

# ON THE PREVALENCE OF THE PERIODICITY OF MAXIMIZING MEASURES

JIAN DING, ZHIQIANG LI, YIWEI ZHANG

ABSTRACT. For a continuous map  $T: X \rightarrow X$  on a compact metric space  $(X, d)$ , we say that a function  $f: X \rightarrow \mathbb{R}$  has the property  $\mathcal{P}_T$  if its time averages along forward orbits of  $T$  are maximized at a periodic orbit. In this paper, we prove that for the one-sided full shift on two symbols, the property  $\mathcal{P}_T$  is prevalent (in the sense of Hunt–Sauer–Yorke) in spaces of Lipschitz functions with respect to metrics with mildly fast decaying rate on the diameters of cylinder sets. This result is a strengthening of [BZ16, Theorem A], confirms the prediction mentioned in the ICM proceeding contribution of J. Bochi ([Boc18, Section 1]) suggested by experimental evidence, and is another step towards the Hunt–Ott conjectures in the area of ergodic optimization.

## CONTENTS

1. Introduction	2
Acknowledgments	8
2. Preliminaries	8
2.1. Full shift	8
2.2. Haar series	9
2.3. Gauges	10
3. A concrete formulation of Theorem 1.1	11
4. A gap criterion for the locking property	12
4.1. Graph theory	12
4.2. Gap criterion	13
5. Random perturbations on the arc-weight	15
6. Proof of Theorem 3.1	21
7. Proof of Theorem 1.2	22
References	24

---

2010 *Mathematics Subject Classification*. Primary: 37A99; Secondary: 05C80, 37A50, 37C40, 37D35.

*Key words and phrases*. ergodic optimization, maximizing measure, prevalence, (random) maximum mean cycle, ground state.

## 1. INTRODUCTION

Mathematicians have hoped for an analog of the notions of “Lebesgue almost every” and “Lebesgue measure zero” in infinite-dimensional spaces such as various Banach spaces of functions studied in dynamics. Such a desire was recorded in the contribution of A. N. Kolmogorov to the 1954 International Congress of Mathematicians [Ko54].

A natural notion to fulfill this quest called *prevalence* was introduced by B. R. Hunt, T. Sauer, and J. A. Yorke [HSY92]. This notion has since proven to play a central role in understanding generic behaviors in dynamics from a probabilistic (measure-theoretic) perspective. For more history, related notions, and applications in analysis, dynamics, economics, etc., we refer the readers to the surveys by B. R. Hunt and V. Yu. Kaloshin [HK10] and by W. Ott and J. A. Yorke [OY05].

Following [HK10], we recall the notion of prevalence as follows.

Let  $V$  be a topological vector space over  $\mathbb{R}$  equipped with a complete metric. A Borel measure  $\mu$  on  $V$  is called *transverse* to a Borel set  $B \subseteq V$  if  $\mu(\{x + v : x \in B\}) = 0$  for every  $v \in V$ . We say that a Borel subset  $B$  of  $V$  is *shy* if there exists a compactly supported Borel probability measure transverse to  $B$ . A subset  $S$  of  $V$  is called *prevalent* (resp. *shy*) if  $V \setminus S$  (resp.  $S$ ) is contained in a shy Borel subset of  $V$ .

Prevalence has the following properties (see for example, [HK10, Section 2.1]): it is preserved under translation and countable intersections, it implies density, and it coincides with the notion of having full Lebesgue measure when the ambient space is a finite-dimensional Euclidean space.

The aim of this paper is to investigate the prevalence of some ergodic properties in the study of dynamical systems.

Let  $T: X \rightarrow X$  be a continuous transformation. Let  $f: X \rightarrow \mathbb{R}$  be a continuous function, called a potential. We denote the maximal potential energy by

$$\beta(f) := \sup_{\mu \in \mathcal{M}(X, T)} \int f \, d\mu = \max_{\mu \in \mathcal{M}(X, T)} \int f \, d\mu.$$

The last identity follows from the weak\*-compactness of the set  $\mathcal{M}(X, T)$  of  $T$ -invariant Borel probability measures on  $X$ . The quantity  $\beta(f)$  can also be expressed in terms of the maximal time average of  $f$  (see for example, [Je06, Proposition 2.1]). We call a measure  $\mu \in \mathcal{M}(X, T)$  that maximizes the potential energy  $\int f \, d\mu$  a *measure of maximal potential energy* with respect to  $f$  or a  *$f$ -potential-energy-maximizing measure* (or a  *$f$ -maximizing measure*).

If one of the  $f$ -maximizing measures is supported on a periodic orbit of  $T$ , then we say that  $f$  has *property*  $\mathcal{P}_T$  and call such a measure *periodic*. We define the following subsets of the set  $C(X)$  of real-valued continuous functions:

$$\begin{aligned}\mathcal{P}(T) &:= \{f \in C(X) : f \text{ has property } \mathcal{P}_T\}, \\ \mathcal{PU}(T) &:= \{f \in \mathcal{P}(T) : f \text{ has a unique } f\text{-maximizing measure}\}.\end{aligned}$$

The set of functions with property  $\mathcal{P}_T$  has been investigated in various Banach spaces of continuous functions for many dynamical systems, see for example, [Bou00, Bou01, CLT01, QS12, Co16, BZ16, HLMXZ19, LZ23]. The study of typical properties of maximizing measures for Lipschitz and Hölder functions has been particularly fruitful.

Assume that  $(X, d)$  is a compact metric space with infinite cardinality. Let  $\text{Lip}(X, d)$  be the space of real-valued Lipschitz functions on  $(X, d)$ . A function  $f: X \rightarrow \mathbb{R}$  is a *little Lipschitz function* or *locally flat Lipschitz function* if

$$\sup\{|f(x) - f(y)| : x, y \in X, d(x, y) \leq r\} = o(r) \quad \text{as } r \rightarrow 0.$$

The space of real-valued little Lipschitz functions on  $(X, d)$  is denoted by  $\text{lip}(X, d)$  and called the *little Lipschitz space*. We equip both spaces with the *Lipschitz norm*

$$\|f\|_{\text{Lip}} := \|f\|_{\infty} + \text{LIP}_d(f),$$

i.e., the sum of the sup-norm and the minimal Lipschitz constant  $\text{LIP}_d(f)$  of  $f$  given by  $\text{LIP}_d(f) := \sup\{|f(x) - f(y)|/d(x, y) : x, y \in X, x \neq y\}$ . The snowflake  $d^\alpha$  of  $d$  given by  $d^\alpha(x, y) := d(x, y)^\alpha$  is also a metric for each  $\alpha \in (0, 1]$ . The space  $C^{0,\alpha}(X, d)$  of  $\alpha$ -Hölder continuous functions is precisely  $\text{Lip}(X, d^\alpha)$ . The space  $c^{0,\alpha}(X, d) := \text{lip}(X, d^\alpha)$  is called the space of *little  $\alpha$ -Hölder functions* by some authors.

We say that a function  $f \in \text{Lip}(X, d)$  (resp.  $\text{lip}(X, d)$ ) has the *locking*<sup>1</sup> *property* if it has a unique maximizing measure  $\nu \in \mathcal{M}(X, T)$  that is supported on a periodic orbit of  $T$  (in particular,  $f$  has property  $\mathcal{P}_T$ ) and moreover  $\nu$  is the unique maximizing measure for every function in  $\text{Lip}(X, d)$  (resp.  $\text{lip}(X, d)$ ) sufficiently close to  $f$  in the Lipschitz norm. The set of Lipschitz functions  $f \in \text{Lip}(X, d)$  (resp.  $\text{lip}(X, d)$ ) with the locking property is denoted by  $\text{Lock}_T(X, d)$  (resp.  $\text{lock}_T(X, d)$ ). The locking property clearly depends on the metric  $d$ .

The set of Lipschitz functions with the locking property is exactly the interior of the set of Lipschitz functions with property  $\mathcal{P}_T$ . One inclusion is obvious from the definition, and the other was shown by

<sup>1</sup>This is translated from the term *verrouillage* used by T. Bousch in [Bou00, Section 8].

G. Yuan and B. R. Hunt [YH99, Remark 4.5] (see also [BZ15]). The same property holds for little Lipschitz functions, see [LZ23, Theorem 4.1]. In particular, the set of functions with the locking property is dense among the functions with property  $\mathcal{P}_T$ .

Consider the full shift  $(\Sigma^+, \sigma)$  on the set  $\Sigma^+ := \{0, 1\}^{\mathbb{N}}$  of binary sequences. Here the *left-shift operator*  $\sigma: \Sigma^+ \rightarrow \Sigma^+$  is given by

$$\sigma(\{w_i\}_{i \in \mathbb{N}}) = \{w_{i+1}\}_{i \in \mathbb{N}} \quad \text{for } \{w_i\}_{i \in \mathbb{N}} \in \Sigma^+.$$

For a strictly decreasing sequence  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  of positive numbers converging to zero, define a function  $d_{\mathbf{a}}: \Sigma^+ \times \Sigma^+ \rightarrow \mathbb{R}$  by assigning, for each pair of distinct  $\mathbf{w} = \{w_n\}_{n \in \mathbb{N}}, \mathbf{z} = \{z_n\}_{n \in \mathbb{N}} \in \Sigma^+$ ,

$$(1.1) \quad d_{\mathbf{a}}(\mathbf{w}, \mathbf{z}) := a_m$$

where  $m$  is the smallest positive integer with the property that  $w_n \neq z_n$ , and by setting  $d_{\mathbf{a}}(\mathbf{z}, \mathbf{z}) := 0$  for each  $\mathbf{z} \in \Sigma^+$ . It is easy to see that  $d_{\mathbf{a}}$  is a metric on  $\Sigma^+$ .

We consider the Banach spaces  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$  and  $\text{lip}(\Sigma^+, d_{\mathbf{a}})$  with respect to the metric  $d_{\mathbf{a}}$ . Observe that the faster the sequence  $\mathbf{a}$  decays, the more regular the functions in  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$  are.

We are now ready to state our main theorem below.

