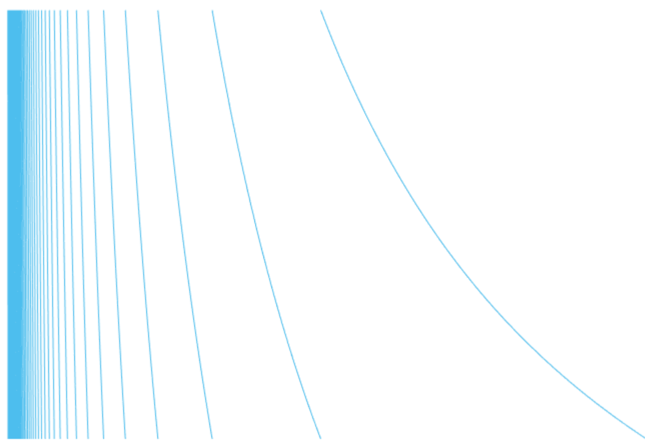
Decay of Correlations for the Gauss Map

Samuel Ekström

June 2024



Acknowledgements

I would like to express my deepest gratitude to Tomas Persson for his invaluable supervision throughout the writing of this thesis. His endless patience and teaching in mathematics have been a great source of comfort and motivation. Additionally, I extend my heartfelt thanks to those who provide stability in my life; you made this possible, and no words are sufficient to convey my gratitude.

Contents

Abstract

We study the Gauss map and derive certain statistical properties, specifically upper bounds on the decay of correlations. The method is due to Liverani [7], which in turn is based on foundational work by Birkhoff [1]. One defines a cone of functions, and then considers the induced Hilbert metric associated to the cone. A strict contraction in this metric gives exponential decay of correlations. First functions belonging to $C^1([0,1])$ are considered and the result is extended to $\mathrm{Lip}([0,1])$. Second we focus on functions of bounded total variation on $[0,1]$. A result in metric number theory regarding growth of continued fractions coefficients follows from the exponential decay of correlations for BV. The proof makes use of a dynamical Borel–Cantelli lemma.

1 Introduction

Dynamical systems and ergodic theory are research fields that deal with the long term behaviour of systems, where some sort of quantity is preserved. This is modeled mathematically by some space X describing the states, a transformation T describing the evolution in time, and a measure μ related to the time-invariant property. One says that the measure μ is invariant under T if $\mu(T^{-1}A) = \mu(A)$ for all μ-measurable $A \subset X$. The pair (X, T) is called a dynamical system. When we include a T invariant measure μ, the triple (X, T, μ) is referred to as a measurable dynamical system. Classical example often come from mechanics, where the invariant quantity is the energy of the system. When studying dynamical systems, it is of interest to ask how the future behaviour of the system depends upon the choice of initial conditions or what is often called *observables*. Assuming one has a transformation T and an invariant measure μ, we define the *correlation* of the system given two observables f, g at time n as

$$C(f,g,n) \coloneqq \left| \int_X f \cdot g \circ T^n \, d\mu - \int_X f \, d\mu \cdot \int_X g \, d\mu \right| .$$

The speed at which this quantity goes to zero gives information regarding the systems statistical properties. Of course one must specify some collection of observables. For example, if characteristic functions on Borel sets are allowed, then the correlation becomes

$$C(\chi_A, \chi_B, n) = \left| \mu(A \cap T^{-n}B) - \mu(A)\mu(B) \right|$$

and if this goes to zero, the system is *mixing*. In this thesis we shall explore statistical properties of the Gauss map $T(x) = \frac{1}{x} \mod 1$ and how the rate of decay differs for different sets of observables. More specifically, we shall employ a method giving *explicit* bounds on the speed of decay, meaning we look at

$$C(f,g,n) = \left| \int_0^1 f \cdot g \circ T^n \, d\mu - \int_0^1 f \, d\mu \cdot \int_0^1 g \, d\mu \right| \leq D \, \|f\| \, \|g\| \, \rho^n$$

and find upper bounds on ρ. Given that the Gauss is intimately associated with regular continued fraction expansions, these statistical properties then give information regarding the distribution and growth of the associated continued fraction coefficients.

The method relies first on defining the Perron–Frobenius operator $\mathcal{L}_T$, which is the adjoint operator of $U_T g = g \circ T$, or simply stated composition with T. Questions regarding the statistical properties of T can then be rephrased as spectral properties of $\mathcal{L}_T$. One then defines a cone of functions $\mathcal{C}$, which one hopes is mapped strictly inside itself by the operator. What was shown by Birkhoff in [1] is that this map is a strict contraction in the projective Hilbert (pseudo) metric

Θ, provided that the image is of finite diameter. This gives exponential decay of correlations for functions lying in the cone. The results can then be extended to functions lying outside the cone. This program is carried out by Liverani in both [7] and [6]. The novel part of this thesis is that the Gauss map does not lie in the collection of maps considered in either of these papers. We first study a cone of C^1 functions, follow the program and establish exponential decay of correlations. The result is then extended further to all Lipschitz functions on $[0, 1]$. We move on to functions of bounded total variation, and yet again show exponential decay. This case is slightly more technical due to necessity of Lemma 4.7, a lemma which has no analogue in the cases considered in [7].

Finally we prove a result regarding the frequency of large continued fraction coefficients through, which can be proved using the exponential decay of correlations for functions of bounded variation, and a dynamical Borel–Cantelli lemma.

2 Theory

The reader will be expected to be familiar with measure theory, at least to the level expected after an introductory course in the subject. Some major results and definitions will however be stated for the readers convenience. The main tool of this thesis, the Perron–Frobenius operator, is considered vital enough that it is given its own section to shine. To begin however, some results from projective geometry are showcased.

2.1 Projective geometry and Lattice Theory

In Birkhoff's book *Lattice Theory* [1] a general theory of lattices is presented. While the scope of this field is wide, the chapters on vector lattices and positive linear operators are sufficient for our purposes. They form the basis of both [7] and [6], which are the main sources of inspiration for this thesis. Presenting a concise, yet sufficiently thorough, introduction to the needed background will be the goal of this section.

A subset $\mathscr{C}$ of a vector space $\mathbb{V}$ is called a *cone* if $f \in \mathscr{C}$ implies that $\lambda f \in \mathscr{C}$ for every positive λ in $\mathbb{R}$. Birkhoffs presentation is for general group structures. We shall limit ourselves to the theory of vector spaces. By convex positive cone, we shall mean any cone $\mathscr{C}$ that satisfies

$$\mathscr{C} \cap -\mathscr{C} = \{0\}, \quad \mathscr{C} + (-\mathscr{C}) = \mathbb{V}, \quad \mathscr{C} + \mathscr{C} \subset \mathscr{C}.$$

An example of such a cone that is a subset of $\mathbb{R}^n$, given the canonical basis, is the set of vectors with non-negative coordinates. It should be clear this cone fulfills all the conditions above. In fact, this is true given any choice of basis defining the coordinate system. We introduce a partial ordering on a cone $\mathscr{C}$ and denote it by $\preceq$. The relation is induced by the cone in the following manner:

$$f \preceq g \iff g - f \in \mathscr{C}$$

where it becomes clear that the relation is dependent on the choice of cone. When one studies maps from one cone to another, it will be important to keep in mind which relation one is referencing, and perhaps if one where to be completely clear, the relation would be denoted by $\preceq_{\mathscr{C}}$. We will disregard this and hope that it shall be clear from context. We shall call two vectors *comparable* if $\lambda f \preceq g \preceq \mu f$ for some choice of scalars λ and μ. From this notion of comparability, it becomes natural to define

$$\begin{aligned}\alpha(f,g) &= \sup\left\{\, \lambda \in \mathbb{R}^+ \;\middle|\; \lambda f \preceq g \,\right\} \\ \beta(f,g) &= \inf\left\{\, \mu \in \mathbb{R}^+ \;\middle|\; g \preceq \mu f \,\right\}.\end{aligned} \tag{2.1}$$

If either set is empty, set $\alpha = 0$ or $\beta = \infty$ respectively. Together they can be used to define Hilbert's projective (pseudo)-metric

$$\Theta(f,g) = \log\left[\frac{\beta(f,g)}{\alpha(f,g)}\right] \tag{2.2}$$

where if $\beta(f,g) = \infty$ or $\alpha(f,g) = 0$, we put $\Theta(f,g) = \infty$. It is not a true metric, since $\Theta(f, af) = 0$, and the distance between points may be infinite, but it does fulfill the triangle inequality and it is symmetric. This we shall prove. The assumption that f, g are comparable $\lambda f \prec g \prec \mu f$ implies that $\mu^{-1} g \prec f \prec \lambda^{-1} g$, which gives that $\alpha(g,f) = \beta(f,g)^{-1}$ and that $\beta(g,f) = \alpha(f,g)^{-1}$. So

$$\Theta(g,f) = \log\left[\frac{\alpha(f,g)^{-1}}{\beta(f,g)^{-1}}\right] = \Theta(f,g)$$

and hence Θ is symmetric. Now for the triangle inequality. Let $\{\lambda_n\}_{n=1}^{\infty}, \{\tilde{\lambda}_n\}_{n=1}^{\infty}$ be two increasing sequences with $\lambda_n \uparrow \alpha(f,h), \tilde{\lambda}_n \uparrow \alpha(h,g)$ and analogously two decreasing sequences $\mu_n \downarrow \beta(f,h),\ \tilde{\mu}_n \downarrow \beta(h,g)$. Then

$$\lambda_n f \prec h \prec \mu_n f \quad \text{and} \quad \tilde{\lambda}_n h \prec g \prec \tilde{\mu}_n h$$

which can be combined into

$$\lambda_n \tilde{\lambda}_n f \prec g \prec \mu_n \tilde{\mu}_n f \quad \text{where} \quad \lambda_n \tilde{\lambda}_n \uparrow \alpha(f,h)\alpha(h,g) \quad \text{and} \quad \mu_n \tilde{\mu}_n \downarrow \beta(f,h)\beta(h,g)$$

finally yielding

$$\Theta(f,g) \leq \lim \log\left[\frac{\mu_n \tilde{\mu}_n}{\lambda_n \tilde{\lambda}_n}\right] = \lim \log\left[\frac{\mu_n}{\lambda_n}\right] + \log\left[\frac{\tilde{\mu}_n}{\tilde{\lambda}_n}\right] = \Theta(f,h) + \Theta(h,g)$$

where we get inequality due to $\beta(f,h) + \beta(h,g) \geq \beta(f,g)$ and analogously $\alpha(f,h) + \alpha(h,g) \leq \alpha(f,g)$. We leave it to the reader to ponder when the triangle inequality gives equality. The previous discussion is summarised into the following theorem.

Theorem 2.1. *The function Θ is non-negative, symmetric and fulfills the triangle inequality. Hence, it is a pseudo metric.*

We shall now present a result originally by Birkhoff in [1]. For the proof however, we follow a more modern approach by Liverani in [6] which does not rely on projective geometry. To understand the result we must however first introduce the concept of being *integrally closed*. It is desirable for the topology of the vector space and the partial ordering to be well behaved in relation to each other. Let f_n be a sequence of functions in $\mathscr{C}$ converging to f in the metric induced by $\mathscr{C}$, with $f_n \preceq g$. We say that $\mathscr{C}$ is integrally closed if for all such sequences, $f \preceq g$. By vector lattice we mean a vector space $\mathbb{V}$ with a partial order relation $\leq$ such that for all $f, g \in \mathbb{V}$ both $f \leq g \iff 0 \leq g - f$ and $0 \leq f \implies 0 \leq \lambda f$ $(\lambda \in \mathbb{R}^+)$ hold.

Theorem 2.2. *Let $\mathbb{V}_1$, and $\mathbb{V}_2$ be two integrally closed vector lattices; $P : \mathbb{V}_1 \to \mathbb{V}_2$ a linear map such that $P\mathscr{C}_1 \subset \mathscr{C}_2$, for two closed convex cones $\mathscr{C}_1 \subset \mathbb{V}_1$ and $\mathscr{C}_2 \subset \mathbb{V}_2$ with $\mathscr{C}_i \cap -\mathscr{C}_i = \{0\}$. Let Θ_i be the Hilbert metric corresponding to the cone $\mathscr{C}_i$. Setting $\Delta = \sup_{f,g \in P\mathscr{C}_1} \Theta_2(f,g)$ we have*

$$\Theta_2(Pf, Pg) \leq \tanh\left(\frac{\Delta}{4}\right)\Theta_1(f,g) \quad \forall f, g \in \mathscr{C}_1$$

Note that the diameter of $P\mathscr{C}_1$ may very well be infinite. In that case one makes the natural interpretation that $\tanh\infty = 1$.

