Existence, mixing and approximation of invariant densities for expanding maps on $\mathbb{R}^r$

Rua Murray
Department of Mathematics, University of Waikato,
Private Bag 3105, Hamilton NEW ZEALAND.

Final version: June 14, 1999

Abstract

This paper generalises Góra and Boyarsky's [16] bounded variation (BV) approach to the ergodic properties of expanding transformations, and analyses the convergence of Ulam's method for the numerical approximation of absolutely continuous invariant measures. We first prove an existence theorem for BV invariant densities for piecewise expanding maps on subsets of $\mathbb{R}^r$; the maps must be C^2, but may have infinitely many branches and need not be Markov. Under an additional "onto" assumption, explicit bounds on the spectral gap in the associated Perron–Frobenius operator are proved. The corresponding contraction rates are in the BV norm, rather than a projective metric. With this quantitative information, we are then able to prove convergence and explicit upper bounds on the approximation error in Ulam's method for approximating invariant measures. Because the BV approach is rather concrete, the methods of this paper can be applied in practice; this is illustrated by an application of the main results to the Jacobi–Perron transformation on $\mathbb{R}^2$.

Keywords: Absolutely continuous invariant measures, decay of correlations, approximation of invariant measures, Jacobi–Perron transformation

Introduction

This paper concerns the construction of probability measures which are absolutely continuous with respect to Lebesgue measure m and invariant under the action of a piecewise expanding map on a subset of $\mathbb{R}^r$. Such measures are called absolutely continuous invariant measures (acims), and are determined by their corresponding *invariant densities*. A map T on a domain $A \subset \mathbb{R}^r$ is *piecewise expanding* if there exists a constant $\lambda > 1$ and a countable partition $\{B_\alpha\}$ of A such that each $T|_{B_\alpha}$ is one–to–one, and $|T(x) - T(y)| > \lambda|x - y|$ whenever x, y are in the same B_α. In this paper we give an existence theorem for invariant densities for such maps subject to λ being large enough, T being locally C^2 and the image sets $T(B_\alpha)$ not being too wild. T may have infinitely many branches. The proof uses the multi–dimensional bounded variation (BV) approach of Góra and Boyarsky [16]. Next, some explicit quantitative bounds are derived for the exponential rate at which arbitrary densities approach the invariant measure under iterates of the Perron–Frobenius operator. In contrast to other approaches [12, 27, 33], these estimates are in the BV–norm, and lead to explicit quantitative error bounds on Ulam's method [35] for the numerical approximation of acims. Related, and better known decay of correlation results [12, 27, 33] are, so far as we know, not sufficient to obtain such bounds. In the final section of the paper, the approach is applied to the classical Jacobi–Perron transformation.

Motivation and pedagogy

This work is motivated in part by a specific problem: to construct an acim for the Jacobi–Perron (JP) transformation on $\mathbb{R}^2$ [34]. The JP transformation has an infinite Markov partition, and since each branch has bounded distortion, the existence of an invariant density is known; however, no formula for it exists [2]. A natural question is therefore whether the density can be approximated numerically. Recent work on Ulam's method [35] for the numerical approximations of acims in multi–dimensions [11] suggests a positive answer to this question. Ulam's method consists in replacing the

Perron–Frobenius operator $\mathcal{L}$ by a certain finite–dimensional approximation; the fixed point of the finite dimensional operator is then an approximation to an invariant density for T. One needs several ingredients to analyse Ulam's method: (i) existence of an invariant density in a reasonably well behaved subspace of $L^1(A)$; (ii) explicit control of the effects of finite–dimensional approximations to that subspace; (iii) knowledge of the "stability" of the acim to perturbations. The BV approach from [16] almost satisfies the first of these requirements, but contains a more or less technical obstruction for transformations with infinitely many branches. By considering the geometry of the images of one–to–one branches of T, rather than one–to–one domains themselves, and making a few other adjustments to the approach in [16], we obtain a suitable solution to problem (i) in Section 1 below. The rest of the paper is then devoted to problems (ii) and (iii). The central tool is the derivation of explicit quantitative bounds on the non–peripheral part of the spectrum of $\mathcal{L}$, and this is the most substantial part of the paper. Before giving a precise statement of the main results, we give a few historical comments.

Historical comments on the acim problem

Invariant measures are the fundamental objects in the ergodic theory of dynamical systems, and are therefore of considerable importance in statistical descriptions of complicated dynamics [23]. For piecewise expanding maps on domains in $\mathbb{R}^r$, an important question is: how much expansion is necessary to guarantee the existence of an acim? When $r = 1$, Lasota and Yorke [24] proved that $\lambda > 1$ is sufficient when all branches of the map are C^2. Subsequent work by Rychlik [31] (and others) required less regularity of each branch, and in all cases $\lambda > 1$ suffices.

In higher dimensions, considerably more delicacy is required. For piecewise complex analytic maps of the plane, Keller [20] proved that $\lambda > 1$ is sufficient to ensure the existence of an acim. Buzzi [10] has recently proved the existence of acim under the same expansivity assumption for maps which are $\mathbb{R}$–analytic, but not holomorphic. In higher dimensions (r not necessarily 1 or 2) and without analyticity, the situation is less satisfactory: one generally requires "sufficient expansivity" conditions, where λ is bigger than some number determined by either the shape of the boundaries of one–to–one branches [16], or the density with which ϵ–neighbourhoods of these boundaries "fill out" the space [33, 10]. These "boundary effects" arise from discontinuities of the invariant densities on $\cup_{n,\alpha} T^n(\partial B_\alpha)$. When T admits a Markov partition, boundary effects can be avoided (for example [28]). At least two approaches to the non–Markov problem are present in the literature: one based on bounded mean local oscillation [5, 6, 33, 10] of densities, and another on bounded variation in the sense of distributions [16, 1]. In both cases, a suitable relatively compact subset of $L^1(A)$ is used, and a quasi–compactness argument based on the Ionescu–Tulcea and Marinescu ergodic theorem [18] yields a spectral gap for the Perron–Frobenius operator; this is important for the stochastic stability and numerical approximation of acims [21]. More detailed commentary can be found in [5, 10].

In each of the above approaches, the choice of function space determines the precise formulation of the "sufficient expansivity" condition. Despite these attempts, a general answer as to when $\lambda > 1$ is sufficient to admit an acim for piecewise C^2 maps on $\mathbb{R}^r$ is still lacking. We prove below that if the range structure of T is uniformly non-degenerate then $\lambda > 1$ is sufficient (Theorem 2). However, for general piecewise expanding maps, Theorem 2 retains a "sufficient expansivity" condition which is constrained by the geometry of boundaries of images of one–to–one branches. This is a consequence of the technology we exploit: the behaviour of BV. Nonetheless, the BV approach has the virtue of concreteness, and this allows us to write the statements of the theorems in a way which makes the hypotheses verifiable for specific transformations. In the final section of the paper this is illustrated with an application to the JP transformation.

Main results

Let $A \subset \mathbb{R}^r$ be a bounded subset of $\mathbb{R}^r$, consisting of an at most countable union of connected components, where the boundary of each component has finite $(r-1)$–dimensional measure. We consider transformations $T : A \to A$ satisfying the following conditions:

One–to–one branches: There exists a partition (possibly countably infinite) $\xi = \{B_\alpha\}$ of A such that T is C^2 and non–singular on each $int(B_\alpha)$, and the boundary ∂B_α of each B_α is piecewise Lipschitz with finite $(r-1)$–dimensional measure. The inverse branches will be denoted by $T_\alpha^{-1} : TB_\alpha \to B_\alpha$. The inverse branches of T^n will be denoted

$$T^{-n}_{\alpha^{(n)}} = T^{-n}_{\alpha_n\cdots\alpha_1} \triangleq T^{-1}_{\alpha_n} \circ \cdots \circ T^{-1}_{\alpha_1}$$

where each α_i denotes a one–to–one branch of T.

Expansivity: There exist global constants $C_1 \geq 0$, $\lambda > 1$ such that

$$\left|DT^{-n}_{\alpha(n)} \cdot v\right| \leq C_1 \lambda^{-n} |v| \tag{1}$$

where $DT^{-n}_{\alpha(n)}$ is the Jacobian matrix of $T^{-n}_{\alpha(n)}$.

Bounded Distortion: There exists a constant $C_2 \geq 0$ such that

$$\left| \frac{\nabla \left(\det(DT_\alpha^{-1})\right)}{\det(DT_\alpha^{-1})} \right| \leq C_2. \tag{2}$$

We now give geometric conditions which allow the formulation of our "sufficient expansivity" conditions.

Definition. A connected set $S \subset \mathbb{R}^r$ with piecewise Lipschitz boundary of finite $(r-1)$–dimensional measure whose faces meet at angles strictly bounded away from zero will be called *non–degenerate* provided that both $a(S) > 0$ (where $a(S)$ is the sine of half the smallest interior angle at corners of S), and the following geometric condition holds:

Geometric condition *Let $a' \leq a(S)$. A number $\epsilon \geq 0$ will be called* small enough for (S, a') *if (i) $S \setminus B_\epsilon(\partial S) \neq \emptyset$ ($B_\epsilon(X)$ denotes the ϵ–neighbourhood of a set X), and (ii) each line segment L_x connecting $x \in \partial S$ to the nearest point in $\partial(B_\epsilon(\partial S)) \cap S$ is wholly contained in S and makes an angle with ∂S at x whose sine is bounded below by a'. Then S is* non–degenerate *if for every $a' < a(S)$,*

$$\epsilon(S, a') = \sup_{\epsilon \geq 0} \{\epsilon : \epsilon \text{ is small enough for } (S, a')\} > 0.$$

A collection $\{S_\alpha\}$ of subsets will be called *uniformly non–degenerate* if each S_α is non–degenerate, and both $a = \inf_\alpha a(S_\alpha) > 0$ and $\inf_\alpha \epsilon(S_\alpha, a') > 0$ for all $a' < a(S)$. □

By following the reasoning from [16] we first prove:

Theorem (GB^+) *Suppose that the images of one–to–one branches $\{TB_\alpha\}$ are uniformly non–degenerate. Then*
(i) if $C_1 \lambda^{-1}(1 + 1/a) < 1$ then T possesses an acim;
(ii) if $C_1 \lambda^{-1}(1 + 1/a) < 1/\sqrt{r}$ then Ulam's method converges in $L^1(A)$;
(iii) if the collection $\cup_n \{T^n B_{\alpha(n)}\}$ is uniformly non–degenerate and $\lambda > 1$, then parts (i) and (ii) of the theorem apply to an iterate of T.

This result is called (GB^+) because of the close resemblance to Góra and Boyarsky's existence result [16]; the main difference is that our sufficient expansivity condition is on one–to–one images, and therefore allows infinitely many such pieces. Parts (i) and (iii) of the theorem are proved as Theorem 2 in section 1 below. Part (ii) of the theorem is given in section 3. In section 2, we prove the more substantial result:

Mixing theorem *If the condition in (GB^+) (i) holds for some T^{n_1}, and for each k there exists a large enough non–degenerate set W_k such that $W_k \subset T^{kn_1}(B_{\alpha(kn_1)})$ for all one–to–one branches of T^{kn_1}, then there exist constructively defined constants $C < \infty$ and $\rho < 1$ such that*

$$\left\|\mathcal{L}^n(f - \textstyle\int_A f\, dm)\right\|_{BV} \leq C\rho^n \left\|f - \textstyle\int_A f\, dm\right\|_{BV}$$

for all $n > 0$ where $\mathcal{L}$ is the Perron–Frobenius operator for T, and $\|\cdot\|_{BV}$ denotes the BV norm.

The mixing theorem is proved as Theorems 3 and 4. The importance of the "onto" assumption is also discussed. The main argument has already appeared for one–dimensional maps [30], and comparison with that paper may be helpful. In section 3, the results of the first two sections can be applied to obtain explicit quantitative error bounds for Ulam's method. In section 4, all the results of the paper are illustrated with the Jacobi–Perron transformation [34].

Remark. The definition of non–degeneracy adopted here is a reworking of [16], where a more detailed description of the constant $a(S)$ can be found. The geometric definition of $\epsilon(S, a')$ is relatively easy to compute for reasonably well behaved sets, such as star–shaped domains. The utility of the definition can be seen in 4.2, where computations are made for an explicit example. □

1 Existence of invariant densities

In this section we follow Góra and Boyarsky's multi–dimensional bounded variation approach to obtain an existence theorem for acims. While the approach used here is similar to the one from [16], we need to make a number of small changes to get a result which is explicitly applicable to the JP transformation. Therefore, much of this section is rather standard; the aim has been to provide sufficient detail to elucidate applications of the theorems. Consequently, in this section, as in the remainder of the paper, we are rather careful to give formulas for the various constants required in the theorems.

Throughout the paper, m denotes the Lebesgue measure on $\mathbb{R}^r$, $T : A \to A$ ($A \subset \mathbb{R}^r$), and we assume $m(A) = 1$. The characteristic (or indicator) function of a subset $E \subset \mathbb{R}^r$ is denoted by χ_E, and $|\cdot|$ denotes Euclidean distance on $\mathbb{R}^r$.