**Theorem 1.1** (Main Theorem). *Let  $(\Sigma^+, \sigma)$  be the full shift on binary sequences. Let  $\theta$  be a number in  $(0, 1/4)$  and  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  be a strictly decreasing sequence of positive numbers satisfying*

$$(1.2) \quad \frac{a_{n+1}}{a_n} = O(\theta^n) \quad \text{as } n \rightarrow +\infty.$$

*Then  $\mathcal{P}\mathcal{U}(\sigma) \cap \text{Lip}(\Sigma^+, d_{\mathbf{a}})$  is a prevalent set in the space  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$  of real-valued Lipschitz functions, where  $\mathcal{P}\mathcal{U}(\sigma)$  consists of continuous functions  $f: \Sigma^+ \rightarrow \mathbb{R}$  with the property that there is a unique  $f$ -maximizing measure  $\mu$  and that  $\mu$  is supported on a periodic orbit of  $\sigma$ . Moreover,  $\text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$  is a prevalent set in  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$ .*

In general, prevalence does not guarantee topological genericity, neither does the latter imply the former. However, due to the fact that  $\text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$  is known to be an open set in  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$ , topological genericity can be guaranteed by the density alone, which is always a direct consequence of prevalence. Therefore, Theorem 1.1 represents, albeit in our special setting, a strengthening of the kind of topological result obtained in [Co16, Theorem A].

As explained in [Boc18, Section 1], the type of results like Theorem 1.1 provide some confirmation of the experimental findings of B. R. Hunt and E. Ott published over two decades ago [HO96a, HO96b].

In the late 1990s, B. R. Hunt, E. Ott, and G. Yuan conjectured that for a typical chaotic system  $T$ , a typical smooth function  $f$  admits a maximizing measure supported on a periodic orbit, where the typicality on the function  $f$  was both in the sense of probability (in terms of the Lebesgue measure of the parameter(s) for the potential  $f$  in a parameter space) ([HO96a, Conjecture] and [HO96b, Conjecture 2]) and in the sense of topology (topological genericity) ([YH99, Conjecture 1.1]). We recall that such a function  $f$  has property  $\mathcal{P}_T$ . Moreover, B. R. Hunt and E. Ott observed based on numerical evidence that the set of the functions  $f$  without property  $\mathcal{P}_T$  has fractal dimension zero. Despite the recent breakthroughs in the direction of topological genericity initiated by [Co16], the Hunt–Ott conjecture has been much less thoroughly investigated.

For a specific map  $T(x) = 2x$  on the circle, and a one-parameter family  $\rho_\tau(x) = \cos 2\pi(x - \tau)$ , Corollary 1 in [Bou00] shows that the maximizing measure is unique and that it is supported on a periodic orbit for every  $\tau$  except on a set which has zero Lebesgue measure and zero Hausdorff dimension (see also [Je00]). Therefore, the Hunt–Ott conjecture is confirmed in this specific one-parameter family. The notion of full Lebesgue measure coincides in finite dimension with the notion of prevalence in Theorem 1.1.

In fact, Theorem 1.1 is a strengthened version of [BZ16, Theorem A] by J. Bochi and the third-named author, where they assume the *evanescent* condition on  $\mathbf{a}$ , i.e.,  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  is a strictly decreasing sequence of positive numbers satisfying

$$(1.3) \quad \frac{a_{n+1}}{a_n} = O(2^{-2^{n+2}}) \quad \text{as } n \rightarrow +\infty.$$

Theorem 1.1 guarantees the prevalence of continuous functions with the locking property in bigger subspaces of Hölder continuous functions (than the space of functions satisfying (1.3)), which confirms the prediction mentioned in [Boc18, Section 1] suggested by experimental evidence, and is another step towards the Hunt–Ott conjectures.

As a counterpart to Theorem 1.1, the following result for little Lipschitz functions holds.

**Theorem 1.2.** *Assume the hypotheses of Theorem 1.1. Then  $\mathcal{P}\mathcal{U}(\sigma) \cap \text{lip}(\Sigma^+, d_{\mathbf{a}})$  is a prevalent set in  $\text{lip}(\Sigma^+, d_{\mathbf{a}})$ . Moreover,  $\text{lock}_\sigma(\Sigma^+, d_{\mathbf{a}})$  is a prevalent set in  $\text{lip}(\Sigma^+, d_{\mathbf{a}})$ .*

**Remark 1.3.** Since  $d_{\mathbf{a}}^\alpha = d_{\mathbf{a}^\alpha}$  where  $\mathbf{a}^\alpha := \{a_n^\alpha\}_{n \in \mathbb{N}_0}$ , it is easy to see that we can replace the metric  $d_{\mathbf{a}}$  by  $d_{\mathbf{a}}^\alpha$  for  $\alpha \in (-\log_\theta 4, 1]$  in Theorems 1.1 and 1.2 to extend these results to Hölder continuous functions with exponent sufficiently close to 1.

**Main ideas of the proof.** The strategy of the proof of Theorem 1.1 relies on an approach of finite-dimensional approximations via Haar wavelets for ergodic optimization, which can be conveniently re-expressed as a maximum mean cycle problem on de Bruijn–Good directed graphs. Due to the hypothesis (1.2), the coefficients in the Haar series for a Lipschitz function (with respect to  $d_a$ ) decay fast in a controlled way. Our main goal is to show that with high probability, the maximizing measure for a function  $f$  coincides with that for a step function obtained from truncating the Haar series for  $f$  at some finite step in the finite-dimensional approximations. To this end, we adapt some idea from [CCK15, Section 5], and make refined estimates to get a lower bound for the fluctuation on the (conditional) gap between the mean-weights of the two heaviest directed cycles on the sequence of de Bruijn–Good directed graphs. These estimates ensure that the hypothesis (1.2) is sufficient to guarantee the prevalence of the periodicity of the maximizing measure. For a comparison between our strategy and that from [CCK15], see a discussion before the proof of Proposition 5.1 in Section 5.

De Bruijn–Good directed graphs were independently introduced by N. G. de Bruijn [deBr46] and I. J. Good [Go46] in 1946, and have played important roles in both pure mathematics (for example, enumerative combinatorics) and applied mathematics (for example, genome assembly). The appearance of de Bruijn–Good directed graphs in this paper is natural, as walks on these graphs correspond to orbits under the shift operator  $\sigma$  (see [BZ16]).

Investigations on the directed cycle with the maximum mean-weight in a directed graph are sometimes known as the *maximum mean cycle problem*. This problem is one of the most fundamental problems in combinatorial optimization and has a long history of algorithmic studies, for example, [DIG99, Ka78].

It is also worth mentioning that our methods should be extendable to shifts on an arbitrary finite number of symbols or more generally to subshifts of finite type by investigating natural generalizations of de Bruijn–Good directed graphs. Favoring clarity at the expense of generality, we opted to work with the full shift on 2 symbols. In this context, we are able to express with relative ease the random perturbations key to the proof of Theorem 1.1.

To compare Theorem 1.1 with [BZ16, Theorem B], we note that the proof of Theorem B in [BZ16] bounds the probability  $\mathbb{P}_{\mathbf{b}}\{g \in \mathcal{H}_{\mathbf{b}} : \text{Gap}_n(f_0 + g) \leq \epsilon_n\}$  by the maximum number of pairs of neighbours in  $\mathcal{C}(\mathcal{BG}_n)$  (see [BZ16, Equation (5.3)]). By [BZ16, Subsection 3.5], such

a number is no more than  $2^{2^n}$ . Hence the approach in [BZ16] requires the evanescent condition to offset the impact of  $2^{2^n}$ .

In contrast, Proposition 5.1 in this paper bounds a similar probability with more sophisticated probabilistic tools, leading to an improvement from the evanescence condition to condition (1.2) in the current paper. Two main new ingredients are discussed as follows. First, in Propositions 5.1 and 5.2, we obtain the desired bound on the probability of small gap from a bound on some conditional probability  $\mathbb{P}(\{\text{Gap}_{n,t} \leq \epsilon_n\})$  of small gap. Here  $\text{Gap}_{n,t}$  is the absolute value of the difference between the largest two values in  $\{X_t(\mathcal{C}) : \mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)\}$ , where  $\{X_t(\mathcal{C}) : \mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)\}$  has the distribution of the collection of random variables  $\{M_n^{f_0+g}(\mathcal{C}) : \mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)\}$  under the conditioning of  $\{Y_w = t_w : w \in \Sigma_{\leq n-2}\}$ . Second, by bounding the probability density function of  $X_t(\mathcal{C})$  and applying the concentration inequality (for the uniform distribution), we get an upper bound for  $\mathbb{P}(\{\text{Gap}_{n,t} \leq \epsilon_n\})$  under an assumption on exponential decay of  $\epsilon_n$  (rather than the double-exponential decay in [BZ16]).

In view of the connection to the coding of hyperbolic smooth systems, it is natural to consider the case where  $a_n = O(\theta^n)$ , for  $\theta \in (0, 1)$ . In such a case,  $d_a$  coincides with the classical Cantor distance on  $\Sigma^+$ . However, a major obstacle for proving Theorem 1.1 in such a case lies in obtaining an effective lower bound for  $\mathbb{P}(\{\text{Gap}_{n,t} \leq \epsilon_n\})$  when  $\epsilon_n$  decays to zero at a speed slower than an exponential rate (c.f. Proposition 5.1).

In addition, the upper bound  $1/4$  on  $\theta$  in Theorem 1.1 comes from the constants  $2^{-n}$  and  $2^n$  in (5.6) from Proposition 5.1. These constants come from our choices of  $\{Y_w\}_{w \in \Sigma^*}$  as i.i.d. random variables with the uniform distribution on  $[-1, 1]$ . We suspect that the constant  $1/4$  might be improvable by using other distributions for  $\{Y_w\}_{w \in \Sigma^*}$ . However, the alternative proof could be significantly lengthier and too complicated to showcase the main ideas of our strategy. Given that the case where  $a_n = O(\theta^n)$  for  $\theta \in (0, 1)$  remains open, such technical considerations may be better suited for subsequent investigations.

The paper is organized as follows. Section 2 is devoted to preliminaries, where we recall some basic facts about full shifts, Haar series, and the compactness of Hilbert bricks (see Lemma 2.1). Section 3 provides a more concrete formulation (see Theorem 3.1) of our main result Theorem 1.1, with the proof of the implication included at the end of the section. In Section 4 we recall the de Bruijn–Good directed graphs and the gap criterion for the locking property that was developed in [BZ16], see Lemma 4.1. In Section 5 we adapt some idea from [CCK15, Section 5] to make refined estimates for the (conditional) gap between the

mean-weights of the two heaviest directed cycles, see Proposition 5.1 and Proposition 5.2. Using these estimates, we show in Section 6 that hypothesis (1.2) guarantees that with probability 1, the gap criterion is satisfied at some finite step in the finite-dimensional approximations, establishing Theorem 3.1. Thus we complete the proof of Theorem 1.1. Finally, Section 7 is devoted to the proof of Theorems 1.2.

