ERGODIC OPTIMIZATION OF PREVALENT SUPER-CONTINUOUS FUNCTIONS

JAIRO BOCHI AND YIWEI ZHANG

Abstract. Given a dynamical system, we say that a performance function has property P if its time averages along orbits are maximized at a periodic orbit. It is conjectured by several authors that for sufficiently hyperbolic dynamical systems, property P should be typical among sufficiently regular performance functions. In this paper we address this problem using a probabilistic notion of typicality that is suitable to infinite dimension: the concept of prevalence as introduced by Hunt, Sauer, and Yorke. For the one-sided shift on two symbols, we prove that property P is prevalent in spaces of functions with a strong modulus of regularity. Our proof uses Haar wavelets to approximate the ergodic optimization problem by a finite-dimensional one, which can be conveniently restated as a maximum cycle mean problem on a de Bruijn graph.

1. Introduction

1.1. Are optimal orbits typically periodic? Consider a continuous transformation $T: X \to X$ on a compact metrizable space, and a continuous “performance” function $f: X \to \mathbb{R}$. The main questions of ergodic optimization are: Along which orbits of $T$ are the time averages of $f$ largest? What is the statistics of those orbits? Equivalently, for which invariant probability measures $\mu$ is average performance $\int f \, d\mu$ maximized? What properties do these so-called maximal measures have? What happens as $f$ changes? The subject was nicely surveyed by Jenkinson [Je3]. Some maximal measures of special interest in thermodynamic formalism are the so-called ground states: see [BL].

If the dynamical system $T$ has few invariant measures then the questions above become uninteresting. So it is natural to focus the study on chaotic systems, as for example uniformly hyperbolic or uniformly expanding ones. These systems possess many invariant probability measures, the simplest of which are the periodic measures (i.e. those supported on periodic orbits). Let us say that the performance function has property P with respect the dynamical system $T$ if it posses a periodic maximal measure.

Based on experimental results and heuristic arguments, Hunt and Ott [HO] conjectured that for typical chaotic systems and typical parameterized families of smooth functions, property P holds for most values of the parameter. Here “most” means full Lebesgue measure in the parameter space (but the meanings of “typical” are not specified). One of the examples studied by these authors consisted on the doubling map $T(\theta) = 2\theta$ on the circle $\mathbb{R}/2\pi\mathbb{Z}$ together with the family of functions

\textit{Date:} January 6, 2015.

2010 Mathematics Subject Classification. 37D20, 90C05, 60B11, 52B12.

J.B. was partially supported by projects Fondecyt 1140202 and Anillo ACT1103 (Center of Dynamical Systems and Related Fields). Y.Z. was supported by project Fondecyt 3130622.
formed by linear combinations of cosines and sines. For this specific case, the conjecture was later confirmed by Bousch \cite{Bo1}. (See also \cite{Je1}.)

A different conjecture by Yuan and Hunt \cite{YH} is that for every uniformly hyperbolic or uniformly expanding map, property $P$ holds for \textit{topologically generic $C^1$} performance functions, i.e., it holds for every function in a dense $G_\delta$ subset. As these authors observe, the interior of property $P$ is dense in itself, so if the conjecture is true then property $P$ actually holds for an open and dense subset of the space of $C^1$ functions.

Regularity assumptions are important. Assume that $T$ is a “hyperbolic” (e.g., expanding) map. Then maximizing measures of topologically generic continuous have full support (see \cite{Bo2, Je3, Mo2}); in particular the Yuan–Hunt conjecture fails dramatically in the space of continuous functions. On the other hand, every sufficiently regular function (e.g., Hölder ones) has the following agreeable property: there exists a compact invariant set $K$ such that an invariant probability measure $\mu$ is maximizing if and only if $\mu(K) = 1$. This fact and related ones (especially the so-called Mañé lemma) were discovered independently by many authors in various settings: see \cite{Bo2, Bo3} and references therein.

Much research was done about property $P$ and related ones on spaces of sufficiently regular functions: \cite{YH, CLT, Bo2, Mo1, QS}. We highlight the recent work of Contreras \cite{Co1}, who showed that this property is open and dense among Lipschitz functions when the dynamics is expanding.

Let us also mention that many of the problems and results we have alluded to were originally formulated in the context of classical mechanics, where optimization is naturally related with the principle of least action: see \cite{CI} and references therein. There is a celebrated conjecture by Mañé \cite{Ma} on the genericity of property $P$ in those contexts, which was proved very recently by Contreras \cite{Co2}.

In this paper we further investigate the genericity of property $P$ in abstract ergodic optimization, not in the topological sense but in a \textit{probabilistic} one, thus in a spirit more akin to \cite{HO}. The first step is to define “probabilistic genericity” in a space of functions. Though a possible approach for this is suggested in \cite{HO}, we follow another route which seems more appropriate.

Probabilistic genericity should have something to do with having full measure. Unfortunately, topological vector spaces do not have any natural reference measures, except when they are finite-dimensional. Worse, there is no natural equivalence class of measures. On the positive side, what does exist (under minor technical assumptions) is a natural class of “fat” subsets, which corresponds in the finite-dimensional case to the class of subsets of full Lebesgue measure. This is the class of \textit{prevalent} sets discovered by Hunt, Sauer, and Yorke \cite{HSY}, though it was previously introduced in a slightly less general setting by Christensen \cite{Ch}. The surveys \cite{OY, HK} contain much information on the subject, including generalizations, applications, and open problems in many areas – especially Dynamical Systems.

In general, prevalence and topological genericity are neither one stronger than the other; indeed the references above contain examples of many properties of natural interest that are generic in one sense but not in the other. However, for properties known to be open (or to contain an open dense subset), topological genericity means only denseness, which is always weaker than prevalence.
Thus we are inevitably led to the following question: Does a prevalent version of the theorem of Contreras [Co1] hold? In other words, if the dynamics is expanding, is property $P$ prevalent among Lipschitz functions?

Though being currently unable to solve the problem in full generality, we do answer the question positively under further assumptions. First, we deal with the simplest interesting dynamics in this context: the full one-sided shift on two symbols. Second, we work on spaces of super-continuous functions, i.e. functions that have a very strong modulus of regularity. The setting comes from the work of Quas and Siefken [QS], who proved denseness of property $P$ in some of these spaces.

Let us proceed with the exact definitions and statements.

### 1.2. Precise statements

Let $V$ be a completely metrizable vector space. Following [HSY], we say that a Borel set $S \subset V$ is shy if there exists a compactly supported Borel probability measure $\mu$ such that $\mu(S + v) = 0$ for every $v \in V$.

In that case we say that $\mu$ is transverse to $S$.$^1$

It is easy to see that if a set $S$ is shy then it has measures transverse to it with support of arbitrarily small diameter. Therefore, given any element of $V$ it is possible to add to it a small random perturbation so that it falls outside of $S$ with probability 1.

A set whose complement is contained in a shy set is called a prevalent set. Prevalence has the following properties (see [HSY]): it is preserved under translation, augmentation, and countable intersections, it implies denseness, and it coincides with the notion of having full Lebesgue measure when the ambient space is finite-dimensional.

Next, let us state more formally the setting for ergodic optimization. Let $(X, d)$ be a compact metric space, and $T: X \to X$ be a continuous map. Let $\mathcal{M}_T$ denote the set of $T$-invariant Borel probability measures. If $\mu \in \mathcal{M}_T$ is supported on a periodic orbit then it is called a periodic measure. Let $C^0(X)$ denote the space of real-valued continuous functions on $X$. The ergodic supremum of $f \in C^0(X)$ is defined as

\[
\text{erg sup}(f) := \sup_{\mu \in \mathcal{M}_T} \langle f, \mu \rangle ,
\]

where angle brackets denote integration. A measure $\mu \in \mathcal{M}_T$ where this supremum is attained is called a maximizing measure for $f$. By continuity and compactness, maximizing measures always exist. The ergodic supremum can also be expressed in terms of maximal time averages of $f$: see [Je3]. If one of these if periodic then we say that $f$ has property $P$.

Let $C^{\text{Lip}}(X) \subset C^0(X)$ be the set of Lipschitz functions, endowed with the Lipschitz norm $\| \cdot \|_{\text{Lip}} := \| \cdot \|_\infty + \text{Lip}(\cdot)$, i.e. the sum of the supremum norm and the least Lipschitz constant. Then $C^{\text{Lip}}(X)$ is a Banach space, and it is nonseparable unless $X$ is countable, because it contains the discrete subset $\{d(x, \cdot) : x \in X\}$.

$^1$If the space $V$ is separable then the compactness hypothesis is basically avoidable. However we will deal with non-separable spaces.
We say that a function \( f \in C^{\text{Lip}}(X) \) has the locking\(^2\) property if it has a unique maximizing measure \( \mu \) that is periodic (in particular, \( f \) has property \( P \)) and moreover \( \mu \) is the unique maximizing measure for every function sufficiently close to \( f \) in the Lipschitz norm. We can also say that \( f \) and all functions close to it are locked to the measure \( \mu \). Of course, the locking property depends on the metric \( d \).

The set of functions with the locking property is exactly the interior of the set of functions with property \( P \): one inclusion is obvious from the definition, and the other was shown by Yuan and Hunt \([YH, \text{Remark 4.5}]^3\). In particular, the locking property is dense among the functions with property \( P \). We will henceforth focus our attention on the locking property.

