J. Eur. Math. Soc. 28, 3983–4046 (2026) DOI 10.4171/JEMS/1545

Harold Erazo · Rodolfo Gutiérrez-Romo · Carlos Gustavo Moreira · Sergio Romaña

Fractal dimensions of the Markov and Lagrange spectra near 3

Received October 3, 2022; revised June 29, 2023

Abstract. The Lagrange spectrum $\mathcal{L}$ and Markov spectrum $\mathcal{M}$ are subsets of the real line with complicated fractal properties that appear naturally in the study of Diophantine approximations. It is known that the Hausdorff dimensions of the intersections of these sets with any half-line coincide, that is, $\dim_H(\mathcal{L} \cap (-\infty, t)) = \dim_H(\mathcal{M} \cap (-\infty, t)) =: d(t)$ for every $t \geq 0$. It is also known that $d(3) = 0$ and $d(3+\varepsilon) > 0$ for every $\varepsilon > 0$. We show that, for sufficiently small values of $\varepsilon > 0$, one has the approximation $d(3+\varepsilon) = 2 \cdot \frac{W(e^{c_0}|\log \varepsilon|)}{|\log \varepsilon|} + O(\frac{\log|\log \varepsilon|}{|\log \varepsilon|^2})$, where W denotes the Lambert function (the inverse of $f(x) = xe^x$) and $c_0 = -\log\log((3+\sqrt{5})/2) \approx 0.0383$. We also show that this result is optimal for the approximation of $d(3+\varepsilon)$ by "reasonable" functions, in the sense that, if $F(t)$ is a C^2 function such that $d(3+\varepsilon) = F(\varepsilon) + o(\frac{\log|\log \varepsilon|}{|\log \varepsilon|^2})$, then its second derivative $F''(t)$ changes sign infinitely many times as t approaches 0.

Keywords: Lagrange spectrum, Markov spectrum, Hausdorff dimension, continued fraction, Gauss–Cantor set.

1. Introduction

1.1. The Lagrange spectrum

The Lagrange spectrum is a subset of the real line which appears naturally in the study of Diophantine approximations of real numbers.

Harold Erazo: IMPA – Instituto de Matemática Pura e Aplicada, 22460-320 Rio de Janeiro, Brazil; harolderaz@gmail.com

Rodolfo Gutiérrez-Romo: Departamento de Ingeniería Matemática & Centro de Modelamiento Matemático, Universidad de Chile, 8370458 Santiago, Chile; g-r@rodol.fo

Carlos Gustavo Moreira: SUSTech International Center for Mathematics, 518055 Shenzhen, P. R. China; IMPA - Instituto de Matemática Pura e Aplicada, 22460-320 Rio de Janeiro, Brazil; gugu@impa.br

Sergio Romaña: Department of Mathematics, Universidade Federal do Rio de Janeiro, 21941-909 Rio de Janeiro, Brazil; sergiori@im.ufrj.br

Mathematics Subject Classification 2020: 11J06 (primary); 11J70, 28A78, 28A80, 37A44, 37B10 (secondary).

Consider an irrational real number $x \in \mathbb{R} \setminus \mathbb{Q}$. By Dirichlet's approximation theorem, there exist infinitely many pairs of integers p, q with $q > 0$ satisfying

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^2}.$$

This result is not tight. Indeed, Hurwitz's theorem states that for infinitely many such pairs p, q,

$$\left| x - \frac{p}{q} \right| < \frac{1}{\sqrt{5} q^2}.$$

This is the best possible inequality of this type that holds for every irrational x. Indeed, if $x = \frac{1+\sqrt{5}}{2}$, the constant $\sqrt{5}$ *cannot* be replaced by a larger constant while preserving the existence of infinitely many such pairs p, q for which the corresponding inequality holds. However, for other irrational values of x we may hope for better results. Following this idea, we define $L(x)$ as the supremum of the set of all $\ell > 0$ such that

$$\left| x - \frac{p}{q} \right| < \frac{1}{\ell q^2}$$

holds for infinitely many pairs of integers p, q with $q > 0$ (possibly with $L(x) = \infty$). The number $L(x)$ is known as the *Lagrange value* of x, and the *Lagrange spectrum* is defined as the set of all finite Lagrange values:

$$\mathcal{L} = \{L(x) < \infty \mid x \in \mathbb{R} \setminus \mathbb{Q}\}.$$

By means of the continued-fraction expansion of x, it is possible to obtain a symbolic-dynamical characterization of the Lagrange spectrum. Indeed, consider the infinite sequence $(c_n)_{n \geq 0}$ such that

$$x = [c_0; c_1, c_2, c_3, \ldots] = c_0 + \cfrac{1}{c_1 + \cfrac{1}{c_2 + \cfrac{1}{c_3 + \ddots}}},$$

that is, $(c_n)_{n \geq 0}$ is the continued-fraction expansion of x. It is well-known that

$$x - \frac{p_n}{q_n} = (-1)^n \frac{1}{(\alpha_{n+1} + \beta_{n+1}) q_n^2},$$

where we set $\alpha_{n+1} = [c_{n+1}; c_{n+2}, c_{n+3}, \ldots]$, $\beta_{n+1} = [0; c_n, c_{n-1}, \ldots, c_1]$, and where $p_n/q_n = [c_0; c_1, c_2, \ldots, c_n]$. It is also known that these *convergents* p_n/q_n of the continued-fraction expansion of x are the best rational approximations of x for instance in the following sense: if p, q are integers with $q > 0$ and $|x - p/q| < \frac{1}{2q^2}$, then $p/q = p_n/q_n$ for some $n \in \mathbb{N}$. From these facts, we obtain the following expression for the Lagrange value of x:

$$L(x) = \limsup_{n \to \infty} (\alpha_{n+1} + \beta_{n+1}).$$

If we define $\beta'_{n+1} = [0; c_n, c_{n-1}, \ldots, c_1, 1, 1, \ldots]$, we also have

$$L(x) = \limsup_{n\to\infty}(\alpha_{n+1} + \beta'_{n+1}),$$

since the trailing sequence of 1's does not change the value in the limit.

It follows that

$$\mathcal{L} = \left\{\limsup_{n\to\infty} \lambda(\sigma^n(\omega)) \;\middle|\; \omega \in (\mathbb{N}^*)^{\mathbb{Z}}\right\},$$

where, for $\omega = (\omega_n)_{n\in\mathbb{Z}} \in (\mathbb{N}^*)^{\mathbb{Z}}$, $\lambda(\omega) = [\omega^+] + [0; \omega^-]$ with $\omega^+ = (\omega_n)_{n\geq 0}$ and $\omega^- = (\omega_{-n})_{n\geq 1}$.

We refer the reader to the expository article by Bombieri [1] and to the books by Cusick–Flahive [3] and by Lima–Matheus–Moreira–Romaña [7] for a more detailed account of these constructions.

1.2. The Markov spectrum

The Markov spectrum is another fractal subset of the real line which is closely related to the Lagrange spectrum. Using the symbolic-dynamical definition of the Lagrange spectrum as a starting point, it can be defined similarly as

$$\mathcal{M} = \left\{\sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega)) \;\middle|\; \omega \in (\mathbb{N}^*)^{\mathbb{Z}}\right\}.$$

We denote by $\mathsf{m}(\omega) = \sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega))$ the *Markov value* of $\omega \in (\mathbb{N}^*)^{\mathbb{Z}}$.

This set is also related to some Diophantine approximation problems. Indeed, it encodes the (inverses of) minimal possible values of real indefinite quadratic forms with normalized discriminants (equal to 1). Nevertheless, throughout this article we will only use the symbolic-dynamical definitions of $\mathcal{L}$ and $\mathcal{M}$.

1.3. Structure of the Lagrange and Markov spectra

Both the Lagrange and Markov spectra have been intensively studied since the seminal work of Markov [8]. In particular, it is well-known that

$$\mathcal{L} \cap [0, 3) = \mathcal{M} \cap [0, 3) = \{\sqrt{5} < \sqrt{8} < \sqrt{221}/5 < \cdots\},$$

that is, $\mathcal{L}$ and $\mathcal{M}$ coincide below 3 and consist of a sequence of *explicit* quadratic surds accumulating only at 3. Moreover, it is also possible to explicitly characterize the sequences $\omega \in (\mathbb{N}^*)^{\mathbb{Z}}$ associated with Markov values less than or equal to 3 [1, Theorem 15].

On the other hand, the behavior of these sets after 3 remains somewhat mysterious. Indeed, it is known that $\mathcal{L} \subseteq \mathcal{M}$ and some authors conjectured that these sets are *equal*; Freĭman [4] disproved this conjecture only in 1968. Much more is now known in this regard: the Hausdorff dimension of the complement $\mathcal{M} \setminus \mathcal{L}$ lies strictly between 0 and 1 [9].

Even if the previous paragraph suggests that these sets are somewhat different, they are known to coincide before 3 and after large enough values. Indeed, Hall [6] showed in 1947 that $\mathcal{L}$ (and thus also $\mathcal{M}$) contains a half-line $[c,\infty)$; any such ray is hence known as a *Hall ray*. After several years, Freĭman [5] found the largest Hall ray to be $[c_{\mathrm{F}},\infty)$, where $c_{\mathrm{F}} \approx 4.5278\ldots$ is an explicit quadratic surd known as Freĭman's constant. These results in turn imply that $\mathcal{L}$ and $\mathcal{M}$ coincide starting at c_{F}, so they both contain the half-line $[c_{\mathrm{F}},\infty)$.

There are more striking similarities between these two sets. In particular, their Hausdorff dimensions coincide when truncated: the third author showed that

$$\dim_{\mathrm{H}}(\mathcal{L}\cap(-\infty,t)) = \dim_{\mathrm{H}}(\mathcal{M}\cap(-\infty,t))$$

for every $t>0$ [10]. Clearly, this result shows that when studying the Hausdorff dimension of such truncated versions, one can choose to use either $\mathcal{L}$ or $\mathcal{M}$.

Let

$$d(t) := \dim_{\mathrm{H}}(\mathcal{L}\cap(-\infty,t)) = \dim_{\mathrm{H}}(\mathcal{M}\cap(-\infty,t)).$$

Moreira [10] also proved the following nice formula:

$$d(t) = \min\{1, 2\cdot D(t)\},$$

where $D(t)=\dim_{\mathrm{H}}(K_t)$, and

$$\begin{aligned} K_t = \{[0;c_1,\ldots,c_n,\ldots] \mid \text{there exists } (c_{-n})_{n\geq 0} \in (\mathbb{N}^*)^{\mathbb{N}} \text{ such that} \\ [c_k;c_{k+1},\ldots,] + [0;c_{k-1},c_{k-2},\ldots] \leq t, \forall k\in\mathbb{Z}\}. \end{aligned}$$

In fact, he showed [10, Lemma 2] that

$$D(t) = \overline{\dim}_{\mathrm{B}}(K_t) = \dim_{\mathrm{H}}(K_t) \tag{1.1}$$

and

$$\overline{\dim}_{\mathrm{B}}(K_t+K_t) = \dim_{\mathrm{H}}(K_t+K_t) = \min\{1, 2\cdot\dim_{\mathrm{H}}(K_t)\}, \tag{1.2}$$

where $\overline{\dim}_{\mathrm{B}}$ denotes the upper box dimension. Indeed, that lemma states that, given any $\eta>0$, there is a Gauss–Cantor set $K(B)\subseteq K_t$ such that

$$\dim_{\mathrm{H}}(K(B)) > (1-\eta)\overline{\dim}_{\mathrm{B}}(K_t),$$

so

$$(1-\eta)\overline{\dim}_{\mathrm{B}}(K_t) \leq \dim_{\mathrm{H}}(K(B)) \leq \dim_{\mathrm{H}}(K_t) \leq \overline{\dim}_{\mathrm{B}}(K_t).$$

Letting $\eta\to 0$ shows (1.1), while (1.2) follows from the fact that

$$\mathcal{M}\cap(-\infty,t) \subseteq (\mathbb{N}^*\cap[1,t]) + K_t + K_t$$

and the inequalities

$$\begin{aligned} d(t) &= \dim_{\mathrm{H}}(\mathcal{M}\cap(-\infty,t)) \\ &\leq \dim_{\mathrm{H}}(K_t+K_t) \leq \overline{\dim}_{\mathrm{B}}(K_t+K_t) \leq 2\cdot\overline{\dim}_{\mathrm{B}}(K_t) = 2\cdot\dim_{\mathrm{H}}(K_t). \end{aligned}$$

1.4. The Hausdorff dimension near 3

The goal of this article is to determine the behavior of $d(t)$ near $t = 3$. By work of the third author [10], we know that $d(t) > 0$ for every $t > 3$. By contrast, $d(t) = 0$ for every $t \leq 3$, as $\mathcal{L} \cap (-\infty, 3] = \mathcal{M} \cap (-\infty, 3]$ is countable.

Our main objective is to determine the modulus of continuity of $d(t)$ near 3. The first result we obtained in this direction was the following:

There exist constants $C_1, C_2 > 0$ such that, for any sufficiently small $\varepsilon > 0$, one has

$$C_1 \frac{\log|\log \varepsilon|}{|\log \varepsilon|} \leq d(3+\varepsilon) \leq C_2 \frac{\log|\log \varepsilon|}{|\log \varepsilon|}. \tag{1.3}$$

Let us explain how this partial result is obtained. Our methods are mainly combinatorial and the proofs of the upper and lower bounds on $d(t)$ are done in separate sections.

To establish the upper bound, we extend some results in Bombieri's article [1] to (factors of) sequences with Markov value slightly larger than 3. In this way, we can analyze the sequences $\omega \in \{1,2\}^{\mathbb{Z}} \subseteq (\mathbb{N}^*)^{\mathbb{Z}}$ that produce such Markov values; we show that they are not that different from those with Markov value less than or equal to 3.

To make this more precise, let $\Sigma(t) = \{\omega \in (\mathbb{N}^*)^{\mathbb{Z}} \mid \sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega)) \leq t\}$. We define $\Sigma(t,n)$ to be the set of length-n subwords of sequences in $\Sigma(t)$. We have the following:

Theorem 1.1. *There exists a constant $B > 1$ such that*

$$\Sigma(3+B^{-n}, n) = \Sigma(3,n) = \Sigma(3-B^{-n}, n)$$

for every sufficiently large integer n. In fact, we can take $B = 6^3 = 216$ and $n \geq 68$.

This theorem can be interpreted as follows: given a bi-infinite word whose Markov value is exponentially close to 3 (smaller than $3 + B^{-n} = 3 + 6^{-3n}$), its length-n subwords are indistinguishable from those in $\Sigma(3,n)$. That is, a length-n window cannot detect the patterns of symbols that make their Markov values different from 3; they are only present when considering windows of larger lengths.

Since the words before 3 are well understood, we will construct alphabets that allow us to write words in $\Sigma(3+B^{-n}, n)$ as *weakly renormalizable words* (see Definition 3.18). This construction is inspired by the "exponent-reducing" construction by Bombieri, which is detailed in Section 2.1. Indeed, the inductive procedure of reducing exponents can also be regarded as replacing the alphabet in which a word is written with a more complicated alphabet (so some exponents are "captured" by the letters of the new alphabet). The construction is inductive, so we will develop it as a *renormalization algorithm* (Lemma 3.21). This algorithm is used to obtain a proof of Theorem 1.1.

Theorem 1.1 allows us to reduce the proof of the upper bound to a simple counting. Indeed, we show in Corollary 3.13 that $|\Sigma(3,n)| = O(n^3)$, which implies that $|\Sigma(3+B^n, n)| = O(n^3)$. This is enough to establish the upper bound by covering K_t with small intervals in the standard way and using this counting.

To show that the lower bound holds, we prove that $d(3+e^{-r})$ (where $r \in \mathbb{N}^*$) is larger than the Hausdorff dimension of a suitable Gauss–Cantor set; recall that a Gauss–Cantor set is a subset of the real line defined by numbers with continued-fraction expansions that obey certain patterns. Finally, the Hausdorff dimension of a Gauss–Cantor set can be estimated by the (relatively elementary) methods in the book by Palis–Takens [11, Chapter 4], and hence the proof of (1.3) is complete.

While these methods are enough to prove inequalities (1.3), they are actually sufficient to obtain an asymptotic approximation of $d(t)$. In fact, to prove (1.3), only the results in Section 3 and (a simplification of the results) in Section 5 are needed.

We will now state our main results, which give more precise estimates of $d(t)$ for t close to 3. Let $f\colon[-1,+\infty) \to [-e^{-1},+\infty)$ be given by $f(x) = xe^x$ and recall that the Lambert W function is the function $W\colon[-e^{-1},+\infty) \to [-1,+\infty)$ given by $W = f^{-1}$. Our main result is the following:

Theorem 1.2. *Let $d(t) = \dim_{\mathrm{H}}(\mathcal{L} \cap [0,t)) = \dim_{\mathrm{H}}(\mathcal{M} \cap [0,t))$. Then, for all sufficiently small ε, we have*

$$d(3+\varepsilon) = 2 \cdot \frac{W(e^{c_0}|\log \varepsilon|)}{|\log \varepsilon|} + \mathrm{O}\left(\frac{\log|\log \varepsilon|}{|\log \varepsilon|^2}\right),$$

where $c_0 = -\log\log((3+\sqrt{5})/2) \approx 0.0383$.

The main idea behind the upper bound of Theorem 1.2 is again the construction of alphabets that allow us to write finite subwords of $\Sigma(3+e^{-r})$ as *weakly renormalizable words*. Then, using the fact that windows of sizes comparable to r must have a very similar structure to those before 3 (which are well understood because of the work of Bombieri [1]), we can find long forced continuations of finite subwords of size comparable to r of words of $\Sigma(3+e^{-r})$. Here, by *size* we no longer mean the length of a word, but rather the size of the interval it induces by continued-fraction expansions. Using the covering of K_t constructed with finite subwords of $\Sigma(3+e^{-r})$, we can control the size of a subcovering by smaller intervals (associated with longer words), depending on the structure of each word, so intervals with few continuations contribute less to the dimension. It turns out that there are some configurations which contribute more than others, namely configurations obtained by alternate concatenations of large blocks of 1's with blocks 22.

To be more precise, define the Gauss–Cantor set

$$C_n := K(\{221^n, 1\}) = \{[0;\gamma_1,\gamma_2,\dots] \mid \gamma_i \in \{221^n,1\},\ \forall i \geq 1\},$$

and let $\varepsilon_n := \max L(C_n)$, so ε_n is of the order of $((3+\sqrt{5})/2)^{-n}$. From Theorem 1.2 and from the proof of its lower bound (Section 5), we have

$$\begin{aligned} d(3+\varepsilon_n) &= 2 \cdot \frac{W(e^{c_0}|\log \varepsilon_n|)}{|\log \varepsilon_n|} + \mathrm{O}\left(\frac{\log|\log \varepsilon_n|}{|\log \varepsilon_n|^2}\right) = 2 \cdot \dim_{\mathrm{H}}(C_n) + \mathrm{O}\left(\frac{\log|\log \varepsilon_n|}{|\log \varepsilon_n|^2}\right) \\ &= \dim_{\mathrm{H}}(L(C_n)) + \mathrm{O}\left(\frac{\log|\log \varepsilon_n|}{|\log \varepsilon_n|^2}\right) = d(3+\varepsilon_{n-1}) + \mathrm{O}\left(\frac{\log|\log \varepsilon_n|}{|\log \varepsilon_n|^2}\right). \end{aligned}$$

One natural follow-up question is whether it is possible to find a better approximation of $d(t)$ near 3. The next theorem shows that this is not possible for "reasonable" (or explicit) approximations: for such approximations, the error term is optimal. We prove the following:

Theorem 1.3. *Let* $d(t) = \dim_{\mathrm{H}}(\mathcal{L} \cap [0,t)) = \dim_{\mathrm{H}}(\mathcal{M} \cap [0,t))$. *There exist sequences* $(x_k), (y_k)$ *and constants* $0 < C_1 < C_2$, *with* $0 < C_1\varphi^{-4k} = x_k < \frac{3}{2}x_k < y_k = C_2\varphi^{-4k}$, *where* $\varphi = (1+\sqrt{5})/2$ *is the golden mean, such that*

$$d(3+y_k) - d(3+x_k) = \mathrm{O}\left(\frac{1}{k^2}\right).$$

In particular, if F *is a twice continuously-differentiable function satisfying*

$$d(3+\varepsilon) = F(\varepsilon) + \mathrm{o}\left(\frac{\log|\log\varepsilon|}{|\log\varepsilon|^2}\right),$$

then its second derivative $F''(\varepsilon)$ *changes sign infinitely many times as* ε *approaches* 0.

In fact, we will prove that $d(3+y_k) - d(3+x_k) = \mathrm{O}(\frac{1}{k^2}) = \mathrm{o}(\frac{\log k}{k^2})$, while

$$\frac{W(e^{c_0}|\log y_k|)}{|\log y_k|} - \frac{W(e^{c_0}|\log x_k|)}{|\log x_k|} > \tilde{c}\,\frac{\log k}{k^2}$$

for a positive constant $\tilde{c}$, which implies that the error term in the approximation of $d(3+\varepsilon)$ by any reasonable function of ε is at least of the order of $\frac{\log|\log\varepsilon|}{|\log\varepsilon|^2}$. In this sense, $(3+x_k, 3+y_k)$ is an "almost plateau" for the dimension function $d(t)$ (the variation of $d(t)$ in these intervals is much smaller than the variation of its reasonable approximations). Indeed, we have proven that $d(3+\varepsilon)$ is very well approximated by

$$g_1(\varepsilon) = 2\cdot\frac{W(e^{c_0}|\log\varepsilon|)}{|\log\varepsilon|},$$

and that it is also asymptotic to the simpler function $g_2(\varepsilon) = 2\cdot\frac{\log|\log\varepsilon|}{|\log\varepsilon|}$. Moreover, given constants $0 < c_1 < c_2$, we have

$$g_j(c_2\varepsilon) - g_j(c_1\varepsilon) = (2\log(c_2/c_1) + \mathrm{o}(1))\frac{\log|\log\varepsilon|}{|\log\varepsilon|^2}$$

for $j \in \{1,2\}$, so reasonable functions $\tilde{g}(\varepsilon)$ which are asymptotic to g_1 and g_2 should satisfy $\tilde{g}(c_2\varepsilon) - \tilde{g}(c_1\varepsilon) \geq \log(c_2/c_1)\frac{\log|\log\varepsilon|}{|\log\varepsilon|^2}$ for $\varepsilon > 0$ small enough.

While the estimates in the third author's work [10] in principle would allow us to obtain some information regarding the modulus of continuity, those estimates are very far from being optimal (this is particularly true for the upper estimates). Thus, we rely on the methods described above instead of the general methods in the third author's previous work.

This article is organized as follows: Section 2 contains some preliminary notations and facts that we will use later on. By analyzing the combinatorics of finite words, we develop a renormalization algorithm which we use to prove Theorem 1.1 in Section 3. Using the understanding of finite subwords, we will find large forced extensions which, by a delicate analysis of the sizes and counting of them, will give us the upper bound of Theorem 1.2 in Section 4. In Section 5 we present the construction and analysis of a suitable Gauss–Cantor set, which allows us to establish the lower bound in Theorem 1.2, and thus to finish the proof of the main theorem. Finally, we study how the bad cuts produce gaps in their respective Markov values in Section 6, which allows us to prove the optimality of our approximation in Theorem 1.3.

2. Preliminaries

Our goal is to study the function

$$d(t) := \dim_{\mathrm{H}}(\mathcal{L} \cap (-\infty, t)) = \dim_{\mathrm{H}}(\mathcal{M} \cap (-\infty, t))$$

near $t = 3$. If a sequence $\omega \in (\mathbb{N}^*)^{\mathbb{Z}}$ contains 3, then $\lambda(\omega) > 3.52$, which is “much larger” than 3, so we can ignore such sequences. Thus, throughout the entire article, a *word* is made up of letters of the alphabet $\{1, 2\}$. Words can be either finite, infinite or bi-infinite. If w is a finite word, we denote its *length* by $|w|$, that is, the number of letters 1 or 2 that are needed to write w.

We will also consider *sections* of words, which consist of a word together with a choice of a *splitting point* marked with a vertical bar. A section of a bi-infinite word can be interpreted as a *shift* of the original word. We usually write sections as $\omega = P^*|Q$, where $P \in (\mathbb{N}^*)^{\mathbb{N}^*}$ and $Q \in (\mathbb{N}^*)^{\mathbb{N}}$ are infinite words, and $P^* \in (\mathbb{N}^*)^{-\mathbb{N}^*}$ denotes the *transpose* of P, that is, $P^*_{-k} = P_k$ for every $k \in \mathbb{N}^*$.

2.1. Words in $\Sigma(3)$

Bombieri [1] showed that bi-infinite words in $\Sigma(3)$ have to follow very special patterns (which is essentially a restatement of much older results by Markov [8], as stated in the book by Cusick–Flahive [3]). Indeed, Bombieri showed [1, Lemma 9] that ω is a word in the letters $a = 22$ and $b = 11$ (that is, the number of consecutive ones or twos is always even or infinite), and he also showed [1, Lemma 11] that if $\omega \in \Sigma(3)$, then ω has to be of one of four possible forms:

- *constant*, that is, $\omega = a^\infty$ or $\omega = b^\infty$;
- *degenerate*, that is, $\omega = b^\infty a b^\infty$ or $\omega = a^\infty b a^\infty$;
- *Type I*, that is, $\omega = \dots a b^{e_i} a b^{e_{i+1}} a \dots$ with every $e_i \geq 1$; or
- *Type II*, that is, $\omega = \dots b a^{e_i} b a^{e_{i+1}} b \dots$ with every $e_i \geq 1$.

The *exponents* $(e_i)_{i \in \mathbb{Z}}$ that appear in Type I and Type II elements of $\Sigma(3)$ also have to be of some special forms, but we will not use them explicitly.

Now, let U and V be the *Nielsen substitutions* given by

$$U\colon \begin{array}{ccc} a & \mapsto & ab, \\ b & \mapsto & b, \end{array} \qquad V\colon \begin{array}{ccc} a & \mapsto & a, \\ b & \mapsto & ab. \end{array}$$

This substitutions have inverses defined in the free group $\mathsf{F}\langle a, b\rangle$, given by

$$U^{-1}\colon \begin{array}{ccc} a & \mapsto & ab^{-1}, \\ b & \mapsto & b, \end{array} \qquad V^{-1}\colon \begin{array}{ccc} a & \mapsto & a, \\ b & \mapsto & a^{-1}b. \end{array}$$

Bombieri also proved [1, Lemma 14] that if $\omega \in \Sigma(3)$, then both $U(\omega)$ and $V(\omega)$ belong to $\Sigma(3)$. These words can be described explicitly. Indeed, if we write $\omega = \dots ab^{e_i}ab^{e_{i+1}}a\dots$ where $e_i \geq 0$ for each i, then

$$U(\omega) = \dots ab^{e_i+1}ab^{e_{i+1}+1}a\dots.$$

Similarly, if we write $\omega = \dots ba^{e_i}ba^{e_{i+1}}b\dots$ with each $e_i \geq 0$, then

$$V(\omega) = \dots ba^{e_i+1}ba^{e_{i+1}+1}b\dots.$$

Furthermore, if ω is of Type I, then $U^{-1}(\omega)$ is well-defined and belongs to $\Sigma(3)$. Similarly, if ω is of Type II, then $V^{-1}(\omega)$ is well-defined and belongs to $\Sigma(3)$. These can be described by

$$\begin{aligned} U^{-1}(\omega) &= \dots ab^{e_i-1}ab^{e_{i+1}-1}a\dots, \\ V^{-1}(\omega) &= \dots ba^{e_i-1}ba^{e_{i+1}-1}b\dots, \end{aligned}$$

where $e_i \geq 1$ for each $i \in \mathbb{Z}$ by definition.

We will now include a useful lemma which is implicit in Bombieri's work.

Lemma 2.1. *A nonempty finite word w belongs to $\Sigma(3, |w|)$ if and only if there exists $W \in \langle U, V\rangle$ such that w is a factor of $W(ab)$.*

Proof. We will first show that if w is a factor of $W(ab)$ for some W in $\langle U, V\rangle$, then it belongs to $\Sigma(3, |w|)$. This is shown by Bombieri [1, Theorem 15], as the word $\omega = \dots W(ab)W(ab)W(ab)\dots$ belongs to $\Sigma(3)$.

We will now show that if $w \in \Sigma(3, |w|)$, then it is a factor of $W(ab)$ for some $W \in \langle U, V\rangle$.

Let ω be a bi-infinite word in $\Sigma(3)$ containing w as a factor. We know that ω can only be constant, degenerate, of Type I or of Type II.

Assume first that ω is constant. Then w is a factor of a^k or b^k for some $k \geq 1$. Observe that $U^{k-1}(ab) = ab^k$ and $V^{k-1}(ab) = a^kb$, so the result follows in this case. Assume now that ω is degenerate. If w is constant, we reduce to the previous case. Otherwise, w is a factor of b^kab^k or a^kba^k for some $k \geq 0$. Since $U^kV(ab) = ab^kab^{k+1}$ and $V^kU(ab) = a^{k+1}ba^kb$, we also obtain the result in this case.

Finally, suppose that ω is of Type I or II. Hence, $U^{-1}(\omega)$ or $V^{-1}(\omega)$ is well-defined and belongs to $\Sigma(3)$. Recall that these automorphisms act by reducing all exponents by 1.

By iteratively applying the appropriate automorphism, U^{-1} or V^{-1}, we obtain a (possibly finite) sequence of bi-infinite words $\omega = \omega^{(1)}, \omega^{(2)}, \dots$. This process only stops if $\omega^{(n)}$ is constant or degenerate for some $n \in \mathbb{N}^*$. In the latter case, take $W' \in \langle U, V\rangle$ such that $\omega = W'(\omega^{(n)})$. By definition, there exists a factor θ of $\omega^{(n)}$ such that $W'(\theta)$ contains w. Since $\omega^{(n)}$ is constant or degenerate, we know that its factors satisfy the statement of the lemma. Thus, θ is contained in a word of the form $W''(ab)$ for some $W'' \in \langle U, V\rangle$. We find that w is then a factor of $W'W''(ab)$.

Finally, assume that the process never stops, so we obtain an infinite sequence $(\omega^{(k)})_{k\in\mathbb{N}^*}$ of bi-infinite words. Possibly by first making w longer so it can be written in the alphabet $\{a, b\}$, we can apply the same sequence of operations to the finite word w, that is, reduce its exponents by 1 in the same way so that the exponents of the bi-infinite words in the sequence $(\omega^{(k)})_{k\in\mathbb{N}^*}$ get reduced by 1. In this way, we obtain a sequence $w = w^{(1)}, w^{(2)}, \dots$ of (possibly empty) finite words. We claim that $w^{(n)}$ is constant and nonempty for some $n \geq 1$. Indeed, if $w^{(k)}$ is not constant for some $k \geq 1$, then $w^{(k+1)}$ is nonempty, since only the exponents of exactly one of the letters, a or b, are reduced by the operation of taking $w^{(k)}$ to $w^{(k+1)}$. Moreover, $|w^{(k+1)}| < |w^{(k)}|$, as some exponents are reduced. If $w^{(k+1)}$ is again not constant, we can continue the process inductively. Since w is finite, this process has to stop, so some word in the sequence must be constant. This completes the proof as we already know that constant words satisfy the statement of the lemma. ∎

2.2. *Constraints for words*

For a finite word $u = u_1 \dots u_n \in (\mathbb{N}^*)^n$, we define u^* as the transpose of u, that is, $u^* = u_n u_{n-1} \dots u_1 \in (\mathbb{N}^*)^n$. Moreover, we set

$$M_u = \begin{cases} 12, & |u| \text{ is even,} \\ 21, & |u| \text{ is odd,} \end{cases} \qquad m_u = \begin{cases} 21, & |u| \text{ is even,} \\ 12, & |u| \text{ is odd.} \end{cases}$$

Now, given a section $w = u^*|v$ of a finite word w, we define

$$\lambda^+(w) = [v m_u^\infty] + [0 u M_v^\infty], \quad \lambda^-(w) = [v M_u^\infty] + [0 u m_v^\infty].$$

These quantities are the largest and smallest values of λ that a section of a bi-infinite word containing w can attain, respectively. Thus, they induce restrictions on which finite words can be factors of bi-infinite words whose Markov values are known to be bounded in some way.

2.3. *Useful notation*

We will set some notation that will be used throughout the article; some of it is borrowed from [10].

For a finite word $\alpha \in (\mathbb{N}^*)^n$ written as $\alpha = c_1 \dots c_n$, we define its *size* by $\mathsf{s}(\alpha) := |I(\alpha)|$, where $I(\alpha)$ is the interval

$$I(\alpha) := \{x \in [0, 1] \mid x = [0; c_1, \dots, c_n, t],\ t \geq 1\} \cup \{[0, c_1, \dots, c_n]\}$$

consisting of the numbers in [0, 1] whose continued fractions start with α. The set $I(\alpha)$ is a closed interval in [0, 1].

