FRACTAL DIMENSIONS OF THE MARKOV AND LAGRANGE SPECTRA NEAR 3

HAROLD ERAZO, RODOLFO GUTIÉRREZ-ROMO, CARLOS GUSTAVO MOREIRA, AND SERGIO ROMAÑA

Abstract. The Lagrange spectrum $L$ and Markov spectrum $M$ 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 dimension of the intersection of these sets with any half-line coincide, that is, $\dim_H(L \cap (-\infty, t)) = \dim_H(M \cap (-\infty, t)) \equiv d(t)$ 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 $d(3 + \varepsilon) = 2 \cdot W(e^{c_0 |\log \varepsilon|}) + \mathcal{O}\left(\frac{|\log \varepsilon|^2}{|\log \varepsilon|^2}\right)$, 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 $d(3 + \varepsilon) = F(t) + \mathcal{O}\left(\frac{|\log \varepsilon|^2}{|\log \varepsilon|^2}\right)$, then its second derivative $F''(t)$ changes sign infinitely many times as $t$ approaches 0.

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.

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

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

The previous result is not tight. Indeed, Hurwitz’s theorem states that the following holds 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 number $x$. 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 \( (a_n)_{n \geq 0} \) such that

\[
x = [a_0; a_1, a_2, a_3, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}},
\]

that is, \( (a_n)_{n \geq 0} \) 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 \( \alpha_{n+1} = [a_{n+1}; a_{n+2}, a_{n+3}, \ldots] \), \( \beta_n = [0; a_n, a_{n-1}, \ldots, a_1] \), and where \( p_n/q_n = [a_0; a_1, a_2, \ldots, a_n] \). 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 \( |x - \frac{p}{q}| < \frac{1}{2q^2} \), 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 \( \beta'_{n+1} = [0; a_n, a_{n-1}, \ldots, a_1, 1, \ldots, 1, \ldots] \), we also have that

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

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

It follows that

\[
\mathcal{L} = \left\{ \limsup_{n \to \infty} \lambda(\sigma^n(\omega)) \mid \omega \in (\mathbb{N}^+)^\mathbb{Z} \right\},
\]

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 [Bomb07] and to the books by Cusick and Flahive [CF89], and by Lima, Matheus, Moreira and Romaña [Lim+21] for a more detailed account on these constructions.
1.2. The Markov spectrum. The Markov spectrum is another fractal subset of the real line which is very closely related to the Lagrange spectrum. Using the symbolic-dynamical definition of the Lagrange spectrum a starting point, it can be defined similarly as

\[ \mathcal{M} = \left\{ \sup_{n \in \mathbb{Z}} \lambda (\sigma^n(\omega)) \mid \omega \in (\mathbb{N}^*)^\mathbb{Z} \right\}. \]

We denote by \( m(\omega) = \sup_{n \in \mathbb{Z}} \lambda (\sigma^n(\omega)) \) 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 [Mar80]. In particular, it is well-known that

\[ \mathcal{L} \cap [0, 3) = \mathcal{M} \cap [0, 3) = \left\{ \sqrt{5} < \sqrt{8} < \frac{221}{5} < \cdots \right\}, \]

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 [Bom07, 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; Freiman disproved this conjecture only in 1968 [Fre68]. 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 [MM20].

Even if the previous paragraph suggests that these sets are somewhat different, they are known to coincide before 3 after large enough values. Indeed, Hall showed in 1947 that \( \mathcal{L} \) (and thus also \( \mathcal{M} \)) contains a half-line \([c, \infty)\) [Hal47]; any such ray is hence known as a Hall ray. After several years, Freiman found the largest Hall ray to be \([c_F, \infty)\), where \( c_F \approx 4.5278 \ldots \) is an explicit quadratic surd known as Freiman’s constant [Fre73]. These results in turn imply that \( \mathcal{L} \) and \( \mathcal{M} \) coincide starting at \( c_F \), so they both contain the half-line \([c_F, \infty)\).

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

\[ \dim_H(\mathcal{L} \cup (-\infty, t)) = \dim_H(\mathcal{M} \cup (-\infty, t)) \]

for every \( t > 0 \) [Mor18]. Clearly, this result shows that, when studying the Hausdorff dimension of such truncated versions, one can choose to use either \( \mathcal{L} \) or \( \mathcal{M} \).
1.4. The Hausdorff dimension near 3. Let
\[ d(t) := \dim_H(\mathcal{L} \cup (-\infty, t)) = \dim_H(\mathcal{M} \cup (-\infty, t)). \]

The goal of this article is to determine the behavior of \( d(t) \) near \( t = 3 \). By work of the third author [Mor18], we have that \( d(t) > 0 \) for every \( t > 3 \). On the contrary, \( d(t) = 0 \) for every \( t \leq 3 \), as \( \mathcal{L} \cap (-\infty, 3] = \mathcal{M} \cap (-\infty, 3] \) is at most countable.

Our main objective is to determine the modulus of continuity of \( d(t) \) near \( t = 3 \). 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
\[
C_1 \frac{\log |\log \varepsilon|}{|\log \varepsilon|} \leq d(3 + \varepsilon) \leq C_2 \frac{\log |\log \varepsilon|}{|\log \varepsilon|}.
\]

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 [Bom07] to (factors of) sequences with Markov value slightly larger than 3. In this way, we can analyze the sequences \( \omega \in \{1, 2\}^\mathbb{Z} \subseteq (\mathbb{N}^\ast)^\mathbb{Z} \) 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 \( \Sigma(t) = \{ \omega \in (\mathbb{N}^\ast)^\mathbb{Z} \mid \sup_{n \in \mathbb{Z}} A(\sigma^n(\omega)) \leq t \} \). 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 \( B = 6^3 = 216 \).

The previous theorem can be interpreted as follows: given a bi-infinite word, whose Markov value is exponentially close to 3 (smaller than \( 3 + B^{-n} = 3 + 6^{-3n} \)), then its length-\( n \) subwords are indistinguishable from those in \( \Sigma(3, n) \). That is to say, 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 3.19). The construction is inductive, so we will develop it as a renormalization algorithm (Lemma 3.20). This gives us 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 3.14 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.

To show that the lower bound holds, we prove that \( d(3 + e^{-r}) \) (where \( r \in \mathbb{N}^\ast \)) 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 [PT93, Chapter 4], and, hence the proof of (1.1) is complete.

While these methods are enough to prove inequalities (1.1), they are actually sufficient to obtain an asymptotic approximation of $d(t)$. In fact, to prove (1.1), 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 $f_0 : [-1, +\infty) \to [-e^{-1}, +\infty)$ be given by $f_0(x) = xe^x$ and recall that the Lambert $W$ function is the function $W : [-e^{-1}, +\infty) \to [-1, +\infty)$ given by $W = f_0^{-1}$. Our main result is the following:

**Theorem 1.2.** Let $d(t) = \dim_H(L \cap [0, t)) = \dim_H(M \cap [0, t))$. Then for all, sufficiently small $\varepsilon$, we have

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

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 with those before 3 (that are well understood because of the work of Bombieri [Bom07]), 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 the Markov spectrum 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 to the dimension of these sets, namely configurations obtained by alternate concatenations of large blocks of 1’s with blocks 22.

One natural follow-up question is if 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 reasonable approximations, the error term is optimal. We prove the following:

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

$$d(3 + y_k) - d(3 + x_k) = O \left( \frac{1}{k^2} \right).$$
In particular, if \( F \) is a twice continuously-differentiable function satisfying

\[
\frac{d(3 + \varepsilon)}{\log \varepsilon} = F(\varepsilon) + o\left(\frac{\log \varepsilon}{\log \varepsilon^2}\right),
\]

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

In fact, we will prove that

\[
\frac{W(\varepsilon^0 \log y_k)}{|\log y_k|} - \frac{W(\varepsilon^0 \log x_k)}{|\log x_k|} > \tilde{c} \frac{\log k}{k^2},
\]

for a positive constant \( \tilde{c} \), which implies that the error term in the approximation of \( d(3 + \varepsilon) \) by any reasonable function of \( \varepsilon \) is at least of the order of \( \frac{\log \varepsilon}{|\log \varepsilon|} \).

In this sense, \((3 + x_k, 3 + y_k)\) is an “almost plateau” for the dimension function \( d(i) \) (the variation of \( d(i) \) 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

\[
g_1(\varepsilon) = 2 \cdot \frac{W(\varepsilon^0 |\log \varepsilon|)}{|\log \varepsilon|},
\]

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

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

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 [Mor18] 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 on 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 allow us to prove the optimality of our approximation in Theorem 1.3.

Acknowledgements: We would like to thank Carlos Matheus and Jamerson Bezerra for helpful conversations about the subject of this paper. We also would like to thank Moubariz Garaev, Harald Helfgott and Lola Thompson for organizing and inviting us to the meeting Number Theory in the Americas/Teoría
2. Preliminaries

Our goal is to study the function

\[ d(t) := \dim_{\mathbb{H}}(\mathcal{L} \cup (-\infty, t)) = \dim_{\mathbb{H}}(\mathcal{M} \cup (-\infty, t)) \]

near \( t = 3 \). If a sequence \( \omega \in (\mathbb{N}^*)^2 \) 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. 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 \( \omega = P^*|Q \) can be interpreted as a shift of the original word.

Given a finite word \( u \), we define

\[ p_u = \begin{cases} 12 & |u| \text{ is even} \\ 21 & |u| \text{ is odd} \end{cases}, \quad q_u = \begin{cases} 21 & |u| \text{ is even} \\ 12 & |u| \text{ is odd} \end{cases}. \]

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

\[ \lambda^+(w) = [vp_u^\infty] + [0uq_v^\infty], \quad \lambda^-(w) = [vq_u^\infty] + [0uq_v^\infty]. \]

These quantities are the largest and smallest values of \( \lambda \) that a section of an infinite word containing \( w \) can attain, respectively.

We take \( a = 22 \) and \( b = 11 \).

We will set some notation that will be used throughout all the article; some of it was borrowed from the third author’s work [Mor18]. Given a finite sequence \( \alpha = (a_1, a_2, \ldots, a_n) \in (\mathbb{N}^*)^n \), we define its size by \( s(\alpha) := |I(\alpha)| \), where \( I(\alpha) \) is the interval

\[ \{x \in [0, 1] \mid x = [0; a_1, a_2, \ldots, a_n, a_{n+1}], a_{n+1} \geq 1 \}. \]

If we take \( p_0 = 0, q_0 = 1, p_1 = 1, q_1 = a_1 \) and, for each integer \( k \geq 0 \), we take \( p_{k+2} = a_{k+2}p_{k+1} + p_k \) and \( q_{k+2} = a_{k+2}q_{k+1} + q_k \), then \( I(\alpha) \) is the interval with endpoints \([0; a_1, a_2, \ldots, a_n] = p_n/q_n \) and \([0; a_1, a_2, \ldots, a_{n-1}, a_n+1] = p_{n+1}q_n/q_{n+1} \). Thus,

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

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

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

\[ P_r = \{ \alpha = (a_1, a_2, \ldots, a_n) \mid r(\alpha) \geq r, r((a_1, a_2, \ldots, a_{n-1})) < r \}. \]
Let us recall some estimates from the third author’s work [Mor18] that will be useful for us. Indeed, for any finite words \( \alpha, \beta \), we have that
\[
\frac{1}{2} s(\alpha) s(\beta) < s(\alpha \beta) < 2 s(\alpha) s(\beta);
\]
and it follows that \( r(\alpha) + r(\beta) - 1 \leq r(\alpha \beta) \leq r(\alpha) + r(\beta) + 2 \) [Mor18, Lemma A.2]. By Euler’s property of continuants, if \( \alpha = a_1 a_2 \cdots a_m \) and \( \beta = b_1 b_2 \cdots b_n \) are finite words, then we have
\[
q_{m+n}(\alpha \beta) = q_m(\alpha) q_n(\beta) + q_{m-1}(a_1 a_2 \cdots a_{m-1}) q_{n-1}(b_2 b_3 \cdots b_n),
\]
and, thus,
\[
q_m(\alpha) q_n(\beta) < q_{m+n}(\alpha \beta) < 2 q_m(\alpha) q_n(\beta).
\]

Finally, recall that \( \Sigma(t) = \{ \omega \in (\mathbb{N}^n)^Z \mid \sup_{n \in \mathbb{Z}} \lambda(\sigma^n(\omega)) \leq t \} \) 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 the words \( w \in P_r \) belonging to \( \Sigma(3 + \delta, |w|) \).

### 3. Weakly renormalizable words

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

#### 3.1. Basic facts about \( \lambda \)

We start by showing some basic facts about the function \( \lambda \) that will be useful throughout the article.