We work in the symbolic space \( \Omega := \{0, 1\}^\mathbb{N} \), where \( \mathbb{N} \) includes 0. Suppose \( a = (a_n)_{n \in \mathbb{N}} \) is a strictly decreasing sequence of positive numbers converging to zero, or in symbols \( a \searrow 0 \). Then we define a metric \( d_a \) (actually an ultrametric) on \( \Omega \) as follows: given two sequences \( x = (x_n), y = (y_n) \in \Omega \), if they are different we set \( d_a(x, y) := a_m \) where \( m \) is their first disagreement (i.e., the least position where the symbols differ), otherwise we set \( d_a(x, y) := 0 \). All metrics of this form induce the product topology on \( \Omega \).

Let \( C^a(\Omega) := C^{\text{Lip}}(\Omega) \) where \( \Omega \) is endowed with the metric \( d_a \); in other words \( C^a(\Omega) \) is the space of functions on \( \Omega \) that are Lipschitz with respect to the metric \( d_a \). As explained above, this is a non-separable Banach space.

Of course, the faster the sequence \( a \) decays, the more regular are the functions in \( C^a(\Omega) \). As an example, consider geometric sequences \( (\theta^n)_{n \in \mathbb{N}} \) with factor \( \theta \in (0, 1) \); then \( f \) is \( \alpha \)-Hölder with respect to the metric \( d_{\theta^n} \) if and only if \( f \in C^{(\theta^n)}(\Omega) \).

Let us consider ergodic optimization on the function spaces \( C^a(\Omega) \) with respect to the shift dynamics \( \sigma : \Omega \to \Omega \). Let \( L_a \subset C^a(\Omega) \) denote the set of \( d_a \)-Lipschitz functions with the locking property.

It was proved by Quas and Siefken \([QS]\) that if the sequence \( a = (a_n)_{n \in \mathbb{N}} \) is such that \( \lim a_{n+1}/a_n = 0 \) then \( L_a \) is dense (and thus topologically generic) in \( C^a(\Omega) \). Contreras \([Co1]\) obtains the same conclusion under the weaker condition \( \sup a_{n+1}/a_n < 1 \). (Actually, Quas and Siefken deal with one-sided shifts in an arbitrary finite number of symbols, while Contreras deals with arbitrary expanding maps.)

Our aim is to replace denseness by prevalence in these kind of results. However we need an even stronger modulus of regularity. Let us say that a sequence \( a = (a_n)_{n \in \mathbb{N}} \) is evanescent if \( a \searrow 0 \) and

\[
\frac{a_{n+1}}{a_n} = O(2^{-2^{n+2}}).
\]

Our main result is as follows:

**Theorem A.** If \( a \) is an evanescent sequence then \( L_a \) is a prevalent set in \( C^a(\Omega) \).

\(^2\)This is a free translation of the verrouillage term used in \([Bo2, \S 8]\).

\(^3\)Actually this exact statement is not proved in the paper \([YH]\) but it follows by an adaptation of the arguments. The interested reader may try to figure out the details himself or check our note [www.mat.uc.cl/~jairo.bochi/docs/lock.pdf](http://www.mat.uc.cl/~jairo.bochi/docs/lock.pdf).
Since the set $L_a$ is open by definition, and every prevalent set is dense, we reobtain its topological genericity under the strong assumption (1.2). We stress that our methods are very different from those of \[QS, \text{Co1}\], however.

Currently we are not able to weaken the evanescence condition (1.2) considerably. The difficulties will be discussed later. On the other hand, our methods should be easily extendable to shifts on an arbitrary finite number of symbols or more generally to subshifts of finite type. Favoring clarity at the expense of generality, we decided to work with shifts on 2 symbols only. This choice has the additional advantage of being able to express the random perturbations leading to the proof of Theorem A in a quite convenient way.

1.3. The random perturbations. Let us describe explicitly certain probability measures that are transverse to the complement of $L_a$. Therefore adding to any given function in $C^a(\Omega)$ a random perturbation chosen according to one of those probability measures one almost surely obtains a function in $L_a$.

Let $\Omega^*$ denote the set of all finite words in the letters 0, 1, including the empty word. Each $\omega \in \Omega^*$ determines a cylinder $[\omega]$ which is the set of elements of $\Omega$ that have $\omega$ as an initial subword. The level of the cylinder $[\omega]$ is defined as the length of the word $\omega$, which is denoted by $|\omega|$.

For each $\omega \in \Omega^*$, we define the Haar function:

$$h_\omega := \frac{\chi_{[\omega 0]} - \chi_{[\omega 1]}}{2},$$

where $\chi$ denotes characteristic functions and $\omega 0, \omega 1$ denote concatenated words. These functions are a straightforward adaptation of the classical Haar functions on $[0,1]$ (see e.g. \[Pi, \S 6.3\]). Working on the Cantor set $\Omega$ has an advantage over working on the interval $[0,1]$: our Haar functions (1.3) are continuous. Our unusual choice of the normalization makes subsequent formulas simpler.

Let us mention a few facts about these functions, leaving the details for §2. Haar functions, together with the constant function 1, form an orthogonal set with respect to a natural inner product. An arbitrary continuous function $f \in C^0(\Omega)$ can be written as a Haar series:

$$f = c(f) + \sum_{\omega \in \Omega^*} c_\omega(f) h_\omega,$$

where the convergence is uniform. (In this sense Haar series are better behaved than Fourier series.) Moreover, the Haar coefficients $c(f), c_\omega(f)$ are uniquely determined by $f$.

A gauge is a family $b = (b_\omega)_{\omega \in \Omega^*}$ of positive numbers indexed by words. Suppose $a = (a_n)$ is an evanescent sequence in the sense (1.2). We say that a gauge $b = (b_\omega)$ is admissible with respect to $a$ if

$$\bar{b}_n = o(a_n) \quad \text{and} \quad \log(a_n/b_n) = O(n),$$

where

$$\bar{b}_n := \max_{|\omega| = n} b_\omega \quad \text{and} \quad \underline{b}_n := \min_{|\omega| = n} b_\omega.$$

In other words, $b_\omega$ becomes much (but not superexponentially) smaller than $a_n$ as $n = |\omega| \to \infty$. Examples of admissible gauges are $\bar{b}_\omega = n^{-1}a_n$ and $\underline{b}_\omega = 2^{-n}a_n$. 

Lemma 1.1. Suppose that $a = (a_n)$ is an evanescent sequence, and that $b = (b_\omega)$ is an admissible gauge with respect to $a$. Given numbers $c_\omega$ satisfying $|c_\omega| \leq b_\omega$ for all $\omega \in \Omega^*$, the Haar series $\sum_{\omega \in \Omega^*} c_\omega h_\omega$ converges uniformly to a function in $C^a(\Omega)$. Moreover, the functions obtained in this way form a compact subset $\mathcal{H}_b$ of $C^a(\Omega)$.

The set of functions $\mathcal{H}_b$ as in the lemma above is called the Hilbert brick with gauge $b$. By means of the Haar coefficients, it can be identified with the product space $\prod_{\omega \in \Omega^*} [-b_\omega, b_\omega]$. We endow each interval $[-b_\omega, b_\omega]$ with the normalized Lebesgue measure, take the product probability, and using the identification obtain a probability measure $P_b$ on the Banach space $C^a(\Omega)$ supported on the compact subset $\mathcal{H}_b$.

The main technical result of this paper is the following:

Theorem B. Suppose that $a$ is an evanescent sequence and that $b$ is an admissible gauge with respect to $a$. Then for any $f_0 \in C^a(\Omega)$,

$$P_b(\{g \in \mathcal{H}_b : f_0 + g \in L_a\}) = 1.$$ 

The conclusion of the theorem is that the measure $P_b$ is transverse of the complement of the set $L_a$. In particular $L_a$ is prevalent. So Theorem B implies Theorem A.

Remark 1.2. Based on experimental evidence, Hunt and Ott [HO] also conjecture that for typical parameterized families of functions, the Lebesgue measure of the parameters corresponding to maximizing measures with period $p$ or greater decreases exponentially with $p$. Using our methods it is straightforward to construct spaces of Lipschitz functions and measures on Hilbert bricks (which can be regarded as families on infinitely many parameters) such that the measure of the parameters corresponding to maximizing measures with period $p$ or greater decays faster than a given function of $p$. Such result does not seem especially interesting, so we will not provide details.

1.4. Main ideas of the proof of Theorem B and organization of the paper.

Due to high regularity, the functions considered in this paper are well-approximated by step functions. Equivalently, their Haar coefficients decay fast. The strategy of the proof is to show that with high probability the maximal measure for the function $f$ is the same as for a step function obtained by a truncation of the Haar series.

The paper is organized as follows: In Section 2 we prove a few simple facts about Haar series and the compactness of Hilbert bricks (Lemma 1.1). In Section 3 we study in detail some finite-dimensional projections of the set of invariant measures (rotation sets). We describe these sets in graph-theoretical terms. This information is basic to the arguments that follow. In Section 4 we obtain a sufficient gap condition for the locking property; it is basically a transversality condition in terms of finitely many Haar coefficients. In Section 5 we use evanescence (1.2) and admissibility (1.4) to show that with probability 1 the transversality condition will be satisfied at some level, thus proving Theorem B. The final Section 6 contains a few open questions and discusses the obstacles to improving the evanescence condition.

Acknowledgement. We thank Juan Rivera-Letelier for useful discussions.
2. More about Haar series and Hilbert bricks

In this section we collect a few facts about Haar series and we prove Lemma 1.1.