**Acknowledgments.** The authors want to thank the anonymous referee(s) for valuable comments. Y. Zhang is grateful to Peking University for the hospitality, during his stay when part of this work was done. J. Ding was partially supported by NSFC Key Program Project No. 12231002. Z. Li was partially supported by NSFC Nos. 12101017, 12090010, 12090015, and BJNSF No. 1214021. Y. Zhang was partially supported by NSFC Nos. 12161141002, 11871262, and 12271432.

## 2. PRELIMINARIES

In this section, we will introduce the background and necessary concepts.

**2.1. Full shift.** We follow the convention that  $\mathbb{N} := \{1, 2, 3, \dots\}$  and  $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$  in this paper.

For each  $n \in \mathbb{N}$ , the set  $\Sigma_n := \{0, 1\}^n$  of *words of length  $n$*  (or  *$n$ -words*) in the alphabet  $\{0, 1\}$  is the set of binary sequences of length  $n$ . Write  $\Sigma_0 := \{\emptyset\}$  and call  $\emptyset$  the word of length 0. Define  $|\mathbf{w}| := n$  for each  $n$ -word,  $n \in \mathbb{N}_0$ . The set  $\Sigma^+ := \{0, 1\}^{\mathbb{N}}$  of *infinite words* in the alphabet  $\{0, 1\}$  is the set of one-sided infinite binary sequences. Denote the set of *finite words* by  $\Sigma_* := \bigcup_{n \in \mathbb{N}_0} \Sigma_n$  and the set of words of length at most  $k$ , for  $k \in \mathbb{N}_0$ , by  $\Sigma_{\leq k} := \bigcup_{n=0}^k \Sigma_n$ . Note that  $\Sigma_*$  includes  $\emptyset$ , i.e., the word of length 0. We often write  $\{w_n\}_{n \in \mathbb{N}} = w_1 w_2 \dots w_n \dots$  for an infinite word  $\mathbf{w} = \{w_n\}_{n \in \mathbb{N}} \in \Sigma^+$ , and similarly, write  $\{z_n\}_{n=1}^m = z_1 z_2 \dots z_m$  for an  $m$ -word  $\mathbf{z} = \{z_n\}_{n=1}^m \in \Sigma_m$ ,  $m \in \mathbb{N}$ .

The *left-shift operator*  $\sigma: \Sigma^+ \rightarrow \Sigma^+$  is given by

$$\sigma(\{w_i\}_{i \in \mathbb{N}}) = \{w_{i+1}\}_{i \in \mathbb{N}} \quad \text{for } \{w_i\}_{i \in \mathbb{N}} \in \Sigma^+.$$

The pair  $(\Sigma^+, \sigma)$  is called the *full shift* on  $\Sigma^+ = \{0, 1\}^{\mathbb{N}}$ .

For each pair of finite words  $\mathbf{w}, \mathbf{z} \in \Sigma_*$ , define  $\mathbf{wz}$  to be their concatenation if  $\mathbf{w} \neq \emptyset$  and set  $\mathbf{wz} := \mathbf{z}$  otherwise, and denote by  $[\mathbf{w}] := \{\mathbf{wy} : \mathbf{y} \in \Sigma^+\}$  the *cylinder set* beginning at  $\mathbf{w}$ .

An infinite word  $\mathbf{w} \in \Sigma^+$  is a *periodic point* of  $(\Sigma^+, \sigma)$  if  $\sigma^n(\mathbf{w}) = \mathbf{w}$  for some  $n \in \mathbb{N}$ , and the smallest such positive integer  $n$  is called the *period* of  $\mathbf{w}$ . Let  $P_\sigma$  denote the set of all periodic points of  $(\Sigma^+, \sigma)$ .



Let  $\mathbf{x} = \{x_i\}_{i \in \mathbb{N}_0}$  and  $\mathbf{y} = \{y_i\}_{i \in \mathbb{N}_0}$  be two elements of  $\Sigma^+$ . Denote by  $\mathbf{x} \dagger \mathbf{y}$  the position of first disagreement between the sequences  $\mathbf{x}$  and  $\mathbf{y}$ , that is, the least  $m \in \mathbb{N}$  such that  $x_m \neq y_m$ , with the convention  $\mathbf{x} \dagger \mathbf{x} = +\infty$ . The following properties hold:

$$\mathbf{x} \dagger \mathbf{y} = \mathbf{y} \dagger \mathbf{x}, \quad \mathbf{x} \dagger \mathbf{z} \geq \min\{\mathbf{x} \dagger \mathbf{y}, \mathbf{y} \dagger \mathbf{z}\}.$$

For each  $n \in \mathbb{N}$ , we define the  $n$ -th variation of a function  $f: \Sigma^+ \rightarrow \mathbb{R}$  as:

$$(2.1) \quad \text{var}_n(f) := \sup_{\mathbf{x} \dagger \mathbf{y} \geq n} |f(\mathbf{x}) - f(\mathbf{y})|.$$

**2.2. Haar series.** As mentioned in [BZ16, Section 1.3], for each  $\mathbf{w} \in \Sigma_*$ , we define the *Haar function*:

$$(2.2) \quad h_{\mathbf{w}} := \frac{\mathbb{1}_{[\mathbf{w}0]} - \mathbb{1}_{[\mathbf{w}1]}}{2}.$$

These functions are straightforward adaptations of the classical Haar functions on  $[0, 1]$  (see for example, [Pi09, Section 6.3]). It has an advantage to work on the Cantor set  $\Sigma^+$  over working on the interval  $[0, 1]$ : the Haar functions defined in (2.2) are continuous. We follow the choice of the normalization in [BZ16], which makes subsequent formulas simpler.

If a function  $f: \Sigma^+ \rightarrow \mathbb{R}$  satisfies  $\text{var}_n(f) = 0$  for some  $n \in \mathbb{N}_0$ , i.e., it is constant on each cylinder of length  $n$ , then it is called a *step function of level  $n$* . Such functions form a vector space  $\mathcal{S}_n$  of dimension  $2^n$ . For each  $n \in \mathbb{N}$ , the set  $\{1\} \cup \{h_{\mathbf{w}} : |\mathbf{w}| < n\}$  forms a basis of the vector space  $\mathcal{S}_n$ .

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

$$f = c_{\emptyset}(f) + \sum_{\mathbf{w} \in \Sigma_*} c_{\mathbf{w}}(f) h_{\mathbf{w}} \quad (\text{equality in } L^2(\beta)),$$

where the *Haar coefficients* are defined as:

$$(2.3) \quad c_{\emptyset}(f) := \int f \, d\beta, \quad c_{\mathbf{w}}(f) := 2^{|\mathbf{w}|+2} \int f h_{\mathbf{w}} \, d\beta,$$

for each  $\mathbf{w} \in \Sigma_n$ . It follows from [BZ16, Equations (2.2), (2.3), and (2.4)] that the  $n$ -th approximation  $A_n f$  of  $f \in L^2(\beta)$  has the following equivalent characterizations:

- the projection of  $f$  on the subspace  $\mathcal{S}_n$  along the orthogonal complement of  $\mathcal{S}_n$ ;
- the sum of the truncated Haar series:

$$(2.4) \quad A_n f = c_\emptyset(f) + \sum_{|\mathbf{w}| < n} c_{\mathbf{w}}(f) h_{\mathbf{w}};$$

- the function obtained by averaging  $f$  on cylinders of length  $n$ :

$$(2.5) \quad A_n f = \sum_{\mathbf{w} \in \Sigma_n} \left( 2^n \int_{[\mathbf{w}]} f \, d\beta \right) \mathbb{1}_{[\mathbf{w}]}.$$

Thus we have

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

which, under the additional assumption that  $f$  is continuous, converges to 0 as  $n$  tends to  $+\infty$ . Therefore every continuous function  $f$  can be written as a uniformly convergent series:

$$f = c_\emptyset(f) + \sum_{n=0}^{+\infty} \sum_{\mathbf{w} \in \Sigma_n} c_{\mathbf{w}}(f) h_{\mathbf{w}},$$

which, with some abuse of language, we also call a *Haar series*.

**2.3. Gauges.** Let  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  be a strictly decreasing sequence of positive numbers converging to 0.

Consider a function  $f \in C(\Sigma^+)$ . It follows from (2.1) and (2.3) that for each  $n \in \mathbb{N}$  and each  $\mathbf{w} \in \Sigma_n$ , we have

$$(2.6) \quad c_{\mathbf{w}}(f) \leq \text{var}_n(f).$$

Similarly, if  $f$  is Lipschitz with respect to  $d_{\mathbf{a}}$ , then

$$(2.7) \quad \text{LIP}_{d_{\mathbf{a}}}(f) = \sup_{n \in \mathbb{N}} \frac{\text{var}_n(f)}{a_n}.$$

A *gauge* is a family  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  of positive numbers indexed by finite words. We say that a gauge  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  is weakly  $\mathbf{a}$ -admissible if

$$(2.8) \quad \bar{b}_n = o(a_n) \quad \text{as } n \rightarrow +\infty,$$

where

$$\bar{b}_n := \max_{\mathbf{w} \in \Sigma_n} \{b_{\mathbf{w}}\} \quad \text{for each } n \in \mathbb{N}_0.$$

The proof of [BZ16, Lemma 1.1] in [BZ16, Section 2] establishes the following lemma, which is stronger than the statement of [BZ16, Lemma 1.1].

**Lemma 2.1.** *Let  $(\Sigma^+, \sigma)$  be the full shift on binary sequences. Suppose that  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  is a strictly decreasing sequence of positive numbers converging to zero satisfying  $\sum_{k \geq n} a_k = O(a_n)$  as  $n \rightarrow +\infty$ . Let  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  be a weakly  $\mathbf{a}$ -admissible gauge. For a family  $\mathbf{e} = \{e_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  of real numbers satisfying  $|e_{\mathbf{w}}| \leq b_{\mathbf{w}}$  for each  $\mathbf{w} \in \Sigma^*$ , the Haar series  $\sum_{\mathbf{w} \in \Sigma^*} e_{\mathbf{w}} h_{\mathbf{w}}$  converges in the uniform norm to a function  $f_{\mathbf{e}}$  in  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$ . Moreover, the set*

$$(2.9) \quad \mathcal{H}_{\mathbf{b}} := \{f_{\mathbf{e}} : \mathbf{e} = \{e_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*} \text{ with } |e_{\mathbf{w}}| \leq b_{\mathbf{w}} \text{ for each } \mathbf{w} \in \Sigma^*\}$$

*is a compact subset of  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$ .*

The set  $\mathcal{H}_{\mathbf{b}}$  as in the lemma above is called the *Hilbert brick* with gauge  $\mathbf{b}$ . By means of Haar coefficients, it can be identified with the product space  $\prod_{\mathbf{w} \in \Sigma^*} [-b_{\mathbf{w}}, b_{\mathbf{w}}]$ . In particular, we can endow each interval  $[-b_{\mathbf{w}}, b_{\mathbf{w}}]$ ,  $\mathbf{w} \in \Sigma^*$ , with a probability measure  $\mu_{\mathbf{w}}$ . Then by using the identification from Lemma 2.1, we obtain a probability measure  $\mathbb{P}_{\mathbf{b}} := \prod_{\mathbf{w} \in \Sigma^*} \mu_{\mathbf{w}}$  on the Banach space  $\text{Lip}(\Sigma^+, d_{\mathbf{a}})$  supported on the compact subset  $\mathcal{H}_{\mathbf{b}}$ .