Proof. Since $\Theta_1(f,g) = \log\frac{\beta(f,g)}{\alpha(f,g)}$, when either $\alpha(f,g) = 0$ or $\beta(f,g) = \infty$ then the right hand side is infinite so the inequality holds. We therefore assume that $\alpha \neq 0$ and $\beta \neq \infty$. The fact that $f, g \in \mathscr{C}_1$ gives that $0 \preceq f, g$. The assumption that $\mathscr{C}_i$ is integrally closed gives that $\alpha f \preceq g \preceq \beta f$. This can instead be written as

$$0 \preceq g - \alpha f, \quad 0 \preceq \beta f - g \iff g - \alpha f,\ \beta f - g \in \mathscr{C}_1$$

and by the assumption that $P\mathscr{C}_1 \subset \mathscr{C}_2$ we get that $P(g - \alpha f),\ P(\beta f - g) \in \mathscr{C}_2$ or equivalently

$$0 \preceq P(g - \alpha f),\ 0 \preceq P(\beta f - g) \iff \alpha Pf \preceq Pg \preceq \beta Pg$$

which gives that $\alpha_1(f,g) \leq \alpha_2(Pf,Pg)$ and likewise that $\beta_1(f,g) \geq \beta_2(Pf,Pg)$. Consider first the case when the diameter of the image is infinite. Then we must show that $\Theta_2(Pf,Pg) \leq \Theta_1(f,g)$. This follows immediately by

$$\Theta_2(Pf,Pg) = \log\frac{\beta_2(Pf,Pg)}{\alpha_2(Pf,Pg)} \leq \log\frac{\beta_1(f,g)}{\alpha_1(f,g)} = \Theta_1(f,g).$$

What remains is the case when $\Delta < \infty$. The assumption that $\Delta < \infty$ implies that there must exist λ and μ both bigger than zero, such that $\log\frac{\mu}{\lambda} \leq \Delta$ where λ and μ also satisfy

$$\lambda P(g - \alpha f) \preceq P(\beta f - g) \preceq \mu P(g - \alpha f).$$

Rewriting these inequalities to isolate Pf and Pg leads to

$$\frac{\beta + \alpha\mu}{1+\mu} Pf \preceq Pg \preceq \frac{\beta + \alpha\lambda}{1+\lambda} Pf$$

which in turn gives

$$\begin{aligned}
\Theta_2(Pf,Pg) &\leq \log\frac{(\beta + \lambda\alpha)(1+\mu)}{(\beta + \alpha\mu)(1+\lambda)} = \log\frac{\frac{\beta}{\alpha} + \lambda}{\frac{\beta}{\alpha} + \mu} - \log\frac{1+\lambda}{1+\mu} \\
&= \log\frac{\exp\Theta_1(f,g) + \lambda}{\exp\Theta_1(f,g) + \mu} - \log\frac{1+\lambda}{1+\mu} \\
&= \left[\frac{e^x + \lambda}{e^x + \mu}\right]_0^{\Theta_1(f,g)} = \int_0^{\Theta_1(f,g)} \frac{e^x(\mu - \lambda)}{(e^x + \mu)^2} dx \\
&\leq \int_0^{\Theta_1(f,g)} \frac{e^x(\mu - \lambda)}{(e^x + \mu)(e^x + \lambda)} dx
\end{aligned}$$

where one can find the maximum of the integrand to be $\frac{1 - \lambda/\mu}{(1 + \sqrt{\frac{\lambda}{\mu}})^2}$, giving

$$\Theta_2(Pf,Pg) \leq \Theta_1(f,g)\frac{(1 - \frac{\lambda}{\mu})}{(1 + \sqrt{\frac{\lambda}{\mu}})^2}$$

now, λ and μ where assumed to fulfill $\log\frac{\mu}{\lambda} \leq \Delta$. Which gives that

$$1 - \frac{\lambda}{\mu} \leq 1 - e^{\Delta} \text{ and } \left(1 + \sqrt{\frac{\lambda}{\mu}}\right)^{-2} \leq (1 + e^{-\Delta/2})^{-2}$$

and these two estimates along with some algebra produces

$$\Theta_1(f,g)\frac{(1-\frac{\lambda}{\mu})}{(1+\sqrt{\frac{\lambda}{\mu}})^2} \leq \Theta_1(f,g)\frac{1-e^{-\Delta}}{(1+e^{-\Delta/2})^2} = \Theta_1(f,g)\frac{\sinh\frac{\Delta}{4}}{\cosh\frac{\Delta}{4}} = \Theta_1(f,g)\tanh\frac{\Delta}{4}$$

which proves the desired result. □

2.2 Results from measure theory

The reader is assumed to be familiar with the basic results of an introductory course in measure theory and Lebesgue integration. Some selected results and definitions will however still be presented for the readers convenience. We begin by stating a well known result that will be of great importance

Theorem 2.3 (Radon–Nikodym Theorem)**.** *Let $(X, \mathcal{B})$ be a measurable space and let ν and μ be two normalized measures on $(X, \mathcal{B})$. If $\nu \ll \mu$, then there exists a unique $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$ such that for every $A \in \mathcal{B}$,*

$$\nu(A) = \int_A f d\mu.$$

f is called the Radon–Nikodym derivative and is denoted by $\frac{d\nu}{d\mu}$.

The importance of this theorem will be clear in section 2.3 when discussing the Perron–Frobenius operator.

Definition 2.4. *Let $(X_1, \mathcal{B}_1), (X_2, \mathcal{B}_2)$ be two measurable spaces and μ a measure on X_1 and $T : X_1 \to X_2$ be a μ-measurable transformation. Then we can define the **pushforward** measure, $T_*\mu$ by*

$$T_*\mu(A) = \mu(T^{-1}A), \quad \text{for all } A \in \mathcal{B}_2.$$

We will be particularly interested in the case when $X_1 = X_2$ (and $\mathcal{B}_1 = \mathcal{B}_2$)

Definition 2.5. *Let $(X, \mathcal{B}, \mu)$ be a normalized measure space. Then a transformation $T : X \to X$ is said to be nonsingular if and only if $T_*\mu \ll \mu$, or stated explicitly, if for any $A \in \mathcal{B}$ such that $\mu(A) = 0$, we have $T_*\mu(A) = \mu(T^{-1}A) = 0$*

2.3 Perron–Frobenius Operator

The following exposition on the topic of the Perron-Frobenius operator follows that of A. Boyarsky and G. Góra in [2]. They start by introducing the motivation for the PF operator from the point of view of transformations on stochastic variables, and its consequence on associated probability density functions. We shall follow in their footsteps. Let X be a random variable on an interval $[a, b]$ with a given probability density function f. Then the probability of any event A happening is

$$\mathbb{P}\{X \in A\} = \int_A f d\lambda$$

where $d\lambda$ denotes the Lebesgue measure, normalized such that it is a probability measure. A natural path of inquiry is the following. If one transforms the random variable X trough some mapping $T : I \to I$, then what is the probability density of $T(X)$? To deduce this we start by considering again the probability of an event A.

$$\mathbb{P}\{T(X) \in A\} = \mathbb{P}(X \in T^{-1}A) = \int_{T^{-1}A} f d\lambda.$$

So what we need is to establish the existence of a function ϕ such that for any measurable set A, we have that

$$\int_{T^{-1}A} f d\lambda = \int_A \phi d\lambda.$$

We begin by defining a new measure μ by

$$\mu(A) = \int_{T^{-1}A} f d\lambda,$$

where T is assumed to be non-singular. We will now argue that μ must be absolutely continuous with respect to the Lebesgue measure. Then the Radon–Nikodym theorem 2.3 gives the existence of a unique L^1 function ϕ such that $\mu(A) = \int_A \phi \, d\lambda$ for all $A \in \mathcal{B}$. This will in turn justify Definition 2.7 defining the Perron–Frobenius operator $\mathcal{L}_T : L^1 \to L^1$ by

$$\mathcal{L}_T f = \phi$$

where ϕ is the density given by Theorem 2.3 where ϕ fulfills

$$\int_{T^{-1}A} f d\lambda = \int_A f \circ T d\lambda = \int_A \mathcal{L}_T f d\lambda = \int_A \phi \, d\lambda.$$

Lemma 2.6. *The measure μ defined by*

$$\mu(A) = \int_{T^{-1}A} f d\lambda$$

is absolutely continues with respect to the Lebesgue measure λ for any non-singular T.

Proof. Let A be any λ-measurable set with $\lambda(A) = 0$. Then

$$\mu(A) = \int_{T^{-1}A} f d\lambda$$

but $\lambda(T^{-1}A) = 0$ since T is non-singular by assumption. Hence $\mu(A) = 0$. $\square$

The existence of a function ϕ such that $\mathcal{L}_T f = \phi$ now follows from the Radon–Nikodym Theorem 2.3. Note that ϕ is unique up to almost everywhere equivalence, and hence $\mathcal{L}_T$ is well defined as an operator from L^1 to L^1.

Definition 2.7 (Perron–Frobenius Operator)**.** *Let $I = [a, b]$, $\mathcal{B}$ be the Borel σ-algebra of subsets of I and let λ be the normalized Lebesgue measure on I. Let $T : I \to I$ be a non-singular transformation. We define the Perron–Frobenius operator $\mathcal{L}_T : L^1 \to L^1$ as follows: For any $f \in L^1$, $\mathcal{L}_T f$ is the unique (up to a.e. equivalence) function in L^1 such that*

$$\int_A \mathcal{L}_T f d\lambda = \int_{T^{-1}A} f d\lambda$$

for any $A \in \mathcal{B}$

We shall prove some properties of the operator.

Proposition 2.8. *The operator $\mathcal{L}_T$ is linear*

Proof. From the definition of $\mathcal{L}_T$ and linearity of the integral we get immediately

$$\int_A \mathcal{L}_T(\alpha f + \beta g) d\lambda = \int_{T^{-1}A} \alpha f + \beta g d\lambda = \alpha \int_{T^{-1}A} f d\lambda + \beta \int_{T^{-1}A} g d\lambda = \int_A \alpha \mathcal{L}_T f d\lambda + \int_A \beta \mathcal{L}_T g d\lambda$$

which holds for all $A \in \mathcal{B}$, meaning that

$$\mathcal{L}_T(\alpha f + \beta g) = \alpha \mathcal{L}_T f + \beta \mathcal{L}_T g$$

holds λ-a.e. □

Proposition 2.9. *If $f \in L^1$ is non-negative a.e., then so is $\mathcal{L}_T f$.*

Proof. Simply using the definition of $\mathcal{L}_T f$

$$\int_A \mathcal{L}_T f d\lambda = \int_{T^{-1}A} f d\lambda \geq 0$$

for all $A \in \mathcal{B}$. Hence $\mathcal{L}_T f$ is non-negative a.e. □

'Note that one can interchange non-negative with positive and the above proposition still holds. Simply use that $T^{-1}A$ will have non-zero measure as long as $\lambda(A) \neq 0$.

Proposition 2.10. *For any $f \in L^1$, $\mathcal{L}_T$ preserves the integral over the entire space I. Meaning*

$$\int_I \mathcal{L}_T f d\lambda = \int_I f d\lambda$$

Proof. Trivially

$$\int_I \mathcal{L}_T f d\lambda = \int_{T^{-1}I} f d\lambda = \int_I f d\lambda$$

where we use that $T^{-1}I = I$. □

Proposition 2.11. *The Perron–Frobenius operator $\mathcal{L}_T : L^1 \longrightarrow L^1$ is a (weak) contraction*

$$\|\mathcal{L}_T f\|_{L^1} \leq \|f\|_{L^1}.$$

Proof. Take any $f \in L^1$ and write it as the difference of two non-negative functions $f = f^+ - f^-$. Then $|f| = |f^+ - f^-| = f^+ + f^-$ with complementary support. For $|\mathcal{L}_T f|$ use Proposition 2.9 to get

$$|\mathcal{L}_T f| = |\mathcal{L}_T f^+ - \mathcal{L}_T f^-| \leq \mathcal{L}_T f^+ + \mathcal{L}_T f^- = \mathcal{L}_T(f^+ + f^-) = \mathcal{L}_T|f|$$

so one can now estimate the L^1 norm by

$$\|\mathcal{L}_T f\|_1 = \int_I |\mathcal{L}_T f| d\lambda \leq \int_I \mathcal{L}_T |f| d\lambda = \int_I |f| d\lambda = \|f\|_1.$$

using that $\mathcal{L}_T$ preserves integrals over I. □

We wish to establish a relationship between the Perron-Frobenius operator associated with the n-th iterate of a map T^n, and the one associated with T. What about compositions of different maps?

Proposition 2.12. *Let $T : I \to I$ and $S : I \to I$ be non-singular maps. Then $\mathcal{L}_{T \circ S} f = \mathcal{L}_T \mathcal{L}_S f$ a.e..*

Proof. Let $f \in L^1$ and define the measure μ by

$$\mu(A) = \int_{(T \circ S)^{-1} A} f d\lambda$$

where by the dual assumption that both T and S are non-singular, we get that μ is absolutely continuous with respect to the Lebesgue measure λ. Hence by Theorem 2.3 there exists a L^1 function, $\mathcal{L}_{T \circ S} f$ such that

$$\mu(A) = \int_A \mathcal{L}_{T \circ S} f d\lambda = \int\limits_{S^{-1}(T^{-1}A)} f d\lambda.$$

However we also have

$$\int_A \mathcal{L}_T(\mathcal{L}_S) f d\lambda = \int_{T^{-1}A} \mathcal{L}_S f d\lambda = \int\limits_{S^{-1}(T^{-1}A)} f d\lambda$$

which holds for every $A \in \mathcal{B}$, and therefore $\mathcal{L}_{T \circ S} f = \mathcal{L}_T(\mathcal{L}_S) f$ a.e. One immediately gets $\mathcal{L}_T^n f = \mathcal{L}_{T^n} f$ a.e. □

Definition 2.13. *We define the so called Koopman operator* $U_T : L^\infty \to L^\infty$ *by*

$$U_T f = f \circ T.$$

Proposition 2.14. *The Koopman operator* U_T *is the adjoint operator of* $\mathcal{L}_T$*, meaning that*

$$\int_I (\mathcal{L}_T f) \cdot g d\lambda = \int_I f \cdot (U_T g) d\lambda.$$

Proof. We first prove the result for characteristic functions and extend the result to all $g \in L^\infty$ by density. Choose A as any measurable subset of I and define $g = \chi_A$. We then wish to prove that

$$\int_A \mathcal{L}_T f d\lambda = \int_I f \cdot \chi_A \circ T d\lambda \tag{2.3}$$

but the fact that this is true becomes clear when writing the right hand side as

$$\int_I f \cdot \chi_A \circ T d\lambda = \int_{T^{-1}A} f d\lambda$$

which is equal the to left hand side of (2.3) by the definition of the Perron–Frobenius operator. Equality for the general $g \in L^\infty$ follows directly from the density of simple functions in L^∞. □

Proposition 2.14 will be of vital importance in the study of decay of correlations. In this field, the object of study will be the following difference of integrals:

$$C(f, g, n) = \left| \int_I f \cdot g \circ T^n d\mu - \int_I f d\mu \int_I g d\mu \right| \tag{2.4}$$

where μ is measure preserving. The goal is to understand how fast this quantity goes to zero, which measures how quickly our system forgets its initial distribution. The approach that shall be taken in this thesis lives and dies with the fact that the Koopman operator is adjoint to the Perron–Frobenius operator, meaning we can instead study

$$C(f, g, n) = \left| \int_I \mathcal{L}_{T,\mu}^n f \cdot g d\mu - \int_I f d\mu \int_I g d\mu \right| \tag{2.5}$$

where $\mathcal{L}_{T,\mu}$ is the Perron–Frobenius operator associated to T under the measure μ. However, it is at this time not clear why this is preferable to the original formulation. This confusion shall be alleviated when we show that in many circumstances, one can write down workable explicit expressions for $\mathcal{L}_T$. The following representation is due to Mayer in [8], which will be suitable for our purposes.

