Omri Sarig

Introduction to the transfer operator method

Winter School on Dynamics

Hausdorff Research Institute for Mathematics, Bonn, January 2020

Contents

Lecture 1
The transfer operator

1.1 Motivation

A thought experiment Drop a little bit of ink into a glass of water, and then stir it with a tea spoon.

1. *Can you predict where individual ink particles will be after one minute?* NO: the motion of ink particles is chaotic.
2. *Can you predict the density profile of ink after one minute?* YES: it will be nearly uniform.

Gibbs's insight: For chaotic systems, it may be easier to predict the behavior of large collections of initial conditions, than to predict the behavior of individual initial conditions.

The transfer operator: The action of a dynamical system on mass densities of initial conditions.

1.2 Definition, basic properties, and examples

Setup. Let $T : X \to X$ be a *non–singular* measurable map on a σ–finite measure space $(X, \mathscr{B}, \mu)$. Non-singularity means that $\mu(T^{-1}E) = 0 \Leftrightarrow \mu(E) = 0$ $(E \in \mathscr{B})$. All the maps we consider in these notes are *non*–invertible.

The action of T on mass densities. Suppose we distribute mass on X according to the mass density $f d\mu$, $f \in L^1(\mu)$, $f \geq 0$, and then apply T to every point in the space. What will be the new mass distribution?

(The mass of points which land at E) $= \int 1_E(Tx)f(x)d\mu(x)$, ($1_E$ =indicator of E)

$$= \int (1_E \circ T)d\mu_f(x), \text{ where } \mu_f := f d\mu$$

$$= \int 1_E d\mu_f \circ T^{-1} = \int_E (\tfrac{d\mu_f \circ T^{-1}}{d\mu})d\mu \quad \text{(Radon-Nikodym derivative)}$$

Exercise 1.1. $\mu_f \circ T^{-1} \ll \mu$, therefore the Radon-Nikodym derivative exists.

Definition: The *transfer operator* of a non-singular map $(X, \mathscr{B}, \mu, T)$ is the operator $\widehat{T} : L^1(\mu) \to L^1(\mu)$ given by

$$\widehat{T} f = \frac{d\mu_f \circ T^{-1}}{d\mu}, \text{ where } \mu_f \text{ is the (signed) measure } \mu_f(E) := \int_E f\, d\mu.$$

The previous definition is difficult to work with. In practice one works with the following characterization of $\widehat{T} f$:

Proposition 1.1. *$\widehat{T} f$ is the unique element of $L^1(\mu)$ s.t. that for all test functions $\varphi \in L^\infty$, $\int \varphi \cdot (\widehat{T} f)\, d\mu = \int (\varphi \circ T) \cdot f\, d\mu$.*

Proof. The identity holds: For every $\varphi \in L^\infty$,

$$\int \varphi \cdot (\widehat{T} f)\, d\mu = \int \varphi \cdot \tfrac{d\mu_f \circ T^{-1}}{d\mu}\, d\mu = \int \varphi d\mu_f \circ T^{-1} \overset{!}{=} \int (\varphi \circ T) d\mu_f \overset{!}{=} \int (\varphi \circ T) f d\mu$$

(make sure you can justify all $\overset{!}{=}$).

The identity characterizes $\widehat{T} f$: Suppose $\exists h_1, h_2 \in L^1$ s.t. $\int \varphi h_i\, d\mu = \int (\varphi \circ T) f\, d\mu$ for all $\varphi \in L^\infty$. Choose $\varphi = \operatorname{sgn}(h_1 - h_2)$, then $\int |h_1 - h_2| d\mu = \int \varphi (h_1 - h_2) d\mu = \int \varphi h_1 d\mu - \int \varphi h_1 d\mu = \int \varphi \circ T f d\mu - \int \varphi \circ T f d\mu = 0$, whence $h_1 = h_2$ a.e. □

Proposition 1.2 (Basic properties).

1. *$\widehat{T}$ is a positive bounded linear operator with norm equal to one.*
2. *$\widehat{T}[(g \circ T) \cdot f] = g \cdot (\widehat{T} f)$ a.e. ($f \in L^1, g \in L^\infty$),*
3. *Suppose μ is a T-invariant probability measure, then $\forall f \in L^1$,*

$$(\widehat{T} f) \circ T = \mathbb{E}_\mu(f | T^{-1}\mathscr{B}) \text{ a.e.}$$

Proof of part 1: Linearity is trivial. Positivity means that if $f \geq 0$ a.e., then $\widehat{T} f \geq 0$ a.e. Let $\varphi := 1_{[\widehat{T} f < 0]}$, then $0 \geq \int_{[\widehat{T} f < 0]} (\widehat{T} f) d\mu = \int \varphi (\widehat{T} f) d\mu = \int \underbrace{(\varphi \circ T) f}_{\geq 0} d\mu \geq 0$.
It follows that $\int_{[\widehat{T} f < 0]} (\widehat{T} f) d\mu = 0$. This can only happen if $\mu[\widehat{T} f < 0] = 0$.

$\widehat{T}$ is bounded: Let $\varphi := \operatorname{sgn}(\widehat{T} f)$, then $\|\widehat{T} f\|_1 = \int \varphi (\widehat{T} f) d\mu = \int (\varphi \circ T) f\, d\mu \leq \|\varphi \circ T\|_\infty \|f\|_1 = \|f\|_1$, whence $\|\widehat{T} f\|_1 \leq \|f\|_1$. If $f > 0$, $\|\widehat{T} f\|_1 = \int |\widehat{T} f| d\mu = \int \widehat{T} f d\mu = \int (1 \circ T) f d\mu = \|f\|_1$, so $\|\widehat{T}\| = 1$. □

Exercise 1.2. Prove parts 2 and 3 of the proposition. (Hint for part 3: Show first that every $T^{-1}\mathscr{B}$–measurable function equals $\varphi\circ T$ with φ $\mathscr{B}$-measurable.)

Here are some examples of transfer operators.

Angle doubling map If $T:[0,1]\to[0,1]$ is $T(x)=2x \mod 1$, then $(\widehat{T}f)(x)=\frac{1}{2}[f(\frac{x}{2})+f(\frac{x+1}{2})]$.

Proof. For every $\varphi\in L^\infty$,

$$\begin{aligned}\int_0^1 \varphi(Tx)f(x)dx &= \int_0^{\frac{1}{2}}\varphi(2x)f(x)dx+\int_{\frac{1}{2}}^1\varphi(2x-1)f(x)dx\\ &= \int_0^1\varphi(t)f(\tfrac{t}{2})d(\tfrac{1}{2}t)+\int_0^1\varphi(s)f(\tfrac{s+1}{2})d(\tfrac{s+1}{2})\\ &= \int_0^1\varphi(x)\cdot\tfrac{1}{2}[f(\tfrac{x}{2})+f(\tfrac{x+1}{2})]dx.\end{aligned}$$

Exercise 1.3 (Gauss map). Let $T:[0,1]\to[0,1]$ be the map $T(x)=\{\frac{1}{x}\}$. Show that $(\widehat{T}f)(x)=\sum\limits_{n=1}^{\infty}\frac{1}{(x+n)^2}f(\frac{1}{x+n})$.

Exercise 1.4 (General piecewise monotonic map). Suppose $[0,1]$ is partitioned into finitely many intervals $I_1,\ldots,I_N$ and $T|_{I_k}:I_k\to T(I_k)$ is one-to-one and has continuously differentiable extension with non-zero derivative to an ε–neighborhood of I_k. Let $v_k:T(I_k)\to I_k$, $v_k:=(T|_{I_k})^{-1}$. Show that $\widehat{T}f=\sum\limits_{k=1}^{N}1_{T(I_k)}\cdot|v_k'|\cdot f\circ v_k$.

1.3 The transfer operator method

What dynamical information can we extract from the behavior of $\widehat{T}$?

Recall that $f_n\xrightarrow[n\to\infty]{}f$ weakly in L^1, if $\int\varphi f_n d\mu\xrightarrow[n\to\infty]{}\int\varphi f d\mu$ for all $\varphi\in L^\infty$. This is weaker than convergence in L^1 (give an example!).

Proposition 1.3 (Dynamical meaning of convergence of $\widehat{T}^n$).

1. *If $\widehat{T}^n f\xrightarrow[n\to\infty]{}h\int f d\mu$ weakly in L^1 for <u>some</u> non-negative $0\neq f\in L^1$ then T has an absolutely continuous invariant probability measure, and h is the density.*
2. *If $\widehat{T}^n f\xrightarrow[n\to\infty]{}\int f d\mu$ weakly in L^1 for <u>all</u> $f\in L^1$ then T is a mixing probability preserving map.*
3. *If $\widehat{T}^n f\xrightarrow[n\to\infty]{L^1}\int f d\mu$, then for every $\varphi\in L^\infty$,*

$$|\mathrm{Cov}(f,\varphi\circ T^n)|:=|\int f\varphi\circ T^n d\mu-\int f d\mu\int\varphi d\mu|\leq\|\widehat{T}^n f-\int f d\mu\|_1\|\varphi\|_\infty,$$

so the rate of decay of correlations against f is $O(\|\widehat{T}^n f-\int f d\mu\|_1)$.

Proof. 1. Assume w.l.o.g. that $\int f d\mu = 1$, then $\widehat{T}^n f \xrightarrow[n\to\infty]{w} h$. For every $\varphi \in L^\infty$, $\int \varphi h d\mu = \lim \int \varphi \cdot \widehat{T}^{n+1} f d\mu = \lim \int (\varphi \circ T)[\widehat{T}^n f] d\mu = \int (\varphi \circ T) h d\mu$. So $\mu_h := h d\mu$ is T-invariant.
2. exercise
3. $|\mathrm{Cov}(f, \varphi \circ T^n)| = |\int \widehat{T}^n f \varphi d\mu - \int (\int f d\mu) \varphi d\mu| = |\int (\widehat{T}^n f - \int f d\mu) \varphi d\mu|$. So $|\mathrm{Cov}(f, \varphi \circ T^n)| \leq \|\widehat{T}^n f - \int f d\mu\|_1 \|\varphi\|_\infty$. □

Exercise 1.5 (Dynamical interpretation of eigenvalues). Show:

1. All eigenvalues of the transfer operator have modulus less than or equal to one.
2. The invariant probability densities of T are the non-negative $h \in L^1(\mu)$ s.t. $\widehat{T} h = h$ and $\int h d\mu = 1$. We call $h d\mu$ an *acip*(=absolutely continuous invariant probability measure).
3. If $\widehat{T}$ has an acip and 1 is a simple eigenvalue of $\widehat{T}$, then the acip is ergodic. "Simple" means that $\dim\{g \in L^1 : \widehat{T} g = g\} = 1$.
4. If $\widehat{T}$ has an acip and 1 is simple, and all other eigenvalues of $\widehat{T}$ have modulus strictly smaller than one, then the acip is weak mixing.
5. If T is probability preserving and mixing, then $\widehat{T}$ has exactly one eigenvalue on the unit circle, equal to one, and this eigenvalue is simple. (Be careful not to confuse L^1-eigenvalues with L^2-eigenvalues.)

Further reading

1. *J. Aaronson:* An introduction to infinite ergodic theory, *Math. Surv. & Monographs* **50**, *Amer. Math. Soc., xii+284pp (1997)*
2. *V. Baladi:* Positive transfer operators and decay of correlations, *Adv. Ser. in Nonlinear Dynam.* **16** *World Scientific x+314pp. (2000)*

Lecture 2
Spectral gap

The transfer operator $\widehat{T} : L^1(\mu) \to L^1(\mu)$ of a non-singular transformation $(X, \mathscr{B}, \mu, T)$ describes the action of the map on mass densities. The density $f\,d\mu$ is moved after n iterations to $\widehat{T}^n f\,d\mu$. In this lecture we discuss a powerful method for analyzing the asymptotic behavior of $\widehat{T}^n f$ as $n \to \infty$ for "nice" functions f.

2.1 Quasi–compactness and spectral gap

Some operator theory. Suppose $\mathscr{L}$ is a Banach space and $L : \mathscr{L} \to \mathscr{L}$ is a bounded linear operator. We are interested in the behavior of L^n as $n \to \infty$. We review some relevant notions.

1. **Eigenvalues:** λ s.t. $Lv = \lambda v$ for some $0 \neq v \in \mathscr{L}$
2. **Spectrum:** $\text{spect}(L) := \{\lambda : (\lambda I - L) \text{ has no } \underline{\text{bounded}} \text{ inverse}\}$. Every eigenvalue belongs to the spectrum, but if $\dim(\mathscr{L}) = \infty$ then there could be points in the spectrum which are not eigenvalues.[1]
3. **Spectral radius:** $\rho(\mathscr{L}) := \sup\{|z| : z \in \text{Spect}(L)\}$.
4. **Spectral radius formula:** $\rho(L) = \lim_{n\to\infty} \sqrt[n]{\|L^n\|} = \inf_n \sqrt[n]{\|L^n\|}$. In particular, for every $\varepsilon > 0$, $\|L^n v\|/\|v\| = O(e^{n\varepsilon}\rho(L)^n)$ uniformly on $\mathscr{L} \setminus \{0\}$.

Spectral gap. $L : \mathscr{L} \to \mathscr{L}$ has *spectral gap* if we can write $L = \lambda P + N$ where

1. P is a projection (i.e. $P^2 = P$), and $\dim(\text{Im}(P)) = 1$;
2. N is a bounded operator s.t. $\rho(N) < |\lambda|$;
3. $PN = NP = 0$.

The commutation relations imply that $L^n = \lambda^n P + N^n$. Since $\rho(N) < |\lambda|$, for every $v \in \mathscr{L}$, $\|L^n v - \lambda^n P v\| = \|N^n v\| = o(|\lambda|^n)$. Therefore, if L has spectral gap, then

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

[1] Example: $L : L^1[0,1] \to L^1[0,1]$, $L[f(t)] = tf(t)$ has no eigenvalues, but its spectrum equals $[0,1]$.

Exercise 2.1 (Why call this "spectral gap"?). Use the following steps to show that λ is a simple eigenvalue and $\exists\gamma_0 > 0$ (the "gap") s.t.

$$\mathrm{Spect}(L) = \{\lambda\} \cup \text{ subset of } \{z : |z| \leq e^{-\gamma_0}|\lambda|\}.$$

1. $\mathrm{Im}(P) = \{h \in \mathscr{L} : Lh = \lambda h\}$. Consequently, λ is a simple eigenvalue.
2. Suppose $|z| > \rho(N)$, $z \neq \lambda$
   a. Solve the equation $(zI - L)v = w$ for $v \in \mathrm{Im}(P)$
   b. Solve the equation $(zI - L)v = w$ for $v \in \ker(P)$ (Hint: use $|z| > \rho(N)$)
   c. Show that $\mathscr{L} = \mathrm{Im}(P) \oplus \ker(P)$ and find a an explicit formula for the components of a vector according to this decomposition.
   d. Show that $(zI - L)$ has a bounded inverse on $\mathscr{L}$ whenever $z \neq \lambda$, $|z| > \rho(N)$.
3. Find a γ_0.

Quasi-compactness: This is a slightly weaker notion than spectral gap, which is easier to handle theoretically. A bounded linear operator L on a Banach space $\mathscr{L}$ is called *quasi–compact*, if there is a direct sum decomposition $\mathscr{L} = F \oplus H$ and $0 < \rho < \rho(L)$ where

1. F, H are closed and L–invariant: $L(F) \subset F, L(H) \subset H$
2. $\dim(F) < \infty$ and all eigenvalues of $L|_F : F \to F$ have modulus larger than ρ
3. the spectral radius of $L|_H$ is smaller than ρ

Quasi-compactness and spectral gap: If L is quasi-compact, and L has a unique eigenvalue on $\{z : |z| = \rho(L)\}$, and this eigenvalue has algebraic multiplicity one as an eigenvalue of the $\dim(F) \times \dim(F)$–matrix representing $L|_F : F \to F$, then L has spectral gap.

Exercise 2.2. Prove this using the following steps:

1. Show that if V is a Banach space, and $V = W_1 \oplus W_2$ where W_i are closed linear spaces, then the maps π_1, π_2 defined by $v = \pi_1(v) + \pi_2(v)$, $\pi_i(v) \in W_i$, are continuous linear maps s.t. $\pi_i^2 = \pi_i$, $\pi_1\pi_2 = \pi_2\pi_1 = 0$. (Hint: closed graph theorem)
2. Show that the Jordan form of $L|_F : F \to F$ consists of a 1×1 block with eigenvalue λ s.t. $|\lambda| = \rho(L)$, and (possibly) other Jordan blocks with eigenvalues λ_i s.t. $|\lambda_i| < |\lambda|$.
3. $\mathscr{L} = \mathrm{span}\{v\} \oplus H'$ where $Lv = \lambda v$, $L(H') \subset H'$, $\rho(L|_{H'}) < |\lambda|$
4. Deduce that L has spectral gap.

Exercise 2.3. Suppose $\widehat{T}$ is the transfer operator of a non-singular map $(X, \mathscr{B}, \mu, T)$, and assume $\mathscr{L} \subset L^1(\mu)$ possesses a norm $\|\cdot\|_{\mathscr{L}} \geq \|\cdot\|_1$ such that

1. $(\mathscr{L}, \|\cdot\|_{\mathscr{L}})$ is a Banach space which contains the constant functions
2. $\widehat{T}(\mathscr{L}) \subset \mathscr{L}$,
3. $\widehat{T} : \mathscr{L} \to \mathscr{L}$ is quasi-compact, with non-zero spectral radius.

If T is has a mixing absolutely continuous invariant probability density h then $\widehat{T}$ has spectral gap on $\mathscr{L}$ with $\lambda = 1$, and $Pf = h\int f d\mu$.

2.2 Sufficient conditions for quasi-compactness

The problem: The transfer operator typically does *not* have spectral gap on L^1.