We fix some notation. Let \( x = (x_0, x_1, \ldots) \) and \( y = (y_0, y_1, \ldots) \) be two elements of \( \Omega = \{0, 1\}^\mathbb{N} \). Let \( x \uparrow y \) denote the position of first disagreement between the sequences \( x \) and \( y \), that is, the least \( m \in \mathbb{N} \) such that \( x_m \neq y_m \), with the convention \( x \uparrow y = \infty \). For each \( n \in \mathbb{N} \), we define the \( n \)-th variation of a function \( f : \Omega \to \mathbb{R} \) as:

\[
\text{var}_n(f) := \sup_{x \uparrow y \geq n} |f(x) - f(y)|.
\]

If the function \( f \) satisfies \( \text{var}_n(f) = 0 \), i.e., it is constant in each cylinder of level \( n \), then it is called a step functions of level \( n \). Such functions form a vector space \( S_n \) of dimension \( 2^n \).

Recall the definition (1.3) of the Haar functions \( h_{\omega} \), where \( \omega \in \Omega^* \). For each \( n \in \mathbb{N} \), the set \( \{1\} \cup \{h_{\omega} : |\omega| < n\} \) forms a basis of the vector space \( S_n \).

Let \( \beta \) be the unbiased Bernoulli measure on \( \Omega \), i.e. the probability measure that assigns equal weights to all cylinders of the same level. Let \( L^2(\beta) \) denote the Hilbert space of functions that are square-integrable with respect to the measure \( \beta \). Then the set \( \{1\} \cup \{h_{\omega} : \omega \in \Omega^*\} \) is an orthogonal basis of \( L^2(\beta) \). Thus every \( f \in L^2(\beta) \) can be represented by a Haar series:

\[
f = c(f) + \sum_{\omega \in \Omega^*} c_{\omega}(f) h_{\omega} \quad \text{(equality in } L^2(\beta)),
\]

where the Haar coefficients are defined as:

\[
c(f) := \int f \, d\beta, \quad c_{\omega}(f) := 2^{|\omega|+2} \int f h_{\omega} \, d\beta,
\]

The \( n \)-th approximation of \( f \in L^2(\beta) \) is defined in any of the following equivalent ways:

- the projection of \( f \) on the subspace \( S_n \) along its orthogonal complement;
- the sum of the truncated Haar series:

\[
A_n f := c(f) + \sum_{|\omega| < n} c_{\omega}(f) h_{\omega};
\]

- the function obtained by averaging \( f \) on cylinders of level \( n \):

\[
A_n f = \sum_{|\omega| = n} \left( 2^n \int_{[\omega]} f \, d\beta \right) \chi_\omega.
\]

As a consequence of the last characterization we have, in terms of notation (2.1),

\[
\|f - A_n f\|_\infty \leq \text{var}_n(f),
\]

which converges to 0 as \( n \to \infty \) if \( f \) is continuous. So we have proved that the Haar series of a continuous function converges uniformly to it.

Given a sequence of coefficients, when do they form the Haar coefficients of a continuous function? Notice that, as an immediate consequence of definition (2.2),

\[
|c_{\omega}(f)| \leq \text{var}_{|\omega|}(f);
\]

therefore a necessary condition for the continuity of a function is that its Haar coefficients tend to 0. The next proposition gives a sufficient condition:
Proposition 2.1. Given a family of numbers \((c_\omega)_{\omega \in \Omega^*}\), denote \(\bar{c}_n := \sup_{|\omega|=n} |c_\omega|\). If \(\sum \bar{c}_n < \infty\) then the Haar series \(f = c + \sum_{\omega \in \Omega^*} c_\omega h_\omega\) is uniformly convergent and defines a continuous function \(f\) such that
\[
\|f\|_\infty \leq |c| + \frac{1}{2} \sum_n \bar{c}_n ,
\]
(2.6)
and
\[
\var_n(f) \leq \sum_{k \geq n} \bar{c}_k \quad \forall n \in \mathbb{N}.
\]
(2.7)
Proof. Notice that if \(\omega\) and \(\omega'\) are two different words of same length then the corresponding Haar functions \(h_\omega, h_{\omega'}\) have disjoint support. Using this fact, the convergence claim and estimate (2.6) follow. Let us proceed to prove (2.7). Fix distinct points \(x, y \in \Omega\) and estimate
\[
|f(x) - f(y)| \leq \sum_{k=0}^{\infty} \bar{c}_k \sum_{|\omega|=k} |h_\omega(x) - h_\omega(y)|.
\]
Let \(n := x \uparrow y\) (the position of first disagreement between \(x\) and \(y\)). Notice that:
\begin{itemize}
  \item if \(k < n\) then all terms in the inner sum are 0;
  \item if \(k = n\) then the inner sum contains exactly one nonzero term, namely the one corresponding to \(\omega = x_0 \cdots x_{n-1} = y_0 \cdots y_{n-1}\) (which by convention is the empty word when \(n = 0\)), whose value is 1;
  \item if \(k > n\) then the inner sum contains exactly two nonzero terms, namely the ones corresponding to \(\omega = x_0 \cdots x_{n-1}\) and \(\omega = y_0 \cdots y_{n-1}\), both of which have value 1/2.
\end{itemize}
Therefore we obtain that \(|f(x) - f(y)| \leq \sum_{k \geq n} \bar{c}_k\), completing the proof of (2.7). \(\Box\)

Now let us specialize to the Banach spaces \(C^a(\Omega)\) considered on §1.2. Fix a sequence \(a = (a_n) \searrow 0\). Notice that the least Lipchitz constant \(\text{Lip}_a(\cdot)\) with respect to the metric \(d_a\) can be expressed in terms of the variations (2.1) as
\[
\text{Lip}_a(f) = \sup_{n \in \mathbb{N}} \frac{\var_n(f)}{a_n}.
\]
(2.8)
By (2.5), the Haar coefficients of any \(f \in C^a(\Omega)\) satisfy the bounds:
\[
|c_\omega(f)| \leq a_n \text{Lip}_a(f) \quad \text{for all } \omega \in \Omega^* \text{ with } |\omega| = n.
\]
(2.9)
Conversely, if the coefficients \((c_\omega)\) in Proposition 2.1 satisfy \(\sum_{k \geq n} \bar{c}_k = O(a_n)\) then they are the Haar coefficients of a function \(f \in C^a(\Omega)\).

Given a gauge \(b = (b_\omega)_{\omega \in \Omega^*}\) (i.e., a family of positive numbers indexed by words) such that
\[
\sum \bar{b}_n < \infty , \quad \text{where } \bar{b}_n := \sup_{|\omega|=n} b_\omega ,
\]
(2.10)
the Hilbert brick with gauge \(b\) is:
\[
\mathcal{H}_b := \left\{ \sum_{\omega \in \Omega^*} c_\omega h_\omega : |c_\omega| \leq b_\omega \right\}.
\]
(2.11)
Then \(\mathcal{H}_b\) is a well-defined set of continuous functions, as an immediate application of Proposition 2.1. By the observations above, if \(\sum_{k \geq n} \bar{b}_k = O(a_n)\) then \(\mathcal{H}_b \subset C^a(\Omega)\). We have the following compactness criterion:
Proposition 2.2. If \( \sum_{k \geq n} b_k = o(a_n) \) then the Hilbert brick \( \mathcal{H}_b \) is a compact subset of the Banach space \( C^a(\Omega) \).

Proof. Assume that the tail \( \sum_{k \geq n} b_k = o(a_n) \). Consider an arbitrary sequence of functions \( f^{(i)} = \sum_{\omega \in \Omega} c^{(i)}(\omega) + \omega \in \mathcal{H}_b \). Using a diagonal argument, we can pass to a subsequence such that for every word \( \omega \), the limit \( c_\omega := \lim_{i \to \infty} c^{(i)}(\omega) \) exists. Then \( f := \sum_\omega c_\omega + \omega \in \mathcal{H}_b \). Let us show that \( f^{(i)} \) converges to \( f \) in norm. Fix \( \varepsilon > 0 \). Choose \( n_0 \in \mathbb{N} \) such that \( n \geq n_0 \) implies \( \sum_{k \geq n} b_k \leq \varepsilon a_n \), and then choose \( i_0 \in \mathbb{N} \) such that

\[
\left\{ i \geq i_0 \mid |\omega| < n_0 \right\} \Rightarrow \left| c^{(i)}(\omega) - c_\omega \right| < \frac{\varepsilon a(n_0)}{2^n - 1}.
\]

A calculation shows that for each \( n \),

\[
\sum_{k \geq n} \sup_{|\omega| = k} \left| c^{(i)}(\omega) - c_\omega \right| < 2\varepsilon a_n.
\]

Using (2.6) and (2.7), we obtain \( \|f^{(i)} - f\|_\infty < \varepsilon a_0 \) and \( \text{Lip}_a(f^{(i)}(\omega) - f) < 2\varepsilon \) respectively. This proves the desired convergence and establishes compactness. \( \square \)

Proof of Lemma 1.1. If \( a = (a_n) \) is an evanescent sequence then \( \sum_{k \geq n} a_k = O(a_n) \). So if \( b = (b_\omega) \) is an admissible gauge with respect to \( a \) then \( \sum_{k \geq n} b_k = o(a_n) \). Applying Proposition 2.2 we conclude that \( \mathcal{H}_b \) is a compact subset of \( C^a(\Omega) \). \( \square \)

As the reader will have noticed, the second admissibility condition in (1.4) is not needed for the validity of Lemma 1.1. It will be fundamental for the proof of Theorem B, however.