If we take $p_0 = 0$, $q_0 = 1$, $p_1 = 1$, $q_1 = c_1$ and, for each integer $k \geq 0$, we take $p_{k+2} = c_{k+2}p_{k+1} + p_k$ and $q_{k+2} = c_{k+2}q_{k+1} + q_k$, then the endpoints of $I(\alpha)$ are $[0; c_1, \ldots, c_n] = p_n/q_n$ and $[0; c_1, \ldots, c_{n-1}, c_n + 1] = \frac{p_n+p_{n-1}}{q_n+q_{n-1}}$. Thus,

$$\mathsf{s}(\alpha) = \left| \frac{p_n}{q_n} - \frac{p_n + p_{n-1}}{q_n + q_{n-1}} \right| = \frac{1}{q_n(q_n + q_{n-1})},$$

since $p_n q_{n-1} - p_{n-1} q_n = (-1)^{n-1}$. We define $\mathsf{r}(\alpha) = \lfloor \log(\mathsf{s}(\alpha)^{-1}) \rfloor$, which controls the order of magnitude of the size of $I(\alpha)$. Observe that $\mathsf{r}(\alpha) \leq r$ if and only if $\mathsf{s}(\alpha) > e^{-r-1}$.

We also define, for $r \in \mathbb{N}$, the set

$$Q_r = \{\alpha = c_1 \ldots c_n \mid \mathsf{r}(\alpha) \geq r, \, \mathsf{r}(c_1 \ldots c_{n-1}) < r\}.$$

Observe that $\alpha \in Q_r$ if and only if $\mathsf{s}(\alpha) \leq e^{-r}$ and $\mathsf{s}(\alpha') > e^{-r}$, where α' is the word obtained by removing the last letter from α. Informally, this means that the interval $I(\alpha)$ is "small", while the interval $I(\alpha')$ is "not so small", so the last letter cannot be removed from α without changing the order of magnitude of $|I(\alpha)|$.

Let us recall some estimates from [10] that will be useful for us. Indeed, for any finite words α, β, we have

$$\tfrac{1}{2}\mathsf{s}(\alpha)\mathsf{s}(\beta) < \mathsf{s}(\alpha\beta) < 2\mathsf{s}(\alpha)\mathsf{s}(\beta);$$

it follows that $\mathsf{r}(\alpha) + \mathsf{r}(\beta) - 1 \leq \mathsf{r}(\alpha\beta) \leq \mathsf{r}(\alpha) + \mathsf{r}(\beta) + 2$ [10, Lemma A.2]. By Euler's property of continuants (Lemma A.1), if $\alpha = c_1 \ldots c_m$ and $\beta = d_1 \ldots d_n$ are finite words, then

$$q_{m+n}(\alpha\beta) = q_m(\alpha)q_n(\beta) + q_{m-1}(c_1 \ldots c_{m-1})q_{n-1}(d_2 \ldots d_n),$$

and thus

$$q_m(\alpha)q_n(\beta) < q_{m+n}(\alpha\beta) < 2q_m(\alpha)q_n(\beta).$$

Finally, recall that $\Sigma(t) = \{\omega \in (\mathbb{N}^*)^{\mathbb{Z}} \mid \sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega)) \leq t\}$ and that $\Sigma(t, n)$ is the set of length-n subwords of sequences in $\Sigma(t)$. In this context, we define $\Sigma^{(r)}(3 + \delta)$ as the set of words $w \in Q_r$ belonging to $\Sigma(3 + \delta, |w|)$.

3. Weakly renormalizable words

The main goal of this section is to prove Theorem 1.1. For this, we will prove several lemmas that allow us to understand the structure of $\Sigma(3, n)$.

3.1. *Basic facts about λ*

We start by showing some basic facts about the function λ that will be useful throughout the article.

Lemma 3.1. *Let $\omega \in \Sigma(3.06)$. Then ω does not contain 121 or 212 as subwords.*

Proof. Observe that $\lambda^-(1|21) > 3.15$, so the word 121 does not appear in ω. Now, if 212 is a subword of ω, so is 2212. This is not possible since $\lambda^-(2|212) > 3.06$. ∎

Lemma 3.2. *Let ω be a bi-infinite word in 1 and 2 not containing 121 and 212 and such that $\omega = R^* w^* b|awS$, where w is a finite word, $R = R_1 R_2 \dots$, $S = S_1 S_2 \dots$ and $R_1 \neq S_1$, with $R_i, S_i \in \{1, 2\}$ for each i. Then*

$$\mathsf{s}(bwb) < \operatorname{sign}([w, S] - [w, R])(\lambda(\omega) - 3) < \mathsf{s}(bw1).$$

In particular, if w has even length, $R_1 = 1$ and $S_1 = 2$, then

$$\mathsf{s}(bwb) < \lambda(\omega) - 3 < \mathsf{s}(bw1).$$

Proof. First observe that $[2; 2, w, R] + [0; 1, 1, w, R] = 3$. Thus,

$$\begin{aligned}\lambda(R^* w^* 11|22wS) &= [2; 2, w, S] + [0; 1, 1, w, R] \\ &= 3 + [0; 1, 1, w, R] - [0; 1, 1, w, S].\end{aligned}$$

We obtain

$$\begin{aligned}\lambda(R^* w^* 11|22wS) - 3 &= [0; 1, 1, w, R] - [0; 1, 1, w, S] \\ &= \operatorname{sign}([w, S] - [w, R]) \cdot |[0; 1, 1, w, R] - [0; 1, 1, w, S]|.\end{aligned}$$

Let $x = [0; 1, 1, w, R]$ and $y = [0; 1, 1, w, S]$. We will write the continued-fraction expansion of these numbers as

$$\begin{aligned}x &= [0; u_1, \dots, u_\ell, u_{\ell+1}, u_{\ell+2}, \dots], \\ y &= [0; u_1, \dots, u_\ell, v_{\ell+1}, v_{\ell+2}, \dots],\end{aligned}$$

where $u_1 = u_2 = 1$ and $u_{\ell+1} \neq v_{\ell+1}$. With this notation, we have

$$\operatorname{sign}([w, S] - [w, R])(\lambda(\omega) - 3) = |x - y|.$$

Let $(p_n/q_n)_{n\in\mathbb{N}}$ be the sequence of convergents of x. More explicitly, $p_n/q_n = [0; u_1, \dots, u_n]$.

If we put $\alpha_{\ell+1} = [u_{\ell+1}; u_{\ell+2}, u_{\ell+3}, \dots]$, then

$$x = [0; u_1, \dots, u_\ell, \alpha_{\ell+1}] = \frac{\alpha_{\ell+1} p_\ell + p_{\ell-1}}{\alpha_{\ell+1} q_\ell + q_{\ell-1}}.$$

Similarly, let $\beta_{\ell+1} = [v_{\ell+1}; v_{\ell+2}, v_{\ell+3}, \dots]$. We then have

$$y = \frac{\beta_{\ell+1} p_\ell + p_{\ell-1}}{\beta_{\ell+1} q_\ell + q_{\ell-1}},$$

since the sequence of convergents of y coincides with $(p_n/q_n)_{n\in\mathbb{N}}$ up to $n=\ell$. Thus,

$$\begin{aligned} |x-y| &= \left|\frac{\alpha_{\ell+1}p_\ell + p_{\ell-1}}{\alpha_{\ell+1}q_\ell + q_{\ell-1}} - \frac{\beta_{\ell+1}p_\ell + p_{\ell-1}}{\beta_{\ell+1}q_\ell + q_{\ell-1}}\right| \\ &= \left|\frac{(\alpha_{\ell+1}-\beta_{\ell+1})(p_\ell q_{\ell-1} - p_{\ell-1}q_\ell)}{(\alpha_{\ell+1}q_\ell + q_{\ell-1})(\beta_{\ell+1}q_\ell + q_{\ell-1})}\right| \\ &= \left|\frac{(\alpha_{\ell+1}-\beta_{\ell+1})(-1)^{\ell-1}}{(\alpha_{\ell+1}q_\ell + q_{\ell-1})(\beta_{\ell+1}q_\ell + q_{\ell-1})}\right| \\ &= \frac{|\alpha_{\ell+1}-\beta_{\ell+1}|}{(\alpha_{\ell+1}q_\ell + q_{\ell-1})(\beta_{\ell+1}q_\ell + q_{\ell-1})}, \end{aligned} \tag{3.1}$$

where we have used $p_\ell q_{\ell-1} - p_{\ell-1}q_\ell = (-1)^{\ell+1}$.

Since we are only interested in continued fractions whose partial quotients are 1 or 2, we can assume, without loss of generality, that $u_{\ell+1} = 2$ and $v_{\ell+1} = 1$. We denote $\alpha = \alpha_{\ell+1}$, $\beta = \beta_{\ell+1}$ and $\lambda = q_{\ell-1}/q_\ell \in (0,1)$. Thus,

$$|x-y| = \frac{\alpha-\beta}{q_\ell^2(\alpha+\lambda)(\beta+\lambda)} = \frac{1}{q_\ell^2}\left(\frac{1}{\beta+\lambda} - \frac{1}{\alpha+\lambda}\right). \tag{3.2}$$

We see that $|x-y|$ is (for fixed $q_{\ell-1}$ and q_ℓ) an increasing function of α, and a decreasing function of β. By analyzing (3.1) and (3.2) we deduce the following:

- $|x-y|$ is minimized when α is minimized and β is maximized. This happens when

$$\alpha = \alpha_0 := [2;\overline{2,1,1,1,2,2}] = \frac{21+2\sqrt{210}}{21} \approx 2.3801,$$
$$\beta = \beta_0 := [1;\overline{1,2,2,2,1,1}] = \frac{6+\sqrt{210}}{12} \approx 1.7076.$$

- $|x-y|$ is maximized when α is maximized and β is minimized. This happens when

$$\alpha = \alpha_1 := [2;\overline{1,1,1,2,2,2}] = \frac{21+2\sqrt{210}}{19} \approx 2.6306,$$
$$\beta = \beta_1 := [1;\overline{2,2,2,1,1,1}] = \frac{12+2\sqrt{210}}{29} \approx 1.4132,$$

On the other hand, $bwb = u_1 \dots u_\ell 11$, so

$$\begin{aligned} \mathsf{s}(bwb) &= |[0;u_1,\dots,u_\ell,1,1] - [0;u_1,\dots,u_\ell,1,1,1]| \\ &= \left|\frac{2p_\ell + p_{\ell-1}}{2q_\ell + q_{\ell-1}} - \frac{3p_\ell + 2p_{\ell-1}}{3q_\ell + 2q_{\ell-1}}\right| \\ &= \frac{1}{(2q_\ell + q_{\ell-1})(3q_\ell + 2q_{\ell-1})} = \frac{1}{q_\ell^2}\frac{1}{(2+\lambda)(3+2\lambda)}. \end{aligned}$$

Similarly,

$$\begin{aligned}\mathsf{s}(bw1) &= |[0; u_1, \dots, u_\ell, 1] - [0; u_1, \dots, u_\ell, 1, 1]| \\ &= \left| \frac{p_\ell + p_{\ell-1}}{q_\ell + q_{\ell-1}} - \frac{2p_\ell + p_{\ell-1}}{2q_\ell + q_{\ell-1}} \right| \\ &= \frac{1}{(q_\ell + q_{\ell-1})(2q_\ell + q_{\ell-1})} = \frac{1}{q_\ell^2} \frac{1}{(1+\lambda)(2+\lambda)}.\end{aligned}$$

We then have

$$\begin{aligned}\frac{|x-y|}{\mathsf{s}(bwb)} &\geq (\alpha_0 - \beta_0) \frac{(2+\lambda)(3+2\lambda)}{(\alpha_0+\lambda)(\beta_0+\lambda)} \\ &\geq (\alpha_0 - \beta_0) \frac{(2+1/3)(3+2/3)}{(\alpha_0+1/3)(\beta_0+1/3)} \approx 1.03895 > 1,\end{aligned}$$

since the maps $f_1(\lambda) = \frac{2+\lambda}{\alpha_0+\lambda}$ and $f_2(\lambda) = \frac{3+2\lambda}{\beta_0+\lambda}$ are increasing and

$$\lambda = q_{\ell-1}/q_\ell = q_{\ell-1}/(u_\ell q_{\ell-1} + q_{\ell-2}) \geq q_{\ell-1}/(2q_{\ell-1} + q_{\ell-2}) \geq 1/3.$$

Analogously,

$$\begin{aligned}\frac{|x-y|}{\mathsf{s}(bw1)} &\leq (\alpha_1 - \beta_1) \frac{(1+\lambda)(2+\lambda)}{(\alpha_1+\lambda)(\beta_1+\lambda)} \\ &\leq (\alpha_1 - \beta_1) \frac{(1+1)(2+1)}{(\alpha_1+1)(\beta_1+1)} \approx 0.83374 < 1,\end{aligned}$$

since the maps $g_1(\lambda) = \frac{1+\lambda}{\beta_1+\lambda}$ and $g_2(\lambda) = \frac{2+\lambda}{\alpha_1+\lambda}$ are increasing and $\lambda \leq 1$. ∎

Remark 3.3. The Markov value of $\omega = R^*11|22S$ coincides with the Markov value of $\sigma(\omega)^* = S^*2|211R$ [1, Lemma 5].

It is not difficult to adapt the proof above to obtain a more explicit (but weaker) version of this lemma which depends only on the length of w:

Lemma 3.4. *Let ω be a bi-infinite word in 1 and 2 not containing 121 and 212 and such that $\omega = R^*11|22S$ with $R = R_1R_2\dots$ and $S = S_1S_2\dots$ and $R \neq S$. Let ℓ be the smallest nonnegative integer such that $R_\ell \neq S_\ell$. Then*

$$\frac{1}{7}(3-2\sqrt{2})^\ell < \operatorname{sign}([S]-[R])(\lambda(R^*11|22S)-3) < \frac{1}{7}\left(\frac{3-\sqrt{5}}{2}\right)^\ell.$$

In particular, if $w = w^*$ and $\ell = |w|$ is even, then

$$3 - \frac{1}{7}\left(\frac{3-\sqrt{5}}{2}\right)^{\ell+1} < \lambda((wba)^\infty wb|aw(baw)^\infty) < 3 - \frac{1}{7}(3-2\sqrt{2})^{\ell+1}.$$

We will usually use this lemma in the following way. Consider a finite word w in the alphabet $\{a, b\}$. Assume that ba is a factor of w. Then we write $w = u^*b|av$, where the vertical bar indicates a *cut*, that is, the position at which we compute the Markov value.

Now, let ℓ be the smallest nonnegative integer such that $u_\ell \neq v_\ell$ and assume that $u_\ell = b$ and $v_\ell = a$. In other words, w contains the factor $b\theta^* b|a\theta a$, where the vertical bar marks the same position as the cut in w. By the lemma, the Markov value of *any* infinite word in the alphabet $\{a, b\}$ containing w is at least $3 + \frac{1}{7}(3 - 2\sqrt{2})^{2\ell-1}$. Similarly, if w contains ab as a factor, then we can also write $w = u^* a|bv$. Assume now that the smallest nonnegative integer ℓ such that $u_\ell \neq v_\ell$ satisfies $u_\ell = a$ and $v_\ell = b$. Then the Markov value of *any* infinite word in the alphabet $\{a, b\}$ containing w is at least $3 + \frac{1}{7}(3 - 2\sqrt{2})^{2\ell-1}$.

In particular, if we assume that ω is an infinite word in the alphabet $\{a, b\}$ and that its Markov value is sufficiently small, then no finite factor w of ω can contain patterns as above. This ultimately allows us deduce that some letters are *forced* inside an infinite word containing a finite word.

For concreteness, we will demonstrate the usage of the previous lemma by showing that no bi-infinite word in $\Sigma(3.0007)$ contains the factor $w = bbab|aa$. Let ω be a bi-infinite word containing w. We start by considering the cut $bb|abaa$. By the lemma, if aa does not appear to the left of w in ω, then $\lambda(\omega) > 3 + \frac{1}{7}(3 - 2\sqrt{2})^3 > 3.0007$. Thus, we assume that ω contains $aabbabaa$ as a factor. We can now consider a second cut, $aa|bbabaa$. This cut shows that $\lambda(\omega) > 3 + \frac{1}{7}(3 - 2\sqrt{2})^1 > 3.0007$, which completes the example.

We now show that sequences of 1's or 2's of odd length are forbidden if we assume that the Markov value of a word is sufficiently close to 3 (relative to the size of the interval it defines).

Lemma 3.5. *Let $r \in \mathbb{N}$ with $r \geq 5$. Let $c, c' \in \{1, 2\}$ with $c \neq c'$. Let $w = c'c^n c'$ for some integer $n \geq 1$, and suppose that $w \in \Sigma(3 + e^{-r}, |w|)$. If $\mathsf{r}(c^n) \leq r - 4$ then n is even.*

Proof. Note that $w \neq 121$ and $w \neq 212$ by Lemma 3.1, so $n > 1$. Without loss of generality, we can assume that w is the shortest word of this form satisfying $w \in \Sigma(3 + e^{-r}, |w|)$. Let $\omega \in \Sigma(3 + e^{-r})$ be a bi-infinite word such that w is a factor of ω. Assume for a contradiction that $n = 2k + 1$. We will show that $\lambda(\omega) > 3 + e^{-r}$.

Suppose $c = 1$. We have a section $\omega = R^* 11|22S$ with

$$R = R_1 R_2 R_3 \ldots = 1^{2k-1} 2 \ldots, \qquad S = S_1 S_2 S_3 \ldots = 2^p 1^q 2^r \ldots.$$

By Lemma 3.2, $p > 0$ implies that $\lambda(\omega) > 3 + \mathsf{s}(bb) = 3 + \frac{1}{40}$, which contradicts the assumption on w. Thus, that $p = 0$. Let ℓ be the smallest positive integer such that $R_\ell \neq S_\ell$. We have two cases:

- If $q > 2k - 1$, then $\ell = 2k$. Since we are assuming that $n = 2k + 1$, we have $\ell < n$. Moreover, $[S] > [R]$ since $S_\ell < R_\ell$ and ℓ is even.
- If $q \leq 2k - 1$, then q is even as, otherwise, it would contradict the assumption on k. Thus, $q \leq 2k - 2$ and $\ell = q + 1 < n$. Hence, $[S] > [R]$ as $S_\ell > R_\ell$ and ℓ is odd.

In any case, by the assumption on n we deduce from Lemma 3.2 that

$$\lambda(\omega) > 3 + \mathsf{s}(111^{\ell-1}11) \geq 3 + \mathsf{s}(1^{n+3}) \geq 3 + \mathsf{s}(1^n)e^{-3} > 3 + e^{-r},$$

where the last inequality holds as $\mathsf{r}(1^n) \leq r - 4$.

Now suppose $c = 2$, so we have a section $\omega = R^* 11|22S$ with

$$R = R_1 R_2 R_3 \ldots = 1^p 2^q 1^r \ldots, \quad S = S_1 S_2 S_3 \ldots = 2^{2k-1} 1 \ldots.$$

If $p > 0$, Lemma 3.2 shows that $\lambda(\omega) > 3 + \mathsf{s}(bb) = 3 + \frac{1}{40}$, so we must have $p = 0$. Let ℓ be the smallest positive integer such that $R_\ell \neq S_\ell$. We have two cases:

- If $q > 2k - 1$, then $\ell = 2k$. Since we are assuming that $n = 2k + 1$, we have $\ell < n$. Moreover, $[S] > [R]$ since $S_\ell < R_\ell$ and ℓ is even.
- If $q \leq 2k - 1$, then q is even as before. Thus, $q \leq 2k - 2$ and $\ell = q + 1 < n$. Hence, $[S] > [R]$ as $S_\ell > R_\ell$ and ℓ is odd.

In any case, by the assumption on n we deduce from Lemma 3.2 that

$$\lambda(\omega) > 3 + \mathsf{s}(112^{\ell-1}11) \geq 3 + \mathsf{s}(2^n)e^{-2} > 3 + e^{-r},$$

where the last inequality holds as $\mathsf{r}(2^n) \leq r - 3$. ∎

Whenever we want a version of some lemma that depends only on the length of a word instead of on the size of the interval that it defines (since we want to prove Theorem 1.1 which is stated in terms of lengths of words), we can either repeat the proof using Lemma 3.4 instead of Lemma 3.2, or directly compare r with the length using Lemma A.2. For example, we can show that sequences of 1's or 2's of odd length are forbidden:

Lemma 3.6. *Let n be so large that*

$$\frac{1}{6^n} < \frac{1}{7}(3 - 2\sqrt{2})^n;$$

for definiteness, we can take $n \geq 68$. Let $\omega \in \Sigma(3 + 6^{-n})$. Then ω does not contain $12^{2k+1}1$ or $21^{2k+1}2$ as subwords if $2k + 1 < n$.

3.2. Nielsen substitutions and sequences with Markov value close to 3

Recall the Nielsen substitutions

$$U\colon \begin{matrix} a & \mapsto & ab, \\ b & \mapsto & b, \end{matrix} \qquad V\colon \begin{matrix} a & \mapsto & a, \\ b & \mapsto & ab. \end{matrix}$$

Let T be the tree obtained by successive applications of U and V, starting at the root ab. Let P be the set of vertices of T and let P_n, for $n \geq 0$, be the set of elements of P whose distance to the root ab is exactly n. Recall from Lemma 2.1 that a finite word w belongs to $\Sigma(3, |w|)$ if and only if it is a factor of a word in P.

Given a pair (u, v) of words, we also define the operations $\bar{U}(u, v) = (uv, v)$ and $\bar{V}(u, v) = (u, uv)$. Let $\bar{T}$ be the tree obtained by successive applications of $\bar{U}$ and $\bar{V}$, starting at the root (a, b). Let $\bar{P}$ be the set of vertices of $\bar{T}$ and let $\bar{P}_n$, for $n \geq 0$, be the set of elements of $\bar{P}$ whose distance to (a, b) is exactly n.

Let c be the concatenation operator, that is, $\mathsf{c}(u, v) = uv$.

Lemma 3.7. *Let $(\alpha, \beta) \in \overline{P}$. Then there exists $W \in \langle U, V\rangle$ such that $\alpha = W(a)$ and $\beta = W(b)$. In particular, the sets $\mathsf{c}(\overline{P})$ and P are equal.*

Proof. We will prove a stronger equality: $\mathsf{c}(\overline{P}_n) = P_n$ for each $n \geq 0$. It is enough to show one inclusion as both sets have cardinality 2^n.

We proceed by induction. We claim that, for every $n \geq 0$ and $(u, v) \in \overline{P}_n$, there exists $W \in \langle U, V\rangle$ such that $u = W(a)$ and $v = W(b)$. The base case, for $n = 0$, is clear.

Now, let $(u, v) \in \overline{P}_{n-1}$ for $n \geq 1$. We will prove the claim for $(uv, v) \in \overline{P}_n$. Indeed, there exists $W \in \langle U, V\rangle$ such that $u = W(a)$ and $v = W(b)$. Observe that $WU(a) = W(ab) = W(a)W(b) = uv$ and $WU(b) = W(b) = v$. The proof for $(u, uv) \in \overline{P}_n$ is analogous. ∎

To state the following lemmas, we need to fix some useful notation. Let α and β be finite words and assume that α starts with a, while β ends with b. We write $\alpha = a\alpha^+$ and $\beta = \beta^- b$. Then we define $\alpha^b = b\alpha^+$ and $\beta_a = \beta^- a$. That is, α^b is obtained by replacing the first letter of α (which is a by assumption) with b, and similarly β_a is obtained by replacing the last letter of β (which is b by assumption) with a.

Lemma 3.8. *For every $(\alpha, \beta) \in \overline{P}$, α starts with a and β ends with b. Moreover, every word $\alpha^k\beta$ with $k \geq 1$ starts with β_a, and every word $\alpha\beta^k$ with $k \geq 1$ ends with α^b. In particular, every sufficiently large word in (α, β) starts with β^- and ends with α^+, and we always have $\alpha\beta = (\beta_a)(\alpha^b)$.*

Proof. For $(\alpha, \beta) = (a, b)$, it is clear that α starts with a, β ends with b, $\alpha^+ = \beta^- = \varnothing$, $\alpha^b = b$, $\beta_a = a$, $\alpha^k\beta = a^k b$ starts with $a = \beta_a$ for every $k \geq 1$, and $\alpha\beta^k = ab^k$ ends with $b = \alpha^b$ for every $k \geq 1$.

By induction, if $(A, B) = (\alpha, \alpha\beta)$ then $A = \alpha$ starts with a and $B = \alpha\beta$ ends with b. Since $B = \alpha\beta$ ends with $\alpha^b = A^b$, we see that, for every $k \geq 1$, AB^k also ends with A^b. Now, fix $k \geq 1$. By induction, $\alpha^k\beta$ starts with β_a, so $A^k B = \alpha^{k+1}\beta = \alpha\alpha^k\beta$ starts with $\alpha\beta_a = (\alpha\beta)_a = B_a$.

On the other hand, if $(A, B) = (\alpha\beta, \beta)$ then clearly A starts with a and B ends with b. Since $A = \alpha\beta$ starts with $\beta_a = B_a$, we find that, for every $k \geq 1$, $A^k B$ starts with B_a. Furthermore, since $\alpha\beta^k$ ends with α^b, $AB^k = \alpha\beta^{k+1} = \alpha\beta^k\beta$ ends with $\alpha^b\beta = (\alpha\beta)^b = A^b$ for every $k \geq 1$. The inductive argument is therefore complete.

Finally, the remaining equality $\alpha\beta = (\beta_a)(\alpha^b)$ follows immediately since $|\alpha\beta| = |(\beta_a)(\alpha^b)|$ (and, as we have just proved, $\alpha\beta$ starts with β_a and ends with α^b). ∎

Remark 3.9. Every word in P is of the form $a\theta b$ with θ palindromic, i.e., θ coincides with its transpose θ^*, as stated in Bombieri's article [1, proof of Theorem 15]. Since α starts with a and β ends with b, this is equivalent to $(\alpha\beta)^* = ((\alpha\beta)^b)_a$. In other words, both α^b and β_a are palindromic for every pair $(\alpha, \beta) \in \overline{P}$. We will now present an alternative proof of this fact.

As in the previous lemma, we will proceed by induction; the base case is clear. Suppose that α^b and β_a are palindromic. Then $(\alpha\beta)^b$ is palindromic, since both $(\alpha\beta)^b = \alpha^b\beta$ and $((\alpha\beta)^b)^* = (\alpha^b\beta)^* = \beta^*\alpha^b$ are obtained from $\alpha\beta = (\beta_a)(\alpha^b)$ by replacing the first letter (which is a) with b, and therefore coincide. Similarly, $(\alpha\beta)_a$ is also palindromic. Thus, the result holds for both $(\alpha\beta, \beta)$ and $(\alpha, \alpha\beta)$, which completes the inductive proof.

Lemma 3.10. *Suppose that a word w can be written as a concatenation $\tau\alpha\beta\tau'$ for some words τ, τ', α and β with $(\alpha, \beta) \in \overline{P}_n$ and $n \in \mathbb{N}$. If there exist $(A, B) \in \overline{P}_n$, $k \geq 1$ and $w_1, \ldots, w_k \in \{A, B\}$ such that $w = w_1 \ldots w_k$, then $(A, B) = (\alpha, \beta)$ and there exists $1 \leq j < k$ such that $w_1 \ldots w_{j-1} = \tau$, $w_j = \alpha$, $w_{j+1} = \beta$ and $w_{j+2} \ldots w_k = \tau'$.*

Proof. As usual, we proceed by induction. The result is trivial for the base case $(\alpha, \beta) = (a, b) \in \overline{P}_0$. Assume now that $(\alpha, \beta) = (uv, v)$ for $(u, v) \in \overline{P}_{n-1}$, where $n \geq 1$. Let w be a word such that $w = \tau\alpha\beta\tau'$ for some words τ, τ' and assume that there exist $(A, B) \in \overline{P}_n$, $k \geq 1$ and $w_1, \ldots, w_k \in \{A, B\}$ such that $w = w_1 \ldots w_k$.

Since $w = \tau\alpha\beta\tau'$ and $\alpha\beta = uvv$, there exist $\sigma = \tau$ and $\sigma' = v\tau'$ such that $w = \sigma uv\sigma'$. Thus, by induction, if w can be written as concatenation of words from a pair in $\overline{P}_{n-1}$, then the pair is necessarily (u, v) and the words u, v and v appear consecutively in this decomposition. This is indeed the case as each w_j for $1 \leq j \leq k$ is a concatenation of words from a pair in $\overline{P}_{n-1}$, so w can be written in this way as well.

We conclude that $(A, B) = \overline{U}(u, v) = (uv, v)$ or $(A, B) = \overline{V}(u, v) = (u, uv)$. Indeed, if this did not hold, then we would be able to find a different pair in $\overline{P}_{n-1}$ whose words can be concatenated to obtain w. Finally, if $(A, B) = \overline{V}(u, uv)$, then it would not be possible for the words u, v, v to appear consecutively. We conclude that $(A, B) = (\alpha, \beta)$.

The case where $(\alpha, \beta) = (u, uv)$ for $(u, v) \in \overline{P}_{n-1}$ is analogous. ∎

We can now relate the length of a factor of a word in P to the length of the smallest word in P containing it:

Lemma 3.11. *Let w be a factor of a word in P. Then the length of the shortest word in P containing w is strictly smaller than $3|w|$.*

Proof. Let $(\alpha, \beta) \in \overline{P}$ be such that $\alpha\beta$ contains w and $|\alpha\beta|$ is minimal for this property. We will assume $|\alpha| > |\beta|$ (the case $|\alpha| < |\beta|$ is analogous, and the case $|\alpha| = |\beta|$ only occurs in the trivial case $\alpha = a, \beta = b$, in which we may replace the constant 3 with 2). Hence, we may write $\alpha = \tilde{\alpha}\beta^r$ for some $r \geq 1$, where $(\tilde{\alpha}, \beta) \in \overline{P}$ and $|\tilde{\alpha}| \leq |\beta|$. We then have the bounds $(r + 1)|\beta| < |\alpha\beta| \leq (r + 2)|\beta|$.

Observe that w must intersect both α and β by minimality of $|\alpha\beta|$. Indeed, if w only intersects $\alpha = \tilde{\alpha}\beta^r$ or β, then the shorter word $\tilde{\alpha}\beta^{r-1}\beta$ corresponding to the pair $(\tilde{\alpha}\beta^{r-1}, \beta) \in \overline{P}$ contradicts the minimality of $|\alpha\beta|$. Now, if w intersects the prefix $\tilde{\alpha}$ of α, then it contains β^r strictly, and so the ratio $|w|/|\alpha\beta|$ is larger than $r/(r + 2) \geq 1/3$. Thus, from now on we may assume that w is contained in β^{r+1}. Moreover, r is minimal for this property, as otherwise the pair $(\tilde{\alpha}\beta^{r-1}, \beta) \in \overline{P}$ again contradicts the minimality of $|\alpha\beta|$. Thus, $w = u\beta^{r-1}v$, where u is a nonempty suffix of β and v is a nonempty prefix of β.

Assume that $r \geq 2$. By Lemma 3.8, we know that $\tilde{\alpha}\beta = (\beta_a)(\tilde{\alpha}^b)$. We now claim that $|u| \geq |\tilde{\alpha}^b|$. Indeed, assume that this is not the case. Then w is contained in $\alpha = \tilde{\alpha}\beta^r$, since any proper suffix of $\tilde{\alpha}^b$ is also a proper suffix of $\tilde{\alpha}$. This contradicts the minimality of $|\alpha\beta|$ as before. Thus, since $|\tilde{\alpha}^b| = |\tilde{\alpha}|$, the ratio $|w|/|\alpha\beta|$ is at least

$((r-1)|\beta|+|\tilde{\alpha}|)/((r+1)|\beta|+|\tilde{\alpha}|)$, which is larger than $(r-1)/(r+1)\geq 1/3$ since $r\geq 2$.

We will now address the remaining case where $r=1$. First observe that if $|\tilde{\alpha}|=|\beta|$, then $\tilde{\alpha}=a$ and $\beta=b$. Hence, $\alpha=ab$ and $w=b^2$. We then have $|w|/|\alpha\beta|=2/3$. We can therefore assume from now on that $|\beta|>|\tilde{\alpha}|$ and we may write $\beta=\tilde{\alpha}^j\tilde{\beta}$ for some $j\geq 1$, where $(\tilde{\alpha},\tilde{\beta})\in\bar{P}$ and $|\tilde{\beta}|\leq|\tilde{\alpha}|$.

We see that w is a factor of $\beta^2=\tilde{\alpha}^j\tilde{\beta}\tilde{\alpha}^j\tilde{\beta}$ and it intersects both copies of β. Hence, $w=uv$, where u is a nonempty suffix of $\beta=\tilde{\alpha}^j\tilde{\beta}$ and v is a nonempty prefix of $\beta=\tilde{\alpha}^j\tilde{\beta}=\tilde{\alpha}\tilde{\alpha}^{j-1}\tilde{\beta}$.

