Stationary measures associated to analytic iterated function schemes

Italo Cipriano*and Mark Pollicott†

November 17, 2016

Abstract

We study how the stationary measure associated to analytic contractions on the unit interval behaves under changes in the contractions and the weights. Firstly we give a simple proof of the fact that the integrals of analytic functions with respect to the stationary measure vary analytically if we perturb the contractions and the weights analytically. Secondly, we consider the special case of affine contractions and we prove a conjecture of J. Fraser in [3] on the Kantorovich-Wasserstein distance between two stationary measures associated to affine contractions on the unit interval with different rates of contraction.

1 Introduction

We begin by presenting a standard definition of a stationary probability measure on the unit interval. Let $T_1, T_2 : \mathbb{R} \to \mathbb{R}$ be real analytic orientation preserving contractions, then we can assume without loss of generality that $T_1, T_2 : [0,1] \to [0,1]$, with a suitable choice of coordinates. Let $g_1, g_2 : [0,1] \to (0,1)$ be real analytic functions with $g_1(x) + g_2(x) = 1$.

*I. Capriano, Department of Mathematics, Warwick University, Coventry CV8 2AH, UK; email: icipriano@dim.uchile.cl

†M. Pollicott, Department of Mathematics, Warwick University, Coventry CV8 2AH, UK.; email: masdbl@warwick.ac.uk

Definition 1.1. *The unique stationary probability measure $\mu = \mu(T_1, T_2, g_1)$ associated to the contractions $\{T_1, T_2\}$ and the weights $\{g_1, g_2\}$, is the unique solution of*

$$\int_0^1 f(x)d\mu(x) = \int_0^1 g_1(x)f \circ T_1(x)d\mu(x) + \int_0^1 g_2(x)f \circ T_2(x)d\mu(x) \qquad (1)$$

for every continuous function $f : [0,1] \to \mathbb{R}$. (See [4]).

Our first result describes the dependence of the stationary probability measure on the contractions and weights.

Theorem 1.2. *Suppose that $T_1^\lambda, T_2^\lambda : [0,1] \to [0,1]$ are real analytic families of contractions and $g^\lambda : [0,1] \to (0,1)$ is a real analytic family of weights, for $\lambda \in (-\epsilon, \epsilon)$. If the map $(-\epsilon, \epsilon) \ni \lambda \mapsto (T_1^\lambda, T_2^\lambda, g^\lambda)$ is real analytic then the associated stationary probability measure $\mu^\lambda = \mu(T_1^\lambda, T_2^\lambda, g^\lambda)$ on the unit interval satisfies that*

$$(-\epsilon, \epsilon) \in \lambda \mapsto \int_0^1 f(x)d\mu^\lambda(x)$$

is real analytic for any real analytic function $f : [0,1] \to \mathbb{R}$.

In the case of even more general C^k contractions and weights there are related results, albeit with some reduced differentiability in the dependence (see [2]). The dependence of the measure is analogous to that of the natural measure associated to an expanding map of the circle. This is a much studied area known as Linear Response.

We next consider a standard notion of distance.

Definition 1.3. *The Kantorovich-Wasserstein distance between two probability measures μ and ν on $[0,1]$ can be defined by*

$$d_{W_1}(\mu, \nu) := \sup \left\{ \int_0^1 f(x)d\mu(x) - \int_0^1 f(x)\nu(x) : \|f\|_{Lip} \leq 1 \right\}, \qquad (2)$$

where $\|f\|_{Lip}$ is the Lipschitz constant of f.

(This is sometimes referred to as the first Kantorovich-Wasserstein distance, due to the use of the L^1-norm, and it is related to another standard definition via the Kantorovich-Rubinstein duality theorem. See [9]).

We recall a special case of analytic contractions.