1.1 Multi-dimensional bounded variation

As in [16], we use the definition of BV from [14]:

Definition. (**Bounded variation in** $\mathbb{R}^r$) Let $A \subset \mathbb{R}^r$ be an open set with piecewise Lipschitz boundary, and let $C_0^1(A;\mathbb{R}^r)$ denote the collection of compactly supported smooth vector fields on A. Let div denote the divergence and for each $f \in L^1(A)$ put

$$V_A(f) = \sup\left\{\int_A f(x)\mathrm{div}\ w(x)\, dm(x) : w \in C_0^1(A;\mathbb{R}^r), |w(x)| \leq 1\ \forall x \in A\right\},$$

Then $V_A(f)$ is called the *variation* of f over A. If $V_A(f) < \infty$, then f is said to have *bounded variation over* A and it is usual to write $f \in BV(A)$. The set of $f \in BV(A)$ equipped with the norm

$$\|f\|_{BV} = \|f\|_{L^1} + V_A(f)$$

is a Banach space [14, 1.12]. □

$V_A(\cdot)$ generalises most familiar properties of one–dimensional variation, including subadditivity. The computation of $V_A(\cdot)$ can be elucidated by considering the following class of functions:

Definition. A function $f \in BV(A)$ is called *piecewise* C^1 if there exists a countable partition of A into subsets $\{A_\beta\}$ such that f is differentiable on each $int(A_\beta)$, and ∂A_β can be written as a countable collection of piecewise Lipschitz surfaces of finite $(r-1)$–dimensional measure. □

If f is piecewise C^1, contributions to $V_A(f)$ arise in two ways: as integrals of $|\nabla f|$ over the open sets on which f is C^1, and from discontinuities across the boundaries of such sets. To formalise this, we require a little more standard notation: Suppose that Γ is a piece of Lipschitz $(r-1)$ dimensional surface across which f is discontinuous. If $x \in \Gamma$ and ϵ is sufficiently small, then $B_\epsilon(x)$ (the ball of radius ϵ at x) partitions $B_\epsilon(x)$ into two connected components. Denote these by B_ϵ^+ and B_ϵ^-. Then, following [14, 2.5]:

Definition. (**Trace of a function on an oriented surface**) For each point $x \in \Gamma$, put

$$\left(tr_\Gamma^+ f\right)(x) = \lim_{\epsilon\to 0}\frac{\int_{B_\epsilon^+} f\, dm}{m(B_\epsilon^+)} \quad \text{and} \quad \left(tr_\Gamma^- f\right)(x) = \lim_{\epsilon\to 0}\frac{\int_{B_\epsilon^-} f\, dm}{m(B_\epsilon^-)}.$$

(If $x \in \partial A$, extend f to $\mathbb{R}^r$ by setting $f|_{\mathbb{R}^r\setminus A} = 0$.) □

Thus, for example, if $f \in C^1(A;\mathbb{R})$ and $E \subset A$ has piecewise Lipschitz boundary, let H denote the $(r-1)$–dimensional Hausdorff measure on ∂E. Then

$$V_A(f) = \int_{int(E)} |\nabla f|\, dm + \int_{\partial E\setminus\partial A} \left|tr_{\partial E}^+ f - tr_{\partial E}^- f\right|\, dH.$$

Finally, we define a family of *cones* in L^1:

Definition. Let the subset $A \subset \mathbb{R}^r$ be fixed. For each $M > 0$ let

$$\mathcal{C}_M = \left\{0 \leq f \in L^1(A) : V_A(f) \leq M\|f\|_{L^1}\right\}.$$

Each $\mathcal{C}_M$ is a *cone of uniformly bounded variation*. Convexity follows immediately from the subadditivity of $V_A(\cdot)$. □

Remarks:

1. In fact, $V_A(\cdot)$ is lower semicontinuous with respect to L^1 [14, 1.9], every $f \in BV(A)$ can be approximated by a sequence $\{f_j\} \subset \mathcal{C}^\infty(A;\mathbb{R})$ such that $\|f-f_j\|_{L^1} \to 0$ and $V_A(f_j) \to V_A(f)$ [14, 1.17], and subsets of uniformly bounded BV–norm are relatively compact in $L^1(A)$ [14, 1.19]. In view of these facts, all the estimates in this paper are done for piecewise C^1 functions; the case of general BV functions follows by an approximation argument.
2. The cones $\mathcal{C}_M$ are a multi–dimensional generalisation of those introduced by Liverani [27], and exploited in the one–dimensional version of the current paper [30].
3. For $M > 0$, subsets $\mathcal{C}_M|_{\{f:\|f\|_{L^1}=const\}}$ are compact in L^1. This follows from the relative compactness of $BV \subset L^1$ since $\|f\|_{BV} = \|f\|_{L^1} + V_A(f) \le (1+M)\|f\|_{L^1} \le (1+M)const$ whenever $f \in \mathcal{C}_M|_{\{f:\|f\|_{L^1}=const\}}$. This fact is of fundamental importance in the construction of invariant densities. Indeed, the main idea below is to find an M_0 such that $\mathcal{C}_{M_0}$ is preserved by the Perron–Frobenius operator for the dynamics.
4. Unfortunately, the BV class admits some rather wild functions: for example, a positive $BV(\mathbb{R}^r)$ function can have infinite essential supremum, or even have non-zero integral, while its essential infimum on every open set is zero. Fortunately, BV functions are L^1–close to reasonably regular functions, and this fact is exploited in section 2 below. □

1.2 The Perron–Frobenius operator and preliminary BV estimates

We next recall the Perron–Frobenius operator for T whose fixed points are densities of acims. We will make some standard estimates about the behaviour of $V_A(f)$ under iteration by the operator. The idea is to obtain a Lasota–Yorke type inequality, thus implying the invariance of a suitably chosen cone $\mathcal{C}_M$. The existence of an acim then follows almost immediately.

Since T is piecewise non-singular, we may define the *Perron–Frobenius operator* $\mathcal{L}$ for T (see [23]). For $f \in L^1(A)$ we have the formula

$$\mathcal{L}f(x) = \sum_\alpha \frac{f \circ T_\alpha^{-1}}{\left|\det(DT) \circ T_\alpha^{-1}\right|} \chi_{T(B_\alpha)}. \tag{3}$$

We assume throughout that T is fixed, and therefore do not denote the dependence of $\mathcal{L}$ on T. Indeed, the only other transformation we will deal with will be T^n (for some n); then, the corresponding Perron–Frobenius operator is $\mathcal{L}^n = \mathcal{L} \circ \cdots \circ \mathcal{L}$ (n times). Consequently, the formula for the action of $\mathcal{L}^n$ on L^1 functions is similar to (3), except that the sum is taken over inverse branches $T_{\alpha(n)}^{-n}$ of T^n.

The operator $\mathcal{L}$ is a Markov operator on L^1; that is $\|\mathcal{L}f\|_{L^1} \le \|f\|_{L^1}$ with equality if $f \ge 0$.

Lemma 1.1 *For f which are C^1, and T satisfying (1) and (2),*

1.

$$\begin{aligned} V_A(\mathcal{L}f) \quad \le \quad & \sum_\alpha \left(\int_{int(TB_\alpha)} \left| \nabla \left(\frac{f \circ T_\alpha^{-1}}{\det DT \circ T_\alpha^{-1}} \right) \right| dm \right. \\ & \left. + \int_{\partial(TB_\alpha)} \left| tr^+_{\partial(TB_\alpha)} \left(\frac{f \circ T_\alpha^{-1}}{\det DT \circ T_\alpha^{-1}} \right) \right| dH_{\partial TB_\alpha} \right) \end{aligned}$$

(we have adopted the convention that each $\partial(TB_\alpha)$ is oriented towards the interior of TB_α);

2. for each α,

$$\int_{int(TB_\alpha)} \left| \nabla \frac{f \circ T_\alpha^{-1}}{\det DT \circ T_\alpha^{-1}} \right| dm \le C_1 \lambda^{-1} \int_{int(B_\alpha)} |\nabla f|\, dm + C_2 \int_{B_\alpha} |f|\, dm.$$

Proof. See appendix B. □

Next, the trace terms appearing in Lemma 1.1 (1) must be dealt with. This is where the geometry of the one–to–one pieces contributes to the "sufficient expansivity" condition mentioned in the introduction. Here, we borrow a geometric lemma from [16].

Lemma 1.2 (Góra–Boyarsky type inequality [16, 15]) *Let S be a non–degenerate set. For every $a' < a(S)$ there exists $\delta' > 0$ such that*

$$\int_{\partial S} \left| tr^+_{\partial S} g \right| dH \le \frac{1}{a'} \left(\int_{int(S)} |\nabla g|\, dm + \frac{1}{\delta'} \int_S |g|\, dm \right)$$

for any $g \in C^1(S;\mathbb{R})$. The constant δ' depends on a' and $\epsilon(S,a')$ and may approach 0 as $a' \to a(S)$.

Rather than give a proof of Lemma 1.2, we note that the construction from [16] can be realized in specific cases to give explicit bounds on the numbers $a(S)$ and δ'. In Section 4.2 below we do this for a triangle in $\mathbb{R}^2$. The proof can be duplicated for general non–degenerate sets.

1.3 Lasota–Yorke type inequalities

A Lasota–Yorke type inequality now follows easily from Lemmas 1.1 and 1.2 under sufficient expansivity assumptions. It is then a standard matter to deduce the existence of an acim with a density of bounded variation. First of all, we collect together the lemmas:

Proposition 1.3 *Let T satisfy (1) and (2) and suppose that $\{TB_\alpha\}$ is a collection of uniformly non–degenerate sets. Let $a' \le a(T(B_\alpha))$ for all α. Then there exists $\delta' > 0$ such that*

$$V_A(\mathcal{L}f) \le C_1\lambda^{-1}(1+1/a')V_A(f) + \left(C_2(1+1/a') + 1/(a'\delta')\right)\|f\|_{L^1}.$$

Proof. See appendix B. □

To produce a Lasota–Yorke type inequality in which the coefficient of $V_A(f)$ is strictly less than one, it is necessary to impose a "sufficient expansivity" condition on T. We give two ways to achieve this:

DEFINITION. (**Non-degenerate range structure**) The collection of sets $\cup_{n=1}^{\infty}\{T^n B_{\alpha^{(n)}}\}$ will be called the *range structure* of T. The range structure will be called *uniformly non–degenerate* if each collection $\{T^n B_{\alpha^{(n)}}\}$ is uniformly non–degenerate, and $\inf_{n\ge 1}\inf_{\alpha^{(n)}} a(T^n B_{\alpha^{(n)}}) > 0$. □

Theorem 1 (Lasota–Yorke type inequality) *Let T satisfy (1) and (2), and let $f \in BV(A)$.*

(i) Suppose that $\{TB_\alpha\}$ are uniformly non–degenerate sets and that λ is large enough that $C_1\lambda^{-1}(1+1/a) < 1$ where $a = \inf_\alpha a(TB_\alpha)$. Then for every $\sigma \in (C_1\lambda^{-1}(1+1/a), 1)$ there exists a constant $K_\sigma > 0$ such that

$$V_A(\mathcal{L}f) \le \sigma V_A(f) + K_\sigma\|f\|_{L^1}.$$

(ii) If T has a uniformly non–degenerate range structure, then for every $\sigma \in (0,1)$ there exists $n_\sigma > 0$ and a constant K_σ such that

$$V_A(\mathcal{L}^{n_\sigma} f) \le \sigma V_A(f) + K_\sigma\|f\|_{L^1}.$$

Moreover, the constants K_σ, n_σ have explicit formulas.

Proof. *(i)* Pick a' such that $\sigma = C_1\lambda^{-1}(1+1/a')$ and apply Proposition 1.3. *(ii)* Let $0 < a' < a(T^n B_{\alpha^{(n)}})$ for all range sets $\alpha^{(n)}$. Pick $n_\sigma = \min\{n : C_1\lambda^{-n}(1+1/a') < \sigma\}$. Now, since $\{T^{n_\sigma} B_{\alpha^{(n_\sigma)}}\}$ is a uniformly non–degenerate collection, let δ_σ be such that Lemma 2.1 holds with δ_σ replacing δ' for each set $S = T^{n_\sigma} B_{\alpha^{(n_\sigma)}}$. Since we are working with T^{n_σ}, replace in the conclusion of Proposition 1.3, $C_1\lambda^{-1}$ by $C_1\lambda^{-n_\sigma}$, C_2 by $C_2\frac{1-\lambda^{-n_\sigma}}{1-\lambda^{-1}}$ and δ' by δ_σ, thus obtaining a formula for K_σ. □

REMARKS

1. In the derivation of the Lasota–Yorke inequality above, we followed the approach of Góra and Boyarsky [16], except that we have applied the trace inequality to image sets TB_α, rather than the one–to–one domains B_α. This approach resolves a technical difficulty ([16, Lemma 1]), and allows more general transformations, as the images of one–to–one branches may be uniformly non–degenerate, even if the original domains are not.
2. The formulation of "sufficient expansivity" in terms of one–to–one images allows infinitely many branches (not possible in [16]), but it is at the price of a "large images" condition.
3. The uniform non–degeneracy of the range structure may in general be hard to obtain. One special case is the following: T is said to have a *finite range structure* if the range structure contains only finitely many sets. That is there exists a finite collection $\{U_1, \ldots U_k\}$ of subsets of A such that for every $\alpha^{(n)}$, $T^n B_{\alpha^{(n)}} = U_i$ for some $i = i(\alpha^{(n)})$. By introducing the finite range structure condition, Yuri [36, 37] has shown the existence and decay of correlations for invariant densities for large classes of transformations. While the BV approach does not require this condition (as seen in the first part of Theorem 1 below), it does play a practical role in quantitative estimates on the invariant density, as well as in computational approximations. If T has a finite range structure, and each range set U_i is non–degenerate, then Theorem 2 (ii) applies. □

We now complete this section with a simple example illustrating the advantages of looking at image sets in the BV approach:

EXAMPLE (IMPORTANCE OF FINITE RANGE STRUCTURE) Let $A = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ be the usual 2–torus, and let

$$T : \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto P \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad (\text{mod } 1),$$

where

$$P = \begin{cases} \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} & \text{if } 1 > x_1, x_2 \geq 1/2, \\ \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix} & \text{if } 0 \leq x_2 < x_1 < 1/2, \\ \begin{bmatrix} 2 & 0 \\ -2 & 2 \end{bmatrix} & \text{if } 0 \leq x_1 < x_2 < 1/2. \end{cases}$$

Suppose now that we attempt to construct a Lasota–Yorke inequality by applying Lemma 1.2 to one–to–one *domains* of T, rather than their images. The smallest interior angle in the corner of a one–to–one branch of T is $\pi/4$, and under inverse iteration this shrinks as the angle between

$$\left(\begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix}^{-1} \right)^n \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \left(\begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix}^{-1} \right)^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

One can check that $a(n)$, the sine of half this angle, is slightly less that 2^{-n}, so that $1/a(n) > 2^n$. On the other hand, $2^{-n} \leq |P_n^{-1}|$ for any n, where P_n is any n–fold product of the three matrices above. Since each matrix DT^n can be written as such a composition, $|DT^{-n}_{\alpha(n)}|/a(n) > 1$ for all n. Therefore, a Lasota–Yorke inequality cannot be obtained be using the geometry of one–to–one domains. On the other hand, T has a finite range structure consisting of the whole square, and the triangle $\{(x_1, x_2) : 0 \leq x_1 + x_2 < 1\}$. Since each of these sets is non–degenerate, Theorem 1 (ii) applies. □

1.4 Existence of invariant densities

The existence of an invariant density for T now follows in a standard fashion:

Theorem 2 *Let T be as in Theorem 1 (i) or (ii). Then T has an absolutely continuous invariant measure whose density has bounded variation. Upper bounds on the variation can be calculated explicitly.*

Proof. Suppose that T is as in part (i) of Theorem 1. Fix $\sigma \in (C_1\lambda^{-1}(1 + 1/a), 1)$ and let K_σ be as in that theorem. Then, by definition of the cones $\mathcal{C}_M$,

$$\mathcal{L}\mathcal{C}_M \subset \mathcal{C}_{\sigma M + K_\sigma},$$

in view of the fact that $\|\mathcal{L}f\|_{L^1} = \|f\|_{L^1}$ whenever $f \geq 0$. Thus $X = \mathcal{C}_{K_\sigma/(1-\sigma)} \cap \{f : \|f\|_{L^1} = 1\}$ is a compact, convex set, invariant under $\mathcal{L}$. It follows that $\mathcal{L}$ has a fixed point in X; that is, an invariant density f whose variation is bounded above by $K_\sigma/(1-\sigma)$.