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

**Proof.** Assume that \( \omega \in \Sigma(3.06) \). Observe that \( \lambda^{-}(1|21) > 3.15 \), so the word 121 does not appear in \( \omega \). Now, if 212 is a subword of \( \omega \), so is 2212. This is not possible since \( \lambda^{-}(2|212) > 3.06 \). \( \square \)

**Lemma 3.2.** Let \( \omega \) 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 and \( R = R_1 R_2 \cdots, S = S_1 S_2 \cdots \) and \( R_1 \neq S_1 \). Then
\[
s(bw1) < \sign([w, S] - [w, R]) (\lambda(\omega) - 3) < s(bw1).
\]

In particular if \( w \) has even length, \( R_1 = 1 \) and \( S_1 = 2 \), then
\[
s(bw1) < \lambda(\omega) - 3 < s(bw1).
\]

**Proof.** First observe that \( [2; 2, w, R] + [0; 1, 1, w, R] = 3 \). Thus, we have that
\[
\lambda(R^{*}w^{*}11|22wS) = [2; 2, w, S] + [0; 1, 1, w, R]
= 3 + [0; 1, 1, w, R] - [0; 1, 1, w, S].
\]

We obtain that
\[
\lambda(R^{*}w^{*}11|22wS) - 3
= [0; 1, 1, w, R] - [0; 1, 1, w, S]
= \sign([w, S] - [w, R]) \cdot ([0; 1, 1, w, R] - [0; 1, 1, w, S]).
\]
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

\[
x = [0; a_1, a_2, \ldots, a_\ell, a_{\ell+1}, a_{\ell+2}, \ldots],
\]

\[
y = [0; a_1, a_2, \ldots, a_\ell, b_{\ell+1}, b_{\ell+2}, \ldots].
\]

where \( \ell \) is even, \( a_{\ell+1} = 1, b_{\ell+1} = 2 \).

Let \( (p_n/q_n)_{n \in \mathbb{N}} \) be the sequence of convergents of \( x \). More explicitly, we have that \( p_n/q_n = [0; a_1, a_2, \ldots, a_n] \). It is well-known that, for each \( n \in \mathbb{N} \), one has that \( p_{n+1}q_n - p_nq_{n+1} = (-1)^n \). Moreover, if we put \( a_{\ell+1} = [a_{\ell+1}; a_{\ell+2}, a_{\ell+3}, \ldots] \), then

\[
x = [0; a_1, a_2, \ldots, a_\ell, \alpha_{\ell+1}] = \frac{\alpha_{\ell+1} p_\ell + p_{\ell-1}}{\alpha_{\ell+1} q_\ell + q_{\ell-1}}.
\]

Similarly, let \( \beta_{\ell+1} = [b_{\ell+1}; b_{\ell+2}, b_{\ell+3}, \ldots] \). We then have that

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

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

\[
|x - y| = \frac{|\alpha \ell+2 p_\ell+1 + p_\ell - \beta \ell+2 q_\ell+1 + q_\ell|}{|\alpha \ell+2 q_\ell+1 + q_\ell| (\alpha + \lambda) (\beta + \lambda)} = \frac{1}{q_{\ell+1}} \left( \frac{\alpha \ell+2 - \beta \ell+2}{(\alpha \ell+2 + \beta \ell+2)} \right),
\]

(3.1)

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

\[
|x - y| = \frac{\alpha - \beta}{q_{\ell+1} (\alpha + \lambda) (\beta + \lambda)} = \frac{1}{q_{\ell+1}} \left( \frac{1}{\beta + \lambda} - \frac{1}{\alpha + \lambda} \right).
\]

(3.2)

We obtain that \( |x - y| \) is (for fixed \( q_\ell \) and \( q_{\ell+1} \)) an increasing function of \( \alpha \), and a decreasing function of \( \beta \). By analyzing Equations (3.1) and (3.2) we deduce that:

- the quantity \( |x - y| \) is minimized when \( \alpha \) is minimized, \( \beta \) is maximized. This happens when

\[
\alpha = \alpha_0 := [2; 2, 1, 1, 1, 2, 2] = \frac{21 + 2 \sqrt{210}}{21} \approx 2.3801,
\]

\[
\beta = \beta_0 := [1; 1, 2, 2, 2, 1, 1] = \frac{6 + \sqrt{210}}{12} \approx 1.7076.
\]
• the quantity \(|x - y|\) is maximized when \(\alpha\) is maximized, \(\beta\) is minimized. This happens when

\[
\alpha = \alpha_1 := \left[ 2; 1, 1, 1, 2, 2, 2 \right] = \frac{21 + 2\sqrt{210}}{19} \approx 2.6306,
\]

\[
\beta = \beta_1 := \left[ 1; 2, 2, 2, 1, 1 \right] = \frac{12 + 2\sqrt{210}}{29} \approx 1.4132,
\]

On the other hand, \(bwb = (a_1, a_2, \ldots, a_\ell, 1, 1)\), so

\[
s(bwb) = [[0; a_1, a_2, \ldots, a_\ell, 1, 1] - [0; a_1, a_2, \ldots, a_\ell, 1, 1]]
\]

\[
= \left| \frac{2p_\ell + p_{\ell-1}}{2q_\ell + q_{\ell-1}} - \frac{3p_\ell + 2p_{\ell-1}}{3q_\ell + 2q_{\ell-1}} \right|
\]

\[
= \frac{1}{(2q_\ell + q_{\ell-1})(3q_\ell + 2q_{\ell-1})} = \frac{1}{q_\ell^2 (2 + \lambda)(3 + 2\lambda)}.
\]

Similarly

\[
s(bw1) = [[0; a_1, a_2, \ldots, a_\ell, 1] - [0; a_1, a_2, \ldots, a_\ell, 1, 1]]
\]

\[
= \left| \frac{p_\ell + p_{\ell-1}}{q_\ell + q_{\ell-1}} - \frac{2p_\ell + p_{\ell-1}}{2q_\ell + q_{\ell-1}} \right|
\]

\[
= \frac{1}{(q_\ell + q_{\ell-1})(2q_\ell + q_{\ell-1})} = \frac{1}{q_\ell^2 (1 + \lambda)(2 + \lambda)}.
\]

We then have

\[
\frac{|x - y|}{s(bwb)} \geq (\alpha_0 - \beta_0) \frac{(2 + \lambda)(3 + 2\lambda)}{(\alpha_0 + \lambda)(\beta_0 + \lambda)}
\]

\[
\geq (\alpha_0 - \beta_0) \frac{(2 + 1/3)(3 + 2/3)}{(\alpha_0 + 1/3)(\beta_0 + 1/3)} \approx 1.03895 > 1,
\]

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

\[
\lambda = q_{\ell-1}/q_\ell = q_{\ell-1}/(a_\ell q_{\ell-1} + q_{\ell-2}) \geq q_{\ell-1}/(2q_{\ell-1} + q_{\ell-2}) \geq 1/3.
\]

Analogously,

\[
\frac{|x - y|}{s(bw1)} \leq (\alpha_1 - \beta_1) \frac{(1 + \lambda)(2 + \lambda)}{(\alpha_1 + \lambda)(\beta_1 + \lambda)}
\]

\[
\leq (\alpha_1 - \beta_1) \frac{(1 + 1)(2 + 1)}{(\alpha_1 + 1)(\beta_1 + 1)} \approx 0.83374 < 1,
\]

since the maps \(g_1(\lambda) = \frac{1 + \lambda}{\beta_1 + \lambda}\) and \(g_2(\lambda) = \frac{2\lambda + 3}{\alpha_1 + \lambda}\) are increasing and \(\lambda \leq 1\). \(\square\)

**Remark 3.3.** The Markov value of \(\omega = R^{*11}\{22S\} \text{ coincides with the Markov value of } \sigma(\omega)^* = S^2\{211R\} \text{ [Bom07, 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 $\omega$ be a bi-infinite word in 1 and 2 not containing 121 and 212 and such that $\omega = R^*1|22S$ with $R = R_1R_2\ldots$ and $S = S_1S_2\ldots$ and $R \neq S$. Let $\ell$ be the smallest nonnegative integer such that $R_\ell \neq S_\ell$. Then,

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

In particular, if $w = w^*$ and $\ell = |w|$, 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 the previous 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 $\ell$ 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 previous lemma, the Markov value of any infinite word in the alphabet $\{a, b\}$ containing $w$ is at least $3 + \frac{1}{7}(3 - 2\sqrt{2})^\ell$. Similarly, if $w$ contains $ab$ as a factor, then we can also write $w = u^*a|bv$. Assume now that the smallest nonnegative integer $\ell$ 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 $3 + \frac{1}{7}(3 - 2\sqrt{2})^\ell$.

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

For the sake of concreteness, we will demonstrate an usage of the previous lemma by showing that no bi-infinite word in $\Sigma(3.0007)$ contains the factor $w = bbab|aa$. Let $\omega$ be a bi-infinite word containing $w$. We start by considering the cut $bb|aba$. By the previous lemma, if $aa$ does not appear at the left of $w$ in $\omega$, then $\lambda(\omega) > 3 + \frac{1}{7}(3 - 2\sqrt{2})^2 > 3.0007$. Thus, we assume that $\omega$ contains $aababaa$ as a factor. We can now consider consider a second cut, $aa|bbabaa$. This cut shows that $\lambda(\omega) > 3 + \frac{1}{7}(3 - 2\sqrt{2})^0 > 3.0007$, 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 $w = c^n c'$, for some integer $n \geq 1$, and suppose that $w \in \Sigma(3 + e^{-r}, |w|)$. If $r(c^n) \leq r - 4$ then $n$ is even.

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

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

$$R = R_1R_2R_3 \cdots = 1^{2k-1}2 \cdots$$

$$S = S_1S_2S_3 \cdots = 2^p1^q2^r \cdots$$

By Lemma 3.2, $p > 0$ implies that $\lambda(\omega) > 3 + s(bb) = 3 + \frac{1}{17}$, which contradicts the assumption on $w$. Thus, we have that $p = 0$. Let $\ell$ 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 $2k + 1 < n$, we have that $\ell < n$. Moreover, we have that $[S] > [R]$ since $S_\ell < R_\ell$ and $\ell$ is even.
- If $q \leq 2k - 1$, then it is even as, otherwise, it would contradict the assumption on $k$. Thus, $q \leq 2k - 2$ and $\ell = q + 1 < n$. Hence, we have that $[S] > [R]$ as $S_\ell > R_\ell$ and $\ell$ is odd.

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

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

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

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

$$R = R_1R_2R_3 \cdots = 1^p2^q1^r \cdots$$

$$S = S_1S_2S_3 \cdots = 2^{2k-1}1 \cdots$$

If $p > 0$, Lemma 3.2 shows that $\lambda(\omega) > 3 + s(bb) = 3 + \frac{1}{17}$, so we have that $p = 0$. Let $\ell$ 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 $2k + 1 < n$, we have that $\ell < n$. Moreover, we have that $[S] > [R]$ since $S_\ell < R_\ell$ and $\ell$ is even.
- If $q \leq 2k - 1$, then it is even as, otherwise, it would contradict the assumption on $k$. Thus, $q \leq 2k - 2$ and $\ell = q + 1 < n$. Hence, we have that $[S] > [R]$ as $S_\ell > R_\ell$ and $\ell$ is odd.

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

$$\lambda(\omega) > 3 + s(11^\ell 11) \geq 3 + s(2^n)e^{-2} > 3 + e^{-\tau}$$

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

□

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 7.1. For example, we can show that sequences of 1’s or 2’s of odd length are forbidden:
Lemma 3.6. Let $n$ be sufficiently large so that

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

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

Lemma 3.7. Let $w \in \Sigma(3 + 6^{-n}, n)$, where $n \geq 126$ is an integer. Then, every section $w = u^*11|22v$ of $w$ satisfies $[v|_{k-1}] \leq [u|_{k-1}]$, where $k = \min\{|u|, |v|\}$.

Proof. This is a direct consequence of Lemma 3.3. Indeed, if $[v|_{k-1}] > [u|_{k-1}]$ then

$$\lambda(u^*11|22v) > 3 + \frac{1}{7}(3 - 2\sqrt{2})\ell > 3 + \frac{1}{7}(3 - 2\sqrt{2})^n > 3 + \frac{1}{6^n}$$

where $\ell$ is the smallest nonnegative integer such that $u\ell \neq v\ell$. \qed

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

$$U : a \begin{array}{c} \mapsto \ a b \\ b \mapsto \ b \end{array}, \quad V : a \begin{array}{c} \mapsto \ a \\ b \mapsto \ a b \end{array}.$$

Let $T$ be the tree obtained by successive applications of the substitutions $U$ 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 $P$ that whose distance to the root $ab$ is exactly $n$.

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

Let $g$ be the concatenation operator, that is, $g(u, v) = uv$.

Lemma 3.8. 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 $g(\overline{P})$ and $P$ are equal.