Example 1.4. *We can consider affine maps*

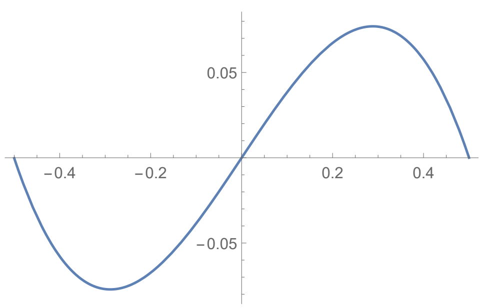
$$T_i(x) = \rho_i x + t_i \quad (i = 1, 2)$$

which are contractions (i.e., $\rho_1, \rho_2 \in (0,1)$) and $0 \le t_1 < t_2 \le 1$. For definiteness we can consider two simple families:

(1) Firstly, we assume that the contractions $T_1(x) = \frac{x}{3}$, $T_2(x) = \frac{2}{3} + \frac{x}{3}$ are fixed but the weights $g^\lambda = \frac{1}{2} + \lambda$ vary (with $-\frac{1}{2} < \lambda < \frac{1}{2}$).

(2) Secondly, we assume that the weight $g = \frac{1}{2}$ is fixed but the contractions $T_1(x) = (\frac{1}{3} + \lambda)x$, $T_2(x) = (\frac{2}{3} + \lambda) + (\frac{1}{3} - \lambda)x$ vary (with $-\frac{1}{3} < \lambda < \frac{1}{3}$).

In each case we can associate a family of stationary measures μ^λ. In Figure 1 we plot $\int \sin(2\pi x) d\mu^\lambda(x)$ for each of these two families. In Figure 2 we plot $\int \cos(2\pi x) d\mu^\lambda(x)$ for each of these two families. These provide a nice illustration of the analytic dependence of the integrals.



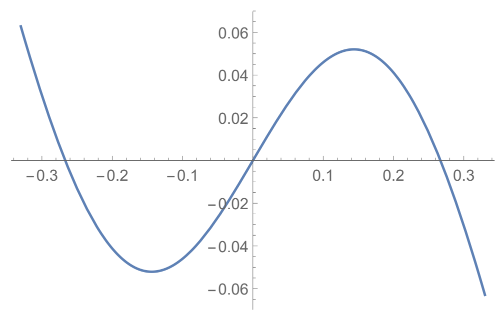
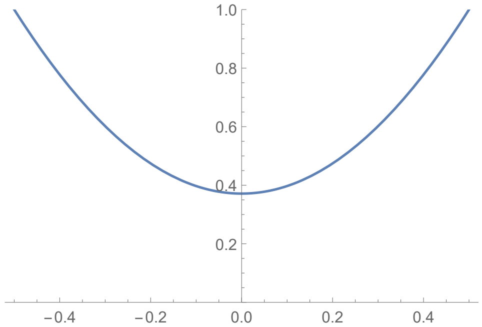


Figure 1: These two plots are for the integral $\int \sin(2\pi x) d\mu^\lambda(x)$ with respect to μ^λ for (1) the first family; and (2) the second family.

In the particular case of affine contractions such that $\rho_1, \rho_2 \in (0,1)$ and $T_1(0,1) \cap T_2(0,1) = \emptyset$, we have an explicit formula for the Kantorovich-Wasserstein distance $d_{W_1}(\mu^p, \mu^q)$ where $\mu^p := \mu(T_1, T_2, p)$ and $\mu^q := \mu(T_1, T_2, q)$ for $p, q \in (0,1)$.

Theorem 1.5. *Let $\mu^p := \mu(T_1, T_2, p)$ and $\mu^q := \mu(T_1, T_2, q)$ for $p, q \in (0,1)$. Then*

$$d_{W_1}(\mu^p, \mu^q) = \left| \frac{(qt_1 + (1-q)t_2)(1 - p\rho_1 - (1-p)\rho_2) - (pt_1 + (1-p)t_2)(1 - q\rho_1 - (1-q)\rho_2)}{(1 - q\rho_1 - (1-q)\rho_2)(1 - p\rho_1 - (1-p)\rho_2)} \right|.$$



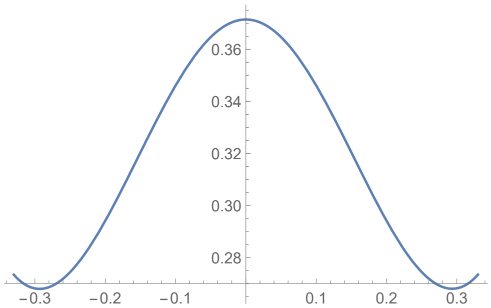


Figure 2: These two plots are for the integral $\int \cos(2\pi x) d\mu^\lambda(x)$ with respect to μ^λ for (1) the first family; and (2) the second family.