Lemma 2.15. *Let $T : I \to I$ be a transformation. Assume that there exists a countable collection of intervals $I_l, l \in \mathcal{F}$ such that $\bigcup_{l\in\mathcal{F}} I_l = I$ and the different I_l have disjoint interior. Also assume that $T_l \coloneqq T|_{I_l}$ is monotone and C^k for some $k \geq 1$. Then*

$$\mathcal{L}_T f = \sum_{l\in\mathcal{F}} |\psi_l'(x)| f \circ \psi_l(x) \chi_{T(I_l)}(x)$$

where ψ_l is the local inverse of the map T restricted to the interval I_l.

Proof. By the definition of the PF operator we have

$$\int_A \mathcal{L}_T f d\lambda = \int_{T^{-1}A} f d\lambda. \tag{2.6}$$

Using that T_l is invertible, we define $T_l^{-1} = \psi_l$ (if $x \notin I_l$ we will say $\psi_l(x) = 0$). Then the right hand side of (2.6) can be written as

$$\int_{T^{-1}A} f d\lambda = \int_{\bigcup_{l\in\mathcal{F}} \psi_l(A\cap I_l)} f d\lambda = \sum_{l\in\mathcal{F}} \int_{\psi_l(A\cap I_l)} f d\lambda$$

where the second equality comes from that all $\psi_l(A \cap I_l)$ are mutually disjoint. The individual terms can now be rewritten using a change of variables to get

$$\int_{\psi_l(A\cup I_l)} f d\lambda = \int_A f \circ \psi_l(x) |\psi_l'(x)| \chi_{TI_l}(x) d\lambda$$

giving that

$$\int_A \mathcal{L}_T f d\lambda = \int_A \sum_{l\in\mathcal{F}} f \circ \psi_l(x) |\psi_l'(x)| \chi_{TI_l}(x) d\lambda = \int_A \sum_{l\in\mathcal{F}} \frac{f \circ \psi_l(x)}{|T_l' \circ \psi_l(x)|} \chi_{TI_l}(x) d\lambda$$

which holds for all $A \in \mathscr{B}$ giving that the integrands are equal a.e. □

An immediate corollary is that if each $T_l : I_l \to I$ is surjective (except at perhaps the endpoints of the interval), then we get the following.

Corollary 2.15.1. *If each T_l is onto, then*

$$\mathcal{L}_T f = \sum_{l\in\mathcal{F}} \frac{f \circ \psi_l(x)}{|T_l' \circ \psi_l(x)|}.$$

Hereafter, we will be working under the assumption that T is piecewise surjective. Hence this is the expression which shall be used from now on.
In (2.4) and (2.5) the correlation is formulated in terms of the measure μ. This is in some sense the most natural formulation, but it is simultaneously somewhat cumbersome. Since $d\mu = \phi d\lambda$, we can write $\tilde{f} = f \cdot \phi$ and transform (2.4) to

$$\begin{aligned} C(f, g, n) &= \left| \int_I \tilde{f} \cdot g \circ T^n d\lambda - \int_I \tilde{f} d\lambda \int_I g d\mu \right| \\ &= \left| \int_I \mathcal{L}_T^n \tilde{f} \cdot g d\lambda - \int_I \tilde{f} d\lambda \int_I g d\mu \right| \end{aligned} \tag{2.7}$$

so one can use the established formulation in section 2.15.1 which is only valid for the Lebesgue measure.

2.4 The Gauss map and continued fractions

The most natural context in which to introduce the Gauss map is when discussing regular continued fractions. In the work of A. Ya. Khinchin [3], it is established that any positive real number x can be written as a continued fraction

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\ddots}}}}, \quad a_n \in \mathbb{Z}^+ \tag{2.8}$$

where for rational x the expansion terminates after a finite number of coefficients, and for irrational x the expansion is infinite. The expression in (2.8) is often written more compactly as

$$x = [a_0; a_1, a_2, a_3, \dots, a_n]$$

for rational x and

$$x = [a_0; a_1, a_2, a_3, \dots]$$

if x is irrational. In most contexts we are not interested in the integer part of x, so one simply assumes that $a_0 = 0$ and write

$$x = [0; a_1, a_2, a_3, \dots] = [a_1, a_2, a_3, \dots].$$

Now the natural question is the following: given an arbitrary $x \in [0, 1]$, how does one find its corresponding expansion? If one wishes to find a_1, simply take the reciprocal and round down. This motivates the definition of the *Gauss map*.

Definition 2.16. *We define the Gauss map $T : \mathbb{R} \to [0, 1]$ by*

$$T(x) = \begin{cases} 0, & \textit{if } x = 0 \\ \frac{1}{x} \mod 1, & \textit{otherwise} \end{cases}$$

Using the mapping we can write

$$T([a_1, a_2, a_3, \dots]) = a_1 + [a_2, a_3, \dots] - a_1 = [a_2, a_3, \dots]$$

so from this point of view, the Gauss map acts as a left shift on the sequence of continued fraction coefficients. Studying the Gauss map as a measure preserving transformation, and ascertaining its dynamical properties, will therefore in turn yield information regarding continued fraction expansions. One then wishes to study the dynamical system $([0, 1], T, \mu)$, where μ is such that $\mu(A) = \mu(T^{-1}A)$ for all Borel-sets A. There are many such measures, and it is well known that

$$\mu(A) = \int_A \frac{1}{\log 2} \cdot \frac{1}{1 + x} d\lambda(x)$$

is one such [3]. It also has the additional property of being absolutely continuous with respect to the Lebesgue measure λ. We call $\phi = \frac{1}{\log 2} \cdot \frac{1}{1+x}$ the *invariant density* of μ and write $d\mu = \phi d\lambda$.

This particular choice of μ is of particular interest to us. Equation (2.5) tells us that studying the decay of correlations for for this system is essentially the same as asking: "how quickly does an arbitrary observable f converge to the invariant density" (this is why we prefer formulation (2.5) over (2.4)). In full generality it is of course not possible to answer such a question, but given some restriction on the observable we can make progress.

3 Decay of Correlations for Differentiable Functions

We shall follow the program established by Liverani and attempt to apply the methods presented for expanding maps on finite partitions of $[0, 1]$. The central object of study is the Perron–Frobenius operator, for which the general form is

$$\mathcal{L}_T f(x) = \sum_{y \in T^{-1}(x)} f(y)|T'(y)|^{-1}$$

but using Corollary 2.15.1 and choosing T to be the Gauss map, we acquire the nicer expression

$$\mathcal{L}_T f(x) = \sum_{n=1}^{\infty} \frac{1}{(x+n)^2} f\left(\frac{1}{x+n}\right).$$

The first interesting step of the program is to find a cone $\mathscr{C}$ of functions that the Perron–Frobenius operator maps strictly inside itself.

3.1 Establishing an Initial Cone

Some general heuristics were kept in mind when searching for reasonable choices of cones. Firstly, the goal was to have few and simple conditions. The expressions quickly become unmanageable when imposing either multiple or complex requirements. Secondly, the conditions must impose some sort of regularity on the admissible functions. Without such a condition, there is no reason to hold any confidence in the image of the transformation $\mathcal{L}_T\mathscr{C}$ having finite diameter. The initial attempt shall be the one parameter family of cones $\mathscr{C}_a$ defined by

$$\mathscr{C}_a = \left\{ f \in C^1[0,1] \middle| f \not\equiv 0 \, , f \geq 0, \; 0 \leq -f' \leq af \right\}$$

where the cones are thought of as subsets of the Banach space of $C^1[0, 1]$ functions. One initial observation, which gives credence to this being a reasonable choice, is that the invariant density lies in this cone as long as a is greater than or equal to 1. This observation is of course a luxury. In a general setting there is no expectation of being a priori able to verify that the invariant density lies in the cone. However, investigation shows that for the Perron–Frobenius operator associated to the Gauss map, if one uses crude estimates, the only choice of parameter that can be shown to be admissible is $a = 2$. To be clear, other choices of a may also leave $\mathscr{C}_a$ invariant, but we shall not spend more time on this. Hence we get the choice

$$\mathscr{C}_2 = \mathscr{C} = \left\{ f \in C^1[0,1] \middle| f \not\equiv 0, \; f \geq 0, \; 0 \leq -f' \leq 2f \right\}$$

where we have suppressed the parameter index. We show that this choice of cone is invariant.

Lemma 3.1. *The cone $\mathscr{C}$ as defined above is invariant under $\mathcal{L}_T$, meaning $\mathcal{L}_T\mathscr{C} \subset \mathscr{C}$.*

Proof. We must show that for any function $f \in \mathscr{C}$ we have $\mathcal{L}_T f \in \mathscr{C}$. That $\mathcal{L}_T f \not\equiv 0$ and $\mathcal{L}_T f \geq 0$ follow immediately from the expression for $\mathcal{L}_T f$. Continuity follows from that $\mathcal{L}_T f$ can be written as a series of continuous functions that converges uniformly on $[0, 1]$. Furthermore, differentiating the series term by term gives

$$(\mathcal{L}_T f)' = \sum_{n=1}^{\infty} \frac{-f'\left(\frac{1}{x+n}\right) - 2(x+n)f\left(\frac{1}{x+n}\right)}{(x+n)^4}$$

and this series converges uniformly by the Weierstrass M-test. Hence $(\mathcal{L}_T f)'$ exists and is continuous. What remains is showing that

$$(\mathcal{L}_T f)' \leq 0 \text{ and } -(\mathcal{L}_T f)' \leq 2\mathcal{L}_T f, \tag{3.1}$$

both of which follow from simple calculations. Starting with the leftmost inequality

$$\begin{aligned}(\mathcal{L}_T f)' &= \sum_{n=1}^{\infty} \frac{-f'\left(\frac{1}{x+n}\right) - 2(x+n)f\left(\frac{1}{x+n}\right)}{(x+n)^4} \\ &\leq \sum_{n=1}^{\infty} \frac{2f\left(\frac{1}{x+n}\right) - 2(x+n)f\left(\frac{1}{x+n}\right)}{(x+n)^4} \leq \sum_{n=1}^{\infty} 2f\left(\frac{1}{x+n}\right)\frac{1-(x+n)}{(x+n)^4} \leq 0\end{aligned}$$

where we have used that $f \geq 0$ and $-f' \leq 2f$. Now a final computation treating the rightmost inequality in (3.1) shows the lemma. We proceed with

$$\begin{aligned}-(\mathcal{L}_T f)' &= \sum_{n=1}^{\infty} \frac{f'\left(\frac{1}{x+n}\right) + 2(x+n)f\left(\frac{1}{x+n}\right)}{(x+n)^4} \\ &\leq \sum_{n=1}^{\infty} \frac{2(x+n)}{(x+n)^4} f\left(\frac{1}{x+n}\right) \leq \sum_{n=1}^{\infty} \frac{2}{(x+n)^2} f\left(\frac{1}{x+n}\right) = 2\mathcal{L}_T f\end{aligned}$$

which uses only that $f \geq 0$ and $f' \leq 0$. □

Now that $\mathscr{C}$ has been shown to be invariant under $\mathcal{L}_T$, we shall proceed to show that the image $\mathcal{L}_T\mathscr{C}$ has finite diameter under the induced Hilbert metric $\Theta(f, g)$, as defined in equation (2.2). It is clear that in order to bound the diameter, one must find upper and lower bounds for β and α respectively. A useful trick in this endeavor employed by Liverani in [7] is to note that

$$\text{diam}\mathcal{L}_T\mathscr{C} = \sup_{f,g\in\mathscr{C}} \Theta(\mathcal{L}_T f, \mathcal{L}_T g) \leq \sup_{f,g\in\mathscr{C}} \leq \Theta(\mathcal{L}_T f, h) + \Theta(h, \mathcal{L}_T g) \leq 2 \sup_{f\in\mathscr{C}} \Theta(\mathcal{L}_T f, h)$$

where h is fixed, so our analysis becomes much more manageable. Choosing $h = \frac{1}{1+x}$ which is the (scaled) invariant density reduces our problem to finding suitable expressions to control $\alpha\big(\mathcal{L}_T f, (1+x)^{-1}\big)$ and $\beta\big(\mathcal{L}_T f, (1+x)^{-1}\big)$. We shall start by deriving expressions for α and β for general $f, g \in \mathscr{C}$. We write explicitly how the parts of the expression for α is found. By definition $\alpha(f, g)$ is the supremum over all λ that satisfy $\lambda f \preceq g$, where $\preceq$ is the order relation induced by inclusion in $\mathscr{C}$. Therefore, the condition $\lambda f \preceq g$ on λ is equivalent to

$$0 \leq g - \lambda f, \quad 0 \leq -(g' - \lambda f'), \text{ and } -(g' - \lambda f') \leq 2(g - \lambda f)$$

where solving the system of inequalities leads to

$$\alpha(f, g) = \min\left\{\inf \frac{g}{f}, \inf \frac{g'}{f'}, \inf \frac{2g + g'}{2f + f'}\right\}$$

and a similar procedure for β gives

$$\beta(f,g) = \max\left\{\sup \frac{g}{f}, \sup \frac{g'}{f'}, \sup \frac{2g+g'}{2f+f'}\right\}.$$

We pause and reiterate that these expressions for α and β hold for our specific choice of cone $\mathscr{C}$. A different choice does not only give a different object to bound the diameter of, but also the way in which we measure distances. Now we evaluate these expressions for $\mathcal{L}_T g$ and the invariant density ϕ, and think of the three arguments as functions of g, denoting them by α_i and β_i respectively. Since Θ is a projective metric, the distance is invariant under scaling. Therefore we use $\frac{1}{1+x}$ over ϕ for simplicity and get

$$\alpha\left(\frac{1}{1+x}, \mathcal{L}_T g\right) = \min\{\alpha_1(g), \alpha_2(g), \alpha_3(g)\}$$
$$\beta\left(\frac{1}{1+x}, \mathcal{L}_T g\right) = \max\{\beta_1(g), \beta_1(g), \beta_1(g)\}.$$