Proof. We will prove a stronger equality: $g(\overline{P}_n) = P_n$ 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, we have that 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. \qed

To state the following lemmas, we need to fix some useful notation. Let $\alpha$ and $\beta$ be finite words and assume that $\alpha$ starts with $a$, and that $\beta$ ends with $b$. We write $\alpha = a\alpha^+$ and $\beta = \beta^-b$. Then, we define $a^\beta = ba^+\alpha$ and $b_\alpha = \beta^-a$. That is, $a^\beta$ is obtained by replacing the first letter of $\alpha$ (which is $a$ by assumption) with
Lemma 3.11. For every \((a, b) \in \overline{P}\), \(a\) starts with \(a\), \(b\) ends with \(b\). Moreover, every word \(a^k b^l\), with \(k \geq 1\), starts with \(a^k\) and every word \(a^{l+1} b^l\), with \(k \geq 1\), ends with \(b^l a^{k-1}\). In particular, every sufficiently large word in \((a, b)\) starts with \(b\) and ends with \(a\), and we always have the equality \(a^i b^j = (a^i)(b^j)\).

Proof. For \((a, b) = (a, b)\), we clearly have that \(a\) starts with \(a\), \(b\) ends with \(b\),\(a^+ = b^- = \emptyset\), \(a^b = b\), \(a^b a = a^k b = a^k\) starts with \(a = a^k\) for every \(k \geq 1\), and \(a^k b = a^k b\) ends with \(b = b^k\) for every \(k \geq 1\).

By induction, if \((A, B) = (a, b)\) then \(A = a\) starts with \(a\), and \(B = b\) ends with \(b\). Since \(B = a b\) ends with \(a^b = A^b\), then, for every \(k \geq 1\), \(A^b\) also ends with \(A^b\). Now, fix \(k \geq 1\). By induction, have that \(a^k b\) starts with \(b\), and \(B^k = a^{k+1} b = a^k b\) starts with \(a\), \((a^i) = A^i = B^i\).

On the other hand, if \((A, B) = (a, b)\), then clearly \(A\) starts with \(a\), and \(B\) ends with \(b\). Since \(A = a b\) starts with \(b\), \(B = B_a\), then, for every \(k \geq 1\), \(A^b B\) starts with \(B\). Furthermore, since \(a^k b\) ends with \(a^b\), \(A^b b = A^b b\) ends with \(b\) \(a^k b = (a^b) = A^b\) for every \(k \geq 1\). The inductive argument is therefore complete.

Finally, the equality remaining \(a^i b = (a^i)\) follows immediately since \(|a^i b| = |(a^i)(b^j)|\) and, as we have just proved, \(a^i b\) starts with \(b\) and ends with \(a^i\).

\[\square\]

Remark 3.10. Every word \(w\) in \(P\) is of the form \(a^i b^j\), with \(a\) palindrome, i.e., \(a\) coincides with its transpose \(a^t\), as stated in Bombieri’s article [Bom07, Proof of Theorem 15]. Since \(a\) starts with \(a\) and \(b\) ends with \(b\), this is equivalent to \((a b)^* = ((a b)^*)^*\). In other words, both \(a^i b^j\) and \(a^i b^j\) are palindrome for every pair \((a, b) \in \overline{P}\). We will now present an alternative proof of this fact.

Observe that \(a^i b^j\) is palindrome if and only if \(b\) \(a^i b^j\) is palindrome. As in the previous lemma, we will proceed by induction; the base case is clear. Suppose that \(a^i b^j\) and \(a\) are palindrome. Then \((a b)^* = (a b)^*\) is palindrome, since both the word \((a b)^*\) starts with \(a\) and \(b\) and the word \((a b)^* = (a b)^*\) are obtained from \(a b = (a b)(a b)^*\) by replacing the first letter (which is \(a\)) with \(b\), and therefore coincide. Similarly, \((a b)^* = (a b)^*\) is also palindrome. Thus, the result holds for both \((a b, a)\) and \((a, a b)\), which completes the inductive proof.

Lemma 3.11. Suppose that a word \(w\) can be written as a concatenation \(\tau a^i b^j \tau'\) for some words \(\tau, \tau', a\) and \(b\), with \((a, b) \in \overline{P}\), \(n \in \mathbb{N}\). If there exist \((A, B) \in \overline{P}\), \(k \geq 1\) and \(w_1, \ldots, w_k \in \{A, B\}\) such that \(w = w_1 \cdots w_k\), then \((A, B) = (a, b)\) and there exists \(1 \leq j < k\) such that \(w_1 \cdots w_{j-1} = \tau, w_j = a, w_{j+1} = b\) and \(w_{j+2} \cdots w_k = \tau'\).

Proof. As usual, we proceed by induction. The result is trivial for the base case \((a, b) = (a, b) \in \overline{P}\). Assume now that \((a, b) = (a, b)\) for \((u, v) \in \overline{P}\), where \(n \geq 1\). Let \(w\) be a word such that \(w = \tau a^i b^j \tau'\) for some words \(\tau, \tau'\) 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 \(w = w_1 \cdots w_k\).

Since \(w = \tau \alpha \beta \tau'\) and \(\alpha \beta = uvw\), 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 \((A, B) = \overline{F}(u, v) = (u, uv)\). 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 \((A, B) = \overline{F}(u, uv)\), then it would not be possible for the words \(u, v\) 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. \(\square\)

We can now relate the length of a factor of a word in \(P\) with the length of the smallest word in \(P\) containing it:

**Lemma 3.12.** 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 \((\alpha, \beta) \in \overline{P}\) such that \(\alpha \beta\) contains \(w\) and such that \(|\alpha \beta|\) 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 \(\alpha = \tilde{\alpha} \beta'\) 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 \(\alpha\) and \(\beta\) by minimality of \(|\alpha \beta|\). Indeed, if \(w\) only intersects \(\alpha = \tilde{\alpha} \beta'\) or \(\beta\), then the shorter word \(\tilde{\alpha} \beta^{r-1} \beta\) corresponding to the pair \((\tilde{\alpha} \beta^{r-1} \beta, \beta) \in \overline{P}\) contradicts the minimality of \(|\alpha \beta|\). Now, if \(w\) intersects the prefix \(\tilde{\alpha}\) of \(\alpha\), then it contains \(\beta'\) 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 \(\beta^{r+1}\). Moreover, \(r\) is minimal for this property as, otherwise, the pair \((\tilde{\alpha} \beta^{r-1} \beta, \beta) \in \overline{P}\) again contradicts the minimality of \(|\alpha \beta|\). Thus, \(w = u \beta^{r+1} v\), where \(u\) is a nonempty suffix of \(\beta\) and \(v\) is a nonempty prefix of \(\beta\).

Assume that \(r \geq 2\). By Lemma 3.9, we have 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 \(\alpha = \tilde{\alpha} \beta'\), 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 that \(|w|/|\alpha \beta| = 2/3\). 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 \((\tilde{\alpha}, \tilde{\beta}) \in \overline{P}\) and \(|\tilde{\beta}| \leq |\tilde{\alpha}|\).
We have that $w$ is a factor of $\beta^2 = \tilde{a}^j \tilde{\beta} \tilde{a}^j \tilde{\beta}$ and that it intersects both copies of $\beta$. Hence, $w = uv$, where $u$ is a nonempty suffix of $\beta = \tilde{a}^j \tilde{\beta}$ and $v$ is a nonempty prefix of $\beta = \tilde{a}^j \tilde{\beta} = \tilde{a} \tilde{a}^{j-1} \tilde{\beta}$.

By Lemma 3.9, $\tilde{a}^j \tilde{\beta}$ ends with $\tilde{a}^b$. We claim that $\tilde{a}^b$ is a suffix of $u$. Indeed, if this were not the case, then $u$ would be a suffix of $\tilde{a}$ and, hence, $w$ would be contained in the shorter word $\tilde{a} \tilde{a}^j \tilde{\beta}$ corresponding to the pair $(\tilde{a}, \tilde{a}^j \tilde{\beta}) \in \mathcal{P}$, which is not possible by the minimality of $|a\beta|$. Similarly, Lemma 3.9 implies that $\tilde{a}^j \tilde{\beta} = \tilde{a} \tilde{a}^{j-1} \tilde{\beta}$ starts with $\tilde{a} \tilde{a}^{j-1} \tilde{\beta}$. We claim that $\tilde{a} \tilde{a}^{j-1} \tilde{\beta}$ is a prefix of $v$. Indeed, if this were not the case, then $v$ would be a prefix of $\tilde{a} \tilde{a}^{j-1} \tilde{\beta}$ and, hence, $w$ would be contained in the shorter word $\tilde{a} \tilde{a}^j \tilde{\beta} \tilde{a}^{j-1} \tilde{\beta}$ corresponding to the pair $(\tilde{a} \tilde{a}^j \tilde{\beta}, \tilde{a} \tilde{a}^{j-1} \tilde{\beta}) \in \mathcal{P}$, which is not possible by the minimality of $|a\beta|$. Finally, we conclude from $|\tilde{a}^b| = |\tilde{a}|$ and $|\tilde{\beta}| = |\tilde{\beta}|$, that the ratio $|w|/|a\beta|$ is at least $(j|\tilde{a}| + |\tilde{\beta}|)/(((2j+1)|\tilde{a}| + 2|\tilde{\beta}|))$, which is larger than $j/(2j+1) \geq 1/3$.

Remark 3.13. The general bound in the previous lemma cannot be improved. Indeed, for the word $w = babb^{k+1}a$ for $k \geq 0$, we have that $|a\beta|$ is minimal for the pair $(\alpha, \beta) = (ab^k a^{k+1}, ab^{k+1}) \in \mathcal{P}$. Since $|w| = k + 4$ and $|a\beta| = 3k + 5$, the ratio $w/|a\beta|$ is arbitrarily close to $1/3$ when $k$ is sufficiently large.

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

Assume then that $w$ does not intersect $\tilde{a}$. As in the previous proof, we first consider the case where $r \geq 2$. Then, we may replace the constant 3 with 2. Indeed, observe first that if $|\tilde{a}| = |\beta|$, then $|\tilde{a}| = a$ and $|\beta| = b$, so $\alpha = ab^r$ and $w = b^{r+1}$. Thus, the ratio $|w|/|a\beta|$ is $(r+1)/(r+2) \geq 3/4 \geq 1/2$. Otherwise, if $|\beta| > |\tilde{a}|$, we write $\beta = \tilde{a}^j \tilde{\beta}$ for $j \geq 1$ and $(\tilde{a}, \tilde{\beta}) \in \mathcal{P}$. We have that $w = uv$, where $u$ is a suffix of $b^r = \tilde{a}^j \tilde{\beta}^{r-1}$ and $v$ is prefix of $\beta = \tilde{a}^j \tilde{\beta}$. By Lemma 3.9, $\tilde{a}^j \tilde{\beta}$ ends with $\tilde{a}^b$ and we claim that $\tilde{a}^h \beta^{r-1}$ is a suffix of $u$. Indeed, if this were not the case, then $u$ would be a suffix of $\tilde{a} \beta^{r-1}$, so $w$ would be contained in the shorter word $\alpha = \tilde{a} \beta^r$ corresponding to the pair $(\tilde{a} \beta^{r-1}, \beta) \in \mathcal{P}$. Similarly, if we put $\beta = \tilde{a}^j \tilde{\beta}$, we have that $(\tilde{a}, \tilde{\beta}) = (\tilde{a}, \tilde{a}^{j-1} \tilde{\beta}) \in \mathcal{P}$, so Lemma 3.9 implies that $\tilde{a} \beta$ starts with $\beta_a$. We claim that $\beta_a$ is a prefix of $v$. Indeed, if we assume otherwise, then $v$ is a prefix of $\beta$ and, thus, $w$ is contained in the shorter word $(\tilde{a} \beta)^r \tilde{\beta}$ corresponding to the pair $(\tilde{a} \tilde{\beta}, (\tilde{a} \beta)^r \tilde{\beta}) \in \mathcal{P}$, a contradiction. Therefore, the ratio $|w|/|a\beta|$ is at least $((r-1)|\beta| + |\tilde{a}| + |\tilde{\beta}|)/((r+1)|\beta| + |\tilde{a}|) = r|\beta|/((r+1)|\beta| + |\tilde{a}|)$, 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. Recall that $\beta = \tilde{a}^j \tilde{\beta}$, so the result is clear when $j \geq 2$ as $j/(2j+1) \geq 2/5$. Thus, we will assume that $j = 1$, so $\alpha = \tilde{a} \tilde{a} \beta$ and that $\beta = \tilde{a} \tilde{\beta}$. If $j = 1$ and $|\tilde{a}| = |\beta|$, then $\alpha = a$, $\beta = b$, $\alpha = aab$ and $\beta = ab$. Since $w$ intersects both $\alpha$ and $\beta$, we have that $|w| \geq 2$, so we obtain $|w|/|a\beta| \geq 2/5$ once again. We will then assume that $|\tilde{\beta}| < |\tilde{a}|$. We have that $w$ is a factor of $\tilde{a}^j \tilde{\beta} \tilde{a} \tilde{\beta}$, which is in turn a factor of the shorter word $\tilde{a}^j \tilde{\beta} \tilde{a} \tilde{\beta} \tilde{a}$ corresponding to the pair $(\tilde{a} \tilde{\beta}, \tilde{a} \tilde{\beta} \tilde{a}) \in \mathcal{P}$, a contradiction.