The solution: Look for *smaller* Banach spaces $\mathscr{L} \subset L^1$ with $\|\cdot\|_{\mathscr{L}} \geq \|\cdot\|_1$ such that $\widehat{T}|_{\mathscr{L}} : \mathscr{L} \to \mathscr{L}$ has spectral gap. The result will be information on $\widehat{T}^n f$ for $f \in \mathscr{L}$.

The following theorem (a generalization of earlier results by Doeblin & Fortet and Ionescu–Tulcea & Marinescu) is a sufficient criterion for quasi–compactness. See the appendix for proof.

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

1. **Continuity:** $\|\cdot\|'$ *is continuous on* $\mathscr{L}$
2. **Pre-compactness:** *for any sequence of* $f_n \in \mathscr{L}$*, if* $\sup \|f_n\| < \infty$ *then there exists a subsequence* n_k *and* $g \in \mathscr{L}$ *s.t.* $\|Lf_{n_k} - g\|' \xrightarrow[k\to\infty]{} 0$
3. **Boundness:** $\exists M > 0$ *s.t.* $\|Lf\|' \leq M\|f\|'$ *for all* $f \in \mathscr{L}$
4. **Doeblin–Fortet inequality:** *there are* $k \geq 1$, $0 < r < \rho(L)$*, and* $R > 0$ *s.t.*

$$\|L^k f\| \leq r^k \|f\| + R\|f\|'. \tag{DF}$$

Then $L : \mathscr{L} \to \mathscr{L}$ *is quasi–compact.*

2.3 Application to continued fractions

Every $x \in [0,1] \setminus \mathbb{Q}$ can be uniquely expressed in the form $\cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cdots}}$ with $a_i(x) \in \mathbb{N}$. What can be said on the distribution of the $a_n(x)$ for $n \gg 1$?

Theorem (Gauss, Kuzmin, Lévy) *Let m denote Lebesgue's measure. For every natural number N,*

$$m\{x \in [0,1] : a_n(x) = N\} \xrightarrow[n\to\infty]{} \frac{1}{\ln 2} \int_{\frac{1}{N+1}}^{\frac{1}{N}} \frac{dx}{1+x} \text{ exponentially fast.}$$

Idea of proof: We use the Gauss map $T : [0,1] \to [0,1]$, $T(x) = \{\frac{1}{x}\}$. For every $x \in (0,1)$ irrational, $T^n(x) = \cfrac{1}{a_n(x) + \cfrac{1}{a_{n+1}(x) + \cdots}}$. So $a_n(x) = N$ iff $T^n(x) \in (\frac{1}{N+1}, \frac{1}{N})$, whence $m\{x : a_n(x) = N\} = \int 1_{(\frac{1}{N+1}, \frac{1}{N})} \circ T^n \, dx$.

We write the last expression in terms of the transfer operator of T:

$$m\{x : a_n(x) = N\} = \int 1_{(\frac{1}{N+1},\frac{1}{N})} \circ T^n \, dx = \int (\widehat{T}^n 1) 1_{(\frac{1}{N+1},\frac{1}{N})} dx = \int_{\frac{1}{N+1}}^{\frac{1}{N}} \widehat{T}^n 1 \, dx.$$

The idea is to find a Banach space $\mathscr{L}$ which contains the constant functions, and where $\widehat{T}$ is quasi-compact. The Gauss map has a mixing absolutely continuous invariant measure equal to $\frac{1}{\ln 2}\frac{dx}{1+x}$ (see appendix), so quasi-compactness implies spectral gap. Consequently, $\widehat{T}^n 1 \xrightarrow[n\to\infty]{\mathscr{L}} \frac{1}{\ln 2}\frac{1}{1+x}$ exponentially fast. If we can arrange $\|\cdot\|_{\mathscr{L}} \geq \|\cdot\|_1$, then $\widehat{T}^n 1 \xrightarrow[n\to\infty]{L^1} \frac{1}{\ln 2}\frac{1}{1+x}$ exponentially, and the theorem follows.

The Banach space: Let $\mathscr{L}$ denote the space of Lipschitz functions on $[0,1]$, with the norm $\|f\| := \|f\|_\infty + \mathrm{Lip}(f)$, where $\mathrm{Lip}(f) := \sup\left\{\frac{|f(x)-f(y)|}{|x-y|} : x \neq y\right\}$. Let $\|\cdot\|'$ denote the L^1 norm: $\|f\|' := \int_0^1 |f(x)|dx$.

Exercise 2.4. $\mathscr{L}$ is a Banach space, and for all $f,g \in \mathscr{L}$,

1. $\|fg\| \leq \|f\| \cdot \|g\|$
2. $\|\frac{1}{(a+x)^2}\| \leq 3/a^2$ for all $a \geq 1$.
3. $\|f(\frac{1}{a+x})\| \leq \|f\|$

Recall that $\widehat{T}f = \sum_{a\geq 1} \frac{1}{(a+x)^2} f(\frac{1}{a+x})$. We claim that $\widehat{T}(\mathscr{L}) \subset \mathscr{L}$ and $T : \mathscr{L} \to \mathscr{L}$ is bounded. The sum converges absolutely in norm, because $\sum_{a\geq 1} \|\frac{1}{(a+x)^2} f(\frac{1}{a+x})\| \leq \sum_{a\geq 1} \|\frac{1}{(a+x)^2}\| \cdot \|f(\frac{1}{a+x})\| \leq (3\sum \frac{1}{a^2})\|f\|$. So $\widehat{T}(\mathscr{L}) \subset \mathscr{L}$ and $\|\widehat{T}\| \leq 3\sum \frac{1}{a^2}$.

Next we check the conditions of Hennion's theorem.

1. **Continuity:** If $\|f_n - f\| \xrightarrow[n\to\infty]{\mathscr{L}} 0$, then $\|f_n - f\|_\infty \xrightarrow[n\to\infty]{} 0$, so $\|f_n - f\|_1 \xrightarrow[n\to\infty]{} 0$. It follows that $\|f_n\|' = \|f_n\|_1 \xrightarrow[n\to\infty]{} \|f\|_1 = \|f\|'$.
2. **Pre-compactness:** Suppose $\{f_n\}$ is bounded in the Lipschitz norm. By the Arzelà–Ascoli theorem there is a subsequence n_k s.t. $f_{n_k} \xrightarrow[k\to\infty]{} f$ uniformly on $[0,1]$. Necessarily $\mathrm{Lip}(f) \leq \sup \mathrm{Lip}(f_{n_k}) < \infty$.

   Uniform convergence implies convergence in $L^1[0,1]$, so $\|f_{n_k} - f\|_1 \to 0$. Since $\widehat{T}$ is a bounded operator on L^1, $\|\widehat{T}f_{n_k} - \widehat{T}f\|_1 \to 0$, equivalently, $\|\widehat{T}f_{n_k} - \widehat{T}f\|' \to 0$. The limit $\widehat{T}f$ is in $\mathscr{L}$ because $f \in \mathscr{L}$ and $\widehat{T}(\mathscr{L}) \subset \mathscr{L}$.
3. **Boundness:** $\|\widehat{T}f\|' = \|\widehat{T}f\|_1 \leq \|f\|_1 = \|f\|'$.
4. **Doeblin–Fortet Inequality:** The proof is based on the following facts.

   Exercise 2.5. Let $v_a(x) := \frac{1}{a+x}$, $v_{a_1,\ldots,a_n} := v_{a_0} \circ \cdots \circ v_{a_n}$, and $[\underline{a}] := v_{\underline{a}}[0,1)$.

   a. $\widehat{T}^n f = \sum_{a_1,\ldots,a_n=1}^{\infty} |v'_{a_1,\ldots,a_n}| f \circ v_{a_1,\ldots,a_n}$ (Hint: start with $n=1$ and iterate)
   b. $\exists C > 0$ and $0 < \theta < 1$ s.t. for all $n \geq 1$ and $\underline{a} = a_1 a_2 \cdots a_n$, $|v_{\underline{a}}(x) - v_{\underline{a}}(y)| < C\theta^n |x-y|$. (Hint: T^2 is expanding).

c. $\exists H > 1$ s.t. for all $x, y \in [0,1]$, $n \geq 1$, and $\underline{a} = a_1 a_2 \cdots a_n$, $\left|\frac{v'_{\underline{a}}(x)}{v'_{\underline{a}}(y)} - 1\right| \leq H|x-y|$.
   (Hint: $\ln|v'_{\underline{a}}(x)| = \sum_{i=0}^{n-1} \ln|v'_{a_i}(v_{a_1 \cdots a_{i-1}}(x))|$, $\mathrm{Lip}(v'_a) \leq 1$.)
d. $\exists G > 1$ s.t. $\forall x \in [0,1]$ and $\underline{a}$, $G^{-1} \cdot m[\underline{a}] \leq |v'_{\underline{a}}(x)| \leq G \cdot m[\underline{a}]$ (Hint: Use (c) to relate $|v'_{\underline{a}}(x)|$ to $\int_0^1 |v'_{\underline{a}}(t)|dt$ and calculate the integral.)
e. $[\underline{a}]$ are non-overlapping sub-intervals of $[0,1)$.

Proof of the Doeblin-Fortet Inequality: Suppose f is Lipschitz, we estimate the Lipschitz constant of $\widehat{T}^n f$:

$$|(\widehat{T}^n f)(x) - (\widehat{T}^n f)(y)| \leq \sum_{a_1,\ldots,a_n=1}^{\infty} \big(|v'_{\underline{a}}(x) - v'_{\underline{a}}(y)||f(v_{\underline{a}}(x))| + |v'_{\underline{a}}(y)||f(v_{\underline{a}}(x)) - f(v_{\underline{a}}(y))|\big)$$

$$\leq \sum_{a_1,\ldots,a_n=1}^{\infty} |v'_{\underline{a}}(y)| \left|\frac{v'_{\underline{a}}(x)}{v'_{\underline{a}}(y)} - 1\right| |f(v_{\underline{a}}(x))| + \sum_{a_1,\ldots,a_n=1}^{\infty} \|v'_{\underline{a}}\|_\infty \mathrm{Lip}(f)|v_{\underline{a}}(x) - v_{\underline{a}}(y)|$$

Using the exercise and the trivial fact that if f is Lipschitz on an interval J, then for every $x \in J$, $|f(x)| \leq \frac{1}{|J|}\int_J |f(t)|dt + \mathrm{Lip}(f)|J|$, we obtain

$$|(\widehat{T}^n f)(x) - (\widehat{T}^n f)(y)| \leq \sum_{a_1,\ldots,a_n=1}^{\infty} G m[\underline{a}] \cdot H|x-y| \cdot \left(\frac{1}{m[\underline{a}]}\int_{[\underline{a}]} |f(t)|dt + \mathrm{Lip}(f)C\theta^n\right)$$
$$+ \sum_{a_1,\ldots,a_n=1}^{\infty} G \cdot m[\underline{a}]\mathrm{Lip}(f)C\theta^n|x-y|.$$

Since $[\underline{a}]$ are non-overlapping sub-intervals of $[0,1]$, $\sum m[\underline{a}] \leq 1$. It follows that $|(\widehat{T}^n f)(x) - (\widehat{T}^n f)(y)| \leq \Big(GH\|f\|_1 + GC(H+1)\theta^n \mathrm{Lip}(f)\Big)|x-y|$, whence

$$\mathrm{Lip}(\widehat{T}^n f) \leq \Big(\text{const.}\, \|f\|_1 + \text{const.}\, \theta^n \mathrm{Lip}(f)\Big).$$

Next we estimate $\|\widehat{T}^n f\|_\infty$. Since $|\widehat{T}^n f(x)| \leq \int |(\widehat{T}^n f)(y)|dy + \mathrm{Lip}(\widehat{T}^n f)$,

$$\|\widehat{T}^n f(x)\|_\infty \leq \|f\|_1 + \mathrm{Lip}(\widehat{T}^n f)$$

In summary, $\|\widehat{T}^n f\| \leq \text{const.}\, \theta^n \mathrm{Lip}(f) + \text{const.}\, \|f\|_1$. The Doeblin–Fortet inequality follows by slightly increasing θ and taking n sufficiently large. ☐

Further reading

1. *J. Aaronson:* An introduction to infinite ergodic theory, *Math. Surv. & Monographs* **50**, *Amer. Math. Soc., xii+284pp (1997)*
2. *V. Baladi:* Positive transfer operators and decay of correlations, *Adv. Ser. in Nonlinear Dynam.* **16** *World Scientific x+314pp. (2000).*

3. *H. Hennion and L. Hervé:* Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness, *Lect. Notes in Math.* **1766**,*Springer, 145pp (2000)*
4. *A. Ya. Khinchin:* Continued fractions. Third edition. *Dover, xi+95pp (1992)*
5. D.H. Mayer: *Continued fractions and related transformations*, "Ergodic Theory, Symbolic Dynamics and Hyperbolic spaces", *edited by T. Bedford, M. Keane, and C. Series. Oxford Science Publications*, 175–229 (1991)

Lecture 3
Analytic perturbation theory

Our next application of the transfer operator method is the central limit theorem. This requires studying (complex) one–parameter families of transfer operators. In this lecture, we develop the tools from functional analysis needed to do this.

3.1 Calculus in Banach spaces

Setup: $\mathscr{L}$ is a Banach space, $B = B(\mathscr{L})$ is the space of all bounded linear operators $L : \mathscr{L} \to \mathscr{L}$ with the norm $\|L\| = \sup \frac{\|Lx\|}{\|x\|}$, and $\mathscr{L}^*$ and B^* are the spaces of all bounded linear functionals on L and B, with the norm $\|\varphi\| = \sup \frac{|\varphi(x)|}{\|x\|}$.

We are interested in (complex) one–parameter families $L_z \in B$, $(z \in U)$, where $U \subset \mathbb{C}$ is open. Formally these are functions $L : U \to B$, $L(z) = L_z$.

Line integrals: Let $\gamma \subset \mathbb{C}$ be a curve with smooth parametrization $z(t)$, $a \leq t \leq b$, and let $L : \gamma \to B$ be continuous. We define the *line integral* of L along γ by

$$\int_\gamma L(z)dz := \text{the limit (in } B\text{) of the Riemann sums } \sum_{i=1}^{N} L(z(\xi_i))[z(t_{i+1}) - z(t_i)],$$

where $a < t_1 < \cdots < t_n = b$, $\xi_i \in [t_i, t_{i+1}]$, and $\max\{|t_{i+1} - t_i| : 1 \leq i \leq n\} \to 0$.

As in the case of complex valued functions, if $L : \gamma \to B$ is continuous, then the limit exists and is independent of the parametrization (exercise).

Exercise 3.1. Suppose $L : \gamma \to B$ is continuous. For every $\varphi \in \mathscr{L}^*$ and $T \in B$, $\varphi[\int_\gamma L(z)dz] = \int_\gamma \varphi[L(z)]dz$ and $T[\int_\gamma L(z)dz] = \int_\gamma T[L(z)]dz$.

Analyticity and derivatives: Suppose $U \subset \mathbb{C}$ is open and z_0 is a point in U. We call $L : U \to B$ *analytic* (or *holomorphic*) at z_0, if there is an element $L'(z_0) \in B$ such that $\|\frac{L(z_0+h) - L(z_0)}{h} - L'(z_0)\| \xrightarrow[|h| \to 0]{} 0$. $L'(z_0)$ is called the *derivative* at z_0.

Exercise 3.2 (Rules of differentiation). If $L, L_1, L_2 : U \to B$ are analytic, then

1. $(L_1 + L_2)' = L_1' + L_2'$
2. $(L_1 L_2)' = L_1' L_2 + L_1 L_2'$
3. in case L is invertible, $(L^{-1})' = -L^{-1} L' L^{-1}$
4. for every bounded linear functional $\varphi : B \to \mathbb{C}$, $\frac{d}{dz}[\varphi \circ L] = \varphi \circ L'$

Analyticity Theorem (Dunford): *Suppose $U \subset \mathbb{C}$ is open. $L(z)$ is analytic on U iff for every $\varphi \in B^*$, $\varphi[L(z)]$ is holomorphic on U in the usual sense of complex functions.* (See the appendix for proof).

Cauchy's integral formula (Wiener): *If $L : U \to B$ is analytic on U, then L is differentiable infinitely many times on U, and for every $z \in U$ and every simple closed smooth curve $\gamma \subset U$ around z, $L(z) = \frac{1}{2\pi i} \oint_\gamma \frac{L(\xi)}{\xi - z} d\xi$* and $L^{(n)}(z) = \frac{n!}{2\pi i} \oint_\gamma \frac{L(\xi)}{(\xi - z)^{n+1}} d\xi$.

Proof. For every bounded linear functional φ, $\varphi[L(z)]$ is holomorphic. Therefore $\frac{d}{dz}\varphi[L(z)] = \varphi[L'(z)]$ is holomorphic. Therefore $L'(z)$ is analytic. By induction, $L(z)$ is differentiable infinitely many times.

Next, for every bounded linear functional φ, we have by Cauchy's integral formula for the complex valued holomorphic function $\varphi[L(z)]$ that

$$\varphi\left[\frac{1}{2\pi i} \oint_\gamma \frac{L(\xi)}{\xi - z} d\xi\right] = \frac{1}{2\pi i} \oint_\gamma \frac{\varphi[L(\xi)]}{\xi - z} d\xi = \varphi[L(z)].$$

Bounded linear functionals separate points, so $\frac{1}{2\pi i} \oint_\gamma \frac{L(\xi)}{\xi - z} d\xi = L(z)$. The identity for higher derivatives is proved the same way and is left as an exercise. $\square$

Exercise 3.3. If $L(z)$ is analytic on U and $\gamma \subset U$ is a simple closed smooth curve, then $\oint_\gamma L(z) dz = 0$.