In this paper, we are most interested in the special case when all  $\mu_{\mathbf{w}}$  are uniform measures on  $[-b_{\mathbf{w}}, b_{\mathbf{w}}]$ .

### 3. A CONCRETE FORMULATION OF THEOREM 1.1

Theorem 1.1 follows from an explicit construction of a compactly supported probability measure that is transverse to the complement of  $\text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$ . Therefore, in order to establish Theorem 1.1, we prove Theorem 3.1 below, which has concrete information on such a measure.

**Theorem 3.1.** *Let  $(\Sigma^+, \sigma)$  be the full shift on binary sequences. Let  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  be a strictly decreasing sequence of positive numbers satisfying  $a_{n+1}/a_n = O(\theta^n)$  as  $n \rightarrow +\infty$  for some  $\theta \in (0, 1/4)$ . Set  $b_n := a_n/n$  for each  $n \in \mathbb{N}_0$ . Consider the weakly  $\mathbf{a}$ -admissible gauge  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  given by  $b_{\mathbf{w}} := b_{|\mathbf{w}|}$  for each  $\mathbf{w} \in \Sigma^*$ . Then there exists a probability measure  $\mathbb{P}_{\mathbf{b}}$  supported on the compact set  $\mathcal{H}_{\mathbf{b}}$  such that for each  $f_0 \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$ ,*

$$\mathbb{P}_{\mathbf{b}}(\{g \in \mathcal{H}_{\mathbf{b}} : f_0 + g \in \text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})\}) = 1,$$

*and consequently,  $\mathbb{P}_{\mathbf{b}}(\{g \in \text{Lip}(\Sigma^+, d_{\mathbf{a}}) : f_0 + g \notin \text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})\}) = 0$ .*

A proof of Theorem 3.1 will be presented in Section 6.

*Proof of Theorem 1.1 assuming Theorem 3.1.* Theorem 3.1 concludes that the measure  $\mathbb{P}_{\mathbf{b}}$  is transverse (as defined in Section 1) to the complement of the set  $\text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$  and is supported on the compact set

$\mathcal{H}_b$ . Thus  $\text{Lock}_\sigma(\Sigma^+, d_a)$  is a prevalent set (as defined in Section 1) in  $\text{Lip}(\Sigma^+, d_a)$ . Therefore Theorem 3.1 implies Theorem 1.1.  $\square$

#### 4. A GAP CRITERION FOR THE LOCKING PROPERTY

From now on, our aim is to establish Theorem 3.1. To this end, we need to utilize a criterion for the locking property in terms of some probabilistic objects for the maximum mean cycle problem on de Bruijn–Good digraphs.

Let us make some remarks to explain the relations between (infinite dimensional) ergodic optimization, finite-dimensional ergodic optimization, and the maximum mean cycle problem on digraphs. The facts stated below until the start of Subsection 4.1 serve merely to assist the reader in comprehending our approach and will not be used in the proofs.

Recall that  $P_\sigma$  denotes the set of all periodic points of  $(\Sigma^+, \sigma)$ . Each periodic orbit in  $P_\sigma$  is associated to a unique invariant probability measure supported on it. It is well-known that the set of these measures is a dense subset in  $\mathcal{M}(\Sigma^+, \sigma)$ . Consider a sequence of maps  $\{\pi_n\}_{n \in \mathbb{N}_0}$  where  $\pi_n: \mathcal{M}(\Sigma^+, \sigma) \rightarrow \mathbb{R}^{2^n}$  is given by  $\pi_n(\mu) := \{\mu([\mathbf{w}])\}_{\mathbf{w} \in \Sigma_n}$  for  $\mu \in \mathcal{M}(\Sigma^+, \sigma)$ . Denote  $R_n := \pi_n(\mathcal{M}(\Sigma^+, \sigma))$ . It then follows from Sections 3 and 4 in [Zi95] (see also Subsection 3.5 in [BZ16]) that  $R_0 \subset R_1 \subset \dots$  form a nested sequence of finite-dimensional polyhedra, whose vertices are measures supported on some periodic orbits in  $P_\sigma$ . Moreover, each polyhedron  $R_n$  is a projection of the next one  $R_{n+1}$ , and  $\mathcal{M}(\Sigma^+, \sigma)$  can be recovered as the inverse limit of the sequence  $\{R_n\}_{n \in \mathbb{N}_0}$ .

Since  $\mathcal{M}(\Sigma^+, \sigma)$  is a Poulsen simplex, the polyhedra in  $\{R_n\}_{n \in \mathbb{N}_0}$  are not regular simplices. On the contrary, they have huge numbers of vertices, and their faces are small. By Proposition 3.5 in [BZ16],  $R_n$  is isomorphic to the so-called ‘‘circulation polytope’’ of the de Bruijn–Good digraph  $\mathcal{BG}_n$ . With the aid of this isomorphism, the finite-dimensional ergodic optimization problem over  $R_n$  can be conveniently restated as the maximum mean cycle problem on the de Bruijn–Good digraph  $\mathcal{BG}_n$ , and a criterion for the locking property is expressed in terms of the gap between the mean-weights of two heaviest directed cycles in the de Bruijn–Good digraph accordingly, see the gap criterion in Subsection 4.2. This step requires some notions from graph theory.

**4.1. Graph theory.** In this subsection, we go over some key concepts and notation from graph theory. For a more detailed introduction to graph theory, we refer to [BM08].

A *directed graph* (or *digraph*)  $\mathcal{D}$  is an ordered pair  $(V(\mathcal{D}), A(\mathcal{D}))$  of sets, where  $V(\mathcal{D})$  is the set of *vertices* and  $A(\mathcal{D})$  is the set of *arcs*, together with an *incidence function*  $\psi_{\mathcal{D}}: A(\mathcal{D}) \rightarrow V(\mathcal{D}) \times V(\mathcal{D})$  that sends an arc  $a$  to an ordered pair  $(u, v)$  of vertices. For an arc  $a \in A(\mathcal{D})$  with  $\psi_{\mathcal{D}}(a) = (u, v)$ , we call the vertex  $u$  the *tail* of  $a$ , denoted by  $\text{tail}(a)$ , and call  $v$  the *head* of  $a$ , denoted by  $\text{head}(a)$ .

A digraph  $\mathcal{D}'$  is a *subdigraph* of a digraph  $\mathcal{D}$  if  $V(\mathcal{D}') \subseteq V(\mathcal{D})$ ,  $A(\mathcal{D}') \subseteq A(\mathcal{D})$ , and  $\psi_{\mathcal{D}'}$  is the restriction of  $\psi_{\mathcal{D}}$  to  $A(\mathcal{D}')$ . We then say that  $\mathcal{D}$  *contains*  $\mathcal{D}'$  or that  $\mathcal{D}'$  is *contained in*  $\mathcal{D}$ , and write  $\mathcal{D} \supseteq \mathcal{D}'$  or  $\mathcal{D}' \subseteq \mathcal{D}$ , respectively.

A *directed cycle*  $\mathcal{C}$  is a digraph satisfying that  $\text{card } A(\mathcal{C}) = \text{card } V(\mathcal{C}) =: n \in \mathbb{N}$ , that  $V(\mathcal{C}) = \bigcup_{a \in A(\mathcal{C})} \{\text{tail}(a), \text{head}(a)\}$ , and that there is an enumeration  $a_0, a_1, \dots, a_{n-1}$  of  $A(\mathcal{C})$  such that  $\text{head}(a_i) = \text{tail}(a_{i+1})$  for each  $i \in \{0, 1, \dots, n-1\}$ , where  $a_n := a_0$ . For a directed cycle  $\mathcal{C}$  we write  $\text{len}(\mathcal{C})$  for the length (i.e., the number of arcs) of  $\mathcal{C}$ . For a digraph  $D$ , we denote by  $\mathcal{C}(D)$  the collection of all directed cycles contained in  $D$ .

The de Bruijn–Good digraph  $\mathcal{BG}_n$ ,  $n \in \mathbb{N}$ , is a digraph whose set of vertices  $V(\mathcal{BG}_n) := \Sigma_{n-1}$  and set of arcs  $A(\mathcal{BG}_n) := \Sigma_n$  consists of words of length  $n-1$  and words of length  $n$ , respectively, which satisfies that  $\text{tail}(w_1 \dots w_n) = w_1 \dots w_{n-1}$  and  $\text{head}(w_1 \dots w_n) = w_2 \dots w_n$  for each  $n$ -word  $w_1 \dots w_n \in A(\mathcal{BG}_n)$ . The first five de Bruijn–Good digraphs are shown in Figure 4.1.

**4.2. Gap criterion.** Fix a real-valued continuous function  $f \in C(\Sigma^+)$  and an integer  $n \in \mathbb{N}$ . We define inductively weights  $W_n^f(\mathbf{w})$  associated to  $f$  on the arcs  $\mathbf{w}$  of the de Bruijn–Good digraph  $\mathcal{BG}_n$  as follows:

- If  $n = 1$ , then  $W_1^f(\alpha) := \frac{(-1)^\alpha}{2} c_\alpha(f)$  for each  $\alpha \in \{0, 1\}$ ;
- If  $n \geq 2$ , then  $W_n^f(\mathbf{w}\alpha) := W_{n-1}^f(\mathbf{w}) + \frac{(-1)^\alpha}{2} (c_{\mathbf{w}\alpha}(f))$  for all  $\mathbf{w} \in \Sigma_{n-1}$  and  $\alpha \in \{0, 1\}$ .

Thus

$$(4.1) \quad W_n^f(\underbrace{x_1 x_2 \dots x_n}_{\mathbf{w}}) = \frac{1}{2} \sum_{i=0}^{n-1} (-1)^{x_{i+1}} c_{x_1 x_2 \dots x_i}(f).$$

The pair  $(\mathcal{BG}_n, W_n^f)$  induces the *mean-weight* (induced by  $f$ )

$$(4.2) \quad M_n^f(\mathcal{C}) := \frac{1}{\text{len}(\mathcal{C})} \sum_{\mathbf{w} \in A(\mathcal{C})} W_n^f(\mathbf{w})$$

on each directed cycle  $\mathcal{C}$  contained in  $\mathcal{BG}_n$ .