By Lemma 3.8, $\tilde{\alpha}^j\tilde{\beta}$ ends with $\tilde{\alpha}^b$. We claim that $\tilde{\alpha}^b$ is a suffix of u. Indeed, if this were not the case, then u would be a suffix of $\tilde{\alpha}$, and hence w would be contained in the shorter word $\tilde{\alpha}\tilde{\alpha}^j\tilde{\beta}$ corresponding to the pair $(\tilde{\alpha},\tilde{\alpha}^j\tilde{\beta})\in\bar{P}$, which is not possible by the minimality of $|\alpha\beta|$. Similarly, Lemma 3.8 implies that $\tilde{\alpha}^j\tilde{\beta}=\tilde{\alpha}\tilde{\alpha}^{j-1}\tilde{\beta}$ starts with $\tilde{\alpha}^{j-1}\tilde{\beta}_a$. We claim that $\tilde{\alpha}^{j-1}\tilde{\beta}_a$ is a prefix of v. Indeed, otherwise v would be a prefix of $\tilde{\alpha}^{j-1}\tilde{\beta}$, and hence w would be contained in the shorter word $\tilde{\alpha}^j\tilde{\beta}\tilde{\alpha}^{j-1}\tilde{\beta}$ corresponding to the pair $(\tilde{\alpha}^j\tilde{\beta},\tilde{\alpha}^{j-1}\tilde{\beta})\in\bar{P}$, which is not possible by the minimality of $|\alpha\beta|$. Finally, we conclude from $|\tilde{\alpha}^b|=|\tilde{\alpha}|$ and $|\tilde{\beta}_a|=|\tilde{\beta}|$ that the ratio $|w|/|\alpha\beta|$ is at least $(j|\tilde{\alpha}|+|\tilde{\beta}|)/((2j+1)|\tilde{\alpha}|+2|\tilde{\beta}|)$, which is larger than $j/(2j+1)\geq 1/3$. ■

Remark 3.12. The general bound in the previous lemma cannot be improved. Indeed, for the word $w=bab^{k+1}a$ for $k\geq 1$, we find that $|\alpha\beta|$ is minimal for the pair $(\alpha,\beta)=(ab^kab^{k+1},ab^{k+1})\in\bar{P}$. Since $|w|=2(k+4)$ and $|\alpha\beta|=2(3k+5)$, the ratio $|w|/|\alpha\beta|$ is arbitrarily close to $1/3$ when k is sufficiently large.

This example corresponds to the first case of the proof of the previous lemma, namely when w intersects the prefix $\tilde{\alpha}$ of $\alpha=\tilde{\alpha}\beta^r$. In the remaining two cases of the proof, nevertheless, the bound *can* be improved as we do below.

Assume then that w does not intersect $\tilde{\alpha}$. As in the previous proof, we first consider the case where $r\geq 2$. Then we may replace the constant 3 with $2+\varepsilon$ for any $\varepsilon>0$. Indeed, observe first that if $|\tilde{\alpha}|=|\beta|$, then $\tilde{\alpha}=a$ and $\beta=b$, so $\alpha=ab^r$ and $w=b^{r+1}$. Thus, the ratio $|w|/|\alpha\beta|$ is $(r+1)/(r+2)\geq 3/4\geq 1/2$. Otherwise, if $|\beta|>|\tilde{\alpha}|$, we write $\beta=\tilde{\alpha}^j\tilde{\beta}$ for $j\geq 1$ and $(\tilde{\alpha},\tilde{\beta})\in\bar{P}$. We have $w=uv$, where u is a suffix of $\beta^r=\tilde{\alpha}^j\tilde{\beta}\beta^{r-1}$ and v is prefix of $\beta=\tilde{\alpha}^j\tilde{\beta}$. By Lemma 3.8, $\tilde{\alpha}^j\tilde{\beta}$ ends with $\tilde{\alpha}^b$ and we claim that $\tilde{\alpha}^b\beta^{r-1}$ is a suffix of u. Indeed, otherwise u would be a suffix of $\tilde{\alpha}\beta^{r-1}$, so w would be contained in the shorter word $\alpha=\tilde{\alpha}\beta^r$ corresponding to the pair $(\tilde{\alpha}\beta^{r-1},\beta)\in\bar{P}$. Similarly, if we put $\hat{\beta}=\tilde{\alpha}^{j-1}\tilde{\beta}$ we have $(\tilde{\alpha},\hat{\beta})=(\tilde{\alpha},\tilde{\alpha}^{j-1}\tilde{\beta})\in\bar{P}$, so Lemma 3.8 implies that $\tilde{\alpha}\hat{\beta}$ starts with $\hat{\beta}_a$. We claim that $\hat{\beta}_a$ is a prefix of v. Indeed, otherwise v is a prefix of $\hat{\beta}$, and thus w is contained in the shorter word $(\tilde{\alpha}\hat{\beta})^r\hat{\beta}$ corresponding to the pair $(\tilde{\alpha}\hat{\beta},(\tilde{\alpha}\hat{\beta})^{r-1}\hat{\beta})\in\bar{P}$, a contradiction. Therefore, the ratio $|w|/|\alpha\beta|$ is at least $((r-1)|\beta|+|\tilde{\alpha}|+|\hat{\beta}|)/((r+1)|\beta|+|\tilde{\alpha}|)=r|\beta|/((r+1)|\beta|+|\tilde{\alpha}|)$, which is larger than $r/(r+2)\geq 1/2$.

Finally, we analyze the case where $r=1$ and show that we can replace the constant 3 with $5/2+\varepsilon$ for any $\varepsilon>0$. Recall that $\beta=\tilde{\alpha}^j\tilde{\beta}$, so the result is clear when $j\geq 2$ as $j/(2j+1)\geq 2/5$. Thus, we will assume that $j=1$, so $\alpha=\tilde{\alpha}\tilde{\alpha}\tilde{\beta}$ and $\beta=\tilde{\alpha}\tilde{\beta}$. If $j=1$

and $|\tilde{\alpha}| = |\tilde{\beta}|$, then $\tilde{\alpha} = a$, $\tilde{\beta} = b$, $\alpha = aab$ and $\beta = ab$. Since w intersects both α and β, we have $|w| \geq 4$, so we obtain $|w|/|\alpha\beta| \geq 2/5$ once again. We will then assume that $|\tilde{\beta}| < |\tilde{\alpha}|$. We see that w is a factor of $\tilde{\alpha}\tilde{\beta}\tilde{\alpha}\tilde{\beta}$, which is in turn a factor of the shorter word $\tilde{\alpha}\tilde{\beta}\tilde{\alpha}\tilde{\beta}\tilde{\beta}$ corresponding to the pair $(\tilde{\alpha}\tilde{\beta}, \tilde{\alpha}\tilde{\beta}\tilde{\beta}) \in \overline{P}$, a contradiction.

The lemma allows us to control the size of the set $\Sigma(3, n)$:

Corollary 3.13. *For all $n \geq 1$, we have $|\Sigma(3, n)| \leq 9n^3$.*

Proof. If $w \in \Sigma(3, n)$, then there is $(\alpha, \beta) \in \overline{P}$ with $|\alpha\beta| < 3n$ such that w is a factor of $\alpha\beta$ by Lemmas 2.1 and 3.11. Notice now that the pair $(\alpha, \beta) \in \overline{P}$ is determined by the irreducible fraction $|\alpha|/|\beta|$: indeed, $|\alpha| = |\beta|$ if and only if $\alpha = a$ and $\beta = b$; if $|\alpha| > |\beta|$, then $\alpha = \tilde{\alpha}\beta^k$ for some positive integer k with $(\tilde{\alpha}, \beta) \in \overline{P}$ and $|\tilde{\alpha}| \leq |\beta|$, and thus $|\alpha|/|\beta| = k + |\tilde{\alpha}|/|\beta|$; and if $|\alpha| < |\beta|$, then $\beta = \alpha^k\tilde{\beta}$ and $|\alpha|/|\beta| = 1/(k + |\tilde{\beta}|/|\alpha|)$. Hence, our claim follows by induction on the number of elements of the continued fraction of $|\alpha|/|\beta|$.

The number of such fractions $|\alpha|/|\beta|$ is bounded by the number of pairs (i, j) of positive numbers with $i + j \leq 3n$, which is $3n(3n - 1)/2 < 9n^2/2$. Since a word of size smaller than $3n$ has at most $2n$ factors of size n, there are at most $2n \cdot 9n^2/2 = 9n^3$ elements in $\Sigma(3, n)$. ∎

Recall that U and V are the Nielsen operators given by $U(a) = ab$, $U(b) = b$, $V(a) = a$ and $V(b) = ab$.

Lemma 3.14. *For any finite word w in the alphabet $\{a, b\}$, we have the identities*

$$bU(w^*) = U(w)^*b \quad \textit{and} \quad V(w^*)a = aV(w)^*.$$

In particular, if w is a palindrome, then so are $bU(w)$ and $V(w)a$.

Proof. This was already shown by Bombieri [1, proof of Theorem 15], but for completeness we include a short proof by induction.

The identities are trivial if $|w| = 0$. Assume that they hold for words of length $n - 1$ for $n \geq 1$, and let w be a word of that length. If $\tilde{w} = aw$, then

$$bU(\tilde{w}^*) = bU(w^*a) = bU(w^*)ab = U(w)^*bab = U(\tilde{w})^*b,$$
$$V(\tilde{w}^*)a = V(w^*a)a = V(w^*)aa = aV(w)^*a = aV(\tilde{w})^*.$$

On the other hand, if $\tilde{w} = bw$, then

$$bU(\tilde{w}^*) = bU(w^*b) = bU(w^*)b = U(w)^*bb = U(\tilde{w})^*b,$$
$$V(\tilde{w}^*)a = V(w^*b)a = V(w^*)aba = aV(w)^*ba = aV(\tilde{w})^*.$$

Assume now that w is a palindrome. Then

$$(bU(w))^* = U(w)^*b = bU(w^*) = bU(w),$$
$$(V(w)a)^* = aV(w)^* = V(w^*)a = V(w)a.$$ ∎

The following lemma shows that bi-infinite words with Markov value exponentially close to 3 (relative to the size of the interval they induce) cannot contain both $\alpha\alpha$ and $\beta\beta$ if $(\alpha, \beta) \in \bar{P}$. Recall that

$$\mathsf{r}(w) = \lfloor \log(\mathsf{s}(\alpha)^{-1}) \rfloor = \lfloor \log(|I(\alpha)|^{-1}) \rfloor.$$

Lemma 3.15. *Let $(\alpha, \beta) \in \bar{P}$. If w is a finite word in the alphabet $\{\alpha, \beta\}$ starting with $\alpha\alpha$ and ending with $\beta\beta$ such that $\mathsf{r}(w) \leq r$, then the Markov value of any bi-infinite word containing w as a factor is larger than $3 + e^{-r}$. Moreover, if w contains $\alpha\alpha\beta\beta$ as a factor and $\mathsf{r}(w) \leq 2r$, then the Markov value of any bi-infinite word containing w as a factor is larger than $3 + e^{-r}$.*

Proof. The proof is in several steps. The first step, labeled as "Step 0", is not strictly necessary; it is contained in the other more general steps. In this step we make an estimate depending on $|w|$, and it is weaker than the estimate of the statement, which depends on $\mathsf{r}(w)$. However, we have included it since it contributes to the understanding of the overall strategy.

Step 0: Assume that $\alpha = a$ and $\beta = b$. Without loss of generality, we can assume that $w = aa(ba)^k bb$, since otherwise there is a factor of w of this form and we may replace w by this factor. We will consider two cuts of this word. One cut, to which we will refer as the "first cut", is $aa(ba)^k|bb$, while the "second cut" is $aa(ba)^k bb|$. We start by applying Lemma 3.4 to the first cut. This immediately shows that $k \geq 1$, as otherwise any bi-infinite word containing w has a Markov value of at least $3 + \frac{1}{7}$ (in the general case this is not immediate; it is treated in Step 2). Hence, we assume that $k \geq 1$.

Let ω be a bi-infinite word containing w and assume towards a contradiction that its Markov value is smaller than $3 + \frac{1}{7}(3 - 2\sqrt{2})^{|w|}$. We continue drawing conclusions from Lemma 3.4: the first cut shows that ω must contain an a to the right of w. Thus, ω contains $w' = aa(ba)^k|bb|a$, where we have again marked both cuts. We now use these cuts to conclude inductively that w' must be followed by $(ba)^{k-1}$ in ω: each b is forced by the second cut (since there is a b at the symmetric position with respect to the second cut), and it is followed by an a by the first cut (since there is an a at the symmetric position with respect to the first cut).

Set $\gamma = (ba)^{k-1}$. Between both cuts, we have the word bb which we will write as $b\theta b$ with $\theta = \varnothing$ (in the general case, θ can be more complicated). On the left of the first cut, we have a word of the form $(\theta ba\gamma a)^* a$, while the second cut is followed by $a\gamma$. Thus, ω contains the word $w'' = (\theta ba\gamma a)^* a|b\theta b|a\gamma$, where we have again marked the first and second cuts.

The structure above is precisely the configuration that we will try to replicate in the general case, as it already leads to a large Markov value. Indeed, using the first cut again, we find that w'' is followed by an a in ω. Finally, w'' can also be written as $w'' = (\theta^* ba\theta ba\gamma a)^* b|a\gamma$ (where only the second cut is marked). Since $\theta^* ba\theta ba\gamma$ starts with γb, we deduce that w'' is followed by a b inside ω, which contradicts the fact that it is followed by an a as established before. In other words, we have shown that any bi-infinite word containing w has a Markov value of at least $3 + \frac{1}{7}(3 - 2\sqrt{2})^{|\gamma|+1} > 3 + \frac{1}{7}(3 - 2\sqrt{2})^{|w|}$.

Step 1: We now start treating the general case, so assume that w starts with $\alpha\alpha$ and ends with $\beta\beta$. Since $(\alpha, \beta) \in \bar{P}$, Lemma 3.7 shows that there exists some $W \in \langle U, V\rangle$ such that $\alpha = W(a)$ and $\beta = W(b)$. Thus, w is the image by W of a word in the alphabet $\{a, b\}$ starting with aa and ending with bb. Without loss of generality, we assume that $w = \alpha\alpha(\beta\alpha)^k\beta\beta$ with $k \geq 0$, as otherwise w contains a factor of this form and we may replace w with this factor.

Step 2: In this step, we assume that $k = 0$, so $w = \alpha\alpha\beta\beta$. We claim that w contains a cut of the form $\tau a|b\theta b$, where τ^* starts with θa and θ is a palindromic word. This leads to a contradiction by Lemma 3.2, as then the Markov value of any bi-infinite word containing w is at least $3 + \mathsf{s}(b\theta b)$, which is larger than $3 + e^{-r}$ by the following computation.

By hypothesis, we have $\mathsf{s}(w) \geq e^{-2r-1}$, so $\mathsf{s}(\alpha\alpha\beta\beta) \geq e^{-2r-1}$. Write $\theta = \theta_1 \dots \theta_n$. By (A.1) we have $\mathsf{s}(a\theta a)^{-1} \geq 961 q_n(\theta)^2$ and $\mathsf{s}(b\theta b)^{-1} \leq 162 q_n(\theta)^2$. Therefore,

$$e^{-2r-1} \leq \mathsf{s}(\alpha\alpha\beta\beta) \leq \mathsf{s}(a\theta ab\theta b) \leq 2\mathsf{s}(a\theta a)\mathsf{s}(b\theta b) \leq \frac{324}{961}\mathsf{s}(b\theta b)^2,$$

hence $\mathsf{s}(b\theta b) \geq e^{-r}$.

We proceed by induction: in the base case, we have $\theta = \varnothing$ and $\tau = a$. Now, observe that

$$U(\tau ab\theta b) = U(\tau)a|bbU(\theta)b = \tilde{\tau}a|b\tilde{\theta}b,$$

where $\tilde{\tau} = U(\tau)$ and $\tilde{\theta} = bU(\theta)$, and we have adjusted the position of the cut. We claim that $\tilde{\tau}^* = U(\tau)^*$ starts with $bU(\theta)a = \tilde{\theta}a$. Indeed, since τ^* starts with θa, we see that τ ends with $a\theta^*$. Thus, $U(\tau)$ ends with $abU(\theta^*)$. Therefore, $U(\tau)^*$ starts with $U(\theta^*)^*ba$, which is equal, by Lemma 3.14, to $bU(\theta)a = \tilde{\theta}a$ (since θ is a palindrome).

On the other hand, observe that

$$V(\tau ab\theta b) = V(\tau)aa|bV(\theta)ab = \tilde{\tau}a|b\tilde{\theta}b,$$

where $\tilde{\tau} = V(\tau)a$ and $\tilde{\theta} = V(\theta)a$, and we have adjusted the position of the cut. We claim that $\tilde{\tau}^* = aV(\tau)^*$ starts with $V(\theta)aa = \tilde{\theta}a$. Indeed, first observe that, by Lemma 3.14, $\tilde{\tau}^* = V(\tau^*)a$. Now, we consider two cases. If $\tau^* = \theta a$, then $\tilde{\tau}^* = V(\theta a)a = V(\theta)aa = \tilde{\theta}a$. Otherwise, τ^* starts with θac where $c \in \{a, b\}$, so $\tilde{\tau}^*$ starts with $V(\theta ac) = V(\theta)aV(c)$. Since $V(c)$ starts with a whether $c = a$ or $c = b$, we conclude that $\tilde{\tau}^*$ starts with $V(\theta)aa = \tilde{\theta}a$.

Since, by Step 1, there exists $W \in \langle U, V\rangle$ such that $W(a) = \alpha$ and $W(b) = \beta$, this concludes the proof when $k = 0$.

Step 3: In this step we leverage the structure found in Step 0 when $k \geq 1$ and show that it also leads to a large Markov value in a more general context. Assume now that we have a word w with two cuts of the form $w = \tau a|b\theta b|$ such that

(1) there exists a word γ such that τ ends with $(\theta ba\gamma a)^*$;

(2) $\theta^*ba\theta ba\gamma$ starts with γb.

We have shown that (1) and (2) hold for the base case $w = aa(ba)^k bb$ with $\tau = aa(ba)^{k-1}b$ and $\theta = \varnothing$.

Then, as before, the Markov value of any bi-infinite word ω containing w is at least $3 + e^{-r}$. To see this, we will again use the fact that, by Lemma 3.2, some of the letters surrounding w are forced in ω for the Markov value to remain below this value; eventually this will not be possible anymore. Indeed, an a is forced after w by the first cut, since τ ends with $(\theta ba)^*$. Moreover, Lemma 3.5 shows that the configuration $\tau a|b\theta b|a$ is followed by γ: each a of γ is forced by the first cut (since τ ends with $(\theta ba\gamma)^*$), while each b of γ is forced by the second cut (since $\theta^* ba\theta ba\gamma$ starts with γb). Finally, the first cut forces an a after $\tau a|b\theta b|a\gamma$ (since τ ends with $(\theta ba\gamma a)^*$), while, on the contrary, the second cut forces a b after $\tau a|b\theta b|a\gamma$ (since $\theta^* ba\theta ba\gamma$ starts with γb). Thus, the Markov value of any bi-infinite word containing w is at least $3 + e^{-r}$.

To be more precise about this last part, observe that γ cannot be followed by 12 or 21, because otherwise we will find a sequence of the form $c'c^s c'$ where $c, c' \in \{1, 2\}$ with $c \neq c'$ and s odd, but using Lemma A.4 and the fact that s is monotone and w ends with $(\theta ba\gamma a)^*(ab\theta b)$, we get

$$\mathsf{s}(c^s) \geq \mathsf{s}(a\gamma a) \geq 2^{-1}\mathsf{s}(a\gamma^* ab\theta^*) \geq 2^{-2}\mathsf{s}(abb)^{-1}\mathsf{s}(w) = (775/2)\mathsf{s}(w),$$

whence $\mathsf{r}(c^s) \leq r - 4$, contradicting Lemma 3.5. If γ is followed by b, then writing the first cut as $\omega^* = R^* b\eta^* b|a\eta a S$ with $\eta = \theta ba\gamma$ we have, by Lemma 3.2,

$$\lambda(\omega) = \lambda(\omega^*) \geq 3 + \mathsf{s}(b\eta b).$$

Since $a\eta^* ab\theta b$ is a subword of w, we see that $\mathsf{r}(a\eta^* ab\theta b) \leq r$ by Lemma A.3. In particular, $\mathsf{s}(a\eta^* ab\theta b) \geq e^{-r-1}$. On the other hand, by Lemma A.4, $\mathsf{s}(a\eta^* ab\theta b) \leq 4\mathsf{s}(a\eta a)\mathsf{s}(b\theta b) \leq \mathsf{s}(b\eta b)/3$, whence $\mathsf{s}(b\eta b) \geq e^{-r}$.

Similarly, if the word γ is followed by a, then, by writing the second cut as $\omega = R^* b\gamma^* b|a\gamma a S$, we have

$$\lambda(\omega) \geq 3 + \mathsf{s}(b\gamma b).$$

Finally, since γ is a subword of $\eta = \theta ba\gamma$, by Lemma A.3 again we get $\mathsf{s}(b\gamma b) \geq \mathsf{s}(b\eta b) \geq e^{-r}$.

Step 4: We now show inductively that the previous structure (namely properties (1) and (2)) persists when we apply U or V to $w = \tau a|b\theta b|$. First, observe that, after adjusting the position of the cuts, we have

$$U(w) = U(\tau)a|bbU(\theta)b| \quad \text{and} \quad V(w) = V(\tau)aa|bV(\theta)ab|. \tag{3.3}$$

Thus, $U(w) = \tilde{\tau} a|b\tilde{\theta} b|$ with $\tilde{\tau} = U(\tau)$ and $\tilde{\theta} = bU(\theta)$. Let $\tilde{\gamma} = bU(\gamma)$. Then, since w satisfies (1), $\tilde{\tau} = U(\tau)$ ends with

$$\begin{aligned} U((\theta ba\gamma a)^*) &= U(a\gamma^* ab\theta^*) = abU(\gamma^*)abbU(\theta^*) \\ &= aU(\gamma)^* babU(\theta)^* b = (\tilde{\theta} ba\tilde{\gamma} a)^*, \end{aligned}$$

where we have used Lemma 3.14. This shows that (1) holds for $U(w)$. Similarly, this lemma shows that

$$\begin{aligned}\tilde{\theta}^*ba\tilde{\theta}ba\tilde{\gamma} &= U(\theta)^*bbabU(\theta)babU(\gamma)\\ &= bU(\theta^*)babU(\theta)babU(\gamma)\\ &= bU(\theta^*ba\theta ba\gamma).\end{aligned}$$

This word starts with $bU(\gamma b) = bU(\gamma)b = \tilde{\gamma}b$, since $\theta^*ba\theta ba\gamma$ starts with γb, as w satisfies (2). Hence, (2) also holds for $U(w)$.

Now, from (3.3), we have $V(w) = \tilde{\tau}a|b\tilde{\theta}b|$ with $\tilde{\tau} = V(\tau)a$ and $\tilde{\theta} = V(\theta)a$. Let $\tilde{\gamma} = V(\gamma)a$. Then, since w satisfies (1), $\tilde{\tau} = V(\tau)a$ ends with

$$V((\theta ba\gamma a)^*)a = aV(\theta ba\gamma a)^* = a(V(\theta)abaV(\gamma)a)^* = (\tilde{\theta}ba\tilde{\gamma}a)^*,$$

where we have used Lemma 3.14. This shows that (1) holds for $V(w)$. Similarly, this lemma shows that

$$\begin{aligned}\tilde{\theta}^*ba\tilde{\theta}ba\tilde{\gamma} &= (V(\theta)a)^*baV(\theta)abaV(\gamma)a = aV(\theta)^*baV(\theta)abaV(\gamma)a\\ &= V(\theta^*)abaV(\theta)abaV(\gamma)a = V(\theta^*ba\theta ba\gamma)a.\end{aligned}$$

This word starts with $V(\gamma b) = V(\gamma)ab = \tilde{\gamma}b$, since $\theta^*ba\theta ba\gamma$ starts with γb, as (2) holds for w. Hence, (2) also holds for $V(w)$.

Since, by Step 1, there exists $W \in \langle U, V\rangle$ such that $W(a) = \alpha$ and $W(b) = \beta$, this concludes the proof when $k \geq 1$. ■

In order to consider other possible cases, such as words starting with $\beta\beta$ and ending with $\alpha\alpha$, we will show some symmetry properties of the pairs in $\overline{P}$.

Lemma 3.16. *Let $(u, v) \in \overline{P}$. If $(\alpha, \beta) = (u, uv)$, then $\alpha^k\beta = (u^b\alpha^k v_a)^*$. Similarly, if $(\alpha, \beta) = (uv, v)$, then $\alpha\beta^k = (u^b\beta^k v_a)^*$.*

Proof. Assume first that $(\alpha, \beta) = (u, uv)$. We have $(u, u^k v) \in \overline{P}$ for any $k \geq 1$. Now, recall that, by Lemma 3.8, $uu^k v = (u^k v)_a u^b = u^k v_a u^b$. Moreover, both u^b and $u^k v_a$ are palindromic by Remark 3.9. Thus,

$$\alpha^k\beta = uu^k v = u^k v_a u^b = (u^k v_a)^*(u^b)^* = (u^b u^k v_a)^* = (u^b\alpha^k v_a)^*.$$

Similarly, if $(\alpha, \beta) = (uv, v)$, we have $(uv^k, v) \in \overline{P}$ for any $k \geq 1$. Now, using Lemma 3.8 again, we obtain $uv^k v = v_a(uv^k)^b = v_a u^b v^k$, where both v_a and $u^b v^k$ are palindromic by Remark 3.9. Hence,

$$\alpha\beta^k = uv^k v = v_a u^b v^k = (v_a)^*(u^b v^k)^* = (u^b v^k v_a)^* = (u^b\beta^k v_a)^*. \qquad ■$$

Lemma 3.17. *Let $(u, v) \in \overline{P}$ and let $e_1, \dots, e_k \geq 1$. If $(\alpha, \beta) = (u, uv)$, then*

$$u^b\beta\alpha^{e_1}\beta\alpha^{e_2}\beta \dots \alpha^{e_k} v_a = (\alpha^{e_k}\beta\alpha^{e_{k-1}}\beta \dots \beta\alpha^{e_1}\beta\beta)^*,$$

while if $(\alpha, \beta) = (uv, v)$, then

$$u^b\beta^{e_1}\alpha\beta^{e_2}\alpha \dots \beta^{e_k}\alpha v_a = (\alpha\alpha\beta^{e_k}\alpha\beta^{e_{k-1}}\alpha \dots \alpha\beta^{e_1})^*.$$

Proof. Assume first that $(\alpha, \beta) = (u, uv)$. Then, by Lemmas 3.8 and 3.16,

$$\begin{aligned} u^b \beta \alpha^{e_1} \beta \alpha^{e_2} \beta \ldots \alpha^{e_k} v_a &= u^b (uv) \alpha^{e_1} (uv) \ldots (uv) \alpha^{e_k} v_a \\ &= (u^b)(v_a u^b) \alpha^{e_1} (v_a u^b) \ldots (v_a u^b) \alpha^{e_k} v_a \\ &= (u^b v_a)(u^b \alpha^{e_1} v_a) u^b \ldots v_a (u^b \alpha^{e_k} v_a) \\ &= (uv)^* (\alpha^{e_1} \beta)^* \ldots (\alpha^{e_k} \beta)^* \\ &= (\alpha^{e_k} \beta \alpha^{e_{k-1}} \beta \ldots \beta \alpha^{e_1} \beta \beta)^*. \end{aligned}$$

Now, take $(\alpha, \beta) = (uv, v)$. Then, by Lemmas 3.8 and 3.16,

$$\begin{aligned} u^b \beta^{e_1} \alpha \ldots \beta^{e_k} \alpha v_a &= (u^b) \beta^{e_1} (uv) \ldots \beta^{e_k} (uv) v_a \\ &= (u^b) \beta^{e_1} (v_a u^b) \ldots \beta^{e_k} (v_a u^b) v_a \\ &= (u^b \beta^{e_1} v_a) u^b \ldots \beta^{e_k} v_a (u^b v_a) \\ &= (\alpha \beta^{e_1})^* \ldots (\alpha \beta^{e_k})^* (uv)^* \\ &= (\alpha \alpha \beta^{e_k} \alpha \beta^{e_{k-1}} \ldots \alpha \beta^{e_1})^*. \end{aligned}$$ ■

The previous three lemmas imply that we obtain a large Markov value when $(\alpha, \beta) = (u, uv)$ for any word of the form $u^b \beta \ldots \alpha \alpha v_a$, and when $(\alpha, \beta) = (uv, v)$ for any word of the form $u^b \beta \beta \ldots \alpha v_a$.

We now define the notion of a weakly renormalizable word, which is central to our methods as it is used to find suitable alphabets in which words can be written.

Definition 3.18. Let $(\alpha, \beta) \in \overline{P}$ and $w \in \langle a, b \rangle$ be a finite word. We say that w is (α, β)-*weakly renormalizable* if we can write $w = w_1 \gamma w_2$ where γ is a word (called the *renormalization kernel*) in the alphabet $\{\alpha, \beta\}$ and w_1, w_2 are (possibly empty) finite words with $|w_1|, |w_2| < \max\{|\alpha|, |\beta|\}$ such that w_2 is a prefix of $\alpha\beta$ and w_1 is a suffix of $\alpha\beta$, with the following restrictions:

If $(\alpha, \beta) = (u, uv)$ for some $(u, v) \in \overline{P}$ and γ ends with α, then $|v| \le |w_2|$. If $(\alpha, \beta) = (uv, v)$ for some $(u, v) \in \overline{P}$ and γ starts with β, then $|u| \le |w_1|$.

Definition 3.19. Let $(\alpha, \beta) \in \overline{P}$ and $w \in \langle 1, 2 \rangle$ be a finite word. We say that w is (α, β)-*semirenormalizable* if there is an extension $\tilde{w}$ of at most two digits, one to the left and one to the right such that $\tilde{w}$ is (α, β)-*weakly renormalizable*.

The definition is motivated by the following ideas. Given an alphabet $\{\alpha, \beta\}$ with $(\alpha, \beta) \in \overline{P}$, it may not be possible to write a word w in terms of α and β. Nevertheless, it may very well be possible to write "most" of w in terms of α and β, preceded by and followed by some short trailing words. These words are w_1 and w_2 in the previous definition, and the condition ensuring that they are short is that $|w_1|, |w_2| < \max\{|\alpha|, |\beta|\}$. Indeed, if for example $|w_1| \ge \max\{|\alpha|, |\beta|\}$, then either w_1 ends with α or β in $\{\alpha, \beta\}$ (so our choice of renormalization kernel was spurious; it should be longer), or it does not (so w is actually not well described by the alphabet $\{\alpha, \beta\}$). To further ensure that w_1 and w_2 are well-adjusted to the chosen alphabet, we also require them to be a prefix or suffix of $\alpha\beta$; then w is contained in $\alpha\beta\gamma\alpha\beta$, where the renormalization kernel γ can be written in the alphabet $\{\alpha, \beta\}$.

Finally, we need to ensure that the first and last letters of the renormalization kernel are chosen appropriately. This follows from the following lemma (which is essentially already contained in Definition 3.18).

Lemma 3.20. *Let $(\alpha, \beta) \in \overline{P}$ and $w \in \langle a, b\rangle$ be an (α, β)-weakly renormalizable word. Write $w = w_1\gamma w_2$ as in Definition 3.18.*

If $(\alpha, \beta) = (u, uv)$ for some $(u, v) \in \overline{P}$ and γ ends with $\alpha = u$, then w_2 starts with $v_a \neq v$. Moreover, the word θ consisting of the last $|u|$ letters of γ followed by the first $|v|$ letters of w_2 is different from β.

Similarly, if $(\alpha, \beta) = (uv, v)$ for some $(u, v) \in \overline{P}$ and γ starts with $\beta = v$, then w_1 ends with $u^b \neq u$. Moreover, the word θ consisting of the last $|u|$ letters of w_1 followed by the first $|v|$ letters of γ is different from α.

Proof. Assume first that $(\alpha, \beta) = (u, uv)$ and that γ ends with $\alpha = u$. Then Definition 3.18 ensures that $|v| \leq |w_2|$. Since w_2 is a prefix of $\alpha\beta$ of length at least $|v|$, Lemma 3.16 implies that w_2 starts with $v_a \neq v$ (since v ends with b, and v_a is palindromic by Remark 3.9). Now, θ ends with the first $|v|$ letters of w_2, so it ends with $v_a \neq v$. Therefore, it cannot be equal to $\beta = uv$.

Similarly, if $(\alpha, \beta) = (uv, v)$ and γ starts with $\beta = v$, then Definition 3.18 ensures that $|u| \leq |w_1|$. Since w_1 is a suffix of $\alpha\beta$ of length at least $|u|$, Lemma 3.16 implies that w_1 ends with $u^b \neq u$ (since u starts with a, and u^b is palindromic by Remark 3.9). Now, θ starts with the last $|u|$ letters of w_1, so it starts with $u^b \neq u$. Therefore, it cannot be equal to $\alpha = uv$. ∎

The previous lemma can be understood as follows. Since the renormalization kernel γ is the part of $w = w_1\gamma w_2$ that can be written in the alphabet (α, β), it should be as long as possible (in the sense that w_1 and w_2 are just "short trailing words"). Hence, if $(\alpha, \beta) = (u, uv)$ and γ ends with $\alpha = u$, then the word w_2 should not start with v, since otherwise γ should instead end with $\beta = uv$ (and w_2 should be shorter). Similarly, if $(\alpha, \beta) = (uv, v)$ and γ starts with $\beta = v$, then the word w_1 should not start with u, since otherwise γ should instead start with $\alpha = uv$ (and w_1 should be shorter). All of these undesirable cases are ruled out by the lemma.