Exercise 3.4. If $\|T_n\| = O(r^n)$, then $\sum (z - a)^n T_n$ is analytic on $\{z : |z - a| < 1/r\}$.

Exercise 3.5. $L : U \to B$ is analytic on an open subset U iff $\forall a \in U$ $\exists L_n(a) \in B$, $r(a) > 0$ s.t. $\|L_n(a)\| = O(r(a)^n)$ and $L(z) = \sum (z - a)^n L_n(a)$ on $\{z : |z - a| < r(a)\}$. (Hint: Expand the integrand in Cauchy's formula in powers of $z - a$)

3.2 Resolvents and eigenprojections

Spectrum: The *spectrum* of a bounded linear operator L is

$$\mathrm{Spect}(L) = \{z \in \mathbb{C} : (zI - L) \text{ has no } \underline{\text{bounded}} \text{ inverse}\}.$$

Exercise 3.6. Show that $\mathrm{Spect}(L)$ is compact, using the following steps:

1. If $\|L\| < 1$, then $I - L$ has bounded inverse and $(I - L)^{-1} = I + L + L^2 + L^3 + \cdots$.
2. $zI - L$ has bounded inverse for all $|z|$ large enough.
3. If $I - L$ has bounded inverse, so does $I - L_1$ whenever $\|L_1 - L\|$ is small enough.
4. $\mathrm{Spect}(L)$ is compact.

Resolvent: On the complement of the spectrum, one can define the *resolvent*: $R(z) := (zI - L)^{-1} \quad (z \notin \mathrm{Spect}(L))$.

Exercise 3.7 (Properties of the resolvent). Show

1. **Commutation:** $R(z)L = LR(z)$
2. **Resolvent identity:** $R(w) - R(z) = (z - w)R(z)R(w)$
3. **Analyticity:** $R(z)$ is analytic on the complement of $\mathrm{Spect}(L)$, with expansion $R(z) = \sum_{n=0}^{\infty}(-1)^n(z - z_0)^n R(z_0)^{n+1}$ for all $z_0 \notin \mathrm{Spect}(L)$ and $|z - z_0|$ small.

Separation of Spectrum Theorem (Sz.-Nagy, Wolf): *Suppose* $\mathrm{Spect}(L) = \Sigma_{in} \uplus \Sigma_{out}$ *where* $\Sigma_{in}, \Sigma_{out}$ *are compact, and let* γ *be a smooth closed curve which does not intersect* $\mathrm{Spect}(L)$*, and which contains* Σ_{in} *in its interior, and* Σ_{out} *in its exterior. Then:*

1. $P := \frac{1}{2\pi i}\oint_\gamma (zI - L)^{-1}dz$ *is a projection (*$P^2 = P$*), therefore* $\mathscr{L} = \ker(P) \oplus \mathrm{Im}(P)$.
2. $PL = LP$*, therefore* $L(\ker(P)) \subset \ker(P)$ *and* $L(\mathrm{Im}(P)) \subset \mathrm{Im}(P)$.
3. $\mathrm{Spect}(L|_{\mathrm{Im}(P)}) = \Sigma_{in}$ *and* $\mathrm{Spect}(L|_{\ker(P)}) = \Sigma_{out}$.

(The proof is in the appendix. It's worth reading.)

Eigenprojections: $P := \frac{1}{2\pi i}\oint_\gamma (zI - L)^{-1}dz$ is called the *eigenprojection* of Σ_{in}.

Exercise 3.8. Suppose L has spectral gap with representation $L = \lambda P + N$. Show that the eigenprojection of λ equals P.

3.3 Analytic perturbations of operators with spectral gap

Setup: Let $\{L_z\}_{z\in U}$ be a family of bounded linear operators on a Banach space $\mathscr{L}$, such that $z \mapsto L_z$ is analytic.

Analytic perturbation theorem (Rellich, Sz.-Nagy, Wolf, Kato): *Suppose* L_0 *has spectral gap with representation* $\lambda P + N$*, then there are* $\varepsilon, \kappa > 0$ *s.t. for all* $|z| < \varepsilon$*,* L_z *has spectral gap with representation* $\lambda_z P_z + N_z$*, where* λ_z, P_z, N_z *are analytic on* $\{z : |z| < \varepsilon\}$*, and* $\rho(N_z) < |\lambda_z| - \kappa$.

Sketch of proof: We saw that when L_0 has spectral gap, $\mathrm{Spect}(L_0) = \{\lambda_0\} \uplus \Sigma$ where $\Sigma \subset \{z : |z| < \rho(L_0)\}$. Let γ be a small circle around λ_0 s.t. Σ is outside γ.

Step 1. $\exists \varepsilon_1 > 0$ *s.t.* γ *does not intersect* $\mathrm{Spect}(L_z)$ *for any* $|z| < \varepsilon_1$.

Proof. For every $\xi \in \gamma$, $\xi I - L_0$ has a bounded inverse. The property of having a bounded inverse is open (exercise 3.6(3)), therefore

$$\Lambda := \{(\xi, z) \in \mathbb{C} \times \mathbb{C} : \xi I - L_z \text{ has a bounded inverse}\}$$

is an open neighborhood of the compact set $\gamma \times \{0\}$. By compactness, there is a positive ε s.t. $\Lambda \supset \gamma \times \{z : |z| < \varepsilon\}$. This is ε_1.

Step 2: *For every* $|z| < \varepsilon_1$, $P_z := \frac{1}{2\pi i}\oint_\gamma (\xi I - L_z)^{-1} d\xi$ *is a projection and* $P_z L_z = L_z P_z$. *There exists* $0 < \varepsilon_2 < \varepsilon_1$ *s.t.* P_z *is analytic on* $\{z : |z| < \varepsilon_2\}$.

Proof. P_z is a projection, because of the theorem on separation of spectrum and the last step which says that γ does not intersect $\mathrm{Spect}(L_z)$. The analyticity of P_z is shown by direct expansion of the integrand in terms of z. We omit the details which are routine, but tedious.

Step 3: $\exists 0 < \varepsilon_3 < \varepsilon_2$ *s.t.* $\dim(\mathrm{Im}(P_z)) = 1$ *for all* $|z| < \varepsilon_3$.

Proof. Two linear operators P, Q are called *similar*, if there is a linear isomorphism π s.t. $P = \pi^{-1} Q \pi$. The step is based on the the following lemma due to Kato (appendix): *Suppose P is a projection. Any projection Q s.t.* $\|Q - P\| < 1$ *is similar to P.* It follows that if $|z|$ is so small that $\|P_z - P_0\| < 1$, then $\dim(\mathrm{Im}(P_z)) = \dim(\mathrm{Im}(P_0))$. Since L_0 has spectral gap, this dimension is one.

Step 4: $L_z P_z = \lambda_z P_z$ *where* $z \mapsto \lambda_z$ *is analytic on a neighborhood of zero.*

Proof: Suppose $|z| < \varepsilon_3$. Since $P_z L_z = L_z P_z$, $\mathrm{Im}(P_z)$ is L_z–invariant. Since $\dim \mathrm{Im}(P_z) = 1$, $L_z : \mathrm{Im}(P_z) \to \mathrm{Im}(P_z)$ takes the form $f \mapsto \lambda_z f$ for some scalar λ_z. So $L_z P_z = \lambda_z P_z$.

The eigenvalue λ_z depends analytically on z on some neighborhood of zero: Take some $f \in \mathscr{L}$ and $\varphi \in \mathscr{L}^*$ s.t. $\varphi(P_0 f) > 0$. There exists $0 < \varepsilon_4 < \varepsilon_3$ s.t. $\varphi(P_z f) > 0$ for all $|z| < \varepsilon_4$. The formula

$$\lambda_z = \frac{\varphi(L_z P_z f)}{\varphi(P_z f)}$$

shows that λ_z is analytic on $\{z : |z| < \varepsilon_4\}$.

Step 5: *There's a neighborhood of zero where* $N_z := L_z(I - P_z)$ *is analytic, and where* $N_z P_z = P_z N_z = 0$*, and* $\rho(N_z) < |\lambda_z|$.

Proof: $N_z = L_z(I - P_z)$ is analytic on $\{z : |z| < \varepsilon_3\}$, because L_z, P_z are analytic there. $P_z^2 = P_z$ and $L_z P_z = P_z L_z = \lambda_z P_z$ imply that $P_z N_z = N_z P_z = 0$ and $L_z = \lambda_z P_z + N_z$.

The spectral radius formula states that $\rho(N_z) = \lim_{n\to\infty} \sqrt[n]{\|N_z^n\|}$. Since $\|N_z^{n+m}\| \le \|N_z^n\|\|N_z^m\|$, $\rho(N_z) = \inf \sqrt[n]{\|N_z^n\|}$. Fix some $\delta > 0$ (to be determined later). Pick some n s.t. $\sqrt[n]{\|N_0^n\|} < e^{\delta}\rho(N_0)$. Since $z \mapsto \|N_z^n\|$ is continuous, there exists $0 < \varepsilon_5 < \varepsilon_4$ s.t. $\sqrt[n]{\|N_z^n\|} < e^{2\delta}\rho(N_0)$ for all $|z| < \varepsilon_5$.

Similarly, there is $0 < \varepsilon_6 < \varepsilon_5$ s.t. $|\lambda_z| > e^{-\delta}|\lambda_0|$ for all $|z| < \varepsilon_5$. Choosing δ so small that $e^{3\delta}\rho(L_0) < |\lambda_0|$ we get a neighborhood of zero where $\rho(N_z) < |\lambda_z|$. □

Further reading

T. Kato: Perturbation theory for linear operators, *Classics in Math., Springer, xxi+619pp (1980)*

Lecture 4
Application to the Central Limit Theorem

4.1 Spectral gap and the central limit theorem

Setup: Let $(X,\mathscr{B},T,\mu)$ be a mixing, probability preserving map. Suppose $\widehat{T}$ has spectral gap on some Banach space of functions $\mathscr{L}$ which contains the constants, is closed under multiplication, and which satisfies the inequalities

$$\|fg\| \leq \|f\|\|g\| \text{ and } \|\cdot\| \geq \|\cdot\|_1.$$

(Example: The transfer operator of the Gauss map, acting on the space of Lipschitz functions on $[0,1]$.) In this lecture we show:

Central Limit Theorem: *Let $\psi \in \mathscr{L}$ be bounded with integral zero. If $\nexists v \in \mathscr{L}$ s.t. $\psi = v - v\circ T$ a.e., then $\exists \sigma > 0$ s.t. $\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\psi\circ T^n \xrightarrow[n\to\infty]{dist} N(0,\sigma^2)$, i.e.*

$$\mu\left\{x : \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\psi\circ T^n \in [a,b]\right\} \to \frac{1}{\sqrt{2\pi\sigma^2}}\int_a^b e^{-t^2/2\sigma^2}dt \text{ for all intervals } [a,b].$$

Here and throughout, $N(0,\sigma^2)$ denotes the Gaussian distribution with mean zero and standard deviation σ. The CLT as stated and proved here is an abstraction of results due to Doeblin & Fortet, Nagaev, Rousseau-Egele, and Guivarc'h & Hardy.

4.2 Background from probability theory

Distribution functions: Suppose X is a real valued random variable. The *distribution function* of X is $F_X : \mathbb{R} \to [0,1]$, $F_X(t) := \mathbb{P}[X < t]$.

Convergence in distribution: Let X_n, Y denote random variables (possibly defined on different probability spaces). We say that $X_n \xrightarrow[n\to\infty]{\text{dist}} Y$, if $\mathbb{P}[X_n < t] \xrightarrow[n\to\infty]{} \mathbb{P}[Y < t]$ for all t where $F_Y(t) = \mathbb{P}[Y < t]$ is continuous.

The reason we only ask for convergence at continuity points of $F_Y(t)$ is to deal with cases such as $X_n = 2 - \frac{1}{n}$, $Y = 2$. We would like to say that $X_n \xrightarrow[n\to\infty]{dist} Y$, eventhough $\mathbb{P}[X_n < 2] \not\to \mathbb{P}[Y < 2]$.

Characteristic functions: The *characteristic function* of a real valued random variable X is $\varphi_X(t) = \mathbb{E}(e^{itX})$.

The characteristic function is the Fourier transform of the unique measure μ_F on $\mathbb{R}$ such that $\mu_F([a,b)) = \text{Prob}(a \leq X < b)$. Characteristic functions are useful because of the following result, which connects the theory of convergence in distribution to harmonic analysis:

Lévy's continuity theorem: *A sequence of random variables X_n converges in distribution to a random variable Y iff $\mathbb{E}(e^{itX_n}) \xrightarrow[n\to\infty]{} \mathbb{E}(e^{itY})$ for all $t \in \mathbb{R}$.*

If $F_Y(t)$ is continuous (e.g. Y gaussian), there is even a way to estimate $\|F_X - F_Y\|_\infty$ in terms of the distance between φ_X, φ_N (see appendix):

The "smoothing inequality" (Berry & Esseen): $\exists C > 0$ *s.t. for every pair of real valued random variables X, N s.t. that F_N is differentiable,* $\sup |F_N'| < \infty$*, and* $\int |F_X - F_N| dx < \infty$*, then*

$$\|F_X - F_N\|_\infty \leq C \left(\frac{1}{2\pi} \int_{-T}^{T} \frac{|\varphi_X(t) - \varphi_N(t)|}{|t|} dt + \frac{\sup |F_N'|}{T} \right) \text{ for all } T > 0.$$

T is a free parameter which we are free to choose to optimize the bound.

Exercise 4.1. Use the smoothing inequality to prove Lévy's continuity theorem in the particular case $Y = N(0, \sigma^2)$. You may use the fact that the characteristic function of $N(0, \sigma^2)$ is $e^{-\frac{1}{2}\sigma^2 t^2}$.

4.3 The proof of the central limit theorem (Nagaev's method)

Let $\psi_n := \psi + \psi \circ T + \cdots + \psi \circ T^{n-1}$. By Lévy's continuity theorem (or exercise 4.1), it is enough to show that $\mathbb{E}(e^{i\frac{t}{\sqrt{n}}\psi_n}) \equiv \int e^{i\frac{t}{\sqrt{n}}\psi_n} d\mu \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}\sigma^2 t^2}$.

Nagaev's perturbation operators: Define a new operator by $\widehat{T}_t f = \widehat{T}(e^{it\psi} f)$. We think of these as of perturbations of $\widehat{T} \equiv \widehat{T}_0$ for $t \approx 0$.

$\widehat{T}_t$ are bounded linear operators on $\mathscr{L}$, because $\widehat{T}_t f = \widehat{T}(\sum_{k=0}^{\infty} \frac{(it)^k}{k!} \psi^k f)$, whence by our assumptions on $\mathscr{L}$ $\|\widehat{T}_t f\| \leq \|\widehat{T}\| \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \|\psi\|^k \|f\|$, and $\|\widehat{T}_t\| \leq e^{|t| \|\psi\|} \|\widehat{T}\|$.

Exercise 4.2 (Nagaev's identity). $\widehat{T}_t^n f = \widehat{T}^n (e^{it\psi_n} f)$.

Note that $\mathbb{E}(e^{it\psi_n}) = \int e^{it\psi_n} d\mu = \int 1 \circ T^n e^{it\psi_n} d\mu = \int \widehat{T}^n(e^{it\psi_n}) d\mu$, whence

$$\mathbb{E}(e^{it\psi_n}) = \int \widehat{T}_t^n 1 d\mu.$$

Nagaev's method is to use analytic perturbation theory of $\widehat{T}_0 \equiv \widehat{T}$ to show that $\mathbb{E}(e^{it\psi_n/\sqrt{n}}) \equiv \int \widehat{T}^n_{\frac{t}{\sqrt{n}}} 1 d\mu \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}\sigma^2 t^2}$ for some σ.

Analytic perturbation theory. We replace $t \in \mathbb{R}$ by $z \in \mathbb{C}$ and claim that $z \mapsto \widehat{T}_z$ is analytic. This can be seen from the expansion

$$\widehat{T}_z = I + \sum_{n=1}^{\infty} \frac{(iz)^n}{n!} \widehat{T} M_\psi^n, \text{ where } M_\psi : \mathscr{L} \to \mathscr{L} \text{ is } M_\psi f = \psi f.$$

M_ψ is bounded, because $\|M_\psi f\| \le \|\psi\| \|f\|$. Therefore $\|\widehat{T} M_\psi^n\| \le \|\widehat{T}\| \|\psi\|^n$ and the series converges in norm on $\mathbb{C}$. By exercise 3.4, $\widehat{T}_z$ is analytic on $\mathbb{C}$.

Exercise 4.3. $\widehat{T}_z' = i\widehat{T}_z M_\psi, \widehat{T}_z'' = -\widehat{T}_z M_\psi^2$, $(\widehat{T}_z^n)' = i\widehat{T}_z^n M_{\psi_n}, (\widehat{T}_z^n)'' = -\widehat{T}_z^n M_{\psi_n}^2$.
(Hint: To find the derivatives for $n > 1$, use exercise 3.2 and proposition 1.2(1).)

By our assumptions, $\widehat{T}$ has spectral gap. We saw in the last lecture that spectral gaps survive small analytic perturbations. Therefore there is κ positive such that for every $|z| < \kappa$, $\widehat{T}_z = \lambda_z P_z + N_z$, where $P_z^2 = P_z$, $\dim \mathrm{Im}(P_z) = 1$, $N_z P_z = P_z N_z = 0$, and there exists θ s.t. $\rho[N_z] < \theta < |\lambda_z|$.

Since $\widehat{T}_0 = \widehat{T}$, $\lambda_0 = 1$ and $P_0 f = (\int f d\mu)1$ (exercises 2.3, 3.8) .

Expansion of the eigenvalue around zero: Let $\lambda_z', P_z', \widehat{T}_z'$ denote the derivatives of $\lambda_z, P_z, \widehat{T}_z$ at z.