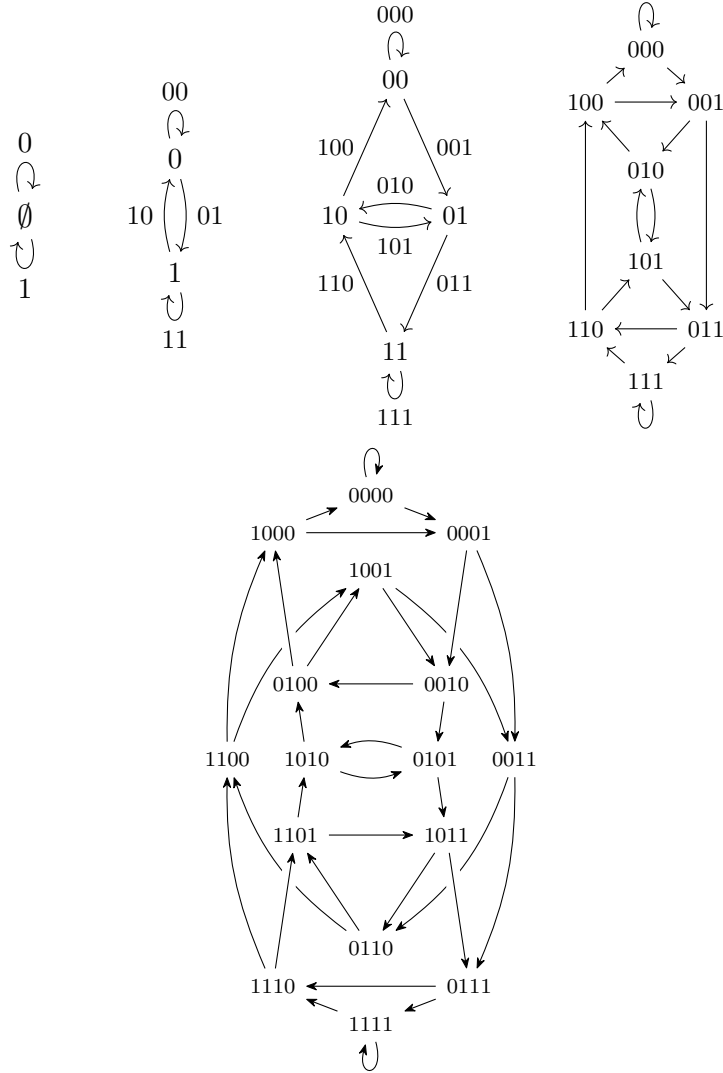


FIGURE 4.1. The de Bruijn–Good digraphs  $\mathcal{BG}_n$  for  $1 \leq n \leq 5$ .

Denote by  $\mathcal{C}_1$  and  $\mathcal{C}_2$  the two heaviest (in terms of mean-weight, in descending order) directed cycles contained in  $\mathcal{BG}_n$ . Then define

$$\text{Gap}_n(f) := M_n^f(\mathcal{C}_1) - M_n^f(\mathcal{C}_2) \geq 0$$

to be the gap in the mean-weights of  $\mathcal{C}_1$  and  $\mathcal{C}_2$  in  $\mathcal{BG}_n$ .

It is easy to check that

$$(4.3) \quad W_n^f(\mathbf{w}) = W_n^{A_n f}(\mathbf{w}),$$

$$(4.4) \quad M_n^f(\mathcal{C}) = M_n^{A_n f}(\mathcal{C}),$$

$$(4.5) \quad \text{Gap}_n(f) = \text{Gap}_n(A_n f)$$

for  $\mathbf{w} \in \Sigma_n$  and  $\mathcal{C} \in A(\mathcal{BG}_n)$ .

In the notations above, we have the following gap criterion from [BZ16, Lemma 4.2].

**Lemma 4.1** (Gap criterion for the locking property [BZ16]). *Let  $(\Sigma^+, \sigma)$  be the full shift on binary sequences. Consider a strictly decreasing sequence  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  of positive numbers satisfying*

$$(4.6) \quad \sum_{n=1}^{+\infty} n a_n < +\infty.$$

If a function  $f \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$  satisfies the following inequality:

$$(4.7) \quad \text{Gap}_n(f) > \sum_{k=n}^{+\infty} (k - n + 1) \max_{\mathbf{w} \in \Sigma_k} \{|c_{\mathbf{w}}(f)|\}.$$

for some  $n \in \mathbb{N}$ , then  $f \in \text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$ .

In informal terms, if the tail of the Haar series is small compared to the gap of its initial part (up to the length  $n - 1$  (see (2.4), (4.1), and (4.5))), so it does not influence the maximizing measure. Note that  $f \in \text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$  then in particular  $f \in \mathcal{PU}(\sigma)$ .

## 5. RANDOM PERTURBATIONS ON THE ARC-WEIGHT

In this section, we will describe the random weights on arcs of the de Bruijn–Good digraphs  $\mathcal{BG}_n$ ,  $n \in \mathbb{N}$ . Namely, we deal with the Haar expansion of  $f_0 + g$ , with  $f_0 \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$  (the deterministic part) and  $g \in \mathcal{H}_{\mathbf{b}}$  (the random part).

Let  $\{Y_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma_*}$  be a collection of i.i.d. random variables, having the uniform distribution on  $[-1, 1]$ , indexed by the set  $\Sigma_*$  of finite words.

Suppose that  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  is a strictly decreasing sequence of positive numbers converging to zero satisfying  $\sum_{k \geq n} a_k = O(a_n)$  as  $n \rightarrow +\infty$ . Let  $\{b_n\}_{n \in \mathbb{N}_0}$  be a sequence of positive numbers satisfying  $b_n = o(a_n)$  as  $n \rightarrow +\infty$ .

We formally define a random function

$$(5.1) \quad g = \sum_{m=0}^{+\infty} \sum_{\mathbf{w} \in \Sigma_m} b_m \cdot Y_{\mathbf{w}} \cdot h_{\mathbf{w}},$$

which has random coefficients on its Haar series. By Lemma 2.1, the series above always converge uniformly, and the limit  $g$  is Lipschitz

with respect to  $d_{\mathbf{a}}$ . For each  $n \in \mathbb{N}$ , the  $n$ -th approximation  $A_n g$  of  $g$  is given by

$$(5.2) \quad A_n g = \sum_{m=0}^{n-1} \sum_{\mathbf{w} \in \Sigma_m} b_m Y_{\mathbf{w}} h_{\mathbf{w}}.$$

Fix an arbitrary function  $f_0 \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$ . With abuse of notation, we use similar symbols for the random arc-weight, mean-weight, and gap induced by random functions that are continuous. In particular, the (random) arc-weights  $W_n^{f_0+g}(\mathbf{w})$ ,  $n \in \mathbb{N}$ , on  $\mathcal{BG}_n$  induced by  $f_0 + g$  are defined by the following inductive process:

- If  $n = 1$ , then  $W_1^{f_0+g}(\alpha) = \frac{(-1)^\alpha}{2}(c_\emptyset(f_0) + b_0 \cdot Y_\emptyset)$  for each  $\alpha \in \{0, 1\}$ ;
- If  $n \geq 2$ , then  $W_n^{f_0+g}(\mathbf{w}\alpha) = W_{n-1}^{f_0+g}(\mathbf{w}) + \frac{(-1)^\alpha}{2}(c_{w_1 w_2 \dots w_{n-1}}(f_0) + b_{n-1} \cdot Y_{\mathbf{w}})$  for all  $\mathbf{w} \in \Sigma_{n-1}$  and  $\alpha \in \{0, 1\}$ .

Fix an integer  $n \in \mathbb{N}$ . We then have

$$(5.3) \quad W_n^{f_0+g}(\underbrace{w_1 w_2 \dots w_n}_{\mathbf{w}}) = \frac{1}{2} \sum_{i=0}^{n-1} (-1)^{w_{i+1}} (c_{w_1 w_2 \dots w_i}(f_0) + b_i Y_{w_1 w_2 \dots w_i}),$$

for  $\mathbf{w} \in \Sigma_n$ . Moreover, the (random) mean-weight of a directed cycle  $\mathcal{C}$  in  $\mathcal{BG}_n$  induced by  $f_0 + g$  is then given by

$$(5.4) \quad M_n^{f_0+g}(\mathcal{C}) = \frac{1}{\text{len}(\mathcal{C})} \sum_{\mathbf{w} \in A(\mathcal{C})} W_n^{f_0+g}(\mathbf{w}).$$

Finally, the (random) gap, induced by  $f_0 + g$ , between the mean-weights of the two heaviest (in terms of mean-weight) directed cycles  $\mathcal{C}_1$  and  $\mathcal{C}_2$  in  $\mathcal{BG}_n$  is

$$(5.5) \quad \text{Gap}_n(f_0 + g) = M_n^{f_0+g}(\mathcal{C}_1) - M_n^{f_0+g}(\mathcal{C}_2) \geq 0.$$

Since uniform measures are continuous, almost surely there is a unique directed cycle with the maximal mean-weight, i.e.,  $\text{Gap}_n(f_0 + g) > 0$ .

The main technical result is formulated below.