To control the $\alpha_i(g)$ we shall make use that $e^{-2x}g(0) \leq g(x)$, which one gets by solving the differential inequality $-g' \leq 2g$. Now, since $\mathscr{C}$ is a cone, we can without loss of generality assume $g(0) = 1$. Starting with $\alpha_1(g)$

$$\alpha_1(g) = \inf(1+x)\sum_{n=1}^{\infty}\frac{1}{(x+n)^2}g\left(\frac{1}{x+n}\right) \geq \inf\sum_{n=1}^{\infty}\frac{g(0)(1+x)\exp\left(\frac{-2}{x+n}\right)}{(x+n)^2}$$

and if one differentiates this series term by term, one discovers that each term is increasing. Hence we can bound the series by evaluating at $x = 0$ and remembering that $g(0) = 1$ gives

$$\alpha_1(g) \geq \sum_{n=1}^{\infty}\frac{\exp\left(\frac{-2}{n}\right)}{n^2} \geq \int_1^{\infty}\frac{\exp\left(\frac{-2}{x}\right)}{x^2}dx = \int_0^1 e^{-2u}du = \frac{1}{2}(1-e^{-2}) \geq 0.43.$$

We proceed with $\alpha_2(g)$

$$\begin{aligned}\alpha_2(g) &= \inf(1+x)^2\sum_{n=1}^{\infty}\frac{g'(\frac{1}{x+n}) + 2(x+n)g(\frac{1}{x+n})}{(x+n)^4} \geq \inf(1+x)^2\sum_{n=1}^{\infty}\frac{-2g+2(x+n)g}{(x+n)^4}\\ &\geq 2g(0)e^{-2}\sum_{n=1}^{\infty}\frac{(x+n-1)(1+x)^2}{(x+n)^4} \geq 2e^{-2}\sum_{n=1}^{\infty}\frac{n-1}{n^4}\end{aligned}$$

where in the last step we use that the series is increasing by Lemma A.2 and that $g(0) = 1$. Hence we get that $\alpha_2(g) \geq 2e^{-2}(\zeta(3) - \pi^4/90) \geq 0.032$. Ending the arduous process of producing lower bounds, we handle $\alpha_3(g)$ as such

$$\begin{aligned}\alpha_3(g) &= \inf(1+x)^2\frac{2\mathcal{L}_T g + (\mathcal{L}_T g)'}{2(1+x)-1}\\ &\geq \inf\left[\frac{(x+1)^2}{2(x+1)-1}\right]\cdot\left[\sum_{n=1}^{\infty}\frac{2(x+n)^2 g}{(x+n)^4} + \sum_{n=1}^{\infty}\frac{-g' - 2(x+n)g}{(x+n)^4}\right]\\ &\geq \sum_{n=1}^{\infty}\frac{2(x+n)^2 g - 2(x+n)g}{(x+n)^4} = \sum_{n=1}^{\infty}\frac{2(x+n)(x+n-1)g}{(x+n)^4}\\ &\geq 2e^{-2}g(0)\sum_{n=2}^{\infty}\frac{(x+n)(x+n-1)}{(x+n)^4} \geq 2e^{-2}\sum_{n=1}^{\infty}\frac{n(n-1)}{(n+1)^4} = 2e^{-2}\left(\frac{\pi^2}{6} + \frac{\pi^4}{45} - 3\zeta(3)\right)\end{aligned}$$

so we are satisfied with finally getting $\alpha_3(g) \geq 0.055$. Summarising the results for each α_i

$$\alpha\left(\frac{1}{1+x}, \mathcal{L}_T g\right) \geq \min\{0.43, 0.032, 0.055\} = 0.032.$$

Despite being repetitive, and labour intensive we shall venture forward and find upper bounds for β_i as such:

$$\beta_1(g) = \sup(x+1)\mathcal{L}_T g \leq \sup \sum_{n=1}^{\infty} \frac{1+x}{(x+n)^2} g\left(\frac{1}{x+n}\right) \leq g(0) \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

where we have used Lemma A.1 when evaluating the sum.

$$\begin{aligned}\beta_2(g) = \sup -(x+1)^2(\mathcal{L}_T g)' &= \sup(x+1)^2 \sum_{n=1}^{\infty} \frac{g'(\frac{1}{x+n}) + 2(x+n)g(\frac{1}{x+n})}{(x+n)^4} \\ &\leq \sup \sum_{n=1}^{\infty} \frac{2(x+n)(x+1)^2}{(x+n)^4} = \sup \sum_{n=1}^{\infty} \frac{2(x+1)^2}{(x+n)^3}\end{aligned}$$

using that g' is negative for all x in $[0,1]$ we get the estimate. To estimate the last term we use Lemma A.3 and evaluate at $x = 0$, giving

$$\beta_2(g) \leq \sum_{n=1}^{\infty} \frac{2}{n^3} \leq 2.41.$$

The final estimate is that of β_3 which is controlled by

$$\begin{aligned}\beta_3(g) &= \sup\, (x+1)^2 \frac{2\mathcal{L}_T g + (\mathcal{L}_T g)'}{2(x+1)-1} \\ &\leq \sup \frac{4}{3} \sum_{n=1}^{\infty} 2g\left(\frac{1}{x+n}\right) \frac{(x+n)^2 - (x+n) + 1}{(x+n)^4} \leq \sup \frac{8}{3} \sum_{n=1}^{\infty} \frac{(x+n)^2 - (x+n) + 1}{(x+n)^4}\end{aligned}$$

where one first uses that $(x+1)^2/(2(x+1)-1) \leq \frac{4}{3}$ and then that each term in the series is decreasing. We evaluate at zero to get

$$\beta_3(\mathcal{L}_T g) \leq \frac{8}{3}(\zeta(2) - \zeta(3) + \zeta(4)) \leq 4.1$$

Summarising gives an upper bound for β in the form of

$$\beta\left(\frac{1}{1+x}, \mathcal{L}_T g\right) \leq \max\left\{\frac{\pi^2}{6}, 2.41, 4.1\right\} = 4.1.$$

We can from this give a finite upper bound on the diameter of the image of $\mathscr{C}$ by

$$\operatorname{diam}(\mathcal{L}_T \mathscr{C}) \leq 2 \sup_{g \in \mathscr{C}} \Theta\left(\frac{1}{1+x}, \mathcal{L}_T g\right) = 2 \sup_{g \in \mathscr{C}} \log \frac{\beta\left(\frac{1}{1+x}, \mathcal{L}_T g\right)}{\alpha\left(\frac{1}{1+x}, \mathcal{L}_T g\right)} \leq 2 \log \frac{4.1}{0.032} = \Delta$$

where $\Delta \approx 9.70601$ or less than 9.71. Now making use of Theorem 2.2 where in our case $\Theta_1 = \Theta_2$, showing the strict contraction in the Hilbert metric. Now setting $\tanh\left(\frac{\Delta}{4}\right) = \Lambda \approx 0.9845 < 1$ we get

$$\Theta\left(\mathcal{L}_T f, \mathcal{L}_T g\right) \leq \tanh\left(\frac{\Delta}{4}\right) \Theta(f, g) = \Lambda \cdot \Theta(f, g).$$

From this point onward, we shall show how to apply this result in order to bound (2.7)

$$C(f,g,n) \coloneqq \left| \int_0^1 f \cdot g \circ T^n - \int_0^1 f d\mu \int_0^1 g d\mu \right| = \left| \int_0^1 \mathcal{L}_T^n f \cdot g - \int_0^1 f d\mu \int_0^1 g d\mu \right|$$

assume wlog that $\int_0^1 f d\lambda = 1$. Then using that $d\mu = \phi dx$ the above can be written as

$$C(f,g,n) = \left| \int_0^1 g(\mathcal{L}_T^n f - \phi) dx \right|$$

which can be trivially bounded by

$$\left| \int_0^1 g(\mathcal{L}_T^n f - \phi) dx \right| \leq \|g\|_1 \|\mathcal{L}_T^n f - \phi\|_\infty \leq \|g\|_1 \left\| \frac{\mathcal{L}_T^n f}{\phi} - 1 \right\|_\infty \|\phi\|_\infty \tag{3.2}$$

In order to proceed we need to relate the supremum norm in (3.2) to the projective distance between $\mathcal{L}_T^n f$ and ϕ.

Lemma 3.2. *For $f \in \mathscr{C}$ with $\int_0^1 f dx = 1$ and $\phi = \frac{1}{\log 2} \cdot \frac{1}{1+x}$ one has*

$$\left\| \frac{\mathcal{L}_T^n f}{\phi} - 1 \right\|_\infty \leq \exp\left[\Theta_+(\mathcal{L}_T^n f, \phi)\right] - 1 \leq \exp\left[\Theta(\mathcal{L}_T^n f, \phi)\right] - 1.$$

Proof. The proof makes use of a trick. We start by considering $\frac{\mathcal{L}_T^n f(x)}{\phi(x)}$, for which we wish to find a bound in terms of the projective distance Θ_+. Note that Θ_+ corresponds to the cone of non-negative C^1 functions $\mathscr{C}_+$, and Θ without subscript to $\mathscr{C}$. This gives us that $\mathscr{C} \subset \mathscr{C}_+$ where these sets are *not* equal. We start by considering

$$\frac{\mathcal{L}_T^n f(x)}{\phi(x)} = \frac{\mathcal{L}_T^n f(x)}{\phi(x)} \cdot \frac{\mathcal{L}_T^n f(y)}{\phi(y)} \cdot \frac{\phi(y)}{\mathcal{L}_T^n f(y)}$$

and remembering the definition (2.1) of μ and λ gives that

$$\frac{\lambda}{\mu} \cdot \frac{\mathcal{L}_T^n f(y)}{\phi(y)} \leq \frac{\mathcal{L}_T^n f(x)}{\phi(x)} \leq \frac{\mu}{\lambda} \cdot \frac{\mathcal{L}_T^n f(y)}{\phi(y)}.$$

Since this inequality holds for all possible μ and λ, we take the inf over μ and sup over λ to get

$$\exp\left[-\Theta_+(\mathcal{L}_T^n f, \phi)\right] \cdot \frac{\mathcal{L}_T^n f(y)}{\phi(y)} \leq \frac{\mathcal{L}_T^n f(x)}{\phi(x)} \leq \exp\left[\Theta_+(\mathcal{L}_T^n f, \phi)\right] \cdot \frac{\mathcal{L}_T^n f(y)}{\phi(y)} \tag{3.3}$$

which holds for all x, y in $[0,1]$ Since we assumed that $\int_0^1 f dx = 1$ and ϕ is the density of a probability measure, we have $\int_0^1 (\mathcal{L}_T^n f - \phi)\, dx = 0$. This implies that there for each n must exist points x_n, y_n such that $\mathcal{L}_T^n f(x_n) \leq \phi(x_n)$ and $\mathcal{L}_T^n f(y_n) \geq \phi(y_n)$, giving that $\frac{\mathcal{L}_T^n f}{\phi}$ takes values both greater and less than one. Hence (3.3) gives

$$\exp\left[-\Theta_+(\mathcal{L}_T^n f, \phi)\right] \leq \frac{\mathcal{L}_T^n f(x)}{\phi(x)} \leq \exp\left[\Theta_+(\mathcal{L}_T^n f, \phi)\right].$$

which is equivalent to

$$\exp\left[-\Theta_+(\mathcal{L}_T^n f, \phi)\right] - 1 \leq \frac{\mathcal{L}_T^n f(x)}{\phi(x)} - 1 \leq \exp\left[\Theta_+(\mathcal{L}_T^n f, \phi)\right] - 1. \tag{3.4}$$

but we also have that $e^x + e^{-x} \geq 2$ or equivalently, $e^{-x} - 1 \geq 1 - e^x$. Applying to the leftmost inequality of (3.4) gives

$$-(\exp[\Theta_+(\mathcal{L}_T^n f, \phi)] - 1) \leq \frac{\mathcal{L}_T^n f(x)}{\phi(x)} - 1 \leq (\exp[\Theta_+(\mathcal{L}_T^n f, \phi)] - 1)$$

finally giving

$$\left\| \frac{\mathcal{L}_T^n f}{\phi} - 1 \right\|_\infty \leq \exp[\Theta_+(\mathcal{L}_T^n f, \phi)] - 1 \leq \exp[\Theta(\mathcal{L}_T^n f, \phi)] - 1.$$

The second inequality comes from using Theorem 2.2 with P equal to the identity operator from $\mathscr{C}$ to $\mathscr{C}_+$. This proves the lemma. □

Remembering that we defined $\Lambda = \tanh\left(\frac{\Delta}{4}\right)$ together with the lemma gives the bound

$$\left\| \frac{\mathcal{L}_T^n f}{\phi} - 1 \right\|_\infty \leq \exp[\Lambda^n \Theta(f, \phi)] - 1 \leq \exp[\Lambda^n \Delta] - 1 \leq (\Lambda^n \Delta) \cdot \exp[\Lambda^n \Delta]$$

which finally shows the exponential decay of correlations for $f \in \mathscr{C}$

$$C(f, g, n) = \left| \int_0^1 \mathcal{L}_T^n f \cdot g d\lambda - \int_0^1 g d\mu \right| \leq D \|f\|_1 \|g\|_1 \cdot \Lambda^n$$

where D is some constant and $\Lambda \leq 0.985$.

3.2 Extending Result

At this time, the decay of correlations has only been established for $f \in \mathscr{C}$. We shall by some simple arguments extend this result to a larger set of functions.