Exhibiting a word as being (α, β)-weakly renormalizable is nontrivial in general and, to complicate matters even further, the choice of alphabet $(\alpha, \beta) \in \overline{P}$ is not clear to begin with. Nevertheless, any word in the alphabet $\{a, b\}$ is trivially (a, b)-weakly renormalizable (by setting the renormalization kernel equal to the entire word).

On the other hand, there are subwords of words in $\langle a, b\rangle$ that can fail to be weakly renormalizable (for any alphabet) with nontrivial kernel, because they miss one digit at one (or both) of their ends. For example, the word $w = 21\ldots1$ of even length is a subword of $b^\infty ab^\infty$, and hence it belongs to $\Sigma(3, n)$. However, it can only be exhibited as an (α, β)-weakly renormalizable word by $w = w_1w_2$. That is why we introduce the notion of (α, β)-semirenormalizable in Definition 3.19. Indeed, the w above is (a, b)-semirenormalizable (with nontrivial kernel), since $2w1 \in \langle a, b\rangle$.

With these considerations, we will now present a renormalization algorithm: if we have a (u, v)-weakly renormalizable word with a nonempty renormalization kernel, we can exhibit this word as being (α, β)-weakly renormalizable for $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$ chosen appropriately.

Lemma 3.21 (Renormalization algorithm). *Let $w \in \Sigma(3 + e^{-r}, |w|)$ satisfy $\mathsf{r}(w) \leq r$. If w is (u, v)-weakly renormalizable as $w = w_1\gamma w_2$ with $\gamma \neq \varnothing$, then w is (α, β)-weakly renormalizable for some $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$. Moreover, if γ starts with u or ends with v, then w_1 or w_2, respectively, does not change for the renormalization with alphabet (α, β).*

Before proving the lemma, we will discuss the intuition behind this algorithm. The main inspiration is the "exponent-reducing" procedure discussed in Section 2.1. Indeed, if a word w is (u, v)-weakly renormalizable, then it is of the form $w = w_1\gamma w_2$, where γ is written in terms of u and v. The word γ cannot contain factors of the form $uu \dots vv$ or $vv \dots uu$ (as discussed in the proof below), so it is written as powers of u (respectively, v) followed by single instances of v (respectively, u). Hence, we can choose a new alphabet $(\alpha, \beta) = (u, uv)$ (respectively, $(\alpha, \beta) = (uv, v)$) so that all exponents are now reduced by 1 when γ is written in the alphabet (α, β). This simplifies the structure of the renormalization kernel at the cost of making the alphabet more complex. The renormalization algorithm should hence be applied inductively a certain number of times to ensure that the complexity of both the renormalization kernel and the alphabet remains reasonable (see for example Corollary 3.23 and the proof of Theorem 1.1 to see how this is used).

Proof of Lemma 3.21. We will explicitly exhibit w as being (α, β)-renormalizable as $w = \tilde{w}_1\tilde{\gamma}\tilde{w}_2$ for some $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$.

By Lemma 3.15 and the comments after Lemma 3.17, some patterns on a weakly renormalizable word imply that $w \notin \Sigma(3 + e^{-r}, |w|)$, and so are forbidden: this holds if γ contains both the factors uu and vv (in any order), and also in the following situations:

(1) $(u, v) = (\eta, \eta\theta)$ for some $(\eta, \theta) \in \bar{P}$, γ starts with v and contains the factor uu, and $|w_1| \geq |u|$.

(2) $(u, v) = (\eta\theta, \theta)$ for some $(\eta, \theta) \in \bar{P}$, γ ends with u and contains the factor vv, and $|w_2| \geq |v|$.

We first assume that w does not contain the factor vv and we analyze the following subcases (where s and e_j are positive integers for $1 \leq j \leq k$):

Case 1: If $\gamma = u^{e_1}vu^{e_2}v \dots u^{e_k}v$, we take $\alpha = u$, $\beta = uv$ and

$$\tilde{\gamma} = \alpha^{e_1-1}\beta\alpha^{e_2-1}\beta \dots \alpha^{e_k-1}\beta, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = w_2.$$

Indeed, $\tilde{w}_1 = w_1$ is a suffix of uv by hypothesis, so it is also a suffix of $\alpha\beta = u^2v$. Moreover, $\tilde{w}_2 = w_2$ is a prefix of uv by hypothesis and to show that it is also a prefix of $\alpha\beta$ we consider two cases. If $|w_2| < |v|$, then w_2 is a prefix of v_a, since uv starts with v_a by Lemma 3.8. The same lemma also shows that u^2v starts with v_a, so w_2 is a prefix of $\alpha\beta = u^2v$. Otherwise, we must have $|w_2| < |u|$, since $|w_2| < \max\{|u|, |v|\}$. Thus, $\tilde{w}_2 = w_2$ is a proper prefix of u, and hence of $\alpha\beta = u^2v$.

Case 2: If $\gamma = vu^{e_1}vu^{e_2}v \dots u^{e_k}v$ we consider two cases. If $|w_1| < |u|$, we take $\alpha = u$, $\beta = uv$ and

$$\tilde{\gamma} = \alpha^{e_1-1}\beta\alpha^{e_2-1}\beta \dots \alpha^{e_k-1}\beta, \quad \tilde{w}_1 = w_1v, \quad \tilde{w}_2 = w_2.$$

Indeed, recall that uv ends with u^b and w_1 is a suffix of uv. Since $|w_1| < |u|$, w_1 is also a suffix of u (as u^b and u are equal up to the first letter). Consequently, $\tilde{w}_1 = w_1v$ is a suffix of $\alpha\beta = u^2v$. Moreover, $\tilde{w}_2$ is a prefix of $\alpha\beta = u^2v$ by the same proof as in the previous case: it is either shorter than v (in which case it is a proper prefix of v_a, and hence of u^2v by Lemma 3.8), or shorter than u (in which case it is a prefix of u, and hence of u^2v).

Otherwise, we have $|u| \le |w_1| < |v|$, so $(u, v) = (\eta, \eta\theta)$ for some pair $(\eta, \theta) \in \overline{P}$. Since w_1 is a suffix of $uv = \eta^2\theta$ and $|w_1| \ge |u| = |\eta|$, we find that w_1 ends with η^b by Lemma 3.8. If $e_j > 1$ for some $1 \le j \le k$, then w contains a factor of the form $\eta^b v \dots uu\theta_a$. In fact, since w_1 ends with η^b, it follows that w contains a word of the form $w' = \eta^b v \dots u^{e_j-2}uuv$, where $1 \le j \le k$ is chosen so that $e_j > 1$. Moreover, $v = \eta\theta$ starts with θ_a by Lemma 3.8, so w' contains, in turn, a word of the form $\eta^b v \dots uu\theta_a$. This contradicts $w \in \Sigma(3 + e^{-r}, |w|)$ by Lemmas 3.15 to 3.17.

We next assume that $e_j = 1$ for every $1 \le j \le k$ and take $\alpha = uv$, $\beta = v$ and

$$\tilde{\gamma} = \beta\alpha^k, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = w_2.$$

Indeed, $\tilde{w}_2 = w_2$ is a prefix of uv^2 since it is a prefix of uv. Moreover, if $|w_1| < |v|$ then $\tilde{w}_1 = w_1$ is a suffix of uv^2 as it is a suffix of uv, and if $|w_1| < |u|$ then w_1 is a proper suffix of u^b (by Lemma 3.8), so it is also a suffix of uv^2 (by Lemma 3.8 again). Finally, since $\tilde{\gamma}$ starts with β and $(\alpha, \beta) = (uv, v)$, we have to check that $|u| \le |\tilde{w}_1| = |w_1|$, but this holds by hypothesis.

Case 3: If $\gamma = u^{e_1}vu^{e_2}v \dots u^{e_k}vu^s$, we must have $|v| \le |uw_2|$. Indeed, if $|v| > |uw_2|$, then $(u, v) = (\eta, \eta\theta)$ for some alphabet (η, θ). Since γ ends with u, by definition of (u, v)-weak renormalizability we have $|\theta| \le |w_2|$. Hence, $|v| = |\eta\theta| = |u\theta| \le |uw_2|$, a contradiction.

Let $r \in \{0, 1\}$ now be such that $|v| \le |u^r w_2| < |uv| \le |u^{r+1}w_2|$. Then we choose $\alpha = u$, $\beta = uv$ and

$$\tilde{\gamma} = \alpha^{e_1-1}\beta\alpha^{e_2-1}\beta \dots \alpha^{e_k-1}\beta\alpha^{s-r}, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = u^r w_2.$$

Indeed, $\tilde{w}_1 = w_1$ is a suffix of $\alpha\beta = u^2v$ since it is a suffix of uv. Now, if $r = 0$, then $|w_2| < |u|$ since $|w_2| < \max\{|u|, |v|\}$ and $|v| \le |w_2|$ by hypothesis. Since w_2 is a prefix of uv, it is actually a prefix of u, and hence of $\alpha\beta = u^2v$. If $r = 1$, we see that w_2 is a prefix of uv, and thus $\tilde{w}_2 = uw_2$ is a prefix of $\alpha\beta = u^2v$.

Since $\tilde{\gamma}$ ends with α if $r = 0$, we have to check that $|v| \leq |\tilde{w}_2|$. This holds since $|v| \leq |\tilde{w}_2| = |w_2|$ in this case.

Case 4: Finally, if $\gamma = vu^{e_1}vu^{e_2}v\dots u^{e_k}vu^s$, we combine the discussions of the previous two cases. More precisely, we assume first that $|u| \leq |w_1| < |v|$. If $e_j > 1$ for some $1 \leq j \leq k$ or $s > 1$, then we obtain a contradiction with the hypothesis that $w \in \Sigma(3 + e^{-r}, |w|)$ by Lemmas 3.15 to 3.17. Indeed, in this case $(u, v) = (\eta, \eta\theta)$ for some $(\eta, \theta) \in \bar{P}$, so w contains a factor of the form $\eta^b v\dots uu\theta_a$ as in the second case if $e_j > 1$ for some $1 \leq j \leq k$. On the other hand, if $s > 1$, then w contains a word of the form $w' = \eta^b v\dots u^{s-2}uuw_2$. Now, observe that the fact that γ ends with u and the definition of (u, v)-renormalizability imply that w_2 starts with θ_a. Hence, w' contains a word of the form $\eta^b v\dots uu\theta_a$. This leads to the same contradiction with Lemmas 3.15 to 3.17.

When $|u| \leq |w_1| < |v|$, $e_j = 1$ for every $1 \leq j \leq k$ and $s = 1$, we take $\alpha = uv$, $\beta = v$ and

$$\tilde{\gamma} = \beta\alpha^k, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = uw_2.$$

Since w_1 is a suffix of uv, it follows that $\tilde{w}_1 = w_1$ is a suffix of $\alpha\beta = u^2v$. Now, observe that $|u| \leq |w_1| < |v|$ implies that $|w_2| < |v|$, since $|w_1|, |w_2| < \max\{|u|, |v|\}$ by hypothesis. Thus, by Lemma 3.8, w_2 is a proper prefix of v_a, so it is also a prefix of v. We then conclude that $\tilde{w}_2 = uw_2$ is a prefix of $\alpha\beta = uv^2$.

Otherwise, if $|w_1| < |u|$ we take $\alpha = u$, $\beta = uv$ and argue as in the third case. More precisely, let $r \in \{0, 1\}$ be such that $|v| \leq |u^r w_2| < |uv| \leq |u^{r+1}w_2|$ and take

$$\tilde{\gamma} = \alpha^{e_1-1}\beta\alpha^{e_2-1}\beta\dots\alpha^{e_k-1}\beta\alpha^{s-r}, \quad \tilde{w}_1 = w_1v, \quad \tilde{w}_2 = u^r w_2.$$

We find that $\tilde{w}_2 = u^r w_2$ is a prefix of $\alpha\beta$ by the same arguments as in the third case, and r is chosen so $|v| \leq |\tilde{w}_2|$. Moreover, $\tilde{w}_1 = w_1v$ is a suffix of $\alpha\beta = u^2v$ since Lemma 3.8 and the fact that $|w_1| < |u|$ imply that w_1 is a proper suffix of u^b, so that it is also a suffix of u. This finishes the last subcase.

We now assume that w contains the factor vv, so in particular it does not contain the factor uu. We analyze the following subcases (where s and e_j are positive integers for $1 \leq j \leq k$):

Case 1: If $\gamma = uv^{e_1}uv^{e_2}\dots uv^{e_k}$, we take $\alpha = uv$, $\beta = v$ and

$$\tilde{\gamma} = \alpha\beta^{e_1-1}\alpha\beta^{e_2-1}\dots\alpha\beta^{e_k-1}, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = w_2.$$

Case 2: If $\gamma = uv^{e_1}uv^{e_2}\dots uv^{e_k}u$, we take $\alpha = uv$, $\beta = v$ and

$$\tilde{\gamma} = \alpha\beta^{e_1-1}\alpha\beta^{e_2-1}\dots\alpha\beta^{e_k-1}, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = uw_2.$$

Case 3: If $\gamma = v^s uv^{e_1}ue^{e_2}\dots uv^{e_k}$, we take $r \in \{0, 1\}$ such that $|u| \leq |w_1v^r| < |uv| \leq |w_1v^{r+1}|$, and define $\alpha = uv$, $\beta = v$ and

$$\tilde{\gamma} = \beta^{s-r}\alpha\beta^{e_1-1}\alpha\dots\beta^{e_{k-1}-1}\alpha\beta^{e_k-1}, \quad \tilde{w}_1 = w_1v^r, \quad \tilde{w}_2 = w_2.$$

Case 4: If $\gamma = v^s u v^{e_1} u e^{e_2} \dots u v^{e_k} u$, we take $r \in \{0, 1\}$ such that $|u| \leq |w_1 v^r| < |uv| \leq |w_1 v^{r+1}|$, and define $\alpha = uv$, $\beta = v$ and

$$\tilde{\gamma} = \beta^{s-r} \alpha \beta^{e_1 - 1} \alpha \dots \beta^{e_{k-1}-1} \alpha \beta^{e_k - 1}, \quad \tilde{w}_1 = w_1 v^r, \quad \tilde{w}_2 = u w_2.$$

Observe that the cases where $e_j = 1$ for all $1 \leq j \leq k$ cannot arise in the previous subcases, since we are explicitly assuming that w contains the factor vv. The arguments showing that these choices satisfy the definition of (α, β)-renormalizabilty are analogous to those of the previous cases (where the factor vv was not present). Thus, this concludes the proof. ∎

Once again, this lemma could be stated in terms of the length of w, as in the following corollary:

Corollary 3.22. *Let* $w \in \Sigma(3 + (3 + 2\sqrt{2})^{-(|w|+1)}, |w|)$ *be a finite word. If* w *is* (u, v)*-weakly renormalizable as* $w = w_1 \gamma w_2$ *with* $\gamma \neq \varnothing$*, then* w *is* (α, β)*-weakly renormalizable for some* $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$.

Proof. By Lemma A.2 we have $\mathsf{r}(w) \leq (n + 1) \log(3 + 2\sqrt{2}) =: r$ and applying Lemma 3.21 we obtain the result. ∎

We will now present a series of corollaries of the renormalization algorithm. We start with the version that is needed for the proof of Theorem 1.1.

Corollary 3.23. *Let* $n \geq 68$ *and let* $w \in \Sigma(3 + 6^{-3n}, 3n)$*. Then there exists an alphabet* $(\alpha, \beta) \in \overline{P}$ *satisfying* $|\alpha|, |\beta| < n$ *and* $|\alpha\beta| \geq n$ *such that* w *is* (α, β)*-semirenormalizable.*

Proof. Since $n \geq 68$, the conclusion of Lemma 3.6 holds, so, possibly up to adding one letter to the left and one to the right, w is a word in the alphabet $\{a, b\}$. As previously discussed, w is trivially (a, b)-weakly renormalizable with $w_1 = w_2 = \varnothing$ and $\gamma = w$. Observe that w satisfies the first hypothesis of Corollary 3.22. Indeed, this follows from the fact that $6^{-3n} \leq (3 + 2\sqrt{2})^{-(3n+3)}$ for every $n \geq 61$. By Corollary 3.22, we can apply the renormalization algorithm inductively as long as the renormalization kernel is nonempty; this produces a a finite sequence of alphabets. We will show that the sought-after alphabet (α, β) is the first alphabet in the sequence that satisfies $|\alpha\beta| \geq n$.

We will first show that such an alphabet exists. Assume that $(u, v) \in \overline{P}$ is one of the alphabets of the sequence. If $|uv| < n$, then $|w_1|, |w_2| < n$, since

$$|w_1|, |w_2| \leq \max\{|u|, |v|\} < |uv| < n,$$

where w_1 and w_2 are the words obtained in this step of the algorithm by the decomposition $w = w_1 \gamma w_2$. Hence, $\gamma \neq \varnothing$, since

$$|\gamma| = |w| - |w_1| - |w_2| > 3n - n - n = n.$$

Thus, we can apply the algorithm again if $|uv| < n$. Since the length of an alphabet increases with each inductive application of the algorithm, we will eventually find an alphabet $(\alpha, \beta) \in \overline{P}$ satisfying $|\alpha\beta| \geq n$. Assume that $(\alpha, \beta) \in \overline{P}$ is the first alphabet in the sequence satisfying this condition.

It remains to show that $|\alpha|, |\beta| < n$. Assume this is false. Assume further that $(\alpha, \beta) = (uv, v)$ for some alphabet $(u, v) \in \overline{P}$; the case where $(\alpha, \beta) = (u, uv)$ is similar. We then have $|\alpha| \geq n$.

Observe that the alphabet (u, v) satisfies $|uv| = |\alpha|$, which contradicts the assumption that (α, β) is the first alphabet in the sequence of inductive applications of the algorithm satisfying this inequality. Thus, the proof of the corollary is complete. ■

Remark 3.24. Clearly, the previous corollary holds for any $w \in \Sigma(3 + B^{-3n}, 3n)$, where $B > 3 + 2\sqrt{2}$ and $n \in \mathbb{N}^*$ is large enough (depending on B).

The following corollaries are straightforward consequences of the renormalization algorithm and are thus presented here. Nevertheless, they are not used in the proof of Theorem 1.1 and will only be used in the next section. Recall that a word belongs to $\Sigma^{(r-2)}(3 + e^{-r})$ if it belongs to both $\Sigma(3 + e^{-r}, |w|)$ and Q_{r-2}.

Corollary 3.25. *Let $r \in \mathbb{N}$ and let $w \in \Sigma^{(r-2)}(3 + e^{-r})$ be a finite word. If w is (u, v)-weakly renormalizable as $w_1\gamma w_2$ with $\gamma \neq \varnothing$, then w is (α, β)-weakly renormalizable for some $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$.*

Proof. Observe that if $w = c_1 \dots c_n \in Q_{r-2}$, then $\mathsf{r}(c_1 \dots c_{n-1}) \leq r - 3$, so

$$\mathsf{s}(w)^{-1} \leq 2\mathsf{s}(c_n)^{-1}\mathsf{s}(c_1 \dots c_{n-1})^{-1} \leq 12e^{r-2},$$

which implies that $\mathsf{r}(w) \leq r$. We then use Lemma 3.21. ■

Corollary 3.26. *Let $\theta \in P$ and let $w \in \Sigma^{(r-2)}(3 + e^{-r})$ be a word starting with θ. Then, by extending w by at most one digit to the right, w is (α, β)-weakly renormalizable for some alphabet (α, β) satisfying $|\alpha\beta| \geq r/6$.*

Proof. By Lemma 3.5 we know that w does not contain "internal" blocks of 1's or 2's of odd length, that is, words of the form $c'c^nc'$ for $c, c' \in \{1, 2\}$ with $c \neq c'$ for some odd $n \in \mathbb{N}$. Since w starts with θ, it starts with an even block as well. On the other hand, w can possibly end with an odd block of 1's or 2's. If w ends with an odd block of 2's, then $w = \gamma w_2$ is (a, b)-weakly renormalizable, where $\gamma \in \langle a, b\rangle$ and $w_2 = 2$. In case it ends with an odd block of 1's, we just need to extend $w = c_1 \dots c_n$ to $\tilde{w} = w1 = c_1 \dots c_n 1$. In this case

$$\mathsf{s}(c_1 \dots c_n 1)^{-1} \leq 2\mathsf{s}(c_1 \dots c_{n-1})^{-1}\mathsf{s}(1, 1)^{-1} \leq 12e^{r-2},$$

which gives $\mathsf{r}(c_1 \dots c_n 1) \leq r$.

We claim that w or $\tilde{w}$ is (α, β)-weakly renormalizable for an alphabet (α, β) satisfying $|\alpha|, |\beta| < |w|$ with $|\alpha\beta| \geq |w|/2$. Indeed, if $|\alpha\beta| < |w|/2$, then writing $w = w_1\gamma w_2$ gives $|w_1| + |w_2| < 2|\alpha\beta| \leq |w|$. We obtain $\gamma \neq \varnothing$, so we can continue applying the algorithm. Here, we have skipped most details as this is very similar to the proof of Corollary 3.23.

We remark that if, for some iteration of the algorithm, we obtain $\gamma = \alpha^r$ (respectively, $\gamma = \beta^r$), then the algorithm increases the size of the alphabet, but does not change the renormalization kernel γ. In these cases, w is a subword of $\alpha\beta\alpha^r\alpha\beta$ and of $\alpha^{r+1}\beta\alpha^{r+1}\beta$ (respectively, of $\alpha\beta\beta^r\alpha\beta$ and of $\alpha\beta^{r+1}\alpha\beta^{r+1}$), and so it belongs to $\Sigma(3,|w|)$.

Now, let (α,β) be such an alphabet. If $r \le 24$, then $|\alpha\beta| \ge 4 \ge r/6$ since $|\alpha|, |\beta| \ge 2$. If $r > 24$, we have

$$|\alpha\beta| \ge |w|/2 \ge (r-8)/(2\log(3+2\sqrt{2})) - 1/2 \ge r/6,$$

where we are using Lemma A.2. ■

Corollary 3.27. *Let $(\alpha,\beta) \in \bar{P}$ with $|\alpha\beta| < r/6$, let $\theta \in P$ and let $w \in \Sigma^{(r-2)}(3+e^{-r})$ be a word starting with θ. If w contains $\alpha\beta$, then w is (α,β)-semirenormalizable, say $\tilde{w} = w_1\gamma w_2$. Moreover, if w starts (resp. ends) with $\alpha\beta$, then $w_1 = \varnothing$ (resp. $w_2 = \varnothing$).*

Proof. First note that w is trivially (a,b)-semirenormalizable, say $\tilde{w} = \gamma_0$ where $\gamma_0 \in \langle a,b\rangle$. Now we inductively apply the renormalization algorithm (Lemma 3.21) to obtain a sequence of alphabets $(A_j,B_j) \in \bar{P}_j$ such that for all $0 \le j \le m$, the word $\tilde{w}$ is (A_j,B_j)-weakly renormalizable for each j and $|A_m B_m| \ge r/6$.

On the other hand, since $(\alpha,\beta) \in \bar{P}$, there exists a sequence of alphabets $(\alpha_i,\beta_i) \in \bar{P}_i$ such that $\alpha\beta \in \langle\alpha_i,\beta_i\rangle$ for all $0 \le i \le n$ and $(\alpha_n,\beta_n) = (\alpha,\beta)$. Since $\alpha\beta$ starts with $a = \alpha_0$ and ends with $b = \beta_0$ (Lemma 3.8), inductively we find that $\alpha\beta$ starts with α_i and ends with β_i. In particular, $\alpha\beta$ contains $\alpha_i\beta_i$.

Write $\tilde{w} = w_1\gamma_j w_2$ as in the definition of (A_j,B_j)-weakly renormalizable. Using the fact that $\alpha\beta$ contains $\alpha_j\beta_j$, gluing some words τ and τ' we get

$$\tau\alpha_j\beta_j\tau' = A_jB_j\gamma_jA_jB_j \in \langle A_j,B_j\rangle,$$

hence by Lemma 3.10 we obtain $(A_j,B_j) = (\alpha_j,\beta_j)$ for all $0 \le j \le n$. In particular, $m > n$, because otherwise $r/6 \le |A_mB_m| = |\alpha_m\beta_m| < r/6$. This shows that $\tilde{w}$ is (α,β)-weakly renormalizable.

Now assume that w starts with $\alpha\beta$ (the other case is analogous). Observe that there is no need to complete the word to the left. We will show that $w_1 = \varnothing$ for all $0 \le j \le n$. Note that we have already shown that w_1 is empty for $(\alpha_0,\beta_0) = (a,b)$. If w_1 becomes nonempty for $k+1$ for some $0 \le k \le n$, it must happen that $\tilde{w} = \gamma_k w_2$ starts with β_k (because of the renormalization algorithm). But w starts with $\alpha\beta$, which in turn starts with $\alpha_k^s\beta_k$, which leads to a contradiction because it starts with $(\beta_k)_a$ by Lemma 3.8. Since $(\alpha_n,\beta_n) = (\alpha,\beta)$, this finishes the proof. ■

Proof of Theorem 1.1. Consider $n \ge 68$. We claim that $\Sigma(3+6^{-3n},n) = \Sigma(3,n)$. Indeed, let $\theta \in \Sigma(3+6^{-3n},n)$. By definition, θ can be continued indefinitely to the left and right, so in particular there exists a word $\tau \in \Sigma(3+6^{-3n},3n)$ obtained by gluing words of size n on each side of θ. By Corollary 3.23, there exists $(\alpha,\beta) \in \bar{P}$ with $|\alpha|,|\beta| < n$ and $|\alpha\beta| \ge n$ such that τ is (α,β)-semirenormalizable. Writing $\tilde{\tau} = w_1\gamma w_2$ as in the definition of weak renormalization, we have $|w_1|,|w_2| < \max\{|\alpha|,|\beta|\} < n$, so θ is a factor of γ. Considering the smallest sequence η of (α,β)-letters of γ containing θ as a factor, the sequence obtained by removing the first and the last (α,β)-letter of η has size smaller than n and thus cannot contain $\alpha\beta$ or $\beta\alpha$ as factors, and thus η is of the form α^r, β^r, $\alpha^r\beta$, $\beta^r\alpha$, $\beta\alpha^r\beta$, $\alpha\beta^r\alpha$, $\beta\alpha^r$ or $\alpha\beta^r$ for some positive integer r.

In each of these cases, $\eta \in \Sigma(3, |\eta|)$. Indeed, since $(\alpha, \beta) \in \bar{P}$, all of these words are factors of words in $\mathsf{c}(\bar{P}) = P$ (where recall that c is the concatenation operator $\mathsf{c}(u, v) = uv$). Since, by Lemma 2.1, the set of factors of words in P coincides with the set of words w satisfying $w \in \Sigma(3, |w|)$, we deduce that $\eta \in \Sigma(3, |\eta|)$. Therefore, $\theta \in \Sigma(3, n)$.

To complete our proof, we need to show that, for every sufficiently large integer n, we have

$$\Sigma(3 - 6^{-3n}, n) = \Sigma(3, n).$$

Indeed, given $w \in \Sigma(3, n)$, by Lemmas 2.1 and 3.11, there exists $\Pi \in P$ containing w such that $|\Pi| \le 3|w|$. Since $(3 + 2\sqrt{2})^3 < 6^3$, if n is sufficiently large then Lemma 3.4 shows that $\Pi^\infty \in \Sigma(3 - 6^{-3n})$, so $w \in \Sigma(3 - 6^{-3n}, n)$. ■

4. Improving the estimates

Bombieri [1, Lemma 13] characterized the words in $\Sigma(3)$ by stating the conditions that the sequence $(e_i)_{i\in\mathbb{Z}}$ of exponents has to satisfy for a Type I or Type II bi-infinite word to belong to $\Sigma(3)$ (where we are using the terminology of Section 2.1). We begin this section by stating an analogue of this fact for words in $\Sigma(3 + e^{-r}, n)$. The proof is essentially applying the renormalization algorithm to a word of the form $w = \alpha^{e_i}\beta\alpha^{e_{i+1}}\beta$ or $w = \beta^{e_i}\alpha\beta^{e_{i+1}}\alpha$, but we need to be careful about the magnitude of $\mathsf{r}(w)$.

Lemma 4.1 (Bombieri's characterization). *Let $(\alpha, \beta) \in \bar{P}$. Consider a word γ of the form $\gamma = \alpha^{e_0}\beta\alpha^{e_1}\beta\dots\beta\alpha^{e_\ell}$ or $\gamma = \beta^{e_0}\alpha\beta^{e_1}\alpha\dots\alpha\beta^{e_\ell}$ with $e_i \ge 1$ for all $1 \le i \le \ell - 1$. Assume that $\gamma \in \Sigma(3 + e^{-r}, |\gamma|)$ and let $\theta = \alpha$ in the first case and $\theta = \beta$ in the second case. If $\mathsf{r}(\theta^{e_i}) \le r - 2|\alpha\beta|$, then*

- *for $1 \le i \le \ell - 2$, we have $|e_i - e_{i+1}| \le 1$,*
- *for $i = 0$, we have $e_1 \ge e_0 - 1$ when $\theta = \alpha$; if $\theta = \beta$, and moreover $\mathsf{r}(\beta^{e_0}) \le r - 6|\alpha\beta|$ or $|\alpha| \le |\beta|$, then $e_1 \ge e_0 - 1$.*
- *for $i = \ell - 1$, we have $e_\ell \le e_{\ell-1} + 1$ when $\theta = \beta$; if $\theta = \alpha$, and moreover $\mathsf{r}(\alpha^{e_{\ell-1}}) \le r - 6|\alpha\beta|$ or $|\beta| \le |\alpha|$, then $e_\ell \le e_{\ell-1} + 1$.*

Before proceeding to the proof, we must comment why we need $\mathsf{r}(\theta^{e_i})$ to be smaller at the end of the word in the last two bullet points. Observe that if $\beta = \alpha^s v$ for some $(\alpha, v) \in \bar{P}$, then clearly e_ℓ can be much larger than $e_{\ell-1}$, because all powers $\alpha^{e_\ell - e_{\ell-1}}$ could belong to the (potential) next letter β. Similarly, when $\alpha = u\beta^s$ for some $(u, \beta) \in \bar{P}$, the power $\beta^{e_0 - e_1}$ could belong to the (potential) preceding letter α.

Proof of Lemma 4.1. Let ω be a bi-infinite word containing γ and such that $\omega \in \Sigma(3 + e^{-r})$. Note that if $\{\theta, \tilde{\theta}\} = \{\alpha, \beta\}$ then $\mathsf{r}(\theta^{e_i}\tilde{\theta}) < r - 2|\alpha\beta| + 2|\theta| + 4 \le r$ by Lemma A.2.

Suppose $\gamma = \alpha^{e_0}\beta\alpha^{e_1}\beta\dots$. Take $k \le e_{i+1}$ maximal such that $\mathsf{r}(\alpha^{e_i}\beta\alpha^k) \le 2r$. If $e_i \ge k$ then $\mathsf{r}(\alpha^k) \le r - 2|\alpha\beta|$ as well, so actually $k = e_{i+1}$ because, otherwise,

$$\begin{aligned}\mathsf{r}(\alpha^{e_i}\beta\alpha^{k+1}) &\le \mathsf{r}(\alpha^{e_i}) + \mathsf{r}(\beta\alpha) + \mathsf{r}(\alpha^k) + 4\\ &\le r - 2|\alpha\beta| + \mathsf{r}(\beta\alpha) + r - 2|\alpha\beta| + 4 \le 2r,\end{aligned}$$

where we have used Lemma A.2 to guarantee that $\mathsf{r}(\alpha\beta) \le 1.8|\beta\alpha| + 1.8$. Similarly, we use $\mathsf{r}(\beta) \le 1.8|\beta| + 1.8$ (for $\beta = b$ use $\mathsf{r}(b) = 1$ instead) to get

$$\mathsf{r}(\alpha^{e_i}\beta\alpha^{e_{i+1}}\beta) \le 2r - 4|\alpha\beta| + 2\,\mathsf{r}(\beta) + 6 \le 2r.$$

Hence, letting $(\tilde{\alpha}, \tilde{\beta}) = (\alpha, \alpha^{e_{i+1}}\beta)$ we find that

$$\tilde{\gamma} = \alpha^{e_i}\beta\alpha^{e_{i+1}}\beta = \tilde{\alpha}^{e_i - e_{i+1}}\tilde{\beta}\tilde{\beta}$$

is a subword of a word $\omega \in \Sigma(3 + e^{-r})$, so if $e_i - e_{i+1} \ge 2$ this will contradict the second part of Lemma 3.15. If $e_i < k$, then let $(u, v) = (\alpha, \alpha^{e_i - 1}\beta)$ and $(\tilde{\alpha}, \tilde{\beta}) = (u, uv)$. Thus by the first case of Lemma 3.17,

$$\alpha\beta\alpha^{e_i}\beta\alpha^{e_{i+1}}\beta = \alpha\beta\tilde{\beta}\alpha^{e_{i+1}-e_i}\tilde{\beta} = \beta_a u^b\tilde{\beta}\tilde{\alpha}^{e_{i+1}-e_i}v_a u^b = \beta_a(\tilde{\alpha}^{e_{i+1}-e_i}\tilde{\beta}\tilde{\beta})^*\alpha^b$$

is a subword of γ when $i < \ell - 1$. If $e_{i+1} - e_i \ge 2$, then $\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta}$ would be a subword of γ^* with $\mathsf{r}(\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta}) = \mathsf{r}(\alpha^{e_i+2}\beta\alpha^{e_i}\beta) \le 2r$, which contradicts Lemma 3.15. This settles the first bullet point for $\theta = \alpha$. In the particular case where $i = \ell - 1$, we do not necessarily have β after α^{e_ℓ}. If $|\beta| \le |\alpha|$, then $(\alpha^b)^*(\tilde{\alpha}^{e_{i+1}-e_i}\tilde{\beta}\tilde{\beta})(\beta_a)^*$ is a subword of γ^* after removing a β^* at the beginning, so $\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta}$ is still a subword of γ^*.