In case (ii), use Theorem 1 (ii) to pick an n such that a contractive Lasota–Yorke inequality holds for T^n, and produce a normalised fixed point f_n for $\mathcal{L}^n$ as in case (i) above. It is easy to verify that

$$g = \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{L}^k f_n$$

is an invariant density for T. We finish by computing an upper bound on its variation. Suppose that $V_A(f_n) \leq M$. For each $k < n$, the same argument as in Proposition 1.3 shows that there exists $\delta_k > 0$ such that

$$V_A(\mathcal{L}^k f) \leq C_1\lambda^{-k}(1 + 1/a')M + (C_2 \frac{1 - \lambda^{-k}}{1 - \lambda^{-1}}(1 + 1/a') + 1/(a'\delta_k))1.$$

Summing this expression over K and dividing by n one obtains

$$V_A(g) \leq \frac{(1 + 1/a')}{1 - \lambda^{-1}} \left(\frac{C_1 M}{n} + C_2 \right) + \frac{1}{a' \min\{\delta_1, \ldots, \delta_k\}}.$$

□

2 Mixing rates

In this section we establish explicit bounds on the non–peripheral spectrum of the Perron–Frobenius operator for T. By the classical Ionescu–Tulcea and Marinescu ergodic theorem [18], the existence of such a spectral gap follows almost immediately from Theorem 1. However, our present concern is to obtain constructively derived expressions for this gap. Consequently, we avoid use of this general theorem, and adopt a more "first–principles" approach.

Throughout this section we will use a Renyi–type distortion constant C_3 satisfying

$$\left| \frac{\det DT^{-n}_{\alpha^{(n)}}(x)}{\det DT^{-n}_{\alpha^{(n)}}(y)} \right| \leq C_3 \tag{4}$$

whenever $x, y \in TB_{\alpha^{(n)}}$ for some $\alpha^{(n)}$. By (1) and (2),

$$Lip\left(\log|\det(DT^{-n}_{\alpha^{(n)}})|\right) \leq C_1 C_2 \frac{1-\lambda^{-n}}{1-\lambda^{-1}}, \tag{5}$$

so that (4) automatically holds with $C_3 \leq e^{\frac{C_1 C_2}{1-\lambda^{-1}}}$.

For the remainder of the paper we assume that $A \subset [0,1]^r$, and make the following "onto" assumption:

Assumption: Let T have a uniformly non–degenerate range structure, and suppose that for every $M \geq 0$ there exists an $n_0(M)$ such that for each $n \geq n_0$ there exists a set $W_n \subset A$ such that (i) $H(\partial W_n) < \infty$, (ii) for each one–to–one branch $\alpha^{(n)}$, $W_n \subset T^n B_{\alpha^{(n)}}$ and (iii) W_n is not too small: in particular,

$$m(W_n) \geq 8rMC_1C_3\lambda^{-n}$$

where C_1, λ are as in (1) and C_3 is as in (4). □

REMARK. If T has a non–degenerate finite range structure $\{U_1, \ldots, U_k\}$ and there exists $i_0 \in 1 \ldots k$ such that

$$U_{i_0} \subset T^n\left(B_{\alpha^{(n)}}\right)$$

for all one–to–one branches of T^n, then the onto assumption will be satisfied for large enough n_0. □

The argument in this part of the paper is in several stages. First of all, we define a suitable family of norms (equivalent to the usual BV norm) in which to obtain contraction rates. We then establish a contraction principle which requires that certain iterates of arbitrary pairs of BV functions satisfy a mutual "lower bound condition". The remainder of the section is devoted to showing that this lower bound condition holds under the "onto assumption".

2.1 Choice of norms for mixing estimates

First of all, we need an appropriate norm in which to compute mixing estimates: As above, let $\mathcal{C}_M$ denote a cone of uniformly bounded variation, and define a *difference cone* Γ_M by

$$\Gamma_M = \left\{ \begin{array}{r} g \in BV(A) : g = f^{(+)} - f^{(-)} \text{where } f^{(\pm)} \in \mathcal{C}_M \text{ are s. t.} \\ \|f^{(+)}\|_{L^1} = \|f^{(-)}\|_{L^1} \end{array} \right\}.$$

Note that $f^{(\pm)}$ are not the usual positive and negative parts $g^{\pm}$ of g (unless it happens that these both lie in the cone $\mathcal{C}_M$). One can easily check that Γ_M is a linear subspace of $BV_0 = \{g \in BV(A) : \int_A g\, dm = 0\}$. We define a norm on each Γ_M by:

$$\|g\|_M = \inf\left\{ \|f^{(\pm)}\|_{L^1} : g = f^{(+)} - f^{(-)}, f^{(\pm)} \in \mathcal{C}_M \right\}.$$

Indeed, one can prove the following lemma [30]:

Lemma 2.1 *Each* $(\Gamma_M, \|\cdot\|_M)$ *is a Banach space. Moreover,*

1. *if* $M_1 < M_2$ *then* $\Gamma_{M_1} \subset \Gamma_{M_2}$ *and* $\|g\|_{M_2} \leq \|g\|_{M_1}$ *whenever* $g \in \Gamma_{M_1}$*;*
2. *if* $\mathcal{L}$ *is a Perron–Frobenius operator and* $n > 0$ *then*

$$\mathcal{L}^n \mathcal{C}_{M_2} \subset \mathcal{C}_{M_1} \quad \Rightarrow \quad \|\mathcal{L}^n g\|_{M_1} \leq \|g\|_{M_2} \text{ whenever } g \in M_2;$$

3. if $g \in \Gamma_M$ *then*

$$\|g\|_{L^1} \leq 2\|g\|_M \text{ and } V_A(g) \leq 2M\|g\|_M;$$

4. if $g \in \Gamma_M$ *then*

$$\|g\|_M \leq \max\left\{\frac{\|g\|_{L^1}}{2}, \frac{V_A(g)}{M}\right\}.$$

In view of parts (3) and (4) of the above lemma, each Γ_M agrees with the zero–mean subspace BV_0 of $BV(A)$, and the norms $\|g\|_M$ and $\|g\|_{BV}$ are equivalent up to a multiplicative constant (depending on M). The norms $\|\cdot\|_M$ have been specifically constructed because they are ideal for estimating rates of contraction in BV_0 under the action of $\mathcal{L}$; we now establish the appropriate contraction principle. For clarity, we state the hypotheses of the proposition first:

Condition: *An operator $\mathcal{L}$ is said to satisfy a* uniform lower bound condition *on the cone $\mathcal{C}_M$ if there exists an $\alpha \in (0,1)$ and $n_M > 0$ such that for each pair of functions $f^{(1)}, f^{(2)} \in \mathcal{C}_M$ with $\|f^{(1)}\|_{L^1} = \|f^{(2)}\|_{L^1}$, there exists a non–negative function $\psi = \psi_{f^{(1)},f^{(2)}}$ such that*

$$\mathcal{L}^{n_M} f^{(1)} - \psi \in \mathcal{C}_M, \qquad \mathcal{L}^{n_M} f^{(2)} - \psi \in \mathcal{C}_M$$

and $\|\psi\|_{L^1} \geq \alpha \|f^{(1,2)}\|_{L^1}$. □

Proposition 2.2 (Contraction on a difference cone) *If $\mathcal{L}$ satisfies a uniform lower bound condition on $\mathcal{C}_M$ then*

$$\|\mathcal{L}^{n_M} g\|_M \leq (1-\alpha)\|g\|_M$$

for all $g \in BV_0$.

Proof. Let $g = f^{(+)} - f^{(-)}$ where $f^{(\pm)} \in \mathcal{C}_M$ and $\|g\|_M = \|f^{(\pm)}\|_{L^1}$. Let $\psi = \psi_{f^{(+)},f^{(-)}}$ be as in the lower bound condition. Then

$$\begin{aligned}\mathcal{L}^{n_M} g = \mathcal{L}^{n_M}(f^{(+)} - f^{(-)}) &= \mathcal{L}^{n_M} f^{(+)} - \mathcal{L}^{n_M} f^{(-)} \\ &= \left(\mathcal{L}^{n_M} f^{(+)} - \psi\right) - \left(\mathcal{L}^{n_M} f^{(-)} - \psi\right).\end{aligned} \tag{6}$$

By assumption, $\mathcal{L}^{n_M} f^{(\pm)} - \psi \in \mathcal{C}_M$, so that $0 \leq \mathcal{L}^{n_M} f^{(\pm)} - \psi$ and hence

$$\left\|\mathcal{L}^{n_M} f^{(\pm)} - \psi\right\|_{L^1} = \int_A \left(\mathcal{L}^{n_M} f^{(\pm)} - \psi\right) dm = \int_A \mathcal{L}^{n_M} f^{(\pm)}\, dm - \int_A \psi\, dm = \left\|\mathcal{L}^{n_M} f^{(\pm)}\right\|_{L^1} - \|\psi\|_{L^1}.$$

Moreover, $f^{(\pm)} \geq 0 \Rightarrow \left\|\mathcal{L}^{n_M} f^{(\pm)}\right\|_{L^1} = \left\|f^{(\pm)}\right\|_{L^1}$, so that

$$\begin{aligned}\left\|\mathcal{L}^{n_M} f^{(\pm)} - \psi\right\|_{L^1} &= \left\|f^{(\pm)}\right\|_{L_1} - \|\psi\|_{L^1} \\ &\leq \left\|f^{(\pm)}\right\|_{L^1} - \alpha\left\|f^{(\pm)}\right\|_{L^1} \\ &= (1-\alpha)\|g\|_M.\end{aligned}$$

The result now follows from (6) and the definition of $\|\cdot\|_M$. □

This contraction principle is a generalisation of the classical Doeblin condition [29], and forms the basis of the argument below. To apply Proposition 2.2, most of the work consists in establishing the existence of lower bound functions for suitable iterates of the Perron–Frobenius operator for T.

REMARK. Using Lemma 2.1, the contraction principle in Proposition 2.2 leads to explicit contraction rates for iterates of $\mathcal{L}$ in the $\|\cdot\|_{BV}$–norm on the BV–cone. This should be compared with the work of Schmitt [12], Liverani [26] and Saussol [33] where contraction rates are obtained in a projective metric (rather than the underlying norm) on suitable cones. □

2.2 Lower bounds for iterates of certain functions

In the next three sections we prove that $\mathcal{L}$ satisfies a uniform lower bound condition. We begin by calculating lower bound functions for certain simple functions.

Recall that $A \subset [0,1]^r$, let $h \in (0,1)$ and let η be a partition of $[0,1]^r$ into cubes of side-length h. Let

$$\mathcal{C}_{M,h} = \mathcal{C}_M \cap \left\{\phi : \phi = \sum_{C_\beta \in \eta} \phi_\beta\, \chi_{C_\beta} \qquad (\phi_\beta \in \mathbb{R}^+)\right\}.$$

Our initial target is to construct uniform lower bound functions for $\mathcal{L}^n \mathcal{C}_{M,h}$, but we begin with a general lemma:

Lemma 2.3 *Let T satisfy (1) and (2), have a uniformly non–degenerate range structure, and let $O_1 = W_{n_1}$ be an "onto set" for T^{n_1}. If $\mathcal{L}$ is the Perron–Frobenius operator for T, then for a.e. $x \in O_1$ and $0 \le \phi \in L^1$:*

$$\mathcal{L}^{n_1}\phi(x) \ge \frac{1}{C_3}\left\| \sum_{\alpha^{(n_1)}} \left(\operatorname*{ess\,inf}_{B_{\alpha^{(n_1)}}} \phi \right) \chi_{B_{\alpha^{(n_1)}}} \right\|_{L^1},$$

where C_3 is the constant from (4), and $\{B_{\alpha^{(n_1)}}\}$ is the partition of A into one–to–one branches of T^{n_1}.