This answers in affirmative a conjecture of J. Fraser, from Section 4 of [3]. Corollary 2.6 of that paper contains the special case of Theorem 1.5 when the contractions rates of T_1 and T_2 are the same (i.e. $\rho_1 = \rho_2$). [1]

We are very grateful to both Jairo Bochi and Anthony Quas for suggesting the proofs of two lemmas that we use.

2 Proof of Theorem 1.2

We begin the proof with some notation. For each $\underline{i} = (i_1, \cdots, i_n) \in \{1,2\}^n$ we can consider the unique fixed points $x^\lambda_{\underline{i}} = T^\lambda_{\underline{i}}(x^\lambda_{\underline{i}})$ for the concatenation $T^\lambda_{\underline{i}} := T^\lambda_{i_1} \circ \cdots \circ T^\lambda_{i_n}$, for $(-1,1)$. We denote $|\underline{i}| = n$ and for any $0 \leq j \leq n-1$ we write $\sigma^j \underline{i} = (i_j, \cdots, i_n, i_1, \cdots, i_{j-1})$.

We can use this information on fixed points to define a complex function

$$Z(z,t,\lambda) := \exp\left(- \sum_{n=1}^{\infty} \frac{z^n}{n} \sum_{|\underline{i}|=n} \frac{\prod_{j=0}^{n-1} g^\lambda_{i_j}(x^\lambda_{\sigma^j \underline{i}}) e^{t \sum_{j=0}^{n-1} f\left(x^\lambda_{\sigma^j \underline{i}}\right)}}{1 - \prod_{j=0}^{n-1} \frac{dT^\lambda_{i_j}}{dx}\left(x^\lambda_{\sigma^j \underline{i}}\right)} \right)$$

which converges to an analytic function for $|z|$ sufficiently small.

Lemma 2.1. *The complex function $Z(z,t,\lambda)$:*

[1]There it is also written “It was crucial to our argument that the contraction ratios of both maps were the same . . . We do not believe the situation is hopeless, but would perhaps require a different approach”.

(1) has an analytic extension to $(z,t,\lambda) \in \mathbb{C} \times U \times V$, where U is a neighbourhood of $(-1,1)$ and V is a neighbourhood of $(-\epsilon,\epsilon)$;

(2) has a simple zero at $(z,t) = (1,0)$ and we can write

$$\int_0^1 f(x) d\mu^\lambda(x) = \frac{\frac{\partial Z(1,t,\lambda)}{\partial t}\big|_{t=0}}{\frac{\partial Z(z,0,\lambda)}{\partial z}\big|_{z=1}}. \quad (3)$$

Proof. We begin with the proof of part (1). The analytic dependence in t and z follows from [7]. To extend this to $\lambda \in V$ we observe that there exists $V_{\underline{i}} \supset (-1,1)$ such that $V_{\underline{i}} \ni \lambda \mapsto x_{\underline{i}}^\lambda$ is analytic. Moreover, since the maps are expanding we have that $V = \cap_{\underline{i}} V_{\underline{i}}$ is still an open neighbourhood of $(-1,1)$. In particular, when $Z(z,t,\lambda)$ converges we have the analytic dependence on $\lambda \in V$. Moreover, by a standard application of Hartog's theorem from several complex variable theory we deduce the analyticity stated in the lemma.

For part (2), we can deduce these results from basic ideas in thermodynamic formalism. The complex function $Z(z,t,\lambda)$ has a zero at

$$e^{P(-tf+\log g^\lambda)} = \lim_{n\to+\infty} \left(\sum_{|\underline{i}|=n} \prod_{j=0}^{n-1} g_{i_j}^\lambda \left(x_{\sigma^j \underline{i}}^\lambda\right) e^{t\sum_{j=0}^{n-1} f\left(x_{\sigma^j x_{\underline{i}}}^\lambda\right)} \right)^{\frac{1}{n}}.$$

The expression (3) follows from well known properties of the pressure [6] and the implicit function theorem. $\square$

The neighbourhoods U and V arise from the analyticity hypothesis for the contractions. In particular, this approach is non-constructive and does not provide explicit estimates on the size of these neighbourhoods.