**Proposition 5.1.** *Let  $(\Sigma^+, \sigma)$  be the full shift on binary sequences. Let  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  be a strictly decreasing sequence of positive numbers satisfying  $a_{n+1}/a_n = O(\theta^n)$  as  $n \rightarrow +\infty$  for some  $\theta \in (0, 1/2)$ . Let  $\{b_n\}_{n \in \mathbb{N}_0}$  be a sequence of positive numbers satisfying  $b_n = o(a_n)$  as  $n \rightarrow +\infty$ . Consider a weakly  $\mathbf{a}$ -admissible gauge  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  satisfying  $b_{\mathbf{w}} = b_{|\mathbf{w}|}$  for all  $\mathbf{w} \in \Sigma^*$ . Fix an arbitrary function  $f_0 \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$ . Consider the random function  $g$  defined in (5.1).*

*Consider an integer  $n \geq 2$ , a real number  $\epsilon \in (0, 1/2)$ , and a collection  $\mathbf{t} = \{t_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma_{\leq n-2}}$  of real numbers  $t_{\mathbf{w}} \in \mathbb{R}$ . Let  $\{X_{\mathbf{t}}(\mathcal{C}) : \mathcal{C} \in$*



$\mathcal{C}(\mathcal{BG}_n)$  be a collection of random variables with joint distribution given by the conditional distribution of  $\{M_n^{f_0+g}(\mathcal{C}) : \mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)\}$  given  $\{Y_{\mathbf{w}} = t_{\mathbf{w}} : \mathbf{w} \in \Sigma_{\leq n-2}\}$ . Let  $\text{Gap}_{n,t}$  be the absolute value of the difference between the largest two values in  $\{X_t(\mathcal{C}) : \mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)\}$ . Then

$$(5.6) \quad \mathbb{P}(\{\text{Gap}_{n,t} \leq 2^{-n}b_{n-1}\epsilon\}) \leq 2^n\epsilon.$$

*Remark.* Note that  $A(\mathcal{C}_1) \not\subseteq A(\mathcal{C}_2)$  for each pair of distinct directed cycles contained in a directed graph  $\mathcal{D}$ . Then by (5.3), (5.4), and the definition of  $Y_{\mathbf{w}}$  for  $\mathbf{w} \in \Sigma_*$ , we get that the random functions  $X_t(\mathcal{C})$  and  $X_t(\mathcal{C}')$  are distinct linear combinations of random variables  $Y_{\mathbf{w}}$ ,  $\mathbf{w} \in \Sigma_{\leq n-1}$ , (with scalar multiples of Haar functions as coefficients) for distinct directed cycles  $\mathcal{C}$  and  $\mathcal{C}'$  contained in  $\mathcal{BG}_n$ . Moreover,  $\text{Gap}_{n,t}$  in the proposition is well-defined almost surely.

The strategy of our proof of Proposition 5.1 is inspired by the proof of Theorem 3 in [CCK15]. The proof in [CCK15] states the anti-concentration inequalities of Gaussian random vectors and the proof lies in bounding the probability density function of the maximum of a Gaussian random vector. Our proof has three major differences:

- (i) Due to the definition of prevalence, it requires a probability measure which is compactly supported. Thus we cannot directly use the jointly Gaussian random variables (which are not compactly supported) and the resulting estimates in the proof of Theorem 3 in [CCK15].
- (ii) While in [CCK15] special properties of jointly Gaussian random variables are used in an essential way in the general case, we have to work with uniform distributions relying on the combinatorics of de Bruijn–Good digraphs and specific edge weights.
- (iii) Our strategy concentrates on the upper bound on the probability of small conditional gap  $\text{Gap}_{n,t}$ , rather than the maximum of a Gaussian random vector in the proof of Theorem 3 in [CCK15].

*Proof of Proposition 5.1.* Consider  $n$ ,  $\epsilon$ , and  $t$  as given in the hypothesis.

Let  $\mathcal{E}$  be the event that  $|Y_{\mathbf{w}}| \leq 1 - 2\epsilon$  for all  $\mathbf{w} \in \Sigma_{n-1}$ .

We will investigate, for each  $\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)$ , the standard deviation  $\sigma_{\mathcal{C}} := \sigma(X_t(\mathcal{C}))$ , the mean  $\mu_{\mathcal{C}} := \mathbb{E}[X_t(\mathcal{C})]$ , the probability density function  $\phi_{\mathcal{C}}$  for  $X_t(\mathcal{C})$ , and the probability density function  $\tilde{\phi}_{\mathcal{C}}$  for  $X_t(\mathcal{C})$  restricted to  $\mathcal{E}$ . Then it follows from the definition of  $X_t$ , the fact that

$\mathbb{E}[Y_{\mathbf{w}}] = 0$  for  $\mathbf{w} \in \Sigma_*$ , (5.3), and (5.4) that

$$(5.7) \quad X_t(\mathcal{C}) = \mu_{\mathcal{C}} + \frac{b_{n-1}}{2 \text{len}(\mathcal{C})} \sum_{\mathbf{w} \in V(\mathcal{C})} (-1)^{z_{\mathcal{C}}(\mathbf{w})} Y_{\mathbf{w}}$$

for each  $\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)$ . Here and henceforth by  $z_{\mathcal{C}}(\mathbf{w})$  we denote, for each  $\mathbf{w} \in V(\mathcal{C})$ , the unique number in  $\{0, 1\}$  such that  $\mathbf{w}z_{\mathcal{C}}(\mathbf{w}) \in A(\mathcal{C})$ . Note that

$$(5.8) \quad |X_t(\mathcal{C}) - \mu_{\mathcal{C}}| \leq b_{n-1}/2.$$

Define, for each  $\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)$  and each pair of numbers  $x, x' \in \mathbb{R}$ , the event

$$\mathcal{R}_t(\mathcal{C}, x') := \{X_t(\mathcal{C}') \leq x' \text{ for all } \mathcal{C}' \in \mathcal{C}(\mathcal{BG}_n) \setminus \{\mathcal{C}\}\},$$

and the conditional probabilities

$$\begin{aligned} Q_{\mathcal{C}}(x, x') &:= \mathbb{P}(\mathcal{R}_t(\mathcal{C}, x') \mid X_t(\mathcal{C}) = x), \\ \tilde{Q}_{\mathcal{C}}(x, x') &:= \mathbb{P}(\mathcal{E} \cap \mathcal{R}_t(\mathcal{C}, x') \mid X_t(\mathcal{C}) = x), \\ Q_{\mathcal{C}}(x) &:= Q_{\mathcal{C}}(x, x), \\ \tilde{Q}_{\mathcal{C}}(x) &:= \tilde{Q}_{\mathcal{C}}(x, x). \end{aligned}$$

We first verify the following two claims.

*Claim 1.*  $\phi_{\mathcal{C}}(x + \Delta) \geq \tilde{\phi}_{\mathcal{C}}(x)$  for all  $\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)$ ,  $\Delta \in [0, b_{n-1}\epsilon]$ , and  $x \in \mathbb{R}$  satisfying  $|x - \mu_{\mathcal{C}}| \leq 1 - b_{n-1}\epsilon$ .

To establish Claim 1, we fix  $\mathcal{C}$ ,  $\Delta$ , and  $x$  as in the claim. We consider a collection  $\{y_{\mathbf{w}}\}_{\mathbf{w} \in V(\mathcal{C})}$  of real numbers satisfying both

$$(5.9) \quad \mu_{\mathcal{C}} + \frac{b_{n-1}}{2 \text{len}(\mathcal{C})} \sum_{\mathbf{w} \in V(\mathcal{C})} (-1)^{z_{\mathcal{C}}(\mathbf{w})} y_{\mathbf{w}} = x$$

and  $|y_{\mathbf{w}}| \leq 1 - 2\epsilon$  for each  $\mathbf{w} \in V(\mathcal{C})$ . We set

$$y'_{\mathbf{w}} := y_{\mathbf{w}} + (-1)^{z_{\mathcal{C}}(\mathbf{w})} 2\Delta/b_{n-1}$$

for each  $\mathbf{w} \in V(\mathcal{C})$ . Then

$$|y'_{\mathbf{w}}| \leq 1 - 2\epsilon + \frac{2\Delta}{b_{n-1}} \leq 1$$

for each  $\mathbf{w} \in V(\mathcal{C})$ , and

$$\begin{aligned} & \mu_{\mathcal{C}} + \frac{b_{n-1}}{2 \operatorname{len}(\mathcal{C})} \sum_{\mathbf{w} \in V(\mathcal{C})} (-1)^{z_{\mathcal{C}}(\mathbf{w})} y'_{\mathbf{w}} \\ &= \mu_{\mathcal{C}} + \frac{b_{n-1}}{2 \operatorname{len}(\mathcal{C})} \left( \sum_{\mathbf{w} \in V(\mathcal{C})} (-1)^{z_{\mathcal{C}}(\mathbf{w})} y_{\mathbf{w}} + \sum_{\mathbf{w} \in V(\mathcal{C})} \frac{2\Delta}{b_{n-1}} \right) \\ &= x + \Delta. \end{aligned}$$

Claim 1 then follows from (5.7) and the translation invariance of uniformly distributed random variables.

By a similar argument as in the proof of Claim 1, we get the following claim.

*Claim 2.*  $Q_{\mathcal{C}}(x + \Delta, x + (1 - 2^{-n+1})\Delta) \geq \tilde{Q}_{\mathcal{C}}(x)$  for all  $\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)$ ,  $\Delta \in [0, b_{n-1}\epsilon]$ , and  $x \in \mathbb{R}$  satisfying  $|x - \mu_{\mathcal{C}}| \leq 1 - b_{n-1}\epsilon$ .

To establish Claim 2, we fix  $\mathcal{C}$ ,  $\Delta$ , and  $x$  as in the claim. We consider a collection  $\{y_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma_{n-1}}$  of real numbers satisfying both the equality in (5.9) and  $|y_{\mathbf{w}}| \leq 1 - 2\epsilon$  for each  $\mathbf{w} \in \Sigma_{n-1}$ . We set

$$y'_{\mathbf{w}} := y_{\mathbf{w}} + (-1)^{z_{\mathcal{C}}(\mathbf{w})} 2\Delta/b_{n-1}$$

for each  $\mathbf{w} \in V(\mathcal{C})$ , and set  $y'_{\mathbf{u}} := y_{\mathbf{u}}$  for each  $\mathbf{u} \in \Sigma_{n-1} \setminus V(\mathcal{C})$ . Then  $|y'_{\mathbf{w}}| \leq 1$  for all  $\mathbf{w} \in \Sigma_{n-1}$  as

$$|y'_{\mathbf{w}}| \leq 1 - 2\epsilon + \frac{2\Delta}{b_{n-1}} \leq 1$$

for each  $\mathbf{w} \in V(\mathcal{C})$ . On the other hand, consider an arbitrary directed cycle  $\mathcal{C}' \in \mathcal{C}(\mathcal{BG}_n) \setminus \{\mathcal{C}\}$ . Since  $\mathcal{C}'$  and  $\mathcal{C}$  are distinct directed cycles, it is clear that  $A(\mathcal{C}') \setminus A(\mathcal{C}) \neq \emptyset$ . Consequently,

$$\begin{aligned} & \mu_{\mathcal{C}'} + \frac{b_{n-1}}{2 \operatorname{len}(\mathcal{C}')} \sum_{\mathbf{w} \in V(\mathcal{C}')} (-1)^{z_{\mathcal{C}'}(\mathbf{w})} y'_{\mathbf{w}} \\ & \leq \mu_{\mathcal{C}'} + \frac{b_{n-1}}{2 \operatorname{len}(\mathcal{C}')} \left( \sum_{\mathbf{w} \in V(\mathcal{C}')} (-1)^{z_{\mathcal{C}'}(\mathbf{w})} y_{\mathbf{w}} - \frac{2\Delta}{b_{n-1}} + \sum_{\mathbf{w} \in V(\mathcal{C}')} \frac{2\Delta}{b_{n-1}} \right) \\ & = x + \frac{\Delta}{\operatorname{len}(\mathcal{C}')} (-1 + \operatorname{card} V(\mathcal{C}')) \\ & \leq x + (1 - 2^{-n+1})\Delta. \end{aligned}$$

Claim 2 then follows from (5.7) and the translation invariance of uniformly distributed random variables.