3. Finite-dimensional ergodic optimization

3.1. Rotation sets. As in §1.2, consider a general continuous transformation \( T \) of a compact metric space \( X \). The set \( \mathcal{M}_T \) of invariant probabilities is a compact convex subset of the space \( \mathcal{M}_+^\pm(X) \) of all finite signed Borel measures, endowed with the usual weak topology.

Let \( F \subset C^0(\Omega) \) be a finite-dimensional space of continuous functions. Given a basis \( \{f_1, \ldots, f_n\} \) of \( F \), consider the continuous linear map

\begin{equation}
\pi: \mathcal{M}_\pm(X) \to \mathbb{R}^n \text{ defined by } \pi(\mu) = (\langle f_1, \mu \rangle, \ldots, \langle f_n, \mu \rangle).
\end{equation}

A change of basis produces a conjugate map \( \pi \). The image of \( \mathcal{M}_T \) under \( \pi \) is a convex compact subset of \( \mathbb{R}^n \), called a generalized rotation set, or simply a rotation set. The terminology is not especially good, but since it seems to be established in the literature [Zi, Je2, KW] we will conform with it. The prime example is when \( (f_1, \ldots, f_n) \) is the displacement vector of a map of the \( n \)-torus homotopic to the identity. An example which is much more related to the theme of this paper is the Bousch’s “fish” [Bo1], where the dynamics is the doubling map \( T(\theta) = 2\theta \) on the circle \( \mathbb{R}/2\pi \mathbb{Z} \) and \( F \) is spanned by the functions \( \cos \theta \) and \( \sin \theta \), so the projection \( \pi \) consists on computing two Fourier coefficients.

It is shown in [KW] that arbitrary compact convex sets in \( \mathbb{R}^n \) appear as rotation sets of shift transformations. On the other hand, a situation where rotation sets are relatively simple was uncovered by Ziemian [Zi]: if \( T \) is a transitive subshift of finite type and \( F \) is composed by step functions constant on cylinders of level 2 then the associated rotation set is a polytope.
Let us say that a rotation set \( R = \pi(M_T) \) has the *injectivity property* if for each extreme point \( \rho \) of \( R \) there exists an unique \( \mu \in M_T \) such that \( \pi(\mu) = \rho \). Examples are given by Ziemian’s polytopes (where the preimages of extreme points are certain periodic measures) and Bousch’s fish (where the preimages are the Sturmian measures).

Let us discuss the relevance of rotation sets in ergodic optimization. The map \( \pi \) has the following obvious projection property:

\[
\pi(\mu) = \pi(\tilde{\mu}) \Rightarrow \langle f, \mu \rangle = \langle f, \tilde{\mu} \rangle \quad \text{for all } f \in F.
\]

Using this fact, we can define a bilinear map \( \langle \cdot, \cdot \rangle : F \times V \to \mathbb{R} \), where \( V \) is the “rotation space” \( \pi(M_\pm(X)) \), by

\[
\langle f, v \rangle := \langle f, \mu \rangle \quad \text{where } \mu \text{ is such that } \pi(\mu) = v.
\]

In particular, the ergodic supremum (1.1) of a function \( f \in F \) can be expressed as:

\[
\text{erg sup } f = \sup_{v \in R} \langle f, v \rangle.
\]

Therefore \( R \) and \( \langle \cdot, \cdot \rangle \) encode basically all the information necessary to do ergodic optimization for functions in \( F \), which becomes a finite-dimensional problem. In the particular case that \( R \) is a polytope, the computation of \( \text{erg sup } f \) becomes a linear programming problem. If the supremum in (3.4) is attained at a unique point, which is often the case, and \( R \) happens to have the injectivity property, then the original ergodic optimization problem (1.1) has a unique solution, i.e., the function \( f \in F \) has a unique maximizing measure.

### 3.2. A family of rotation sets for the shift

Let us focus on the shift transformation \( \sigma \) on \( \Omega = \{0,1\}^\mathbb{N} \). For each \( n \in \mathbb{N} \), consider the vector space \( S_n \) of step functions of level \( n \), which has dimension \( 2^n \). The characteristic functions of cylinders of level \( n \), ordered lexicographically, form a basis of \( S_n \). Let \( \pi_n : M_\pm(\Omega) \to \mathbb{R}^{2^n} \) be the corresponding projection map, as defined in the previous subsection, that is,

\[
\pi_0(\mu) := (\mu(\Omega)), \quad \pi_1(\mu) := (\mu([0]), \mu([1])), \quad \pi_2(\mu) := (\mu([00]), \mu([01]), \mu([10]), \mu([11])), \quad \text{etc.}
\]

If \( \mu \) is a probability measure then \( \pi_n(\mu) \) is a subset of the unit simplex \( \Delta^{2^n-1} \). We define a rotation set

\[
R_n := \pi_n(M_n).
\]

**Proposition 3.1.** The rotation sets \( R_n \) are polytopes, and satisfy the injectivity property.

The proposition (or at least its first part) can be deduced from the aforementioned result of Ziemian [Zi]. It is also related to work by Fukš [Fu]. Since we will need finer information about the sets \( R_n \) we will provide a self-contained proof. As in [Zi], we will use some elementary graph theory.
3.3. Preliminaries on convexity and graph theory. Let us review some terminology of convex analysis. Let $C$ be a closed convex subset of Euclidean space. A subset $\tilde{C} \subset C$ is called a face of $C$ if any segment in $C$ whose relative interior intersects $\tilde{C}$ is actually contained in $\tilde{C}$. The proper faces of $C$ are those different from $\emptyset$ and $C$. If a singleton $\{v\}$ is a face of $C$ then $v$ is called an extreme point of $C$. If $C$ has finitely many extreme points then they are called vertices, and $C$ is called a (convex) polytope. The faces of a polytope are also polytopes; among them, 1-dimensional faces are called edges, and proper faces of maximal dimension are called facets.

A closed convex set is called a (convex) polyhedron an intersection of finitely many affine-half spaces. A fundamental theorem [Sc, Corollary 7.1c] states that bounded polyhedra are polytopes, and vice-versa.

Next let us fix some graph theory terminology. Let $G$ be a directed graph, composed of finitely many nodes and arcs. We will always assume that no two arcs have the same source and target nodes.

A walk in $G$ is a finite sequence of the form $n_0a_1n_1 \cdots n_{\ell-1}a_\ell n_\ell$ where each $a_i$ is an arc from the node $n_{i-1}$ to the node $n_i$. The nonnegative integer $\ell$ is called the length of the walk. A closed walk is an equivalence class of walks with equal ending and starting nodes, where $n_0a_1n_1 \cdots n_{\ell-1}a_\ell n_0$ is considered equivalent to $n_1a_1n_2 \cdots n_{\ell-1}a_\ell n_1$. A closed walk with no repeated nodes (nor arcs) is called a cycle.

A circulation on $G$ is a function that attributes a weight to each arc of $G$ in such a way that:

- weights are nonnegative;
- the total weight is 1;
- the total weight entering any node is equal to the total weight leaving that node (Kirchhoff’s current law).

The support of a circulation is the subgraph where all weights are positive.

The set of functions on the arcs of $G$ can be identified with some $\mathbb{R}^4$. One such function is a circulation if it satisfies certain linear equalities and inequalities; therefore the set of circulations is a polyhedron. Moreover, it is obviously bounded, so it is a polytope. It is called the circulation polytope of $G$.

**Proposition 3.2.** Let $G$ be directed graph, and let $C$ be its circulation polytope. Then the faces of $C$ are the circulation polytopes of the subgraphs of $G$, and the vertices of $C$ are the circulations supported on the cycles of $G$.

**Proof.** If a polytope is expressed in terms of linear equalities and inequalities, then its faces can be obtained by setting some of the inequalities to equalities: see [Sc, §8.2]. The inequalities in our case are “weights are nonnegative”. Forcing some of the arcs to have zero weights gives rise to circulations in a subgraph. This proves the first claim.

Given a subgraph $G'$, there exists a circulation with support contained in $G'$ if and only if $G'$ contains a cycle. Moreover, each cycle supports a unique circulation. So the second claim follows. \hfill $\square$

3.4. Graphical itineraries. For each integer $n \geq 1$, the de Bruijn graph $G_n$ is the directed graph defined as follows:

\[4\text{The terms vertices and edges are reserved to polytopes.}\]
• the nodes of $G_n$ are exactly the words of length $n - 1$ in the alphabet $\{0, 1\}$;
• the arcs of $G_n$ are exactly the words of length $n$ in the alphabet $\{0, 1\}$;
• for each word $\omega$ of length $n$, the source node and the target node of the arc $\omega$ are respectively its initial and final subwords of length $n - 1$.

These graphs (together with their analogues for larger alphabets) were introduced independently by de Bruijn [dB] and Good [Go], and are relevant to many problems in combinatorics and applied mathematics: see [Mau] and references therein. The first four de Bruijn graphs are pictured in Fig. 1.