Lemma 3.3. *We have exponential decay of correlations for Lip([0, 1]) vs $L^1[0,1]$) where the speed of decay is at least Λ^n as in the previous section.*

Proof. Let f be a non-negative decreasing C^1 function. If $f \in \mathscr{C}$ then we are done. Hence, assume that f does not lie in our cone. Then we must have that $-f' > 2f$ at some point in $[0, 1]$. Since f is continuously differentiable, f' is bounded by $\|f'\|_\infty < \infty$. Then we write $f = f + \frac{1}{2}\|f'\|_\infty - \frac{1}{2}\|f'\|_\infty$. Now both $f + \frac{1}{2}\|f'\|_\infty$ and $\frac{1}{2}\|f'\|_\infty$ lie in $\mathscr{C}$ so for a general positive $h \in C^1$ function, we get by the triangle inequality

$$\begin{aligned} C(h, g, n) &= C(h + \frac{1}{2}\|h'\|_\infty - \frac{1}{2}\|h'\|_\infty, g, n) \\ &\leq C(h + \frac{1}{2}\|h'\|_\infty, g, n) + C(\frac{1}{2}\|h'\|_\infty, g, n) \\ &\leq D(\|h\|_1 + \frac{1}{2}\|h'\|_\infty)\|g\|_\infty \Lambda^n + D\frac{1}{2}\|h'\|_\infty \|g\|_\infty \Lambda^n \\ &= D(\|h\|_1 + \|h'\|_\infty)\|g\|_\infty \Lambda^n \end{aligned} \tag{3.5}$$

For an arbitrary non-negative $f \in C^1$, not necessarily decreasing, we can simply decompose f as

$$f = (f + \|f'\|_\infty - \|f'\|_\infty x) - (\|f'\|_\infty - \|f'\|_\infty x)$$

and the exponential decay of correlations follows much the same in this case as in (3.5). Now we will make use of the fact that any Lipschitz function f_L can be approximated arbitrarily well in L^1 norm by a C^1 function f, while simultaneously ensuring that $\|f'\|_\infty \leq \mathrm{Lip}(f_L)$, where $\mathrm{Lip}(f_L)$ is the Lipschitz constant of f_L. Using this we can bound the correlation decay of a non-negative $f_L \in \mathrm{Lip}([0,1])$ by

$$\begin{aligned} C(f_L - f + f, g, n) &\leq C(f_L - f, g, n) + C(f, g, n) \\ &\leq \|f_L - f\|_\infty \|g\|_1 + \|f_L - f\|_\infty \|\phi\|_\infty^2 + D(\|f\|_\infty + \|f'\|_\infty) \|g\|_1 \Lambda^n \\ &\leq \varepsilon(\|g\|_1 + \|\phi\|_\infty^2) + D(\|f_L\|_\infty + \mathrm{Lip}(f_L)) \|g\|_1 \Lambda^n. \end{aligned}$$

Taking the infimum over all epsilon gives that

$$C(f_L, g, n) \leq D(\|f_L\|_\infty + \mathrm{Lip}(f_L)) \|g\|_1 \Lambda^n = D(\|f_L\|_L) \|g\|_1 \Lambda^n$$

for all $0 \leq f_L \in \mathrm{Lip}([0,1])$ and $g \in L^\infty([0,1])$. Since every functions f in $\mathrm{Lip}([0,1])$ can be written as the difference of non-negative Lipschitz functions $f = f^+ - f^-$, with $\mathrm{Lip}(f^+), \mathrm{Lip}(f^-) \leq \mathrm{Lip}(f)$. Using this we immediately get

$$\begin{aligned} C(f_L, g, n) &\leq C(f^+, g, n) + C(f^-, g, n) \\ &\leq D\left(\|f^+\|_\infty + \mathrm{Lip}(f^+)\right) \|g\|_1 \Lambda^n + D\left(\|f^-\|_\infty + \mathrm{Lip}(f^-)\right) \|g\|_1 \Lambda^n \\ &\leq D\left(\|f^+\|_\infty + \|f^-\|_\infty + \mathrm{Lip}(f)\right) \|g\|_1 \Lambda^n \leq 2D \|f\|_L \|g\|_1 \Lambda^n. \end{aligned}$$

This extends the exponential rate of decay for the correlation to all $\mathrm{Lip}([0,1])$ functions. □

The established program has now run it's course and no more will be said regarding the cone $\mathscr{C}$ or Lipschitz functions. However, before we move on we make a comment. The achievement of this section is more than anything a proof of concept. It confirms that the methods work in the context of the Gauss map. While this will not be surprising to those familiar with the field, it is nice to have it confirmed. Not yet satisfied, we continue onward.

4 Decay of Correlations for Functions of Bounded Variation

In section 3 we pose highly restrictive requirements on the cone of functions considered, imposing both a requirement on being decreasing and continuously differentiable, but also a limitation on the rate of decrease. This forces us to, in some sense, start with observables qualitatively similar to the limiting distribution ϕ. This regularity has the effect of giving a fast decay of correlation for $\mathrm{Lip}[(0,1)]$ vs $L^1[(0,1)]$. A natural question to ask is then, if we expand the admissible class of function, what will happen to the decay of correlations? What can be said is that there is in essence a trade-off between how general of a result can be achieved, and how fast the correlation will decay. In the upcoming section, we shall consider functions of bounded variation $BV[(0,1)]$.

4.1 The Gauss-map

In appendix B we define the notion of an expanding map and we wish to show that the Gauss map fits into this category. The first three properties are trivial, and as such we shall focus on showing that for some n one has $|T^n(x)| \geq \lambda > 1$ for all $x \in [0,1]$. It will turn out that $n = 2$ will suffice, and it will be worthwhile to give this operator its own. We define $P \coloneqq T \circ T = T^2$.

Lemma 4.1. *The square map $P : [0,1] \to [0,1]$ fulfills*

$$|P'(x)| \geq 4$$

for all $x \in [0,1]$ where the derivative exists.

Proof. The Gauss-map can be written as

$$T(x) = \frac{1}{x} \pmod 1 = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$$

where $\lfloor \cdot \rfloor$ is the floor function. We now write the composition $T \circ T(x)$ as

$$T^2(x) = \frac{1}{T(x)} - \left\lfloor \frac{1}{T(x)} \right\rfloor = \frac{1}{\frac{1}{x} - \lfloor \frac{1}{x} \rfloor} - \left\lfloor \frac{1}{\frac{1}{x} - \lfloor \frac{1}{x} \rfloor} \right\rfloor .$$

We first show that the absolute value of the derivative of the square map is greater than one where it exists. Differentiating gives

$$(T^2)'(x) = \left(\frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \right)' \frac{-1}{(\frac{1}{x} - \lfloor \frac{1}{x} \rfloor)^2} = \frac{-1}{x^2} \frac{-1}{(\frac{1}{x} - \lfloor \frac{1}{x} \rfloor)^2} = \frac{1}{(1 - x \lfloor \frac{1}{x} \rfloor)^2}.$$

Since this expression is non-negative, we wish to prove that

$$(T^2)'(x) = \frac{1}{(1 - x \lfloor \frac{1}{x} \rfloor)^2} \geq \lambda$$

or equivalently

$$(1 - x \left\lfloor \frac{1}{x} \right\rfloor)^2 \leq \lambda^{-1}.$$

Due to $\lfloor \frac{1}{x} \rfloor \leq \frac{1}{x}$ we can conclude that $1 - x \lfloor \frac{1}{x} \rfloor \geq 0$. Now using that $x \lfloor \frac{1}{x} \rfloor \geq x(\frac{1}{x} - 1)$ we get that

$$\left(1 - x \left\lfloor \frac{1}{x} \right\rfloor\right) \leq \left(1 - x \left(\frac{1}{x} - 1\right)\right) = x,$$

which on the interval $[0, 0.5]$ is less than 0.5, giving that

$$0.5^{-1} \leq \frac{1}{(1 - x \lfloor \frac{1}{x} \rfloor)}, \ x \in [0, 0.5] \iff 4 \leq \frac{1}{(1 - x \lfloor \frac{1}{x} \rfloor)^2}, \ x \in [0, 0.5]$$

Furthermore, on $x \in (0.5, 1]$ we have $\lfloor \frac{1}{x} \rfloor = 1$, so

$$\left(1 - x \left\lfloor \frac{1}{x} \right\rfloor\right) = 1 - x \leq 0.5$$

again giving the lower bound of 4. Hence the square of the Gauss map satisfies $(P)'(x) \geq 4$ for all $x \in [0, 1]$ such that the derivative exists. $\square$

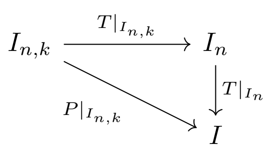
We wish to find the smallest partition such that P is monotone and continuous on each partition element. It suffices to find the points of discontinuity of P. The points of discontinuity and hence the points defining the partition that T is monotone on are the reciprocals of the natural numbers $\frac{1}{n}, n \in \mathbb{N}^+$. These can be characterised by the solutions to $\lfloor \frac{1}{x} \rfloor = 0$. This collection makes up part of the critical points of the square map T^2. The rest can be found by studying

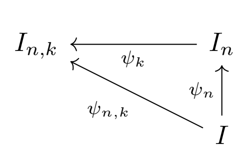
$$\left\lfloor \frac{1}{\frac{1}{x} - \lfloor \frac{1}{x} \rfloor} \right\rfloor = 0$$

or more easily

$$\frac{1}{\frac{1}{x}-n} = k \implies x = \frac{1}{n+\cfrac{1}{k}} \quad n,k \in \mathbb{N}^+$$

Hence the partition for T^2 is the collection of intervals $I_{n,k} = \left[\frac{1}{n+\frac{1}{k}}, \frac{1}{n+\frac{1}{k+1}}\right]$. The facts that these intervals have disjoint interior and that their union covers $[0,1]$ (mod 0) are both obvious. To be clear, two sets being equal mod 0 means the measure of $(A \cup B) \setminus (A \cap B)$ is zero. That T^2 is monotone on $I_{n,k}$ follows immediately from $(T^2)'$ being positive in the interiors of $I_{n,k}$. That the square map is surjective on each interval, follows immediately from applying the map twice on any interval $I_{n,k}$. For convenience let $P_{n,k} = P|_{I_{n,k}}$. Of course P has a well defined inverse on $I_{n,k}$. Define $\psi_{n,k} = P^{-1}_{n,k}$. The situation can be summarised with the diagram





where we have explicitly

$$P_{n,k} = \frac{1}{\cfrac{1}{x}-k} - n, \quad \psi_{n,k} = \frac{1}{k+\cfrac{1}{n+x}}$$

and it will also be necessary to have the derivative of $P_{n,k}$

$$P'_{n,k} = \frac{1}{(1-kx)^2}.$$

Composing with $\psi_{n,k}$, taking the reciprocal and absolute value gives

$$\frac{1}{|P'_{n,k} \circ \psi_{n,k}|} = \frac{1}{(1+k(n+x))^2}$$

So we can now write

$$\mathcal{L}_P f(x) = \sum_{n,k \geq 1} \frac{1}{(1+k[n+x])^2} f\left(\frac{1}{k+\frac{1}{n+x}}\right).$$

Of course this expression could be more simply obtained by applying $\mathcal{L}_T$ twice by proposition 2.12. However, it is useful to see explicitly how one can get the intervals $I_{n,k}$, and the functions $\psi_{n,k}$. As hinted in the title of the current section, we shall be dealing with functions of bounded (total) variation. Then next theorem is the crucial ingredient in this endeavour.

Theorem 4.2. *If $f \in BV([0,1])$, then*

$$\bigvee_0^1 \mathcal{L}_P|f| \leq 6\int_0^1 f dx + \frac{1}{2}\bigvee_0^1 f$$

where $\bigvee_a^b f$ is the variation of the function f from a to b.

Proof.

$$\bigvee_0^1 \mathcal{L}_P f(x) \le \sum_{n,k\ge 1} \bigvee_0^1 \frac{1}{(1+k[n+x])^2} f\left(\frac{1}{k+\frac{1}{n+x}}\right) \tag{4.1}$$
$$= \sum_{n,k\ge 1} \underbrace{\bigvee_{I_{n,k}} \frac{f}{|P'|}}_{L} + \underbrace{\frac{|f\circ\psi_{n,k}(0)|}{(1+kn)^2} + \frac{|f\circ\psi_{n,k}(1)|}{(1+k(n+1))^2}}_{R}$$

where we now bound L and R respectively. We begin with R as follows

$$\begin{aligned}
R &\le \frac{1}{(1+kn)^2}\big(|f\circ\psi_{n,k}(0)| + |f\circ\psi_{n,k}(1)|\big) \le \frac{1}{(1+kn)^2}\big(2\inf_{x\in I_{n,k}}|f(x)| + \bigvee_{I_{n,k}} f\big)\\
&\le \frac{1}{(1+kn)^2}\left(\frac{2}{|I_{n,k}|}\int_{I_{n,k}}|f|dx + \bigvee_{I_{n,k}} f\right) = \frac{2(1+k+kn)}{(1+kn)}\int_{I_{n,k}}|f|dx + \frac{1}{(1+kn)^2}\bigvee_{I_{n,k}} f\\
&\le 4\int_{I_{n,k}}|f|dx + \frac{1}{4}\bigvee_{I_{n,k}} f
\end{aligned}$$

where one uses that $|I_{n,k}|^{-1} = (1+kn)(1+k+kn)$. Now for L we proceed as follows:

$$\begin{aligned}
L &= \bigvee_{I_{n,k}} \frac{f}{|P'_{n,k}|} = \int_{I_{n,k}}\left|d\left(\frac{f}{|P'_{n,k}|}\right)\right| = \int_{I_{n,k}}\left|fd\left(\frac{1}{|P'_{n,k}|}\right) + \frac{1}{|P'_{n,k}|}d(f)\right|\\
&\le \int_{I_{n,k}}|f|\left|\frac{P''_{n,k}}{(P'_{n,k})^2}\right|dx + \int_{I_{n,k}}\left|\frac{1}{|P'_{n,k}|}d(f)\right| \le \sup_{x\in I_{n,k}}\left|\frac{P''_{n,k}(x)}{P'_{n,k}(x)^2}\right|\int_{I_{n,k}}|f|dx + \frac{1}{4}\bigvee_{I_{n,k}} f\\
&\le 2\int_{I_{n,k}}|f|dx + \frac{1}{4}\bigvee_{I_{n,k}} f.
\end{aligned}$$

We now apply the above to (4.1) to get

$$\begin{aligned}
\bigvee_0^1 \mathcal{L}_P f(x) &\le \sum_{n,k\ge 1}\left[2\int_{I_{n,k}}|f|dx + \frac{1}{4}\bigvee_{I_{n,k}} f + 4\int_{I_{n,k}}|f|dx + \frac{1}{4}\bigvee_{I_{n,k}} f\right]\\
&\le 6\int_0^1|f|dx + \frac{1}{2}\bigvee_0^1 f
\end{aligned}$$

finishing the proof. □

The end result and the argument are both similar to that achieved in [7], and [5] originally. The inequality allows the same choice of cones as Liverani. Namely, we can choose $\mathscr{C}_a$ as

$$\mathscr{C}_a = \left\{ f\in BV([0,1]) \;\middle|\; f\not\equiv 0, f\ge 0, \bigvee_0^1 f \le a\int_0^1 f \right\}.$$

which is highly desirable for two reasons. Firstly, we follow the heuristic of imposing regularity which we mentioned in Subsection 3.1. This regularity is ensured by Theorem 4.2. Secondly, there is the practical consideration of much of the the analysis becoming identical. There is however a non-trivial difficulty that is yet to be seen. We only comment on this now and return to it when bounding the projective diameter of $\mathcal{L}_P\mathscr{C}$. The first step is showing that our chosen cone is mapped strictly inside itself for a suitable choice of parameter a.