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

Corollary 3.14. 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 \( \alpha\beta \). 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 \( \beta = a^k\beta \) and \( |\alpha|/|\beta| = 1/(k + |\beta|/|\alpha|) \), so 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 3.15.** For any finite word \( w \) in the alphabet \( \{a, b\} \), we have the identities \( bU(w^*) = U(w^*)b \) and \( V(w^*)a = aV(w)^* \).

**Proof.** This was already done by Bombieri [Bom07, Proof of Theorem 15], but for the sake of completeness we include a short proof by induction.

These identities are trivial if \( |w| = 0 \). We then assume that they hold for words of length \( n - 1 \) for \( n \geq 1 \); let \( w \) be a word of such length. If \( \tilde{w} = aw \), then

\[
\begin{align*}
    bU(\tilde{w}^*) &= bU(w^*a) = bU(w^*)ab = U(w)^*b, \\
    V(\tilde{w})a &= V(w)^*a = V(w)aa = aV(w)^*a = aV(\tilde{w})^*.
\end{align*}
\]

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

\[
\begin{align*}
    bU(\tilde{w}^*) &= bU(w^*b) = bU(w^*)b = U(w)^*bb = U(\tilde{w})^*b, \\
    V(\tilde{w})a &= V(w^*)b = V(w)aba = aV(w)^*ba = aV(\tilde{w})^*.
\end{align*}
\]

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 \((\alpha, \beta) \in \overline{P}\).

**Lemma 3.16.** Let \((\alpha, \beta) \in \overline{P}\). If \( w \) is a finite word in the alphabet \( \{\alpha, \beta\} \) starting with \( \alpha\alpha \) and ending by \( \beta\beta \) such that \( |w| \leq r \), 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 \( |w| \leq 2r \), we have that the Markov value of any bi-infinite word containing \( w \) as a factor is larger than \( 3 + e^{-r} \).

**Proof.** We will present this proof in several steps. The first step, that we label as “Step 0”, is not strictly necessary; it is contained in the other more general steps. However, we include 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 \( w = aa(ba)^kbb \), since, otherwise there is a factor of \( w \) of this form and we may replace \( w \) with 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 $3 + \frac{1}{\gamma}$ (in the general case this is not immediate; it is treated in Step 2). Hence, we assume that $k \geq 1$.

Let $\omega$ be a bi-infinite word containing $w$ and assume by contradiction that its Markov value is smaller than $3 + \frac{1}{\gamma}(3 - 2\sqrt{2})|w|$. We continue drawing conclusions from Lemma 3.4: the first cut shows that $\omega$ must contain an $a$ to the right of $w$. Thus, $\omega$ contains $w' = aa(ba)^k | bb | a$, where we again marked both cuts. We now use these cuts to conclude inductively that $w'$ must be followed with $(ba)^{k-1}$ in $\omega$: 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 with 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 $\theta = \emptyset$ (in the general case, $\theta$ can be more complicated). At the left of the first cut, we have a word of the form $(\theta b a \gamma a)^* a$, while the second cut is followed with $a \gamma$. Thus, $\omega$ contains the word $w'' = (\theta b a \gamma a)^* a | b b b | a \gamma$, where we again marked the first and second cuts.

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

**Step 1:** We now start treating the general case, so assume that $w$ starts with $\alpha \alpha$ and ends with $\beta \beta$. Since $(\alpha, \beta) \in \mathcal{P}$, Lemma 3.8 shows that there exists some $W \in (U, V)$ 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 $\tau^*$ starts with $\theta a$ and $\theta$ 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 $3 + s(b \theta b)$ which is larger than $3 + e^{-1}$ by the following computation.

By hypothesis, $s(w) \geq e^{-2 r - 1}$, so $s(\alpha \alpha \beta \beta) \geq e^{-2 r - 1}$. Write $\theta = (\theta_1, \ldots, \theta_n)$. By using Euler’s property of continuants one can get that $s(\alpha \theta a)^{-1} \leq 961 q_n(\theta)^2$, and $s(b \theta b)^{-1} \leq 162 q_n(\theta)^2$. Therefore

$$
e^{-2 r - 1} \leq s(\alpha \alpha \beta \beta) \leq s(\alpha \theta a b \theta b) \leq 2 s(\alpha \theta a) s(b \theta b) \leq \frac{324}{961} s(b \theta b)^2,$$
hence $s(bθb) \geq e^{-r}$.

We proceed by induction: in the base case, we have $θ = 0$ and $τ = a$. Now, observe that

$$U(τabθb) = U(τ)a|bbU(θ)b = ˜τa|bθb,$$

where $˜τ = U(τ)$ and $˜θ = bU(θ)$, and we have adjusted the position of the cut. We claim that $˜τ* = U(τ)*$ starts with $bU(θ)a = ˜θa$. Indeed, since $τ*$ starts with $θa$, we have that $τ$ ends with $aθ*$. Thus, $U(τ)$ ends with $abU(θ*)$. Therefore, $U(τ)*$ starts with $U(θ*)ba$, which is equal, by Lemma 3.15 to $bU(θ)a = ˜θa$.

On the other hand, observe that

$$V(τabθb) = V(τ)aa|bV(θ)ab = ˜τa|bθb,$$

where $˜τ = V(τ)a$ and $˜θ = V(θ)a$, and we have adjusted the position of the cut. We claim that $˜τ* = V(τ)*$ starts with $V(θ)aa = ˜θa$. Indeed, first observe that, by Lemma 3.15, $˜τ* = V(τ*)a$. Now, we consider two cases. If $τ* = θa$, then $˜τ* = V(θ)a = V(θ)aa = ˜θa$. Otherwise, $τ*$ starts with $θac$ where $c \in \{a, b\}$, so $˜τ*$ starts with $V(θac) = V(θ)aV(c)$. Since $V(c)$ starts with $a$ whether $c = a$ or $c = b$, we obtain that $˜τ*$ starts with $V(θ)aa = ˜θa$.

Since, by Step 1, there exists $W ∈ (U, V)$ such that $W(a) = α$ and $W(b) = β$, 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 shows 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 $w = τa|bθb$ such that:

1. there exists a word $γ$ such that $τ$ ends with $(θbγa)^*$; and
2. $θbaθbαγ$ starts with $γb$.

We have shown that (1) and (2) hold for the base case $w = aa(ba)^kbb$ with $τ = a(a/ba)^k-1b$ and $θ = 0$.

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 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 $(θba)^*$. Moreover, the configuration $τa|bθb|a$ is followed by $γ$: each $a$ of $γ$ is forced by the first cut (since $τ$ ends with $(θbγa)^*$), while each $b$ of $γ$ is forced by the second cut (since $θ*baθbαγ$ starts with $γb$). Finally, the first cut forces an $a$ after $τa|bθb|aγ$ (since $τ$ ends with $(θbaγa)^*$), while, on the contrary, the second cut forces a $b$ after $τa|bθb|aγ$ (since $θ*baθbαγ$ starts with $γb$). Thus, we obtain that the Markov value of any bi-infinite word containing $w$ is at least $3 + e^{-r}$.

To be more precise with this last part, observe that $γ$ cannot be followed by 12 or 21, because otherwise we will find a sequence of the form $c'c'c'$ where $c, c' \in \{1, 2\}$ with $c \neq c'$ and $s$ odd, but using Lemma 7.3 and the fact that $s$ is monotone

$$s(c') ≥ s(γa) ≥ s(ay*) ≥ 2^{-1}s(ab)^{-1}s(ay*abθ*) ≥ 114s(w)$$
whence \( r(c') \leq r - 4 \), a contradiction with Lemma 3.3. If \( \gamma \) is followed by \( b \), then writing the first cut as \( \omega^* = R^*b\eta^*b|\eta\alpha S \) with \( \eta = \theta\beta\gamma \) we have, by Lemma 3.2 that

\[
\lambda(\omega) = \lambda(\omega^*) \geq 3 + s(b\eta b).
\]

Since \( a\eta^*ab\theta b \) is a subword of \( w \), we have that \( r(a\eta^*ab\theta b) \leq r \) by Lemma 7.2. In particular, \( s(a\eta^*ab\theta b) \geq e^{-r-1} \). On the other hand, by Lemma 7.3 one has that \( s(a\eta^*ab\theta b) \leq 4s(a\eta\alpha)s(b\theta b) \leq s(b\eta b)/3 \), whence \( s(b\eta b) \geq e^{-r} \).

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

\[
\lambda(\omega) \geq 3 + s(b\gamma b).
\]

Finally, since \( \gamma \) is a subword of \( \eta = \theta\beta\gamma \), by Lemma 7.2 again we get that \( s(b\gamma b) \geq s(b\eta b) \geq e^{-r} \).

**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 that

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

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

\[
U((\theta ba\gamma\alpha)^*) = U(a\gamma^*ab\theta^*) = abU(\gamma^*)abU(\theta^*),
\]

\[
= aU(\gamma^*)babU(\theta)^* = (\tilde{\theta}ba\tilde{\gamma}a)^*.
\]

where we used Lemma 3.15. This shows that (1) holds for \( U(w) \). Similarly, this lemma shows that

\[
\tilde{\theta}^*ba\tilde{\theta}ba\tilde{\gamma} = U(\theta)^*bbabU(\theta)babU(\gamma)
\]

\[
= bU(\theta^*)babU(\theta)babU(\gamma)
\]

\[
= bU(\theta^*ba\theta\gamma).
\]

This word starts with \( bU(\gamma b) = bU(\gamma) b = \tilde{\gamma} b \), since \( \theta^*ba\theta\gamma \) starts with \( \gamma b \), as \( w \) satisfies (2). Hence, we obtain that (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), \( \tilde{\gamma} = V(\gamma)a \) ends with

\[
V((\theta ba\gamma\alpha)^*)a = aV(\theta ba\gamma\alpha)^* = a(V(\theta)abaV(\gamma)a)^* = (\tilde{\theta}ba\tilde{\gamma}a)^*.
\]

where we used Lemma 3.15. This shows that (1) holds for \( V(w) \). Similarly, this lemma shows that

\[
\tilde{\theta}^*ba\tilde{\theta}ba\tilde{\gamma} = (V(\theta)a)^*baV(\theta)abaV(\gamma)a
\]

\[
= aV(\theta)^*baV(\theta)abaV(\gamma)a
\]

\[
= V(\theta^*)abaV(\theta)abaV(\gamma)a
\]

\[
= V(\theta^*ba\theta\gamma a).
\]
This word starts with $V(\gamma b) = V(\gamma)ab = \gamma b$, since $\theta^*ba\theta b\gamma$ starts with $\gamma b$, as (2) holds for $w$. Hence, we obtain that (2) also holds for $V(w)$.

Since, by Step 1, there exists $W \in (U, V)$ 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 3.17.** Let $(u, v) \in \overline{P}$. If $(\alpha, \beta) = (u, uv)$, then $\alpha^k \beta = (u^k \alpha^k v_a)^*$. Similarly, if $(\alpha, \beta) = (uv, v)$, then $\alpha^k \beta = (u^k \beta^k v_a)^*$.

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

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

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

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

□

**Lemma 3.18.** Let $(u, v) \in \overline{P}$ and let $e_1, \ldots, e_k \geq 1$. If $(\alpha, \beta) = (u, uv)$, then

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

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

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

**Proof.** Assume first that $(\alpha, \beta) = (u, uv)$. Then, by Lemmas 3.9 and 3.17

$$u^b \beta \alpha^{e_1} \beta \alpha^{e_2} \beta \cdots \alpha^{e_k} v_a = u^b (uv) \alpha^{e_1} (uv) \cdots (uv) \alpha^{e_k} v_a$$

$$= (u^b) (v_a u^b) \alpha^{e_1} (v_a u^b) \cdots (v_a u^b) \alpha^{e_k} v_a$$

$$= (u^b v_a) (u^b \alpha^{e_1} v_a) u^b \cdots v_a (u^b \alpha^{e_k} v_a)$$

$$= (uv)^* (\alpha^{e_1} \beta)^* \cdots (\alpha^{e_k} \beta)^*$$

$$= (\alpha \alpha \beta^\beta \alpha \alpha^{e_1-1} \beta \cdots \beta \alpha^{e_k} \beta)^*.$$

Now, take $(\alpha, \beta) = (uv, v)$. Then, by Lemmas 3.9 and 3.17

$$u^b \beta \alpha^{e_1} \alpha \cdots \beta^{e_k} \alpha \beta v_a = (u^b) \beta \alpha^{e_1} (uv) \cdots \beta \alpha^{e_k} (uv) v_a$$

$$= (u^b) \beta \alpha^{e_1} (v_a u^b) \cdots \beta \alpha^{e_k} (v_a u^b) v_a$$

$$= (u^b \beta^{e_1} v_a) u^b \cdots \beta \alpha^{e_k} v_a (u^b v_a)$$

$$= (\alpha \beta^{e_1})^* \cdots (\alpha \beta^{e_k})^* (uv)^*$$

$$= (\alpha \alpha \beta^\beta \alpha \alpha^{e_1-1} \alpha \cdots \alpha \beta^{e_k})^*.$$  

□
The three previous lemmas imply that we obtain a large Markov value in the case where \((\alpha, \beta) = (u, wv)\) for any word of the form \(u^b \beta \cdots \alpha v_u\), and in the case where \((\alpha, \beta) = (uv, v)\) for any word of the form \(u^b \beta \cdots \alpha v_u\).