Proof. Since $x \in O_1$, it has a unique pre-image under each one–to–one branch of T^{n_1}: $x_{\alpha^{(n_1)}} = T^{-n_1}_{\alpha^{(n_1)}}(x) \in B_{\alpha^{(n_1)}}$. Hence,

$$\mathcal{L}^{n_1}\phi(x) = \sum_{\alpha^{(n_1)}} \frac{\phi(x_{\alpha^{(n_1)}})}{|\det DT^n(x_{\alpha^{(n_1)}})|}.$$

In view of equation (4) and the mean value theorem,

$$|\det DT^n(x_{\alpha^{(n_1)}})| \le C_3 \frac{m\left(T^n(B_{\alpha^{(n_1)}})\right)}{m(B_{\alpha^{(n_1)}})} \le \frac{C_3}{m(B_{\alpha^{(n_1)}})}.$$

Thus, for a.e. $x \in O_1$,

$$\begin{aligned} \mathcal{L}^n\phi(x) &\ge \sum_{\alpha^{(n_1)}} \phi(x_{\alpha^{(n_1)}}) \frac{m(B_{\alpha^{(n_1)}})}{C_3} \\ &\ge \frac{1}{C_3} \sum_{\alpha^{(n_1)}} \left(\operatorname*{ess\,inf}_{B_{\alpha^{(n_1)}}} \phi \right) m(B_{\alpha^{(n_1)}}) \\ &= \frac{1}{C_3} \left\| \sum_{\alpha^{(n_1)}} \left(\operatorname*{ess\,inf}_{B_{\alpha^{(n_1)}}} \phi \right) \chi_{B_{\alpha^{(n_1)}}} \right\|_{L^1}, \end{aligned}$$

since for a.e. $x \in O_1$, $0 \le \operatorname*{ess\,inf}_{B_{\alpha^{(n_1)}}} \phi \le \phi(x_{\alpha^{(n_1)}})$ for each $\alpha^{(n_1)}$. □

We now estimate the norm of the function on the right in the conclusion of Lemma 2.3. In the one–dimensional case, this is very easy as BV functions on the interval are well approximated by the infimum over fine partitions [30]. Unfortunately this is not the case with $BV(\mathbb{R}^r)$ when a function can have essential infimum equal to zero over any reasonable partition, while still retaining positive mass and finite variation. Since our method uses $\sum_{\alpha^{(n)}} \operatorname*{ess\,inf}_{B_{\alpha^{(n)}}} \phi \chi_{b_{\alpha^{(n)}}}$ to capture a fixed proportion of the mass of ϕ, we cannot use just any initial functions $\phi \in BV(A)$. For this reason, we restrict to simple functions, which cannot be too wild.

Lemma 2.4 *If $\phi \in \mathcal{C}_{M,h}$ and n is sufficiently large that $C_1\lambda^{-n}\sqrt{r} < h/2$ (cf. (1)) then there exists a function $\tilde{\phi} \ge 0$ such that*

$$\tilde{\phi} \le \sum_{\alpha^{(n)}} \left(\operatorname*{ess\,inf}_{B_{\alpha^{(n)}}} \phi \right) \chi_{B_{\alpha^{(n)}}}$$

and

$$\left\|\tilde{\phi}\right\|_{L^1} \ge \left(1 - 2^{r-1}hM\right) \|\phi\|_{L^1}.$$

Proof. If $1 < 2^{r-1}Mh$ then there is nothing to prove. Otherwise, let ζ be the partition of $[0,1]^r$ into rectangles which is obtained by using the central points of the cubes from η as vertices. Then each partition element $E \in \zeta$ is a rectangle with side-length h (unless part of its boundary is formed by ∂A, in which case some sides have length $h/2$). Put

$$\tilde{\phi} = \sum_{E \in \zeta} \left(\min_E \phi \right) \chi_E.$$

We first prove the norm estimate, and then the lower bound.

Note that since $\phi \in \mathcal{C}_M$, $V_A(\phi) \le M\|\phi\|_{L^1}$. Since ϕ is constant on elements of η, all contributions to $V_A(\phi)$ come from discontinuities at the boundary of partition elements. In particular, since H–almost all[1] such boundary points are contained in $\cup_{E\in\zeta} int(E)$,

$$M\|\phi\|_{L^1} \ge V_A(\phi) \ge \sum_{E\in\zeta} V_{int(E)}(\phi). \tag{7}$$

Now, fix $E \in \zeta$. If $\{\Gamma_i\}$ denotes the (finite) collection of oriented boundaries of cells from η that intersect E, then

$$V_{int(E)}(\phi) = \sum_i \left| tr^+_{\Gamma_i}\phi - tr^-_{\Gamma_i}\phi \right| \, H(E\cap\Gamma_i).$$

However, since each $E\cap\Gamma_i$ is an $(r-1)$–dimensional rectangle of side-length $h/2$, we can immediately calculate $H(E\cap\Gamma_i) = (h/2)^{r-1}$. Moreover, since $\phi|_E$ is piecewise constant with discontinuities along Γ_i,

$$\max_E \phi - \min_E \phi \le \sum_i \left| tr^+_{\Gamma_i}\phi - tr^-_{\Gamma_i}\phi \right| = V_{int(E)}(\phi)\,(2/h)^{r-1}.$$

In particular, it follows that

$$\left\| (\phi - \tilde{\phi})\Big|_E \right\|_{L^1} = \left\| (\phi - \min_E \phi)\Big|_E \right\|_{L^1} \le \left(\max_E \phi - \min_E \phi \right) m(E) \le V_{int(E)}(\phi) 2^{r-1} h.$$

Summing this expression over $E \in \zeta$ and comparing with (7), we obtain

$$\|\tilde{\phi}\|_{L^1} = \|\phi\|_{L^1} - \|\phi - \tilde{\phi}\|_{L^1} \ge \left(1 - 2^{r-1} hM\right) \|\phi\|_{L^1}.$$

We now prove the lower bound. This follows from:

Claim: Let $B_{\alpha(n)}$ be a one–to–one domain of T^n and let $E \in \zeta$ be such that $m(E\cap B_{\alpha(n)}) > 0$. Then

$$\{C \in \eta : m(B_{\alpha(n)} \cap C) > 0\} \subset \{C \in \eta : m(C\cap E) > 0\}.$$

Proof of claim: By the construction of ζ, whenever $C \in \eta$ and $E \in \zeta$, either $m(C\cap E) > 0$, or $|x_C - x_E| \ge h/2$ for all $x_C \in C$ and $x_E \in E$. Therefore, if $m(C\cap E) = 0$ and $y \in C \in \eta$, $z \in E \in \zeta$ then $|y-z| \ge h/2$. Now, suppose that C is such that $m(B_{\alpha(n)} \cap C) > 0$. Let $y \in B_{\alpha(n)} \cap C$ and $z \in E \cap B_{\alpha(n)}$. Since both $y, z \in B_{\alpha(n)}$, (1) implies that $|y-z| < C_1\lambda^{-n}|T^n(y) - T^n(z)| \le C_1\lambda^{-n}\sqrt{r} < h/2$. Therefore $m(C\cap E) > 0$, and the claim follows.

Proof of lemma continued: Now, if $E \in \zeta$, $m(B_{\alpha(n)} \cap E) > 0$ and $x \in B_{\alpha(n)} \cap E$ then, by the claim,

$$\left(\operatorname*{ess\,inf}_{B_{\alpha(n)}} \phi \right) = \min_{\{C\in\eta : m(C\cap B_{\alpha(n)})>0\}} \phi|_C \ge \min_{\{C\in\eta : m(C\cap E)>0\}} \phi|_C = \min_E \phi = \tilde{\phi}(x).$$

Repeating the argument for each E with $m(B_{\alpha(n)} \cap E) > 0$ completes the proof of the lemma. □

Proposition 2.5 *Let T, n_1 and O_1 be as in Lemma 2.3. If $\phi \in \mathcal{C}_{M,h}$ and and n_1 is such that $C_1\lambda^{-n_1}\sqrt{r} < h/2$ (cf. (1)), then*

$$\mathcal{L}^{n_1}\phi \ge \frac{1}{C_3}\left(1 - 2^{r-1}hM\right) \|\phi\|_{L^1} \chi_{O_1}.$$

Proof. Follows immediately from Lemmas 2.3 and 2.4. □

Proposition 2.5 gives a lower bound function for iterates under $\mathcal{L}$ of functions from $\mathcal{C}_{M,h}$. To construct lower bound functions for iterates of general functions from the cone $\mathcal{C}_M$ (as needed in the application of Proposition 2.2), it is essential to study how $\mathcal{C}_{M,h}$ can be used to approximate $\mathcal{C}_M$.

[1] H denotes the $(r-1)$–dimensional Hausdorff measure.

2.3 Approximation of $BV([0,1]^r)$ by simple functions

Here we collect several facts about the behaviour of $BV([0,1]^r)$ under approximation by simple functions.

DEFINITION. (**Rectangular partition**) A partition η of $[0,1]^r$ is a *rectangular partition* if every element of η is a rectangle. If there exist positive integers $N_1, \ldots, N_r$ such that if $B \in \eta$ then

$$B = B_{k_1 \ldots k_r} = [k_1/N_1, (k_1+1)/N_1) \times \ldots \times [k_r/N_r, (k_r+1)/N_r)$$

where[2] $0 \le k_i < N_i$, then η is called a *regular rectangular partition.* Let $\mathcal{D}_\eta$ denote the collection of L^1 functions that are constant on each $B \in \eta$. □

REMARK. If η is a regular rectangular partition let $h(\eta) = \min\{1/N_1, \ldots, 1/N_r\}$. If $\phi \in \mathcal{D}_\eta$, then

$$V_A(\phi) \le \frac{2r\|\phi\|_{L^1}}{h(\eta)},$$

as is easily checked. □

Let η be a regular rectangular partition of a rectangular domain, and define a projection operator $\Pi_\eta : L^1(A) \to \mathcal{D}_\eta$ by the formula

$$(\Pi_\eta \phi)(x) = \sum_{B \in \eta} \frac{\int_B \phi \, dm}{m(B)} \chi_B(x).$$

One can check that $\int_A \phi \, dm = \int_A (\Pi_\eta \phi) \, dm$, and further key properties are summarised in the following lemma.

Lemma 2.6 *If $A \subset \mathbb{R}^r$ is a rectangle, and η a regular rectangular partition with maximum side-length h, $\phi \in BV(A)$, then*

1. $V_A(\Pi_\eta \phi) \le \sqrt{r} V_A(\phi)$;
2. $\|\phi - \Pi_\eta \phi\|_{L^1} \le h\sqrt{r} V_A(\phi)$.

REMARK. The facts reported in Lemma 2.6 generalise elementary properties of variation in one–dimension and depend on the fact that the partition is rectangular. A sketch of the proof is given in Appendix A. □

Lemma 2.6 gives quantitative control over the approximation of $BV(A)$ by $\mathcal{D}_\eta$. We can now use these estimates to prove the lower bound condition for iterates of $\mathcal{C}_M$.

2.4 Lower bounds for iterates of general BV densities

The construction of lower bound functions proceeds by "boot–strapping" off Proposition 2.5. We also need Theorem 1 the Lasota–Yorke inequality to show that the variation of the difference between a general BV function and an L^1–close simple function decays under iteration by $\mathcal{L}$. Therefore, suppose that Theorem 1 holds for an iterate of T. That is, there exists $n_1 \ge 1$ and constants $\sigma_1 < 1, K_1 < \infty$ such that

$$V_A(\mathcal{L}^{n_1} f) \le \sigma_1 V_A(f) + K_1 \|f\|_{L^1}$$

for $f \in BV(A)$. Iterated application of this inequality yields:

$$V_A(\mathcal{L}^{kn_1}) \le \sigma_1^k V_A(f) + \frac{1-\sigma_1^k}{1-\sigma_1} K_1 \|f\|_{L^1}. \tag{8}$$

Proposition 2.7 (Lower bound functions for $\mathcal{C}_M$) *Let $M > 0$, $A = [0,1]^r$ and suppose that $\phi^{(1)}, \phi^{(2)} \in \mathcal{C}_M$, are such that $\|\phi^{(1)}\|_{L^1} = \|\phi^{(2)}\|_{L^1}$. Suppose also that n_2 is large enough that*

$$\sqrt{r}\, C_1 \lambda^{-n_1 n_2} \le \frac{1}{2^{r+1}\sqrt{r} M},$$

and $n_1 n_2 > n_0(M)$ where $n_0(M)$ is as in the definition of the onto assumption. Let O_2 be an onto set for $T^{n_1 n_2}$. Then there exists a function $\psi \ge 0$ such that

$$\mathcal{L}^n \phi^{(1)} \ge \psi \quad \text{and} \quad \mathcal{L}^n \phi^{(2)} \ge \psi,$$

[2] If any $k_i = N_i$, then the corresponding interval in the product should be $[1 - 1/N_i, 1]$.

while

$$\|\phi^{(1,2)}\|_{L^1}\frac{m(O_2)}{2C_3} \geq \|\psi\|_{L^1} \geq \|\phi^{(1,2)}\|_{L^1}\frac{m(O_2)}{4C_3}$$

and

$$V_A(\psi) \leq M_\psi \|\phi^{(1,2)}\|_{L^1},$$

where

$$M_\psi = \frac{H(O_2)}{2C_3} + 2\left(\sigma_1^{n_2}(1+\sqrt{r})M + \frac{K_1}{1-\sigma_1}\frac{m(O_2)}{4C_3}\right).$$

Proof. By multiplying ψ by $\|\phi^{(1,2)}\|_{L^1}$, the proposition will hold for general $\phi^{(1,2)}$ if it holds for $\phi^{(1)}/\|\phi^{(1)}\|_{L^1}$ and $\phi^{(2)}/\|\phi^{(2)}\|_{L^1}$. We therefore assume without loss of generality that $\|\phi^{(1,2)}\|_{L^1} = 1$. The proof involves approximating $\phi^{(1)}$ and $\phi^{(2)}$ by piecewise constant densities.