Remark 2.2. *In fact the expression (3) can be used to compute $\int_0^1 f d\mu^\lambda$ with great accuracy. More precisely, if we truncate the formal expansion for $Z(z,t,\lambda) = 1 + \sum_{n=1}^\infty a_n(t,\lambda) z^n$ to those terms with $n \le N$ then we have an approximation which only requires a knowledge of the first 2^N fixed points. However, there exists $\alpha > 0$ so that the error in the approximation is only $O(2^{-\alpha N^2})$. The proof follows the same lines as in [5].*

3 Proof of Theorem 1.5

We begin with a lemma, the statement and proof of which was suggested to us by Jairo Bochi.

Lemma 3.1. *Let μ, ν be probability measures on $[0,1]$. Then*

$$d_{W_1}(\mu,\nu) = \int_0^1 \left(\int_0^x C_{\mu,\nu}(t)dt \right) d(\mu-\nu)(x),$$

where

$$C_{\mu,\nu}(x) := \begin{cases} 1 & \text{if } (\mu-\nu)[x,1] > 0, \\ -1 & \text{if } (\mu-\nu)[x,1] < 0. \end{cases}$$

Proof. Suppose that f with $\|f\|_{\mathrm{Lip}} \le 1$ realises the supremum in $d_{W_1}(\mu,\nu)$. Then $f(x) = \int_0^x g(x)dx$, where $g : [0,1] \to [-1,1]$ is an integrable function. By an application of Fubini's theorem we have

$$\begin{aligned} \int_0^1 f(x)d\mu(x) - \int_0^1 f(x)d\nu(x) &= \int_0^1 f(x)d(\mu-\nu)(x) \\ &= \int_0^1 \int_0^x g(t)dtd(\mu-\nu)(x) \\ &= \int_0^1 \int_t^1 g(t)d(\mu-\nu)(x)dt \\ &= \int_0^1 g(t) \int_t^1 d(\mu-\nu)(x)dt \\ &= \int_0^1 g(t)(\mu-\nu)[t,1]dt. \end{aligned}$$

Because of our assumption that f realises the supremum in $d_{W_1}(\mu,\nu)$, we have that $g(x) = C_{\mu,\nu}(x)$. □

Remark 3.2. *After completing this paper we discovered that Lemma 3.1 was proved independently by Dall'Aglio [1] and Vallender [8] in a more general form: if μ and ν are probability measures on $\mathbb{R}$, then $d_{W_1}(\mu,\nu) = \int_{-\infty}^{\infty} |F(t) - G(t)|dt$, where F and G are the cumulative distribution functions of μ and ν, respectively.*

We have the following lemma in probability theory, which was suggested to us by Anthony Quas.

Lemma 3.3. *Suppose that $p \neq q$, then the function $D : [0,1] \to [0,1]$ defined by $D(x) := (\mu^p - \mu^q)[0,x]$ does not change sign.* [2]

[2] This result does not require that the contractions are affine.

Proof. We assume without lost of generality that $p < q$. We prove the lemma in two steps. Firstly, under the assumption that $T_1[0,1] \cap T_2[0,1] = \emptyset$. Secondly, under the assumption that $T_1(0,1) \cap T_2(0,1) = \emptyset$.

1. Assume that T_1 and T_2 satisfy $T_1[0,1] \cap T_2[0,1] = \emptyset$. Let $\Sigma_2 := \prod_{n=0}^{\infty}\{1,2\}$ and $\Sigma_3 := \prod_{n=0}^{\infty}\{1,2,3\}$. Let Λ denote the limit set for T_1, T_2, then there is a natural monotone bijection $f : \Sigma_2 \to \Lambda$ given by $f(x) = \lim_{n\to+\infty} T_{x_0} \circ \cdots \circ T_{x_n}([0,1])$. In particular, for the Bernoulli measures $\mu_p := \prod_{n=0}^{\infty}(p, 1-p)$ and $\mu_q := \prod_{n=0}^{\infty}(q, 1-q)$ on Σ_2 we have that $\mu^p = \mu_p \circ f^{-1}$ and $\mu^q = \mu_q \circ f^{-1}$, respectively. We can define maps $\pi_p : \Sigma_3 \to \Sigma_2$ by