Finally, we apply the two claims to finish the proof of the proposition.

Provided  $x' < x$ , we define the events  $\mathcal{Q}(\mathcal{C}) := \mathcal{R}_t(\mathcal{C}, x') \cap \{X_t(\mathcal{C}) = x\}$ ,  $\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)$ . We observe that  $\mathcal{Q}(\mathcal{C})$  and  $\mathcal{Q}(\mathcal{C}')$  are disjoint for distinct directed cycles  $\mathcal{C}$  and  $\mathcal{C}'$  in  $\mathcal{C}(\mathcal{BG}_n)$ .

Define  $\iota := b_{n-1}\epsilon$ . Then by the two claims, the observation above, (5.7), (5.8), and the definitions of  $Q_{\mathcal{C}}$ ,  $\tilde{Q}_{\mathcal{C}}$ ,  $\mathcal{R}_t$ , and  $\mathcal{E}$ , we get

$$\begin{aligned}
& \mathbb{P}(\text{Gap}_{n,t} \geq 2^{-n+1}\iota) \\
& \geq \sum_{\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)} \int_{\iota + \mu_{\mathcal{C}} - 2^{-1}b_{n-1}}^{-\iota + \mu_{\mathcal{C}} + 2^{-1}b_{n-1}} Q_{\mathcal{C}}(x + \iota, x + (1 - 2^{-n+1})\iota) \phi_{\mathcal{C}}(x + \iota) dx \\
& \geq \sum_{\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)} \int_{\iota + \mu_{\mathcal{C}} - 2^{-1}b_{n-1}}^{-\iota + \mu_{\mathcal{C}} + 2^{-1}b_{n-1}} \tilde{Q}_{\mathcal{C}}(x) \tilde{\phi}_{\mathcal{C}}(x) dx \\
& \geq \sum_{\mathcal{C} \in \mathcal{C}(\mathcal{BG}_n)} \mathbb{P}(\mathcal{E} \cap \{\mathcal{C} \text{ is the cycle with maximal mean-weight}\}) \\
& = \mathbb{P}(\mathcal{E}).
\end{aligned}$$

We remark that in the last inequality, we used the fact that  $\tilde{\phi}_{\mathcal{C}}$  is zero outside of  $[\mu_{\mathcal{C}} - 2^{-1}b_{n-1}, \mu_{\mathcal{C}} + 2^{-1}b_{n-1}]$  by (5.8).

As a straightforward consequence of the uniform distribution and the hypothesis that  $\epsilon \in (0, 1/2)$ , we have

$$\mathbb{P}(\mathcal{E}) = (1 - 2\epsilon)^{2^{n-1}} \geq 1 - 2^n\epsilon,$$

where the inequality follows from the well-known inequality  $1 - m\alpha \leq (1 - \alpha)^m$  for  $m \in \mathbb{N}$  and  $\alpha \in (0, 1)$  from the concavity of the logarithm. The proposition is therefore established.  $\square$

The following proposition brings us a step closer to the gap criterion in Lemma 4.1.

**Proposition 5.2.** *Let  $(\Sigma^+, \sigma)$  be the full shift on binary sequences. Let  $\mathbf{a} = \{a_n\}_{n \in \mathbb{N}_0}$  be a strictly decreasing sequence of positive numbers satisfying  $a_{n+1}/a_n = O(\theta^n)$  as  $n \rightarrow +\infty$  for some  $\theta \in (0, 1/2)$ . Let  $\{b_n\}_{n \in \mathbb{N}_0}$  be a sequence of positive numbers satisfying  $b_n = o(a_n)$  as  $n \rightarrow +\infty$ . Consider a weakly  $\mathbf{a}$ -admissible gauge  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma^*}$  satisfying  $b_{\mathbf{w}} = b_{|\mathbf{w}|}$  for all  $\mathbf{w} \in \Sigma^*$ .*

*Then there exists a measure  $\mathbb{P}_{\mathbf{b}}$  supported on  $\mathcal{H}_{\mathbf{b}}$  with the property that for every  $f_0 \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$ , we have*

$$\mathbb{P}_{\mathbf{b}}(\{g \in \mathcal{H}_{\mathbf{b}} : \exists N \in \mathbb{N}, \forall n \geq N, \text{Gap}_n(f_0 + g) \geq 4^{-n}n^{-3}b_{n-1}\}) = 1.$$

*Proof.* Fix an arbitrary integer  $n \geq 2$ .

It follows from Proposition 5.1 that

$$\mathbb{P}(\{\text{Gap}_{n,t} \leq 2^{-n}b_{n-1}\epsilon\}) \leq 2^n\epsilon$$

for all  $\epsilon \in (0, 1/2)$ .

We denote by  $\phi_{n-2}$  the joint probability density function for all the random variables in  $\{Y_{\mathbf{w}} : \mathbf{w} \in \Sigma_{\leq n-2}\}$ . Then for each  $\epsilon \in (0, 1/2)$ ,

$$\begin{aligned} & \mathbb{P}\{\text{Gap}_n(f_0 + g) \leq 2^{-n}b_{n-1}\epsilon\} \\ & \leq \int \mathbb{P}\{\text{Gap}_{n,t} \leq 2^{-n}b_{n-1}\epsilon\} \phi_{n-2}(\mathbf{t}) \, d\mathbf{t} \\ & \leq 2^n \epsilon. \end{aligned}$$

Taking  $\epsilon := 2^{-n}n^{-3}$ , then, we have

$$\sum_{n=2}^{+\infty} \mathbb{P}\{\text{Gap}_n(f_0 + g) \leq 4^{-n}n^{-3}b_{n-1}\} \leq \sum_{n=2}^{+\infty} n^{-3} < +\infty.$$

By the Borel–Cantelli lemma, therefore, the proposition follows.  $\square$

## 6. PROOF OF THEOREM 3.1

We are now ready to establish the more concrete formulation Theorem 3.1 of our main theorem (Theorem 1.1).

*Proof of Theorem 3.1.* Fix an arbitrary function  $f_0 \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$ . Define

$$\Lambda_{f_0} := \{g \in \mathcal{H}_{\mathbf{b}} : \exists N \in \mathbb{N}, \forall n \geq N, \text{Gap}_n(f_0 + g) \geq 4^{-n}n^{-3}b_{n-1}\}.$$

Next, for each  $g \in \Lambda_{f_0}$ , by the definition of  $\Lambda_{f_0}$ , for all sufficiently large  $n \in \mathbb{N}$ ,

$$(6.1) \quad \text{Gap}_n(f_0 + g) \geq 4^{-n}n^{-3}b_{n-1}.$$

On the other hand, since by the hypothesis on  $\mathbf{a}$  and  $\mathbf{b}$ ,

$$(6.2) \quad \sum_{k=1}^{+\infty} ka_k < +\infty,$$

and since  $\theta \in (0, 1/4)$ , for all sufficiently large integers  $n$ , one has

$$(6.3) \quad \frac{b_{n-1}}{b_n} > \max\{\text{LIP}_{d_{\mathbf{a}}}(f_0), 1\}4^{n+1}n^4$$

and

$$(6.4) \quad a_n \geq \sum_{k=n+1}^{+\infty} (k - n + 1)a_k.$$

Combining (6.1), (6.3),  $b_n = a_n/n$ , (6.4), (2.6), (2.7), and the triangle inequality, we deduce that for each  $N \in \mathbb{N}$ , there exists an integer  $n \geq N$  such that

$$\begin{aligned}
& \text{Gap}_n(f_0 + g) \\
& \geq 4^{-n} n^{-3} b_{n-1} && \text{(by (6.1))} \\
& > 4 \max\{\text{LIP}_{d_a}(f_0), 1\} n b_n && \text{(by (6.3))} \\
& = 4 \max\{\text{LIP}_{d_a}(f_0), 1\} a_n && \text{(by } b_n = a_n/n) \\
& \geq 2 \max\{\text{LIP}_{d_a}(f_0), 1\} \sum_{k=n}^{+\infty} (k - n + 1) a_k && \text{(by (6.4))} \\
& \geq \sum_{k=n}^{+\infty} (k - n + 1) (a_k \text{LIP}_{d_a}(f_0) + b_k) && \text{(by } b_k = a_k/k).
\end{aligned}$$

Note that by (2.6) and (2.7),  $a_k \text{LIP}_{d_a}(f_0) \geq \max_{\mathbf{w} \in \Sigma_k} \{|c_{\mathbf{w}}(f_0)|\}$ . On the other hand, since  $g \in \Lambda_{f_0}$ , by the definition of  $\Lambda_{f_0}$  and Lemma 2.1, we have  $b_k \geq \max_{\mathbf{w} \in \Sigma_k} \{|c_{\mathbf{w}}(g)|\}$ . Hence

$$\begin{aligned}
& \text{Gap}_n(f_0 + g) \\
& \geq \sum_{k=n}^{+\infty} (k - n + 1) \left( \max_{\mathbf{w} \in \Sigma_k} \{|c_{\mathbf{w}}(f_0)|\} + \max_{\mathbf{w} \in \Sigma_k} \{|c_{\mathbf{w}}(g)|\} \right) \\
& \geq \sum_{k=n}^{+\infty} (k - n + 1) \max_{\mathbf{w} \in \Sigma_k} \{|c_{\mathbf{w}}(f_0 + g)|\}.
\end{aligned}$$

This means that  $f_0 + g$  satisfies the gap criterion (4.7) at length  $n$ . Together with (6.2) and Lemma 4.1 (on  $f_0 + g$ ), it then follows that  $f_0 + g \in \text{Lock}_\sigma(\Sigma^+, d_a)$ .