Since $n_1 n_2 > n_0(M)$, the onto assumption guarantees that

$$2\sqrt{r}C_1\lambda^{-n_1 n_2} \leq \frac{m(O_2)}{4C_3\sqrt{r}M}.$$

Hence, it is possible to choose $h > 0$ such that

$$2\sqrt{r}C_1\lambda^{-n_1 n_2} \leq h \leq \frac{1}{\sqrt{r}M}\min\left\{\frac{m(O_2)}{4C_3}, \frac{1}{2^r}\right\}, \tag{9}$$

and η a regular rectangular partition of side-length h. Then, since $\phi^{(1,2)} \geq 0$,

$$\|\Pi_\eta\phi^{(1)}\|_{L^1} = \|\phi^{(1)}\|_{L^1} = 1 = \|\phi^{(2)}\|_{L^1} = \|\Pi_\eta\phi^{(2)}\|_{L^1}.$$

Moreover, by Lemma 2.6 and (9), for $i = 1, 2$,

$$\begin{aligned} \|(\Pi_\eta - Id)\phi^{(i)}\|_{L^1} &\leq h\sqrt{r}V_A(\phi^{(i)}) \\ &\leq h\sqrt{r}M\|\phi^{(i)}\|_{L^1} \\ &\leq \frac{m(O_2)}{4C_3}, \end{aligned} \tag{10}$$

while

$$V_A(\Pi_\eta\phi^{(i)}) \leq \sqrt{r}M. \tag{11}$$

Then, by (9),

$$2^{r-1}h\,V_A(\Pi_\eta\phi^{(i)}) \leq \frac{1}{2} \quad \text{and} \quad \sqrt{r}\,C_1\lambda^{-n_1 n_2} \leq \frac{h}{2}.$$

Therefore, by Proposition 2.5,

$$\mathcal{L}^{n_1 n_2}(\Pi_\eta\phi^{(i)}) \geq \frac{1}{2C_3}\chi_{O_2}.$$

Now, because $\mathcal{L}$ is a linear operator,

$$\begin{aligned} \mathcal{L}^{n_1 n_2}\phi^{(i)} &= \mathcal{L}^{n_1 n_2}(\Pi_\eta\phi^{(i)}) - \mathcal{L}^{n_1 n_2}(\Pi_\eta\phi^{(i)} - \phi^{(i)}) \\ &\geq \frac{1}{2C_3}\chi_{O_2} - \mathcal{L}^{n_1 n_2}\left(\Pi_\eta\phi^{(i)} - \phi^{(i)}\right)^+ \end{aligned}$$

where $f^+ = \max\{f, 0\}$ denotes the positive part of a function.

Now put

$$\psi = \max\left\{0, \frac{1}{2C_3}\chi_{O_2} - \mathcal{L}^{n_1 n_2}\left(\Pi_\eta\phi^{(1)} - \phi^{(1)}\right)^+ - \mathcal{L}^{n_1 n_2}\left(\Pi_\eta\phi^{(2)} - \phi^{(2)}\right)^+\right\}.$$

Then,

$$\mathcal{L}^{n_1 n_2}\phi^{(1)} \geq \psi, \qquad \mathcal{L}^{n_1 n_2}\phi^{(2)} \geq \psi, \qquad \psi \geq 0;$$

we need only check the norm and variation estimates. Clearly,

$$\begin{aligned} \frac{m(O_2)}{2C_3} \geq \|\psi\|_{L^1} \geq \frac{m(O_2)}{2C_3} &- \left\|\left(\Pi_\eta\phi^{(1)} - \phi^{(1)}\right)^+\right\|_{L^1} \\ &- \left\|\left(\Pi_\eta\phi^{(2)} - \phi^{(2)}\right)^+\right\|_{L^1}, \end{aligned} \tag{12}$$

and

$$V_A(\psi) \le \frac{H(\partial O_2)}{2C_3} \quad + V_A\left(\mathcal{L}^{n_1 n_2}\left(\Pi_\eta \phi^{(1)} - \phi^{(1)}\right)^+\right)$$
$$+ V_A\left(\mathcal{L}^{n_1 n_2}\left(\Pi_\eta \phi^{(2)} - \phi^{(2)}\right)^+\right). \qquad (13)$$

Now, because $\int_A \mathcal{L}^{n_1 n_2}(\Pi_\eta \phi^{(i)} - \phi^{(i)}) = 0$,

$$\left\|\mathcal{L}^{n_1 n_2}\left(\Pi_\eta \phi^{(i)} - \phi^{(i)}\right)^+\right\|_{L^1} = \frac{\|\mathcal{L}^{n_1 n_2}(\Pi_\eta \phi^{(i)} - \phi^{(i)})\|_{L^1}}{2}$$
$$\le \frac{1}{2}\frac{m(O_2)}{4C_3} \qquad \text{by (10).}$$

The norm estimate now follows from (12). Finally, by (11)

$$V_A(\Pi_\eta \phi^{(i)} - \phi^{(i)}) \le V_A(\Pi_\eta \phi^{(i)}) + V_A(\phi^{(i)}) \le (1 + \sqrt{r})M,$$

so that equations (8) and (10) together show that

$$V_A(\mathcal{L}^{n_1 n_2}(\Pi_\eta \phi^{(i)} - \phi^{(i)})) \le \sigma_1^{n_2}(1 + \sqrt{r})M + \frac{K_1}{1 - \sigma_1}\frac{m(O_2)}{4C_3}.$$

The bound on M_ψ now follows from (13). □

2.5 Mixing rates

We now give a spectral estimate for $\mathcal{L}$.

Let the transformations T be as above. That is, there is a Lasota–Yorke inequality for T^{n_1} with constants σ_1, K_1, and the onto assumption holds. Fix

$$M_* = \frac{3K_1}{1 - \sigma_1}.$$

Let n_2 be large enough that $n_1 n_2 > n_0(M_*)$ and the other hypotheses of Proposition 2.7 hold. Let M_ψ be the constant from Proposition 2.7, and let n_3 be the minimal integer such that

$$\sigma_1^{n_3}\left(\sigma_1^{n_2} M_* + \frac{K_1}{1 - \sigma_1} + M_\psi\right) \le \frac{1}{2}\frac{K_1}{1 - \sigma_1}.$$

Let

$$n_* = n_1(n_2 + n_3).$$

Theorem 3 (Mixing Theorem for $\mathcal{L}$) *Let O_2 be the onto set from Proposition 2.7. Then for any function $g \in BV_0(A)$,*

$$\|\mathcal{L}^{n_*} g\|_{M_*} \le (1 - m(O_2)/4C_3)\|g\|_{M_*},$$

where $\|\cdot\|_{M_*}$ is the norm on the difference cone Γ_{M_*}.

Proof. The theorem follows more or less immediately from Propositions 2.2 and 2.7. Let $f^{(1,2)} \in \mathcal{C}_{M_*}$ be such that $\|f^{(1)}\|_{L^1} = \|f^{(2)}\|_{L^1}$, then Proposition 2.7 applies. Let ψ be the given function. Then, for $i = 1, 2$,

$$\left\|\mathcal{L}^{n_1 n_2} f^{(i)} - \psi\right\|_{L^1} \ge \|f^{(i)}\|_{L^1} - \|\psi\|_{L^1} \ge (1 - m(O_2)/2C_3)\,\|f^{(i)}\|_{L^1}$$

while

$$V_A(\mathcal{L}^{n_1 n_2} f^{(i)} - \psi) \le \sigma_1^{n_2} V_A(f^{(i)}) + \frac{K_1}{1 - \sigma_1}\|f^{(i)}\|_{L^1} + V_A(\psi) \le M'\|f^{(i)}\|_{L^1},$$

where $M' = \sigma_1^{n_2} M_* + M_\psi + K_1/(1 - \sigma_1)$. Now, let

$$\psi_* = \mathcal{L}^{n_1 n_3}\psi.$$

Since ψ is a positive function,

$$\|\psi_*\|_{L^1} = \|\psi\|_{L^1} \ge \frac{m(O_2)}{4C_3}\|f^{(1,2)}\|_{L^1}.$$

To apply Proposition 2.2 we must check that $\mathcal{L}^{n_*}f^{(i)}-\psi_* \in \mathcal{C}_{M_*}$ for $i=1,2$. Since $\mathcal{L}^{n_1n_2}f^{(i)}-\psi \geq 0$,

$$\|\mathcal{L}^{n_*}f^{(i)}-\psi_*\|_{L^1} = \left\|\mathcal{L}^{n_1n_3}\left(\mathcal{L}^{n_1n_2}f^{(i)}-\psi\right)\right\|_{L^1} \geq \left(1-\frac{m(O_2)}{2C_3}\right)\|f^{(i)}\|_{L^1} \geq \frac{\|f^{(i)}\|_{L^1}}{2}.$$

On the other hand, by (8),

$$V_A(\mathcal{L}^{n_*}f^{(i)}-\psi_*) = V_A(\mathcal{L}^{n_1n_3}(\mathcal{L}^{n_1n_2}f^{(i)}-\psi)) \leq \left(\sigma_1^{n_3}M' + \frac{K}{1-\sigma_1}\right)\|f^{(i)}\|_{L^1} \leq \frac{1}{2}M_*\|f^{(i)}\|_{L^1}.$$

by the choice of n_3. The inclusion of $\mathcal{L}^{n_*}f^{(i)}-\psi_*$ in $\mathcal{C}_{M_*}$ follows. Thus $\mathcal{L}^{n_*}$ satisfies a uniform lower bound condition on $\mathcal{C}_{M_*}$ and the theorem now follows from Proposition 2.2. □

To illustrate that this theorem is fairly easy to apply, we give a mixing result for the special case when T has a finite range structure of non–degenerate sets. Suppose the sets $\{U_1,\dots U_k\}$ form the range structure, and let

$$a = \frac{1}{2}\min\{a(U_1),\dots,a(U_k)\}$$

where $a(U_i)$ is the smallest interior angle of the non–degenerate set U_i. Let $\delta > 0$ be such that Lemma 1.2 holds uniformly with $\delta = \delta(U_i)$ for each i. Next put

$$n_1 = \left\lceil \frac{-\log(2C_1(1+1/a))}{\log\lambda} \right\rceil \qquad \text{and} \qquad K_1 = C_2\frac{1}{1-\lambda^{-1}}\left(1+\frac{1}{a}\right)+\frac{1}{a\delta}.$$

Now, by Theorem 1 and this choice of n_1, K_1,

$$V_A(\mathcal{L}^k f) \leq C_1\lambda^{-i}V_A(f) + K_1\|f\|_{L^1},$$

and in particular

$$V_A(\mathcal{L}^{n_1}f) \leq 0.5\, V_A(f) + K_1\|f\|_{L^1}.$$

Therefore, put $M_* = 6K_1$ and $\beta = \min\{m(U_i)\}$. Then the minimal n_0 in the onto assumption is:

$$n_0 = \left\lceil \frac{\log(8rM_*C_1C_3/\beta)}{\log\lambda} \right\rceil.$$

Finally, choose n_2, n_3, n_* as preceding[3] Theorem 3. Then:

Theorem 4 *Let T have a finite range structure of non–degenerate sets and let $f \in BV(A)$. Then for any $n > 0$,*

$$\|\mathcal{L}^n(f - \textstyle\int f\,dm)\|_{BV} \leq C\rho^n\|f - \textstyle\int f\,dm\|_{BV},$$

where

$$\rho = (1-\beta/(4C_3))^{1/n_*}, \qquad \text{and} \qquad C = \frac{8(1+M_*)}{3}\max\left\{\frac{1}{2}, \frac{1+C_1/C_2}{6}\right\}$$

where $n_*, M_*, \beta, C_1, C_2, C_3$ are as in Theorem 3.

Proof. First of all, set $f - \int f\,dm = g \in BV_0(A)$. Then, by Theorem 3,

$$\|\mathcal{L}^{n_*}g\|_{M_*} \leq (1-m(O_2)/4C_3)\|g\|_{M_*}$$

where O_2 is the onto set for $T^{n_1n_2}$. Since this onto set must be one of the sets U_i, it follows automatically that $(1-m(O_2)/2C_3) \leq \rho^{n_*}$. Now, write $n = n_*k + n_r$ where $n_r < n_*$. Then,

$$\|\mathcal{L}^n g\|_{M_*} \leq \rho^{n_*k}\|\mathcal{L}^{n_r}g\|_{M_*} \leq \frac{\rho^n}{\rho^{n_*}}\max_{0\leq i<n_*}\|\mathcal{L}^i g\|_{M_*}.$$

By Lemma 2.1 (4) and the Lasota–Yorke inequality above,

$$\max_i\|\mathcal{L}^i g\|_{M_*} \leq \max_i\max\left\{\frac{\|\mathcal{L}^i g\|_{L^1}}{2}, \frac{C_1\lambda^{-i}(1+1/a)V_A(g)+K_1\|g\|_{L^1}}{M_*}\right\}.$$

Since $(1+1/a) < K_1/C_2$ and $M_* = 6K_1$, the rhs is bounded by $\{1/2, (1+C_1/C_2)/6\}\|g\|_{BV}$. Since $m(U_i)/4C_3 \leq 1/4$, it follows that $\rho^{n_*} > 3/4$, so that the upper bound on C follows from Lemma 2.1 (3). □

REMARKS:

[3] The choice of n_3 depends on the measure of the boundary of the onto sets; one simply uses $\max\{H(\partial U_i)\}$ at this place in the definition.

1. Theorems 3 and 4 are less than ideal. First of all, the choices of n_* and M_* have been made for efficient exposition, rather than minimality. For specific examples, better choices will always be possible. Secondly, the onto assumption is rather heavy–handed. In each of [12, 27, 26, 33], rates of decay of correlations are computed under much weaker hypotheses—essentially that T is mixing. In this paper, the requirement of having an onto set is important for getting explicit numerical bounds on the mixing rates.

2. The lower bound function approach used here means that mixing rates depend sensitively on both the distortion of the map (as expressed by the constants C_2, C_3), and the size of the contractive constant in the Lasota–Yorke inequality: the faster that variation is shrunk, the sooner mixing will occur.

3. For the applications presented in sections 3 and 4 below, explicit bounds on the spectral gap of the operator $\mathcal{L}|_{BV(A)}$ are important. The "lower bound function" method used to establish such bounds (cf. section 2) is related to a general method used by others [12, 27, 33]. In each of these papers, contraction rates are established for iterates of the Perron–Frobenius operator in a projective Hilbert metric on the BV–cone. In this paper, contraction rates have been established in a Banach space norm which is equivalent to the BV–norm (up to multiplicative constants). For the application to the error analysis of Ulam's method, this extension is essential: the projection operation Π_η is well behaved in the L^1–norm (cf. Lemma 2.6), but not in the projective metric used in [12, 27, 33]. □

3 Application to computing invariant measures

We now discuss a well known numerical scheme for approximation of invariant densities: Ulam's method. Initially suggested by Ulam in 1960 [35], and shown to converge for one-dimensional transformations by Li [25], there has been a recent upsurge in interest for multi-dimensional transformations [8, 13, 11]. The idea is to solve an *approximate operator equation.*

Ulam's method. Fix a finite partition η of A. Let Π_n denote the corresponding projection onto $\mathcal{D}_\eta$ (cf. Section 2). Then

$$\mathcal{P}_\eta \triangleq \Pi_\eta \circ \mathcal{L}$$

is an *Ulam approximate operator.* If T has a unique invariant density ϕ_*, then $\mathcal{P}_\eta$ has a one–dimensional fixed point space. Let ϕ_η be such that

$$\mathcal{P}_\eta \phi_\eta = \phi_\eta \quad \text{and} \quad \|\phi_\eta\|_{L^1} = 1.$$

One hopes that $\phi_\eta \to \phi_*$ as the partition η is refined. □

Given a partition η, Ulam's method is easily implementable on a computer; either as a finite–dimensional fixed point problem [25], or by Monte–Carlo simulations of an associated Markov chain [17, 30].