$$\pi_p\left((i_n)_{n=0}^{\infty}\right) = (j_n)_{n=0}^{\infty} \text{ where } j_n = \begin{cases} 1 & \text{if } i_n = 1 \text{ or } i_n = 2 \\ 2 & \text{if } i_n = 3 \end{cases}$$

and $\pi_q : \Sigma_3 \to \Sigma_2$ by

$$\pi_q\left((i_n)_{n=0}^{\infty}\right) = (j_n)_{n=0}^{\infty} \text{ where } j_n = \begin{cases} 1 & \text{if } i_n = 1 \\ 2 & \text{if } i_n = 2 \text{ or } i_n = 3. \end{cases}$$

Equip Σ_3 with the Bernoulli measure $\mu_{pq} := \prod_{n=0}^{\infty}(p, q-p, 1-q)$. Then $\mu_p = \mu_{pq} \circ \pi_q^{-1}$ and $\mu_q = \mu_{pq} \circ \pi_p^{-1}$. Therefore, we can define a map $g : \Lambda \to \Lambda$ such that for every $x \in \Lambda$

$$g(x) \in f \circ \pi_q \circ \pi_p^{-1} \circ f^{-1}(x)$$

and

$$\mu^q[0,x] \geq \mu^p[0, g(x)].$$

We have that $g(x) \geq x$ for every $x \in \Lambda$, because f is monotone and for every $y \in \Sigma_2$, for every $z \in \pi_q \circ \pi_p^{-1} y$, we have $z \geq y$ with respect to the lexicographic order. This allows to conclude the result.

2. Assume that T_1 and T_2 satisfy $T_1(0,1) \cap T_2(0,1) = \emptyset$ and without lost of generality that $T_1(1) = T_2(0) = \{x_0\}$, $x_0 \in (0,1)$. We define $\{1,2\}^* := \cup_{n\in\mathbb{N}}\{1,2\}^n$ and for $\underline{i} = (i_0, i_1, \ldots, i_n) \in \{1,2\}^*$ we define $T_{\underline{i}} := T_{i_0} \circ T_{i_1} \circ \cdots \circ T_{i_n}$. We now define the orbit of x_0 by $\mathrm{orb}(x_0) := \{T_{\underline{i}}(x_0) : \underline{i} \in \{1,2\}^*\} \cup \{x_0\}$. Let Σ_2, Σ_3 and Λ be defined as before. We

denote $\Sigma_2' := \prod_{i=0}^{\infty}\{2,3\}$, $\tilde{\Sigma}_2 := \{x \in \Sigma_2 : \forall n \in \mathbb{N}, \sigma^n x \notin \{1^\infty, 2^\infty\}\}$, $\tilde{\Sigma}_3 := \{x \in \Sigma_3 : \forall n \in \mathbb{N}, \sigma^n x \notin \Sigma_2 \cup \Sigma_2'\}$ and $\tilde{\Lambda} := \Lambda \setminus \mathrm{orb}(x_0)$. There is a monotone bijection $\tilde{f} : \tilde{\Sigma}_2 \to \tilde{\Lambda}$ given by $\tilde{f} := f|_{\tilde{\Sigma}_2}$. For the Bernoulli measures μ_p, μ_q on Σ_2 and μ_{pq} on Σ_3 we have that $\mu_p(\tilde{\Sigma}_2) = \mu_q(\tilde{\Sigma}_2) = \mu_{pq}(\tilde{\Sigma}_3) = 1$. We notice that $\mu^p|_{\tilde{\Lambda}} = \mu_p \circ \tilde{f}^{-1}$, $\mu^q|_{\tilde{\Lambda}} = \mu_q \circ \tilde{f}^{-1}$ and $\mu^p(\tilde{\Lambda}) = \mu^q(\tilde{\Lambda}) = 1$, where the last equality comes from the fact that μ^p and μ^q are non-atomic measures and the set $\mathrm{orb}(x_0)$ is denumerable. We define the maps $\tilde{\pi}_p : \tilde{\Sigma}_3 \to \tilde{\Sigma}_2$ and $\tilde{\pi}_q : \tilde{\Sigma}_3 \to \tilde{\Sigma}_2$ by $\tilde{\pi}_p = \pi_p|_{\tilde{\Sigma}_3}$ and $\tilde{\pi}_q = \pi_q|_{\tilde{\Sigma}_3}$, respectively. Then $\mu_p|_{\tilde{\Sigma}_2} = \mu_{pq} \circ \tilde{\pi}_q^{-1}$ and $\mu_q|_{\tilde{\Sigma}_2} = \mu_{pq} \circ \tilde{\pi}_p^{-1}$. The map g defined in the previous part of the proof satisfies that $g|_{\tilde{\Lambda}} : \tilde{\Lambda} \to \tilde{\Lambda}$, then for every $x \in \tilde{\Lambda}$ we have $\mu^q[0,x] \geq \mu^p[0,x]$. The fact that $\mu^p(\tilde{\Lambda}) = \mu^q(\tilde{\Lambda}) = 1$ allows to conclude the result.