When $(\alpha, \beta) = (u, uv)$ we need to extend the word $\beta\alpha^{e_{\ell-1}}\beta\alpha^{e_\ell}$ by using Corollary 3.27. We will extend this word to the left and then to the right. Since $|uv| = |\beta| < r/6$ and $\mathsf{r}(\beta) < r - 2$ (because $0 \le \mathsf{r}(\alpha^{e_{\ell-1}}) \le r - 6|\alpha\beta|$), consider the (u, v)-semirenormalizable continuation $w \in \Sigma^{(r-2)}(3 + e^{-r})$ inside ω that contains and ends in the leftmost β of $\beta\alpha^{e_{\ell-1}}\beta\alpha^{e_\ell}$ (one begins with β and then one starts adding the digits of ω that are to the left of that β until one obtains a word w with $\mathsf{r}(w) \ge r - 2$, which by minimality must be in $\Sigma^{(r-2)}(3 + e^{-r})$); in particular, w has a (u, v)-weakly renormalizable extension $\hat{w} = \hat{w}_1\hat{\gamma}$ where $\hat{\gamma} \in \langle u, v\rangle$ and $\hat{w}_1$ is a suffix of uv. We claim that $|u| \le |\hat{w}_1|$. Otherwise, Lemma A.2 yields

$$\frac{r-2}{1.8} - 1 \le |\hat{w}| \le |\hat{w}_1\beta| \le |\alpha\beta| - 1 \le \frac{r}{6} - 1,$$

which is a contradiction. Hence $|u| \le |\hat{w}_1|$ and $\hat{w}_1$ ends with u^b. Therefore there must be a $u^b\alpha^f$ with $f \ge 0$ before the first β.

Now we want to extend the word to the right. Consider now the continuation $w \in \Sigma^{(r-2)}(3 + e^{-r})$ that begins at $\alpha\beta\alpha^{e_\ell}$. In particular, w has an extension $\hat{w}$ that is (α, β)-weakly renormalizable by Corollary 3.27. Since

$$\begin{aligned}\mathsf{r}(\alpha\beta\alpha^{e_{\ell-1}+1}\alpha\beta) &\le \mathsf{r}(\alpha\beta\alpha) + \mathsf{r}(\alpha^{e_{\ell-1}}) + \mathsf{r}(\alpha\beta) + 4\\ &\le 3.6|\alpha\beta| + 1.8|\alpha| + r - 6|\alpha\beta| + 8 \le r - 2,\end{aligned}$$

we deduce that $\hat{w}$ contains all $\alpha\beta\alpha^{e_{\ell-1}+2}$ if $e_\ell \geq e_{\ell-1}+2$ and after it must come an $\alpha^g\beta$ or $\alpha^g\hat{w}_2$ where $\hat{w}_2$ starts with v_a and $g \geq 0$. In conclusion,

$$u^b\alpha^f\beta\alpha^{e_{\ell-1}}\beta\alpha^{e_{\ell-1}+2+g}v_a$$

is a subword of $\omega \in \Sigma(3+e^{-r})$. In this situation the first case of Lemma 3.17 yields

$$u^b\alpha^f\beta\alpha^{e_{\ell-1}}\beta\alpha^{e_{\ell-1}+2+g}v_a = (\alpha^{e_{\ell-1}+2+g}\beta\alpha^{e_{\ell-1}}\beta\alpha^f\beta)^*,$$

so we still find that $\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta} = \alpha^{e_{\ell-1}+2}\beta\alpha^{e_{\ell-1}}\beta$ is a subword of $\omega^* \in \Sigma(3+e^{-r})$, contrary to Lemma 3.15 above. This finishes the case $\theta = \alpha$.

Now assume $\gamma = \beta^{e_0}\alpha\beta^{e_1}\alpha\ldots$. Take the maximal integer $k \leq e_{i+1}$ $\mathsf{r}(\beta^{e_i}\alpha\beta^k) \leq 2r$. If $e_i < k$, then letting $(\tilde{\alpha},\tilde{\beta}) = (\alpha\beta^{e_i},\beta)$ one finds that $\alpha\beta^{e_i}\alpha\beta^k = \tilde{\alpha}\tilde{\alpha}\tilde{\beta}^{k-e_i}$ is a subword of a word $\omega \in \Sigma(3+e^{-r})$. Observe that

$$\begin{aligned}\mathsf{r}(\alpha\beta^{e_i}\alpha\beta^{e_i+2}) &\leq \mathsf{r}(\alpha)+\mathsf{r}(\beta^{e_i})+\mathsf{r}(\alpha\beta^2)+\mathsf{r}(\beta^{e_i})+6\\ &\leq 2r-4|\alpha\beta|+\mathsf{r}(\alpha)+\mathsf{r}(\alpha\beta^2)+6 \leq 2r,\end{aligned}$$

where we have used $\mathsf{r}(\alpha)+\mathsf{r}(\alpha\beta^2) \leq 3.6|\alpha\beta|+3.6$ for $|\alpha\beta| \geq 24$ and the explicit values for $|\alpha\beta| < 24$. Since $k \leq e_{i+1}$, if $e_{i+1}-e_i \geq 2$ we again have a contradiction with Lemma 3.15. If $e_i \geq k$, then $\mathsf{r}(\beta^k) \leq r-2|\alpha\beta|$ so $k = e_{i+1}$ as before, hence let $(u,v) = (\alpha\beta^{e_{i+1}-1},\beta)$ and $(\tilde{\alpha},\tilde{\beta}) = (uv,v)$; then by the second case of Lemma 3.17 one deduces that

$$\alpha\beta^{e_i}\alpha\beta^{e_{i+1}}\alpha\beta = \tilde{\alpha}\tilde{\beta}^{e_i-e_{i+1}}\tilde{\alpha}\alpha\beta = v_a u^b\tilde{\beta}^{e_i-e_{i+1}}\tilde{\alpha}v_a\alpha^b = \beta_a(\tilde{\alpha}\tilde{\alpha}\tilde{\beta}^{e_i-e_{i+1}})^*\alpha^b$$

is a subword of γ. If $e_i-e_{i+1} \geq 2$ then one would infer that $\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta} = \alpha\beta^{e_{i+1}}\alpha\beta^{e_{i+1}+2}$ is a subword of γ^* with $\mathsf{r}(\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta}) = \mathsf{r}(\alpha\beta^{e_{i+1}}\alpha\beta^{e_{i+1}+2}) \leq 2r$, which is impossible. This settles the first bullet point for $\theta = \beta$. In the particular case where $i = 0$, we do not have α before β^{e_0}. If $|\alpha| \leq |\beta|$ this is no problem because then $(\alpha^b)^*\tilde{\alpha}\tilde{\alpha}\tilde{\beta}^{e_i-e_{i+1}}(\beta_a)^*$ is a subword of γ after removing an α^* at the end, so we still find that $\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta}$ is a subword of γ. When $(\alpha,\beta) = (uv,v)$, an analogous argument shows that the word $\beta^{e_0}\alpha\beta$ has a continuation to the left that is (α,β)-weakly renormalizable. So as before there is either an $\alpha\beta^f$ or a $\hat{w}_1\beta^f$, where $\hat{w}_1$ is a suffix of $\alpha\beta$ that ends with u^b. Similarly there is a $\beta^g v_a$ to the right of $\beta^{e_1}\alpha$. In summary, $u^b\beta^{e_0+f}\alpha\beta^{e_1+g}v_a$ is a subword of $\omega \in \Sigma(3+e^{-r})$. But the second case of Lemma 3.17 implies that

$$u^b\beta^{e_0+f}\alpha\beta^{e_1}\alpha\beta^g v_a = (\alpha\alpha\beta^g\alpha\beta^{e_1}\alpha\beta^{e_0+f})^*.$$

So again if $e_0-e_1 \geq 2$, then $\tilde{\alpha}\tilde{\alpha}\tilde{\beta}\tilde{\beta} = \alpha\beta^{e_1}\alpha\beta^{e_1+2}$ is a subword of $\omega^* \in \Sigma(3+e^{-r})$, which contradicts Lemma 3.15 once more. ■

4.1. Constructing renormalizable extensions

We start by considering local extensions. More precisely, if we have a word w that starts (resp. ends) with $\alpha\beta$ where $(\alpha,\beta) \in \overline{P}$, then the alphabet is uniquely determined, and the beginning (resp. end) of w should be (α, β)-weakly renormalizable. This is a consequence of Corollary 3.27.

Now we want to consider extensions of renormalizable words. The next lemma says that if we have a power $(uv)^s$ with $(u, v) \in \bar{P}$, then it will have a large extension w that is "almost" (u, v)-weakly renormalizable consisting mostly of powers $(uv)^{s_i}$. This is explained since exponents can only decrease linearly, in fact, they may only decrease by 1 when below some threshold by Lemma 4.1. We say that $w = \gamma w_2$ is almost (u, v)-weakly renormalizable, because the tail w_2 now satisfies the condition $|w_2| < 2|uv|$ and w_2 is a prefix of a word in $\{uuv, uvv\}$.

Lemma 4.2. *Let $w \in \Sigma^{(Tr-2)}(3 + e^{-r})$ be a finite word starting with θ^s, where $\theta = uv$ and $(u, v) \in \bar{P}$, $\mathsf{r}(\theta^s) \leq r - 4|\theta|$, $|uv| < r/9$ and also*

$$Tr \leq s^2|\theta|\log((3 + \sqrt{5})/2)/2.$$

Then $w = \gamma w_2$, where $\gamma = (uv)^{s_1}\theta_1(uv)^{s_2}\theta_2 \ldots (uv)^{s_\ell}$, $s_j \geq 0$, each θ_j belongs to $\{uuv, uvv\}$ and w_2 is a prefix of a word in $\{uuv, uvv\}$. Moreover, $\ell \leq 2.1\,Tr/|\theta^s| + 1$.

Proof. Let γ be the largest prefix of the word w than can be written in the form $\gamma = (uv)^{s_1}\theta_1(uv)^{s_2}\theta_2 \ldots (uv)^{s_\ell}$ where each θ_j belongs to $\{uuv, uvv\}$ and $s_j \geq 0$. Then we claim that:

- If $\mathsf{r}((uv)^{s_{j+1}}) \leq r - 6|uv|$, then $\theta_j = \theta_{j+1}$ for $1 \leq j \leq \ell - 2$.
- If $\mathsf{r}((uv)^{s_j}) \leq r - 10|uv|$, then $|s_j - s_{j+1}| \leq 1$ for $1 \leq j \leq \ell - 2$.
- $s_j \geq s - j$ for all $1 \leq j \leq \ell - 1$.

Indeed, to see the first claim, if $\theta_j \neq \theta_{j+1}$ note that

$$\begin{aligned}\mathsf{r}(\theta_j(uv)^{s_{j+1}}\theta_{j+1}) &\leq \mathsf{r}(\theta_j) + \mathsf{r}((uv)^{s_{j+1}}) + \mathsf{r}(\theta_{j+1}) + 4\\ &\leq 1.8|uuv| + r - 6|uv| + 1.8|uvv| + 7.6 \leq r - 2,\end{aligned}$$

which is clear for $|uv| \geq 16$, while for $|uv| < 16$ we have computed the explicit values of $\mathsf{r}(uuv) + \mathsf{r}(uvv)$ to check that the inequality $\mathsf{r}(\theta_j(uv)^{s_{j+1}}\theta_{j+1}) \leq r - 2$ still holds. If $\theta_j = uuv$, let $(\alpha, \beta) = (u, uv)$, and if $\theta_j = uvv$, let $(\alpha, \beta) = (uv, v)$. Consequently, $\alpha\beta\tau' = \theta_j(uv)^{s_{j+1}}\theta_{j+1}$ is a subword of a word in $\Sigma^{(r-2)}(3 + e^{-r})$. Then Corollary 3.27 says that this word is contained in a word that can be written in the alphabet $\{\alpha, \beta\}$. But if $\theta_{j+1} \neq \theta_j$, this would be impossible.

To prove the second claim, consider the subword $\theta_{j-1}(uv)^{s_j}\theta_j(uv)^{s_{j+1}}$. Since $\theta_{j-1} = \theta_j$, using the appropriate pair $(\alpha, \beta) \in \bar{P}$ this whole word can be written in that alphabet. An application of Bombieri's characterization (Lemma 4.1) yields $\theta_j = \theta_{j+1}$ and also the second claim. Indeed, if $(\alpha, \beta) = (u, uv)$ then $\theta_{j-1}(uv)^{s_j}\theta_j(uv)^{s_{j+1}}u = \alpha\beta^{s_j+1}\alpha\beta^{s_{j+1}+1}\alpha$ is a subword of a word in $\Sigma(3 + e^{-r})$ so the first bullet point of Lemma 4.1 gives $|s_j - s_{j+1}| \leq 1$. If $(\alpha, \beta) = (uv, v)$ then $\theta_{j-1}(uv)^{s_j}\theta_j(uv)^{s_{j+1}} = \alpha\beta\alpha^{s_j+1}\beta\alpha^{s_{j+1}}$, so the third bullet point of Lemma 4.1 gives $s_{j+1} \leq s_j + 2$. Therefore,

$$\begin{aligned}\mathsf{r}((uv)^{s_{j+1}}) &\leq \mathsf{r}((uv)^{s_j}) + \mathsf{r}((uv)^2) + 2\\ &\leq r - 10|uv| + 1.8|uv| + 3.8 \leq r - 6|uv|,\end{aligned}$$

which is clear for $|uv| \geq 10$, while for $|uv| < 10$ the inequality $\mathsf{r}((uv)^2) + 2 \leq 4|uv|$ can be checked directly. The previous item gives $\theta_j = \theta_{j+1}$, hence $\theta_{j-1}(uv)^{s_j}\theta_j(uv)^{s_{j+1}}\theta_{j+1} = \alpha\beta\alpha^{s_j+1}\beta\alpha^{s_{j+1}+1}\beta$ and finally we use the first bullet point of Lemma 4.1.

We now prove the third claim. Observe that $s_1 \geq s-1$ by construction of γ. If $s_j \geq s-2$ then we are done. Note that if $s_j \leq s-3$, then

$$\mathsf{r}((uv)^{s_j}) \leq r - 4|uv| - \mathsf{r}((uv)^3) + 1 \leq r - 6|uv|$$

where we have used $\mathsf{r}((uv)^3) \geq 2.8|uv| - 3$ for $|uv| > 4$ and for $(u, v) = (a, b)$ we have used $\mathsf{r}((uv)^3) = 16$. Hence the first claim shows that $s_j \leq s-3$ implies $\theta_j = \theta_{j-1}$. In particular, $(uv)^{s_{j-1}}\theta_{j-1}(uv)^{s_j}\theta_j$ can be written in the appropriate alphabet $(\alpha, \beta) \in \overline{P}$ as $\alpha^{s_{j-1}+1}\beta\alpha^{s_j+1}\beta$ or $\beta^{s_{j-1}}\alpha\beta^{s_j+1}\alpha\beta$. Since $s_j \leq s-3$ implies $\mathsf{r}((uv)^{s_j}) \leq r - 2|\alpha\beta|$ by hypothesis, Lemma 4.1 gives $s_j \geq s_{j-1} - 1$. This proves the three claims.

Now, note that $s \geq 2$, since if $s = 1$ we get a contradiction because

$$\mathsf{r}(\theta) \leq \mathsf{r}(w) \leq Tr \leq |\theta|/2$$

gives $|\theta| \leq 8$, but $\mathsf{r}(ab) = 5$ and $\mathsf{r}(ab) \leq \mathsf{r}(\theta) \leq |\theta|/2 \leq 4$.

On the other hand, using Lemma A.2 one sees that

$$\begin{aligned}\left(\log\left(\frac{3+\sqrt{5}}{2}\right)\right)^{-1} Tr + 4 \geq |w| \geq |\gamma| &= (s_1 + \cdots + s_\ell)|uv| + |\theta_1| + \cdots + |\theta_\ell| \\ &\geq (s_1 + \cdots + s_\ell + \ell)|\theta|\end{aligned}$$

If $\ell \geq s$ then one gets

$$\begin{aligned}\left(\log\left(\frac{3+\sqrt{5}}{2}\right)\right)^{-1} Tr + 4 &> (s + (s-1) + \cdots + 1)|\theta| \\ &\geq s^2|\theta|/2 + s|\theta|/2 \geq s^2|\theta|/2 + 4,\end{aligned}$$

contradicting the hypothesis (we have used irrationality). Thus $\ell < s$. Hence

$$\left(\log\left(\frac{3+\sqrt{5}}{2}\right)\right)^{-1} Tr + 4 \geq (\ell - 1)(s - \ell/2)|\theta|.$$

This shows that $\ell \leq 2.1\,Tr/|\theta^s| + 1$ (for $\ell = 2$ we use $\mathsf{r}(\theta^s) \leq Tr$) instead.

Now, write $w = \gamma w_2$. We have two cases to consider. When $s_\ell \leq 1$, we choose $\hat{w} \in \Sigma^{(r-2)}(3 + e^{-r})$ starting with $\theta_{\ell-1}$, while when $s_\ell \geq 2$, we choose $\hat{w} \in \Sigma^{(r-2)}(3 + e^{-r})$ starting at the last occurrence of $uvuv$. Since $|uv| < r/6$, Corollary 3.27 shows that $\hat{w}$ is (u, v)-semirenormalizable. We know that $\hat{w}$ is $(\hat{\alpha}, \hat{\beta})$-semirenormalizable for some $(\hat{\alpha}, \hat{\beta}) \in \overline{P}$ with $|\hat{\alpha}\hat{\beta}| \geq r/6$, because of Corollary 3.26. Since $|uv| < r/6$ and $\hat{w}$ has uv, the word $\hat{w}$ is (α, β)-semirenormalizable for some $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$. Possibly adding one digit to the right, write $\hat{w} = \hat{\gamma}\hat{w}_2$ with $\hat{\gamma} \in \langle u, v\rangle$ and $\hat{w}_2$ a prefix of uv. Observe that $5|uv| < r/1.8 < |\hat{w}|$, thus $|\hat{\gamma}| = |\hat{w}| - |\hat{w}_2| > 4|uv|$.

We will find a continuation of $\gamma\gamma'$ of γ with

$$\gamma' \in \{uv, uuv, uvv\} \cap \{\alpha, \beta, \alpha\beta\}.$$

By the maximality of γ, we find that w_2 is contained in γ', which will finish the proof of the lemma. When $s_\ell \geq 2$, if $uvuv = \alpha\alpha$ we see that $\alpha\alpha\beta = uv(uvv)$ extends and $\alpha\alpha\alpha = (uv)^3$ extends $(uv)^{s_\ell}$ to $(uv)^{s_\ell+1}$. If $uvuv = \beta\beta$ then $\beta\beta\beta = (uv)^3$ extends $(uv)^{s_\ell}$ to $(uv)^{s_\ell+1}$ while if there is $\alpha\alpha$ after $\beta\beta$, then $\beta\beta$ is followed by $\alpha^k v_a$ (if there is no β after $\alpha\alpha$ then $\hat{w}_2$ starts with v_a). But note that

$$\beta\beta\alpha^k v_a = v_a u^b \beta\alpha^k v_a = v_a(\alpha^k\beta\beta)^*,$$

so $\alpha\alpha\beta\beta$ is a subword of w^*, contradicting Lemma 3.15. In the situation where $s_\ell = 1$, we have $uvvuv = \alpha\beta\alpha$, so if there is an α afterwards, then $\alpha\beta\alpha\alpha$ extends $(uv)^{s_\ell}$ while $\alpha\beta\alpha\beta$ gives a contradiction because before θ_ℓ there is $s_{\ell-1} \geq 2$ and we have $\alpha\alpha\alpha\beta\alpha\beta = \alpha\alpha\tilde{\beta}\tilde{\beta}$. When $\theta_{\ell-1}uv = uuvuv = \alpha\beta\beta$, if there is β afterwards then $\alpha\beta\beta\beta = \theta_{\ell-1}(uv)^2$ extends and $\alpha\beta\beta\alpha\beta = \theta_{\ell-1}(uv)\theta_\ell$ also extends, but if there is $\alpha\alpha$ after β we have the same contradiction with Lemma 3.15 in the transpose word. If $s_\ell = 0$, we have $uvuv\theta_{\ell-1} \in \{\alpha\alpha\alpha\beta, \beta\beta\alpha\beta\}$ and we arrive at the same contradictions or extensions as before. In summary, w_2 is a subword of γ', so it is a subword of a word in $\{uv, uuv, uvv\}$. ∎

If we have a power a^e or b^e, then its extensions are not necessarily (a, b)-weakly renormalizable, because we could have odd powers of a digit $\{1, 2\}$ appearing after. Nevertheless, it takes a long time for these powers to decay, which gives us the next extension lemma.

Lemma 4.3. *Let $w \in \Sigma^{(Tr-2)}(3 + e^{-r})$ be a finite word starting with c^s, where $c \in \{1, 2\}$ and $\mathsf{r}(c^s) \leq r - 2$. Suppose that $Tr \leq s^2(\log x)/4$ where $x = (3 + \sqrt{5})/2$ and $x = 3 + 2\sqrt{2}$ for $c = 1$ and $c = 2$ respectively. Then $w = c^{s_1}\theta' c^{s_2}\theta' \ldots \theta' c^{s_\ell}$, $\theta' = c'c'$, $c' \in \{1, 2\}$, $c \neq c'$ and $Tr \geq \ell(s - \ell + 1)\log x$ (in particular $\ell < 2Tr/(s\log x)$). Moreover, if $\mathsf{r}(c^{s_j}) \leq r - 8$ then s_j is even and $|s_{j+1} - s_j| \in \{0, 2\}$.*

Proof. Observe that we only have to prove that in w there is no $c'c'c'c'$, as the sequence $cc'c'c'c$ is forbidden by Lemma 3.5. Define $\gamma = c^{s_1}\theta' c^{s_2}\theta' \ldots \theta' c^{s_\ell}$ to be the longest sequence inside w starting with c^s that does not contain $(\theta')^2$. The fact that $\mathsf{r}(c^{s_j}) \leq r - 8$ implies s_j even and $|s_{j+1} - s_j| \in \{0, 2\}$ follows from Lemmas 3.5 and 4.1. Inequalities (A.4) for $c = 2$ and (A.3) for $c = 1$ give $Tr \geq \ell(s - \ell + 1)\log x$. Since $\ell \leq s/2 - 1$ (by hypothesis), we can bound the length: $\ell < 2Tr/(s\log x)$. If $w \neq \gamma$, then γ must end with $\theta' cc\theta'$, but this would imply that $\ell \geq s/2$. ∎

The next situation is where we have an (α, β)-renormalizable word w that does not contain big powers of α or β. In this case there is an extension that can be written in the same alphabet (α, β). In fact, this extension is almost (α, β)-weakly renormalizable, in the sense that its tail is small; it is a prefix of a word in $\langle\alpha, \beta\rangle$ but it is not necessarily a prefix of $\alpha\beta$.

Lemma 4.4. *Let $w \in \Sigma^{(r-2)}(3 + e^{-r})$ be an (α, β)-weakly renormalizable word with $|\alpha\beta| < r/40$. Suppose that for every factor of the form α^s or β^s of w, we have $|\alpha^s| < \delta|w|$ and $|\beta^s| < \delta|w|$ where $0 < \delta < (1/2)\log((3 + \sqrt{5})/2)$ is a constant. Assume further that*

w contains an $\alpha\beta$. If $\overline{w} \in \Sigma^{(Tr)}(3+e^{-r})$ is an extension of w with $T+2 \le \delta' r/(16|\alpha\beta|)$ where $\delta' = 1 - 2\delta/\log((3+\sqrt{5})/2))$, then $w\overline{w} = w_1\gamma\tau'$ where $\gamma \in \langle\alpha,\beta\rangle$, τ' with $|\tau'| < |\alpha\beta|$ is a prefix of some word in $\langle\alpha,\beta\rangle$ and w_1 is a suffix of $\alpha\beta$.

Proof. Write $w = w_1\gamma w_2$ as in the definition of (α,β)-weakly renormalizable. Take γ' to be the largest word inside $w\overline{w}$ starting with γ which can be written in the alphabet $\langle\alpha,\beta\rangle$. Write $w\overline{w} = w_1\gamma'\tau'$. We will show that τ' is a prefix of some word in $\langle\alpha,\beta\rangle$ and $|\tau'| < |\alpha\beta|$. Take the last factor of $\alpha\beta$ in γ', say $\gamma' = \eta\alpha\beta\eta'$. In particular, by Lemma 3.10 we obtain $\eta' = \theta^s$ with $\theta \in \{\alpha,\beta\}$. Consider the factor $\hat{w} \in \Sigma^{(r-2)}(3+e^{-r})$ (which possibly extends $w\overline{w}$) starting at this last occurrence of $\alpha\beta$. By Corollary 3.27, after possibly adding one digit to the right, it can be written as $\hat{w} = \hat{\gamma}\hat{w}_2$, $\hat{\gamma} = \alpha\beta\ldots$ with $\hat{\gamma} \in \langle\alpha,\beta\rangle$ and $\hat{w}_2$ a prefix of $\alpha\beta$. If $\alpha\beta\theta^s$ is strictly contained in $\hat{\gamma}$, then if $|\tau'| \ge |\alpha\beta|$ we will find that there is at least one more letter $\{\alpha,\beta\}$ of $\hat{\gamma}$ inside $w\overline{w}$ after γ', contradicting the maximality of γ'. Hence $|\tau'| < |\alpha\beta|$ and $\alpha\beta\theta^s\tau'$ is a subword of $\hat{\gamma}\alpha\beta$, so τ' is the prefix of a word in $\langle\alpha,\beta\rangle$ as claimed. Now suppose that $\alpha\beta\theta^s$ contains $\hat{\gamma}$, so $\mathsf{r}(\theta^s) \ge r - 4|\alpha\beta| - 6$. We will get a contradiction by considering an extension in the transpose word $(w\overline{w})^*$.

We need the identity $((uv)^k)^* = u^b(uv)^{k-1}v_a$. For $k=1$ this is a consequence of Lemma 3.8. To prove the identity for $k \ge 2$, let $(\alpha,\beta) = (uv,v)$, so that $(uv)^k = v_a u^b (uv)^{k-2} v_a u^b = v_a u^b \alpha^{k-2}\beta_a u^b$. Now since $u^b, v_a, \alpha^k\beta_a$ are all palindromic, one gets $((uv)^k)^* = u^b\alpha^{k-2}\beta_a u^b v_a = u^b(uv)^{k-1}v_a$.

In summary there is a θ^{s-1} inside $(w\overline{w})^*$. Let $\theta^{s'}$ be the maximal suffix of θ^{s-1} that satisfies $\mathsf{r}(\theta^{s'}) \le r - 4|\theta|$. Since $\mathsf{r}(\theta^{s'}) \ge r - 6|\alpha\beta| - 10$, by Lemma A.2 we see that $s'|\theta| = |\theta^{s'}| > r/2$. Observe that $\mathsf{r}((w\overline{w})^*) \le (T+2)r - 4$ by Lemma A.4, which gives

$$(s')^2|\theta|/4 \ge (r/8)\cdot\frac{r}{2|\theta|} \ge (T+2)r$$

since $T+2 \le \frac{r}{16|\theta|}$ by hypothesis. Note that $(T+2)r \le (s')^2/8$ also holds.

If $\theta = uv$ with $(u,v) \in \overline{P}$ we use Lemma 4.2 to find a $\tilde{\gamma} = \theta^{s_1}\theta_1\theta^{s_2}\ldots\theta^{s_\ell}$ starting with this $\theta^{s'} = (uv)^{s'}$, where each θ_i is in $\{uuv, uvv\}$ and such that $(w\overline{w})^* = \tilde{\eta}\tilde{\gamma}\tilde{w}_2$ with $\tilde{w}_2$ a prefix of some word in $\{uuv, uvv\}$. When $|\theta| = 2$, we write $\theta = cc$ with $c \in \{1,2\}$, $\theta' \in \{a,b\}$ and use Lemma 4.3 to find an extension $\tilde{\gamma} = c^{e_1}\theta' c^{e_2}\theta'\ldots c^{e_\ell}$ starting with $\theta^{s'}$ such that $(w\overline{w})^* = \tilde{\eta}\tilde{\gamma}$.

Since $(w\overline{w})^* = \overline{w}^* w^*$, when $\theta = uv$ we find that θ^{s_ℓ} (or at least some factor of it) is inside w^*. Similarly, when $|\theta| = 2$, we see that c^{e_ℓ} (or at least some factor of it) is inside w^*, so θ^{s_ℓ} is inside w where $s_\ell = 2\lfloor e_\ell/2\rfloor$.

In any case $\theta^{s_\ell - 1}$ is inside w (because of the identity proved above). In any case we will also get $s_\ell \ge s' - 1 - \ell$ and also $\ell \le 2.1\,Tr/|\theta^{s'}| + 1$. But note that using $|\theta^{s'}| > r/2$ and Lemma A.2, we get

$$\begin{aligned}|\theta^{s_\ell - 1}| &\ge |\theta^{s'-1-\ell}| = |\theta^{s'}| - |\theta^{\ell+1}| \ge r/2 - (\ell+1)|\theta| \\ &\ge r/2 - (2.1\,Tr/|\theta^{s'}| + 2)|\theta| \ge r/2 - 2(2.1\,T + 1) \\ &\ge r/2 - \frac{|\theta|}{|\alpha\beta|}\cdot\delta' r + 6|\theta| \ge (1/2)(1-\delta')r + 6|\theta| \ge \delta|w|,\end{aligned}$$

which contradicts the existence of those factors inside w. In conclusion, $w\bar{w} = w_1\gamma\tau'$ where τ' with $|\tau'| < |\alpha\beta|$ is a prefix of some word in $\langle\alpha, \beta\rangle$ and w_1 is a suffix of $\alpha\beta$. ■

4.2. Proof of Theorem 1.2

The main idea of the proof of Theorem 1.2 is to find a subcovering of the natural covering of K_t. Indeed, recall from the introduction that

$$d(t) = \min\{1, 2\cdot\dim_{\mathrm{H}}(K_t)\} = \min\{1, 2\cdot\dim_{\mathrm{B}}(K_t)\}.$$

In order to prove that the Hausdorff dimension of K_t is at most d we will start with a covering of K_t by a finite union of intervals and then replace each of these intervals by a suitable union of smaller subintervals so that the sum of the d-th powers of the sizes of the subintervals is smaller than the d-th power of the size of the initial interval.

The proof is quite long, so it is divided into several subsections. We will need the following combinatorial lemma.

Lemma 4.5. *Let U be a positive integer, and let m be a positive real number. If $U \le m$ then the number of solutions $(\ell, x_1, \dots, x_\ell)$ of $x_1 + \cdots + x_\ell \le (U - \ell)m$ with each $x_i \in \mathbb{N}^*$ is at most*

$$U\left(\frac{em\varepsilon_m}{1-\varepsilon_m}\right)^{(1-\varepsilon_m)(U+1)} = Ue^{(1-\varepsilon_m)(U+1)/\varepsilon_m} = Ue^{W(m)(U+1)},$$

where ε_m is the solution in $(0, 1)$ of the equation $\log(\frac{em\varepsilon}{1-\varepsilon}) = \frac{1}{\varepsilon}$.

In general (including the case when $U > m$), this number of solutions is at most $Ue^{Um/e^{W(m-1)}}$. For $m > 1$, this upper bound is equal to $Ue^{U\cdot\frac{m}{m-1}W(m-1)}$, and for $m \ge 5$, it is at most $Ue^{U\cdot\frac{\log m}{m}}e^{U\cdot W(m-1)} < Ue^{U\cdot\frac{\log m}{m}}e^{U\cdot W(m)}$.

In particular, if $U = \mathrm{o}(\frac{m}{\log m})$, then this number is at most $e^{(\log m - \log\log m + \mathrm{o}(1))(U+1)}$.