An early result on the strong–norm convergence of Ulam's method for a class of multi–dimensional maps appeared in [8]. A partial generalisation for expanding transformations was [11]. We now give a more complete version, with error bounds. The *rate* of approximation first appeared in [21] for one–dimensional maps, but the constructive approach adopted here enables error *bounds* to be derived.

Theorem 5 *Let $T : [0,1]^r \to [0,1]^r$ satisfy (1) and (2), and let η be a partition of $[0,1]^r$ into rectangles of side-length $h = h(\eta)$.*

(i) Let ϕ_η be a normalised fixed point of $\mathcal{P}_\eta$. If the one–to–one images of T are uniformly non–degenerate, and

$$C_1\lambda^{-1}(1 + 1/a) < 1/\sqrt{r}$$

(cf. Theorem 1), then T has an invariant density ϕ_* and $\|\phi_\eta - \phi_*\|_{L^1} \to 0$ as $h \to 0$.

(ii) If T has a uniformly non–degenerate range structure, then there exists $n_0 > 0$ such that if

$$\Pi_\eta \circ \mathcal{L}^{n_0}\phi_\eta = \phi_\eta, \qquad \|\phi_\eta\|_{L^1} = 1,$$

then

$$\|\phi_\eta - \phi_*\|_{L^1} \to 0 \text{ as } h \to 0.$$

In either case, $\|\phi_\eta - \phi_*\|_{L^1} \le O(-h\log h)$. In case (ii), the constants in the $O(\cdot)$ notation are explicitly computable.

Proof. *(i)* By Theorem 1, there exist constants $\sigma < 1/\sqrt{r}$, $K < \infty$ such that

$$\mathcal{L}\mathcal{C}_M \subset \mathcal{C}_{\sigma M+K}.$$

Moreover, by Lemma 2.6 (i), $\Pi_\eta \mathcal{C}_M \subset \mathcal{C}_{\sqrt{r}M}$. Hence,

$$(\Pi_\eta \circ \mathcal{L})\mathcal{C}_M \subset \mathcal{C}_{\sigma\sqrt{r}M+\sqrt{r}K},$$

so there exists $M_0 < \infty$ such that $(\Pi \circ \mathcal{L})\mathcal{C}_{M_0} \subset \mathcal{C}_{M_0}$. It follows that $\phi_\eta \subset \mathcal{C}_{M_0}$ (since ϕ_η is the unique normalised fixed point of $\mathcal{P}_\eta$). Because level sets of $\mathcal{C}_{M_0}$ are relatively compact in L^1, there exists a convergent subsequence of ϕ_η as $h(\eta) \to 0$. Indeed, by Lemma 2.6 (ii),

$$\begin{aligned} \|\mathcal{L}\phi_\eta - \phi_\eta\|_{L^1} &= \|(Id - \Pi_\eta)\mathcal{L}\phi_\eta\|_{L^1} \\ &\le h\sqrt{r}V_A(\mathcal{L}\phi_\eta) < hM_0 \to 0 \qquad \text{as } h \to 0. \end{aligned}$$

Hence, every limit point of $\{\phi_\eta\}$ is an invariant density for T.

(ii) Use Theorem 1 (ii) to pick $n_0 > 0$ such that

$$\mathcal{L}^{n_0}\mathcal{C}_M \subset \mathcal{C}_{\sigma M+K}$$

for some $\sigma < 1/\sqrt{r}$. Then, use the same argument as in part (i).

Finally, the approximation rate follows from a general argument used by Keller [21]. If the constants from Theorems 1 and 3 are used, then the same argument as in [30] will give explicit bounds on the constants in the $O(\cdot)$ notation. □

REMARKS:

1. In [11] the authors showed the existence of a constant c such that $V_A(\Pi_\eta\phi) \le cV_A(\phi)$ for suitable partitions η and $\phi \in BV(A)$. They therefore obtain a version of Theorem 5 (i). With the knowledge of c given by Lemma 2.6, we have been able to give a result with explicitly verifiable hypotheses.

2. There appears to be no way to get a result similar to part (ii) of the theorem without taking several iterates of $\mathcal{L}$ before applying Π_η. While this situation is unsatisfactory, it may be unavoidable: in a recent paper, Blank and Keller [4] discussed conditions for stochastic stability of one-dimensional maps. They proved that for Lasota–Yorke type maps, the expansivity constant λ must exceed 2 in order to guarantee stochastic stability, whereas expansivity exceeding 1 is sufficient for existence of an invariant density. If $1 < \lambda < 2$, a localisation phenomenon can occur, although they were able to show that Ulam's method still approaches the invariant measure in the topology of weak convergence. Whether this convergence can be extended to the L^1 topology without enough expansion to ensure stochastic stability is unclear.

3. It is worth pointing out that the partition η used for implementations of Ulam's method does not need to be a dynamical partition for the map. Certainly, if a Markov partition exists, better numerical performance will be obtained by using it. However, for multi–dimensional maps, calculating Markov partitions of arbitrarily fine resolution is at best extremely difficult. Indeed, this is one of the virtues of using the BV approach for Markov maps: In the Markov case, satisfactory existence results for acims need not require delicate calculations, since the discontinuities across partition boundaries can be ignored. In applications, this simplification cannot be assumed because dynamical partition boundaries are often infeasible to compute. One thus typically uses a non–Markov partition in Ulam's method (for example above, η is a partition into cubes), so that boundary discontinuities appear interior to partition cells; the projection operation Π_η then averages mass across (Markov) partition boundaries. Any analysis of Ulam's method needs to take this "spreading" into account. When the BV approach is analytically tractable (and successful, as in Theorems 1— 5), the exact location of partition boundaries becomes unimportant; only their geometry contributes to the Lasota–Yorke inequality (cf. Lemma 1.2). Since Ulam's method preserves BV (all our computations take place in suitably chosen cones C_M), the "global regularity" implied by inclusion in a uniform BV class is of immense practical importance in numerical calculations: it implies that the Ulam approximations are actually convergent to the acim. □

4 Example: Jacobi–Perron transformation in $\mathbb{R}^2$

We now illustrate the main results of the paper with the classical *Jacobi–Perron* (JP) transformation on $\mathbb{R}^2$. The quantitative information obtained is far from optimal, but the JP transformation has long resisted this kind of analysis [19, 3].

The JP algorithm is one attempt to generalise Diophantine approximation by the one–dimensional continued fraction algorithm to a multi–dimensional setting. Hence, it provides an example of a nontrivial (and uncontrived) multi–dimensional transformation whose ergodic properties are of interest [19, 9, 34]. There is a unique ergodic acim for the JP transformation [34], but no exact formula exists [2]. It also has exponential decay of correlations [28], but no bounds are known on the rate. In [19], the authors use a careful argument to prove an exponential rate of convergence[4] for the "modified Jacobi–Perron algorithm" for simultaneous approximation of points in $\mathbb{R}^2$ by rationals with a common denominator. The argument is based on integrating certain quantities with respect to the (known) invariant density for the modified Jacobi–Perron transformation. Because the invariant density is unknown for the JP transformation, similar results do not exist for the classical JP algorithm.

The transformation

Let $A = [0,1]^2$, and let

$$T(x_1, x_2) = \left(\left\{ \frac{x_2}{x_1} \right\}, \left\{ \frac{1}{x_1} \right\} \right),$$

where $\{\cdot\}$ denotes the fractional part of a number. Thus, there exist unique integers $k_2 \geq 1$ and $0 \leq k_1 \leq k_2$ such that

$$T(x_1, x_2) = \left(\frac{x_2}{x_1} - k_1, \frac{1}{x_1} - k_2 \right).$$

The pairs of integers (k_1, k_2) index the monotonicity sets of the transformation T, we write this partition as $\{B_{(k_1,k_2)}\}$.

The monotonicity components of T consist of two kinds of pieces: trapezia (indexed by pairs (k_1, k_2) with $k_1 < k_2$) and triangles (indexed by pairs (k_1, k_2) with $k_1 = k_2$). Under one application of T, the trapezia map over the entire square, while the triangles map over the triangular region

$$S = \cup_{k_2 \geq k_1 \geq 1} B_{(k_1,k_2)} = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq x_2 \leq 1\}.$$

Therefore, T has a *finite range structure* consisting of two subsets $\{S, [0,1]^2\}$. In fact, the transformation is Markov, and is conjugate to a subshift of finite type on the set of symbols

$$\{(k_1, k_2) : 0 \leq k_1 \leq k_2, 1 \leq k_2\}.$$

The transition

$$(k_1, k_2) \mapsto (k_1', k_2')$$

being admissible if either $k_1 < k_2$ or if $k_1 = k_2$ and $k_1' > 0$ (this is easily checked from the geometry of the map T). A string $\alpha^{(n)} = (k_1^{(0)}, k_2^{(0)}), \ldots, (k_1^{(n-1)}, k_2^{(n-1)})$ is called *admissible* if every transition $(k_1^{(j)}, k_2^{(j)}) \mapsto (k_1^{(j+1)}, k_2^{(j+1)})$ is admissible. It is obvious that $\alpha^{(n)}$ is an admissible string if and only if the cylinder set $T^{-n}_{\alpha^{(n)}}([0,1]^2)$ has non–zero measure. (To make the conjugacy with the subshift well–defined, all preimages of the vertical and integer slope lines which partition the monotonicity components of T must be removed from the square (c.f. [34]). Since these together have Lebesgue measure zero, generic ergodic properties are unaffected.)

Multi–dimensional rational approximation

Before applying our results, we establish some notation, and describe the JP algorithm for simultaneous approximation of two real numbers by rationals with a common denominator.

Fix $x = (x_1, x_2)$, and let $k(x) = (k_1(x), k_2(x))$ be the integers obtained by one application of the JP transformation. Put

$$p_1 = 1, \qquad q_1 = k_1(x), \qquad r_1 = k_2(x)$$

[4]This is defined below, and is not the same thing as *exponential decay of correlations*.

and for each $n > 1$

$$\begin{bmatrix} p_{n+1} \\ q_{n+1} \\ r_{n+1} \end{bmatrix} = \begin{bmatrix} p_{n-2} & p_{n-1} & p_n \\ q_{n-2} & q_{n-1} & q_n \\ r_{n-2} & r_{n-1} & r_n \end{bmatrix} \begin{bmatrix} 1 \\ k_1(T^n(x)) \\ k_2(T^n(x)) \end{bmatrix},$$

where the initial conditions $p_{-1} = p_0 = q_0 = r_{-1} = 0$ and $q_{-1} = r_0 = 1$ have been adopted. Obviously, each of the sequences $\{p_n\}, \{q_n\}, \{r_n\}$ is increasing. Then the nth JP approximation to x is the pair

$$\left(\frac{p_n}{r_n}, \frac{q_n}{r_n}\right).$$

Each triple (p_n, q_n, r_n) is in one to one correspondence with an admissible string $\alpha^{(n)} = (k_1(x), k_2(x)), \ldots, (k_1(T^{n-1}(x)), k_2(T^{n-1}(x)))$, so that each $x \in B_{\alpha^{(n)}} = T^{-n}_{\alpha^{(n)}}[0,1]^2$ has the same nth JP approximation. The *rate of approximation* of the algorithm is the the maximal number γ such that

$$\max\left\{\left|\frac{p_n}{r_n} - x_1\right|, \left|\frac{q_n}{r_n} - x_2\right|\right\} \leq \frac{const}{{r_n}^{1+\gamma}}$$

for m–a.e. pair (x_1, x_2). If $\gamma > 0$, then the JP algorithm is said to *converge exponentially*. (In the case of one–dimensional approximation by ordinary continued fractions, the analogous value of γ is 1, so that the continued fraction algorithm converges exponentially). Ito et. al. [19] prove that $\gamma > 0$ for the modified Jacobi–Perron algorithm, and numerical estimates of γ for the JP algorithm (which suggest exponential convergence) are given in [3]. In view of the argument in [19], knowledge of the invariant density for T is a helpful step towards rigorous estimates of γ for JP.

4.1 Regularity of the Jacobi–Perron transformation

We now show that the T satisfies (1)—(4). The following estimates from Schweiger's monograph are sufficient to derive suitable constants:

Lemma 4.1 (Basic properties of JP [34]) *Let T be the Jacobi–Perron transformation, and for each $n > 0$ let $\alpha^{(n)}$ be an admissible string. If (p_n, q_n, r_n) is the corresponding triple of integers, then*

1. *for each $x = (x_1, x_2)$,*

$$T^{-n}_{\alpha^{(n)}}(x) = \left(\frac{p_{n-2}x_1 + p_{n-1}x_2 + p_n}{r_{n-2}x_1 + r_{n-1}x_2 + r_n}, \frac{q_{n-2}x_1 + q_{n-1}x_2 + q_n}{r_{n-2}x_1 + r_{n-1}x_2 + r_n}\right),$$

2. *also*

$$|\det DT^{-n}_{\alpha^{(n)}}(x)| = \frac{1}{(r_{n-2}x_1 + r_{n-1}x_2 + r_n)^3},$$

3. *and*

$$\max\left\{\left|\frac{p_n}{r_n} - \frac{p_{n-1}}{r_{n-1}}\right|, \left|\frac{p_n}{r_n} - \frac{p_{n-2}}{r_{n-2}}\right|, \left|\frac{q_n}{r_n} - \frac{q_{n-1}}{r_{n-1}}\right|, \left|\frac{q_n}{r_n} - \frac{q_{n-2}}{r_{n-2}}\right|\right\} \leq \left(\frac{8}{9}\right)^{\lfloor n/2 \rfloor}.$$

Remark. Part (1) is Lemma 1.2, Part (2) is Lemma 2.4 and Part (3) follows from the proof of Theorem 9.6 in [34]. □

Similar to [34], we can estimate the expansivity and distortion constants of the JP transformation. These bounds are sufficient for our applications, but not optimal.