□

We can now prove Theorem 1.5.

Proof of Theorem 1.5. Suppose without loss of generality that $p > q$ and $t_1 < t_2$. We can use Lemmas 3.1 and 3.3 with $C_{\mu^p,\mu^q} = -1$ to show that

$$d_{W_1}(\mu^p, \mu^q) = -\int_0^1 x d(\mu^p - \mu^q)(x).$$

To finish the proof it is enough to compute $\int_0^1 x d\mu^p(x)$, for this we have

$$\begin{aligned}
\int_0^1 x d\mu^p(x) &= p\int_0^1 T_1(x) d\mu^p(x) + (1-p)\int_0^1 T_2(x) d\mu^p(x) \\
&= p\int_0^1 (\rho_1 x + t_1)\, d\mu^p(x) + (1-p)\int_0^1 \rho_2 x + t_2 d\mu^p(x) \\
&= p\rho_1 \int_0^1 x d\mu^p(x) + pt_1 + (1-p)\rho_2 \int_0^1 x d\mu^p(x) + (1-p)t_2 \\
&= (p\rho_1 + (1-p)\rho_2)\int_0^1 x d\mu^p(x) + (pt_1 + (1-p)t_2)\,,
\end{aligned}$$

then

$$\int_0^1 x d\mu^p(x) = \frac{pt_1 + (1-p)t_2}{1 - (p\rho_1 + (1-p)\rho_2)}.$$

Finally,

$$d_{W_1}(\mu^p, \mu^q) = \frac{qt_1 + (1-q)t_2}{1 - (q\rho_1 + (1-q)\rho_2)} - \frac{pt_1 + (1-p)t_2}{1 - (p\rho_1 + (1-p)\rho_2)},$$

as claimed. □

References

[1] G. Dall'Aglio, Sugli estremi dei momenti delle funzioni di ripartizione doppia, *Ann. Scuola Norm. Sup. Pisa*, Cl. Sci. 3:1 (1956), 33-74.

[2] I. Cipriano, The smoothness of the stationary measure, *preprint* (2016), arXiv:1607.02680v1.

[3] J. Fraser, First and second moments for self-similar couplings and Wasserstein distances, *Mathematishe Nachrichten*, 288 (2015), 2028-2041.

[4] J. Hutchinson, Fractals and self similarity, *Indiana Univ. Math. J.*, 30 (1981), 713-747.

[5] O. Jenkinson and M. Pollicott, Calculating Hausdorff dimension of Julia sets and Kleinian limit sets, *American Journal of Mathematics*, 124 (2002), 495-545.

[6] W. Parry and M. Pollicott, Zeta Functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* 187-188 (1990), 1-268.

[7] D. Ruelle, Zeta functions for expanding maps and Anosov flows. *Inventiones Math.* 34 (1976), 231-242.

[8] S.S. Vallender, Calculation of the Wasserstein distance between probability distributions on the line, *Teor. Veroyatnost. i Primenen.*, 18:4 (1973), 824-827.

[9] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics, Vol. 58, *Amer. Math. Soc.* 2003.