![Figure 1. The de Bruijn graphs $G_n$ for $1 \leq n \leq 4$.](image)

Any orbit of the one-sided shift $\sigma : \Omega \to \Omega$ induces in the obvious way a semi-infinite walk on the de Bruijn graph $G_n$, called the $n$-itinerary of the orbit. If the orbit is periodic then we obtain a closed walk of length equal to the period. If $\mu$ is the measure supported on the periodic orbit, we denote the induced closed walk on $G_n$ by $W_n(\mu)$. The following sets of measures will play an important role:

\[
C_n := \{ \mu \in M_\sigma \text{ periodic} ; W_n(\mu) \text{ is a cycle in } G_n \}.
\]

There is an obvious bijection between $C_n$ and the set of cycles in $G_n$. The initial terms of the sequence $(C_n)$ are:

$C_1 = \{ \overline{0}, \overline{1} \} \subset C_2 = \{ \overline{0}, \overline{1}, \overline{00} \} \subset C_3 = \{ \overline{0}, \overline{1}, \overline{00}, \overline{01}, \overline{01}, \overline{10}, \overline{11} \},$

where $\overline{\omega}$, for example, denotes the periodic measure supported on the orbit of the point $(0, 1, 0, 1, \ldots) \in \Omega$.

**Remark 3.3.** The nested union $\bigcup_{n \geq 1} C_n$ equals the set of periodic measures in $M_\sigma$. Following Maurer [Mau], we define the recursive complexity of a periodic measure $\mu$ as the least $n$ such that $\mu \in C_n$. Recursive complexity is not monotonic with respect to the period: for example, $000111$ and $01011$ have recursive complexities 4 and 5, respectively.

**Remark 3.4.** It seems to be a difficult task to count precisely the number of elements of $C_n$ (or equivalently the number of cycles in $G_n$): see [Mau] for related results. Let us mention a celebrated result by de Bruijn [dB]: for $n \geq 2$, there are exactly

Note that Maurer’s definition is slightly different since his $G_n$ is our $G_{n+1}$. 

\[2^{2n-2} - n + 1\] cycles in \(G_n\) of maximal length \(2^{n-1}\); these are the Hamiltonian cycles of \(G_n\), i.e., the cycles that visit every node.\(^6\)

3.5. **The structure of the rotation sets** \(R_n\). Let us use the above material to describe precisely the rotation sets \(R_n = \pi_n(M_\sigma)\) introduced in §3.2. The following result is a finer version of Proposition 3.1:

**Proposition 3.5.** For all \(n \geq 1\), the rotation set \(R_n\) is isomorphic to the circulation polytope of the de Bruijn graph \(G_n\). Moreover, \(R_n\) has the injectivity property: the restriction of \(\pi_n\) to the set \(C_n\) is a bijection onto the vertices of \(R_n\).

**Proof.** Suppose \(\rho \in R_n\), that is, \(x = \pi_n(\mu)\) for some \(\mu \in M_\sigma\). The vector \(x\) gives a weight to each cylinders of level \(n\) and thus to each arc of \(G_n\). We claim that this weight function is a circulation. Weights are obviously nonnegative and sum to 1, so we are left to verify Kirchhoff’s current law. Given any node \(\omega\), that is, a word of length \(n-1\), we have the following disjoint unions:

\[
[\omega] = [\omega 0] \cup [\omega 1], \quad \sigma^{-1}([\omega]) = [0\omega] \cup [1\omega].
\]

So \(\mu([\omega])\) is the total weight leaving the node \(\omega\), while \(\mu(\sigma^{-1}([\omega]))\) is the total weight entering it. They are equal by invariance of the measure. So Kirchhoff’s current law is satisfied, and we have shown that \(x\) induces a circulation on \(G_n\).

Conversely, let us check that every circulation on \(G_n\) is of the form above. This is obvious if the circulation is supported on a cycle, and since a general circulation is a convex combination of circulations supported on cycles (by Proposition 3.2), the claim follows.

The injectivity property is clear, and the proposition is proved. \(\square\)

Proposition 3.5 permits us to deduce other properties of rotation set \(R_n\). For example, the dimension of \(R_n\) is \(2^{n-1}\), since each node of \(G_n\) contributes with one degree of freedom. The face structure of \(R_n\) can be studied using Proposition 3.2: an example is shown in Fig. 2. Since \(G_n\) has \(2^n\) arcs, it follows from Propositions 3.5 and 3.2 that \(R_n\) has no more than \(2^{2n}\) faces.

![Figure 2. The polytope \(R_3\) has 6 facets: 4 tetrahedra and 2 pyramids (pictured). The total number of faces is 1 + 6 + 13 + 13 + 6 + 1 = 40.](image)

We close this section mentioning other facts that, despite not being needed for the rest of the paper, have independent interest.

\(^6\)Incidentally, these cycles are exactly the lifts of the Eulerian closed walks on \(G_{n-1}\).
Remark 3.6. In view of the characterization (3.4), the computation of the ergodic supremum of a function in $S_n$ becomes equivalent to a maximum cycle mean problem for the de Bruijn graph $G_n$. These problems have been studied in the combinatorics optimization literature, and efficient algorithms are known: see [Ka, DIG].

Remark 3.7. The polytopes $R_n$ are finite-dimensional approximations the set $M_\sigma$. Let us make this statement precise. Consider the sequence of maps $R_0 \leftarrow R_1 \leftarrow R_2 \leftarrow \cdots$ where each arrow is defined by $(a, b, c, d, \ldots) \mapsto (a + b, c + d, \ldots)$ Then the inverse limit of this sequence is (topologically and affinely) isomorphic to $M_\sigma$. Moreover, the natural projections $M_\sigma \rightarrow R_n$ are the restrictions of the maps $\pi_n$ defined previously.

Remark 3.8. Using the appropriate graphs, it is straightforward to adapt the results above to general subshifts of finite type and in particular reobtain the results of [Zi].

4. A gap criterion for the locking property

4.1. Statement of the criterion. Consider a step function $g \in S_n$ of level $n \geq 1$. By the facts discussed in the previous section, in terms of notation (3.3) we have

$$\text{erg sup } g = \langle g, v \rangle$$

for some vertex $v$ of the polytope $R_n$. Define the $n$-gap of $g$ as (see Fig. 3):

$$\text{gap}_n(g) := \langle g, v \rangle - \max_{w \text{ vertex of } R_n \wedge w \neq v} \langle g, w \rangle.$$ 

Note that in the expression above it is sufficient to consider the vertices $w$ such that $[v, w]$ is an edge of the polytope $R_n$.

Let $\mu$ be the measure in $M_\sigma$ such that $v = \pi_n(\mu)$. Then $\text{erg sup } g = \langle g, \mu \rangle$, i.e. $\mu$ is a maximizing measure for $g$; moreover, $\mu \in C_n$, and in particular it is a periodic measure. The $n$-gap can be written as:

$$\text{gap}_n(g) = \langle g, \mu \rangle - \max_{\nu \in C_n \setminus \mu} \langle g, \nu \rangle.$$

Note that $\text{gap}_n(g)$ is positive if $g$ has no other maximizing measure in $M_\sigma$, and is zero otherwise. In the former case, every $\tilde{g} \in S_n$ sufficiently close to $g$ also has an unique maximizing measure, which is $\mu$.

Figure 3. Gap of a function $g \in S_n$. The lines are the level sets of $\langle g, \cdot \rangle$.

Recall that $A_n f$ denotes the truncated Haar series (2.3) of a function $f$, and so it belongs to $S_n$.

The following lemma plays a pivotal role in this paper:

---

7It is interesting to note that Karp’s algorithm relies on a graph-theoretical version of Mañé lemma. In this regard, a function $f$ on the set of arcs is called a coboundary if there exists a function $h$ on its set of nodes such that $f(a) = h(n_1) - h(n_0)$ whenever $a$ is an arc from node $n_0$ to node $n_1$. 


Lemma 4.1 (Gap criterion). Let \( f \in C^0(\Omega) \) and \( n \in \mathbb{N} \). Assume that the following gap condition is satisfied:

\[
\text{gap}_n(A_n f) > \sum_{k=n}^{\infty} (k - n + 1) \max_{|\omega|=k} |c_\omega(f)|.
\]

Then the maximizing measure for \( A_n f \) (which is unique and periodic) is also the unique maximizing measure for \( f \).

In informal terms, if the tail of the Haar series is small compared to the gap of its initial part, then it does not influence the maximizing measure; in particular property \( P \) is satisfied. The lemma actually permits us to obtain the stronger locking property:

Corollary 4.2 (Gap criterion version 2). Consider a sequence \( a = (a_n) \searrow 0 \) such that:

\[
\sum_{n=0}^{\infty} na_n < \infty.
\]

If a function \( f \in C^a(\Omega) \) satisfies the gap condition (4.1) for some \( n \in \mathbb{N} \) then \( f \in L_a \).

Proof. In view of estimate (2.9), assumption (4.2) implies that, for every \( n \in \mathbb{N} \), the RHS of (4.1) converges for any function \( f \in C^a(\Omega) \), and moreover it depends continuously on \( f \). The LHS is also continuous, so the functions \( f \) in the space \( C^a(\Omega) \) that satisfy the gap condition (4.1) form an open subset \( \mathcal{U}_n \). By Lemma 4.1, each \( f \in \mathcal{U}_n \) has a unique maximizing measure, which coincides with the unique maximizing measure for \( A_n f \) and so is periodic. As observed before, maximizing measures of functions in \( S_n \) are locally constant whenever the gap is positive. It follows that \( \mathcal{U}_n \subset L_a \), as we wanted to show. \( \square \)

The proof of Lemma 4.1 will occupy the rest of this section, and requires a few other lemmas.

4.2. Basins of repellency. Given \( \mu \in M_\sigma \) and \( k \in \mathbb{N} \), let \( B_{\mu,k} \subset \Omega \) be the union of all cylinders of level \( k \) that intersect the support of \( \mu \).