We use exercise 4.3 to find λ_0' and λ_0''. Differentiate both sides of the equation $\widehat{T}_z P_z = \lambda_z P_z$: $\widehat{T}_z' P_z + \widehat{T}_z P_z' = \lambda_z' P_z + \lambda_z P_z'$. Multiply on the right by P_z. Since $P_z^2 = P_z$ and $P_z \widehat{T}_z = \lambda_z P_z$, $P_z \widehat{T}_z' P_z + \lambda_z P_z P_z' = \lambda_z' P_z + \lambda_z P_z P_z'$. Substituting $z = 0$, $\widehat{T}_0' = i\widehat{T} M_\psi$, and $P_0 f = \int f d\mu$, we obtain that

$$\lambda_0' = \int \psi d\mu = 0.$$

Next we claim that $\lambda''(0) = -\lim_{n\to\infty} \frac{1}{n} \int (\psi_n)^2 d\mu$. One differentiation of the identity $\widehat{T}_z^n P_z = \lambda_z^n P_z$ gives $(\widehat{T}_z^n)' P_z + \widehat{T}_z^n P_z' = (\lambda_z^n)' P_z + \lambda_z^n P_z'$. Another differentiation gives $(\widehat{T}_z^n)'' P_z + 2(\widehat{T}_z^n)' P_z' + \widehat{T}_z^n P_z'' = (\lambda_z^n)'' P_z + 2(\lambda_z^n)' P_z' + \lambda_z^n P_z''$. Multiplying on the right by P_z and substituting $z = 0$, we get

$$P_0 (\widehat{T}_0^n)'' P_0 + 2 P_0 (\widehat{T}_0^n)' P_0' = (\lambda_0^n)'' P_0 + 2 (\lambda_0^n)' P_0'.$$

Since $(\widehat{T}_0^n)' = i\widehat{T}^n M_{\psi_n}$, $(\widehat{T}_0^n)'' = -\widehat{T}^n M_{\psi_n}^2$, $(\lambda_0^n)' = n\lambda_0^{n-1}\lambda_0' = 0$, and $(\lambda_0^n)'' = n\lambda_0''$,

$$\lambda_0'' = -\frac{1}{n} \int (\psi_n)^2 d\mu + 2i \int \frac{1}{n} \psi_n P_0' 1 d\mu.$$

The second term tends to zero, because $\frac{1}{n}\psi_n \to 0$ a.e. by the ergodic theorem, and because $\|\frac{1}{n}\psi_n P_0'1\|_1 \leq \sup|\psi|\|P_0'1\|_1 \leq \sup|\psi|\|P_0'1\| < \infty$. It follows that $\lambda_0'' = -\lim_{n\to\infty}\frac{1}{n}\int(\psi_n)^2 d\mu$.

We obtain the following expansion of λ_t near zero:

$$\lambda_t = 1 - \frac{1}{2}\sigma^2 t^2 + O(t^3) \text{ as } t \to 0, \text{ where } \sigma = \sqrt{\lim_{n\to\infty}\frac{1}{n}\int(\psi_n)^2 d\mu} \geq 0.$$

Exercise 4.4 (Green–Kubo formula). $\sigma^2 = \int \psi^2 d\mu + 2\sum_{n=1}^{\infty}\int \psi(\psi\circ T^n)d\mu$.

The limit of the characteristic functions:

$$\begin{aligned}
\mathbb{E}(e^{i\frac{t}{\sqrt{n}}\psi_n}) &= \int e^{i\frac{t}{\sqrt{n}}\psi_n} d\mu = \int \widehat{T}^n_{\frac{t}{\sqrt{n}}} 1 d\mu = \int \left(\lambda^n_{\frac{t}{\sqrt{n}}} P_{\frac{t}{\sqrt{n}}} 1 + N^n_{\frac{t}{\sqrt{n}}} 1\right) d\mu \\
&= \lambda^n_{\frac{t}{\sqrt{n}}}\left(1 + \int (P_{\frac{t}{\sqrt{n}}} 1 - P_0) 1 d\mu + \lambda^{-n}_{\frac{t}{\sqrt{n}}} \int N^n_{\frac{t}{\sqrt{n}}} 1 d\mu\right) \\
&= \lambda^n_{\frac{t}{\sqrt{n}}}\left(1 + O(\|P_{\frac{t}{\sqrt{n}}} - P_0\|) + O(\lambda^{-n}_{\frac{t}{\sqrt{n}}}\|N^n_{\frac{t}{\sqrt{n}}}\|)\right) \quad (\because \|\cdot\| \geq \|\cdot\|_1) \\
&= \lambda^n_{\frac{t}{\sqrt{n}}}[1 + o(1)], \text{ because } z \mapsto P_z \text{ is continuous and } \rho(N_z) < |\lambda_z| \\
&= \left(1 - \tfrac{1}{2}\sigma^2(\tfrac{t}{\sqrt{n}})^2 + O((\tfrac{t}{\sqrt{n}})^3)\right)^n [1 + o(1)] \xrightarrow[n\to\infty]{} e^{-\frac{1}{2}\sigma^2 t^2}.
\end{aligned}$$

This proves that $\frac{1}{\sqrt{n}}\psi_n \xrightarrow[n\to\infty]{dist} N(0,\sigma^2)$. But we still need to show that $\sigma \neq 0$.

Positivity of σ: We assume by contradiction that $\sigma = 0$ and construct a solution $v \in \mathscr{L}$ to the equation $\psi = v - v\circ T$ (this contradicts our assumptions).

First we observe that $\psi = u - \widehat{T}u$ where $u := \psi + \sum_{n\geq 1}\widehat{T}^n\psi$ (the sum converges in norm, because $P_0\psi = \int \psi d\mu = 0$ so $\|\widehat{T}^n\psi\| = \|N_0^n\psi\| \xrightarrow[n\to\infty]{} 0$ exponentially fast). By the Green–Kubo formula,

$$\begin{aligned}
0 = \sigma^2 &= \int\left(\psi^2 + 2\psi\sum_{n=1}^{\infty}\psi\circ T^n\right)d\mu = \int\left((u-\widehat{T}u)^2 + 2(u-\widehat{T}u)\widehat{T}u\right)d\mu \\
&= \int\left((u-\widehat{T}u)(u-\widehat{T}u+2\widehat{T}u)\right)d\mu = \int\left((u-\widehat{T}u)(u+\widehat{T}u)\right)d\mu \\
&= \int\left(u^2 - (\widehat{T}u)^2\right)d\mu = \int\left(\widehat{T}(u^2) - (\widehat{T}u)^2\right)d\mu \quad (\because \forall g, \textstyle\int \widehat{T}g d\mu = \int g d\mu) \\
&= \int\left(\widehat{T}(u^2)\circ T - (\widehat{T}u\circ T)^2\right)d\mu \quad (\because \forall g, \textstyle\int g\circ T d\mu = \int g d\mu) \\
&= \int\left(\mathbb{E}(u^2|T^{-1}\mathscr{B}) - \mathbb{E}(u|T^{-1}\mathscr{B})^2\right)d\mu \quad (\because \forall g, (\widehat{T}g)\circ T = \mathbb{E}(g|T^{-1}\mathscr{B})).
\end{aligned}$$

Jensen's inequality (see appendix) says that $\mathbb{E}(u^2|T^{-1}\mathscr{B}) \geq \mathbb{E}(u|T^{-1}\mathscr{B})^2$ a.e. Necessarily $\mathbb{E}(u^2|T^{-1}\mathscr{B}) = \mathbb{E}(u|T^{-1}\mathscr{B})^2$. Equality in Jensen's inequality can only happen if $u = \mathbb{E}(u|T^{-1}\mathscr{B})$ a.e. (see appendix). So $u = \mathbb{E}(u|T^{-1}\mathscr{B}) = (\widehat{T}u) \circ T$ almost everywhere (proposition 1.2). Thus $\psi = u - \widehat{T}u = (\widehat{T}u) \circ T - (\widehat{T}u)$ almost everywhere, whence $\psi = v - v \circ T$ with $v := -\widehat{T}u$. □

Further reading

1. *B.V. Gnedenko and A.N. Kolmogorov:* Limit distributions for sums of independent random variables, *Addison–Wesley, ix+264pp (1954).*
2. Y. Guivarc'h, J. Hardy: *Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov.* Ann. Inst. H. Poincaré Probab. Statist. **24** (1988), no. 1, 73–98.
3. *W. Parry and M. Pollicott:* Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* **187-188***, 268pp (1990)*

Lecture 5
Absence of spectral gap

5.1 Absence of spectral gap

Obstructions to spectral gap: Spectral gap implies exponential decay of correlations. Therefore, if $f \in L^1, g \in L^\infty$ and $\mathrm{Cov}(f, g \circ T^n) \xrightarrow[n\to\infty]{} 0$ sub-exponentially, then there is no Banach space $\mathscr{L}$ which contains f s.t. $\widehat{T} : \mathscr{L} \to \mathscr{L}$ has spectral gap.

Example (The Manneville–Pomeau map): $T : [0,1] \to [0,1]$, $T(x) = x(1 + x^{1+s})$ mod 1, $0 < s < \frac{1}{2}$. Here the correlations decay at a rate $\frac{1}{n^{\frac{1}{\alpha}-1}}$ whenever f is Lipschitz, $g \in L^\infty$, and f, g are supported inside [discontinuity, 1] and have non-zero integrals with respect to the absolutely continuous invariant probability measure.

Other obstructions: breakdown of the CLT, non-integrable invariant density, and (for those who understand what this means) a phase transition.

5.2 Inducing

The induced system: Suppose $(X, \mathscr{B}, \mu, T)$ is a probability preserving map, and $A \subset X$ is a measurable subset of positive measure. By Poincaré's Recurrence Theorem, for a.e. $x \in A$ there are infinitely many $n \geq 1$ s.t. $T^n(x) \in A$.

Let $A' := \{x \in A : T^n(x) \in A$ for infinitely many $n\}$, and define

1. **First return time:** $\varphi_A : A' \to \mathbb{N}$, $\varphi_A(x) := 1_A(x) \min\{n \geq 1 : T^n(x) \in A\}$
2. **Induced map (on A):** $T_A : A' \to A'$, $T_A(x) = T^{\varphi_A(x)}(x)$

Exercise 5.1 (Transfer operator of T_A). Show that $\widehat{T_A} f = \sum_{n\geq 1} \widehat{T}^n(f 1_{[\varphi_A = n]})$

Sometimes it is possible to choose A in such a way that $\widehat{T_A}$ has spectral gap on a large Banach space, eventhough $\widehat{T}$ does not.

Example: Induce the Manneville–Pomeau map on $A = [\text{discontinuity}, 1]$. Unlike T, T_A is piecewise uniformly expanding:

$$(T_A)'(x) = T'(x) \cdot \underbrace{[T'(Tx)T'(T^2x)\cdots T'(T^{\varphi_A(x)-1}x)]}_{\geq 1} \geq \min_{[\text{discontinuity},1]} T' > 1.$$

In fact T_A is a piecewise onto, uniformly expanding, interval map on A.

One can show, exactly as in the case of the Gauss map, that $\widehat{T}_A$ has spectral gap on $\mathscr{L} := \{\text{Lipschitz functions on } A\}$.

The question is how to use the spectral gap of $\widehat{T}_A$ to obtain information on the asymptotic behavior of $\widehat{T}^n$ as $n \to \infty$. This is purpose of "operator renewal theory."

5.3 Operator renewal theory

The basic construction: Define operators $T_n, R_n : L^1(A) \to L^1(A)$ by

1. $T_0 = I$, $T_n f = 1_A \cdot \widehat{T}^n(f 1_A)$
2. $R_0 = 0$, $R_n f = 1_A \cdot T^n(f 1_{[\varphi_A = n]})$

These operators satisfy a non-commutative version of the "renewal equation" from probability theory:

The renewal equation: $T_n = T_0 R_n + T_1 R_{n-1} + \cdots + T_{n-1} R_1$ and $T_n = R_n T_0 + R_{n-1} T_1 + \cdots + R_1 T_{n-1}$.

Proof. For every $u \in L^\infty(A)$,

$$\begin{aligned}
\int_A u T_n f d\mu &= \int (1_A u) \circ T^n \cdot 1_A f d\mu = \int (1_A u) \circ T^n \cdot \big(\sum_{k=1}^{\infty} 1_{[\varphi_A = k]} f\big) d\mu \\
&= \sum_{k=1}^{n} \int \big(1_A u\big) \circ T^n \cdot 1_{[\varphi_A = k]} f d\mu \quad (\because (1_A u) \circ T^n = 0 \text{ on } [\varphi_A > n]) \\
&= \sum_{k=1}^{n} \int \big(1_A u\big) \circ T^{n-k} \circ T^k \cdot (1_{[\varphi_A = k]} f) d\mu \\
&= \sum_{k=1}^{n} \int \big(1_A u\big) \circ T^{n-k} \cdot \widehat{T}^k (1_{[\varphi_A = k]} f) d\mu \\
&= \sum_{k=1}^{n} \int \big(1_A u\big) \circ T^{n-k} \cdot 1_A \widehat{T}^k (1_{[\varphi_A = k]} f) d\mu \quad (\because \operatorname{supp}[\widehat{T}^k 1_{[\varphi_A = k]}] \subset A) \\
&= \sum_{k=1}^{n} \int (1_A u) \circ T^{n-k} R_k f d\mu = \sum_{k=1}^{n} \int (1_A u) \widehat{T}^{n-k} [R_k f] d\mu \\
&= \int u \left(\sum_{k=1}^{n} (T_{n-k} R_k) f \right) d\mu.
\end{aligned}$$

Exercise 5.2. Prove the other inequality, using the following decomposition:

$$
\begin{aligned}
(1_A u)\circ T^n \cdot 1_A &= (1_{\{x\in A: T^n(x)\in A\}} u)\circ T^n \cdot 1_A \\
&= \left(\sum_{k=0}^{n-1} (1_{\{\text{last visit to } A \text{ before time } n \text{ is at time } k\}} u)\circ T^n\right)\cdot 1_A
\end{aligned}
$$

Generating functions: Let $T(z) := I + \sum_{n\geq 1} z^n T_n$ and $R(z) = \sum_{n\geq 1} z^n R_n$.

Notice that $R(1) = \sum R_n = \widehat{T}_A$. Since $\|T_n\|, \|R_n\| \leq 1$ as operators on L^1, these power series converge on $\{z : |z| \leq 1\}$ and are analytic on $\{z : |z| < 1\}$. The following exercise gives the generating function form of the renewal equation.

Exercise 5.3. $T(z) = (I - R(z))^{-1}$ for all $|z| < 1$.

The idea: $T(z)f$ is a generating function of $1_A \widehat{T}^n(f 1_A)$, therefore it contains information on the asymptotic behavior of $\widehat{T}^n$. $R(z)$ is a perturbation of $R(1) = \widehat{T}_A$. This suggests the following strategy:

1. Find a set A s.t. $\widehat{T}_A$ has spectral gap on some space
2. Use the spectral gap of $R(1)$ and perturbation theory to analyze $R(z)$ for $z \approx 1$
3. Use the renewal equation $T(z) = (I - R(z))^{-1}$ to deduce information on $T(z)$

The last two steps are handled by the following abstract theorem.

Theorem (Gouëzel, Sarig). *Suppose T_n are bounded linear operators on a Banach space $\mathscr{L}$ s.t. $T(z) = I + \sum_{n\geq 1} z^n T_n$ converges in the operator norm on the open unit disk. Assume further that*

1. **Renewal equation:** *$T(z) = (I - R(z))^{-1}$ on $\{z : |z| < 1\}$, where $R(z) = \sum_{n\geq 1} z^n R_n$ and $\sum \|R_n\| < \infty$.*
2. **Spectral gap:** *$R(1) = P + N$ where $P^2 = P$, $\dim \mathrm{Im}(P) = 1$, $PN = NP = 0$ and $\rho(N) < 1$.*
3. **Aperiodicity:** *$I - R(z)$ is invertible for every $z \neq 1$ s.t. $|z| \leq 1$.*

If $\sum_{k>n} \|R_k\| = O(n^{-\beta})$ for some $\beta > 1$ and $PR'(1)P \neq 0$, then there are bounded linear operators $\varepsilon_n : \mathscr{L} \to \mathscr{L}$ s.t. $\|\varepsilon_n\| = o(n^{-(\beta-1)})$ and

$$
T_n = \frac{1}{a} P + \frac{1}{a^2} \sum_{k=n+1}^{\infty} P_k + \varepsilon_n
$$

where a is given by $PR'(1)P = aP$, and $P_n = \sum_{\ell>n} PR_\ell P$.

Let's calculate a, P, P_k in the dynamical context. Suppose T is a mixing probability preserving transformation whose transfer operator $\widehat{T}$ satisfies the conditions of the theorem with some Banach space $\mathscr{L}$ such that $\mathscr{L} \subset L^1(A)$ and $\|\cdot\|_{\mathscr{L}} \geq \|\cdot\|_1$. In this case $T_n f = 1_A \widehat{T}^n(f 1_A)$, $R_n f = 1_A \widehat{T}^n(f 1_{[\varphi_A = n]})$, and $R(1)f = \widehat{T}_A f$ $(f \in \mathscr{L})$.

- $Pf = \left(\frac{1}{\mu(A)} \int_A f d\mu\right) 1_A$: Since T is ergodic, T_A is ergodic (exercise), therefore $\frac{1}{N}\sum_1^N \widehat{T}_A^n f \xrightarrow{w} \frac{1}{\mu(A)} \int_A f d\mu$. By the spectral gap assumption, $\frac{1}{N}\sum_1^N \widehat{T}_A^n f = \frac{1}{N}\sum_1^N (P+N)^n f \xrightarrow{\mathscr{L}} Pf$. Necessarily $Pf = \frac{1}{\mu(A)} \int_A f d\mu$.