Lemma 4.3. *For $\mathscr{C}_a$ we have $\mathcal{L}_P\mathscr{C}_a \subseteq \mathscr{C}_{\sigma a}$ whenever $a > 12$ and $\sigma = \frac{6}{a} + \frac{1}{2}$.*

Proof. Let $\mathscr{C}_a$ be as above and $a \geq 12$. Then

$$\bigvee_0^1 \mathcal{L}_P f \leq 6\int_0^1 f + \frac{1}{2}\bigvee_0^1 f \leq 6\int_0^1 f + \frac{a}{2}\int_0^1 f = \left(\frac{6}{a} + \frac{1}{2}\right) a\int_0^1 f = \sigma a\int_0^1 f$$

where we define $\sigma = \frac{6}{a} + \frac{1}{2}$ which is less than one if $a \geq 12$. □

Before moving on to the diameter we need to establish the Lemmas 4.4, 4.5 and 4.8.

Lemma 4.4. *The Gauss-map, and hence it's square, is covering in as defined in Definition B.4.*

Proof. Let $n \in \mathbb{N}$. Then pick any $I \in \mathscr{A}_n$. Then it is clear that $T^n I = [0,1]$. Since $n \leq 2(\lceil \frac{n}{2} \rceil)$ we also have

$$P^{(\lceil \frac{n}{2} \rceil)} I = (T^2)^{(\lceil \frac{n}{2} \rceil)} I \supseteq T^n I = [0,1]$$

Showing that P is covering with $N(n) = \lceil \frac{n}{2} \rceil$. □

Lemma 4.5. *Given a partition $\mathcal{P}$ of $[0,1] \mod 0$, if each $p \in \mathcal{P}$ is an interval with Lebesgue measure less than $1/2a$, then, for each $g \in \mathscr{C}_a$, there exists $p_0 \in \mathcal{P}$ such that*

$$g(x) \geq 1/2\int_0^1 g \quad \forall x \in p_0$$

Proof. Suppose that the lemma is false. Then for all intervals $p_0 \in \mathcal{P}$ one must have that there exists a $x_p \in p_0$ such that $g(x_p) < 1/2\int_0^1 g$. In that case we can for an arbitrary $p \in \mathcal{P}$ estimate

$$\int_p g \leq |p|\left(g(x_p) + \bigvee_p g\right) < \frac{|p|}{2}\int_0^1 g + \frac{1}{2a}\bigvee_p g$$

where in the first step one uses that $g(x) \leq g(x_p) + \bigvee_p g$ for all x in p, and in the second our assumption on our partition $\mathcal{P}$. Now summing over all $p \in \mathcal{P}$ together with the fact that $g \in \mathscr{C}_a$ gives

$$\int_0^1 g < \sum_{p\in\mathcal{P}}\left(\frac{|p|}{2}\int_0^1 g + \frac{1}{2a}\bigvee_p g\right) \leq \frac{1}{2}\int_0^1 g + \frac{1}{2a}\bigvee_0^1 g \leq \frac{1}{2}\int_0^1 g + \frac{a}{2a}\int_0^1 g = \int_0^1 g$$

which is a contradiction. □

We shall now give a definition that puts any interval $I \in \mathcal{A}_n$ into one of two categories. One being those we consider well-behaved or "nice" in relation to T and the other of those considered poorly-behaved or "bad".

Definition 4.6. *Let $I \in \mathscr{A}_n$ be an interval such that T^n is continuous and monotone on I and in addition, $T^n I = [0,1]$. Then I is given by a coding $k_1, k_2, \ldots, k_n$, corresponding to iterated application of the piecewise inverses ψ_{k_i} to $[0,1]$. Explicitly, $I = \psi_{k_n} \circ \cdots \circ \psi_{k_1}[0,1]$. Let $m \in \mathbb{N}$. We call I m-nice if $k_i \leq m$ for all $i = 1, 2, \ldots, n$. If $I \in \mathscr{A}_n$ is not m-nice it is m-bad.*

Let $l = \{l_k\}_{k=1}^n$ be a sequence of n positive integers. Then define $I_l = \psi_{l_n} \circ \cdots \circ \psi_{l_1} \circ \psi_{l_1}[0,1]$. It immediately follows that

$$1 = |T^n I_l| \geq \lambda^n |I_l| \quad \text{so} \quad |I_l| \leq \lambda^{-n},$$

and hence choosing $n = \left\lceil \frac{\log 2a}{\log \lambda} \right\rceil = \left\lceil \frac{\log 2a}{\log 4} \right\rceil$ is sufficient to ensure that $|I| < \frac{1}{2a}$ for all $I \in \mathscr{A}_n$. With this choice for n we have the following lemma:

Lemma 4.7. *Let $\mathscr{P}$ be a uniform partition of $[0,1]$ into intervals such that for each $p \in \mathscr{P}$ we have $|p| = \frac{1}{\lceil 2a \rceil}$. Pick $p_0 \in \mathscr{P}$. If $m \geq 2(\lceil 2a \rceil + 1)$, then there exists an n such that there exists an $I \in \mathscr{A}_n$ such that $I \subseteq p_0$ and I is m-nice.*

Proof. Pick $m, n > 0$. Note that no m-bad interval in $\mathscr{A}_n$ is "isolated" in the following sense: if J is m-bad, then either the interval immediately to the left or right of J is also m-bad. This is clear when one remembers how the intervals correspond to coefficients in regular continued fraction expansions.

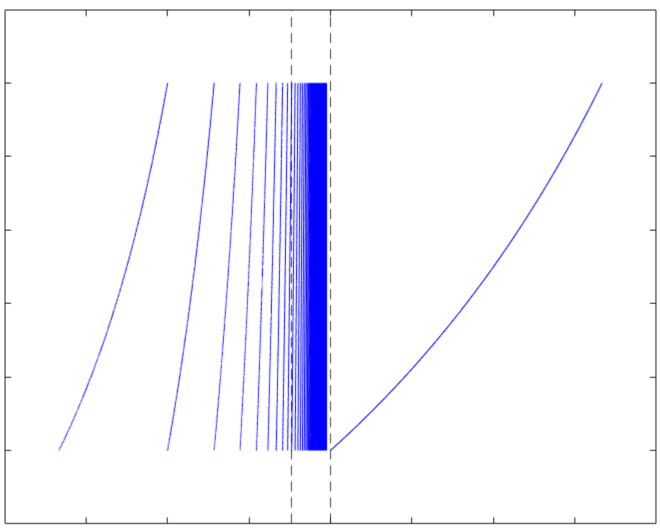


Figure 1: Typical behaviour of a m-bad interval. Interval between dotted lines is m-bad

Any such clump of m-bad intervals has length less than $\frac{1}{m}$, and will necessarily have an m-nice interval adjacent to it. Since any interval in $\mathscr{A}_n$ has measure bounded above by $4 \cdot 2^{-n}$, we can ensure the existence of an m-nice interval in p_0 if the following inequality is fulfilled:

$$|p_0| - \frac{1}{m} - 2(4 \cdot 2^{-n}) > 0.$$

We pick m such that $\frac{1}{m} \ll \frac{1}{2a}$, for example, choosing $m \geq 2(\lceil 2a \rceil + 1)$ gives that

$$|p_0| - \frac{1}{m} - 2 \cdot 4 \cdot 2^{-n} > \frac{1}{\lceil 2a \rceil + 1} - \frac{1}{2(\lceil 2a \rceil + 1)} - 8 \cdot 2^{-n} = \frac{1}{2(\lceil 2a \rceil + 1)} - 8 \cdot 2^{-n}$$

which is strictly bigger than zero if one chooses $n \geq 4 + \left\lceil \frac{\log(\lceil 2a \rceil + 1)}{\log 2}. \right\rceil$. Picking such an n proves the lemma. □

Corollary 4.7.1. *Consider the square of the Gauss map, $P = T^2$. Choose the partition, $\mathscr{P}$ and constant m as in Lemma 4.7, and $p_0 \in \mathscr{P}$. Then it is sufficient to choose $n \geq \frac{1}{2}\left\lceil 1 + \frac{\log(\lceil 2a \rceil + 1)}{\log 2} \right\rceil$ to ensure the existence of an interval $I \in \mathscr{A}_{2n}$ with $I \subseteq p_0$.*

Proof. It suffices to follow the same calculations as in the proof of the lemma, with the additional observation that the measure of $I \in \mathscr{A}_{2n}$ can be bounded above by 4^{-n}. □

Lemma 4.8. *Let $I \in \mathscr{A}_n$ be an m-nice interval. Then for any $x \in I$*

$$|(T^n)'(x)| \leq m^{2n}$$

Proof. The statement follows immediately from the fact that $|T'(x)| \leq m^2$ whenever $x \geq \frac{1}{m}$ and the definition of m-nice. □

Lemma 4.9. *Let $g \in \mathscr{C}_a$ be a function. If $N \geq \frac{1}{2}\left\lceil 1 + \frac{\log(\lceil 2a \rceil + 1)}{\log 2} \right\rceil$, $m \geq 2(\lceil 2a \rceil + 1)$, then*

$$\mathcal{L}_P^N g(x) \geq \frac{\int_0^1 g}{2 \cdot m^{4N}}$$

for any $g \in \mathscr{C}_a$

Proof. If $N \geq \frac{1}{2}\left\lceil 1 + \frac{\log(\lceil 2a \rceil + 1)}{\log 2} \right\rceil$, then Corollary 4.7.1 along with Lemma 4.5 ensure the existence of an interval $p_0 \in \mathscr{P}$ and an m-nice interval $I \in \mathscr{A}_{2N}$ completely contained within p_0 such that $g(x) \geq \frac{1}{2}\int_0^1 g$ for all $x \in p_0$ and hence also in the m-nice interval I. This gives that

$$\mathcal{L}_P^N g(x) = \mathcal{L}_{(T^2)}^N g(x) = \sum_{y \in T^{-2N}\{x\}} \frac{g(y)}{|(P^N)'(y)|} \geq \frac{\int_0^1 g}{2 \cdot m^{4N}}$$

using Lemma 4.8. □

4.2 Diameter for BV case

Lemma 4.10. *The diameter of $\mathcal{L}_T^N \mathscr{C}_a$ is finite for some N and can be bounded by*

$$\operatorname{diam}(\mathcal{L}_P^N \mathscr{C}_a) \leq 2 \log \left[\frac{\max\{1 + \sigma; 1 + a\sigma\}}{\min\{(1 - \sigma); \frac{1}{m^{4N}}} \right] = \Delta$$

where N is as in lemma 4.9 and σ as in Lemma 4.3.

Proof. Lemma B.2 gives that the distance

$$\Theta_a(1, g) \leq \log \left[\frac{\max\{(1 + \sigma)\int_0^1 g; \sup g\}}{\min\{(1 - \sigma)\int_0^1 g; \inf g\}} \right]$$

whenever $g \in \mathscr{C}_{\sigma a}$ with $\sigma < 1$. Now the proof of our lemma relies of choosing appropriate σ and N. The first observation is that for appropriate choices of a and σ it is true that

$$\mathcal{L}^N \mathscr{C}_a \subset \mathscr{C}_{\sigma a}.$$

This can be achieved iteratively applying Theorem 4.2 to the variation of $\mathcal{L}_T^N f$, $f \in \mathscr{C}_a$, giving

$$\begin{aligned}\bigvee_0^1 \mathcal{L}_P^N f &\leq \left(\frac{1}{2}\right)^N \bigvee_0^1 f + 6 \cdot \frac{1-(1/2)^N}{1-1/2} \int_0^1 f \\ &\leq \left[a\left(\frac{1}{2}\right)^N + 12 \cdot \left(1-(1/2)^N\right)\right] \int_0^1 f = \sigma a \int_0^1 f\end{aligned} \tag{4.2}$$

where

$$\sigma = \left(\frac{1}{2}\right)^N + \frac{12}{a}\left(1-(1/2)^N\right)$$

and where $\sigma < 1$ holds as long as $a > 12$. Now using Lemma B.2 to bound the diameter and then applying (4.2) gives

$$\begin{aligned}\mathrm{diam}\mathcal{L}_P^N \mathscr{C}_a &\leq 2\log\left[\frac{\max\{(1+\sigma)\int_0^1 \mathcal{L}_P^N g; \sup \mathcal{L}_P^N g\}}{\min\{(1-\sigma)\int_0^1 \mathcal{L}_P^N g; \inf \mathcal{L}_P^N g\}}\right] \\ &\leq 2\log\left[\frac{\max\{(1+\sigma)\int_0^1 \mathcal{L}_P^N g; \inf \mathcal{L}_P^N g + \bigvee_0^1 \mathcal{L}_P^N g\}}{\min\{(1-\sigma)\int_0^1 \mathcal{L}_P^N g; \inf \mathcal{L}_P^N g\}}\right]\end{aligned}$$

where we use that $\sup|f| \leq \inf|f| + \bigvee f$. Now, remembering that $g \in \mathscr{C}_a$ implies that $\mathcal{L}_T^N g \in \mathscr{C}_{\sigma a}$. This gives

$$\begin{aligned}&\leq 2\log\left[\frac{\max\{(1+\sigma)\int_0^1 \mathcal{L}_P^N g; \int_0^1 \mathcal{L}_P^N g + a\sigma \int_0^1 \mathcal{L}_P^N g\}}{\min\{(1-\sigma)\int_0^1 \mathcal{L}_P^N g; \inf \mathcal{L}_P^N g\}}\right] \\ &= 2\log\left[\frac{\max\{1+\sigma; 1+a\sigma\}}{\min\{(1-\sigma); \inf \mathcal{L}_P^N g \left(\int_0^1 g\right)^{-1}}\right].\end{aligned}$$