Suppose \( \mu \) is periodic, and let \( k_0 \) be the least integer such that \( \mu \in C_{k_0} \). Then, for all \( k \geq k_0 \), the set \( B_{\mu,k} \) is a basin of repellency for the periodic orbit corresponding to \( \mu \), in the sense that this is the only orbit that does not leave \( B_{\mu,k} \). (This fact can be verified by considering itineraries with respect to the graph \( G_k \).)

In this subsection we prove two technical lemmas concerning basins of repellency. Let \( c \) denote the complement operation, that is, \( B^c := \Omega \setminus B \).

Lemma 4.3. Given measures \( \mu \in C_n \) and \( \xi \in C_k \) with \( n \leq k \), we have

\[
\xi(B^c_{\mu,n}) \leq \xi(B^c_{\mu,k}) \leq (k - n + 1)\xi(B^c_{\mu,n}).
\]

For example, if \( \mu = \overline{01} \in C_2 \) and \( \xi = \overline{001} \in C_3 \) then

\[
\xi(B^c_{\mu,2}) = \xi(([01] \cup [10])^c) = 1/2 \quad \text{and} \quad \xi(B^c_{\mu,3}) = \xi(([010] \cup [101])^c) = 1,
\]

so the second inequality in Lemma 4.3 becomes an equality.
Proof of Lemma 4.3. Since $B_{\mu,k} \subset B_{\mu,n}$, the first inequality is trivial.

Recall from §3.4 that $W_k(\xi)$ denotes the closed walk induced by $\xi$ on $G_k$, which is actually a cycle since $\xi \in C_k$. Since $k \geq n$ and $\mu \in C_n$, the closed walks $W_k(\mu)$ and $W_n(\mu)$ are cycles on $G_k$ and $G_n$, respectively.

Given an arc $a$ of $G_k$, represent it as a word $b_1 \ldots b_k$ in zeros and ones, and let $w(a)$ be the walk in $G_n$ that traverses the arcs $b_1 \ldots b_n, b_2 \ldots b_{n+1}, \ldots, b_k \ldots b_{k-n+1} \ldots b_k$, in this order. Then $a \mapsto w(a)$ is a bijection between the arcs of $G_k$ and the walks of length $k - n + 1$ in $G_n$.

Choose randomly an arc $a$ in the cycle $W_k(\xi)$, considering all possibilities equiprobable. Let $E$ be the event “$a$ does not belong to the cycle $W_k(\mu)$”, and let $N$ be the number of arcs of the walk $w(a)$ outside the cycle $W_n(\mu)$. Then the probability of $E$ and the expected value of $N$ are respectively:

$$\mathbb{P}(E) = \xi(B_{\mu,k}^c) \quad \text{and} \quad \mathbb{E}(N) = (k - n + 1)\xi(B_{\mu,n}^c).$$

Note that $a$ is an arc of the cycle $W_k(\mu)$ if and only if the walk $w(a)$ is wholly contained in the cycle $W_n(\mu)$. In other words, if $E$ occurs then $N \geq 1$. Therefore $\mathbb{P}(E) \leq \mathbb{P}[N \leq 1] \leq \mathbb{E}(N)$, which gives the second inequality.

Lemma 4.4. Given measures $\mu \in C_n$ and $\xi \in C_k$ with $n \leq k$ and $\per(\mu) \leq \per(\xi)$, and a function $g \in \mathcal{S}_k$, we have

$$\langle g, \xi \rangle - \langle g, \mu \rangle \leq 2\|g\|_\infty \xi(B_{\mu,k}^c).$$

Notice that $n \leq k$ does not imply that $\per(\mu) \leq \per(\xi)$ in the statement above; an example appear in Remark 3.3.

Proof of Lemma 4.4. Write $m := \per(\mu)$ and $\ell := \per(\xi)$. The step function $g \in \mathcal{S}_k$ induces a function on the arcs of $G_k$, which by simplicity we denote by the same symbol. Let $A$ and $B$ be the sets of arcs in the cycles $W_k(\mu)$ and $W_k(\xi)$, respectively. So $\#A = m$ and $\#B = \ell$. Let $i := \#(A \cap B)$. Then:

$$\langle g, \xi \rangle - \langle g, \mu \rangle = \frac{1}{\ell} \sum_{a \in B} g(a) - \frac{1}{m} \sum_{a \in A} g(a)$$

$$= \frac{1}{\ell} \sum_{a \in B \setminus A} g(a) - \frac{1}{m} \sum_{a \in A \setminus B} g(a) + \left( \frac{1}{\ell} - \frac{1}{m} \right) \sum_{a \in A \cap B} g(a)$$

$$\leq \left\{ \ell - i + \frac{m - i}{m} + \left| \frac{i}{m} - \frac{1}{\ell} \right| \right\} \|g\|_\infty.$$

By assumption, $m \leq \ell$, and so the quantity between curly brackets is $2(\ell - i)/\ell$, which equals $2\xi(B_{\mu,k}^c)$. This proves the lemma.

4.3. Proof of the gap criterion. Fix $f \in C^0(\Omega)$ satisfying the gap condition (4.1) for some $n \in \mathbb{N}$. Let $\mu \in C_n$ be the maximizing measure for $A_n f$, and fix an arbitrary $\nu \in \mathcal{M}_\sigma$. We need to prove that $\langle f, \mu \rangle \geq \langle f, \nu \rangle$, with equality only if $\nu = \mu$.

For each integer $k \geq n$, we will deal with the basin of repellency $B_{k,\mu}$ introduced in §4.2, which we will denote by $B_k$ for simplicity.

For each $k \geq n$, consider convex combinations of the form

$$\nu_k = \sum_{\xi \in C_k} p_k(\xi) \xi$$

(4.3)
(so the numbers $p_k(\xi)$ are nonnegative and their sum over $\xi$ is 1) such that

$$\pi_k(\nu) = \pi_k(\nu).$$

Such combinations always exist, but uniqueness may fail. We choose and fix a combination such that the following quantity:

$$u_k := \sum_{\xi \in C_k \per(\xi) \geq \per(\mu)} p_k(\xi) \xi(B_n)$$

is minimal among all possible decompositions; its existence follows from a compactness argument.

For example, if $\nu = \beta$ (the unbiased Bernoulli measure) and $\per(\mu) = 4$ then $u_3 = 0$, because we can take $\nu_3$ as a convex combination of measures in $C_3$ of periods less than 4, namely $\nu_3 = \frac{1}{8} \beta + \frac{1}{8} \beta + \frac{3}{8} \beta + \frac{3}{8} \beta$.

**Lemma 4.5.** The sequence $(u_k)_{k \geq n}$ is monotone nondecreasing.

**Proof.** Fix $k \geq n$. For each $\eta \in C_{k+1}$, write the circulation $\pi_k(\eta)$ as a convex combination of vertices of $R_k$:

$$\pi_k(\eta) = \sum_{\xi \in C_k} q_\eta(\xi) \pi_k(\xi).$$

Notice that

$$\per(\xi) > \per(\eta) \Rightarrow q_\eta(\xi) = 0;$$

indeed for each $\xi \in C_k$, the closed walk $W_k(\xi)$ visits exactly $\per(\xi)$ nodes of $G_k$, while the closed walk $W_k(\eta)$ visits at most $\per(\eta)$ nodes.

On the other hand,

$$\pi_k(\nu) = \pi_k(\pi_{k+1}(\nu)) = \pi_k \left( \sum_{\eta \in C_{k+1}} p_{k+1}(\eta) \eta \right) = \sum_{\eta \in C_{k+1}} p_{k+1}(\eta) \pi_k(\eta).$$

Substituting the expressions (4.6) and manipulating, we obtain:

$$\pi_k(\nu) = \pi_k(\bar{\nu}), \quad \text{where} \quad \bar{\nu} := \sum_{\xi \in C_k} \left( \sum_{\eta \in C_{k+1}} p_{k+1}(\eta) q_\eta(\xi) \right) \pi_k(\xi),$$
which should be compared with (4.3) and (4.4). Therefore:

\[
u_k \leq \sum_{\xi \in C_k} \left( \sum_{\eta \in C_{k+1}} p_{k+1}(\eta) q_\eta(\xi) \right) \xi(B_n) \quad \text{(by minimality of (4.5))}
\]

\[
\leq \sum_{\eta \in C_{k+1}} \sum_{\xi \in C_k} p_{k+1}(\eta) q_\eta(\xi) \xi(B_n) \quad \text{(by (4.7))}
\]

\[
= \sum_{\eta \in C_{k+1}} p_{k+1}(\eta) \left( \sum_{\xi \in C_k} q_\eta(\xi) \right) \xi(B_n)
\]

\[
= \sum_{\eta \in C_{k+1}} p_{k+1}(\eta) \eta(B_n) \quad \text{(using (4.6), \(\chi_{B_n} \in S_n \subset S_k\), (3.2))}
\]

\[
= u_{k+1},
\]

thus proving monotonicity. \(\square\)

We are ready to prove the gap criterion:

**Proof of Lemma 4.1.** Let \(f, n, \mu, \nu\) be as explained above. Write

\[
f = A_n f + \sum_{k=n}^{\infty} g_k, \quad \text{where} \quad g_k := \sum_{|\omega|=k} c_\omega(f) h_\omega.
\]

Then, using the projection property (3.2) together with (4.3) and (4.4), we have:

\[
(f, \mu - \nu) = \langle A_n f, \mu - \nu \rangle - \sum_{k=n}^{\infty} \langle g_k, \nu - \mu \rangle
\]

\[
= \langle A_n f, \mu - \nu_n \rangle - \sum_{k=n}^{\infty} \langle g_k, \nu_k - \mu \rangle
\]

\[
= \sum_{\xi \in C_n} p_n(\xi) \langle A_n f, \mu - \xi \rangle - \sum_{k=n}^{\infty} \sum_{\xi \in C_k} p_k(\xi) \langle g_k, \xi - \mu \rangle .
\]

(4.8)

We will estimate the terms \(\textcircled{1}\) and \(\textcircled{2}\) from below and from above, respectively.