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 3.19.** Let \((\alpha, \beta) \in \overline{\Gamma}\) and \(w \in \Sigma(3 + \delta, n)\) a finite word. We say that \(w\) is \((\alpha, \beta)\)-weakly renormalizable if we can write \(w = w_1 \gamma w_2\) where \(\gamma\) 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 \((u, v) \in \overline{\Gamma}\) and \(\gamma\) ends with \(\alpha\), then \(|v| \leq |w_2|\) (and \(w_2\) starts with \(v_0\)). If \((\alpha, \beta) = (uv, v)\) for some \((u, v) \in \overline{\Gamma}\) and \(\gamma\) starts with \(\beta\), then \(|u| \leq |w_1|\) (and \(w_1\) ends with \(u^b\)).

The previous definition is motivated by the following ideas. Given an alphabet \(\{\alpha, \beta\}\) with \((\alpha, \beta) \in \overline{\Gamma}\), it may not be possible to write a word \(w\) in terms of \(\alpha\) and \(\beta\). Nevertheless, it may very well be possible to write “most” of \(w\) in terms of \(\alpha\) and \(\beta\), 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, \(|w_1| \geq \max\{|\alpha|, |\beta|\}\), then either \(w_1\) ends with \(\alpha\) or \(\beta\) 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 \(\gamma\) 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. Indeed, if, for example, \((\alpha, \beta) = (u, uv)\) with \((u, v) \in \overline{\Gamma}\) and \(\gamma\) ends with \(\alpha = u\), then we need to ensure that \(w_2\) does not start with \(v\), because, otherwise, \(\alpha v_2 = wv_2\) would start with \(uv\), so the last letter of the renormalization kernel should be \(\beta\) instead of \(\alpha\) and \(w_2\) should be shorter. Similarly, if \((\alpha, \beta) = (uv, v)\) with \((u, v) \in \overline{\Gamma}\) and \(\gamma\) starts with \(\beta = v\), we need to ensure that \(w_1\) does not end with \(u\), since, otherwise, \(w_1 \beta = w_1 v\) would end with \(uv = \alpha\) and, hence, the first letter of the renormalization kernel should be \(\alpha\) instead of \(\beta\) (and \(w_1\) should be shorter). These facts are guaranteed by the previous definition. Indeed, if \((\alpha, \beta) = (u, uv)\) and \(\gamma\) ends with \(\alpha\) we require \(w_2\) to start with \(v_0 \neq v\) (by requiring it to be at least as long as \(v\) and by Lemma 3.9). Similarly, if \((\alpha, \beta) = (uv, v)\) and \(\gamma\) starts with \(\beta\), we require \(w_1\) to end with \(u^b \neq u\) (by requiring it to be at least as long as \(u\) and by Lemma 3.9).

Exhibiting a word as being \((\alpha, \beta)\)-weakly renormalizable is nontrivial in general and, to complicate matters even further, the choice of alphabet \((\alpha, \beta) \in \overline{\Gamma}\) is not clear to begin with. Nevertheless, any word in the alphabet \(\{a, b\}\) is trivially \((a, b)\)-weakly renormalizable (setting the renormalization kernel equal to
the entire word). Thus, 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 \((\alpha, \beta)\)-weakly renormalizable for \((\alpha, \beta) \in \{(u, v), (u, uv)\}\) chosen appropriately.

**Lemma 3.20** (Renormalization algorithm). Let \(w \in \Sigma(3 + e^{-r}, |w|)\) with \(r(w) \leq r\). If \(w\) is \((u, v)\)-weakly renormalizable as \(w = w_1\gamma w_2\) with \(\gamma \neq \emptyset\), then \(w\) is \((\alpha, \beta)\)-weakly renormalizable for some \((\alpha, \beta) \in \{(u, v), (uv, v)\}\). Moreover, if \(w\) starts with \(u\) or ends with \(v\), then \(w_1\) or \(w_2\), respectively, does not change for the renormalization with alphabet \((\alpha, \beta)\).

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

If \(\gamma\) contains the factors \(uv\) and \(vu\), then Lemmas 3.16 and 3.17 imply that \(w \notin \Sigma(3 + e^{-r}, |w|)\). We first assume that \(w\) does not contain the factor \(uv\) and we analyze the following subcases (where \(e_j\) is a positive integer for \(1 \leq j \leq k\)):

**Case 1:** If \(\gamma = u^{e_1}v^{e_2}v \cdots u^{e_k}v\), we take \(\alpha = u, \beta = uv\) and

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

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 \(v\), since \(uv\) starts with \(v\) by Lemma 3.9. The same lemma also shows that \(u^2v\) starts with \(v\), 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 \(\gamma = vu^{e_1}v^{e_2}v \cdots u^{e_k}v\) we consider two cases. If \(|w_1| < |u|\), we take \(\alpha = u, \beta = uv\) and

\[
\tilde{\gamma} = \alpha^{e_1-1} \beta \alpha^{e_2-1} \beta \cdots \alpha^{e_k-1} \beta, \quad \tilde{w}_1 = w_1 v, \quad \tilde{w}_2 = w_2.
\]

Indeed, recall that \(uv\) ends with \(u^b\) and that \(w_1\) 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). We obtain that \(w_1\) is a suffix of \(u\), so \(\tilde{w}_1 = w_1 v\) is a suffix of \(\alpha \beta = u^2v\). Moreover, \(\tilde{w}_2 = w_2\) is a prefix of \(\alpha \beta = u^2v\) by the exact same proof of the previous case: it is either shorter than \(v\) (in which case it is a proper prefix of \(v\), and, hence, of \(u^2v\) by Lemma 3.9), or shorter than \(u\) (in which case it is a prefix of \(u\) and, hence, of \(u^2v\)).

Otherwise, we have \(|u| \leq |w_1| < |v|\), so \((u, v) = (\eta, \eta \theta)\) for some pair \((\eta, \theta) \in \overline{P}\). Since \(w_1\) is a suffix of \(uv = \eta^b \theta\) and \(|w_1| \geq |u| = |\eta|\), we have that \(w_1\) ends with \(\eta^b\) by Lemma 3.9. If \(e_j > 1\) for some \(1 \leq j \leq k\), then \(w\) contains a factor of the form \(\eta^b v \cdots u u \theta_{a}\). In fact, since \(w_1\) ends with \(\eta^b\), we have that \(w\) contains a word of the form \(w' = \eta^b v \cdots u^{e_j-1} u u v\), where \(1 \leq j \leq k\) is chosen so \(e_j > 1\). Moreover, \(v = \eta \theta\) starts with \(\theta_{a}\) by Lemma 3.9, so \(w'\) contains, in turn, a word of the form \(\eta^b v \cdots u u \theta_{a}\). This contradicts that \(w \in \Sigma(3 + e^{-r}, |w|)\) by Lemmas 3.16 to 3.18.

We assume then that \(e_j = 1\) for every \(1 \leq j \leq k\) and take \(\alpha = uv, \beta = v\) and

\[
\tilde{\gamma} = \beta \alpha^k, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = w_2.
\]
we have that \( \tilde{w}_2 = w_2 \) 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 \( |w_2| < |u| \) then \( w_1 \) is a proper suffix of \( u^b \) (by Lemma 3.9), so it is also a suffix of \( uv^2 \) (by Lemma 3.9 again). Finally, since \( \tilde{\gamma} \) starts with \( \beta \) and \((\alpha, \beta) = (uv, v)\), we have to check that \( |\alpha| \leq \tilde{|\tilde{w}_1|} = |w_1| \), but this holds by hypothesis.

**Case 3:** If \( \gamma = u^{e_1}vuv^{e_2}v \cdots u^{e_s}vu^s \), we take the unique integer \( 0 \leq r \leq s \) such that \( |v| \leq |uv^{r+1}w_2| \). Then, we choose \( \alpha = u, \beta = uv \) and

\[
\tilde{\gamma} = \alpha^{e_1-1}\beta\alpha^{e_2-1}\beta \cdots \alpha^{e_s-1}\beta \alpha^{j-r}, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = u^r w_2.
\]

Indeed, \( \tilde{w}_1 = w_1 \) is a suffix of \( \alpha \beta = u^2 v \) since it is a suffix of \( uv \). Now, if \( r = 0 \), then \( |w_2| < |u| \) since \( |w_2| < \max\{|u|, |v|\} \) and \( |v| \leq |w_2| \) by hypothesis. Since \( w_2 \) is a prefix of \( uv \), it is actually a prefix of \( u \) and, hence, of \( \alpha \beta = u^2 v \). If \( r = 1 \), we have that \( w_2 \) is a prefix of \( uv \) and, thus, \( \tilde{w}_2 = w_2v \) is a prefix of \( \alpha \beta = u^2 v \); from now on we assume that \( r > 1 \). We have that \( |v| > (r - 1)|u| + |w_2| \) and, thus, that \( v = u^j \tilde{v} \) for some \( j \geq r - 1 \) with \((u, \tilde{v}) \in P \) and \( |\tilde{v}| < |u| \). Observe that the hypothesis implies that \( |\tilde{v}| > (r - j - 1)|u| + |w_2| \), so if \( j = r - 1 \) we obtain that \( |u| \geq |\tilde{v}| > |w_2| \). Then, since \( w_2 \) is a prefix of \( uv \), it is actually a prefix of \( u \) and, therefore, \( \tilde{w}_2 = u^r w_2 \) is a prefix of \( u^{j+1} = u^{j+2} \). We obtain that \( \tilde{w}_2 \) is a prefix of \( \alpha \beta = u^2 v = u^{j+2}v \). Otherwise, if \( j \geq r \), we have that

\[
(j - r + 2)|u| \geq (j - r + 1)|u| + |\tilde{v}| > |w_2|
\]

so \( w_2 \) is a prefix of \( u^{j+2} \) (since this latter word is a prefix of \( u^{j+1} \) which is, in turn, a prefix of \( uv = u^{j+1} \tilde{v} \)). Hence, \( \tilde{w}_2 = u^r w_2 \) is a prefix of \( u^{j+2} \), so it is also a prefix of \( \alpha \beta = u^2 v = u^{j+2}v \).

Since \( \tilde{\gamma} \) ends with \( \alpha \) if \( r < s \), we have to check that \( |v| \leq |\tilde{w}_2| \). Since \( \tilde{w}_2 = u^r w_2 \) and \( r \) was chosen so \( |v| \leq |u^r w_2| \), this holds in all such cases.

**Case 4:** Finally, if \( \gamma = uv^{e_1}vuv^{e_2}v \cdots u^{e_s}vu^s \), 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 3.16 to 3.18. Indeed, in this case we have that \((u, v) = (\eta, \eta \theta)\) for some \((\eta, \theta) \in P \), so \( w \) contains a factor of the form \( \eta^b v \cdots uu^{e_1}\eta \theta \) 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 \( w' = \eta^b v \cdots uu^{e_2}uuw_2 \). Now, observe that the fact that \( \gamma \) ends with \( u \) and the definition of \((u, v)\)-renormalizability imply that \( w_2 \) starts with \( \theta_a \). Hence, \( w' \) contains a word of the form \( \eta^b v \cdots uu \theta_a \). This leads to the same contradiction with Lemmas 3.16 to 3.18.