Corollary 4.2 *For each inverse branch $T^{-n}_{\alpha^{(n)}}$ of T^n:*

1. *for any $x, y \in T^n(B_{\alpha^{(n)}})$,*

$$\left|\frac{\det DT^{-n}_{\alpha^{(n)}}(x)}{\det DT^{-n}_{\alpha^{(n)}}(y)}\right| \leq 3^3;$$

2. *while*

$$\left|DT^{-n}_{\alpha^{(n)}}\right| \leq 2\left(\frac{8}{9}\right)^{\lfloor n/2 \rfloor},$$

3. *and for each x,*

$$\left|\nabla \det DT^{-n}_{\alpha^{(n)}}(x)\right| \leq 3\sqrt{2}|\det DT^{-n}_{\alpha^{(n)}}(x)|.$$

Proof. The first part follows immediately from Lemma 4.1 (2) since

$$0 \le x_1, x_2, y_1, y_2 \le 1 \qquad \text{and} \qquad r_{n-2} \le r_{n-1} \le r_n$$

(c.f. [34, Lemma 2.6]).

For Part (2), notice that

$$\frac{\partial}{\partial x_1}(T^{-n}_{\alpha^{(n)}}(x))_1 = \frac{r_{n-2}}{r_{n-2}x_1 + r_{n-1}x_2 + r_n}\left(\frac{p_{n-2}}{r_{n-2}} - (T^{-n}_{\alpha^{(n)}}(x))_1\right).$$

By Lemma 4.1 (1), $(T^{-n}_{\alpha^{(n)}}(x))_1$ can be written as a convex combination of

$$\frac{p_{n-2}}{r_{n-2}}, \frac{p_{n-1}}{r_{n-1}}, \frac{p_n}{r_n}.$$

Therefore, since the sequence $\{r_n\}$ is increasing and $0 \le x_1, x_2 \le 1$,

$$\left|\frac{\partial}{\partial x_1}(T^{-n}_{\alpha^{(n)}}(x))_1\right| \le \max\left\{\left|\frac{p_n}{r_n} - \frac{p_{n-1}}{r_{n-1}}\right|, \left|\frac{p_n}{r_n} - \frac{p_{n-2}}{r_{n-2}}\right|\right\} \le \left(\frac{8}{9}\right)^{\lfloor n/2 \rfloor}$$

by Lemma 4.1 (3). The other entries of $DT^{-n}_{\alpha^{(n)}}$ can be estimated in the same way, and this part of the corollary follows from standard properties of the matrix 2–norm.

The last part of the corollary also follows from Lemma 4.1 (1) by differentiation. □

4.2 Boundary estimates on a triangle

To apply the main results of this paper we need knowledge of the remaining constants appearing in Theorem 1. For this, we give a precise statement of Lemma 1.2 for a triangle in $\mathbb{R}^2$:

Lemma 1.2 revisited *Let $S \subset \mathbb{R}^2$ be a non–degenerate triangle. Then if a is the sine of half the minimal angle at a vertex of S and δ is the minimal distance from the central point of S to ∂S:*

$$\int_{\partial S} |tr^+_{\partial S} g| \le \frac{1}{1-c}\frac{1}{a}\left(\int_{int(S)} |dg| + \frac{1}{c(1-c)}\frac{1}{a\delta}\int_S |g|\right)$$

for every $c \in (0,1)$ and $g \in BV(S)$. □

REMARK. One corollary of this lemma is that for $a' < a(S)$, every $\epsilon < \delta$ is small enough for (S, a'). In the notation of the statement of the lemma in section 1, $\delta' = (1 - a'/a(S))\delta$. □

We now prove the lemma. This case is needed for the JP transformation, and is easily extended to arbitrary star-like regions in $\mathbb{R}^r$. The proof is motivated by the geometric construction in [16].

Proof of lemma: Let $S \subset \mathbb{R}^2$ be a non–degenerate triangle. We prove the lemma for $0 \le g \in C^1$; the general case follows by separating positive and negative parts and using an approximation argument. First, we establish some notation: let $\{L_y\}_{y \in \partial S}$ be the field of line segments from points $y \in \partial S$ to the central point of the triangle; see Figure 1.

For each $y \in \partial S$ there exists a vector w_y such that

$$L_y = \{y + t\, w_y : t \in [0,1]\}.$$

Let

$$\delta(S) \triangleq \inf_{y \in \partial S} |w_y| \tag{14}$$

be the minimal length of a segment L_y. If the face of ∂S containing y has unit normal vector v, then w_y meets ∂S at an angle θ_y such that

$$\sin\theta_y = \frac{w_y \cdot v}{|w_y|}.$$

If $\beta(S)$ is half the minimal angle at a vertex of S, then

$$a(S) \triangleq \sin\beta(S) \le |\sin\theta_y| \tag{15}$$

for each $y \in \partial S$. The constant $a(S)$ is the same as in [16].

Now, because g is continuous, for each $y \in \partial S$,

$$tr_{\partial S}^{+} g(y) = \lim_{t \to 0} \frac{1}{t} \int_{s=0}^{t} g(z + s\, w_y)\, ds.$$

Letting $g_y(s) = g(y + s\, w_y)$, it follows from standard properties of one–dimensional variation $\underset{[a,b]}{\mathrm{var}}(\cdot)$ that for $c \in (0,1)$,

$$\begin{aligned} tr_{\partial S}^{+} g(y) &\leq \inf_{s \in [0,c]} g_y(s) + \underset{[0,c]}{\mathrm{var}}(g_y) \\ &\leq \frac{1}{c} \int_{s=0}^{c} g_y(s)\, ds + \int_{s=0}^{c} |g_y'|\, ds \\ &= \frac{1}{c} \int_{s=0}^{c} g_y(s)\, ds + \int_{s=0}^{c} |\nabla g \cdot w_y|\, ds \\ &\leq |w_y| \left(\frac{1}{c\delta(S)} \int_{s=0}^{c} g_y(s)\, ds + \int_{s=0}^{c} |\nabla g|\, ds \right), \end{aligned}$$

by (14). Therefore, by integrating over ∂S and applying Fubini's Theorem, one obtains

$$\begin{aligned} \int_{\partial S} tr_{\partial S}^{+} g &\leq \frac{1}{c\delta(S)} \int_{\{(y,s) \in \partial S \times [0,c]\}} g(y + s\, w_y)\, |w_y|\, d\nu(y,s) \\ &\quad + \int_{\{(y,s) \in \partial S \times [0,c]\}} |\nabla g(y + s\, w_y)|\, |w_y|\, d\nu(y,s), \end{aligned} \tag{16}$$

where $d\nu(y,s) = dm_{2,\partial S \times [0,c]}(y,s) = dm_{1,[0,c]}(s) \times dm_{1,\partial S}(y)$ ($m_{d,W}$ denotes the d–dimensional Lebesgue measure on the subset $W \subset \mathbb{R}^r$).

Next, recall that $0 < c < 1$ and put

$$S' = S'(c) \triangleq \cup_{y \in \partial S} \{y + s\, w_y : s \in [0,c]\} \subset S,$$

and let $\Phi : \partial S \times [0,c] \to S'$ be the diffeomorphism defined by

$$\Phi(y,s) = y + s\, w_y.$$

To bound the rhs of (16) by integrals over $S' \subset S$, we must estimate the distortion of Φ.

Suppose that the situation is as depicted in Figure 1, with $y \in \partial S$ contained a face aligned in the e_1 coordinate direction (horizontal direction, e_2 is the vertical direction). Then

$$w_y = \delta_1 \left(\cot\theta_y\, e_1 + e_2 \right)$$

and

$$\Phi(y,s) = y + s\, \delta_1 \left(\cot\theta_y\, e_1 + e_2 \right).$$

One can easily check that $\cot\theta_{y+\Delta y e_1} - \cot\theta_y = -\Delta y/\delta_1 + o(\Delta y)$ so that

$$\frac{\partial}{\partial y} \Phi(y,s) = \lim_{\Delta y \to 0} \frac{\Phi(y + \Delta y, s) - \Phi(y,s)}{\Delta y} = (1-s) e_1.$$

On the other hand,

$$\frac{\partial}{\partial s} \Phi(y,s) = \lim_{\Delta s \to 0} \frac{\Phi(y, s + \Delta s) - \Phi(y,s)}{\Delta s} = \delta_1 \left(\cot\theta_y\, e_1 + e_2 \right).$$

Therefore,

$$\det D\Phi(y,s) = \begin{vmatrix} (1-s) & 0 \\ \delta_1 \cot\theta_y & \delta_1 \end{vmatrix} = \delta_1 (1-s) = \sin\theta_y\, |w_y|\, (1-s).$$

Then, for all $(y,s) \in \partial S \times [0,c]$,

$$\begin{aligned} |w_y|\, d\nu(y,s) &= \frac{|w_y|}{\det D\Phi(y,s)} dm_{2,S'}(\Phi(y,s)) \\ &= \frac{1}{1-s} \frac{1}{\sin\theta_y} dm_{2,S'}(\Phi(y,s)) \\ &\leq \frac{1}{1-c} \frac{1}{a(S)} dm_{2,S'}(\Phi(y,s)), \end{aligned} \tag{17}$$

by equation (15). Combining (16) and (17) we have (by the change of variables formula for integration):

$$\int_{\partial S} tr^{+}_{\partial S} g \leq \frac{1}{1-c}\frac{1}{a(S)}\int_{S'}|dg| + \frac{1}{c(1-c)}\frac{1}{a(S)\delta(S)}\int_{S'} g.$$

Since $S' \subset S$ and c is arbitrary, the lemma follows. □

EXAMPLE: Let S be the right–angled triangle with side-lengths $1, 1, \sqrt{2}$. Then the minimal interior angle at a vertex is $\pi/4$, so that

$$a(S) = \sin\frac{\pi}{8} = \frac{1}{\sqrt{4+2\sqrt{2}}}.$$

One can check that

$$\delta(S) = 1 - \frac{1}{\sqrt{2}}.$$

□

4.3 Application of results

Finally, we can apply Theorems 1—5 to the JP transformation.

Theorem 6 *Let T be the Jacobi–Perron transformation on $\mathbb{R}^2$, and let $\mathcal{L}$ be its Perron–Frobenius operator. Then*

1. *For every $n > 0$ and $g \in BV$,*

$$\begin{aligned} V(\mathcal{L}^n g) &\leq 2\left(\frac{8}{9}\right)^{\lfloor n/2 \rfloor} \times 5.021 V(g) \\ &\quad + 60.52 \times \|g\|_{L^1}. \end{aligned}$$

2. *T has an invariant density ϕ_* with*

$$V(\phi_*) \leq 60.52\|g\|_{L^1}$$

3. *Let $n_* = 280$, $M_* = 363$. Then*

$$\|\mathcal{L}^n g\|_{BV} \leq 485.3\left(\frac{215}{216}\right)^{n/n_*}\|g\|_{BV}$$

   for all $g \in BV(A)$.

4. *if $n_0 \geq 40$, then the Ulam's method with the operator*

$$\Pi_\eta \circ \mathcal{L}^{n_0}$$

   converges to an invariant density for T as the regular rectangular partition η is refined. The approximation error is at most $O(-h(\eta)\log h(\eta))$.

Proof. The first part follows from Theorem 1, Corollary 4.2 and Lemma 1.2:

$$V_A(\mathcal{L}^n f) \leq 2\left(\frac{8}{9}\right)^{\lfloor n/2 \rfloor}(1 + 1/ca) + \left(3\sqrt{2}(1+1/ca) + \frac{1}{c(1-c)}\delta a\right).$$

For the triangle S above, $a(S) = 1/\sqrt{4+2\sqrt{2}}, \delta(S) = 1/(2+\sqrt{2})$, and we choose $c = 0.65$. Lemma 1.2 also holds for the square with these constants (since it's angles and diameter are larger). Putting these numbers in the above equation yields the required result.

Since T has a finite range structure, the existence of ϕ_* follows from Theorem 2 (ii). Since $\mathcal{L}^n\phi_* = \phi_*$ for all $n \geq 0$,

$$V(\phi_*) = \lim_{n\to\infty} V(\mathcal{L}^n \phi_*) \leq 60.52\|\phi_*\|_{L^1}$$

by part (i).

Rate of mixing: Choose $n_1 = 40$. Part (i) of the theorem gives a Lasota–Yorke inequality with $\sigma_1 = 0.5, K_1 = 60.52$. We thus obtain $M_* = 363$, and using $\beta = 1/2, C_1 = 2, C_2 = 2\sqrt{3}$, we have $n_0 \leq 192$. Therefore, we choose $n_2 = 5$ to ensure that $n_1 n_2 > n_0$. Finally,put $n_3 = 2$. To check that this n_3 is big enough (as preceding Theorem 3), simply observe that for either range set, $H(\partial S) \le 4$ and $m(S) \le 1$. The theorem now follows from Theorem 4 with $n_* = n_1(n_2 + n_3)$.

The final part follows from Theorem 5, because n_0 need only satisfy

$$C_1 \lambda^{-n_0} (1 + 1/a) < 1/\sqrt{2}.$$

With the JP constants, $n_0 = 40$ will suffice. □

From a quantitative point of view, the results presented in Theorem 6 are unimpressive; they have been included primarily to illustrate the practical application of the results in this paper. More detailed study of the expansion constants of the JP transformation would lead to better estimates.

A Proof of Lemma 2.6

The proof of the lemma is derived from a related result using a slightly different notion of variation. We describe this result, and give an indication of how to prove it. Let $\Gamma_i = \{x \in [0,1]^r : x_i = 0\}$, and for each $z \in \Gamma_i$ let $f_{i,z}(s) = f(z + s\,e_i)$, where e_i is the coordinate vector orthogonal to Γ_i. Thus, each $f_z : [0,1] \to \mathbb{R}$. Then let,

$$V_A^1(f) = \sum_{i=1}^{r} \int_{\Gamma_i} \operatorname*{var}_{[0,1]}(f_{i,z})\, dz,$$

where var denotes the usual one–dimensional variation. In fact, from [22],

$$V_A(f) \le V_A^1(f) \le \sqrt{r} V_A(f).$$

Therefore, Lemma 2.6 follows from the following:

Lemma A.1 *Let η be a regular rectangular partition of $[0,1]^r$ into cubes of side–length h, and let Π_η be the associated projection. Then*

1. $V_A^1(\Pi_\eta f) \le V_A^1(f)$;
2. $\|f - \Pi_\eta f\|_{L^1} \le h V_A^1(f)$.