On the one hand, by definition of the gap, for all \(\xi \in C_n\) different from \(\mu\) we have \(\langle A_n f, \mu - \xi \rangle \geq \text{gap}_n(A_n f)\), and therefore

\[
\textcircled{1} \geq (1 - p_n(\mu)) \text{gap}_n(A_n f) \geq (1 - u_n) \text{gap}_n(A_n f).
\]

On the other hand, if \(\xi \in C_k\) has per(\(\xi\)) \geq per(\(\mu\)) then by Lemmas 4.4 and 4.3,

\[
\langle g_k, \xi - \mu \rangle \leq 2\|g_k\|_\infty \xi(B_k^c) \leq 2(k - n + 1)\|g_k\|_\infty \xi(B_n^c).
\]

In the case that per(\(\xi\)) < per(\(\mu\)) we use the trivial estimate

\[
\langle g_k, \xi - \mu \rangle \leq 2\|g_k\|_\infty \leq 2(k - n + 1)\|g_k\|_\infty.
\]
So we have:

\[ \tau_k \leq 2(k - n + 1)\|g_k\|_\infty \left\{ \sum_{\xi \in C_k \atop \text{per}(\xi) < \text{per}(\mu)} p_k(\xi) + \sum_{\xi \in C_k \atop \text{per}(\xi) \geq \text{per}(\mu)} p_k(\xi)\|B^*_k\| \right\} \]

\[ = 2(k - n + 1)\|g_k\|_\infty (1 - u_k) \]

\[ \leq 2(k - n + 1)\|g_k\|_\infty (1 - u_n), \]

where in the last step we used Lemma 4.5. Substituting these estimates into (4.8), we obtain

\[ \langle f, \mu - \nu \rangle \geq \delta (1 - u_n), \quad \text{where} \quad \delta := \text{gap}_n(A_n f) - \sum_{k=n}^{\infty} 2(k - n + 1)\|g_k\|_\infty. \]

Notice that \( \|g_k\|_\infty = \frac{1}{2} \sup_{|\omega| = k} |c_\omega(f)| \), and so assumption (4.1) means that \( \delta > 0 \).

Moreover, by definition \( u_n \leq \nu_n(B_n) \), so

\[ \langle f, \mu - \nu \rangle \geq \delta \nu_n(B_n). \]

In particular \( \langle f, \mu - \nu \rangle \geq 0 \). Assume that equality holds; then \( \nu_n(B_n) = 1 \). Since \( B_n \) is a basin of repellency for \( \mu \), we have \( \nu_n = \mu \). So \( \tau_n(\nu) = \tau_n(\mu) \) and by the injectivity property of the rotation set \( R_n \) we conclude that \( \nu = \mu \). This proves that \( \mu \) is the unique maximizing measure for \( f \), as claimed.

5. PROOF OF PREVALENCE

In this section we prove Theorem B which, as explained before, yields Theorem A as a corollary.

A box \( Q \subset \mathbb{R}^d \) is a product \( I_1 \times \cdots \times I_d \) of compact intervals of positive lengths, the least of which is called the thickness of the box and is denoted by \( \tau(Q) \). Lebesgue measure on \( \mathbb{R}^d \) is denoted by \( \text{Leb}_d \). We will estimate the measure of slices of a box, as in Fig. 4.

![Figure 4. Slicing a box.](image)

If \( L \) is a linear functional on \( \mathbb{R}^d \), say \( L(x) = \sum_{i=1}^d \lambda_i x_i \), then its 1-norm is \( \|L\|_1 := \sum_{i=1}^d |\lambda_i| \).

**Lemma 5.1.** Let \( Q \subset \mathbb{R}^d \) be a box. Let \( \varphi : \mathbb{R}^d \to \mathbb{R} \) be an affine function with linear part \( L \neq 0 \). Then, for all \( \delta > 0 \),

\[ \frac{\text{Leb}_d(Q \setminus \{ x \in Q : |\varphi(x)| \leq \delta \})}{\text{Leb}_d(Q)} \leq \frac{2\delta}{\|L\|_1 \tau(Q)}. \]

**Proof.** Without loss of generality, assume that \( Q = I_1 \times \cdots \times I_d \) has thickness \( \text{Leb}_1(I_1) \). Let \( L(x) = \sum_{i=1}^d \lambda_i x_i \). Then for each \( y \in I_2 \times \cdots \times I_d \), the section \( \{ x_1 \in I_1 : |\varphi(x_1, y)| \leq \delta \} \) has measure at most \( 2\delta/|\lambda_1| \). So, by Fubini theorem,

\[ \text{Leb}_d(Q \setminus \{ x \in Q : |\varphi(x)| \leq \delta \}) \leq \frac{2\delta}{|\lambda_1|} \text{Leb}_{d-1}(I_2 \times \cdots \times I_d) \leq \frac{2\delta}{\|L\|_1} \text{Leb}_d(Q) \frac{\text{Leb}_d(Q)}{\tau(Q)}. \]
Proof of Theorem B. Fix an evanescent sequence \( a \), and an admissible gauge \( b \) with respect to \( a \). Fix an arbitrary \( f_0 \in C^a(\Omega) \). Recall that our aim to prove that for \( P_b \)-almost every function \( g \) in the Hilbert brick \( H_b \), the function \( f = f_0 + g \) has the locking property. For that purpose we will use the gap criterion version 2 (Corollary 4.2). Note that hypothesis (4.2) is automatic from evanescence.

For each \( n \in \mathbb{N} \), let
\[
\delta_n := \sum_{k=n}^{\infty} (k-n+1) (\text{Lip}_a f_0 + b_k) \quad \text{and} \quad G_n := \{ g \in H_b : \text{gap}_n (A_n (f_0 + g)) > \delta_n \}.
\]
In view of (2.9) and the definition of the Hilbert brick \( H_b \), the number \( \delta_n \) is an upper bound for the RHS of (4.1) for all functions of the form \( f = f_0 + g \) with \( g \in H_b \). So by Corollary 4.2 we have:
\[
g \in \bigcup_n G_n \Rightarrow f_0 + g \in L_a.
\]
So in order to prove Theorem B it is sufficient to show that \( P_b(G_n) \to 1 \) as \( n \to \infty \).

Fix a positive integer \( n \). Let \( d_n := 2^n - 1 \) and let \( \tilde{\pi}_n : C^a(\Omega) \to \mathbb{R}^{d_n} \) be the map
\[
f \mapsto (c_\omega(f))_{\omega \in \Omega, |\omega| < n}
\]
that associates to a function \( f \) its Haar coefficients of level less than \( n \) (the initial coefficient \( c(f) \) being disregarded). Then \( \tilde{\pi}_n(H_b) \) equals the box
\[
Q_n := \prod_{|\omega| < n} [-b_\omega, b_\omega];
\]
moreover the push-forward of \( P_b \) under \( \tilde{\pi}_n \) is the normalized Lebesgue measure on \( Q_n \). For each measure \( \mu \in \mathcal{M} \), define a linear map \( I_\mu : \mathbb{R}^{d_n} \to \mathbb{R} \) by
\[
I_\mu((x_\omega)_{\omega}) := \sum_{\omega} \langle h_\omega, \mu \rangle x_\omega.
\]
Then \( \langle A_n(f_0 + g), \mu \rangle = c(f) + I_\mu(f) \) for every \( f \in C^a(\Omega) \).