Proof. We should have $1 \le \ell \le U$. Given such an ℓ, the number of solutions of the inequality is the number of natural solutions of $x_0 + x_1 + \cdots + x_\ell = \lfloor(U-\ell)m\rfloor$, where x_0 is included to transform the inequality into an equality. This is equal to $\binom{\lfloor(U-\ell)m\rfloor+\ell}{\ell}$, and using the inequalities $\binom{n}{k} \le \frac{n^k}{k!} \le (\frac{en}{k})^k$, which hold for $1 \le k \le n$, this number of solutions is at most

$$\left(\frac{e(\lfloor(U-\ell)m\rfloor + \ell)}{\ell}\right)^\ell \le \left(\frac{e((U-\ell)m + \ell)}{\ell}\right)^\ell.$$

If $U \le m$, then $\ell \le U \le m$ and $(U-\ell)m + \ell \le (U+1-\ell)m$, so the previous upper bound is at most

$$\left(\frac{e((\tilde{U}-\ell)m)}{\ell}\right)^\ell,$$

where $\tilde{U} := U + 1$. Let $\varepsilon \in (0, 1)$ be such that $\ell = (1-\varepsilon)\tilde{U}$, so $\frac{\tilde{U}-\ell}{\ell} = \frac{\varepsilon}{1-\varepsilon}$. The derivative of $g(\ell) = \log\left(\frac{e((\tilde{U}-\ell)m)}{\ell}\right)^\ell = \ell\log(\frac{e((\tilde{U}-\ell)m)}{\ell})$ is

$$\log\left(\frac{e((\tilde{U}-\ell)m)}{\ell}\right)-\frac{\tilde{U}}{\tilde{U}-\ell}=\log\left(\frac{em\varepsilon}{1-\varepsilon}\right)-\frac{1}{\varepsilon},$$

and so $g(\ell)$ is maximized for $\ell=(1-\varepsilon_m)\tilde{U}$. Moreover, since there are U possible values of ℓ, the number of solutions we are estimating is at most

$$U\cdot e^{g((1-\varepsilon_m)\tilde{U})}=Ue^{(1-\varepsilon_m)(U+1)/\varepsilon_m},$$

since, by definition, $\frac{em\varepsilon_m}{1-\varepsilon_m}=e^{1/\varepsilon_m}$.

Notice that since ε_m is the solution in (0, 1) of the equation $\log(\frac{em\varepsilon}{1-\varepsilon})=\frac{1}{\varepsilon}$, writing $x_m=\frac{1-\varepsilon_m}{\varepsilon_m}$, we have $\frac{1}{\varepsilon_m}=x_m+1$, and thus $\log(\frac{em}{x_m})=x_m+1$, and $\log(\frac{m}{x_m})=x_m$. It follows that $x_m e^{x_m}=m$, and thus $x_m=W(m)$, and $Ue^{(1-\varepsilon_m)(U+1)/\varepsilon_m}=Ue^{(U+1)x_m}=Ue^{W(m)(U+1)}$.

In the general case, let us estimate

$$h(\ell)=\left(\frac{e((U-\ell)m+\ell)}{\ell}\right)^{\ell}.$$

The derivative of $\log h(\ell)=\log(\frac{e((U-\ell)m+\ell)}{\ell})^{\ell}=\ell\log(\frac{e((U-\ell)m+\ell)}{\ell})$ is

$$\log\left(\frac{(U-\ell)m+\ell}{\ell}\right)-\frac{\ell(m-1)}{(U-\ell)m+\ell}=\log z-\frac{m-1}{z},$$

where $z=\frac{(U-\ell)m+\ell}{\ell}$ is a decreasing function of ℓ and so $h(\ell)$ is maximized when $\log z=\frac{m-1}{z}$, which is equivalent to $z\log z=m-1$ and to $\log z=W(m-1)$. In this case, we have $h(\ell)=(ez)^{\ell}=e^{\ell\log(ez)}$. Since $\ell=\frac{Um}{z+m-1}$ and $\log(ez)=1+\log z=1+\frac{m-1}{z}=\frac{z+m-1}{z}$, we have $\ell\log(ez)=\frac{Um}{z}=\frac{Um}{e^{W(m-1)}}$, which gives the upper estimate for $h(\ell)$:

$$e^{Um/e^{W(m-1)}},$$

and as before, since there are U possible values of ℓ, the number of solutions we are estimating is at most

$$Ue^{Um/e^{W(m-1)}}.$$

For $m>1$, we have $e^{W(m-1)}=\frac{m-1}{W(m-1)}$, so our bound becomes

$$Ue^{UmW(m-1)/(m-1)}=Ue^{U\cdot\frac{m}{m-1}W(m-1)}.$$

We have $\frac{m}{m-1}W(m-1)=W(m-1)+\frac{W(m-1)}{m-1}$, so

$$Ue^{U\cdot\frac{m}{m-1}W(m-1)}=Ue^{U\frac{W(m-1)}{m-1}}e^{UW(m-1)},$$

and for $m\geq 5$, we have $\frac{W(m-1)}{m-1}<\frac{\log m}{m}$ (indeed, this is equivalent to

$$\frac{(m-1)\log m}{m}m^{(m-1)/m}=\frac{(m-1)\log m}{m}e^{\frac{(m-1)\log m}{m}}>m-1,$$

which is equivalent to $\log m > m^{1/m}$, and thus holds for every $m \geq 5$). Thus, in this case, our upper estimate becomes

$$Ue^{U\cdot\frac{\log m}{m}} e^{U\cdot W(m-1)} < Ue^{U\cdot\frac{\log m}{m}} e^{U\cdot W(m)}.$$

If $U = \mathrm{o}(\frac{m}{\log m})$ then $U = \mathrm{o}(e^{\log m - \log\log m})$, and thus, since $\frac{m}{m-1}W(m-1) = W(m-1) + \mathrm{o}(1) = W(m) + \mathrm{o}(1) = \log m - \log\log m + \mathrm{o}(1)$, we have

$$\begin{aligned} e^{(U+1)\cdot\frac{m}{m-1}W(m-1)} &= e^{\log m - \log\log m + \mathrm{o}(1)} e^{U\cdot\frac{m}{m-1}W(m-1)} \\ &= (1+\mathrm{o}(1))\frac{m}{\log m} e^{U\cdot\frac{m}{m-1}W(m-1)} > Ue^{U\cdot\frac{m}{m-1}W(m-1)}, \end{aligned}$$

which was our previous upper estimate, and we have

$$e^{(U+1)\cdot\frac{m}{m-1}W(m-1)} = e^{(\log m - \log\log m + \mathrm{o}(1))(U+1)},$$

which concludes the proof. ∎

Remark 4.6. In the proof of Theorem 1.2 we only use the case $U \leq m$ of Lemma 4.5. We use the general case of Lemma 4.5 only in Section 6.

Proof of Theorem 1.2. We start by recalling some notation. Let $\mathsf{m}(\omega) = \sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega))$ be the Markov value of ω, and we have

$$Q_r = \{\alpha = c_1 \dots c_n \mid \mathsf{r}(\alpha) \geq r,\ \mathsf{r}(c_1 \dots c_{n-1}) < r\},$$

that is, α belongs to Q_r if and only if $\mathsf{s}(\alpha) < e^{-r-1}$ and $\mathsf{s}(\alpha') \geq e^{-r-1}$, where α' is the word obtained by removing the last letter from w.

We now recall how the covering of K_t is constructed. We define the sets of words

$$\begin{aligned} C(t,r) &= \{\alpha = c_1 \dots c_n \in Q_r \mid K_t \cap I(\alpha) \neq \varnothing\} \\ &= \{\alpha \in Q_r \mid \alpha \text{ subword of a word } \omega \in (\mathbb{N}^*)^{\mathbb{Z}} \text{ with } \mathsf{m}(\omega) \leq t\}. \end{aligned}$$

Here, $K_t = \{[0;\gamma] \mid \gamma \in \pi_+(\Sigma(t))\}$ where $\pi_+\colon \Sigma \to \Sigma^+$ is the projection associated with the decomposition $\Sigma = \Sigma^- \times \Sigma^+ = (\mathbb{N}^*)^{\mathbb{Z}_-} \times (\mathbb{N}^*)^{\mathbb{N}}$, that is,

$$\pi_+(\dots c_{-2}c_{-1}c_0c_1c_2\dots) = c_0c_1c_2\dots.$$

Moreover, $\Sigma(t) = \{\omega \in (\mathbb{N}^*)^{\mathbb{Z}} \mid \mathsf{m}(\omega) \leq t\}$. It is clear that

$$\mathcal{M} \cap (-\infty, t) \subseteq (\mathbb{N}^* \cap [1, \lfloor t \rfloor]) + K_t + K_t.$$

Observe that K_t is covered by all $I(\alpha)$ where $\alpha \in C(t,r)$ for any fixed r.

If $r \leq s$, then the set $C(t,r)$ covers $C(t,s)$, in the sense that for any interval $I(\alpha)$ with $\alpha = c_1 \dots c_n \in C(t,s)$ there is $m \leq n$ such that $\tilde{\alpha} = c_1 \dots c_m \in C(t,r)$ and $I(\tilde{\alpha}) \subseteq I(\alpha)$.

Given d depending on r, we can prove that the Hausdorff dimension of K_t is at most d, by replacing an interval I (corresponding to a word in $C(t,r)$) with several intervals I_j contained in it, but smaller and of different sizes, each corresponding to a word in $C(t, Tr)$, where $T \in \{10, \lfloor \log^2 r \rfloor, \lfloor r/5 \rfloor\}$, whose union still contains the intersection of K_t with I and that satisfy $\sum_j |I_j|^d < |I|^d$. Since this process can be iterated, this shows that the d-dimensional Hausdorff measure of K_t is finite for large enough r.

By Corollary 3.26, if $w \in \Sigma^{(r)}(3 + e^{-r-4})$, then there is a sequence of alphabets (α_j, β_j) such that, for all $0 \le j \le m$, w is (α_j, β_j)-semirenormalizable with

$$(\alpha_0, \beta_0) = (a, b) \quad \text{and} \quad (\alpha_{j+1}, \beta_{j+1}) \in \{(\alpha_j \beta_j, \beta_j), (\alpha_j, \alpha_j \beta_j)\}$$

for each $0 \le j < m$, and $|\alpha_m \beta_m| \ge r/6$.

We consider such a renormalization (α_t, β_t) with

$$r/\sqrt{\log r} \le |\alpha_t| + |\beta_t| < 2r/\sqrt{\log r}.$$

We will consider words $\tilde{w} \in \Sigma^{(r)}(3 + e^{-r-4})$ such that $w\tilde{w} \in \Sigma(3 + e^{-r-4}, |w\tilde{w}|)$. Then, depending on $w\tilde{w}$, we will consider continuations $\overline{w} \in \Sigma^{(Tr)}(3 + e^{-r-4})$ for some $T \in \{10, \lfloor \log^2 r \rfloor, \lfloor r/5 \rfloor\}$ such that $w\tilde{w}\overline{w} \in \Sigma(3 + e^{-r-4}, |w\tilde{w}\overline{w}|)$.

The strategy is the following: If w (or, more generally, $w\tilde{w}$) contains a factor $\alpha_t \beta_t$, we may consider the factor $\hat{w} \in Q_{r+2}$ of $w\tilde{w}$ starting at this factor $\alpha_t \beta_t$; it should be (α_t, β_t)-renormalizable. We will attempt to use this argument several times in order to cover the whole word $w\tilde{w}$ by (α_t, β_t)-renormalizable words. To determine $\tilde{w}$, we only need to estimate the number of words in (α_t, β_t) after the last factor equal to $\alpha_t \beta_t$ in w. For this, we consider several cases according to the size of α_t^s, β_t^s as a factor of w, for some integer s.

In Cases 1 and 2 below, we choose $T = 10$, while in Case 3 we initially choose $T = \lfloor \log^2 r \rfloor$. Furthermore, in all the following cases except Case 3.2.2, we take $d = \frac{\log r - \log\log r}{r}$. In Case 3.2.2, corresponding to when $\alpha_t = 11$ and $\tilde{w}w$ contains a relatively long factor α_t^s, we initially choose the estimate $d = \frac{\log r - \log\log r + c_0 + o(1)}{r}$, where $c_0 = -\log\log(\frac{3+\sqrt{5}}{2}) > 0$. This is already enough to obtain the upper bound

$$d(3 + e^{-r}) \le 2 \cdot \frac{\log r - \log\log r + c_0 + o(1)}{r}.$$

The only case that produces a "bad" estimate is then Case 3.2.2. This estimate can actually be improved by a refined analysis using $T = \lfloor r/5 \rfloor$, giving rise to Case 3.2.3. Our final upper bound in Theorem 1.2 is derived in this way.

Case 1: Suppose first that, for every factor of the form α_t^s or β_t^s of w, we have $|\alpha_t^s| < |w|/3$ and $|\beta_t^s| < |w|/3$. In this case we take $T = 10$.

Then there is a factor $\alpha_t \beta_t$ in the first half of w, and until the next appearance of $\alpha_t \beta_t$ (which happens before the end of w), we have a factor with total size smaller than $|w|/2$ of the type $\alpha_t \beta_t \alpha_t^j \beta_t$ or $\alpha_t \beta_t^j \alpha_t \beta_t$ for some positive integer j. Suppose we are in the first case, without loss of generality.

Then, given a continuation $\tilde{w}\overline{w}$ of w with $\tilde{w} \in \Sigma^{(r)}(3 + e^{-r-4})$ and $\overline{w} \in \Sigma^{(10r)}(3 + e^{-r-4})$, Lemma 4.4 shows that $w\tilde{w}\overline{w} = \tau\gamma\tau'$ with $\gamma \in \langle \alpha_t, \beta_t \rangle$, τ a suffix of $\alpha_t \beta_t$ and τ' a prefix of some word in $\langle \alpha_t, \beta_t \rangle$ with $|\tau'| < |\alpha_t \beta_t|$. Thus, the continuation of the first factor of the form $\alpha_t\beta_t$ of w in $w\tilde{w}\overline{w}$ is a concatenation of factors of the form $\alpha_t^{\tilde{j}}\beta_t$ or $\alpha_t\beta_t^{\tilde{j}}$. The number of such factors is at most $|w\tilde{w}\overline{w}|/|\alpha_t\beta_t| \le (13r+10)/|\alpha_t\beta_t| \le 25\sqrt{\log r}$. Moreover, if we have two consecutive such factors $\alpha_t^{\tilde{j}_1}\beta_t$ and $\alpha_t^{\tilde{j}_2}\beta_t$ (or $\alpha_t\beta_t^{\tilde{j}_1}$ and $\alpha_t\beta_t^{\tilde{j}_2}$), then $|\tilde{j}_1 - \tilde{j}_2| \le 1$, and if we have two consecutive factors $\beta_t\alpha_t^{\tilde{j}_1}\beta_t$ and $\alpha_t\beta_t^{\tilde{j}_2}\alpha_t$ then $2 \le |j_1| + |j_2| \le 3$. This implies that each of these factors of the form $\alpha_t^{\tilde{j}}\beta_t$ or $\alpha_t\beta_t^{\tilde{j}}$ has at most three continuations of this form, and so the number of such continuations $\tilde{w}\overline{w}$ of w is at most $3^{25\sqrt{\log r}} < r$. Since the number of possible $w \in \Sigma^{(r)}(3+e^{-r-4})$ is $\mathrm{O}(r^3)$, the number of possible continuations $w\tilde{w}\overline{w}$ in this case is $\mathrm{O}(r^4)$.

Case 2: Suppose now that w has a factor α_t^s with $|\alpha_t^s| \ge |w|/3$ (the case of w having a factor β_t^s with $|\beta_t^s| \ge |w|/3$ will be analogous).

Let us check that we can apply Lemmas 4.2 and 4.3 to this factor. Observe that $r/6 - 1/3 \le |w|/3 \le s|\alpha_t|$, whence $s > \sqrt{\log r}/12$, so the condition $(T+2)r \le s^2|\alpha_t|/8$ holds for $T = 10$ and r large. In particular, $\ell \le 2.1\,Tr/|\alpha_t^s| + 1 < 150$. Hence $(s - \ell/2)|\alpha_t| > r/7$ for sufficiently large r. Going back to the inequality $1.1\,Tr + 4 \ge (\ell-1)(s-\ell/2)|\alpha_t|$ we get a stronger bound $\ell < (1.1\,Tr+4)/((s-\ell/2)|\alpha_t|) + 1 \le 80$. We consider two subcases depending on the length of α_t.

Case 2.1: Suppose that $|\alpha_t| > r^{15/16}$ and $\alpha_t = uv$ with $(u,v) \in \overline{P}$.

For r large we can assume further that $\mathsf{r}(\alpha_t^s) \le r - 4|\alpha_t|$, since $\frac{9}{10}|w| \le r+3$. Then we claim that after $\alpha_t^s = (uv)^s$ the continuation is renormalizable with alphabet $\{u,v\}$, and the first appearance of uu or vv determines the new alphabet ($\{u,uv\}$ or $\{uv,v\}$). To prove that, let $\hat{w} \in \Sigma^{((T+2)r)}(3+e^{-r-4})$ be the factor of $w\tilde{w}\overline{w}$ starting at that factor α_t^s. Then Lemma 4.2 gives $\hat{w} = \hat{\gamma}\hat{w}_2$, where

$$\hat{\gamma} = (uv)^{s_1}\theta_1(uv)^{s_2}\theta_2\ldots(uv)^{s_\ell},$$

and each θ_j belongs to $\{uuv, uvv\}$ and moreover $\ell \le 80$.

In particular, given a continuation $\tilde{w}\overline{w}$ of w with $\tilde{w} \in \Sigma^{(r)}(3+e^{-r-4})$ and $\overline{w} \in \Sigma^{(10r)}(3+e^{-r-4})$, from the first such factor $(uv)^{s_1}$, the sequence should be

$$(uv)^{s_1}\theta_1(uv)^{s_2}\theta_2\ldots(uv)^{s_\ell}$$

with $\theta_j \in \{uuv, uvv\}$, $\ell \le 80$ and $s_1 + \cdots + s_\ell \le 20r^{1/16}$, so we have in total at most 2^{80} choices for the θ_j, and, given $\ell \le 80$, the number of choices for the s_j is at most the number of natural solutions of $x_1 + \cdots + x_{\ell+1} = \lfloor 20r^{1/16}\rfloor =: M$, which is $\binom{M+\ell}{\ell} < (21r^{1/16})^{80} = 21^{80}r^5$, and so the total number of such words $w\tilde{w}\overline{w}$ is $\mathrm{O}(r^3 \cdot 2^{80} \cdot 80 \cdot 21^{80}r^5) = \mathrm{O}(r^8)$.

Case 2.2: Suppose $|\alpha_t| \le r^{15/16}$ and that the largest factor α_t^s of $w\tilde{w}$ satisfies $\mathsf{r}(\alpha_t^s) \le r - 170|\alpha_t|$.

We claim there is β such that $(\alpha_t, \beta) \in \overline{P}$, $\beta(\alpha_t^{s_1})\beta$ is a subword of $w\tilde{w}$, and the continuation $w\tilde{w}\overline{w}$ has the form $\beta(\alpha_t)^{s_1}\beta(\alpha_t)^{s_2}\ldots(\alpha_t)^{s_\ell}$ with $|s_j - s_{j+1}| \le 1$ for all $1 \le j \le \ell \le 80$. By hypothesis there is $\beta_t(\alpha_t)^s\beta_t$ inside $w\tilde{w}$. If $(\alpha_t, \beta_t) = (uv, v)$

for some $(u,v) \in \bar{P}$ or $(\alpha_t, \beta) = (a, b)$ then we set $\beta = \beta_t$, while if $\beta_t = \alpha_t^k \tilde{\beta}$ with $(\alpha_t, \tilde{\beta}) \in \bar{P}$ and $|\tilde{\beta}| \le |\alpha_t|$, then we set $\beta = \tilde{\beta}$.

If $(\alpha_t, \beta) = (uv, v)$ with $(u,v) \in \bar{P}$ then we use Lemma 4.2 to obtain a continuation $\hat{w}$ with $\hat{\gamma} = (uv)^{s_1}\theta_1(uv)^{s_2}\theta_2 \dots$ with $\theta_i \in \{uuv, uvv\}$ and $\hat{w} = \hat{\gamma} w_2$. Moreover, $\ell \le 80$. Since $\mathrm{r}(\alpha_t^{s_1}) \le r - 170|\alpha_t|$, by induction $\mathrm{r}(\alpha_t^{s_j}) \le r - (172 - 2j)|\alpha_t|$, and since $\ell \le 80$, by the proof of Lemma 4.2 we conclude that all θ_j are equal to $\theta_j = \alpha_t \beta = uvv$ (since $\alpha_t^s \beta_t = (uv)^f v$ for some $f \ge 1$, and uuv and uv start with v_a) and that $|s_j - s_{j+1}| \le 1$ for all $j \ge 1$.

When $\alpha_t = a$, we use Lemma 4.3 to find a continuation $\hat{w}$ such that $\hat{w} = 2^{e_1} b 2^{e_2} b \dots$. But observe that $2^{e_1} = \alpha_t^{e_1/2}$ is inside w, so by hypothesis $\mathrm{r}(2^{e_1}) < r - 170|\alpha_t|$ and there is b before 2^{e_1}, so e_2 is even and $|e_1 - e_2| \in \{0, 2\}$. By induction we obtain $\mathrm{r}(2^{e_j}) \le r - (172 - 2j)|\alpha_t|$, which forces all e_j to be even and $|e_{j+1} - e_j| \in \{0, 2\}$. In this case $\beta = b$, so we get $\hat{w} = 2^{e_1} b 2^{e_2} b \dots = \alpha_t^{s_1} \beta \alpha_t^{s_2} \beta \dots$ with $|s_{j+1} - s_j| \le 1$ for all $j \ge 1$.

Therefore, given a continuation $\tilde{w}\bar{w}$ of w with $\tilde{w} \in \Sigma^{(r)}(3 + e^{-r-4})$ and $\bar{w} \in \Sigma^{(10r)}(3 + e^{-r-4})$, there is β with $(\alpha_t, \beta) \in \bar{P}$ such that $w\tilde{w}$ has a factor $\beta(\alpha_t)^{s_1}\beta$, after which the continuation of $w\tilde{w}\bar{w}$ is a concatenation of at most 79 sequences of the type $(\alpha_t)^{s_j}\beta$, $2 \le j \le 80$, with $|s_{j+1} - s_j| \le 1$ for every $j \ge 1$. This gives at most 3^{80} continuations of $\beta(\alpha_t)^{s_1}\beta$, and so, since we have at most $\mathrm{O}((r^3)^2) = \mathrm{O}(r^6)$ choices for $w\tilde{w}$, we have in total $\mathrm{O}(2^{80} \cdot r^6) = \mathrm{O}(r^6)$ such words $w\tilde{w}\bar{w}$.

In all the previous cases (Case 1, Case 2.1 and Case 2.2), if $d = \frac{\log r - \log\log r}{r}$, then

$$(e^{-10r})^d = e^{-10(\log r - \log\log r)} = \left(\frac{r}{\log r}\right)^{-10}.$$

Moreover, in these cases we have $\mathrm{O}(r^8)$ possible such words $w\tilde{w}\bar{w}$. Notice that

$$r^8 (e^{-10r})^d = r^8 \cdot \left(\frac{r}{\log r}\right)^{-10} = \frac{\log^{10} r}{r^2} < \frac{1}{r} \ll 1.$$

Our third case is derived from Case 2.2, but it is more delicate.

Case 3: Under the conditions of Case 2, suppose that $|\alpha_t| \le r^{15/16}$ and that $w\tilde{w}$ has a factor $\alpha_t^{s_1}$ satisfying $\mathrm{r}(\alpha_t^{s_1}) \ge r - 170|\alpha_t|$.

In this case we will consider continuations $\bar{w} \in \Sigma^{(Tr)}(3 + e^{-r-4})$ for $T = \lfloor \log^2 r \rfloor$ such that $w\tilde{w}\bar{w} \in \Sigma(3 + e^{-r-4}, |w\tilde{w}\bar{w}|)$. Again, consider two subcases depending on the length of α_t.

Case 3.1: Suppose that $|\alpha_t| > 2$.

So, $\alpha_t = uv$ with $(u,v) \in \bar{P}$. Now let $T = \lfloor (\log r)^2 \rfloor$. The condition on $\alpha_t^{s_1}$ implies that $s_1 \ge (r - 170|\alpha_t|)/(2|\alpha_t|) \ge (1/2) r^{1/16} - 85$, so $s_1^2 |\alpha_t|/2 \ge s_1(r - 170 r^{15/16})/4 \ge (\log r)^2 r \ge Tr$. Let $\hat{w} \in \Sigma^{((T+2)r)}(3 + e^{-r-4})$ be the factor of $w\tilde{w}\bar{w}$ starting at the factor $\alpha_t^{s_1}$. Lemma 4.2 guarantees that

$$\hat{w} = (uv)^{s_1}\theta_1(uv)^{s_2} \dots (uv)^{s_\ell}\hat{w}_2$$

where each θ_i is in $\{uuv, uvv\}$ and $\ell \le 2.1\, Tr/|\alpha_t^{s_1}| + 1 \le 5T + 2$ for r large enough.

Therefore, from this factor $\alpha_t^{s_1}$, the continuation of $w\tilde{w}\overline{w}$ is an initial factor of a word of the form $(uv)^{s_1}\theta_1(uv)^{s_2}\theta_2\dots\theta_{\ell-1}(uv)^{s_\ell}$, and

$$Tr \le \mathsf{r}((uv)^{s_1}\theta_1(uv)^{s_2}\theta_2\dots(uv)^{s_\ell}) < (T+3)r$$

with $\theta_j \in \{uuv, uvv\}$, $\ell \le 5T+2$ (and such that $(uv)^{s_1}\theta_1(uv)^{s_2}\theta_2\dots\theta_{\ell-1}$ is an initial factor of this word beginning in this factor $\alpha_t^{s_1}$ and going till the end of $w\tilde{w}\overline{w}$) with $\mathsf{r}(\alpha_t^{s_j}) \ge r-(173+T)|\alpha_t| > r - \lfloor 6\log^2 r \cdot r^{15/16}\rfloor =: M$ (notice that if $\mathsf{r}((uv)^{s_j}) < r-10|\alpha_t|$ then $|s_{j+1}-s_j| \le 1$). Let s_0 be the smallest integer satisfying $\mathsf{r}(\alpha_t^{s_0}) \ge M$. Then $s_j = s_0 + \tilde{s}_j$ with $\tilde{s}_j \ge 0$ for each $1 \le j \le \ell$.

Since $q_{2s|\alpha_t|}(\alpha_t^s) \ge q_{2|\alpha_t|}(\alpha_t)^s$ and $q_{2|\alpha_t|}(\alpha_t) \ge q_4(1122) = 12$, we have $\mathsf{r}(\alpha_t^s) \ge \lfloor\log(q_{2s.|\alpha_t|}(\alpha_t^s)^2)\rfloor \ge \lfloor\log((12^s)^2)\rfloor = \lfloor s\log(144)\rfloor > 4s$. Hence

$$\begin{aligned}(T+3)r > \mathsf{r}((uv)^{s_1}\theta_1(uv)^{s_2}\theta_2\dots(uv)^{s_\ell}) &\ge \ell\,\mathsf{r}(\alpha_t^{s_0}) + (\tilde{s}_1+\dots+\tilde{s}_\ell)\,\mathsf{r}(\alpha_t)\\ &\ge \ell M + 4(\tilde{s}_1+\dots+\tilde{s}_\ell).\end{aligned}$$

In particular, $\ell \le T+3$ for r large. Since $(T+4)M = (T+4)(r-\lfloor 6\log^2 r\cdot r^{15/16}\rfloor) > (T+3)r$, it follows that, given $1 \le \ell \le T+3$, the number of choices of the s_j, $j \le \ell$, is at most the number of natural solutions of

$$\tilde{s}_1+\dots+\tilde{s}_\ell \le (T+4-\ell)M/4,$$

which is

$$\binom{\lfloor (T+4-\ell)M/4\rfloor+\ell}{\ell} \le \left(\frac{e((T+4-\ell)M/4+\ell)}{\ell}\right)^\ell < \left(\frac{e((\tilde{T}-\ell)M/4)}{\ell}\right)^\ell,$$

where $\tilde{T} := T+5$ (here we have used the inequalities $\binom{n}{k} \le n^k/k! \le (en/k)^k$, which hold for $1 \le k \le n$). We have at most 2^ℓ choices for the θ_j, $j \le \ell$; let us estimate the maximum of

$$f(\ell) = 2^\ell\left(\frac{e((\tilde{T}-\ell)M/4)}{\ell}\right)^\ell = \left(\frac{e((\tilde{T}-\ell)M/2)}{\ell}\right)^\ell \quad \text{for } 1 \le \ell \le \tilde{T}-1.$$

The derivative of $\log f(\ell)$ is $\log(\frac{e((\tilde{T}-\ell)M/2)}{\ell}) - \frac{\tilde{T}}{\tilde{T}-\ell}$. Since, in this range of ℓ,

$$\log\left(\frac{e((\tilde{T}-\ell)M/2)}{\ell}\right) = (1+\mathrm{o}(1))\log M = (1+\mathrm{o}(1))\log r$$

and $\tilde{T} = T+5 = \log^2 r + \mathrm{O}(1)$, we have the maximum attained for

$$\ell = \tilde{T}\left(1-\frac{1+\mathrm{o}(1)}{\log r}\right) = \log^2 r - (1+\mathrm{o}(1))\log r < T,$$

and, for this value of ℓ,

$$f(\ell) = \left(\frac{e((\tilde{T}-\ell)M/2)}{\ell}\right)^\ell = \left(\frac{e((1+\mathrm{o}(1))M/2)}{\log r}\right)^\ell < \left(\frac{3M}{2\log r}\right)^{T-(1+\mathrm{o}(1))\log r}.$$

We have at most $\tilde{T} = \log^2 r + \mathrm{O}(1)$ choices for ℓ, and at most $\mathrm{O}(r^6)$ choices for $w\tilde{w}$, so we have at most

$$\mathrm{O}\left(r^6 \log^2 r \left(\frac{3M}{2\log r}\right)^{T-(1+\mathrm{o}(1))\log r}\right) < \left(\frac{2r}{\log r}\right)^{T-(1+\mathrm{o}(1))\log r}$$

such words $w\tilde{w}\overline{w}$.

Notice that, for $d = \frac{\log r - \log\log r}{r}$, we have $(e^{-Tr})^d = e^{-T(\log r - \log\log r)}$, so

$$\left(\frac{2r}{\log r}\right)^T (e^{-Tr})^d = \left(\frac{2r}{\log r} e^{-\log r + \log\log r}\right)^T = 2^T,$$

and

$$\begin{aligned}\left(\frac{2r}{\log r}\right)^{T-(1+\mathrm{o}(1))\log r} (e^{-Tr})^d &\leq \left(\frac{2r}{\log r}\right)^{-(1+\mathrm{o}(1))\log r} \cdot 2^{\log^2 r} \\ &= e^{-(1+\mathrm{o}(1))\log^2 r} \cdot e^{\log 2 \log^2 r} \\ &= e^{-(1-\log 2+\mathrm{o}(1))\log^2 r} \\ &< e^{-\frac{1}{4}\log^2 r} \ll 1.\end{aligned}$$

Case 3.2: Suppose that $|\alpha_t| = 2$. Then, for some $c \in \{1,2\}$, $w\tilde{w}$ has a factor c^{s_1} satisfying $\mathsf{r}(c^{s_1}) \geq r - 340$. Let $c' = 3 - c \in \{1,2\}$ and $\theta = c'c'$. Observe that $(s+1)\log x \geq \mathsf{r}(c^s) \geq r - 340$, so $(T+2)r \leq s^2(\log x)/4$. Using Lemma 4.3, from this factor c^{s_1} the continuation of $w\tilde{w}\overline{w}$ has the form $c^{s_1}\theta c^{s_2}\theta \ldots \theta c^{s_\ell}$ with $\ell < 2Tr/((s-4)\log x) < 2Tr/(r-340) < 3T$ for large r. Using this information in the inequality $\ell(s-\ell-3)\log x < Tr$ gives $\ell \leq T+1$ for large r. Therefore, from this factor c^{s_1}, the continuation of $w\tilde{w}\overline{w}$ is an initial factor of a word of the form $c^{s_1}\theta c^{s_2}\theta \ldots \theta c^{s_\ell}$, $Tr \leq \mathsf{r}(c^{s_1}\theta c^{s_2}\theta \ldots c^{s_\ell}) \leq (T+1)r$, $\ell \leq T+1$ (and such that $c^{s_1}\theta c^{s_2}\theta \ldots c^{s_{\ell-1}}\theta$ is an initial factor of this word beginning with this factor c^{s_1} and going till the end of $w\tilde{w}\overline{w}$) with $\mathsf{r}(c^{s_j}) \geq r - (342 + 2T) > r - \lfloor 3\log^2 r \rfloor =: N$ (notice that if $\mathsf{r}(c^{s_j}) < r - 7$ then s_j is even and $|s_{j+1} - s_j| \in \{0,2\}$). Let s_0 be minimal such that $\mathsf{r}(c^{s_0}) \geq N$. Then $s_j = s_0 + \tilde{s}_j$ with $\tilde{s}_j \geq 0$ for each $1 \leq j \leq \ell$. Notice that, given c, $\overline{w}$ is determined by the choice of $(\ell, s_1, \ldots, s_\ell)$.