Now finally applying Lemma 4.9 gives

$$\begin{aligned}\mathrm{diam}\mathcal{L}^N \mathscr{C}_a &\leq 2\log\left[\frac{\max\{1+\sigma; 1+a\sigma\}}{\min\{(1-\sigma); \frac{\int_0^1 g}{2m^{4N}}\left(\int_0^1 g\right)^{-1}}\right] \\ &= 2\log\left[\frac{\max\{1+\sigma; 1+a\sigma\}}{\min\{(1-\sigma); (2m^{4N})^{-1}}\right] = \Delta\end{aligned}$$

where m and N are chosen as in the lemma. □

The coming procedure is very much analogous to the one at the end of section 3 and 3.2 so details will be somewhat glossed over. We begin by restating, with a slight modification but without proof, Lemma 3.2

Lemma 4.11. *For $f \in \mathscr{C}_a$ with $\int_0^1 f d\mu = 1$ and $\phi = \frac{1}{\log 2} \cdot \frac{1}{1+x}$ one has*

$$\left\|\frac{\mathcal{L}_T^n f}{\phi} - 1\right\|_\infty \leq \exp\left[\Theta(\mathcal{L}_T^n f, \phi)\right] - 1.$$

Lemma 4.12. *We have exponential decay of correlations for BV vs L^1 with speed of decay at least $(\Lambda^{\frac{1}{2N}})^n$, $\Lambda = \tanh\left(\frac{\Delta}{4}\right)$ with Δ given by Lemma 4.10 and N chosen in accordance to the lemma.*

Proof. Let $f \in \mathscr{C}_a$ and $g \in L^1$. Assume, for the moment, that $\int_0^1 f = 1$, then

$$\begin{aligned} C(f,g,n) &= \left| \int_0^1 \mathcal{L}_T^n f \cdot g\, dx - \int_0^1 f dx \cdot \int_0^1 g d\mu \right| = \left| \int_0^1 \mathcal{L}_T^n f \cdot g dx - \int_0^1 g\, d\mu \right| \\ &= \left| \int_0^1 (\mathcal{L}_T^n f - \phi) g dx \right| \le \|g\|_1 \cdot \left\| \frac{\mathcal{L}_T^n f}{\phi} - 1 \right\|_\infty \|\phi\|_\infty . \end{aligned}$$

Applying Lemma 2.2 and 4.11 yields

$$C(f,g,n) \le \|g\|_1 \cdot \|\phi\|_\infty (\exp[\Theta(\mathcal{L}_T^n f, \phi)] - 1) \le \|g\|_1 \cdot \|\phi\|_\infty (\exp[\Theta((\mathcal{L}_T^2)^{\lfloor \frac{n}{2} \rfloor} f, \phi)] - 1). \quad (4.3)$$

Now clearly $\Theta((\mathcal{L}_T^2)^{\lfloor \frac{n}{2} \rfloor} f, \phi) = \Theta(\mathcal{L}_P)^{\lfloor \frac{n}{2} \rfloor} f, \phi)$. Now, choosing N as in Lemma 4.10, and writing $\lfloor \frac{n}{2} \rfloor = kN + r$ with $k \in \mathbb{N}$ and $r \in \{ 1, 2, \ldots, N-1 \}$ gives that $k = \left\lfloor \frac{1}{N} \lfloor \frac{n}{2} \rfloor \right\rfloor$ we get

$$\Theta(\mathcal{L}_P^{\lfloor \frac{n}{2} \rfloor} f, \phi) \le \Theta((\mathcal{L}_P^N)^k f, \phi).$$

Since ϕ is invariant under $\mathcal{L}_T$ and $\mathcal{L}_P^N$ is a strict contraction in the projective metric associated to $\mathscr{C}_a$, we get

$$\Theta((\mathcal{L}_P^N)^k f, \phi) \le \tanh\left(\frac{\Delta}{4}\right)^{k-1} \Theta(\mathcal{L}_P^N f, \mathcal{L}_P^N \phi) \le \tanh\left(\frac{\Delta}{4}\right)^{k-1} \Delta$$

where we can now apply the above to (4.3) to get

$$C(f,g,n) \le \|g\|_1 \|\phi\|_\infty (\exp[\Lambda^{k-1}\Delta] - 1) \le \|g\|_1 \|\phi\|_\infty \le \|g\|_1 \|\phi\|_\infty \Lambda^{k-1} \Delta \exp \Lambda^{k-1} \Delta$$

when $\Lambda = \tanh\left(\frac{\Delta}{4}\right)$, Δ is as in Lemma 4.10, and $k = \left\lfloor \frac{1}{N} \lfloor \frac{n}{2} \rfloor \right\rfloor$. Put $\hat{K}_n = \exp \Lambda^{k-1}\Delta$, then some algebra gives that

$$\|g\|_1 \|\phi\|_\infty \Lambda^{k-1}\Delta \exp \Lambda^{k-1}\Delta \le \hat{K}_n \cdot \|g\|_1 \|\phi\|_\infty \Lambda^{-2N} \cdot (\Lambda^{\frac{1}{2N}})^n.$$

Finally, let $K_n = \|\phi\|_\infty \Lambda^{-2N} \hat{K}_n$ (note that $\hat{K}_n$ and K_n are both decreasing, and hence bounded), then

$$C(f,g,n) \le K_n \|g\|_1 (\Lambda^{\frac{1}{2N}})^n. \quad (4.4)$$

Since we assumed that $\int_0^1 f d\mu = 1$, we can scale (4.4) by $\|f\phi\|_1$ to recover the general case and get a bound for any $f \in \mathscr{C}_a$, giving

$$C(f,g,n) \le K_n \|\phi\|_\infty \|f\|_1 \|g\|_1 (\Lambda^{\frac{1}{2N}})^n.$$

Now let

$$\begin{aligned} C(f,g,n) &\le C(f + a^{-1} \bigvee_0^1 f, g, n) + C(a^{-1} \bigvee_0^1 f, g, n) \\ &\le K_n \|\phi\|_\infty \left\| f + a^{-1} \bigvee_0^1 f \right\|_1 \|g\|_1 (\Lambda^{\frac{1}{2N}})^n + K_n \|\phi\|_\infty \left\| a^{-1} \bigvee_0^1 f \right\|_1 \|g\|_1 (\Lambda^{\frac{1}{2N}})^n \\ &\le K_n \|\phi\|_\infty (\|f\|_1 + \frac{2}{a} \bigvee_0^1 f) \|g\|_1 (\Lambda^{\frac{1}{2N}})^n \\ &\le C \|f\|_{BV} \|g\|_1 (\Lambda^{\frac{1}{2N}})^n \end{aligned}$$

finishing the proof. □

One might at this stage want an explicit example of what the speed of decay may be bounded by. It can be verified that choosing $a = 12.5, N = 3, m = 52$ fulfills all the necessary assumptions. From this we can by Lemma 4.3 calculate $\sigma = \frac{6}{12.5} + \frac{1}{2} = \frac{49}{50}$ and then Δ as

$$\Delta = 2\log\left[\frac{\max\{1+\frac{49}{50}; 1+12.5\cdot\frac{49}{50}\}}{\min\{(1-\frac{49}{50}); \frac{1}{2\cdot 52^{12}}\}}\right] = 2\log\left[\frac{\frac{53}{4}}{\frac{1}{2\cdot 52^{12}}}\right] = 2\log\left[\frac{53}{2}\cdot 52^{12}\right],$$

and from this $\Lambda = \tanh\left(\frac{\Delta}{4}\right) \leq 1 - 10^{-21}$ and then finally speed of decay $(\Lambda^{\frac{1}{6}}) \leq 1 - 2\cdot 10^{-23}$.

4.3 Comparison of Results

It is of some note how drastically different the speed of decay achieved in Section 3 for Lip([0, 1]) functions and that of Section 4 where we handle $BV([0,1])$ functions. Of course, since C^1 and Lipschitz functions on $[0, 1]$ are strict subsets of BV, the optimal decay of correlations is smaller in the first case. Furthermore, since the approach to find bounds is the same in each instance, it is from a practical perspective reasonable so suspect better decay estimates as well, although perhaps not to this extent. A pertinent question would then be whether the reason lies in a much smaller spectral gap for the operator, or due to the program relying on very non-optimal bounds. For the first case there is of course nothing to be done, as the spectral gap is the best possible speed of decay. For example it was proved by P. Levi that

$$\lim_{n\to\infty}\int_0^1 \chi_{[0,a]}\mathcal{L}_T^n(g)\,dx \longrightarrow \int_0^1 \chi_{[0,a]}\phi(x)\,dx$$

converges as q^n to the invariant density ϕ with $0 < q < 0.68$, and then later numerically calculated up to 20 decimal places by Wirsing [8] to

$$q = 0.30366300289873265860\ldots.$$

If one is instead in the second case, one would have to answer to what degree more optimal bounds would improve the final result. For example in the BV case, when bounding the infimum of functions in $\mathcal{L}_P\mathscr{C}_a$ we had to introduce m-nice intervals and get the lower bound of $(2\cdot m^{4N})^{-1}$, which is exceptionally small, giving a large diameter. But even if this term could be improved, would it have a meaningful impact on the result? This is unknown.

5 An application of the Decay of Correlations

In this section we shall be showcasing a result that quickly follows from the decay of correlations for functions of bounded variation.

5.1 Borel–Cantelli lemma

In [4], Dong Han Kim states two versions of the classical Borel–Cantelli Lemmas, and later introduces the notion of (Strong) Borel–Cantelli sequences (SBC/BC). We begin by defining these properties. Let (X, μ) be a probability space and $\{A_n\}_{n=1}^{\infty}$ a sequence of subsets of X with $\sum_{n=1}^{\infty}\mu(A_n) = \infty$, and T a measure preserving transformation on X under μ. The sequence is called a Borel–Cantelli sequence if for μ almost every $x \in X$, $T^n x \in A_n$ hold for infinitely many n. We shall now define the notion of a Strong Borel–Cantelli sequence. Let $S_N(x)$ denote the number of times $T^n x \in A_n$ holds for $n = 1, 2, \ldots, N$. We can write $S_N(x)$ as

$$S_N(x) = \sum_{n=1}^{N}\chi_{A_n}\circ T^n(x) = \sum_{n=1}^{N}\chi_{T^{-n}A_n}(x) = \sum_{n=1}^{N}\chi_{B_n}(x)$$

where we define $B_n = T^{-n}A_n$. Now let the quantity E_N be defined by

$$E_N = \mu(S_N) = \int_0^1 S_N d\mu = \sum_{n=1}^{N} \mu(B_n) = \sum_{n=1}^{N} \mu(A_n)$$

where in the last equality we use that T is measure preserving. Now we finally say that a sequence of subsets A_N is a strong Borel–Cantelli sequence (SBC) if

$$\lim_{N\to\infty} \frac{S_N(x)}{E_N} = 1$$

holds for μ almost every $x \in X$. One can think of this as if the orbits of an x is "evenly distributed" in X. Meaning that the likelihood of hitting a target A_n at time n, is on average just the size of A_n. Since we know that E_N must tend to infinity as $N \to \infty$, it is clear that being a SBC sequence implies being a SB sequence. Now we are ready to use this framework for our purposes.

5.2 Growth of continued fraction coefficients

Let $A_n = [0, 1/n)$. A number x is in A_n if its leading continued fraction coefficient is larger than or equal to n. This also means that if $T^n x$ is in A_n if the n-th associated continued fraction coefficient is greater than or equal to n. We shall show that for a.e. x, the orbit $T^n x$ "hits" the collection of shrinking targets A_n infinitely many times. Another way to view it is that $x \in B_n = T^{-n}A_n$ holds for infinitely many n. It shall be nice to establish the following result:

Lemma 5.1. *If A_k is defined as above and $B_k = T^{-k}A_k$, then we have*

$$\mu(B_k \cap B_l) = \mu(B_k)\mu(B_l) + \mathcal{O}\left(\frac{\rho^{|k-l|}}{\max(l,k)}\right)$$

where $\rho = (\Lambda^{\frac{1}{2N}})$ is the rate of exponential decay from section 4.

Proof. Assume without loss of generality that $l > k$. Then

$$\begin{aligned}
\mu(B_k \cap B_l) &= \int_{B_k \cap B_l} d\mu = \int_0^1 \chi_{B_k}(x) \cdot \chi_{B_l}(x) d\mu = \int_0^1 \chi_{T^{-k}A_k}(x) \cdot \chi_{T^{-l}A_l}(x) d\mu \\
&= \int_0^1 \chi_{A_k} \circ T^k(x) \cdot \chi_{A_l} \circ T^l(x) d\mu = \int_0^1 \chi_{A_k}(x) \cdot \chi_{A_l} \circ T^{l-k}(x) d\mu \\
&\le \int_0^1 \chi_{A_k}\, d\mu \cdot \int_0^1 \chi_{A_l}\, d\mu + C \left\|\chi_{A_k}\right\|_{BV} \left\|\chi_{A_l}\right\| \rho^{l-k} \\
&\le \int_0^1 \chi_{A_k}\, d\mu \cdot \int_0^1 \chi_{A_l}\, d\mu + D\frac{\rho^{l-k}}{l} = \mu(A_k)\mu(A_l) + \mathcal{O}\left(\frac{\rho^{l-k}}{l}\right)
\end{aligned}$$

where the first inequality follows from the exponential decay of correlations for functions of bounded variation. The general case follows from interchanging l and k. ☐

Lemma 5.2. *The limit*

$$\lim_{N\to\infty} \frac{S_N(x)}{E_N} = 1$$

holds for μ almost every $x \in X$.