The first part of the lemma is reasonably well–known [32, Lemma 3.3], and can be deduced from the equivalent result in one–dimension. The second part is not too surprising, and also comes from the one–dimensional result. We now give a sketch proof, as the details are a little tedious.

Sketch proof of part (ii). The idea is to write Π_η as a sequence of "one–dimensional" projections. Let $h = 1/N$, and let $B(x) \in \eta$ be the unique B such that $x \in B$. Then $B(x)$ can be written as; $B(x) = \Pi_{i=1}^{r}[k_i(x)/N, (k_i(x)+1)/N)$. Now, for each i, one can write $x = z + s\,e_i$, where $z \in \Gamma_i$ and $k_i(x) \le sN < k_i(x) + 1$, so that $f(x) = f_{i,z}(s)$. Then, define projection operators Q_i by their pointwise action:

$$(Q_i f)(x) = N \int_{k_i(x)/N}^{(k_i(x)+1)/N} f_{i,z}(s)\, ds.$$

For each $z \in \Gamma_i$, the function $(Q_i f)_z$ is simply the mass preserving projection of f_z onto the partition of $[0,1]$ into N equal subintervals. It follows from standard properties of one–dimensional variation that

$$\int_{s=0}^{1} |f_{i,z}(s) - (Q_i f)_{i,z}(s)|\, ds \le \frac{1}{N} \operatorname*{var}_{[0,1]}(f_{i,z}).$$

By integrating over Γ_i, one thus obtains (after using Fubini's theorem)

$$\int_A |f - Q_i f|\, dm \le h \int_{\Gamma_i} \operatorname*{var}_{[0,1]}(f_{i,z})\, dz. \tag{18}$$

With a little care, one can also show [7] that when $j \ne i$,

$$\int_{\Gamma_j} \operatorname*{var}_{[0,1]}((Q_i f)_{j,z})\, dz \le \int_{\Gamma_j} \operatorname*{var}_{[0,1]}(f_{j,z})\, dz. \tag{19}$$

Now, put $g_0 = f$, and for each $i = 1, \ldots, r$ put $g_i = Q_i f_{i-1}$. In this notation, equation (18) says that

$$\int_A |g_i - g_{i-1}|\, dm \le h \int_{\Gamma_i} \operatorname*{var}_{[0,1]}((g_{i-1})_{i,z})\, dz.$$

Successive application of (19) yields

$$\int_{\Gamma_i} \operatorname*{var}_{[0,1]}((g_{i-1})_{i,z})\,dz \le \int_{\Gamma_i} \operatorname*{var}_{[0,1]}((g_0)_{i,z})\,dz.$$

Therefore,

$$\int_A |g_n - g_0|\,dm \le \sum_{i=1}^{r} \int_A |g_i - g_{i-1}|\,dm \le h \sum_{i=1}^{r} \int_{\Gamma_i} \operatorname*{var}_{[0,1]}((g_0)_{i,z})\,dz = h\,V_A^1(f).$$

However, by Fubini's theorem, $\Pi_\eta = Q_r \cdots Q_2 Q_1$. Therefore, $g_n = \Pi_\eta f$ and the lemma follows. □

B Other calculations

Proof of Lemma 1.1 The first part follows immediately from the definitions of variation and $\mathcal{L}$. For the second part,

$$\begin{aligned}
&\int_{int(TB_\alpha)} \left| d\frac{f\circ T_\alpha^{-1}}{|\det DT\circ T_\alpha^{-1}|}\right| \\
&\quad = \int_{int(TB_\alpha)} \left| \frac{d(f\circ T_\alpha^{-1})}{|\det DT\circ T_\alpha^{-1}|} + f\circ T_\alpha^{-1} d\left(\frac{1}{|\det DT\circ T_\alpha^{-1}|}\right)\right| \\
&\quad \le \int_{int(TB_\alpha)} \left| DT_\alpha^{-1}\left(\frac{(df)\circ T_\alpha^{-1}}{\det DT\circ T_\alpha^{-1}}\right)\right| \\
&\qquad + \int_{int(TB_\alpha)} \left| \frac{f\circ T_\alpha^{-1}}{\det DT\circ T_\alpha^{-1}}\frac{d\left(\det DT\circ T_\alpha^{-1}\right)}{\det DT\circ T_\alpha^{-1}}\right| \\
&\quad \le C_1\lambda^{-1}\int_{int(TB_\alpha)} \left|\frac{(df)\circ T_\alpha^{-1}}{\det DT\circ T_\alpha^{-1}}\right| \\
&\qquad + C_2\int_{int(TB_\alpha)} \left|\frac{f\circ T_\alpha^{-1}}{\det DT\circ T_\alpha^{-1}}\right| \\
&\quad = C_1\lambda^{-1}\int_{int(B_\alpha)} |df| + C_2\int_{B_\alpha} |f|,
\end{aligned}$$

by (1), (2) and the change of variables formula for integration. □

Proof of Proposition 1.3 Without loss of generality assume that f is C^1 on each $int(B_\alpha)$. Thus,

$$\sum_\alpha \int_{int(B_\alpha)} |\nabla f|\,dm \le V_A(f). \tag{20}$$

Now, by Lemmas 1.1 (1) and 1.2,

$$V_A(\mathcal{L}f) \le \sum_\alpha \left((1+1/a')\int_{int(TB_\alpha)} |\nabla\mathcal{L}(f\chi_{B_\alpha})|\,dm + 1/(a'\delta')\int_{TB_\alpha} |\mathcal{L}(f\chi_{B_\alpha})|\,dm\right).$$

Since $\int_{TB_\alpha} |\mathcal{L}(f\chi_{B_\alpha})|\,dm \le \int_{B_\alpha} |f|\,dm$, the proposition follows from (20), and Lemma 1.1 (2). □

Acknowledgement: This work was carried out at the Statistical Laboratory, University of Cambridge and the Centre for Nonlinear Dynamics, University College London. The author is grateful to two anonymous referees, whose insightful comments greatly improved the paper.

References

[1] K Adl-Zarabi. Absolutely continuous invariant measures for piecewise expanding $\mathcal{C}^2$ transformations on $\mathbb{R}^n$ on domains with cusps on the boundaries. *Ergod. Th. & Dynam. Sys.*, 16:1–18, 1996.

[2] P Arnoux and A Nogueira. Mesures de Gauss pour des algorithmes do fractions continues multidimensionnelles. *Ann. scient. Éc. Norm. Sup.*, 26:645–664, 1993.

[3] V Baladi and A Nogueira. Lyapunov exponents for non–classical multi–dimensional continued fraction algorithms. *Nonlinearity*, 9:1529–1546, 1996.

[4] M Blank and G Keller. Stochastic stability versus localisation in one–dimensional chaotic dynamical systems. *Nonlinearity*, 10:81–107, 1997.

[5] M L Blank. Stochastic properties of deterministic dynamical systems. *Sov. Sci. Rev. C: Math. Phys*, 6:243–271, 1987.

[6] M L Blank. Ergodic properties of a method of numerical simulation of a chaotic dynamical systems. *Math. Notes*, 45(3–4):267–273, 1989.

[7] C Bose, W Miller, and R Murray. Optimal inequalities for the behaviour of $BV(\mathbb{R}^r)$ under approximation by simple functions. Preprint, 1998.

[8] C J Bose. On the existence and approximation of invariant densities for nonsingular transformations on $\mathbb{R}^d$. *J. Approx. Theory*, 79:260–270, 1994.

[9] A Broise. Fractions continues multidimensionnelles et lois stables. *Bull. Soc. Math. France*, 124:97–139, 1996.

[10] J Buzzi. ACIMs for arbitrary piecewise $\mathbb{R}$–analytic mappings of the plane. Preprint, Institut de Mathématiques de Luminy, http://iml.univ-mrs.fr/ buzzi/, November 1998.

[11] Jiu Ding and Aihui Zhou. Finite approximations of Frobenius–Perron operators. A solution of Ulam's conjecture to multi–dimensional transformations. *Phys. D*, 92:61–68, 1996.

[12] P Ferrero and B Schmitt. Produits aléatoires d'opérateurs matrices de transfert. *Prob. Th. & Rel. Fields*, 79(2):227–248, 1988.

[13] G Froyland. Finite approximation of Sinai–Bowen–Ruelle measures for Anosov systems in two dimensions. *Random Comput. Dynam.*, 3(4):251–264, 1995.

[14] E Giusti. *Minimal surfaces and functions of bounded variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser, 1984.

[15] P Góra, 1997. Private communication.

[16] P Góra and A Boyarsky. Absolutely continuous invariant measures for piecewise expanding $\mathcal{C}^2$ transformations in $\mathbb{R}^n$. *Israel J. Math.*, 67(3):272–286, 1989.

[17] F Y Hunt. A Monte Carlo approach to the approximation of invariant measures. *Random Comput. Dynam.*, 2(1):111–133, 1994.

[18] C T Ionescu-Tulcea and G Marinescu. Theorie ergodique pour des classes d'operations non completement continues. *Annals of mathematics*, 52:140–147, 1950.

[19] S Ito, M Keane, and M Ohtsuki. Almost everywhere exponential convergence of the modified Jacobi–Perron algorithm. *Ergod. Th. & Dynam. Sys.*, 13:319–334, 1993.

[20] G Keller. Ergodicíte et mesures invariantes pour les transformations dilatantes par mairceaux d'une région bornée du plan. *C. R. Acad. Sci. Paris, Série A*, 289:625–627, 1980.

[21] G Keller. Stochastic stability in some chaotic dynamical systems. *Monatsh. Mathematik*, 94:313–333, 1982.

[22] G Keller and M Künzle. Transfer operators for coupled map lattices. *Ergod. Th. & Dynam. Sys.*, 12:297–318, 1992.

[23] A Lasota and M C Mackey. *Chaos, Fractals and Noise: stochastic aspects of deterministic dynamics.* Springer, 2 edition, 1994.

[24] A Lasota and J A Yorke. On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.*, 186:481–488, 1973.

[25] T-Y Li. Finite approximation for the Perron–Frobenius operator. a solution to Ulam's conjecture. *Journal of approximation theory*, 17:177–186, 1976.

[26] C Liverani. Decay of correlations. *Annals of Mathematics*, 142:239–301, 1995.

[27] C Liverani. Decay of correlations for piecewise expanding maps. *J. Statistical Physics*, 78(3/4):1111–1129, 1995.

[28] D H Mayer. Approach to equilibrium for locally expanding maps in $\mathbb{R}^k$. *Commun. Math. Phys.*, 95:1–15, 1984.

[29] S P Meyn and R L Tweedie. *Markov chains and stochastic stability.* Springer, 1993.

[30] R Murray. Approximation error for invariant density calculations. *J. Discrete & Continuous Dyn. Sys.*, 4:535–558, 1998.

[31] M Rychlik. Bounded variation and invariant measures. *Studia Mathematica*, 76:69–80, 1983.

[32] R Sanders. On convergence of monotone finite difference schemes with variable spatial differences. *Mathematics of Computation*, 40:91–106, 1983.

[33] B Saussol. Absolutely continuous invariant measures for multidimensional expanding maps. Preprint, Centre de Physique Théorique, Marseille: http://www.cpt.univ-mrs.fr/, December 1998.

[34] F Schweiger. *The metrical theory of the Jacobi–Perron algorithm*, volume 334 of *LNM.* Springer, 1973.

[35] S Ulam. *A collection of mathematical problems.* Interscience Publishers, 1960.

[36] M Yuri. Invariant measures for certain multi–dimensional maps. *Nonlinearity*, 7:1093–1124, 1994.

[37] M Yuri. Decay of correlations for certain multi–dimensional maps. *Nonlinearity*, 9:1439–1461, 1996.

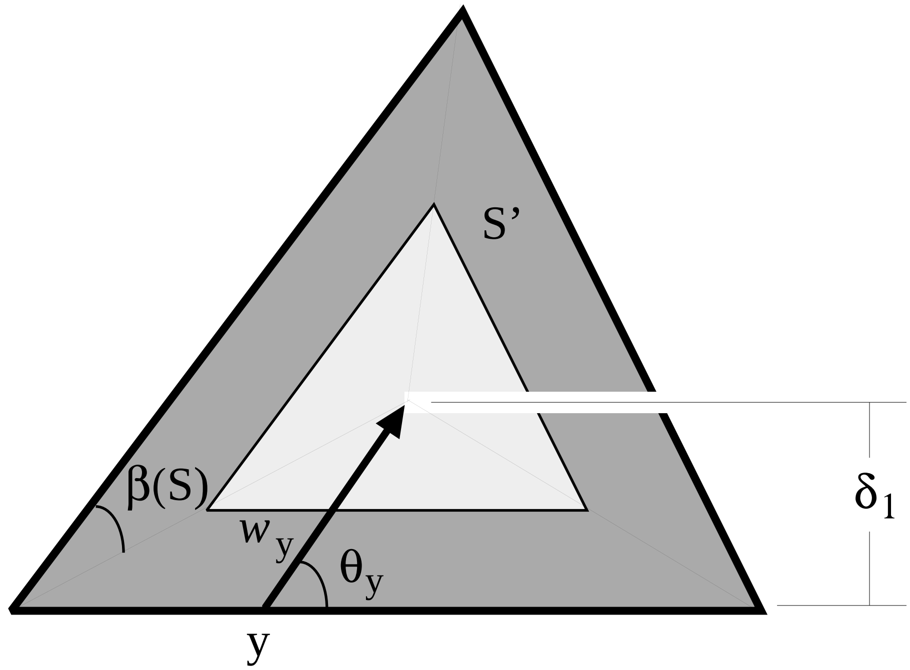

FIGURE 1: Construction of Lemma 1.2 on a triangle S in $\mathbb{R}^2$.