Hence for each  $f_0 \in \text{Lip}(\Sigma^+, d_a)$  we have

$$\Lambda_{f_0} \subseteq \{g \in \mathcal{H}_b : f_0 + g \in \text{Lock}_\sigma(\Sigma^+, d_a)\}.$$

Then it follows from Proposition 5.2,

$$1 = \mathbb{P}_b(\Lambda_{f_0}) \leq \mathbb{P}_b(\{g \in \mathcal{H}_b : f_0 + g \in \text{Lock}_\sigma(\Sigma^+, d_a)\}) \leq 1.$$

The proof of Theorem 3.1 is therefore complete.  $\square$

## 7. PROOF OF THEOREM 1.2

Recall the definition of little Lipschitz spaces reviewed in Section 1.

*Proof of Theorem 1.2.* It suffices to show that  $\text{lock}_\sigma(\Sigma^+, d_a)$  is a prevalent set in  $\text{lip}(\Sigma^+, d_a)$ .

Set  $b_n := a_n/n$  for each  $n \in \mathbb{N}_0$ . Consider the weakly  $\mathbf{a}$ -admissible gauge  $\mathbf{b} = \{b_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma_*}$  given by  $b_{\mathbf{w}} := b_{|\mathbf{w}|}$  for each  $\mathbf{w} \in \Sigma_*$ .

We first show that the Hilbert brick  $\mathcal{H}_{\mathbf{b}}$  defined in (2.9) in Lemma 2.1 is a subset of  $\text{lip}(\Sigma^+, d_{\mathbf{a}})$ .

Indeed, for each  $\mathbf{e} = \{e_{\mathbf{w}}\}_{\mathbf{w} \in \Sigma_*}$  with  $|e_{\mathbf{w}}| \leq b_{\mathbf{w}}$  for each  $\mathbf{w} \in \Sigma_*$ , by Lemma 2.1 the uniform limit  $f_{\mathbf{e}} \in \text{Lip}(\Sigma^+, d_{\mathbf{a}})$  of the Haar series  $\sum_{\mathbf{w} \in \Sigma_*} e_{\mathbf{w}} h_{\mathbf{w}}$  satisfies that for each  $n \in \mathbb{N}$  sufficiently large and each pair of  $\mathbf{x}, \mathbf{y} \in \Sigma^+$  with  $\mathbf{x} \dagger \mathbf{y} = n$ , by (2.2), the hypothesis on  $\mathbf{a}$ , and our choice of  $b_n$ , we have

$$|f_{\mathbf{e}}(\mathbf{x}) - f_{\mathbf{e}}(\mathbf{y})| \leq \sum_{m=n}^{+\infty} b_m \leq 2b_n = 2a_n/n = o(a_n).$$

Thus by the definitions of the metric  $d_{\mathbf{a}}$  and of the little Lipschitz space, we have  $f_{\mathbf{e}} \in \text{lip}(\Sigma^+, d_{\mathbf{a}})$ . It follows from (2.9) that  $\mathcal{H}_{\mathbf{b}} \subseteq \text{lip}(\Sigma^+, d_{\mathbf{a}})$ .

Note that  $\text{lip}(\Sigma^+, d_{\mathbf{a}}) \subseteq \text{Lip}(\Sigma^+, d_{\mathbf{a}})$  and that

$$\text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}}) \cap \text{lip}(\Sigma^+, d_{\mathbf{a}}) \subseteq \text{lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}}).$$

Hence by Theorem 3.1, the probability measure  $\mathbb{P}_{\mathbf{b}}$  from Theorem 3.1 supported on the compact set  $\mathcal{H}_{\mathbf{b}}$  satisfies that for each  $f_0 \in \text{lip}(\Sigma^+, d_{\mathbf{a}})$ ,

$$\begin{aligned} & \mathbb{P}_{\mathbf{b}}(\{g \in \mathcal{H}_{\mathbf{b}} : f_0 + g \in \text{lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})\}) \\ & \geq \mathbb{P}_{\mathbf{b}}(\{g \in \mathcal{H}_{\mathbf{b}} : f_0 + g \in \text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}}) \cap \text{lip}(\Sigma^+, d_{\mathbf{a}})\}) \\ & = \mathbb{P}_{\mathbf{b}}(\{g \in \mathcal{H}_{\mathbf{b}} : f_0 + g \in \text{Lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})\}) = 1. \end{aligned}$$

Therefore,  $\text{lock}_{\sigma}(\Sigma^+, d_{\mathbf{a}})$  is a prevalent set in  $\text{lip}(\Sigma^+, d_{\mathbf{a}})$ , and the theorem is established.  $\square$

As more detailed discussions in [LZ23, Section 1] suggest, there is a “dictionary” for the correspondences between thermodynamic formalism and its tropical counterpart, ergodic optimization. In this dictionary, the existence, uniqueness, and equidistribution of periodic points for the equilibrium state may be considered as the counterparts to the existence, uniqueness, and periodicity of (the support of) the maximizing measure. In the current paper, we investigated in yet another “language”, namely, a probabilistic one through random maximum mean cycle problems on digraphs using probabilistic tools. Here directed cycles on digraphs translate to periodic orbits, and maximum mean cycles to maximizing measures. Similar analogies can be drawn between these three “languages” in the “dictionary.” It would be interesting to investigate further into such connections, especially the probabilistic aspects.

## REFERENCES

- [Boc18] BOCHI, J., Ergodic optimization of Birkhoff averages and Lyapunov exponents. In *Proc. Internat. Congr. Math. (Rio de Janeiro 2018)*, Volume III, World Sci. Publ., Singapore, 2018, pp. 1843–1866.
- [BZ15] BOCHI, J. and ZHANG, Y., Note on robustness of periodic measures in ergodic optimization. Available at <http://www.personal.psu.edu/jzd5895/>.
- [BZ16] BOCHI, J. and ZHANG, Y., Ergodic optimization of prevalent super-continuous functions. *Int. Math. Res. Not. IMRN* 19 (2016), 5988–6017.
- [BM08] BONDY, J.A. and MURTY, U.S.R., *Graph theory*, volume 244 of *Grad. Texts in Math.*, Springer, New York, 2008.
- [Bou00] BOUSCH, T., Le poisson n’a pas d’arêtes. *Ann. Inst. Henri Poincaré Probab. Stat.* 36 (2000), 489–508.
- [Bou01] BOUSCH, T., La condition de Walters. *Ann. Sci. Éc. Norm. Supér. (4)* 34 (2001), 287–311.
- [deBr46] DE BRUIJN, N.G., A combinatorial problem. *Nederl. Akad. Wetensch. Proc.* 49 (1946), 758–764.
- [CCK15] CHERNOZHUKOV, V., CHETVERIKOV, D., and KATO, K., Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* 162 (2015), 47–70.
- [Co16] CONTRERAS, G., Ground states are generically a periodic orbit. *Invent. Math.* 205 (2016), 383–412.
- [CLT01] CONTRERAS, G., LOPES, A.O., and THIEULLEN, P., Lyapunov minimizing measures for expanding maps of the circle. *Ergodic Theory Dynam. Systems* 21 (2001), 1379–1409.
- [DIG99] DASDAN, A., IRANI, S.S., and GUPTA, R.K., Efficient algorithms for optimum cycle mean and optimum cost to time ratio problems. In *Proceedings of the 36th annual ACM/IEEE Design Automation Conference*, ACM, New York, 1999, pp. 37–42.
- [Go46] GOOD, I.J., Normal Recurring Decimals. *J. Lond. Math. Soc.* 21 (1946), 167–169.
- [HLMXZ19] HUANG, W., LIAN, Z., MA, X., XU, L., and ZHANG, Y., Ergodic optimization theory for a class of typical maps. Preprint, (arXiv:1904.01915v3), 2019.
- [HK10] HUNT, B.R. and KALOSHIN, V.YU., Prevalence. In *Handbook of Dynamical Systems*, Volume 3, Elsevier, 2010, pp. 43–87.
- [HO96a] HUNT, B.R. and OTT, E., Optimal periodic orbits of chaotic systems. *Phys. Rev. Lett.* 76 (1996), 2254–2257.
- [HO96b] HUNT, B.R. and OTT, E., Optimal periodic orbits of chaotic systems occur at low period. *Phys. Rev. E* 54 (1996), 328–337.
- [HSY92] HUNT, B.R., SAUER, T., and YORKE, J.A., Prevalence: a translation-invariant “almost every” on infinite-dimensional spaces. *Bull. Amer. Math. Soc.* 27 (1992), 217–238.
- [Je00] JENKINSON, O., Frequency locking on the boundary of the barycentre set. *Exp. Math.* 9 (2000), 309–317.
- [Je06] JENKINSON, O., Ergodic optimization. *Discrete Contin. Dyn. Syst.* 15 (2006), 197–224.
- [Ka78] KARP, R.M., A characterization of the minimum cycle mean in a digraph. *Discrete Math.* 23 (1978), 309–311.



- [Ko54] KOLMOGOROV, A.N., Théorie générale des systèmes dynamiques et mécanique classique. In *Proc. Internat. Congr. Math. (Amsterdam 1954)*, Volume I, North-Holland Publishing Co., Amsterdam, 1957, pp. 315–333.
- [LZ23] LI, Z. and ZHANG, Y., Ground states and periodic orbits for expanding Thurston maps. Preprint, (arXiv:2303.00514), 2023.
- [OY05] OTT, W. and YORKE, J.A., Prevalence. *Bull. Amer. Math. Soc.* 42 (2005), 263–290.
- [Pi09] PINSKY, M.A. *Introduction to Fourier analysis and wavelets*. Reprint of the 2002 original. Graduate Studies in Mathematics, 102. American Mathematical Society, Providence, RI, 2009.
- [QS12] QUAS, A. and SIEFKEN, J., Ergodic optimization of super-continuous functions on shift spaces. *Ergodic Theory Dynam. Systems* (6) 32 (2012), 2071–2082.
- [YH99] YUAN, G. and HUNT, B.R., Optimal orbits of hyperbolic systems. *Nonlinearity* 12 (1999), 1207–1224.
- [Zi95] ZIEMIAN, K., Rotation sets for subshifts of finite type. *Fund. Math.* 146 (1995), 189–201.

JIAN DING, SCHOOL OF MATHEMATICAL SCIENCES, PEKING UNIVERSITY, BEIJING 100871, CHINA

*Email address:* dingjian@math.pku.edu.cn

ZHIQIANG LI, SCHOOL OF MATHEMATICAL SCIENCES & BEIJING INTERNATIONAL CENTER FOR MATHEMATICAL RESEARCH, PEKING UNIVERSITY, BEIJING 100871, CHINA

*Email address:* zli@math.pku.edu.cn

YIWEI ZHANG, DEPARTMENT OF MATHEMATICS, SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY, SHENZHEN, GUANGDONG 518055, CHINA.

*Email address:* zhangyw@sustech.edu.cn