- $a = \frac{1}{\mu(A)}$: This is because $PR'(1)Pf = \frac{1}{\mu(A)}\int_A[\sum n\widehat{T}^n(Pf1_{[\varphi=n]})]d\mu = \frac{1}{\mu(A)}\sum n\mu([\varphi=n]\cap T^{-n}A)Pf = \frac{1}{\mu(A)}\sum n\mu[\varphi=n]Pf = \frac{1}{\mu(A)}Pf$, because $\sum n\mu[\varphi=n]=1$ by Kac formula. So $PR'(1)P = \frac{1}{\mu(A)}P$
- $P_n f = \frac{1}{\mu(A)}\mu[\varphi>n]Pf$: direct calculation as above.

Exercise 5.4. Use this to show that for the Manneville–Pomeau map equipped with its acip μ, for every f,g bounded Lipschitz supported inside $A := [\text{discontinuity}, 1]$ s.t. $\int f d\mu, \int g d\mu \neq 0$,

$$\mathrm{Cov}(f, g\circ T^n) = [1+o(1)](\sum_{k=n+1}^{\infty}\mu[\varphi_A > k])\int f\int g.$$

The estimate we mentioned at the beginning for the polynomial rate of decay of correlations for this map is obtained by further analysis of $\mu[\varphi_A > k]$ as $k\to\infty$.

Further reading

1. S. Gouëzel: *Sharp polynomial estimates for the decay of correlations.* Israel J. Math. **139** (2004), 29–65.
2. O. Sarig: *Subexponential decay of correlations*, Invent. Math. **150** (2002), 629–653.

Appendix A
Supplementary material

A.1 Conditional expectations and Jensen's inequality

σ–algebras and information Recall that a σ*-algebra* on a set X is a collection $\mathscr{B}$ of subsets of X which contains $\varnothing$ and X; is closed under complements ($A \in \mathscr{F} \Rightarrow A^c := X \setminus A \in \mathscr{B}$); and is closed under *countable* unions and intersections: $\{A_n : n \in \mathbb{N}\} \subset \mathscr{B} \Rightarrow \bigcup_{n\geq 1} A_n, \bigcap_{n\geq 1} A_n \in \mathscr{B}$.

A *sub-*σ*-algebra* of $(X,\mathscr{B})$ is a σ-algebra $\mathscr{F}$ on X such that $\mathscr{F} \subseteq \mathscr{B}$.

To understand the heuristic foundations for the definition of the conditional expectation given $\mathscr{F}$, it is useful to think of $\mathscr{F}$ as of a representation of the "partial information" on an unknown point $x \in X$ contained in the answers to all yes/no questions of the form "is $x \in F$?" with $F \in \mathscr{F}$.

Examples: Suppose $X = \mathbb{R}$ and $\mathscr{B}$ is the Borel σ-algebra.

1. Suppose $A \in \mathscr{B}$ is a set and all we know is whether $x \in A$ or not. This partial information is represented by $\mathscr{F} = \{\varnothing, \mathbb{R}, A, A^c\}$
2. Suppose $A, B \in \mathscr{B}$ are sets and all we know is whether $x \in A, B$ or not. This partial information is represented by $\mathscr{F}$:=smallest σ-algebra containing $\{A, B\}$. This is the collection of all sets which can be written as a union of the elements of the partition generated by A, B, namely $\{\varnothing, A\cap B, A\setminus B, B\setminus A, (A\cup B)^c\}$.
3. Suppose we know $|x|$ but not x. This partial information is represented by $\mathscr{F} := \{E \in \mathscr{B} : E = -E\}$
4. Suppose we know nothing on x. The corresponding σ-algebra is $\{\varnothing, \mathbb{R}\}$

A function $f : X \to \mathbb{R}$ is called $\mathscr{F}$*–measurable*, if for every $t \in \mathbb{R}$, $[f < t] := \{x \in X : f(x) < t\}$ belongs to $\mathscr{F}$. Notice that if $f : X \to \mathbb{R}$ is $\mathscr{F}$–measurable, then there are countably many $F_n \in \mathscr{F}$ so that $f(x)$ can be calculated for each x from the answers to the questions "is $x \in F$?". To see this take an enumeration of the rationals $\{t_n\}$, let $F_n := [f < t_n]$, and observe that

$$f(x) := \inf\{t \in \mathbb{Q} : x \in [f < t]\} = \inf\{t_n : x \in F_n\}.$$

The "best estimate" given partial information: Suppose g is *not* $\mathscr{F}$–measurable. What is the "best estimate" for $g(x)$ given the information $\mathscr{F}$?

When g is in L^2, the "closest" $\mathscr{F}$–measurable function (in the L^2–sense) is the projection of g on $L^2(X,\mathscr{F},\mu)$. The defining property of the projection Pg of g is $\langle Pg,h\rangle = \langle g,h\rangle$ for all $h \in L^2(X,\mathscr{F},\mu)$.

In practice, one often needs to work with the larger space L^1. There is only one way to continuously extends the definition from L^2 to L^1 and it is the following:

Definition: The *conditional expectation* of $f \in L^1(X,\mathscr{B},\mu)$ given $\mathscr{F}$ is the unique $L^1(X,\mathscr{F},\mu)$–element $\mathbb{E}_\mu(f|\mathscr{F})$ which satisfies

1. $\mathbb{E}_\mu(f|\mathscr{F})$ is $\mathscr{F}$–measurable;
2. $\forall\varphi \in L^\infty$ $\mathscr{F}$–measurable, $\int \varphi\mathbb{E}_\mu(f|\mathscr{F})d\mu = \int \varphi f d\mu$.

Note: L^1–elements are equivalence classes of functions, not functions. Any *function* which defines the same L^1–element as $\mathbb{E}_\mu(f|\mathscr{F})$ is called a *version* of $\mathbb{E}_\mu(f|\mathscr{F})$. There are many possible versions (all equal a.e.).

Proposition A.1. *The conditional expectation exists for every L^1 element, and is unique up sets of measure zero.*

Proof. Consider the measures $\nu_f := fd\mu|_{\mathscr{F}}$ and $\mu|_{\mathscr{F}}$ on $(X,\mathscr{F})$. Then $\nu_f \ll \mu$. The function $\mathbb{E}_\mu(f|\mathscr{F}) := \frac{d\nu_f}{d\mu}$ (Radon-Nikodym derivative) is $\mathscr{F}$–measurable, and it is easy to check that it satisfies the conditions of the definition of the conditional expectation. The uniqueness of the conditional expectation is left as an exercise. □

Proposition A.2 (Basic properties).

1. $f :\mapsto \mathbb{E}_\mu(f|\mathscr{F})$ is linear, bounded, and has norm one as an operator on L^1;
2. $f \geq 0 \Rightarrow \mathbb{E}_\mu(f|\mathscr{F}) \geq 0$ a.e.;
3. if h is $\mathscr{F}$–measurable, then $\mathbb{E}_\mu(hf|\mathscr{F}) = h\mathbb{E}_\mu(f|\mathscr{F})$;
4. If $\mathscr{F}_1 \subset \mathscr{F}_2$, then $\mathbb{E}_\mu[\mathbb{E}_\mu(f|\mathscr{F}_1)|\mathscr{F}_2] = \mathbb{E}_\mu(f|\mathscr{F}_2)$.

The proof is left as an exercise.

Proposition A.3 (Jensen's inequality). *Suppose $\varphi : \mathbb{R} \to \mathbb{R}$ is twice differentiable with strictly positive second derivative, then for every $f \in L^\infty$,*

$$\mathbb{E}_\mu(\varphi \circ f|\mathscr{F}) \geq \varphi[\mathbb{E}_\mu(f|\mathscr{F})] \text{ a.e.,}$$

and $\mathbb{E}_\mu(\varphi \circ f|\mathscr{F}) = \varphi[\mathbb{E}_\mu(f|\mathscr{F})]$ a.e. iff $f = g$ a.e. where g is $\mathscr{F}$–measurable.

Proof. The content of the assumptions on φ are that φ is strictly convex. In particular, φ lies strictly above its tangent lines:

$$\varphi(t) > \varphi'(x)(t-x) + \varphi(x) \text{ for all } x \in X, t \neq 0.$$

Fix once and for all an $\mathscr{F}$–measurable version of $\mathbb{E}_\mu(f|\mathscr{F})$. Given x, let $m(x) = \varphi'[\mathbb{E}_\mu(f|\mathscr{F})(x)]$. This is a bounded $\mathscr{F}$–measurable function, and

$$\varphi(t) > m(x)(t - \mathbb{E}_\mu(f|\mathscr{F})(x)) + \varphi[\mathbb{E}_\mu(f|\mathscr{F})(x)] \text{ for all } t \neq \mathbb{E}_\mu(f|\mathscr{F})(x).$$

In particular

$$\varphi[f(x)] \geq m(x)(f(x) - \mathbb{E}_\mu(f|\mathscr{F})(x)) + \varphi[\mathbb{E}_\mu(f|\mathscr{F})(x)] \text{ for all } x, \tag{A.1}$$

with equality only at the x where $f(x) = \mathbb{E}_\mu(f|\mathscr{F})(x)$.

Taking conditional expectations on both sides, and recalling that $\mathbb{E}_\mu(\cdot|\mathscr{F})$ is a positive operator, we see that

$$\begin{aligned}
&\mathbb{E}_\mu(\varphi \circ f|\mathscr{F}) \geq \mathbb{E}_\mu(m(f - \mathbb{E}_\mu(f|\mathscr{F}))|\mathscr{F}) + \varphi[\mathbb{E}_\mu(f|\mathscr{F})] \\
&= m\mathbb{E}_\mu(f - \mathbb{E}_\mu(f|\mathscr{F})|\mathscr{F}) + \varphi[\mathbb{E}_\mu(f|\mathscr{F})] \quad (\because\ m \text{ is bounded, } \mathscr{F}\text{–measurable}) \\
&= m\left(\mathbb{E}_\mu(f|\mathscr{F}) - \mathbb{E}_\mu(\mathbb{E}_\mu(f|\mathscr{F})|\mathscr{F})\right) + \varphi[\mathbb{E}_\mu(f|\mathscr{F})] \text{ (prop A.2 part 1)} \\
&= m\left(\mathbb{E}_\mu(f|\mathscr{F}) - \mathbb{E}_\mu(f|\mathscr{F})\right) + \varphi[\mathbb{E}_\mu(f|\mathscr{F})] \text{(prop A. 2 part 4)} \\
&= \varphi[\mathbb{E}_\mu(f|\mathscr{F})(x)].
\end{aligned}$$

So $\mathbb{E}_\mu(\varphi \circ f|\mathscr{F}) \geq \varphi[\mathbb{E}_\mu(f|\mathscr{F})]$ almost everywhere.

The chain of inequalities also shows that $\mathbb{E}_\mu(\varphi \circ f|\mathscr{F}) = \varphi[\mathbb{E}_\mu(f|\mathscr{F})]$ iff there is equality a.e. in (A.1), which happens exactly when $f(x) = \mathbb{E}_\mu(f|\mathscr{F})(x)$. So $\mathbb{E}_\mu(\varphi \circ f|\mathscr{F}) = \varphi[\mathbb{E}_\mu(f|\mathscr{F})]$ a.e. iff $f = \mathbb{E}_\mu(f|\mathscr{F})$ a.e., and this is the same as saying that f has an $\mathscr{F}$–measurable version. □

A.2 Mixing and exactness for the Gauss map

Mixing: A probability preserving map $(X,\mathscr{B},\mu,T)$ is called *mixing*, if for every $A,B\in\mathscr{B}$, $\mu(A\cap T^{-n}B)\xrightarrow[n\to\infty]{}\mu(A)\mu(B)$.

Exactness: A (non-invertible) non-singular map $(X,\mathscr{B},\mu,T)$ is called *exact*, if for every $E\in\bigcap_{n=0}^{\infty}T^{-n}\mathscr{B}$, either $\mu(E)=0$ or $\mu(X\setminus E)=0$.

Proposition: *An exact probability preserving map is mixing.*

Proof. Suppose $(X,\mathscr{B},\mu,T)$ is exact. Since T is measurable, $T^{-n}\mathscr{B}$ is a decreasing sequence of σ–algebras. By the Martingale convergence theorem, $\mathbb{E}(A|T^{-n}\mathscr{B})\xrightarrow[n\to\infty]{L^1}$ $\mathbb{E}\big(A|\bigcap_{n=0}^{\infty}T^{-n}\mathscr{B}\big)=\mathbb{E}(A|\{\varnothing,X\})=\mu(A)$ for all $A\in\mathscr{B}$. So for all $A,B\in\mathscr{B}$,

$$\mu(A\cap T^{-n}B)=\int 1_A(1_B\circ T^n)d\mu=\int\mathbb{E}(1_A|T^{-n}\mathscr{B})1_B\circ T^n d\mu$$
$$=\int\mu(A)1_B\circ T^n d\mu+O\left(\int\big|\mathbb{E}(1_A|T^{-n}\mathscr{B})-\mu(A)\big|\,d\mu\right)\to\mu(A)\mu(B).\qquad\square$$

Theorem (Rényi). *The Gauss map $T(x)=\{\frac{1}{x}\}$ is exact with respect to its absolutely continuous invariant probability measure.*

Proof. It is enough to show that T is exact with respect to Lebesgue's measure m. Let $v_a:[0,1]\to[0,1]$ denote the inverse branches $v_a(x)=\frac{1}{a+x}$ $(a\in\mathbb{N})$, and set for every $\underline{a}=(a_1,\ldots,a_n)$, $v_{\underline{a}}=v_{a_1}\circ\cdots\circ v_{a_n}$. Let $[\underline{a}]:=v_{\underline{a}}([0,1))$. This is the set of all numbers whose continued fraction expansion starts with $\underline{a}$.

Rényi's inequality: $\exists C>1$ s.t. $C^{-1}m[\underline{a}]m[\underline{b}]\le m[\underline{a},\underline{b}]\le Cm[\underline{a}]m[\underline{b}]$ for all $\underline{a},\underline{b}$

Proof: $$m[\underline{a},\underline{b}]=\int 1_{[\underline{a}]}1_{[\underline{b}]}\circ T^{|\underline{a}|}dm=\int_{[\underline{b}]}\widehat{T}^{|\underline{a}|}1_{[\underline{a}]}dm$$
$$=\int_{[\underline{b}]}v'_{\underline{a}}dm\quad(\because\ \underline{a}\neq\underline{b}\Rightarrow 1_{[\underline{a}]}\circ v_{\underline{b}}=0\text{ by exercise 2.5 e})$$
$$=\int_{[\underline{b}]}G^{\pm1}m[\underline{a}]dm=G^{\pm1}m[\underline{a}]m[\underline{b}]\ \ (\text{exercise 2.5 d. })$$

(Here $a=G^{\pm1}b$ means $G^{-1}\le a/b\le G$.)

Standard approximation arguments show that for every $\underline{a}$ and $B\in\mathscr{B}$,

$$C^{-1}m[\underline{a}]m(B)\le m([\underline{a}]\cap T^{-|\underline{a}|}(B))\le Cm[\underline{a}]m(B)$$

We can now show exactness. Suppose $B\in\bigcap_{n\ge0}T^{-n}\mathscr{B}$ and $m(B)\neq0$. For every n, there is $B_n\in\mathscr{B}$ s.t. $B=T^{-n}B_n$, therefore for every $\underline{a}$ with $|\underline{a}|=n$,

$$m(B\cap[\underline{a}])=m(T^{-n}B_n\cap[\underline{a}])\ge C^{-1}m(B_n)m[\underline{a}].$$

Notice that $\frac{1}{2\ln 2} \leq \frac{d\mu}{dm} \leq 2\ln 2$ where $d\mu = \frac{1}{\ln 2}\frac{1}{1+x}dx$ is the absolutely continuous invariant measure of the Gauss map. So $m(B_n) \geq \frac{1}{2\ln 2}\mu(B_n) = \frac{1}{2\ln 2}\mu(B) \geq \frac{1}{4\ln^2 2}\mu(B)$. We see that

$$\frac{m(B\cap[\underline{a}])}{m[\underline{a}]} \geq \frac{m(B)}{4C\ln^2 2} \text{ for all } \underline{a}.$$

Let $\mathscr{F}_n :=$ σ–algebra generated by $\{[\underline{a}] : |\underline{a}| = n\}$, then $\mathbb{E}_m(1_B|\mathscr{F}_n) = \sum_{|\underline{a}|=n} \frac{m(B\cap[\underline{a}])}{m[\underline{a}]} 1_{[\underline{a}]}$ (exercise). Therefore $\mathbb{E}_m(1_B|\mathscr{F}_n) \geq \frac{m(B)}{4C\ln^2 2}$. But $\mathscr{F}_n \uparrow \mathscr{B}$ so by the Martingale convergence theorem $\lim_{n\to\infty} \mathbb{E}_m(1_B|\mathscr{F}_n) = \mathbb{E}_m(1_B|\mathscr{B}) = 1_B$. So $1_B > 0$ a.e., whence $m(X \setminus B) = 0$. □

Reference: J. Aaronson, M. Denker, and M. Urbanski: *Ergodic theory for Markov fibred systems and parabolic rational maps.* Trans. Amer. Math. Soc. **337** (1993), no. 2, 495–548.

A.3 Hennion's theorem on quasi-compactness

Theorem (Doeblin & Fortet, Ionescu-Tulcea & Marinescu, Hennion). *Suppose* $(B, \|\cdot\|)$ *is a Banach space and* $L : B \to B$ *is a bounded linear operator with spectral radius* $\rho(L)$ *for which there exists semi–norm* $\|\cdot\|'$ *s.t.:*

1. $\|\cdot\|'$ *is continuous on B;*
2. *there exists* $M > 0$ *s.t.* $\|Lf\|' \leq M\|f\|'$ *for all* $f \in B$*;*
3. *for any sequence of* $f_n \in B$*, if* $\sup\|f_n\| < \infty$ *then there exists a subsequence* $\{n_k\}_{k\geq 1}$ *and some* $g \in B$ *s.t.* $\|Lf_{n_k} - g\|' \xrightarrow[k\to\infty]{} 0$*;*
4. *there are* $k \geq 1$*,* $0 < r < \rho(L)$*, and* $R > 0$ *s.t.*

$$\|L^k f\| \leq r^k\|f\| + R\|f\|'. \tag{A.2}$$

Then L is quasi-compact.