Let \( G_n := \tilde{\pi}_n(G_n) \). Belongingness to \( G_n \) only depends on the initial Haar coefficients, so \( G_n = \tilde{\pi}_n^{-1}(G_n) \). Write \( G_n^c = H_b \setminus G_n \) and \( G_n^c = Q_n \setminus G_n \). If \( g \in G_n^c \) then there exist distinct elements \( \mu, \nu \) of \( C_n \) such that
\[
|\langle A_n(f_0 + g), \mu - \nu \rangle| \leq \delta_n.
\]
Moreover, \( \mu, \nu \) can be chosen as \textit{neighbors} in the sense that \([\pi_n(\mu), \pi_n(\nu)]\) is an edge of the polytope \( R_n \). It follows that
\[
G_n^c \subset \bigcup_{\mu \neq \nu \in C_n \text{ neighbors}} \left\{ x \in Q_n : \left| (I_\mu - I_\nu)(x^{(0)} + x) \right| \leq \delta_n \right\},
\]
where \( x^{(0)} := \tilde{\pi}_n(f_0) \). In particular, by Lemma 5.1,
\[
\mathbb{P}_b(G_n^c) = \frac{\text{Leb}_{d_n}(G_n^c)}{\text{Leb}_{d_n}(Q_n)} \leq \frac{2\delta_n}{\tau(Q_n)} \sum_{\mu \neq \nu \in C_n \text{ neighbors}} \| I_\mu - I_\nu \|^{-1}_1. \tag{5.2}
\]
Given $\mu \neq \nu$ in $C_n$, there exists a node of the graph $G_n$ that is visited by the cycle $C_n(\mu)$ but not by cycle $C_n(\nu)$. That is, there exists a word $\omega$ of length $n-1$ such that
\[ \langle h_\omega, \mu \rangle = \frac{\pm 1}{2 \text{per}(\mu)} \quad \text{and} \quad \langle h_\omega, \nu \rangle = 0. \]
We have $\text{per}(\mu) \leq 2^{n-1}$ (since the cycle $C_n(\mu)$ contains no repeated nodes), and therefore $\|I_\mu - I_\nu\|_1 \geq 2^{-n}$. On the other hand, the number of pairs of neighbors $\{\mu, \nu\}$ in $C_n$ is equal to the number of edges of the polytope $R_n$. We have seen in §3.5 that $R_n$ has no more than $2^n$ faces of all dimensions. Substituting these estimates back in (5.2) we obtain:
\begin{equation}
\mathbb{P}_b(G_n^\circ) \leq 2^{2^n + n + 1} \delta_n / \tau(Q_n).
\end{equation}
Recall the evanescence and admissibility conditions (1.2), (1.4), which can be rewritten as:
\[ a_n = O \left( 2^{-2^{n+1}} a_{n-1} \right), \quad \bar{b}_n = o(a_n), \quad a_n = 2^{O(n)} \bar{b}_n. \]
Then:
\[ \delta_n = \sum_{k=n}^{\infty} (k - n + 1)O(a_k) = O(a_n) = O \left( 2^{-2^{n+1}} a_{n-1} \right). \]
Also,
\[ \bar{b}_n = o(a_n) = O \left( 2^{-2^{n+1}} a_{n-1} \right) = O \left( 2^{-2^{n+1} + O(n)} \bar{b}_{n-1} \right). \]
In particular, if $n$ is large enough then $\min \{\bar{b}_0, \bar{b}_1, \ldots, \bar{b}_{n-1}\} = \bar{b}_{n-1}$, that is, the box $Q_n$ has thickness $\tau(Q_n) = \bar{b}_{n-1}$. Substituting these estimates back in (5.3) we obtain:
\[ \mathbb{P}_b(G_n^\circ) = 2^{2^n - 2^{n+1} + O(n)} a_{n-1} / \bar{b}_{n-1} = 2^{-2^n + O(n)} = o(1). \]
As remarked before, it follows from (5.1) that $f_0 + g \in L_a$ for $\mathbb{P}_b$-a.e. $g$, as we wanted to prove. \qed

6. Directions for future research

The results of this paper motivate a number of questions.

Consider the property of uniqueness of the maximizing measure (which is of course weaker than the locking property). This property is known to be topologically generic in many function spaces: see [CLT, Je3]. So it is natural to ask whether the uniqueness property is prevalent in function spaces larger than those considered in this paper.

The most obvious open question is whether our prevalence results hold on larger function spaces. The reader will have noticed that several estimates in the proof of Theorem B are very crude. Nevertheless, it is not clear how the evanescence condition (1.2) could be significantly weakened. One of the obstacles is the fact that low-period vertices in $R_n$ belong to many edges: for example the two vertices of period 1 are connected by edges to all other vertices. Such “edgy” vertices give rise to many potential gaps to be controlled. On the other hand, those vertices seem to be “pointy” and should typically generate large gaps, though we have not been able to formalize this. On the other extreme we have the vertices of maximum period, namely the hamiltonian cycles (see Remark 3.4). These vertices seem to be “blunt”. On the positive side it can be shown that each of them belongs to exactly $2^{n-1}$ edges, which is the minimum number allowed by the dimension of the polytope. To summarize, an improvement of our methods seems to require a finer
understanding of the geometry and the combinatorics of the polytopes $R_n$ (or their duals).

Ultimately one would like to go beyond the locking property and formalize the conjecture of Hunt and Ott [HO] (mentioned in Remark 1.2 above) that maximal measures typically have low period.

References

[BLL] BARAVIERA, A.; LEPLAIDEUR, R.; LOPES, A.O. Ergodic optimization, zero temperature limits and the max-plus algebra. IMPA, Rio de Janeiro, 2013. (Cited on page 1.1.)

[Bo1] BOUSCH, T. Le poisson n’a pas d’arêtes. Ann. Inst. H. Poincaré Probab. Statist. 36 (2000), no. 4, 489–508. (Cited on pages 1.1 and 3.1.)

[Bo2] La condition de Walters. Ann. Sci. École Norm. Sup. 34 (2001), no. 2, 287–311. (Cited on pages 1.1 and 2.)

[Bo3] Le lemme de Mañé-Conze-Guivarc’h pour les systèmes amphidynamiques rectifiables. Ann. Fac. Sci. Toulouse Math. 20 (2011), no. 1, 1–14. (Cited on page 1.1.)

[Ch] CHRISTENSEN, J.P.R. On sets of Haar measure zero in abelian Polish groups. Israel J. Math. 13 (1972), 255–260. (Cited on page 1.1.)

[Co1] CONTRERAS, G. Ground states are generically a periodic orbit. Preprint arXiv:1307.0559 (Cited on pages 1.1, 1.2, and 1.2.)

[Co2] Generic Mañé sets. Preprint arXiv:1410.7141 (Cited on page 1.1.)

[CI] CONTRERAS, G.; ITURRIAGA, R. Global minimizers of autonomous Lagrangians. IMPA, Rio de Janeiro, 1999. (Cited on page 1.1.)

[CLT] CONTRERAS, G.; LOPES, A.O.; THEUILLEN, P. Lyapunov minimizing measures for expanding maps of the circle. Ergodic Theory Dynam. Systems 21 (2001), no. 5, 1379–1409. (Cited on page 1.1.)

[CLY] HUNT, B.R.; SAUER, T.; YORKE, J.A. Prevalence: a translation-invariant “almost every” on infinite-dimensional spaces. Bull. Amer. Math. Soc. 27 (1992), no. 2, 217–238. Addendum. Ibid 28 (1993), no. 2, 306–307. (Cited on pages 1.1 and 1.2.)

[Je1] JENKINSON, O. Frequency locking on the boundary of the barycentre set. Experiment. Math. 9 (2000), no. 2, 309–317. (Cited on page 1.1.)

[Je2] Rotation, entropy, and equilibrium states. (English summary) Trans. Amer. Math. Soc. 353 (2001), no. 9, 3713–3739. (Cited on page 3.1.)

[Je3] Ergodic optimization. Discrete Contin. Dyn. Syst. 15 (2006), no. 1, 197–224. (Cited on pages 1.1, 1.2, and 6.)

[Ka] KARP, R.M. A characterization of the minimum cycle mean in a digraph. Discrete Math. 23 (1978), no. 3, 309–311. (Cited on page 3.6.)

[KW] KÜCHERENKO, T.; WOLF, C. Geometry and entropy of generalized rotation sets. Israel J. Math. 199 (2014), no. 2, 791–829. (Cited on page 3.1.)

[Ma] Mañé, R. Generic properties and problems of minimizing measures of Lagrangian systems. Nonlinearity 9 (1996), no. 2, 273–310. (Cited on page 1.1.)
[Mau] Maurer, U.M. Asymptotically tight bounds on the number of cycles in generalized de Bruijn–Good graphs. *Discrete Appl. Math.* 37/38 (1992), 421–436. (Cited on pages 3.4, 3.3, and 3.4.)

[Mo1] Morris, I.D. Maximizing measures of generic Hölder functions have zero entropy. *Nonlinearity* 21 (2008), no. 5, 993–1000. (Cited on page 1.1.)

[Mo2] ———. Ergodic optimization for generic continuous functions. *Discrete Contin. Dyn. Syst.* 27 (2010), no. 1, 383–388. (Cited on page 1.1.)

[OY] Ott, W.; Yorke, J.A. Prevalence. *Bull. Amer. Math. Soc.* 42 (2005), no. 3, 263–290. (Cited on page 1.1.)

[Pi] Pinsker, M.A. *Introduction to Fourier analysis and wavelets*. Reprint of the 2002 original. Graduate Studies in Mathematics, 102. American Mathematical Society, Providence, RI, 2009. (Cited on page 1.3.)

[QS] Quas, A.; Siefken, J. Ergodic optimization of super-continuous functions on shift spaces. *Ergodic Theory Dynam. Systems* 32 (2012), no. 6, 2071–2082. (Cited on pages 1.1, 1.2, and 1.2.)

[Sc] Schrijver, A. *Theory of linear and integer programming*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Ltd., Chichester, 1986. (Cited on pages 3.3 and 3.3.)

[YH] Yuan, G.; Hunt, B.R. Optimal orbits of hyperbolic systems. *Nonlinearity* 12 (1999), no. 4, 1207–1224. (Cited on pages 1.1, 1.2, and 3.)

[Zi] Ziemian, K. Rotation sets for subshifts of finite type. *Fund. Math.* 146 (1995), no. 2, 189–201. (Cited on pages 3.1, 3.2, and 3.8.)

Jairo Bochi jairo.bochi@mat.puc.cl www.mat.uc.cl/~jairo.bochi

Yiwei Zhang yzhang@mat.puc.cl

Facultad de Matemáticas, Pontificia Universidad Católica de Chile. Avenida Vicuña Mackenna 4860, Santiago, Chile