In the case where \( |u| \leq |w_1| < |v|, e_j = 1 \) for every \( 1 \leq j \leq k \) and \( s = 1 \), we take \( \alpha = uv, \beta = v \) and

\[
\tilde{\gamma} = \beta \alpha^k, \quad \tilde{w}_1 = w_1, \quad \tilde{w}_2 = uvw_2.
\]

Since \( w_1 \) is a suffix of \( uv \), then \( \tilde{w}_1 = w_1 \) is a suffix of \( \alpha \beta = u^2 v \). Now, observe that \( |u| \leq |w_1| < |v| \) implies that \( |w_2| < |v| \), since by hypothesis we have that \( |w_1|, |w_2| < \max\{|u|, |v|\} \). Thus, by Lemma 3.9 \( w_2 \) is a proper prefix of \( v_a \), so it is also a prefix of \( v \). We then obtain that \( \tilde{w}_2 = uvw_2 \) is a prefix of \( \alpha \beta = uv^2 \).
Otherwise, if $|w_1| < |u|$ we take $\alpha = u$, $\beta = u\nu$ and argue as in the third case. More precisely, let $0 \leq r \leq s$ be the unique integer that satisfies $|\nu| \leq |u^{s+1}w_2| < |uv| \leq |u^{r+1}w_2|$ and take

$$\tilde{\gamma} = \alpha^{s+1} \beta \alpha^{s-1} \beta \cdots \alpha \beta - r, \quad \tilde{w}_1 = w_1\nu, \quad \tilde{w}_2 = u^r\nu.$$  

We have that $\tilde{w}_2 = u^r\nu$ is a prefix of $\alpha \beta$ by the same arguments of the third case, and $r$ is chosen so $|\nu| \leq |\tilde{w}_2|$. Moreover, $\tilde{w}_1 = w_1\nu$ is a suffix of $\alpha \beta = u^2\nu$ since Lemma 3.9 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 $\nu\nu$, so, in particular, it does not contain the factor $uu$. We analyze the following subcases (where $e_j$ is a positive integer for $1 \leq j \leq k$):

**Case 1**: If $\gamma = u\nu^{e_1}u\nu^{e_2} \cdots u\nu^{e_k}$, we take $\alpha = u\nu$, $\beta = \nu$ and

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

**Case 2**: If $\gamma = u\nu^{e_1}u\nu^{e_2} \cdots u\nu^{e_k}u$, we take $\alpha = u\nu$, $\beta = \nu$ and

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

**Case 3**: If $\gamma = \nu^{e_1}u\nu^{e_2} \cdots u\nu^{e_k}$, we take the unique integer $0 \leq r \leq s$ such that $|\nu| \leq |w_1\nu^r| < |uv| \leq |w^r\nu^{s+1}|$, and define $\alpha = u\nu$, $\beta = \nu$ and

$$\tilde{\gamma} = \nu^{s+r}u\nu^{e_1-1}u \cdots \nu^{s-1}u\nu^{e_k-1}, \quad \tilde{w}_1 = w_1\nu^r, \quad \tilde{w}_2 = w_2.$$  

**Case 4**: If $\gamma = \nu^{e_1}u\nu^{e_2} \cdots u\nu^{e_k}u$, we take the unique integer $0 \leq r \leq s$ such that $|\nu| \leq |w_1\nu^r| < |uv| \leq |w^{r-s+1}\nu^{s+1}|$, and define $\alpha = u\nu$, $\beta = \nu$ and

$$\tilde{\gamma} = \nu^{s+r}u\nu^{e_1-1}u \cdots \nu^{s-1}u\nu^{e_k-1}u, \quad \tilde{w}_1 = w_1\nu^r, \quad \tilde{w}_2 = u\nu w_2.$$  

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 $\nu\nu$. The arguments showing that these choices satisfy the definition of $(\alpha, \beta)$-renormalizability are completely analogous to those of the previous cases (where the factor $\nu\nu$ was not present). Thus, this concludes the proof.

- □

Once again, this lemma could be stated in terms of the length of $w$. That is, if $w \in \Sigma(3 + 6^{-n}, n)$ with $n$ large enough, then $r(w) \leq (n + 1) \log(3 + 2\sqrt{2})$, so $w$ is $(\alpha, \beta)$-weakly renormalizable.

Let $r \in \mathbb{N}$. Observe that if $\alpha = (a_1, \ldots, a_n) \in P_r$, then

$$s(\alpha)^{-1} \leq 2s(a_n)^{-1} s(a_1 \cdots a_{n-1})^{-1} \leq 12e^r,$$

which implies that $r(\alpha) \leq r + 2$. In particular, the renormalization exists for every word $w \in \Sigma^{(r-2)}(3 + e^{-r})$. Moreover, we can show that $w$ is $(\alpha, \beta)$-weakly renormalizable for some alphabet $|\alpha|, |\beta| < |w|$ with $|\alpha| + |\beta| \geq |w|/2$. Indeed, if $|\alpha| + |\beta| < |w|/2$, then writing $w = w_1\nu w_2$ gives $|w_1| + |w_2| < 2(|\alpha| + |\beta|) \leq |w|$. We obtain that $\gamma \neq \emptyset$, so we can continue applying the algorithm. In particular,
if $w \in \Sigma(3 + e^{-r}, |w|)$, then $w$ is $(\alpha, \beta)$-renormalizable for some $\alpha, \beta \in \overline{P}$ with

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

where we are using Lemma \[7.1\] and that the inequality $|\alpha \beta| \geq r/6$ trivially holds for $r \leq 24$ since $|\alpha \beta| \geq 4$.

**Lemma 3.21.** Let $(\alpha, \beta) \in \overline{P}$ with $|\alpha \beta| < r/6$ and $w \in \Sigma^{(r-2)}(3 + e^{-r})$. If $w$ contains $\alpha \beta$, then $w$ is $(\alpha, \beta)$-weakly renormalizable, say $w = w_1 \gamma w_2$. Moreover, if $w$ starts (ends) with $\alpha \beta$, then $w_1 = \emptyset$ ($w_2 = \emptyset$).

**Proof.** First note that $w$ is $(a, b)$-weakly renormalizable. Indeed, writing $w = \eta a \beta \eta'$ we have that $\max\{r(\eta), r(\eta')\} \leq r(w) - r(\alpha \beta) + 1 \leq r - 4$ where we used $r(ab) = 5$. Hence, by Lemma \[3.5\] sequences of 1’s or 2’s odd length are forbidden. Thus $w = \gamma_0$ where $\gamma_0 \in \langle a, b \rangle$. Now we apply inductively the renormalization algorithm (Lemma \[3.20\]) to obtain a sequence of alphabets $(\alpha_j, \beta_j) \in \overline{P}_j$ such that for all $0 \leq j \leq m$, $w$ is $(\alpha_j, \beta_j)$-weakly renormalizable for each $j$ and $|A_n B_n| \geq r/6$.

On the other hand, since $(\alpha, \beta) \in \overline{P}$ there exists a sequence of alphabets $(\alpha_i, \beta_i) \in \overline{P}_i$ such that $\alpha \beta \in \langle \alpha_i, \beta_i \rangle$ for all $0 \leq i \leq n$ and $(\alpha_n, \beta_n) = (\alpha, \beta)$. Since $\alpha \beta$ starts with $a = a_0$ and ends with $b = b_0$ (Lemma \[3.9\]), inductively we obtain that $\alpha \beta$ starts with $\alpha_i$ and ends with $\beta_i$. In particular $\alpha \beta$ contains $\alpha_j \beta_j$.

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

$$\tau \alpha_j \beta_j \tau' = A_j \gamma_j A_j \beta_j \in \langle \alpha_j, \beta_j \rangle,$$

hence by Lemma \[3.11\] we obtain that $(\alpha_j, \beta_j) = (\alpha_j, \beta_j)$ for all $0 \leq j \leq n$. In particular $m > n$, because otherwise $r/6 \leq |A_n B_n| = |\alpha_n \beta_n| < r/6$. This shows that $w$ is $(\alpha, \beta)$-weakly renormalizable.

Now assume that $w$ starts with $\alpha \beta$ (the other case is analogous). We will show that $w_1 = \emptyset$ for all $0 \leq j \leq n$. Note that we already showed that $w_1$ is empty for $(a_0, b_0) = (a, b)$. If $w_1$ becomes nonempty for $k + 1$ for some $0 \leq k \leq n$, it must happen that $w = \gamma_j w_2$ starts with $\beta_k$ (because of the renormalization algorithm). But $w$ starts with $\alpha \beta$, which in turn starts with $\alpha_1' \beta_1'$, which leads to a contradiction because it starts with $(\beta_k)_{\gamma_j}$. Since $(\alpha_n, \beta_n) = (\alpha, \beta)$ this finishes the proof. 

Finally, to end this section we prove Theorem \[1.1\].

**Proof of Theorem 1.1.** We claim that $\Sigma(3 + 6^{-3n}, n) = \Sigma(3, n)$, provided $n$ is large enough. Indeed, let $\theta$ an element of $\Sigma(3 + 6^{-3n}, n)$. It has a $3n$-neighbourhood $\tau$ (gluing sequences of size $n$ at each side of $\theta$ in $\Sigma(3 + 6^{-3n}, 3n)$. By Lemma \[3.20\] there exists $(\alpha, \beta) \in \overline{P}$ with $|\alpha|, |\beta| < n$ and $|\alpha| + |\beta| \geq n$ such that $\tau$ is $(\alpha, \beta)$-weakly renormalizable. Writing $\tau = w_1 \gamma \gamma_2$ as in the definition of weak renormalization, we have $|w_1|, |w_2| < \max\{|\alpha|, |\beta|\} < n$, so $\theta$ is a factor of $\gamma$. Considering the smallest sequence $\eta$ of $(\alpha, \beta)$-letters of $\gamma$ containing $\theta$ as a factor, the sequence obtained by removing the first and the last $(\alpha, \beta)$-letter of $\eta$ has size smaller than $n$ and thus cannot contain $\alpha \beta$ or $\beta \alpha$ as factors,
and thus $\eta$ is of the form $\alpha^r, \beta^r, \alpha^r \beta, \beta^r \alpha, \beta \alpha^r \beta$ or $\alpha \beta^r \alpha$ for some positive integer $r$. In any of these cases, $\eta \in \Sigma(3, |\eta|)$, and 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 Lemma 3.12 there exists $\Pi \in P$ containing $w$ such that $|\Pi| \leq 3|w|$. Since $(3 + 2\sqrt{2})^3 < 6^3$, if $n$ is sufficiently large then Lemma 3.4 shows that $\Pi^\infty \in \Sigma(3 - 6^{-3n}, n)$, so $w \in \Sigma(3 - 6^{-3n}, n)$.

\[\Box\]

4. IMPROVING THE ESTIMATES

We begin this section by stating an analog of Bombieri’s characterization of admissible words [Bom07, Lemma 13]. Its proof is essentially applying the renormalization algorithm to a word of the form $w = \alpha^e \beta \alpha^{e+1} \beta$ or $w = \beta^e \alpha \beta^{e+1} \alpha$, but we need to be careful about the magnitude of $r(w)$.

**Lemma 4.1 (Bombieri’s characterization)**. Let $(\alpha, \beta) \in \overline{P}$. Consider a word $\gamma$ of the form $\gamma = \alpha \beta^e \alpha \beta^{e+1} \cdot \cdot \cdot \beta^e \alpha \beta^{e+1} \cdot \cdot \cdot \beta^e \alpha \beta^{e+1} \alpha$ or $\gamma = \beta^e \alpha \beta^{e+1} \alpha \cdot \cdot \cdot \alpha \beta^e \beta$ with $e_i \geq 1$ for all $1 \leq i \leq \ell - 1$. Assume that $\gamma \in \Sigma(3 + e^{-r})$, $|\gamma|$ and let $\theta \in (\alpha, \beta)$. Then

- If $r(\theta^e) \leq r - 2|\alpha \beta|$ then $|e_i - e_{i+1}| \leq 1$ for $1 \leq i \leq \ell - 2$, $e_1 \geq e_0 - 1$ for $i = 0$ when $\theta = \alpha$ and $e_i \leq e_{i+1} - 1$ for $i = \ell - 1$ when $\theta = \beta$.
- If $r(\theta^{e+1}) \leq r - 6|\alpha \beta|$ or $|\beta| \leq |\alpha|$ then $e_\ell \leq e_{\ell-1} + 1$.
- If $r(\theta^{e+1}) \leq r - 6|\alpha \beta|$ or $|\alpha| \leq |\beta|$ then $e_1 \geq e_0 - 1$.

Before proceeding with the proof, we must comment why we need $r(\theta^e)$ to be smaller at the end of the word in the last two cases. Observe that if $\beta = \alpha^v$ for some $(\alpha, v) \in \overline{P}$, then clearly $e_\ell$ can be much larger than $e_{\ell-1}$, because all the powers $\alpha^{e_{\ell-1}} \beta$ could belong to the (possible) next letter $\beta$. Similarly, when $\alpha = u \beta^i$ for some $(u, \beta) \in \overline{P}$, the power $\beta^{e_{\ell-1}}$ could belong to the (possible) preceding letter $\alpha$.