Proof. We first give the proof in the special case $k = 1$.

Fix $r < \rho \leq \rho(L)$ arbitrarily close to r, and let

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

The plan of the proof is to show that for all $z \in A(\rho, \rho(L))$,

- $K(z) := \bigcup_{\ell\geq 0} \ker(zI - L)^\ell$ is finite dimensional, and $I(z) := \bigcap_{\ell\geq 0} \mathrm{Im}(zI - L)^\ell$ is closed;
- $K(z), I(z)$ are L–invariant and $B = K(z) \oplus I(z)$;
- $(zI - L) : I(z) \to I(z)$ is a bijection with bounded inverse;
- the set of $\lambda \in A(\rho, \rho(L))$ s.t. $K(\lambda) \neq \{0\}$ is finite and non–empty.

This implies that the intersection of the spectrum of L with the annulus $A(\rho, \rho(L))$ is a finite set of eigenvalues with finite multiplicity, because if z is not an eigenvalue, then $K(z) = 0$, whence $B = I(z)$, whence $(zI - L) : B \to B$ is a bijection with a bounded inverse, and z is outside the spectrum of L.

Once we have this spectral information, we let $\{\lambda_1, \ldots, \lambda_t\}$ denote the eigenvalues of L in $A(\rho, \rho(L))$ and form

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

By the properties of $K(z), I(z)$ mentioned above, F, H are L–invariant, F is finite dimensional, and H is closed. We will show, using standard linear algebra techniques, that $B = F \oplus H$, that the eigenvalues of $L|_F$ are $\lambda_1, \ldots, \lambda_t$, and that the spectral radius of $L|_H$ is less than ρ.

The double norm inequality (A.2) and the semi–norm $\|\cdot\|'$ are used in the following statement, which is the main technical tool:

Conditional Closure Lemma: *Fix $|z| > r$ and let $\{g_n\}_{n\geq 1}$ be a sequence in B s.t. $g_n = (zI-L)f_n$ has a solution $f_n \in \mathscr{L}$ for all n. If $g_n \xrightarrow[n\to\infty]{B} g$ and $\sup\|f_n\| < \infty$, then $\{f_n\}_{n\geq 1}$ has a subsequence which converges in B to a solution f of $g = (zI-L)f$.*

Proof. Starting from the equation $(g_n - g_m) = (zI-L)(f_n - f_m)$, we see that

$$|z|\|f_n - f_m\| = \|(g_n - g_m) + L(f_n - f_m)\| \leq \|g_n - g_m\| + r\|f_n - f_m\| + R\|f_n - f_m\|'.$$

Rearranging terms, we obtain

$$\|f_n - f_m\| \leq \frac{1}{|z|-r}\left[\|g_n - g_m\| + \|f_n - f_m\|'\right]. \tag{A.3}$$

1. $\|g_n - g_m\|$ tends to zero as $n,m\to\infty$, because $g_n \xrightarrow[n\to\infty]{B} g$.
2. To deal with $\|f_n - f_m\|'$ we start again from $g_n = (zI-L)f_n$ and deduce

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

   Since $\sup\|f_n\| < \infty$, there is a subsequence $\{Lf_{n_k}\}_{k\geq 1}$ s.t. $\|Lf_{n_k} - h\|' \to 0$ for some $h \in B$. Since $\|\cdot\|'$ is continuous, $\|g_{n_k} - g\|' \to 0$. Thus $\|f_{n_k} - f_{m_k}\|' \leq \frac{1}{|z|}\left(\|g_{n_k} - g_{m_k}\|' + \|Lf_{n_k} - Lf_{m_k}\|'\right) \xrightarrow[k,\ell\to\infty]{} 0$.

Returning to (A.3), we see that $\|f_{n_k} - f_{n_\ell}\| \xrightarrow[k,\ell\to\infty]{} 0$, so $\exists f \in B$ s.t. $f_{n_k} \xrightarrow[k\to\infty]{B} f$. Since $zI - L$ is continuous, $g = (zI-L)f$, and we are done.

Riesz Lemma: *Let $(V,\|\cdot\|)$ be a normed vector space, and suppose $U \subseteq V$ is a subspace. If $\overline{U} \neq V$, then for every $0 < t < 1$ there exists $v \in V$ s.t. $\|v\| = 1$ and* $\mathrm{dist}(v,U) \geq t$.

If V were a Hilbert space, then any unit vector in $U^{\perp}$ would work with $t = 1$. The point of Riesz's Lemma is that it holds in general normed vector spaces.

Proof of Riesz's Lemma. Fix $v_0 \in V\setminus\overline{U}$, and construct $u_0 \in U$ s.t. $\mathrm{dist}(v_0,U) \leq \|v_0 - u_0\| \leq \frac{1}{t}\mathrm{dist}(v_0,U)$. Calculating, we see that for every $u \in U$,

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

Since this holds for all $u \in U$, $v := (v_0 - u_0)/\|v_0 - u_0\|$ is as required.

We are now ready for the proof of Hennion's Theorem. Define

$$K(z) := \bigcup_{\ell>0}\ker(zI-L)^{\ell}\ ,\ I(z) := \bigcap_{\ell>0}\mathrm{Im}(zI-L)^{\ell}.$$

Step 1. *Let $K_\ell(z) := \ker(zI-L)^{\ell}$, $I_\ell(z) := \mathrm{Im}(zI-L)^{\ell}$ and suppose $|z| > r$, then*

1. $K_\ell(z)$ is finite dimensional for all ℓ;

2. *$I_\ell(z)$ is closed for all ℓ;*
3. *there exists ℓ s.t. $K_\ell(z) = K(z)$ and $I_\ell(z) = I(z)$.*

Proof. Fix z s.t. $|z| > r$, and let $K_\ell := K_\ell(z), I_\ell := I_\ell(z)$.

We show by induction that $\dim K_\ell < \infty$ for all ℓ. Suppose by way of contradiction that $\dim K_1 = \infty$. Using the Riesz Lemma with $t = 1/2$, it is not difficult to construct $f_n \in \ker(zI - L)$ s.t. $\|f_n\| = 1$ and $\|f_n - f_m\| \geq 1/2$ for all $n \neq m$. We have for all n, $\sup \|f_n\| < \infty$ and $(zI - L)f_n = 0$, so by the conditional closure lemma $\{f_n\}_{n \geq 1}$ contains a convergent sequence. But this cannot be the case, so we get a contradiction which proves that $\dim K_1 < \infty$.

Next we assume by induction that $\dim K_\ell < \infty$, and show that $\dim K_{\ell+1} < \infty$. Assume by contradiction that $\dim K_{\ell+1} = \infty$, then $\exists f_n \in \ker(zI - L)^{\ell+1}$ s.t. $\|f_n\| = 1$ and $\|f_n - f_m\| \geq 1/2$ for $n \neq m$. By construction $g_n := (zI - L)f_n \in K_\ell$, and $\|g_n\| \leq |z| + \|L\|$. The unit ball in K_ℓ is compact, because $\dim K_\ell < \infty$ by the induction hypothesis, so $\exists n_k \uparrow \infty$ s.t. g_{n_k} converges in norm. By the conditional closure lemma, $\exists n_{k_\ell}$ s.t. $\{f_{n_{k_\ell}}\}$ converges in norm. But this cannot be the case because $\|f_n - f_m\| \geq 1/2$ when $n \neq m$. So $\dim K_{\ell+1}$ must be finite. This concludes the proof that K_ℓ has finite dimension for all ℓ.

Next we show that $I_\ell := \mathrm{Im}(zI - L)^\ell$ is closed for all ℓ. Again we use induction on ℓ, except that this time we start the induction at $\ell = 0$, with the understanding that $(zI - L)^0 = I$, whence $I_0 = \mathrm{Im}(I) = B$. This space, of course, is closed.

We now assume by induction that I_ℓ is closed, and show that $I_{\ell+1} \equiv (zI - L)I_\ell$ is closed. We must show that for every sequence of functions $g_n \in (zI - L)I_\ell$, if $g_n \to g$, then $g \in (zI - L)I_\ell$. Write

$$g_n = (zI - L)f_n, \quad f_n \in I_\ell.$$

We are free to modify f_n by subtracting arbitrary elements of $K_1 \cap I_\ell$. For example, we may subtract the closest element to f_n in $K_1 \cap I_\ell$ (the closest element exists since $\dim K_1 < \infty$ and I_ℓ is closed). Thus we may assume without loss of generality that

$$\|f_n\| = \mathrm{dist}\,(f_n, K_1 \cap I_\ell).$$

We claim that $\sup \|f_n\| < \infty$. Otherwise, $\exists n_k \uparrow \infty$ s.t. $\|f_{n_k}\| \to \infty$, and then $g_{n_k}/\|f_{n_k}\| \to 0$ (because $g_{n_k} \to g$). But

$$\frac{g_{n_k}}{\|f_{n_k}\|} = (zI - L)\frac{f_{n_k}}{\|f_{n_k}\|}$$

so $\exists n_{k_\ell} \uparrow \infty$ s.t. $f_{n_{k_\ell}}/\|f_{n_{k_\ell}}\| \to h$ where $(zI - L)h = 0$ (conditional closure lemma). Since $f_n \in I_\ell$ and I_ℓ is closed, $h \in I_\ell$. Thus $f_{n_{k_\ell}}/\|f_{n_{k_\ell}}\| \to h \in K_1 \cap I_\ell$. But this is impossible, since we have constructed f_n so that $\mathrm{dist}\,(f_n/\|f_n\|, K_1) = 1$ for all n. This contradiction shows that

$$\sup \|f_n\| < \infty.$$

Since $\sup \|f_n\| < \infty$, $g_n \to g$, and $g_n = (zI - L)f_n$, the conditional closure lemma provides a subsequence $n_k \uparrow \infty$ s.t. $f_{n_k} \to f$ where $g = (zI - L)f$. The limit f belongs to I_ℓ, because $f_n \in I_\ell$ and I_ℓ is closed by the induction hypothesis. Thus $g \in (zI - L)I_\ell \equiv I_{\ell+1}$ as required. This concludes the proof that I_ℓ is closed for all ℓ.

We show that $K(z) = K_\ell$ for some ℓ. By definition, $K_1 \subseteq K_2 \subseteq \cdots$, so if the assertion is false, then $K_{n-1} \subsetneq K_n$ infinitely often. Construct, using the Riesz lemma a sequence of vectors $f_{n_k} \in K_{n_k}$ s.t. $n_k \to \infty$, $\|f_{n_k}\| = 1$ and $\operatorname{dist}(f_{n_k}, K_{n_k-1}) \geq \frac{1}{2}$. So $\{f_{n_k}\}_{k\geq 1}$ is $\frac{1}{2}$–separated.

We claim that for every $m \in \mathbb{N}$, $\{L^m f_{n_i}\}_{i\geq 1}$ is $\frac{1}{2}|z|^{m+1}$–separated. To show this we write $z^{-m}L^m f_{n_i+k} - z^{-m}L^m f_{n_i} = f_{n_i+k} - \left[(I - z^{-m}L^m)f_{n_i+k} + z^{-m}L^m f_{n_i}\right]$, and show that the term in the brackets belongs to K_{n_i+k-1}. This means that $\|L^m f_{n_i+k} - L^m f_{n_i}\| \geq |z|^m \operatorname{dist}(f_{n_i+k}, K_{n_i+k-1}) \geq |z|^{m+1}/2$.

We begin with two trivial observations on K_ℓ. Firstly, $L(K_\ell) \subseteq K_\ell$ (because $(zI - L)^\ell L = L(zI - L)^\ell$). Secondly, $(zI - L)K_\ell \subseteq K_{\ell-1}$. The first observation shows that $L^m f_{n_i} \in K_{n_i}$. The second observation shows that

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

Thus the term in the brackets belongs to K_{n_i+k-1}, and $\|L^m f_{n_i+k} - L^m f_{n_i}\| \geq \frac{1}{2}|z|^{m+1}$.

We obtain a contradiction to this fact as follows. Recall that we are assuming that (A.2) holds with $k = 1$. Iterating, we get for all m and $f \in B$,

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

Applying this to $Lf_{n_k} - Lf_{n_\ell}$ we get

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

By our assumptions on $\|\cdot\|'$, since $\sup\|f_{n_k}\| < \infty$, $\exists k_i \uparrow \infty$ s.t. $\|Lf_{n_{k_i}} - h\|' \to 0$ for some $h \in B$. This means that for all $\varepsilon > 0$ and $m \geq 1$, we can find $i \neq j$ so large that

$$\|L^{m+1}f_{n_{k_i}} - L^{m+1}f_{n_{k_j}}\| \leq 2\|L\|r^m + \varepsilon.$$

Choosing m so large that $2\|L\|r^m < \frac{1}{4}|z|^{m+1}$ and $\varepsilon < \frac{1}{4}|z|^{m+1}$, we obtain $k_i \neq k_j$ s.t. $\|L^{m+1}f_{n_{k_i}} - L^{m+1}f_{n_{k_j}}\| < \frac{1}{2}|z|^{m+1}$. But this is impossible, because $\{L^m f_{n_k}\}_{k\geq 1}$ is $\frac{1}{2}|z|^{m+1}$–separated.

This proves that the sequence $K_1 \subseteq K_2 \subseteq \cdots$ stabilizes eventually. A similar argument, applied to $I_1 \supseteq I_2 \supseteq \cdots$ shows that that sequence eventually also stabilizes. The first step is complete.

Step 2. *$LK(z) \subseteq K(z)$, $LI(z) \subseteq I(z)$, and $B = K(z) \oplus I(z)$.*

Proof. The first two statements are because L commutes with $(zI-L)^\ell$. We show the third. The previous step shows that for some m, $K(z)=K_\ell, I(z)=I_\ell$ for all $\ell \geq m$. So it's enough to show that $B=K_m \oplus I_m$.

$B=K_m+I_m$: Suppose $f\in B$, then $(zI-L)^m f \in I_m = I_{2m}$ ($\because 2m>m$), so $\exists g\in B$ s.t. $(zI-L)^m f = (zI-f)^{2m} g$. We have $f=[f-(zI-L)^m g]+(zI-L)^m g \in K_m+I_m$.

$K_m \cap I_m = \{0\}$: Suppose $f\in K_m\cap I_m$, then $f=(zI-L)^m g$ for some $g\in B$. Necessarily $(zI-L)^{2m} g = (zI-L)^m f = 0$, so $g\in K_{2m}$. But $K_{2m}=K_m$, so $g\in K_m$. It follows that $f=(zI-L)^m g = 0$.

Step 3. $(zI-L): I(z)\to I(z)$ *is a bijection with bounded inverse.*

Proof. Let m be a number s.t. $I(z)=I_m, K(z)=K_m$. $(zI-L)$ is one-to-one on $I(z)$, because $\ker(zI-L)\cap I(z) \subseteq K_1 \cap I_m \subseteq K_m\cap I_m = \{0\}$. $(zI-L)$ is onto $I(z)$, because $(zI-L)I(z) = (zI-L)I_m = I_{m+1} = I_m = I(z)$. Thus

$$(zI-L): I(z)\to I(z) \text{ is a bijection.}$$

Since $I(z)$ is a closed subset of a Banach space, $(I(z),\|\cdot\|)$ is complete. By the open mapping theorem, $zI-L: I(z)\to I(z)$ is open. So $(zI-L)^{-1}$ is continuous, and therefore bounded.

Step 4. $K(z)=0$ *for all but at most finitely many* $z\in A(\rho,\rho(L))$. $K(z)\neq 0$ *for at least one* z *s.t.* $|z|=\rho(L)$.

Proof. Suppose by way of contradiction that $K(z)\neq\{0\}$ for infinitely many different points $z_i\in A(\rho,\rho(L))$ $(i\geq 1)$. Since $A(\rho,\rho(L))$ is compact, we may assume without loss of generality that $z_n \xrightarrow[n\to\infty]{} z\in A(\rho,\rho(L))$.

Since $K(z_n)\neq 0$, $\ker(z_n I-L)\neq 0$. Let $F_n := \ker(z_1 I-L)\oplus\cdots\oplus\ker(z_n I-L)$, then $F_1\subsetneq F_2 \subsetneq F_3 \subsetneq \cdots$. We now argue as in step 1. By the Riesz Lemma, $\exists f_n\in F_n$ s.t. $\|f_n\|=1$ and $\operatorname{dist}(f_n,F_{n-1})\geq \frac{1}{2}$. Using the obvious inclusion

$$L^m f_{n+k} - L^m f_n \in z_{n+k}^m f_{n+k} + F_{n+k-1}$$

we see that $\|L^m f_{n+k}-L^m f_n\| \geq \operatorname{dist}(z_{n+k}^m f_{n+k}, F_{n+k-1}) \geq \frac{1}{2}|z_{n+k}|^m \geq \frac{1}{2}\rho^m$. But this is ruled out by (A.2) as in step 1.

Thus $\{z\in A(\rho,\rho(L)) : K(z)\neq 0\}$ is finite. Next we claim that it contains an element on $\{z: |z|=\rho(L)\}$. Otherwise, $\exists\rho'<\rho(L)$ s.t. $K(z)=0$ for all $|z|\geq\rho'$. This means that $I(z)=B$ for all $|z|\geq\rho'$, whence by the previous step, $(zI-L)$ has a bounded inverse for all $|z|\geq\rho'$. It follows that the spectral radius of L is less than or equal to ρ'. But this is not the case, because $\rho'<\rho(L)$.