Proof. Consider the integral

$$\int_0^1 \left(\frac{S_N}{E_N} - 1\right)^2 d\mu = \int_0^1 \left(\frac{S_N^2}{E_N^2} - 2\frac{S_N}{E_N} + 1\right) d\mu \tag{5.1}$$
$$= \int_0^1 \frac{S_N^2}{E_N^2} d\mu - \frac{2}{E_N} \cdot E_N + 1 = \int_0^1 \frac{S_N^2}{E_N^2} d\mu - 1$$

where we use that E_N is a constant with respect to the integration and that $\mu(S_N) = E_N$ by definition. We shall now make use of the exponential decay of correlations for the Gauss map and functions of bounded variation

$$\int_0^1 \frac{S_N^2}{E_N^2} d\mu = \left|\int_0^1 \frac{S_N^2}{E_N^2} d\mu\right| \tag{5.2}$$
$$= \frac{1}{E_N^2} \left|\int_0^1 S_N^2 d\mu - \int_0^1 S_N d\mu \int_0^1 S_N d\mu + \int_0^1 S_N d\mu \int_0^1 S_N d\mu\right|.$$

Let L_N equal the absolute values of the two leftmost terms in (5.2) and then expanding $S_N(x)^2$ gives

$$L_N = \left|\int_0^1 S_N \cdot S_N d\mu - \int_0^1 S_N d\mu \cdot \int_0^1 S_N d\mu\right|$$
$$= \left|\int_0^1 \left(\sum_{k=1}^N \chi_{B_k}(x)\right)\left(\sum_{l=1}^N \chi_{B_l}(x)\right) d\mu - \int_0^1 \sum_{k=1}^N \chi_{B_k}(x) d\mu \cdot \int_0^1 \sum_{l=1}^N \chi_{B_l}(x) d\mu\right|$$
$$\leq \sum_{k,l\geq 1}^N \left|\int_0^1 \chi_{B_k} \cdot \chi_{B_l} d\mu - \int_0^1 \chi_{B_k} d\mu \int_0^1 \chi_{B_l} d\mu\right|.$$

Since the sum is symmetric in k and l we compute it by considering the diagonal $l = k$ and adding twice the sum where $l > k$. Thus we can apply Lemma 5.1 giving

$$\leq \sum_{k=1}^N \mu(B_k)(1 - \mu(B_k)) + 2\sum_{k=1}^{N-1} \sum_{l=k+1}^N C(\chi_{B_l}, \chi_{B_k}, n) \leq E_N + 2\sum_{k=1}^{N-1} \sum_{l=k+1}^N \mathcal{O}\left(\frac{\rho^{l-k}}{l}\right) \tag{5.3}$$
$$\leq E_N + \mathcal{O}\left(\sum_{k=1}^{N-1} \sum_{l=k+1}^N \frac{\rho^{l-k}}{l}\right) \leq E_N + D\sum_{k=1}^{N-1} \frac{1}{k+1} \sum_{l=k+1}^N \rho^{l-k}$$
$$\leq E_N + D\frac{\rho}{1-\rho} \sum_{k=1}^{N-1} \frac{1}{k+1}(1 - \rho^{N-k}) \leq E_N + \tilde{D}\sum_{k=1}^N \mu(B_k) = (1 + \tilde{D})E_N.$$

Summarising, we apply (5.3) to (5.1) to get

$$0 \leq \int_0^1 \left(\frac{S_N^2}{E_N^2} - 1\right)^2 d\mu = \int_0^1 \frac{S_N(x)^2}{E_N^2} d\mu - 1$$
$$\leq \frac{1}{E_N^2}\left[\int_0^1 S_N d\mu \cdot \int_0^1 S_N + (1 + \tilde{D})E_N\right] - 1 = 1 + \frac{1 + \tilde{D}}{E_N} - 1 \longrightarrow 0$$

as $N \to \infty$. Finally we get that

$$\int_0^1 \left(\frac{S_N^2}{E_N^2} - 1\right)^2 d\mu \longrightarrow 1 - 1 = 0$$

as $N \to \infty$. It follows immediately that $\left(\frac{S_N(x)}{E_N} - 1\right)^2 \to 0$ holds μ almost everywhere and the lemma is proved. □

The consequence of this lemma, that of the growth of regular continued fraction coefficients, in some sense fulfills a promise made in the thesis introduction. It was mentioned that the statistical properties of the Gauss gives information about continued fractions. The above lemma achieves exactly this, and the crucial ingredient for the proof is the exponential decay of correlations for functions of bounded variation. This showcases a general fact, that statistical information can be used to prove results within the fields of metric number theory and Diophantine approximation.

A Diameter details

Here we keep some of the more tedious details connected to showing the diameter of

$$\mathscr{C} = \left\{ f \in C^1[0,1] \middle| f \not\equiv 0,\ f \geq 0,\ 0 \leq -f' \leq 2f \right\}$$

is finite.

Lemma A.1. *The function* $t : [0,1] \to \mathbb{R}$ *defined by*

$$t(x) = \sum_{n=1}^{\infty} \frac{1+x}{(x+n)^2}$$

is decreasing on $[0,1]$.

Proof. We show that the derivative is negative everywhere on $[0,1]$. Calculating the derivative gives

$$t'(x) = \sum_{n=1}^{\infty} \frac{(n+x) - 2(1+x)}{(n+x)^3} = \sum_{n=1}^{\infty} \frac{n-2}{(n+x)^3} - \sum_{n=1}^{\infty} \frac{x}{(n+x)^3}$$

where we can no keep only a few of the initial terms and make worst case estimates for the remainders. Write

$$\begin{aligned} t'(x) &= \frac{-1}{(1+x)^3} + \frac{0}{(1+x)^3} + \sum_{n=3}^{\infty} \frac{n-2}{(n+x)^3} - \frac{x}{(1+x)^3} - \frac{x}{(2+x)^3} + \sum_{n=3}^{\infty} \frac{x}{(n+x)^3} \\ &= -\frac{1}{(1+x)^2} - \frac{x}{(2+x)^3} + \sum_{n=3}^{\infty} \frac{n-2}{(n+x)^3} - \sum_{n=3}^{\infty} \frac{x}{(n+x)^3} \\ &\leq -\frac{1}{4} - \frac{1}{27} + \sum_{n=1}^{\infty} \frac{n-2}{n^3} = -2(\zeta(3) - 9/8) - 83/54 + \pi^2/6 < -0.04 \end{aligned}$$

so we have proved that t is (strictly) decreasing. □

Lemma A.2. *The function* $u : [0,1] \to \mathbb{R}$ *defined by*

$$u(x) = \sum_{n=1}^{\infty} \frac{(n-1+x)(1+x)^2}{(x+n)^4}$$

is an increasing function.

Proof. Simply differentiating gives

$$u'(x) = \sum_{n=1}^{\infty} \frac{(x+1)(2n^2 + n(x-5) - x^2 - x + 4)}{(x+n)^5}$$

where each term is greater than or equal to zero, and hence u is increasing on $[0, 1]$. □

Lemma A.3. *The function $p : [0, 1] \to \mathbb{R}$ defined by*

$$p(x) = \sum_{n=1}^{\infty} \frac{(1+x)^2}{(x+n)^3}$$

is decreasing on $[0, 1]$.

Proof. It suffices to show that the derivative p' is negative. Two terms and then a worst case estimate gives

$$p'(x) = \sum_{n=1}^{\infty} \frac{2(1+x)(x+n)^3 - 3(1+x)(x+n)^2}{(x+n)^6} < 0$$

□

Lemma A.4. *The function $l : [0, 1] \to \mathbb{R}$*

$$l(x) = \sum_{n=1}^{\infty} \frac{(x+n)^2 - (x+n) + 1}{(x+n)^4}$$

is a decreasing function.

Proof. The function $g(x) = \frac{x^2-x+1}{x^4}$ is decreasing on $[1, \infty)$. □

B Expanding maps on countable partitions

In this section we shall be attempt to produce explicit bounds on the decay of correlations of a general collection of maps. Note that in the following section, intervals will be considered disjoint if the measure (Lebesgue) of their intersection is zero. (Will mostly/only be single points)

B.1 Assumptions

Definition B.1. *A map $T : I \to I$ is called expanding if there exists a countable collection of intervals $\mathcal{A} = \{I_l\}_{l \in \mathcal{F}}$ with $I_l = [a_l, b_l]$ such that*

(E1) $I = \bigcup_{l \in \mathcal{F}} I_l$

(E2) $\operatorname{Int} I_l \cap \operatorname{Int} I_j = \emptyset$ *if* $i \neq j$

(E3) $T_l \coloneqq T|_{I_l}$ *is monotone and* C^k *for some* $k \geq 1$

(E4) $\inf |(T^n)'(x)| \geq \lambda > 1$ *for some* n

This is a vital ingredient in showing that the diameter of the image is finite.

B.2 Diameter estimates

As in the section 3.1, we shall use that

$$\operatorname{diam}(\mathcal{L}_T\mathscr{C}) = \sup_{f,g\in\mathscr{C}_a} \Theta(\mathcal{L}_T f, \mathcal{L}_T g) \leq 2 \sup_{g\in\mathscr{C}_a} \Theta(1, \mathcal{L}_T g).$$

Lemma B.2. *If $g \in \mathscr{C}_{\nu a}$, $\nu < 1$ then the projective distance $\Theta(1, g)$ associated to the cone $\mathscr{C}_a$ can be bounded by*

$$\Theta(1, g) \leq \log \left[\frac{\max\{(1+\nu)\int_0^1 g; \sup g\}}{\min\{(1-\nu)\int_0^1 g; \inf g\}} \right]$$

Proof. We begin by finding a lower bound for $\alpha(1, g)$, and to do so we solve for λ that satisfy $\lambda \preceq g$. This gives the double condition on λ that $\lambda \leq \inf g$ and $\lambda \leq \int_0^1 g - a^{-1} \bigvee_0^1 g$. Therefore

$$\alpha(1, g) = \min \left\{ \inf g; \int_0^1 g - a^{-1} \bigvee_0^1 g \right\}$$

and using that $g \in \mathscr{C}_{\nu a}$ gives that $-a^{-1} \bigvee_0^1 g \geq \nu \int_0^1 g$. Applying the α gives that

$$\alpha(1, g) \geq \min \left\{ \inf g; \int_0^1 g - \nu \int_0^1 g \right\} = \min \left\{ \inf g; (1-\nu) \int_0^1 g \right\}$$

which is our desired estimate. Now the process is analogous for $\beta(1, g)$. We find that for $g \preceq \mu$ the requirements of μ become $\mu \geq \sup g$ and $\mu \geq \int_0^1 g + a^{-1} \bigvee_0^1 g$ and hence that

$$\beta(1, g) = \max \left\{ \sup g; \int_0^1 g + a^{-1} \bigvee_0^1 g \right\}$$

and now $g \in \mathscr{C}_{\nu a}$ instead gives that $a^{-1} \bigvee_0^1 g \leq \nu \int_0^1 g$. Applying yields

$$\beta(1, g) = \max \left\{ \sup g; (1+\nu) \int_0^1 g \right\}$$

which finishes the proof. $\square$

Definition B.3.

$$\mathscr{A}_n = \bigvee_{j=0}^{n} T^{-j} \mathscr{A}_0$$

which is the partition generated by the preimages of sets in $\mathscr{A}_n$ over T^n.

The notation $\bigvee_{j=0}^n$ should not be confused with that of the variation of a function. Their similarity is unfortunate. Definition B.3 could also equivalently be stated as the coarsest partition such that T^n is continuous and monotone on every $I \in \mathscr{A}_n$. Now we wish to make precise the notion of a map "reaching" or "covering" all areas of its codomain.

Definition B.4. *We call a map T covering iff for each $n \in \mathbb{N}$ there exists $N(n)$ such that, for each $I \in \mathscr{A}_n$,*

$$T^{N(n)} I = [0, 1]$$

where this equality as always should be interpreted as equality except a set of measure zero.

References

[1] Garrett Birkhoff. *Lattice theory*. third. Vol. Vol. 25. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 1979, pp. vi+418. ISBN: 0-8218-1025-1.

[2] Abraham Boyarsky and Paweł Góra. *Laws of chaos*. Probability and its Applications. Invariant measures and dynamical systems in one dimension. Birkhäuser Boston, Inc., Boston, MA, 1997, pp. xvi+399. ISBN: 0-8176-4003-7. DOI: 10.1007/978-1-4612-2024-4. URL: https://doi.org/10.1007/978-1-4612-2024-4.

[3] A. Ya. Khinchin. *Continued fractions*. University of Chicago Press, Chicago, Ill.-London, 1964, pp. xi+95.

[4] Dong Han Kim. "The dynamical Borel-Cantelli lemma for interval maps". In: *Discrete and Continuous Dynamical Systems - DISCRETE CONTIN DYN SYST* 17 (Apr. 2007), pp. 891–900. DOI: 10.3934/dcds.2007.17.891.

[5] A. Lasota and James A. Yorke. "On the existence of invariant measures for piecewise monotonic transformations". In: *Trans. Amer. Math. Soc.* 186 (1973), 481–488 (1974). ISSN: 0002-9947,1088-6850. DOI: 10.2307/1996575. URL: https://doi.org/10.2307/1996575.

[6] Carlangelo Liverani. "Decay of correlations". In: *Ann. of Math. (2)* 142.2 (1995), pp. 239–301. ISSN: 0003-486X,1939-8980. DOI: 10.2307/2118636. URL: https://doi.org/10.2307/2118636.

[7] Carlangelo Liverani. "Decay of correlations for piecewise expanding maps". In: *J. Statist. Phys.* 78.3-4 (1995), pp. 1111–1129. ISSN: 0022-4715,1572-9613. DOI: 10.1007/BF02183704. URL: https://doi.org/10.1007/BF02183704.

[8] Dieter H. Mayer. "Continued fractions and related transformations". In: *Ergodic theory, symbolic dynamics, and hyperbolic spaces (Trieste, 1989)*. Oxford Sci. Publ. Oxford Univ. Press, New York, 1991, pp. 175–222. ISBN: 0-19-859685-5.