 \in \text{int cl } A(B_X(r))$, as claimed. □

Claim 2. $\text{cl } A(B_X(r/2)) \subseteq A(B_X(r))$ **for any** $r > 0$**.**

Proof of Claim 2. Fix $y_1 \in \text{cl } A(B_X(r/2))$. Since $0 \in \text{int cl } A(B_X(r/4))$ according to Claim 1, it follows that

$$y_1 \in [y_1 - \text{cl } A(B_X(r/4))] \cap A(B_X(r/2)) \neq \varnothing.$$

Let $x_1 \in B_X(r/2)$ such that $A(x_1) \in [y_1 - \text{cl } A(B_X(r/4))]$. Then, $A(x_1) = y_1 - y_2$ for some $y_2 \in \text{cl } A(B_X(r/4))$.

Continuing this way, we obtain two sequences

$$\{x_n\}_n \subseteq B_X(r/2^n) \quad \text{and} \quad \{y_n\}_n \subseteq \text{cl } A(B_X(r/2^n))$$

such that

$$A(x_n) = y_n - y_{n+1}.$$

Since $x_n \in B_X(r/2^n)$, it follows that $\|x_n\| < r/2^n$, hence the sequence $\{\sum_{n=1}^{N} x_n\}_N \subseteq X$ is Cauchy, and therefore the limit

$$x := \sum_{n=1}^{\infty} x_n \in X$$

exists with $\|x\| < r$.

Also,

$$\sum_{k=1}^{n} A(x_k) = \sum_{k=1}^{n} y_k - y_{k+1} = y_1 - y_{n+1}.$$

[17] For a great discussion of the Baire Category Theorem and the (strong) Open Mapping and Closed Graph Theorems as its consequences, see `https://www.ucl.ac.uk/~ucahad0/3103_handout_7.pdf`.

But $y_n \in \text{cl } A(B_X(r/2^n))$ implies that $\|y_n\| \leqslant \|A\| r/2^n$ and hence $y_n \to 0$. It follows from the continuity of A and $x \in B_X(r)$ that

$$y_1 = \sum_{k=1}^{\infty} A(x_k) = A(\sum_{k=1}^{\infty} x_k) = A(x) \in A(B_X(r)).$$

Since $y_1 \in \text{cl } A(B_X(r/2))$ was arbitrary, the claim is proven. □

We now return to the proof of the Open Mapping Theorem. Note Claims 1 and 2 together imply that

$$0 \in \text{int } A(B_X(r)), \qquad \forall r > 0.$$

Take any open set $G \subseteq X$. For each $x \in G$, let $r_x > 0$ be such that $B_X(x, r_x) \subseteq G$. Since $0 \in \text{int } A(B_X(r_x))$, it follows that

$$A(x) = A(x) + 0 \in A(x) + \text{int } A(B_X(r_x)) = \text{int}[A(x) + A(B_X(r_x))] = \text{int } A(x + B_X(r_x)) = \text{int } A(B_X(x, r_x)),$$

and hence there is some $s_x > 0$ for which $B_Y(A(x), s_x) \subseteq A(B_X(x, r_x))$. We then have

$$\bigcup_{x\in G} B_Y(A(x), s_x) \subseteq \bigcup_{x\in G} A(B_X(x, r_x)) = A(\bigcup_{x\in G} B_X(x, r_x)) = A(G) \subseteq \bigcup_{x\in G} B_Y(A(x), s_x).$$

This shows $A(G) = \bigcup_{x\in G} B_Y(A(x), s_x)$ is open and completes the proof of the Open Mapping Theorem. □

Corollary C.2 (Inverse Mapping Theorem; [Con85] Theorem 12.5)**.** *If X and Y are Banach spaces and $A : X \to Y$ is a bounded linear bijection, then its inverse A^{-1} is also bounded.*

Index

Bernoulli Map, 6

Characteristic Function, 78
Closed Linear Subspace, 93
Convergence in Distribution, 77

Distribution Function, 77
Doubling Map, 20

Eigenprojection, 69
Eigenvalue, 24
Eigenvector, 24

Gauss Map, 20

Inequality
  Doeblin-Fortet, 31, 37
Invariant Measure, 4
  Absolutely Continuous, 5

Nonsingular Transformation, 13, 18

Piecewise Monotone Interval Map, 20

Quasi-compactness, 28

Resolvent, 66

Spectral Gap, 17, 25
Spectral Radius, 16, 24
Spectrum, 16, 24
Sublinear Functional, 93

Theorem
  Analytic Perturbation, 69
  Central Limit Theorem, 77
  Geometric Hahn-Banach, 94
  Hahn-Banach, 93
  Hennion's, 31, 37
  Lévy's Continuity Theorem, 78
  Radon-Nikodym, 6, 13
  Riesz Representation, 90
Transfer Operator, 6, 18

Weak Convergence, 21

Zorn's Lemma, 93

References

[Cli13] Climenhaga. *Spectral methods in dynamics* Lecture Notes. January 30, 2013.

[Con85] Conway, John. *A Course in Functional Analysis.* Second Edition. Springer. 1985.

[Dur13] Durrett, Rick. *Probability: Theory and Examples*. April, 2013.

[LM94] Lasota, Andrzej & Mackey, Michael. *Chaos, Fractals, and Noise*. Second Edition. Springer New York, 1994.

[PC09] Pesin, Yakov & Climenhaga, Vaughn. *Lectures on Fractal Geometry and Dynamical Systems*. AMS. 2013.

[Rud87] Rudin, Walter. *Real and Complex Analysis.* McGraw-Hill International Editions Mathematics Series. 1987.

[Sar12] Sarig, Omri. *Introduction to the transfer operator method.* Lecture notes for Second Brazilian School on Dynamical Systems in São Carlos, SP, Brazil, October 2012.

[Sar20] Sarig, Omri. *Lecture Notes on Ergodic Theory*. 2020.

[Sau98] Saussol, Benoît. *Absolutely Continuous Invariant Measures for Multidimensional Expanding Maps.* Second version December 1998.

[VO16] Viana, Marcelo & Oliveira, Krerley. *Foundations of Ergodic Theory*. Cambridge University Press. 2016.

[Roc70] Rockafellar, R. Tyrrell *Convex Analysis*, Princeton Mathematical Series, 28, Princeton, N.J.: Princeton University Press, 1970.

[Hen01] Hennion, H. & Herve, L. *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-Compactness*, Lecture Notes in Mathematics, 1766, Springer-Verlag Berlin Heidelberg, 2001.

[Bal18] Baladi, V. *Dynamical Zeta Functions and Dynamical Determinants for Hyperbolic Maps*, A Series of Modern Surveys in Mathematics, 66, Springer International Publishing, 2018.