Step 5. *Let* $\lambda_1,\ldots,\lambda_t$ *denote the complete list of different eigenvalues of* L *in* $A(\rho,\rho(L))$, *then* $F := \bigoplus_{i=1}^t K(\lambda_i)$ *is a direct sum,* $\dim F<\infty$, $L(F)\subseteq F$, *and the eigenvalues of* $L|_F$ *are* $\lambda_1,\ldots,\lambda_t$.

Proof. Suppose $v_i\in K(\lambda_i)\setminus\{0\}$ and $\sum\alpha_i v_i = 0$. We have to show that $\alpha_j=0$ for all j. Suppose by way of contradiction that $\alpha_j\neq 0$ for some j.

Find, using step 1, an $m\geq 1$ s.t. $K(\lambda_i)=\ker(\lambda_i I-L)^m$, and set $p_i(z) := (\lambda_i - z)^m$. For every j, let $q_j(z) := \prod_{i\neq j} p_i(z)$, then $q_j(L)v_i = 0$ for all $i\neq j$, and so

$$0 = q_j(L)\left(\sum_i \alpha_i v_i\right) = \alpha_j q_j(L) v_j.$$

Since $\alpha_j \neq 0$, $q_j(L)v_j = 0$. Obviously, also $p_j(L)v_j = 0$.

The polynomials $q_j(z), p_j(z)$ have no zeroes in common, so they are relatively prime. Find polynomials $a(z), b(z)$ s.t. $a(z)p_j(z) + b(z)q_j(z) = 1$.

So $a(L)p_j(L)v_j + b(L)q_j(L)v_j = v_j$. But the left-hand-side vanishes, so $v_j = 0$ contrary to our assumptions. Thus the sum defining F is direct.

The dimension of F is finite by step 1. Clearly $\lambda_1, \ldots, \lambda_t$ are eigenvalues of $L|_F$. There are no other eigenvalues because $\prod_{i=1}^t (\lambda_i I - L|_F)^m = 0$, so the minimal polynomial of $L|_F$ divides $\prod_{i=1}^t (\lambda_i - t)^m$.

Step 6. *$H := \bigcap_{i=1}^t I(\lambda_i)$ is closed, $L(H) \subseteq H$, and $B = F \oplus H$.*

Proof. H is closed by step 1, and L–invariant by step 2.

For every $i = 1, \ldots, t$ $B = K(\lambda_i) \oplus H(\lambda_i)$ (step 2), so there exist continuous projection operators $\pi_i : B \to K(\lambda_i)$ s.t. for every $f \in B$,

$$\pi_i(f) \in K(\lambda_i) \text{ and } (I - \pi_i)(f) \in I(\lambda_i).$$

(Existence is because of the direct sum decomposition; continuity can be checked using the closed graph theorem.) We have

1. $\pi_i L = L\pi_i$, because $LK(\lambda_i) \subseteq K(\lambda_i), LI(\lambda_i) \subseteq I(\lambda_i)$;
2. $i \neq j \Rightarrow \pi_i \pi_j = 0$: Suppose $u \in B$, and let $v := \pi_j(u)$. Then $v \in K(\lambda_j)$, so $\exists m$ s.t. $(\lambda_j I - L)^m v = 0$. So $((\lambda_j - \lambda_i)I + (\lambda_i I - L))^m v = 0$, whence

$$(\lambda_j - \lambda_i)^m v + \sum_{\ell=1}^m \binom{m}{\ell} (\lambda_j - \lambda_i)^{m-\ell} (\lambda_i I - L)^\ell v = 0.$$

   So $v = -(\lambda_j - \lambda_i)^{-m} \sum_{\ell=1}^m \binom{m}{\ell} (\lambda_j - \lambda_i)^{m-\ell} (\lambda_i I - L)^\ell v$. Iterating this identity we see that for every n

$$v = \left[-(\lambda_j - \lambda_i)^{-m} \sum_{\ell=1}^m \binom{m}{\ell} (\lambda_j - \lambda_i)^{m-\ell} (\lambda_i I - L)^\ell \right]^n v \in \operatorname{Im}(\lambda_i I - L)^n,$$

   whence $v \in I(\lambda_i) \subseteq \ker \pi_i$. It follows that $(\pi_i \circ \pi_j)(u) = \pi_i(v) = 0$.

We can now show that $B = F \oplus H$. Every $f \in B$ can be decomposed into

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

The left summand is in F, the right summand is in $\bigcap_{i=1}^t \ker \pi_i = \bigcap_{i=1}^n I(\lambda_i) = H$. Thus $B = F + H$. At the same time $F \cap H = \{0\}$, because if $f \in F \cap H$, then $\pi_i(f) = 0$ for all i (because $f \in H$), whence $f = 0$ (because $f \in F$).

Step 7. *The spectral radius of $L|_H$ is strictly smaller than ρ.*

Proof. It is enough to show that $(zI-L):H\to H$ has a bounded inverse for all $|z|\geq\rho$. Fix such a z, and let h be some element of H.

Suppose $z\notin\{\lambda_1,\dots,\lambda_t\}$, then $K(z)=\{0\}$ so $I(z)=B$. By step 3, $(zI-L):B\to B$ is invertible with bounded inverse.

Now suppose $z=\lambda_i$ for some i. Recall that $(\lambda_i I-L):I(\lambda_i)\to I(\lambda_i)$ is an isomorphism, so $\exists! f\in I(\lambda_i)$ s.t. $h=(\lambda_i I-L)f$. We show that f belongs to H, by checking that $\pi_j(f)=0$ for all j. If $j=i$, use $f\in I(\lambda_i)=\ker\pi_i$. If $j\neq i$, then

$$0=\pi_j(h)=\pi_j(\lambda_i I-L)f=(\lambda_i I-L)\pi_j(f),$$

so $\pi_j(f)\in K(\lambda_i)\cap K(\lambda_j)=\{0\}$. Thus $f\in\bigcap\ker\pi_j=H$. We see that $\exists! f\in H$ s.t. $h=(zI-L)f$. It follows that $(zI-L):H\to H$ is invertible. Since H is closed, H is a Banach space. By the inverse mapping theorem, $(zI-L)^{-1}$ is bounded.

In summary, $B=F\oplus H$ where F,H are L–invariant spaces such that (a) F is finite dimensional, (b) H is closed, (c) all the eigenvalues of $L|_F$ have modulus larger than or equal to ρ, and (d) the spectral radius of $L|_H$ is strictly less than or equal to ρ. In other words: L is quasi-compact.

Step 7 completes the proof of Hennion's theorem in the special case when (A.2) holds with $k=1$. Suppose now that (A.2) holds with $k>1$. By what we just proved, L^k is quasi-compact, and we can decompose

$$B=F_0\oplus H_0$$

where F_0,H_0 are closed linear spaces s.t. $L^k(F_0)\subset F_0$, $L^k(H_0)\subset H_0$, $\dim(F_0)<\infty$, and there exists $r^k<\rho^k<\rho(L^k)$ arbitrarily close to r^k such that all eigenvalues of $L^k|_{F_0}$ have modulus at least ρ^k and the spectral radius of $L^k|_{H_0}$ is strictly less than ρ^k. Since $\rho(L^k)=\rho(L)^k$, $r<\rho<\rho(L)$ and ρ can be chosen arbitrarily close to r.

We saw in the proof above that $\exists\lambda_1,\dots,\lambda_{t_0}$ s.t. $|\lambda_i|>\rho$ s.t.

$$F_0=\bigoplus_{i=1}^{t_0}\ker[(\lambda_i I-L^k)^m].$$

There is also a useful formula for H_0:

Claim: $H_0=\{v\in B:\limsup\|L^{k\ell}v\|^{1/\ell}<\rho^k\}$.

Proof. The inclusion $\subseteq$ is because $\rho(L^k|_{H_0})<\rho^k$. To see $\supseteq$ we first observe that $L^k:F_0\to F_0$ is invertible, because $\dim(F_0)<\infty$ and $\ker(L^k|_{F_0})=\{0\}$ since zero is not an eigenvalue. So for all $v\in F_0$, $\|v\|\leq\|L^{-k\ell}|_{F_0}\|\|L^{k\ell}v\|$, whence

$$\|L^{k\ell}v\|^{1/\ell}\geq\|v\|^{1/\ell}\|L^{-k\ell}|_{F_0}\|^{-1/\ell}\xrightarrow[\ell\to\infty]{}\frac{1}{\rho(L^{-k}|_{F_0})}=\frac{1}{\max\{|\lambda_i^{-1}|\}}\geq\rho^k.$$

Now suppose $v\neq 0$ satisfies $\limsup\|L^{k\ell}\|^{1/\ell}<\rho^k$, and decompose $v=f+h$ with $f\in F_0, h\in H_0$. Then $f=0$, otherwise $\|L^{k\ell}v\|$ grows too fast. So $v\in H_0$.

Let $F_1 := \sum_{j=0}^{k-1} L^j(F_0)$. This is a closed L-invariant space, and $\dim(F) < \infty$. Suppose the minimal polynomial of $L^k|_{F_0}$ is $p(t)$. For every $v \in L^j(F_0)$, $p(L)L^{k-j}v = 0$, and so $p(L)L^k v = L^j p(L) L^{k-j}(v) = 0$. So the minimal polynomial of $L|_F$ divides $t^k p(t^k)$. It follows that all eigenvalues of $L|_F$ are either zero or are k-th roots of eigenvalues of $L^k|_{F_0}$. As such, they are either zero or have modulus at least ρ. Let $\mu_1, \dots, \mu_s$ denote the non-zero eigenvalues, then

$$F_1 = F \oplus \bigcup_{j \geq 1} \ker(L^j|_{F_1}), \text{ where } F := \bigoplus_{i=1}^{s} \bigcup_{j \geq 1} \ker[(\mu_i I - L|_{F_1})^j].$$

F has finite dimension, $L(F) \subset F$, and all eigenvalues of $L|_F$ have modulus $\geq \rho$. One shows as in the claim that for all $v \in F \setminus \{0\}$, $\liminf \|L^\ell v\|^{1/\ell} \geq \rho$.

Next write $H := H_0 \oplus \bigcup_{j \geq 1} \ker(L^j|_{F_1})$. This is again a closed L-invariant space, and because of the formula for H_0,

$$H = \{v \in B : \limsup \|L^\ell v\|^{1/\ell} < \rho\}.$$

Clearly $F \cap H = \{0\}$, and clearly $F + H = (F + \bigcup_{j \geq 1} \ker(L^j|_{F_1})) + H_0 \supseteq F_1 + H_0 \supseteq F_0 + H_0 = B$. So $B = F \oplus H$ and L is quasi-compact. $\square$

Reference: H. Hennion and L. Hervé: Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness, *Lect. Notes in Math.* **1766**, Springer, 145pp (2000)

A.4 The analyticity theorem

Theorem (Dunford): *Suppose $T : U \to B$ is a function from an open set in $\mathbb{C}$ to a Banach space B. The following conditions are equivalent:*

1. **Weak analyticity:** *For every bounded linear functional $\varphi : B \to \mathbb{C}$, $\varphi[T(z)]$ is holomorphic on U.*
2. **Strong analyticity:** *For every $z \in U$ there is $T'(z) \in B$ (called the* derivative at *z) s.t. $\|\frac{T(z+h)-T(z)}{h} - T'(z)\| \xrightarrow[|h|\to 0]{} 0$*

Proof. (2)⇒(1) is obvious, so we only do (1)⇒(2).

Lemma. *Suppose B is a Banach space and $x_n \in B$ satisfy $\|x_n\| \to \infty$, then there exists a bounded linear functional φ s.t. $\varphi(x_n) \to \infty$.*

Proof: Let B^* denote the space of bounded linear functionals on B equipped with the norm $\|\varphi\| = \sup \frac{|\varphi(x)|}{\|x\|}$. Every $x_n \in B$ defines a bounded linear functional $x_n^* : B^* \to \mathbb{C}$ through $x_n^*(\varphi) = \varphi(x_n)$, and it's an easy consequence of the Hahn–Banach Theorem that $\|x_n^*\| = \|x_n\|$. So $\sup \|x_n^*\| = \infty$. By the Banach–Steinhaus Theorem, there must exist $\varphi \in B^*$ s.t. $\sup x_n^*(\varphi) = \infty$, which is exactly what the lemma asserts.

We now prove (1)⇒(2). Suppose $T(z)$ is weakly differentiable on U. $\|T(z)\|$ must be locally bounded in U, otherwise $\exists z_n \to z \in U$ s.t. $\|T(z_n)\| \to \infty$, and then by the lemma $\varphi[T(z_n)] \to \infty$ for some bounded linear functional φ. But $\varphi[T(z_n)] \to \varphi[T(z)]$ because $\varphi[T(z)]$ is holomorphic and therefore continuous.

We show that $D(h) := \frac{1}{h}[T(z+h) - T(z)]$ satisfies the Cauchy criterion on U as $h \to 0$. Since $\varphi[T(z)]$ is holomorphic on U, it satisfies Cauchy's Integral formula: $\varphi[T(z)] = \frac{1}{2\pi i}\oint_{\partial B_r(z)} \frac{\varphi[T(\xi)]}{\xi - z} d\xi$. Here $B_r(z)$ is a disc with center z and radius r so small that $\overline{B_r(z)} \subset U$. Direct calculations show that

$$\varphi[D(h) - D(k)] = \frac{h-k}{2\pi i}\oint_{B_r(z)} \frac{\varphi[T(\xi)]}{(\xi - (z+h))(\xi - (z+k))(\xi - z)} d\xi.$$

Setting $M := \sup\{\|T(\xi)\| : |\xi - z| \le r\}$, we have for all $|h|, |k| < \frac{1}{2}r$

$$\left|\varphi[D(h) - D(k)]\right| \le |h-k| \cdot \frac{2\pi r}{2\pi} \cdot \frac{4M\|\varphi\|}{r^3}.$$

Since this holds for all bounded linear fuctionals, and by the Hahn–Banach theorem $\|x\| = \sup\{|\varphi(x)| : \varphi \in B^*, \|\varphi\| = 1\}$,

$$\|D(h) - D(k)\| \le |h-k| \cdot \frac{2\pi r}{2\pi} \cdot \frac{4M}{r^3} = O(|h-k|).$$

The Cauchy criterion follows. So $\lim_{h\to 0} \frac{1}{h}[T(z+h) - T(z)]$ exists. □

A.5 Eigenprojections, "separation of spectrum", and Kato's Lemma

Theorem (Sz.-Nagy, Wolf). *Suppose L is a bounded linear operator and* $\mathrm{Spect}(L) = \Sigma_{in} \uplus \Sigma_{out}$ *where $\Sigma_{in}, \Sigma_{out}$ are compact, and let γ be a smooth closed curve which does not intersect* $\mathrm{Spect}(L)$*, and which contains Σ_{in} in its interior, and Σ_{out} in its exterior. Then*

1. $P := \frac{1}{2\pi i}\oint_\gamma (zI-L)^{-1}dz$ *is a projection* ($P^2 = P$)*, therefore* $\mathscr{L} = \ker(P) \oplus \mathrm{Im}(P)$.
2. $PL = LP$*, therefore* $L(\ker(P)) \subset \ker(P)$ *and* $L(\mathrm{Im}(P)) \subset \mathrm{Im}(P)$.
3. $\mathrm{Spect}(L|_{\mathrm{Im}(P)}) = \Sigma_{in}$ *and* $\mathrm{Spect}(L|_{\ker(P)}) = \Sigma_{out}$.

Step 1: *P is a projection.*

Proof. Let $R(z) = (zI-L)^{-1}$. Since γ is compact and outside the spectrum, $\|R(z)\|$ is continuous and bounded on γ. It follows that $\|P\| < \infty$. We show that $P^2 = P$.

"Expand" γ to a larger curve γ^* which contains $\Sigma_{in} \cup \gamma$ in its interior and Σ_{out} in its exterior. P can be calculated by integrating on γ^* instead of γ (prove!), and so

$$
\begin{aligned}
P^2 &= \frac{1}{(2\pi i)^2}\oint_\gamma R(z)dz \oint_{\gamma^*} R(w)dw = \frac{1}{(2\pi i)^2}\oint_\gamma\oint_{\gamma^*} R(z)R(w)dwdz \\
&\qquad\qquad (\because R(\cdot) \text{ is linear and continuous on } \mathbb{C}\setminus\mathrm{Spect}(L)) \\
&= \frac{1}{(2\pi i)^2}\oint_\gamma\oint_{\gamma^*} \frac{R(z)-R(w)}{w-z}dwdz \qquad \text{(Resolvent identity)} \\
&= \frac{1}{(2\pi i)^2}\oint_\gamma\oint_{\gamma^*} \frac{R(z)}{w-z}dwdz - \frac{1}{(2\pi i)^2}\oint_\gamma\oint_{\gamma^*} \frac{R(w)}{w-z}dwdz \\
&= \frac{1}{(2\pi i)^2}\oint_\gamma \left(R(z)\oint_{\gamma^*}\frac{1}{w-z}dw\right)dz - \frac{1}{(2\pi i)^2}\oint_{\gamma^*}\left(R(w)\oint_\gamma \frac{1}{w-z}dz\right)dw.
\end{aligned}
$$

The first inner integral is $2\pi i$ (z is inside γ^*) and the second inner integral is zero (w is outside γ). The net result is $\frac{1}{2\pi i}\oint_\gamma R(z)dz = P$.

Step 2: $PL = LP$, $L(\ker(P)) \subset \ker(P)$, $L(\mathrm{Im}(P)) \subset \mathrm{Im}(P)$.

Proof: The resolvent of L commutes with L.