**Proof.** Let $\omega$ be a bi-infinite word containing $\gamma$ and such that $\omega \in \Sigma(3 + e^{-r})$.

Suppose $\gamma = \alpha^{e_0} \beta \alpha^{e_1} \beta \cdot \cdot \cdot$. Take $k \leq e_{i+1}$ maximal such that $r(\alpha^e \beta \alpha^k) \leq 2r$. If $e_\ell \geq k$ then $r(\alpha^k) \leq r - 2|\alpha \beta|$ as well, so actually $k = e_{i+1}$ because, otherwise,

$$r(\alpha^e \beta \alpha^{e+1}) \leq r(\alpha^e) + r(\beta \alpha) + r(\alpha^{e+1}) + 6 \leq r - 2|\alpha \beta| + r(\beta \alpha) + r - 2|\alpha \beta| + 4 \leq 2r,$$

where we used Lemma 7.1 to guarantee that $r(\alpha^e) \leq 1.8|\beta \alpha| + 1.8$. Similarly, we use $r(\beta) \leq 1.8|\beta| + 1.8$ (for $\beta = b$ use $r(b) = 1$ instead) to get

$$r(\alpha^e \beta \alpha^{e+1} \beta) \leq 2r - 4|\alpha \beta| + 2r(\beta) + 6 \leq 2r.$$

Hence, letting $(\tilde{\alpha}, \tilde{\beta}) = (\alpha, \alpha^{e+1} \beta)$ we obtain that

$$\tilde{\gamma} = \alpha^e \beta \alpha^{e+1} \beta = \tilde{\alpha}^{e-e+1} \tilde{\beta} \tilde{\beta}$$

is a subword of a word $\omega \in \Sigma(3 + e^{-r})$, so if $e_\ell - e_{i+1} \geq 2$ it will contradict the second part of Lemma 3.16. If $e_i < k$, then let $(u, v) = (\alpha, \alpha^{e+1} \beta)$ and
(\tilde{\alpha}, \tilde{\beta}) = (u, vw)$ hence by the first case of Lemma 3.18

$$\alpha \beta \alpha^{\epsilon_i+1} \beta = \alpha \beta \tilde{\beta} \alpha^{\epsilon_i+1} \tilde{\beta} = \beta_a b^i \tilde{\beta} \alpha^{\epsilon_i+1} \tilde{\beta} = \beta_a (\tilde{\alpha}^{\epsilon_i+1} \tilde{\beta} \tilde{\beta})^* \alpha^b$$

is a subword of $\gamma$ when $i < \ell - 1$. If $e_{i+1} - e_i \geq 2$, then we would have that $\tilde{\alpha} \tilde{\alpha} \tilde{\beta} \tilde{\beta}$ is a subword of $\gamma^*$ with $r(\tilde{\alpha} \tilde{\alpha} \tilde{\beta} \tilde{\beta}) = r(\alpha^{\epsilon_i+2} \beta \alpha^{\epsilon_i}) \leq 2r$, which contradicts Lemma 3.16. In the particular case where $i = \ell - 1$, we do not necessarily have $\beta$ after $\alpha^{\epsilon_i}$. If $|\beta| \leq |\alpha|$, then $(\alpha^b)^* (\tilde{\alpha}^{\epsilon_i+1} \tilde{\beta} \tilde{\beta}) (\beta_a)^*$ is a subword of $\gamma^*$ after removing a $\beta^*$ at the beginning, so we still get that $\tilde{\alpha} \tilde{\alpha} \tilde{\beta} \tilde{\beta}$ is a subword of $\gamma^*$.

When $(\alpha, \beta) = (u, vw)$ we need to extend the word $\beta \alpha^{\epsilon_i+1} \beta \alpha^{\epsilon_i}$ by using Lemma 3.21. We will extend this word to the left and then to the right. Since $|\beta| = |vw| < r/6$, consider the $(u, v)$-weakly renormalizable continuation $\hat{w} \in \Sigma^{(r/2)}(3 + \epsilon^{-r})$ that ends in the first $\beta$; in particular $\hat{w} = \hat{w} \hat{v}$ where $\hat{v} \in \{u, v\}$ and $\hat{w}_1$ is a suffix of $u$. We claim that $|u| \leq |\hat{w}_1|$. Otherwise, we use Lemma 7.1 to obtain that

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

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

Now we want to extend the word to the right. Consider now the continuation $\hat{w} \in \Sigma^{(r/2)}(3 + \epsilon^{-r})$ that begins at $\alpha \beta \alpha^{\epsilon_i}$. In particular it is $(\alpha, \beta)$-weakly renormalizable by Lemma 3.21. Since

$$r(\alpha \beta \alpha^{\epsilon_i+1} \alpha \beta) \leq r(\alpha \beta \alpha) + r(\alpha^{\epsilon_i+1}) + r(\alpha \beta) + 4$$

$$\leq 3.6|\alpha \beta| + 1.8|\alpha| + r - 6|\alpha \beta| + 8 \leq r - 2$$

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

$$u^b \tilde{\alpha}^f \beta \alpha^{\epsilon_i+1} \beta \alpha^{\epsilon_i+1+2+g} v_a$$

is a subword of $\omega \in \Sigma^{(3 + \epsilon^{-r})}$. In this situation the first case of Lemma 3.16 yields

$$u^b \tilde{\alpha}^f \beta \alpha^{\epsilon_i} \beta \alpha^{\epsilon_i+1+2+g} v_a = (\alpha^{\epsilon_i+1+2+g} \beta \alpha^{\epsilon_i} \beta \alpha^{\epsilon_i})^*$$

so we still get that $\tilde{\alpha} \tilde{\alpha} \tilde{\beta} \tilde{\beta} = \alpha^{\epsilon_i+1+2} \beta \alpha^{\epsilon_i+1} \beta$ is a subword of $\omega^* \in \Sigma^{(3 + \epsilon^{-r})}$, contrary to Lemma 3.16 above.

Now assume $\gamma = \beta^{\epsilon_0} \alpha^{\epsilon_1} \alpha \cdots$. Take a maximal integer $k \leq e_{i+1}$ satisfying $r(\beta^{\epsilon_i+2} \alpha^{\epsilon_i}) \leq 2r$. If $e_i < k$, then letting $(\tilde{\alpha}, \tilde{\beta}) = (\alpha^{\epsilon_0}, \beta)$ one gets that $\alpha^{\epsilon_i} \beta^{\epsilon_i} = \tilde{\alpha} \tilde{\beta} \tilde{\alpha} \tilde{\beta}$ is a subword of a word $\omega \in \Sigma^{(3 + \epsilon^{-r})}$. Since $k \leq e_{i+1}$, if $e_{i+1} - e_i \geq 2$, we have a contradiction again. If $e_i > k$, then $r(\beta^{\epsilon_i+1}) \leq r - 2|\alpha \beta|$ as well, so $k = e_{i+1}$, hence let $(u, v) = (\alpha^{\epsilon_i+1}, \beta)$ and $(\tilde{\alpha}, \tilde{\beta}) = (uv, v)$ so by the second case of Lemma 3.18 one gets

$$\alpha \beta^{\epsilon_i} \alpha \beta^{\epsilon_i+1} \beta = \tilde{\alpha} \tilde{\beta}^{\epsilon_i+1} \tilde{\alpha} \beta = \alpha v_a b^i \tilde{\beta}^{\epsilon_i+1} \tilde{\alpha} v_a \alpha^b = \beta_a (\tilde{\alpha} \tilde{\alpha} \tilde{\beta}^{\epsilon_i+1} \tilde{\beta})^* \alpha^b$$

is a subword of $\gamma$. If $e_i - e_{i+1} \geq 2$ then one would get that $\tilde{\alpha} \tilde{\alpha} \tilde{\beta} \tilde{\beta} = \alpha^{\epsilon_i+2} \alpha \beta^{\epsilon_i+1} \beta$ is a subword of $\gamma^*$ with $r(\tilde{\alpha} \tilde{\alpha} \tilde{\beta} \tilde{\beta}) = r(\alpha \beta^{\epsilon_i+1} \alpha \beta^{\epsilon_i+2}) \leq 2r$, which is impossible. In the particular case where $i = 0$, we do not have $\alpha$ before $\beta^{\epsilon_0}$. In the case where $|\alpha| \leq |\beta|$ this is no problem because then $(\alpha^b)^* \tilde{\alpha} \tilde{\alpha} \tilde{\beta}^{\epsilon_i+1} (\beta_a)^*$ is a
subword of γ after removing a α∗ at the end, so we still get that ̃a ̃a ̃β ̃β is a subword of γ. When (α, β) = (u, v), an analogous argument as before shows that the word βα αβ has a continuation to the left that is (α, β)-renormalizable. So before there is either a αβ αβ or a ̃w1β ̃β, where ̃w1 is a suffix of αβ that ends with u̅. Similarly there is a βα αβ to the right of βα αβ. In resume the word u̅βα βα αβ αβ is a subword of ω ∈ Σ(3 + e−r). But the second case of Lemma 3.18 implies that

\[ u̅βα βα αβ αβ = (α α β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α β β α beta

So again if e0 − e1 ≥ 2, then ̃a ̃a ̃β ̃β = α β e1 α β e1+2 is a subword of ω∗ ∈ Σ(3 + e−r), which contradicts once more Lemma 3.16.

\[ \square \]

4.1. Constructing renormalizable extensions. We start by considering local extensions. More precisely, if we have a word w that starts (ends) with αβ where (α, β) ∈ P, then the alphabet is uniquely determined, and the beginning (end) of w should be (α, β)-weakly renormalizable. This is a consequence of Lemma 3.21.

Now we want to consider extensions of renormalizable words. The next lemma says that if we have a power (uv)ℓ with (u, v) ∈ P, then it will have a large extension w that is “almost” (u, v)-weakly renormalizable consisting mostly of powers (uv)ℓ. This is explained since exponents can only decrease linearly, in fact, they may only decrease by 1 when below some threshold. We say that w = γw2 is almost (u, v)-weakly renormalizable, because the tail w2 satisfies now the condition |w2| < 2ℓ|w| and w2 is a prefix of a word in {uvw, uuv}.

Lemma 4.2. Let w ∈ Σ(3 + e−r) be a finite word starting with θν, where θ = uv and (u, v) ∈ P, r(θν) ≤ r − 4|θ|, |w| < r/9 and also

\[ Tr ≤ s2[θ | log((3 + √5)/2)/2. \]

Then, w = γw2, where γ = (uv)j1 θ1 (uv)j2 θ2 · · · (uv)jr, each θj belongs to {uvw, uuv} and w2 is a prefix of a word in {uvw, uuv}. Moreover, ℓ ≤ 2Tr/|θν| + 1.

Proof. Let γ be the largest prefix of the word w than can be written in the form γ = (uv)j1 θ1 (uv)j2 θ2 · · · (uv)jr, where each θj belongs to {uvw, uuv}, and where, possibly, sj = 0. Then, we claim that:

- If r((uv)j1) ≤ r − 8|wv|, then θj = θj+1, for 1 ≤ j ≤ ℓ − 2.
- If r((uv)j1) ≤ r − 8|wv|, then |sj − sj+1| ≤ 1 for 1 ≤ j ≤ ℓ − 2.
- sj ≥ s − j, for all 1 ≤ j ≤ ℓ − 1.

Indeed, to see the first claim, if θj ̸= θj+1 note that

\[ r(θj(uv)j1 θj+1) ≤ r(θj) + r((uv)j1) + r(θj+1) + 4 \]

\[ ≤ 1.8|uvw| + r − 8|wv| + 1.8|uvw| + 8 ≤ r − 2. \]

If θj = uuv, let (α, β) = (u, uuv) and, if θj = uuv, let (α, β) = (uv, u).

So α β r = θj(uv)j1 θj+1 is a subword of a word in Σ(r−2)(3 + e−r). Then Lemma 3.21 says that this word is contained in a word that can be written in the alphabet {α, β}. But, if θj+1 ̸= θj, this would be impossible.
To prove the second claim, consider the subword $\theta_{j-1}(uv)^{\ell}\theta_j(uv)^{j+1}$. Since $\theta_{j-1} = \theta_j$, using the appropriate pair $(\alpha, \beta) \in \overline{P}$ 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.

We now prove the third claim. Observe that $s_1 \geq s$ by construction of $\gamma$. Writing $(uv)^{s_1}(uv)^{s_2}$ in the appropriate alphabet $(\alpha, \beta) \in \overline{P}$, we will get a word of the form $\theta^*\alpha\beta\theta^{s_2}$ and by hypothesis $r(\theta^*) \leq r - 2|\alpha\beta|$, so $s_2 + 1 \geq s - 1$. By induction suppose that $s_j \geq s - j$. If $r((uv)^{s_j}) \leq r - 8|uv|$ then, from the previous claim, we get $s_j+1 \geq s - j - 1$. Otherwise $(uv)^{s_j}$ is not a subword of $(uv)^s$, so $s_j+1 \geq s$ which is stronger. This proves the three claims.