To estimate the number of the corresponding possibilities, we will make use of Lemma 4.5. We will consider last two subcases depending on the value of c.

Case 3.2.1: Assume that $c = 2$.

Since $q_s(2^s) \geq 2^s$, we have

$$\mathsf{r}(2^s) \geq \lfloor \log(q_s(2^s)^2) \rfloor \geq \lfloor \log((2^s)^2) \rfloor = \lfloor s\log(4) \rfloor \geq 4s/3 - 1.$$

We have

$$\begin{aligned}(T+1)r > \mathsf{r}(2^{s_1}112^{s_2}11\ldots 2^{s_\ell}) &\geq \ell\,\mathsf{r}(2^{s_0}) + 4(\tilde{s}_1 + \cdots + \tilde{s}_\ell)/3 - \ell \\ &\geq \ell N + 4(\tilde{s}_1 + \cdots + \tilde{s}_\ell)/3 - \ell.\end{aligned}$$

Since

$$(T+2)N = (T+2)(r - \lfloor 3\log^2 r \rfloor) > (T+1)r + (T+1) \geq (T+1)r + \ell,$$

it follows that, given $1 \leq \ell \leq T+1$, the number of choices of the s_j, $j \leq \ell$, is at most the number of natural solutions of $\tilde{s}_1 + \cdots + \tilde{s}_\ell \leq 3(T+2-\ell)N/4$. By Lemma 4.5, it is at most

$$e^{(\log(3N/4-\log\log(3N/4)+\mathrm{o}(1))(T+3)} = e^{(\log N - \log\log N - \log(4/3)+\mathrm{o}(1))T},$$

since $\log N = \mathrm{o}(T)$. We have at most $\mathrm{O}(r^6)$ choices for $w\tilde{w}$, so we have at most

$$\mathrm{O}(r^6 e^{(\log N - \log\log N - \log(4/3)+\mathrm{o}(1))T} = \mathrm{O}(e^{(\log N - \log\log N - \log(4/3)+\mathrm{o}(1))T})$$

such words $w\tilde{w}\overline{w}$.

Notice that, for $d = \frac{\log r - \log\log r}{r}$ we have $(e^{-Tr})^d = e^{-T(\log r - \log\log r)}$, and so, since $\log N = \log r + \mathrm{o}(1)$,

$$\begin{aligned} e^{(\log N - \log\log N - \log(4/3)+\mathrm{o}(1))T}(e^{-Tr})^d &= e^{T(\log r - \log\log r - \log(4/3)+\mathrm{o}(1))}(e^{-Tr})^d \\ &= e^{T(\mathrm{o}(1)-\log(4/3))} < e^{-\frac{\log^2 r}{4}} \ll 1. \end{aligned}$$

Case 3.2.2: Assume that $c = 1$. Observe that

$$(T+1)r > \mathsf{r}(1^{s_1}221^{s_2}22\ldots221^{s_\ell}) \geq \ell N + (\tilde{s}_1 + \cdots + \tilde{s}_\ell)\log\left(\frac{3+\sqrt{5}}{2}\right)$$

by (A.3) and (A.8). Since

$$(T+2)N = (T+2)(r - \lfloor 3\log^2 r \rfloor) > (T+1)r,$$

it follows that, given $1 \leq \ell \leq T+1$, the number of choices of the s_j, $j \leq \ell$, is at most the number of natural solutions of

$$\tilde{s}_1 + \cdots + \tilde{s}_\ell \leq (T+2-\ell)N/\log\left(\frac{3+\sqrt{5}}{2}\right).$$

By Lemma 4.5, it is at most

$$\begin{aligned} &e^{(\log(N/\log(\frac{3+\sqrt{5}}{2}))-\log\log(N/\log(\frac{3+\sqrt{5}}{2}))+\mathrm{o}(1))(T+3)} \\ &\qquad = e^{(\log(N/\log(\frac{3+\sqrt{5}}{2}))-\log\log(N/\log(\frac{3+\sqrt{5}}{2}))+\mathrm{o}(1))T}, \end{aligned}$$

since $\log N = \mathrm{o}(T)$. We have at most $\mathrm{O}(r^6)$ choices for $w\tilde{w}$, so we have at most

$$\begin{aligned} &\mathrm{O}\big(r^6 e^{(\log(N/\log(\frac{3+\sqrt{5}}{2}))-\log\log(N/\log(\frac{3+\sqrt{5}}{2}))+\mathrm{o}(1))T}\big) \\ &\qquad = \mathrm{O}\big(e^{(\log(N/\log(\frac{3+\sqrt{5}}{2}))-\log\log(N/\log(\frac{3+\sqrt{5}}{2}))+\mathrm{o}(1))T}\big) \end{aligned}$$

such words $w\tilde{w}\overline{w}$.

Notice that if $\delta > 0$, for $d = \frac{\log r - \log\log r - \log\log(\frac{3+\sqrt{5}}{2}) + \delta}{r}$ we have

$$(e^{-Tr})^d = e^{-dTr} = e^{-T(\log r - \log\log r - \log\log(\frac{3+\sqrt{5}}{2}) + \delta)},$$

and so, since $\log N = \log r + o(1)$,

$$\begin{aligned} & e^{(\log(N/\log(\frac{3+\sqrt{5}}{2})) - \log\log(N/\log(\frac{3+\sqrt{5}}{2})) + o(1))T} (e^{-Tr})^d \\ &\quad = e^{T(\log r - \log\log r - \log\log(\frac{3+\sqrt{5}}{2}) + o(1) - \log r + \log\log r + \log\log(\frac{3+\sqrt{5}}{2}) - \delta)} \\ &\quad = e^{T(o(1)-\delta)} < e^{-\frac{\delta \log^2 r}{2}} \ll 1. \end{aligned}$$

Since $c_0 := -\log\log(\frac{3+\sqrt{5}}{2}) = 0.03830054\ldots > 0$, it follows that

$$d(3+e^{-r}) \leq 2 \cdot \frac{\log r - \log\log r + c_0 + o(1)}{r}.$$

Up to this point of the proof, we have shown the upper bound

$$d(3+t) \leq 2 \cdot \frac{\log|\log t| - \log\log|\log t| + c_0 + o(1)}{|\log t|},$$

which gives us a different proof of the upper bound on the easier bounds stated in the introduction. In fact, the only case that gives the worst bound is the last one with $c = 1$ (that is, Case 3.2.2).

We can actually obtain a more precise upper estimate by choosing $T = \lfloor r/5 \rfloor$, which is what we will do now. For the sake of exposition, we will consider this improved estimate to be a separate case.

Case 3.2.3: We will derive a more precise estimate for $c = 1$. Observe that it is possible to choose $T = \lfloor r/5 \rfloor$ in Lemma 4.3, because $\mathsf{r}(1^s) \geq r - 340$ gives $s \log x \geq r - 340$, so one has

$$s^2(\log x)/4 \geq (r-340)^2/(4\log x) \geq r^2/5$$

for large r. Moreover,

$$\ell < 2Tr/(s\log x) \leq 2.1\, Tr/(r-340) \leq (5/2)T \leq r/2$$

for large r. Putting this again in the inequality $\ell < Tr/((s-\ell+1)\log x)$ gives further $\ell < 2T+1$, so $\ell \leq 2T$ for large r.

Let $T = \lfloor r/5 \rfloor$. We would have a worst lower estimate for $\mathsf{r}(\alpha_t^{s_i})$: for $i \geq 1$, we have $\mathsf{r}(\alpha_t^{s_i}) \geq r - 2(173+i) \geq r/3$. Indeed,

$$\begin{aligned} r^2/5 + r &\geq \mathsf{r}(1^{s_1}221^{s_2}22\ldots221^{s_\ell}) \\ &\geq \sum_{i=1}^{\min\{\ell, r/2\}} (r - 2(173+i)) \\ &= \min\{\ell, r/2\}\,(r - 347 - \min\{\ell, r/2\}), \end{aligned}$$

which implies $\ell < 3r/10$, and thus $\mathsf{r}(\alpha_t^{s_i}) \geq r - 2(173 + i) > r/3$. We will introduce a parameter j equal to the number of values of i for which $\mathsf{r}(\alpha_t^{s_i}) < r - 3$ in $1^{s_1}221^{s_2}22\dots221^{s_\ell}$, for which we should have $s_{i+1} \in \{s_i, s_i - 2, s_i + 2\}$ (for the other $\ell - j$ values of $1 \leq i \leq \ell$ we have $\mathsf{r}(\alpha_t^{s_i}) \geq r - 3$); if we consider these j values $i_1 < \cdots < i_j$ of i, we have $s_{i_1} \geq s_0 - 100$, so $s_{i_t} > s_0 - 100 - 2t$, $1 \leq t \leq j$, and $\sum_{1 \leq i \leq j} s_{i_t} > j \cdot (s_0 - 100 - j)$.

Let $\hat{\ell} = \ell - j$ and $\{s_i \mid i \in I = \{1, \dots, \ell\} \setminus \{i_t \mid 1 \leq t \leq j\}\} = \{\hat{s}_1, \dots, \hat{s}_{\hat{\ell}}\}$. We have $\ell < 3r/10 < 2T$. Given $\hat{\ell}$ and j there are at most

$$\binom{\ell}{j} = \binom{\hat{\ell} + j}{j} < \left(\frac{e\ell}{j}\right)^j < \left(\frac{2eT}{j}\right)^j$$

choices for the set $\{s_{i_t} \mid 1 \leq t \leq j\}$. Since for $i \in \{i_t \mid 1 \leq t \leq j\}$ we have at most three choices for s_{i+1}, the total number of these choices is at most 3^j. Together with the number of choices for the set $\{s_{i_t} \mid 1 \leq t \leq j\}$, this gives an estimate of $(6eT/j)^j$ for these choices.

Let $\hat{s}_0$ be the smallest integer with $\mathsf{r}(1^{\hat{s}_0}) \geq r - 3$. Then $\hat{s}_0 > (r-5)/\log(\frac{3+\sqrt{5}}{2})$. The number of solutions of the above inequality is at most the number of natural solutions of

$$\begin{aligned} \hat{s}_1 + \cdots + \hat{s}_{\hat{\ell}} &\leq (T + 2 - \hat{\ell})(r-5)/\log\left(\frac{3+\sqrt{5}}{2}\right) - j \cdot (s_0 - 100 - j) \\ &< (T + 2 - \hat{\ell} - j \cdot (r - 104 - j)/r)(r-5)/\log\left(\frac{3+\sqrt{5}}{2}\right) \\ &< (T + 2 - \hat{\ell} - j/2)(r-5)/\log\left(\frac{3+\sqrt{5}}{2}\right). \end{aligned}$$

By Lemma 4.5, the number of solutions of

$$\hat{s}_1 + \cdots + \hat{s}_{\hat{\ell}} \leq (T + 2 - \hat{\ell} - j \cdot (r - 104 - j)/r)(r-5)/\log\left(\frac{3+\sqrt{5}}{2}\right)$$

is at most

$$(T+2)e^{(1-\varepsilon_m)(T+3-j\cdot(r-104-j)/r)/\varepsilon_m},$$

where ε_m is the solution in $(0,1)$ of the equation

$$\log\left(\frac{em\varepsilon}{1-\varepsilon}\right) = \frac{1}{\varepsilon} \quad \text{with } m = (r-5)/\log\left(\frac{3+\sqrt{5}}{2}\right).$$

Since $\frac{em\varepsilon_m}{1-\varepsilon_m} > \frac{r}{2\log r}$, the factor $(\frac{em\varepsilon_m}{1-\varepsilon_m})^{-j\cdot(r-104-j)/r}$ is such that

$$\left(\frac{6eT}{j}\right)^j \left(\frac{em\varepsilon_m}{1-\varepsilon_m}\right)^{-j\cdot(r-104-j)/r} < \left(\frac{6eT}{j}\left(\frac{r}{2\log r}\right)^{-(r-104-j)/r}\right)^j.$$

This is smaller than $(\frac{6eT}{j}(\frac{r}{2\log r})^{-1/2})^j$, and for $j \geq r^{3/4}$ this is $o(1)$ (using $T \leq r/4$). For $10\log r \leq j < r^{3/4}$, the estimate

$$\left(\frac{6eT}{j}\left(\frac{r}{2\log r}\right)^{-(r-104-j)/r}\right)^j$$

will be o(1) since $-(r-104-j)/r < -1 + r^{-1/5}$ and $(\frac{r}{2\log r})^{r^{-1/5}} = 1 + \mathrm{o}(1)$, so the estimate becomes $((3+\mathrm{o}(1))e/10)^{10\log r} = \mathrm{o}(1)$. On the other hand, for $0 \le j < 10\log r$, the estimate $(\frac{6eT}{j}(\frac{r}{2\log r})^{-(r-104-j)/r})^j$ becomes $(\frac{(3+\mathrm{o}(1))e\log r}{j})^j < (\frac{9\log r}{j})^j$. The maximum of the function $v(j) = (\frac{9\log r}{j})^j$ is attained at $j = 9(\log r)/e$, and is equal to $e^{9(\log r)/e} < r^4$. So, using again the fact that we have $\mathrm{O}(r^6)$ choices for $w\tilde{w}$, in any case we get an upper estimate for the total number of words $w\tilde{w}\overline{w}$, which is

$$\mathrm{O}(r^6)\cdot r^4\cdot(T+2)\cdot\left(\frac{em\varepsilon_m}{1-\varepsilon_m}\right)^{(1-\varepsilon_m)(T+3)} = \mathrm{O}(r^{14})\cdot\left(\frac{em\varepsilon_m}{1-\varepsilon_m}\right)^{(1-\varepsilon_m)T}.$$

As before, this gives an upper estimate for the dimension which is

$$\frac{(1-\varepsilon_m)\log\left(\frac{em\varepsilon_m}{1-\varepsilon_m}\right) + \mathrm{O}((\log r)/T)}{r} = \frac{(1-\varepsilon_m)\log\left(\frac{em\varepsilon_m}{1-\varepsilon_m}\right) + \mathrm{O}((\log r)/r)}{r}$$

Since ε_m is the solution in $(0, 1)$ of the equation $\log(\frac{em\varepsilon_m}{1-\varepsilon_m}) = \frac{1}{\varepsilon_m}$ with $m = (r-5)/\log(\frac{3+\sqrt{5}}{2})$, we have $(1-\varepsilon_m)\log(\frac{em\varepsilon_m}{1-\varepsilon_m}) = \frac{1-\varepsilon_m}{\varepsilon_m}$. Writing $z = \frac{1-\varepsilon_m}{\varepsilon_m}$, the equality $\log(\frac{em\varepsilon_m}{1-\varepsilon_m}) = \frac{1}{\varepsilon_m}$ can be written as $\log(\frac{em}{z}) = z + 1$, so $z + \log z = \log m$ and $ze^z = m$, so $z = W(m) = W((r-5)/\log(\frac{3+\sqrt{5}}{2}))$, where W is Lambert's function. Since $W'(x) < 1/x$, we get

$$W\left((r-5)/\log\left(\frac{3+\sqrt{5}}{2}\right)\right) = W\left(r/\log\left(\frac{3+\sqrt{5}}{2}\right)\right) + \mathrm{O}(1/r),$$

and our upper estimate for the dimension is

$$z/r + \mathrm{O}\left(\frac{\log r}{r^2}\right) = W\left(r/\log\left(\frac{3+\sqrt{5}}{2}\right)\right)/r + \mathrm{O}\left(\frac{\log r}{r^2}\right). \qquad \blacksquare$$

5. The lower bound

The statements and definitions below are taken from the third author's work [10].

Definition 5.1. Given $B = \{\beta_1, \dots, \beta_\ell\}$, $\ell \ge 2$, a finite alphabet of finite words $\beta_j \in (\mathbb{N}^*)^{r_j}$, which is primitive (in the sense that β_i does not begin with β_j for all $i \ne j$) then the Gauss–Cantor set $K(B) \subseteq [0, 1]$ associated with B is defined as

$$K(B) := \{[0; \gamma_1, \gamma_2, \dots] \mid \gamma_i \in B\}.$$

The set $K(B)$ is a dynamically defined Cantor set. We will now exhibit its Markov partition and the expanding map which defines it.

For each word $\beta_j \in (\mathbb{N}^*)^{r_j}$, let $I_j = I(\beta_j)$ be the convex hull of $\{[0; \beta_j, \gamma_1, \gamma_2, \dots] \mid \gamma_i \in B\}$ and let $\psi|_{I_j} := G^{r_j}|_{I_j}$ where

$$G(x) = \{1/x\} = 1/x - \lfloor 1/x \rfloor$$

is the Gauss map. This defines an expanding map $\psi\colon I(\beta_1) \cup \dots \cup I(\beta_\ell) \to I$. Let $I = [\min K(B), \max K(B)]$. Then I is the convex hull of $I_1 \cup \dots \cup I_\ell$ and $\psi(I_j) = I$ for every $j \leq \ell$.

Let us describe how to estimate $\dim_{\mathrm{H}}(K(B))$.

According to Palis–Takens [11, Chapter 4], let

$$\lambda_j = \inf \left|\psi'|_{I_j}\right|, \quad \Lambda_j = \sup \left|\psi'|_{I_j}\right|$$

and $\alpha, \beta \geq 0$ be such that

$$\sum_{j=1}^{\ell} \lambda_j^{-\alpha} = 1, \quad \sum_{j=1}^{\ell} \Lambda_j^{-\beta} = 1.$$

Then

$$\beta \leq \dim_{\mathrm{H}}(K(B)) \leq \alpha. \tag{5.1}$$

Let us discuss how to find estimates for α and β.

The iterates of the Gauss map are given explicitly by

$$\psi|_{I_j}(x) = \frac{q_{r_j}^{(j)} x - p_{r_j}^{(j)}}{p_{r_j-1}^{(j)} - q_{r_j-1}^{(j)} x}$$

where $p_k^{(j)}/q_k^{(j)} = [0; b_1^{(j)}, \dots, b_k^{(j)}]$ and $\beta_j = (b_1^{(j)}, \dots, b_{r_j}^{(j)})$. Hence

$$(\psi|_{I_j})'(x) = \frac{(-1)^{r_j-1}}{(p_{r_j-1}^{(j)} - q_{r_j-1}^{(j)} x)^2}.$$

Lemma 5.2. *Let $x = [c_0, c_1, c_2, \dots]$ and $p_n/q_n = [c_0, c_1, \dots, c_n]$. Then*

$$\frac{1}{2q_n q_{n+1}} < \frac{1}{q_n(q_n + q_{n+1})} < \left|x - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}},$$

and therefore

$$\frac{1}{2q_{n+1}} < |q_n x - p_n| < \frac{1}{q_{n+1}}.$$

Lemma 5.2 implies that

$$(q_{r_j}^{(j)})^2 < |(\psi|_{I_j})'(x)| = \frac{1}{(p_{r_j-1}^{(j)} - q_{r_j-1}^{(j)} x)^2} < (2q_{r_j}^{(j)})^2.$$

Thus

$$(q_{r_j}^{(j)})^2 \leq \lambda_j = \inf \left|\psi'|_{I_j}\right| \leq \Lambda_j = \sup \left|\psi'|_{I_j}\right| \leq (2q_{r_j}^{(j)})^2.$$

Let $a = 22$, s the smallest natural number such that $\mathsf{r}(1^s) \geq r$, $k = 2r$, $\beta_1 = 1^k$ and, for $2 \leq j \leq k+1$, $\beta_j = 1^{k+1-j} a\, 1^s = 1^{k+1-j} 22\, 1^s$. Then $B = \{\beta_1, \dots, \beta_{k+1}\}$ is primitive.

The alphabet $B = \{\beta_1, \dots, \beta_{k+1}\}$ as above induces a subshift

$$\Sigma(B) = \{(\gamma_i)_{i\in\mathbb{Z}} \mid \gamma_i \in B\}.$$

Lemma 3.2 implies that, for any $\underline{\theta} \in \Sigma(B)$ and every $n \in \mathbb{Z}$,

$$\lambda(\sigma^n(\underline{\theta})) < 3 + e^{-r}.$$

Recall that if $\alpha = c_1 \dots c_m$ and $\beta = \beta_1 \dots \beta_n$ are finite words, then

$$q_m(\alpha) q_n(\beta) < q_{m+n}(\alpha\beta) < 2 q_m(\alpha) q_n(\beta).$$

The above estimates give $\Lambda_1 = \sup \left|\psi'|_{I(\beta_1)}\right| \leq 4(\frac{1+\sqrt{5}}{2})^{2k}$ and, for $2 \leq j \leq k+1$,

$$\begin{aligned}\Lambda_j = \sup \left|\psi'|_{I(\beta_j)}\right| &\leq 8 \cdot \left(\frac{1+\sqrt{5}}{2}\right)^{2k-2(j-2)} \cdot (10^2 \cdot e^{r+1}) \\ &\leq \left(\frac{1+\sqrt{5}}{2}\right)^{2k-2(j-2)} \cdot e^{r+8}.\end{aligned}$$

Thus, from the above lemma and [10], we conclude that

$$d(3+e^{-r}) \geq \dim_{\mathrm{H}}(\mathsf{m}(\Sigma(B))) = \min\{1, 2\cdot \dim_{\mathrm{H}}(K(B))\} \geq 2\tilde{d},$$

where $\mathsf{m}(\omega) = \sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega))$ denotes the Markov value of $\omega \in \Sigma(B)$, and $\tilde{d}$ is the solution of

$$\left(4\left(\frac{1+\sqrt{5}}{2}\right)^{4r}\right)^{-\tilde{d}} + \sum_{t=0}^{k-1}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{2t}\right)^{-\tilde{d}} \cdot e^{-(r+8)\tilde{d}} = 1.$$

Since $d(3+e^{-r}) = \mathrm{O}(\frac{\log r}{r})$, we also have $\tilde{d} = \mathrm{O}(\frac{\log r}{r}) = \mathrm{o}(1)$. The rest of this section is devoted to finding a lower bound for $\tilde{d}$.

Since $(\frac{1+\sqrt{5}}{2})^4 > e^{3/2}$, we have

$$\left(4\left(\frac{1+\sqrt{5}}{2}\right)^{2k}\right)^{-\tilde{d}} \leq \left(\frac{1+\sqrt{5}}{2}\right)^{-4r\tilde{d}} \leq e^{-\frac{3}{2}r\tilde{d}},$$

and we get

$$e^{-(r+8)\tilde{d}} \cdot \frac{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2k\tilde{d}}}{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}} = 1 - \mathrm{O}(e^{-\frac{3}{2}r\tilde{d}}). \tag{5.2}$$

In particular,

$$\begin{aligned}1 &\geq e^{-(r+8)\tilde{d}} \cdot \frac{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-4\tilde{d}}}{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}} = e^{-(r+8)\tilde{d}} \cdot \left(1 + \left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}\right) \\ &\geq 2e^{-(r+8)\tilde{d}} \cdot \left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}} \geq 2e^{-(r+9)\tilde{d}},\end{aligned}$$

and so $\tilde{d} \geq \frac{\log 2}{r+9} \geq \frac{1}{2r}$. So we have

$$\left(\frac{1+\sqrt{5}}{2}\right)^{-2k\tilde{d}} = \left(\frac{1+\sqrt{5}}{2}\right)^{-4r\tilde{d}} \leq e^{-\frac{3}{2}r\tilde{d}} \leq e^{-3/4} < 1/2$$

and thus

$$1 \geq e^{-(r+8)\tilde{d}} \cdot \frac{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2k\tilde{d}}}{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}} \geq \frac{e^{-(r+8)\tilde{d}}}{2\left(1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}\right)}.$$

Since $\tilde{d} = \mathrm{o}(1)$, writing $c_1 = \log \frac{3+\sqrt{5}}{2} = 0.9624\ldots$ we have

$$\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}} = e^{-c_1\tilde{d}} = 1 - c_1\tilde{d} + \mathrm{O}(\tilde{d}^2),$$

and therefore $1-(\frac{1+\sqrt{5}}{2})^{-2\tilde{d}} = c_1\tilde{d} + \mathrm{O}(\tilde{d}^2) = (1+\mathrm{O}(\tilde{d}))c_1\tilde{d}$. It follows that

$$1 \geq \frac{e^{-(r+8)\tilde{d}}}{2\left(1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}\right)} = \frac{e^{-(r+8)\tilde{d}}}{(2+\mathrm{O}(\tilde{d}))c_1\tilde{d}} \geq \frac{e^{-(r+8)\tilde{d}}}{2\tilde{d}}$$

and thus $0 \geq -(r+8)\tilde{d} - \log 2 - \log \tilde{d}$. It follows that $-r\tilde{d} \leq \log \tilde{d} + \mathrm{O}(1)$, and thus $(\frac{1+\sqrt{5}}{2})^{-4r\tilde{d}} \leq e^{-\frac{3}{2}r\tilde{d}} = \mathrm{O}(\tilde{d}^{3/2})$. From (5.2), we get

$$\begin{aligned} 1 - \mathrm{O}(\tilde{d}^{3/2}) &= e^{-(r+8)\tilde{d}} \cdot \frac{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2k\tilde{d}}}{1-\left(\frac{1+\sqrt{5}}{2}\right)^{-2\tilde{d}}} \\ &= e^{-(r+8)\tilde{d}} \cdot \frac{1-\mathrm{O}(\tilde{d}^{3/2})}{(1+\mathrm{O}(\tilde{d}))c_1\tilde{d}} = (1+\mathrm{O}(\tilde{d}))\frac{e^{-r\tilde{d}}}{c_1\tilde{d}}, \end{aligned}$$

and thus $\mathrm{O}(\tilde{d}^{3/2}) = -r\tilde{d} + \mathrm{O}(\tilde{d}) + c_0 - \log \tilde{d}$ and therefore

$$r\tilde{d} = -\log \tilde{d} + c_0 + \mathrm{O}(\tilde{d}) = |\log \tilde{d}| + c_0 + \mathrm{O}(\tilde{d}), \tag{5.3}$$

where $c_0 = -\log c_1 = 0.03830054\ldots$.

In particular, $r\tilde{d} = (1+\mathrm{O}(1/|\log \tilde{d}|))|\log \tilde{d}| = (1+\mathrm{o}(1))|\log \tilde{d}|$, and thus $\log \tilde{d} + \log r = \log|\log \tilde{d}| + \mathrm{o}(1)$ and

$$\log r = -\log \tilde{d} + \log|\log \tilde{d}| + \mathrm{o}(1) = (1-\mathrm{o}(1))|\log \tilde{d}|.$$

It follows that $|\log \tilde{d}| = (1+\mathrm{o}(1))\log r$ and $\log|\log \tilde{d}| = \log\log r + \mathrm{o}(1)$, and so

$$\log \tilde{d} + \log r = \log|\log \tilde{d}| + \mathrm{o}(1) = \log\log r + \mathrm{o}(1)$$

and $|\log \tilde{d}| = -\log \tilde{d} = \log r - \log\log r + \mathrm{o}(1) = \log r\,(1-(1+\mathrm{o}(1))\log\log r/\log r)$, which implies $\log|\log \tilde{d}| = \log\log r - (1+\mathrm{o}(1))\log\log r/\log r$.

From $r\tilde{d} = (1 + \mathrm{O}(1/|\log \tilde{d}|))|\log \tilde{d}|$ it follows that

$$\log \tilde{d} + \log r = \log|\log \tilde{d}| + \mathrm{O}\left(\frac{1}{|\log \tilde{d}|}\right) = \log|\log \tilde{d}| + \mathrm{O}\left(\frac{1}{|\log r|}\right)$$
$$= \log\log r - \frac{(1+\mathrm{o}(1))\log\log r}{\log r},$$

so $|\log \tilde{d}| = -\log \tilde{d} = \log r - \log\log r + (1+\mathrm{o}(1))\log\log r/\log r$, and from $r\tilde{d} = |\log \tilde{d}| + c_0 + \mathrm{O}(\tilde{d}) = |\log \tilde{d}| + c_0 + \mathrm{O}((\log r)/r)$ we get

$$\tilde{d} = \frac{|\log \tilde{d}| + c_0 + \mathrm{O}((\log r)/r)}{r}$$
$$= \frac{\log r - \log\log r + c_0 + (1+\mathrm{o}(1))\log\log r/\log r}{r}$$
$$> \frac{\log r - \log\log r + c_0}{r},$$

and thus

$$d(3 + e^{-r}) > 2 \cdot \frac{\log r - \log\log r + c_0}{r}.$$

We can give a more precise asymptotic expression for $\tilde{d}$ (and thus for $d(3 + e^{-r})$), using the Lambert function $W\colon [e^{-1}, +\infty) \to [-1, +\infty)$, which is the inverse function of $f\colon [-1, +\infty) \to [e^{-1}, +\infty)$, $f(x) = xe^x$ (which is increasing in $[-1, +\infty)$). Let $g\colon (0, +\infty) \to \mathbb{R}$ be given by $g(x) = rx + \log x$. We have $g(\tilde{d}) = r\tilde{d} + \log \tilde{d} = c_0 + \mathrm{O}(\tilde{d})$. Let $d_0 \in (0, +\infty)$ be the solution of $g(d_0) = c_0$. Since $g'(x) = r + 1/x > r$ for every $x \in (0, +\infty)$, and there exists t between d_0 and $\tilde{d}$ such that $|g(\tilde{d}) - c_0| = |g(\tilde{d}) - g(d_0)| = |g'(t)(\tilde{d} - d_0)| \geq r|\tilde{d} - d_0|$, it follows that

$$|\tilde{d} - d_0| \leq \frac{1}{r}|g(\tilde{d}) - c_0| = \mathrm{O}(\tilde{d}/r) = \mathrm{O}((\log r)/r^2)$$

and $\tilde{d} = d_0 + \mathrm{O}((\log r)/r^2) = (1 + \mathrm{O}(1/r))d_0$. On the other hand, since $rd_0 + \log d_0 = g(d_0) = c_0$, we have $d_0 e^{rd_0} = e^{c_0}$, and so $f(rd_0) = rd_0 e^{rd_0} = re^{c_0}$ and thus $rd_0 = W(re^{c_0})$, which gives a closed expression for d_0: $d_0 = \frac{1}{r}W(re^{c_0})$, from which we get

$$\tilde{d} = \frac{W(re^{c_0})}{r} + \mathrm{O}\left(\frac{\log r}{r^2}\right) = \frac{1 + \mathrm{O}(1/r)}{r} \cdot W(re^{c_0}).$$

(for a detailed discussion of the function W, including its asymptotic expansion, we refer the reader to the work of Corless et al. [2]).

The improved estimates of the previous section (using $T = \lfloor r/5 \rfloor$ in the case of $1^{s_1}221^{s_2}\ldots$) give the same asymptotic expression for $\frac{1}{2}d(3 + e^{-r})$, so the proof of Theorem 1.2 is complete.

6. The error term is optimal

In the case $c = 1$, the Markov values larger than 3 are due to two types of "contradictions" that we analyze as two separate subcases:

Case 1: Words of the form $1^{s_1}221^{2k+1}221^{2k+j}221^{s_2}$, where $2k+1$ is of the order of $\hat{s}_0$, and s_1, s_2 are at least $\hat{s}_0 - 4$. In this case the Markov value associated with the cut $1^{s_1}221^{2k+1}|221^{2k+j}221^{s_2}$ is $3+x$, where

$$\begin{aligned} x &= [0; 1^{2k+1}221^{s_1}\dots] - [0; 1^{2k+2+j}221^{s_2}\dots] \\ &= (1+\mathrm{o}(1))\frac{2(3\varphi-4)}{3\varphi^4}\left(\frac{1}{\varphi^{4k}} + \frac{(-1)^j}{\varphi^{4k+2+2j}}\right), \end{aligned}$$

where $\varphi = \frac{1+\sqrt{5}}{2}$, and so x belongs to an interval of the type

$$\frac{2(3\varphi-4)}{3\varphi^{4k+4}}\left[(1+\mathrm{o}(1))\left(1-\frac{1}{\varphi^4}\right), (1+\mathrm{o}(1))\left(1+\frac{1}{\varphi^2}\right)\right].$$

Indeed, we have

$$\begin{aligned} [0; 1^{2k+1}221^{s_1}\dots] &= [0; 1^{2k+1}22\overline{1}] + \mathrm{O}(\varphi^{-8k}), \\ [0; 1^{2k+2+j}221^{s_2}\dots] &= [0; 1^{2k+2+j}22\overline{1}] + \mathrm{O}(\varphi^{-8k}). \end{aligned}$$

Moreover, denoting the n-th Fibonacci number by F_n,

$$\begin{aligned} [0; 1^n 22\overline{1}] &= \left[0; 1^n, 2 + \frac{1}{2+\varphi^{-1}}\right] = [0; 1^n, 4-\varphi] = \frac{(4-\varphi)\mathrm{F}_n + \mathrm{F}_{n-1}}{(4-\varphi)\mathrm{F}_{n+1} + \mathrm{F}_n} \\ &= \frac{\mathrm{F}_{n-1}/\mathrm{F}_n + (4-\varphi)}{(4-\varphi)\mathrm{F}_{n-1}/\mathrm{F}_n + 5 - \varphi}. \end{aligned}$$

On the other hand, the identity $\frac{au+b}{cu+d} - \frac{av+b}{cv+d} = \frac{(ad-bc)(u-v)}{(cu+d)(cv+d)}$ applied for $a=1$, $b = 4-\varphi$, $c = 4-\varphi$, $d = 5-\varphi$, $u = \mathrm{F}_{2k}/\mathrm{F}_{2k+1}$ and $v = \mathrm{F}_{2k+1+j}/\mathrm{F}_{2k+2+j}$ together with $(cu+d)(cv+d) = (1+\mathrm{o}(1))(c\varphi^{-1}+d)^2 = (1+\mathrm{o}(1))(3\varphi)^2$ gives

$$x = (1+\mathrm{o}(1))\frac{12-6\varphi}{(3\varphi)^2}(v-u) = (1+\mathrm{o}(1))\frac{2}{3\varphi^4}(v-u).$$

In order to estimate $v-u$, let us estimate $\mathrm{F}_n/\mathrm{F}_{n+1} - \varphi^{-1}$: we have

$$\begin{aligned} \frac{\mathrm{F}_n}{\mathrm{F}_{n+1}} - \frac{1}{\varphi} &= \frac{\varphi^n - (-\varphi^{-1})^n}{\varphi^{n+1} - (-\varphi^{-1})^{n+1}} - \frac{1}{\varphi} = (1+\mathrm{o}(1))\frac{(-1)^{n+1}(\varphi+\varphi^{-1})\varphi^{-n}}{\varphi^{n+2}} \\ &= \frac{(-1)^{n+1}(3\varphi - 4 + \mathrm{o}(1))}{\varphi^{2n}}. \end{aligned}$$

Using this for $n = 2k+1+j, n = 2k$ and subtracting, we get the above estimate for x.