Step 3: $\mathrm{Spect}(L|_{\mathrm{Im}(P)}) = \Sigma_{in}$ and $\mathrm{Spect}(L|_{\ker(P)}) = \Sigma_{out}$.

Proof. We claim that $(zI-L)|_{\mathrm{Im}(P)}$ has bounded inverse on Σ_{out}. The idea is to extend $R(z)|_{\mathrm{Im}(P)}$ analytically outside of γ and observe that the extension must still be a bounded inverse.

To build the analytic extension, we note that $P = I$ on $\mathrm{Im}(P)$, therefore $R(z)|_{\mathrm{Im}(P)} = R(z)P|_{\mathrm{Im}(P)}$. For $z \notin \mathrm{Spect}(L)$ outside γ

$$\begin{aligned}R(z)P &= R(z)\left(\frac{1}{2\pi i}\oint_\gamma R(w)dw\right) = \frac{1}{2\pi i}\oint_\gamma R(z)R(w)dw \\ &= \frac{1}{2\pi i}\oint_\gamma \frac{R(w)-R(z)}{z-w}dw = \frac{1}{2\pi i}\oint_\gamma \frac{R(w)}{z-w}dw - R(z)\left(\frac{1}{2\pi i}\oint_\gamma \frac{dw}{z-w}\right) \\ &= \frac{1}{2\pi i}\oint_\gamma \frac{R(w)}{z-w}dw \ \ (z \text{ is outside } \gamma).\end{aligned}$$

The magic is that $\widehat{R}(z) := \frac{1}{2\pi i}\oint_\gamma \frac{R(w)}{z-w}dw$ makes sense and is analytic outside γ, *including on* Σ_{out}, and we have obtained an analytic extension of $R(z)P|_{\mathrm{Im}(P)}$ to the complement of Σ_{in}.

We know that $(zI-L)\widehat{R}(z)|_{\mathrm{Im}(P)}$ is analytic outside γ and equals I outside γ away from Σ_{out}. Two holomorphic functions which agree on a set with an accumulation point agree everywhere (prove using the weak differentiability criterion). It follows that $(zI-L)\widehat{R}(z)|_{\mathrm{Im}(P)} = I$ everywhere in the exterior of γ, including Σ_{out}. We found a bounded inverse for $(zI-L)|_{\mathrm{Im}(P)}$ for $z \in \Sigma_{out}$.

Since (obviously) $\mathrm{Spect}(L|_{\mathrm{Im}(P)}) \subset \mathrm{Spect}(L) = \Sigma_{in} \cup \Sigma_{out}$, and $\mathrm{Spect}(L|_{\mathrm{Im}(P)}) \cap \Sigma_{out} = \varnothing$, $\mathrm{Spect}(L|_{\mathrm{Im}(P)}) \subset \Sigma_{in}$. Similarly one proves that $\mathrm{Spect}(L|_{\ker(P)}) \subset \Sigma_{out}$.

The inequalities must be equalities: If for example $\exists z_0 \in \Sigma_{in} \setminus \mathrm{Spect}(L|_{\mathrm{Im}(P)})$ then we can invert $(zI-L)$ on $\ker(P)$ and on $\mathrm{Im}(P)$, whence on $\ker(P) \oplus \mathrm{Im}(P) = \mathscr{L}$. But we can't. □

Lemma (Kato). *Let $P,Q : B \to B$ be two projections on a Banach space B. If $\|P-Q\| < 1$ then P,Q are similar: $\exists$ bounded linear isomorphism π s.t. $P = \pi^{-1}Q\pi$.*

Proof. First we construct a map $U : B \to B$ which maps $\ker(P)$ into $\ker(Q)$, and $\mathrm{Im}(P)$ into $\mathrm{Im}(Q)$: $U := (I-Q)(I-P) + QP$. Observe that

$$\begin{aligned}UP &= (I-Q)(I-P)P + QP^2 = QP \ \ (\because P^2 = P) \\ QU &= Q(I-Q)(I-P) + Q^2P = QP \ \ (\because Q^2 = Q)\end{aligned}$$

We see that $UP = QU$. If we can show that U has a bounded inverse, then $P = U^{-1}QU$ and P,Q are similar.

Consider the map $V : B \to B$ which maps $\ker(Q)$ into $\ker(P)$, and $\mathrm{Im}(Q)$ into $\mathrm{Im}(P)$: $V := (I-P)(I-Q) + PQ$. This is "almost" an inverse for U:

$$\begin{aligned}UV &= (I-Q)(I-P)(I-Q) + QPQ = \cdots = I - Q - P + PQ + QP = I - (P-Q)^2 \\ VU &= (I-P)(I-Q)(I-P) + PQP = \cdots = I - Q - P + PQ + QP = I - (P-Q)^2\end{aligned}$$

If $\|P-Q\| < 1$, then $I - (P-Q)^2$ is invertible, whence one-to-one and onto. Since UV is onto, U is onto. Since VU is one-to-one, U is one-to-one. It follows that U is invertible. Any invertible map on a Banach space has bounded inverse (open mapping theorem). It follows that U is a bounded linear isomorphism. □

Reference: T. Kato: Perturbation theory for linear operators, *Classics in Math., Springer,* xxi+619pp (1980)

A.6 The Berry–Esseen "Smoothing Inequality"

Theorem (Berry & Esseen): *$\exists C > 0$ s.t. if F, G are two probability distribution functions with characteristic functions $f(t), g(t)$ and if $G(x)$ differentiable, $\sup|G'| < \infty$, and $\int |F(x) - G(x)| dx < \infty$, then*

$$\|F - G\|_\infty \le C \left(\frac{1}{2\pi} \int_{-T}^{T} \frac{|f(t) - g(t)|}{|t|} dt + \frac{\sup|G'|}{T} \right) \text{ for all } T > 0.$$

T is a free parameter which we are free to choose to optimize the bound.

The proof uses several tools from real analysis which we will now review briefly.

Lebesgue–Stieltjes integrals: Any distribution function F determines a unique Borel probability measure on $\mathbb{R}$ by $\mu_F([a,b)) := F(b) - F(a)$. This is called the *Lebesgue–Stieltjes measure* of F. It is customary to use the following notation

$$\int_a^b f(x) F(dx) \text{ or } \int_a^b f(x) dF(x) \text{ for } \int_{[a,b)} f d\mu_F.$$

Note that the right endpoint of the interval is not included. This matters when $F(x)$ has a jump discontinuity at b, because in this case μ_F has an atom at b.

Fourier transforms: The Fourier transform of $f \in L^1(\mathbb{R})$ is $\mathfrak{F}(f)(t) = \int e^{itx} f(x) dx$. This has the following properties:

1. $\mathfrak{F}(\mathfrak{F}(f)) = 2\pi f$
2. $\mathfrak{F}(f * g) = \mathfrak{F}(f) \cdot \mathfrak{F}(g)$, where $(f * g)(x) = \int f(x-y) g(y) dy$ (the *convolution*).

The Fourier transform of a Borel probability measure μ on $\mathbb{R}$ is the function $(\mathfrak{F}\mu)(t) := \int e^{itx} d\mu(x)$. The reader can check that characteristic function of a random variable X is the Fourier transform of the Stieltjes measure of the distribution function of X. This only depends on the distribution function of X. Therefore we can safely speak of the characteristic function of a distribution function.

Lemma. *Suppose $F(x), G(x)$ are two distribution functions with characteristic functions $f(t), g(t)$. If $\int |F(x) - G(x)| dx < \infty$, then $[\mathfrak{F}(F - G)](t) = -\frac{f(t) - g(t)}{it}$.*

Proof. The Fourier transform of $F - G$ exists, because $F - G \in L^1$. Let μ_F and μ_G denote the Lebesgue–Stieltjes measures of F, G, then

$$\begin{aligned}
[\mathfrak{F}(F - G)](t) &= \lim_{T \to \infty} \int_{-T}^{T} e^{itx} [F(x) - G(x)] dx \\
&= \lim_{T \to \infty} \left[\int_{-T}^{T} \int_{-\infty}^{T} e^{itx} 1_{[\xi < x]} d\mu_F(\xi) dx - \int_{-T}^{T} \int_{-\infty}^{T} e^{itx} 1_{[\xi < x]} d\mu_G(\xi) dx \right]
\end{aligned}$$

$$= \lim_{T\to\infty}\int_{-\infty}^{T}\left(\int_{\xi}^{T} e^{itx}dx\right)d\mu_F(\xi) - \lim_{T\to\infty}\int_{-\infty}^{T}\left(\int_{\xi}^{T} e^{itx}dx\right)d\mu_G(\xi)$$
$$= \lim_{T\to\infty}\int_{-\infty}^{T}\frac{e^{itT}-e^{it\xi}}{it}d\mu_F(\xi) - \lim_{T\to\infty}\int_{-\infty}^{T}\frac{e^{itT}-e^{it\xi}}{it}d\mu_G(\xi)$$
$$= \lim_{T\to\infty}\left[\frac{e^{itT}}{it}[F(T)-G(T)] - \int_{-\infty}^{T}\frac{e^{it\xi}}{it}d(\mu_F-\mu_G)(\xi)\right].$$

The first summand tends to zero because $F(T), G(T) \xrightarrow[T\to\infty]{} 1$, and the second summand tends to $-\frac{f(t)-g(t)}{it}$. □

Lemma *There exists a non–negative, even, absolutely integrable function $H(x)$ s.t. $\int H(x)dx = 1$, $b := \int |x|H(x)dx < \infty$, $H(x) \xrightarrow[|x|\to\infty]{} 0$, $\mathfrak{F}(H)$ is real–valued and non–negative, and $\mathfrak{F}(H)$ is supported inside $[-1,1]$.*

Proof. There are many possible constructions. Here is one. Start with the indicator of a symmetric interval $[-a,a]$, and take its Fourier transform

$$H_0(y) = \int_{-a}^{a} e^{ity}dt = \frac{2\sin ay}{y}.$$

The Fourier transform of H_0 is $\mathfrak{F}H_0 = 2\pi 1_{[-a,a]}$, so it has compact support. But H_0 is not non-negative, and $\int |x|H_0(x)dx = \infty$. To correct this we let $H_1(x) := (H_0(x))^4$, and observe that $H_1(x) \geq 0$ and $\int |x|H(x)dx < \infty$. The Fourier transform of H_1 still has compact support (in $[-4a,4a]$), because

$$\mathfrak{F}[(H_0)^4] = \mathfrak{F}[(\mathfrak{F}1_{[-a,a]})^4] = \mathfrak{F}\{\mathfrak{F}[(1_{[-a,a]} * 1_{[-a,a]} * 1_{[-a,a]} * 1_{[-a,a]}]\} = 2\pi(1_{[-a,a]})^{*4},$$

and the convolution of functions with compact support has compact support. H_1 is even, because it is the convolution of even functions. It remains to normalize H_1. □

Proof of the Berry-Esseen Theorem. Let $H(x)$ be the function given by the lemma, and let $h := \mathfrak{F}H$. Set $H_T(x) := TH(Tx)$, then $H_T(x)$ is an even non–negative absolutely integrable function s.t.

1. $\int H_T dx = 1$;
2. $\int |x|H_T(x)dx = b/T$;
3. The Fourier transform of H_T is $h_T(t) := h(t/T)$ where $h = \mathfrak{F}H$.

Note that h_T is supported in $[-T,T]$, and $|h_T| \leq \|H_T\|_1 = 1$.

The proof is based on the following heuristic: For T large, $H_T(x)$ has a sharp peak at $x = 0$, and rapid decay for x far from zero. If we average a "nice" function $\varphi(y)$ with weights $H_T(x-y)$, then we expect the result to be close to $\varphi(x)$. In particular

$$|F(x)-G(x)| \overset{?}{\approx} I_T(x) := \int H_T(x-y)[F(y)-G(y)]dy.$$

We will estimate $I_T(x)$ in terms of $f(t), g(t)$, and relate $M := \sup|F(x)-G(x)|$ to the value of $I_T(\cdot)$ at a point where $|F(x)-G(x)|$ is (nearly) maximal.

Step 1. $I_T(x) \le \frac{1}{2\pi}\int_{-T}^{T} \frac{|f(t)-g(t)|}{|t|}dt$.

Proof. $I_T(x) = \left|\int H_T(x-y)[F(y)-G(y)]dy\right| = |H_T * (F-G)|$

$$\begin{aligned}
&= (2\pi)^{-1}\left|\mathfrak{F}^2[H_T*(F-G)]\right| = 2\pi^{-1}\left|\mathfrak{F}[\mathfrak{F}H_T\cdot\mathfrak{F}(F-G)]\right| \\
&= (2\pi)^{-1}\left|\mathfrak{F}[h_T\cdot\mathfrak{F}(F-G)]\right| && \text{(A.4)}\\
&= (2\pi)^{-1}\left|\int_{-\infty}^{\infty} e^{itx}h_T(t)\frac{f(t)-g(t)}{it}dt\right| \quad \text{(lemma)}\\
&\le \frac{1}{2\pi}\int_{-T}^{T}\frac{|f(t)-g(t)|}{|t|}dt, && \text{(A.5)}
\end{aligned}$$

because $|h_T(t)| \le \|H_T\|_1 = 1$ and h_T is supported in $[-T,T]$.

Step 2. *Relating $\|F-G\|_\infty$ to $I_T(x_0)$ at x_0 where $|F(x_0)-G(x_0)| \approx \|F-G\|_\infty$.*

Let $A := \sup|G'(x)|$ and $M := \sup|F(x)-G(x)|$. Fix some point $x_0 \in \mathbb{R}$ s.t. $M_0 := |F(x_0)-G(x_0)| > \frac{1}{2}M$. Since we are free to translate the distributions F,G by the same amount, we may assume w.l.o.g. that $x_0 = 0$. So $M_0 = |F(0)-G(0)|$ and

$$I_T(x_0) = I_T(0) = \int H_T(y)[F(y)-G(y)]dy.$$

(we have used the fact that H_T is even).

Suppose first $F(0) > G(0)$, and decompose the integral $I_T(0)$ into $\int_0^{M_0} + \int_{-\infty}^{0} + \int_{M_0}^{\infty}$.

1. To analyze $\int_0^{M_0}$ we note that if $y \in [0,M_0]$, then $F(y) \ge F(0)$ and so

$$[F(y)-G(y)]-[F(0)-G(0)] \ge G(0)-G(y) = -\int_0^y G'(y)dy \ge -Ay.$$

   Thus $[F(y)-G(y)] \ge [F(0)-G(0)] - Ay = M_0 - Ay$ ($\because F(0) > G(0)$). So

$$\int_0^{M_0} H_T(y)[F(y)-G(y)]dy \ge \int_0^{M_0}(M_0-Ay)H_T(y)dy.$$

2. We estimate $\int_{-\infty}^{0}$ from below by replacing $[F(y)-G(y)]$ by $-M > -2M_0$:

$$\int_{-\infty}^{0} H_T(y)[F(y)-G(y)]dy \ge -\int_{-\infty}^{0} H_T(y)\cdot 2M_0 dy.$$

3. Similarly, $\int_{M_0}^{\infty} H_T(y)[F(y)-G(y)]dy \ge -\int_{M_0}^{\infty} H_T(y)\cdot 2M_0 dy$.

Putting this all together, and recalling that H_T is even, we obtain

$$
\begin{aligned}
I_T(0) &\geq \int_0^{M_0} (M_0 - Ay) H_T(-y) dy - \int_{-\infty}^0 2M_0 H_T(-y) dy - \int_{M_0}^{\infty} 2M_0 H_T(y) dy \\
&= \int_0^{M_0} (3M_0 - Ay) H_T(y) dy - M_0 \\
&\geq 3M_0 \int_0^{M_0} H_T(y) dy - A \int |y| H_T(y) dy - M_0 \\
&= -M_0 + 3M_0 \int_0^{M_0} H_T(y) dy - \frac{Ab}{T} \quad (\because \int |y| H_T(y) dy = \frac{1}{T} \int |y| H(y) dy = \frac{b}{T}) \\
&= -M_0 + \frac{3M_0}{2} \int_{-M_0}^{M_0} H_T(y) dy - \frac{Ab}{T}
\end{aligned}
$$

In summary $M_0[\frac{3}{2}\int_{-M_0}^{M_0} H_T(y)dy - 1] \leq I_T(0) + \frac{Ab}{T}$.

Fix some $\sigma > 0$ s.t. $\int_{-\sigma}^{\sigma} H(y)dy = \frac{8}{9}$, then $\int_{-\sigma/T}^{\sigma/T} H_T(y)dy = \frac{8}{9}$. It is no problem to choose H from the beginning in such a way that $\sigma < A$. There are two cases:

1. $M_0 \leq \frac{\sigma}{T}$, and then $M < 2M_0 \leq 2\sigma/T < 2A/T$;
2. $M_0 > \frac{\sigma}{T}$, and then $\frac{3}{2}\int_{-M_0}^{M_0} H(y)dy - 1 > \frac{1}{3}$, so $M_0 \leq 3I_T(0) + \frac{3Ab}{T}$.

In both cases, this and step 1 yields

$$
\sup |F(x) - G(x)| < 2M_0 \leq 6 \left(\frac{1}{2\pi} \int \frac{|f(t) - g(t)|}{|t|} dt + \frac{\max\{3b, 2\}A}{T} \right),
$$

and the proposition is proved, under the additional assumption that $F(0) > G(0)$.

If $F(0) \leq G(0)$, then we repeat the same procedure, but with the decomposition $\int_{-M_0}^{0} + \int_{-\infty}^{-M_0} + \int_0^{\infty}$. This leads to $\int H_T(y)[G(y) - F(y)]dy \geq \int_{-M_0}^{0}(3M_0 - A|y|)H_T(y)dy - M_0$. From this point onward, the proof continues as before. □

Reference: *B.V. Gnedenko and A.N. Kolmogorov:* Limit distributions for sums of independent random variables, *Addison–Wesley, ix+264pp (1954).*