Case 2: Words of the form $1^{s_1}221^{2k}221^{2k+3+j}221^{s_2}$, where $2k$ is of the order of $\hat{s}_0$, and s_1, s_2 are at least $\hat{s}_0 - 4$. In this case the Markov value associated with the cut $1^{s_1}221^{2k}22|1^{2k+3+j}221^{s_2}$ is $3+y$, where

$$\begin{aligned} y &= [0; 1^{2k+3+j}221^{s_2}\dots] - [0; 1^{2k+2}221^{s_1}\dots] \\ &= (1+\mathrm{o}(1))\frac{2(3\varphi-4)}{3\varphi^4}\left(\frac{1}{\varphi^{4k+2}} + \frac{(-1)^j}{\varphi^{4k+4+2j}}\right). \end{aligned}$$

The proof of this estimate is analogous to the previous one, applying the above estimate of $\mathrm{F}_n/\mathrm{F}_{n+1} - \varphi^{-1}$ for $n = 2k+1, n = 2k+2+j$ and subtracting.

Hence, y belongs to an interval of the type

$$\frac{2(3\varphi-4)}{3\varphi^{4k+6}}\left[(1+\mathrm{o}(1))\left(1-\frac{1}{\varphi^4}\right),(1+\mathrm{o}(1))\left(1+\frac{1}{\varphi^2}\right)\right].$$

Since

$$1-\frac{1}{\varphi^4} > 0.854 > 0.528 > \left(1+\frac{1}{\varphi^2}\right)\cdot\frac{1}{\varphi^2}$$

and

$$1+\frac{1}{\varphi^2} < 1.382 < 2.236 < \left(1-\frac{1}{\varphi^4}\right)\varphi^2,$$

it follows that, for large k, none of these Markov values belongs to the interval

$$[3+x_k, 3+y_k] = 3 + \frac{2(3\varphi-4)}{3\varphi^{4k+4}}[1.382, 2.236],$$

whose size is comparable to the value of its endpoints, and so there are no sequences of the type $\dots 1^{s_1}221^{s_2}221^{s_3}22\dots$ with $s_j > 3k/2$ for all j whose Markov values belong to $[3+x_k, 3+y_k]$. Indeed, we have the same characterization of sequences of this type whose Markov values are smaller than y_k and whose values are smaller than x_k: for $s = 2k$, if $s_j < s$ then s_j is even and $s_{j-1} - s_j, s_{j+1} - s_j \in \{-2, 0, 2\}$ (and there are no other restrictions).

Let again $s = 2k$ and $T = \lfloor r \log r \rfloor$, where $r = \lfloor |\log y_k| \rfloor$. For each $\tilde{T}$ with $T/2 < \tilde{T} \le T$, let $M(\tilde{T})$ be the number of elements of the set $B(\tilde{T})$ of sequences $1^{s_1}221^{s_2}22\dots221^{s_t}22$ with

$$r\cdot(\tilde{T}-1) < \mathsf{r}(1^{s_1}221^{s_2}22\dots221^{s_t}22) \le r\cdot\tilde{T},$$

$s_j > 3s/4$ for every $j \le t$, $s_1, s_t \ge s$ and such that, for each $j \le t$ with $s_j < s$, s_j is even and $s_{j-1} - s_j, s_{j+1} - s_j \in \{-2, 0, 2\}$. Let $\tilde{d} = \max\{\frac{\log M(\tilde{T})}{r\tilde{T}}\}$. Then $d(3+x_k) \ge 2\tilde{d}$. Indeed, $\mathsf{m}(\Sigma(B(\tilde{T}))) \subseteq \mathcal{M} \cap (-\infty, 3+x_k)$, where $\mathsf{m}(\omega) = \sup_{n\in\mathbb{Z}} \lambda(\sigma^n(\omega))$ denotes the Markov value of $\omega \in \Sigma(B(\tilde{T}))$.

Let us now give upper estimates. Suppose that $w\tilde{w}$ does not have a factor 1^{s_1} satisfying $\mathsf{r}(1^{s_1}) \ge r - 340$ where $r = \lfloor |\log y_k| \rfloor$, and consider an infinite continuation ω of it contained in $\Sigma(3+e^{-r}) \supseteq \Sigma(3+y_k)$. Then the previous discussion provides a continuation $\overline{w} \in \Sigma^{(Tr)}(3+e^{-r})$ for some $T \in \{10, \lfloor \log^2 r \rfloor\}$ depending on $\tilde{w}$ such that $w\tilde{w}\overline{w}$ is the continuation of $w\tilde{w}$ in ω, $w\tilde{w}\overline{w} \in \Sigma(3+e^{-r}, |w\tilde{w}\overline{w}|)$, and the number K of these words $w\tilde{w}\overline{w}$ satisfies $K \cdot e^{-Trd} < 1/r$ for $d = \frac{\log r - \log\log r}{r}$.

Suppose now that $w\tilde{w}$ has a factor 1^{s_1} satisfying $\mathsf{r}(1^{s_1}) \ge r - 340$ where

$$r = \lfloor |\log y_k| \rfloor \in \left(s\log\left(\frac{3+\sqrt{5}}{2}\right), (s+2)\log\left(\frac{3+\sqrt{5}}{2}\right)\right).$$

Let us consider continuations $\overline{w} \in \Sigma^{(m)}(3+e^{-r})$ for some $r^{3/2} < m \le r\lfloor r \log r \rfloor$

such that $w\tilde{w}\overline{w} \in \Sigma(3+e^{-r}, |w\tilde{w}\overline{w}|)$ and the continuation of 1^{s_1} in $w\tilde{w}\overline{w}$ is $1^{s_1}221^{s_2}22\dots 1^{s_t}22$ such that there are at least $r^{15/16}$ values of $i \le t$ with $s_i < s$, and t is minimum with this property. Then $s_t > r - 3r^{15/16}$. We will introduce a parameter j equal to the number of values of $i \le t$ with $s_i < s$; let $i_1 < \dots < i_j$ be those values. We have $j \ge r^{15/16}$. There are at most $\binom{t}{j} < (et/j)^j$ choices for the set $\{i_t \mid 1 \le t \le j\}$. Since for $i \in \{i_v \mid 1 \le v \le j\}$ we have at most three choices for s_{i+1}, the total number of these choices is at most 3^j. Together with the number of choices for the set $\{i_t \mid 1 \le t \le j\}$, this gives an estimate of $(3et/j)^j$ for these choices of the set $\{(i_t, s_{i_t}) \mid 1 \le t \le j\}$. Let $\overline{t} = t - j$. The number of choices of the remaining values of the s_i is at most the number of solutions of $\hat{s}_1 + \dots + \hat{s}_{\overline{t}} \le m/\log(\frac{3+\sqrt{5}}{2}) - s\overline{t} - j(r - 3r^{15/16}) \le (U - \overline{t})s$, where $U = m/(r-2) - j/2 < r\log r$, which is at most $Ue^{U\cdot\frac{\log s}{s}}e^{U\cdot W(s)}$. As before, $e^{W(s)} = (1+\mathrm{o}(1))s/\log s$, and (since $e^{U\cdot\frac{\log s}{s}} \le e^{\frac{r\log r\log s}{s}} = e^{\mathrm{O}(\log^2 s)} = e^{\mathrm{o}(j)}$), the total number $\tilde{K}$ of these sequences is

$$\mathrm{O}\left(r^4\left(\frac{6em}{jr}\right)^j (m/r)((1+\mathrm{o}(1))s/\log s)^{-j/2}e^{U\cdot\frac{\log s}{s}}e^{W(s)m/(r-2)}\right) = \mathrm{O}(s^{-j/4}e^{W(s)m/r}),$$

and since $j \ge r^{15/16}$, for $d = \frac{W(r/\log(\frac{3+\sqrt{5}}{2}))}{r} - \frac{1}{r^{3/2}}$ we have $\tilde{K}\cdot e^{-md} < e^{-\sqrt{r}}$.

Consider now the remaining case where there are less than $r^{15/16}$ values of $i \le t$ with $s_i < s$ and consider the largest continuation of 1^{s_1} in $w\tilde{w}\overline{w} \in \Sigma^{(Tr)}(3+e^{-r})$, $T = \lfloor r\log r\rfloor$ of the form $1^{s_1}221^{s_2}22\dots 1^{s_t}$ with $s_j > s - 3r^{15/16}$ for each j. Taking j_1 minimum and j_2 maximum with $s_{j_1}, s_{j_2} \ge s$ (notice that $j_1 + t - j_2 \le r^{15/16}$), the number $\hat{N}$ of such words is at most $3^{j_1+t-j_2}M < 3^{r^{15/16}}M$, where M is the number of elements of $B(\tilde{T})$, where $r\cdot(\tilde{T}-1) < \mathrm{r}(1^{s_{j_1+1}}221^{s_2}22\dots 1^{s_{j_2-1}}) \le r\cdot\tilde{T}$. We have $\tilde{T} < T - (j_1+t-j_2)/2$ and $M(\tilde{T}) \le e^{r\tilde{d}\tilde{T}} < e^{r\tilde{d}(T-(j_1+t-j_2)/2)}$, so $\hat{N} \le e^{r\tilde{d}T}(3e^{-r\tilde{d}/2})^{j_1+t-j_2}$. Since, by our lower estimates on $d(3+\varepsilon)$, $\frac{\log\hat{N}}{Tr} > \frac{\log r - \log\log r + 0.03}{r}$, it follows that $\tilde{d} \ge \frac{\log M}{r\tilde{T}} > \frac{\log r - \log\log r}{r}$, and thus

$$(3e^{-r\tilde{d}/2})^{j_1+t-j_2} < (3(\log r/r)^{1/2})^{j_1+t-j_2} \le 1;$$

adding these estimates for all possible choices of $(j_1, t-j_2)$, we get $\hat{N} \le 2e^{r\tilde{d}T}$. This, together with the previous estimates, implies that $d(3+y_k) \le 2\tilde{d} + \mathrm{O}(1/r^2)$. Indeed, $(e^{-Tr})^{\tilde{d}+1/r^2} = e^{-T/r}e^{-r\tilde{d}T} < e^{1-\log r}e^{-r\tilde{d}T}$, and thus $2e^{r\tilde{d}T}(e^{-Tr})^{\tilde{d}+1/r^2} \le 2e^{1-\log r} = 2e/r = \mathrm{o}(1)$.

Finally, suppose that F is a twice continuously differentiable function such that

$$d(3+\varepsilon) = F(\varepsilon) + \mathrm{o}\left(\frac{\log|\log\varepsilon|}{|\log\varepsilon|^2}\right).$$

By the mean value theorem there is $\xi_k \in (x_k, y_k)$ such that

$$F'(\xi_k) = \frac{F(y_k) - F(x_k)}{y_k - x_k} = \mathrm{o}\left(\frac{\log|\log y_k|}{y_k|\log y_k|^2}\right).$$

Let $c_1 > 1$ be a constant we will choose later. By Theorem 1.2 we have

$$F(c_1 y_k) - F(y_k) = g_1(c_1 y_k) - g_1(y_k) + \mathrm{O}\left(\frac{\log|\log y_k|}{|\log y_k|^2}\right)$$
$$= (2\log(c_1) + \mathrm{o}(1))\frac{\log|\log y_k|}{|\log y_k|^2} + \mathrm{O}\left(\frac{\log|\log y_k|}{|\log y_k|^2}\right).$$

By choosing $c_1 > 1$ large enough and using the mean value theorem, we obtain $\tilde{\xi}_k \in (y_k, c_1 y_k)$ such that

$$F'(\tilde{\xi}_k) > C \cdot \frac{\log|\log y_k|}{y_k|\log y_k|^2}.$$

Hence for each k, we can find a point in $(\xi_k, \tilde{\xi}_k)$ where the second derivative of F is positive and also a point in $(\tilde{\xi}_\ell, \xi_k)$ (for ℓ large enough) where the second derivative of F is negative.

Appendix A. Basic facts and estimates on continued fractions

Let $\alpha = c_1 \dots c_n \in (\mathbb{N}^*)^n$ be a finite word of length $n > 0$. We define $K(c_1 \dots c_n)$ to be the *continuant* of α, that is, the denominator of the fraction $[0; c_1, \dots, c_n]$. The following lemma can be found in the book by Cusick–Flahive [3, Appendix 2].

Lemma A.1 (Euler's property of continuants). *The continuant $K(c_1 \dots c_n)$ is equal to a sum of certain products of the integers $c_1, \dots, c_n$. Moreover, the products that appear in this sum can be determined in the following way. Start with the product $c_1 \dots c_n$. Now, include all products obtained by removing pairs of adjacent integers. Continue by including all products obtained by removing two separate pairs of adjacent integers, and follow this procedure until no pair remains. Observe that if n is even, then the empty product, equal to* 1, *must also be included.*

As a corollary,

$$K(c_1 \dots c_n) = K(c_1 \dots c_m)K(c_{m+1} \dots c_n) + K(c_1 \dots c_{m-1})K(c_{m+2} \dots c_n)$$

for any $1 \le m < n$.

In particular, the lemma implies that

$$\begin{aligned} K(c_1) &= c_1, \\ K(c_1c_2) &= c_1 \cdot c_2 + 1, \\ K(c_1c_2c_3) &= c_1 \cdot c_2 \cdot c_3 + c_1 + c_3, \\ K(c_1c_2c_3c_4) &= c_1 \cdot c_2 \cdot c_3 \cdot c_4 + c_1 \cdot c_2 + c_1 \cdot c_4 + c_3 \cdot c_4 + 1. \end{aligned}$$

Let $\theta = \theta_1 \dots \theta_n \in (\mathbb{N}^*)^n$, $a = (2,2)$ and $b = (1,1)$. Using Euler's property of continuants we can find a gap between the size of the intervals of the following words:

$$\begin{aligned} \mathsf{s}(b\theta b)^{-1} &\le \left(5 + \frac{2}{\theta_1} + \frac{2}{\theta_n}\right)^2 q_n(\theta)^2, \\ \mathsf{s}(a\theta a)^{-1} &\ge \left(25 + \frac{10}{\theta_1 + 1} + \frac{10}{\theta_n + 1}\right)^2 q_n(\theta)^2. \end{aligned} \tag{A.1}$$

Indeed, using the convention $q_0 = 1$ and $q_{-1} = 0$ we have

$$\begin{aligned} q_{n+4}(b\theta b) &= 4q_n(\theta) + 2q_{n-1}(\theta_1 \dots \theta_{n-1}) + 2q_{n-1}(\theta_2 \dots \theta_n) + q_{n-2}(\theta_2 \dots \theta_{n-1}) \\ &\le \left(5 + \frac{2}{\theta_1} + \frac{2}{\theta_n}\right) q_n(\theta), \\ q_{n+4}(a\theta a) &= 25q_n(\theta) + 10q_{n-1}(\theta_1 \dots \theta_{n-1}) + 10q_{n-1}(\theta_2 \dots \theta_n) + 4q_{n-2}(\theta_2 \dots \theta_{n-1}) \\ &\ge \left(25 + \frac{10}{\theta_1 + 1} + \frac{10}{\theta_n + 1}\right) q_n(\theta), \end{aligned}$$

and finally we use $q_m(a_1 \dots a_m)^2 \le \mathsf{s}(a_1 \dots a_m) \le 2q_m(a_1 \dots a_m)^2$.

Lemma A.2. *Let w be a nonempty finite word in* 1 *and* 2 *of length $n \in \mathbb{N}^*$. Then*

$$(n-3)\log\left(\frac{3+\sqrt{5}}{2}\right) \le \mathsf{r}(w) \le (n+1)\log(3+2\sqrt{2}).$$

Proof. Given $\alpha = c_1 \dots c_n \in (\mathbb{N}^*)^n$, we have

$$\mathsf{s}(\alpha) = \frac{1}{q_n(q_n + q_{n-1})},$$

so $\mathsf{s}(\alpha)$ is minimized when q_n and q_{n-1} are maximized, and maximized when q_n and q_{n-1} are minimized. This happens, respectively, when $q_n = \mathrm{P}_n$ (where P_n is the n-th Pell number) and where $q_n = \mathrm{F}_n$ (where F_n is the n-th Fibonacci number). Hence,

$$\mathsf{r}(1^n) \le \mathsf{r}(w) \le \mathsf{r}(2^n).$$

Moreover,

$$\begin{aligned} \mathsf{s}(1^n)^{-1} &= \mathrm{F}_{n+1}(\mathrm{F}_{n+1} + \mathrm{F}_n) \\ &= -\frac{1}{5}(-1)^{n+1} + \frac{\sqrt{5}+1}{10}\left(\frac{3+\sqrt{5}}{2}\right)^{n+1} - \frac{\sqrt{5}-1}{10}\left(\frac{3-\sqrt{5}}{2}\right)^{n+1} \\ &\ge -\frac{1}{5} + \frac{\sqrt{5}+1}{10}\left(\frac{3+\sqrt{5}}{2}\right)^{n+1} - \frac{\sqrt{5}-1}{2} \\ &= \frac{\sqrt{5}+1}{10}\left(\left(\frac{3+\sqrt{5}}{2}\right)^{n+1} - 1\right) \ge \left(\frac{3+\sqrt{5}}{2}\right)^{n-1}, \end{aligned} \tag{A.2}$$

and, on the other hand,

$$\begin{aligned} \mathsf{s}(2^n)^{-1} &= \mathrm{P}_n(\mathrm{P}_n + \mathrm{P}_{n-1}) = \frac{(3+2\sqrt{2})^n - (3-2\sqrt{2})^n}{4\sqrt{2}} \\ &\le \frac{(3+2\sqrt{2})^{n+1}}{4\sqrt{2}} \le (3+2\sqrt{2})^{n+1}. \end{aligned}$$

Thus,

$$(n-1)\log\left(\frac{3+\sqrt{5}}{2}\right) \le \log \mathsf{s}(w)^{-1} \le (n+1)\log(3+2\sqrt{2}).$$

Finally, since $2\log(\frac{3+\sqrt{5}}{2}) > 1$, we get

$$(n-3)\log\left(\frac{3+\sqrt{5}}{2}\right) \le \mathsf{r}(w) = \lfloor \log \mathsf{s}(w)^{-1} \rfloor \le (n+1)\log(3+2\sqrt{2}). \qquad \blacksquare$$

Lemma A.3. *Let w be a finite word and let v be a factor of w. Then $\mathsf{s}(w) \le \mathsf{s}(v)$ and $\mathsf{r}(w) \ge \mathsf{r}(v)$.*

Proof. Assume first that v is a prefix of w, so $w = v\beta$ for some word β. Then $\mathsf{s}(w) = \mathsf{s}(v\beta) = |I(v\beta)| \le |I(v)| = \mathsf{s}(v)$, since, by definition, $I(v\beta) \subseteq I(v)$.

Assume now that $w = \alpha v \beta$ for some words α, β, where α is nonempty. Then $\mathsf{s}(w) = \mathsf{s}(\alpha v \beta) \le \mathsf{s}(\alpha v) < 2\mathsf{s}(\alpha)\mathsf{s}(v)$. Moreover, if α starts with the letter c, then $\mathsf{s}(\alpha) \le \mathsf{s}(c)$. Since $\mathsf{s}(c) = 1/(c^2+c)$, we have $\mathsf{s}(c) \le 1/2$. We obtain $\mathsf{s}(w) < \mathsf{s}(v)$, as desired. $\blacksquare$

A property that is useful to simplify some computations is

$$\mathsf{r}(w_1 k_1 k_2 w_2) \ge \mathsf{r}(w_1) + \mathsf{r}(w_2)$$

for any positive integers such that $(k_1, k_2) \ne (1,1)$ and any words w_1, w_2. Indeed, it follows from

$$\mathsf{s}(w_1 k_1 k_2 w_2) \le 4\mathsf{s}(k_1 k_2)\mathsf{s}(w_1)\mathsf{s}(w_2) \le \mathsf{s}(w_1)\mathsf{s}(w_2)/3.$$

For $(k_1, k_2) = (1,1)$ we have $\mathsf{r}(w_1 b w_2) \ge \mathsf{r}(w_1) + \mathsf{r}(w_2) - 1$, since $\mathsf{r}(b) = 1$.

Nevertheless, we will prove some sharper bounds that we use to get cleaner statements of the lemmas.

Let $s_1, \dots, s_\ell$ be nonnegative integers with $\ell \ge 2$. We will show that

$$\mathsf{r}(1^{s_1} 2 2 1^{s_2} 2 2 \dots 2 2 1^{s_\ell}) \ge (s_1 + \dots + s_\ell + 3(\ell-2)) \log\left(\frac{3+\sqrt{5}}{2}\right), \tag{A.3}$$

$$\mathsf{r}(2^{s_1} 1 1 2^{s_2} 1 1 \dots 1 1 2^{s_\ell}) \ge (s_1 + \dots + s_\ell + \ell - 2) \log(3 + 2\sqrt{2}). \tag{A.4}$$

First, we will show inductively that

$$q(1^{s_1} 2 2 1^{s_2} 2 2 \dots 2 2 1^{s_\ell}) \ge \mathrm{F}_{s_1 + \dots + s_\ell + 3(\ell-1) + 1}, \tag{A.5}$$

$$q(2^{s_1} 1 1 2^{s_2} 1 1 \dots 1 1 2^{s_\ell}) \ge \mathrm{P}_{s_1 + \dots + s_\ell + \ell}. \tag{A.6}$$

By Euler's property of continuants (Lemma A.1),

$$q(1^{s_1} 2 2 1^{s_2}) = q(1^{s_1}) q(2 2 1^{s_2}) + q(1^{s_1 - 1}) q(2 1^{s_2}).$$

Since $q(1^s) = \mathrm{F}_{s+1}$, one has

$$q(2 2 1^s) = 5q(1^s) + 2q(1^{s-1}) = 5\mathrm{F}_{s+1} + 2\mathrm{F}_s = 3\mathrm{F}_{s+1} + 2\mathrm{F}_{s+2},$$

$$q(2 1^s) = 2q(1^s) + q(1^{s-1}) = 2\mathrm{F}_{s+1} + \mathrm{F}_s.$$

From the identity

$$\mathrm{F}_n \mathrm{F}_m + \mathrm{F}_{n-1} \mathrm{F}_{m-1} = \mathrm{F}_{n+m-1},$$

we get

$$\begin{aligned} \mathrm{F}_{n+1}q(221^m) + \mathrm{F}_n q(21^m) &= \mathrm{F}_{n+1}(2\mathrm{F}_{m+2} + 3\mathrm{F}_{m+1}) + \mathrm{F}_n(2\mathrm{F}_{m+1} + \mathrm{F}_m) \\ &= 2\mathrm{F}_{n+m+2} + \mathrm{F}_{n+m+1} + 2\mathrm{F}_{n+1}\mathrm{F}_{m+1} \\ &= \mathrm{F}_{n+m+4} + 2\mathrm{F}_{n+1}\mathrm{F}_{m+1}. \end{aligned} \tag{A.7}$$

Thus

$$\begin{aligned} q(1^{s_1}221^{s_2}) &= q(1^{s_1})q(221^{s_2}) + q(1^{s_1-1})q(21^{s_2}) \\ &= \mathrm{F}_{s_1+1}q(221^{s_2}) + \mathrm{F}_{s_1}q(21^{s_2}) \\ &= \mathrm{F}_{s_1+s_2+4} + 2\mathrm{F}_{s_1+1}\mathrm{F}_{s_2+1}. \end{aligned}$$

Hence (A.5) is true for $\ell = 2$. Assuming it for ℓ, we use (A.7) with $n = s_1 + \cdots + s_\ell + 3(\ell - 1) + 1$ and $m = s_{\ell+1}$ to obtain

$$\begin{aligned} q(1^{s_1}221^{s_2}22\ldots221^{s_{\ell+1}}) &= q(1^{s_1}221^{s_2}22\ldots221^{s_\ell})q(221^{s_{\ell+1}}) \\ &\quad + q(1^{s_1}221^{s_2}22\ldots221^{s_\ell-1})q(21^{s_{\ell+1}}) \\ &\geq \mathrm{F}_n q(221^{s_{\ell+1}}) + \mathrm{F}_{n-1}q(21^{s_{\ell+1}}) \\ &\geq \mathrm{F}_{s_1+\cdots+s_{\ell+1}+3\ell+1}. \end{aligned}$$

Finally, using (A.2) we obtain

$$\begin{aligned} \mathsf{s}(1^{s_1}221^{s_2}22\ldots221^{s_\ell})^{-1} &\geq \mathrm{F}_{s_1+\cdots+s_\ell+3(\ell-1)+1} \\ &\quad + (\mathrm{F}_{s_1+\cdots+s_\ell+3(\ell-1)+1} + \mathrm{F}_{s_1+\cdots+s_\ell+3(\ell-1)}) \\ &\geq \left(\frac{3+\sqrt{5}}{2}\right)^{s_1+\cdots+s_\ell+3(\ell-1)-1}. \end{aligned}$$

On the other hand, using $\mathrm{F}_{n+2} \leq 3\mathrm{F}_n$ we get

$$\mathsf{s}(1^n)^{-1} = \mathrm{F}_{n+1}\mathrm{F}_{n+2} \leq \frac{3}{4}\mathrm{F}_{2n+2} \leq \left(\frac{3+\sqrt{5}}{2}\right)^n,$$

so

$$\mathsf{r}(1^n) \leq n\log((3+\sqrt{5})/2). \tag{A.8}$$

Similarly, one has $q(2^s) = \mathrm{P}_{s+1}$ and $q(112^s) = \mathrm{P}_{s+2}$. The Pell numbers also satisfy the identity

$$\mathrm{P}_n\mathrm{P}_m + \mathrm{P}_{n-1}\mathrm{P}_{m-1} = \mathrm{P}_{n+m-1}.$$

Hence

$$\mathrm{P}_{n+1}q(112^m) + \mathrm{P}_n q(12^m) = \mathrm{P}_{n+1}\mathrm{P}_{m+2} + \mathrm{P}_n(\mathrm{P}_m + \mathrm{P}_{m+1}) = \mathrm{P}_{n+m+2} + \mathrm{P}_n\mathrm{P}_m.$$

Therefore by induction

$$\begin{aligned} q(2^{s_1}112^{s_2}11\ldots112^{s_{\ell+1}}) &\geq \mathrm{P}_{s_1+\cdots+s_\ell+\ell}q(112^{s_{\ell+1}}) + \mathrm{P}_{s_1+\cdots+s_\ell+\ell-1}q(12^{s_{\ell+1}}) \\ &\geq \mathrm{P}_{s_1+\cdots+s_\ell+s_{\ell+1}+(\ell+1)} \end{aligned}$$

Finally, to show (A.4) we use

$$\mathsf{s}(2^{s_1}112^{s_2}11\dots112^{s_\ell})^{-1} \geq \mathrm{P}_n^2 \geq 4(3+2\sqrt{2})^{n-2}$$

where $n = s_1 + \dots + s_\ell + \ell$.

Lemma A.4. *If* $\alpha = c_1 \dots c_n \in (\mathbb{N}^*)^n$ *then*

$$\frac{[1;c_n+1]}{[1;c_1]} \leq \frac{\mathsf{s}(\alpha^*)}{\mathsf{s}(\alpha)} \leq \frac{[1;c_n]}{[1;c_1+1]},$$

and hence

$$-\log\Big(1+\frac{1}{c_n+1}\Big) - 1 \leq \mathsf{r}(\alpha) - \mathsf{r}(\alpha^*) \leq \log\Big(1+\frac{1}{c_n}\Big) + 1.$$

Proof. By Euler's property of continuants (Lemma A.1) we have $q_n(c_1 \dots c_n) = q_n(c_n \dots c_1)$, and thus

$$\begin{aligned}\mathsf{s}(\alpha^*)^{-1} &= q_n(c_n\dots c_1)(q_n(c_n\dots c_1) + q_{n-1}(c_n\dots c_2)) \\ &\geq \Big(1+\frac{1}{c_1+1}\Big)q_n(c_n\dots c_1)^2 = \Big(1+\frac{1}{c_1+1}\Big)q_n(c_1\dots c_n)^2,\end{aligned}$$

while

$$\mathsf{s}(\alpha)^{-1} = q_n(c_1\dots c_n)(q_n(c_1\dots c_n) + q_{n-1}(c_1\dots c_{n-1})) \leq \Big(1+\frac{1}{c_n}\Big)q_n(c_1\dots c_n)^2.$$

Hence

$$\frac{\mathsf{s}(\alpha^*)}{\mathsf{s}(\alpha)} \leq \frac{[1;c_n]}{[1;c_1+1]}.$$

By symmetry we obtain the lower bound. ■

Acknowledgments. We would like to thank Carlos Matheus and Jamerson Bezerra for helpful conversations about the subject of this paper. We also thank Moubariz Garaev, Harald Helfgott and Lola Thompson for organizing and inviting us to the meeting Number Theory in the Americas/Teoría de Números en América held in Casa Matemática Oaxaca, in 2019, where this work started.

We also thank the anonymous referees for their helpful and insightful comments that greatly improved the exposition of this article.

Funding. The first author is partially supported by CAPES. The second author was supported by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for centers of excellence from ANID-Chile, the MATH-AmSud 21-MATH-07 grant, and by ANID-Chile through the FONDECYT Iniciación 11190034 grant. The third author is partially supported by CNPq and FAPERJ. The fourth author is partially supported by FAPERJ-Bolsa Jovem Cientista do Nosso Estado No. E-26/201.432/2022.

References

[1] Bombieri, E.: Continued fractions and the Markoff tree. Expo. Math. **25**, 187–213 (2007) Zbl 1153.11030 MR 2345177

[2] Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., Knuth, D. E.: On the Lambert W function. Adv. Comput. Math. **5**, 329–359 (1996) Zbl 0863.65008 MR 1414285

[3] Cusick, T. W., Flahive, M. E.: The Markoff and Lagrange spectra. Math. Surveys Monogr. 30, American Mathematical Society, Providence, RI (1989) Zbl 0685.10023 MR 1010419

[4] Freĭman, G. A.: Non-coincidence of the spectra of Markov and of Lagrange. Mat. Zametki **3**, 195–200 (1968) (in Russian) Zbl 0155.09602 MR 0227110

[5] Freĭman, G. A.: Diophantine approximations and the geometry of numbers (Markov's problem). Kalinin. Gosudarstv. Univ., Kalinin (1975) (in Russian) Zbl 0347.10025 MR 0485714

[6] Hall, M., Jr.: On the sum and product of continued fractions. Ann. of Math. (2) **48**, 966–993 (1947) Zbl 0030.02201 MR 0022568

[7] Lima, D., Matheus, C., Moreira, C. G., Romaña, S.: Classical and dynamical Markov and Lagrange spectra—dynamical, fractal and arithmetic aspects. World Scientific Publishing, Hackensack, NJ (2021) Zbl 1484.11001 MR 4274593

[8] Markoff, A.: Sur les formes quadratiques binaires indéfinies. Math. Ann. **17**, 379–399 (1880) JFM 12.0143.02 MR 1510073

[9] Matheus, C., Moreira, C. G.: Fractal geometry of the complement of Lagrange spectrum in Markov spectrum. Comment. Math. Helv. **95**, 593–633 (2020) Zbl 1465.11165 MR 4152626

[10] Moreira, C. G.: Geometric properties of the Markov and Lagrange spectra. Ann. of Math. (2) **188**, 145–170 (2018) Zbl 1404.11095 MR 3815461

[11] Palis, J., Takens, F.: Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations. Cambridge Stud. Adv. Math. 35, Cambridge University Press, Cambridge (1993) Zbl 0790.58014 MR 1237641