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

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

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

On the other hand using Lemma 7.1 one has that

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

$$\geq (s_1 + \cdots + s_\ell + \ell)|\theta|$$

If $\ell \geq s$ then one gets that

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

$$\geq \frac{s^2|\theta|}{2} + \frac{s|\theta|}{2} + |\theta| > \frac{s^2|\theta|}{2} + 4,$$

contradicting the hypothesis. Thus $\ell < s$. Hence

$$Tr \geq (\ell - 1)(s - \ell/2 + 1)|\theta|$$

This shows that $\ell \leq 2Tr/|\theta^*| + 1$.

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 at $\theta_{\ell-1}$. On the other hand, when $s_\ell \geq 2$, we choose $\hat{w} \in \Sigma^{(r-2)}(3 + e^{-r})$ starting at the last $uvw$. Since $|uvw| < r/6$, Lemma 3.21 gives that it is $(u, v)$-renormalizable. We know that $\hat{w}$ is $(\hat{\alpha}, \hat{\beta})$-weakly renormalizable for some $(\hat{\alpha}, \hat{\beta}) \in \overline{P}$ with $|\hat{\alpha}\hat{\beta}| \geq r/6$, because of Lemma 3.20. Since $|uvw| < r/6$ and $\hat{w}$ has $uvw$, the word $\hat{w}$ is $(\alpha, \beta)$-weakly renormalizable for some $(\alpha, \beta) \in \{(uv, v), (u, uv)\}$. Write $\hat{w} = \hat{\gamma}\hat{w}_2$ with $\hat{\gamma} \in \langle u, v \rangle$ with $\hat{w}_2$ a prefix of $uvw$. Observe that $5|uvw| < r/1.8 < |\hat{w}|$, thus $|\hat{\gamma}| = |\hat{w}| - |\hat{w}_2| > 4|uvw|$. 

We will find a continuation of $\gamma\gamma'$ of the form

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

inside of $w_2$ which will contradict the maximality of $\gamma$, because either $w_2$ contains $\gamma'$ or $w_2$ is contained in $\gamma'$. When $s_\ell \geq 2$, then if $uvw = \alpha\alpha\beta$ we have that $\alpha\alpha\beta = uv(uvv)$ extends and $\alpha\alpha\alpha = (uv)^{3\ell+1}$ also extends. If $uvw = \beta\beta$ then
\[ \beta \beta = (uv)v^+ \] extends while if there is \( \alpha \alpha \) after \( \beta \beta \), then it must come \( \alpha^k v_a \)
(if there is no \( \beta \) after \( \alpha \alpha \) then \( \tilde{w}_2 \) starts with \( v_a \)). But note that
\[ \beta \beta \alpha^k v_a = v_a u^k \beta \alpha^k v_a = v_a (\alpha^k \beta \beta)^* \]
so \( \alpha \alpha \beta \beta \) is a subword of \( w^* \), a contradiction with Lemma 3.16. In the situation where \( s_2 = 1 \), we have that \( wuvv \alpha \beta \alpha \beta \), so if there is an \( \alpha \) afterwards, then \( \alpha \beta \alpha \alpha \) extends \( (uv)v^+ \) while \( \alpha \beta \alpha \beta \) gives a contradiction because before \( \theta_\ell \) there is \( s_{\ell-1} \geq 2 \) and we have \( \alpha \beta \alpha \beta \) extends \( \alpha \alpha \beta \beta \). When \( \theta_{\ell-1} uv = wuvv \beta \), if there is \( \beta \) afterwards then \( \alpha \beta \beta = \theta_{\ell-1}(uv)^2 \) extends while \( \alpha \beta \alpha \beta = \theta_{\ell-1} (uv) \alpha \beta \) also extends, but if there is \( \alpha \alpha \) after \( \beta \) we have the same contradiction with Lemma 3.16 in the transpose word. In the situation \( s_2 = 0 \), we have \( wuvv \beta \in \{ \alpha \alpha \beta \beta , \beta \beta \alpha \alpha \} \) and we arrive at the same contradictions as before. In summary, \( w_2 \) is a subword of \( \gamma^* \), so it is a subword of a word in \( \{uv, uv, uuv\} \).

\[ \square \]

In the case where we have a power \( a^c \) or \( b^c \), 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 \( c^r \), where \( c \in \{1, 2\} \) and \( r(c^r) \leq r - 2 \). 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 \( w = c^r \theta c^e \theta' \cdots \theta c^e \), \( \theta = c' c' c' \) and \( \gamma' \notin \{1, 2\} \), \( c = c' \) where \( \ell < 2Tr / (s \log x) \). Moreover if \( r(c^r) \leq r - 8 \) then \( s_j \) is even and \( |s_{j+1} - s_j| \in \{-2, 0, 2\} \).

**Proof.** Observe that we only have to prove that in \( w \) there is no \( c' c' c' c' \), since the sequence \( cc' c' c' \) is forbidden by Lemma 3.3. Define \( \gamma = c^r \theta c^e \theta' \cdots \theta c^e \theta' \cdots \). The fact that \( r(c^r) \leq r - 8 \) implies \( s_j \) odd and \( |s_{j+1} - s_j| \in \{-2, 0, 2\} \), is because of Lemmas 3.3 and 3.1. If \( \ell \geq s/2 \), then the set \( \{s_1, \ldots, s_{\ell-1}, s_\ell\} \) contains all the set \( \{2, 4, \ldots, s\} \). When \( c = 2 \) the inequality (7.3) and when \( c = 1 \) the inequality (7.2) gives us
\[ Tr \geq r(w) \geq r(\gamma) \geq (s + \cdots + 2) \log x = \frac{s(s+2)}{4} \log x \]
In both cases we are getting a contradiction, thus \( \ell < s/2 \). Repeating the computation one gets that \( Tr \geq s(s - \ell + 1) \log x \). Since \( \ell \leq s/2 - 1 \), we can bound the length \( \ell < 2Tr / (s \log x) \).

The next situation is where we have a \( (\alpha, \beta) \)-renormalizable word \( w \) that does not contain big powers of \( \alpha \) or \( \beta \). In this case there is an extension that can be written in the same alphabet \( (\alpha, \beta) \). In fact, this extension is almost \( (\alpha, \beta) \)-weakly renormalizable, in the sense that its tail is small; it is a prefix of a word in \( \langle \alpha, \beta \rangle \) but is not necessarily a prefix of \( \alpha \beta \).
Lemma 4.4. Let \( w \in \Sigma^{(r-2)}(3 + e^{-r}) \) be a \((\alpha, \beta)\)-weakly renormalizable word with \( |\alpha \beta| < r/40 \). Suppose that for every factors of the form \( \alpha^i \) or \( \beta^i \) of \( w \), we have \( |\alpha^i| \leq \delta |w| \) and \( |\beta^i| \leq \delta |w| \) where \( 0 < \delta < (1/2) \log((3 + \sqrt{5})/2) \) is a constant. Assume further that \( w \) contains an \( \alpha \beta \). If \( \overline{w} \in \Sigma(T \ell)(3 + e^{-r}) \) is an extension of \( w \) with \( T + 2 \leq \delta' r/4|\alpha \beta| \) where \( \delta' = 1 - 2\delta \log((3 + \sqrt{5})/2) \), then \( \overline{w} = w_1 \gamma \tau' \) where \( \gamma \in \langle \alpha, \beta \rangle, |\tau'| < |\alpha \beta| \) 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 \((\alpha, \beta)\)-weakly renormalizable. Take \( \gamma' \) to be the largest word inside \( \overline{w} \) starting at \( \gamma \) which can be written in the alphabet \( \langle \alpha, \beta \rangle \). Write \( \overline{w} = w_1 \gamma' \tau' \). We will show that \( \tau' \) is a prefix of some word in \( \langle \alpha, \beta \rangle \) and \( |\tau'| < |\alpha \beta| \). Take the last factor of \( \alpha \beta \) in \( \gamma' \), say \( \gamma' = \eta \alpha \beta \eta' \). In particular \( \eta' = \theta^r \) with \( \theta \in \{\alpha, \beta\} \). Consider the factor \( \hat{w} \in \Sigma^{(r-2)}(3 + e^{-r}) \) (that possibly extends \( \overline{w} \)) starting at this last \( \alpha \beta \). By Lemma 3.21 it can be written as \( \hat{w} = \hat{\gamma} w_2, \hat{\gamma} = \alpha \beta \cdots \) with \( \hat{\gamma} \in \langle \alpha, \beta \rangle \) and \( \hat{w}_2 \) a prefix of \( \alpha \beta \). If \( \alpha \beta \theta^r \) is strictly contained on \( \hat{\gamma} \), then if \( |\tau'| \geq |\alpha \beta| \) we will get that there is at least one more letter \( \{\alpha, \beta\} \) of \( \hat{\gamma} \) inside \( \overline{w} \) after \( \gamma' \), a contradiction with the maximality of \( \gamma' \). Hence \( |\tau'| < |\alpha \beta| \) and \( \alpha \beta \theta^r \tau' \) is a subword of \( \hat{\gamma} \alpha \beta \), so \( \tau' \) is the prefix of a word in \( \langle \alpha, \beta \rangle \) as claimed. Now suppose that \( \alpha \beta \theta^r \) contains \( \hat{\gamma} \), so \( r(\theta^r) \geq r - 4|\alpha \beta| - 6 \). We will get a contradiction by considering an extension in the transpose word \( \overline{(w \overline{w})^r} \).

We need the identity \(((w \overline{w})^r)^* = u^b(v \overline{w})^{k-1}v_a \). To prove the identity, let \((\alpha, \beta) = (uv, v)\), so \((w \overline{w})^k = v_a u^b(v \overline{w})^{k-2}v_a u^b = v_a u^b a^{k-2} \beta_a u^b \). Now using that \(u^b, v_a, \alpha^k \beta_a\) are all palindrome, one gets that \(((w \overline{w})^r)^* = u^b a^{k-2} \beta_a u^b v_a = u^b (v \overline{w})^{k-1} v_a \).

In resume there is a \( \theta^{-1} \) inside \( \overline{(w \overline{w})^r} \). Let \( \theta' \) be the maximal suffix of \( \theta^r \) that satisfies \( r(\theta') \leq r - 4|\theta| \). Since \( r(\theta') \geq r - 4|\alpha \beta| - 6 \), by Lemma 7.1 we have that \( |\theta'| > r/2 \). If \( \theta = uv \) with \((u, v) \in \hat{P}\) we use Lemma 4.2 to find a \( \tilde{\gamma} = \theta^* \theta^r \theta^2 \cdots \theta^r \) inside \( \overline{(w \overline{w})^r} \) starting at this \( \theta' = (uv)^r \) where each \( \theta \in \{u, v, wv\} \). When \( |\theta| = 2 \), we write \( \theta = c c \) with \( c \in \{1, 2\} \), \( \theta' \in \{a, b\} \) and use Lemma 4.3 to find an extension \( \tilde{\gamma} = c^r \theta' \ell c^2 \theta' \cdots c^2 \).

Observe that \( r(\overline{(w \overline{w})^r}) \leq (T + 2)r - 4 \) by Lemma 7.3 which gives

\[
(r')^2 |\theta|/4 \geq (r/8) \cdot \frac{r}{2|\theta|} \geq (T + 2)r
\]

by hypothesis.

When \( \theta = uv \), Lemma 4.2 gives us that \( \tilde{\gamma} \) covers all \( \overline{(w \overline{w})^r} = \overline{w}w^* \), which implies that \( \theta^\nu \) (or at least some factor of it) is inside of \( w^* \). Similarly, when \( |\theta| = 2 \), we will have that \( (T + 2)r \leq (r')^2/8 \) and so again \( \tilde{\gamma} \) covers \( \overline{(w \overline{w})^r} = \overline{w}w^* \), which gives that \( c^\ell \) (or at least some factor of it) is inside \( w \), so \( \theta^\ell \) is inside \( w \) where \( s_\ell = 2|c^\ell|2 \).

In any case \( \theta'^{-1} \) is inside of \( w \) (because of the identity proved above). In any case we will also get that \( s_\ell \geq s' - 1 - \ell \) and also \( \ell \leq 2Tr/|\theta'\sqrt{r}| + 1 \). But note that
using $|\theta^r| > r/2$ and Lemma 7.1

$$|\theta^{r-1}| \geq |\theta^{r-1-\ell}| = |\theta^r| - |\theta^{\ell+1}| \geq r/2 - (\ell + 1)|\theta|$$

$$\geq r/2 - (2Tr/|\theta^r| + 2)|\theta| \geq r/2 - 2(2T + 1)$$

$$\geq r/2 - \frac{|\theta|}{|\alpha\beta|} \cdot \delta' r + 6|\theta| \geq (1/2)(1 - \delta')r + 6|\theta| \geq \delta |w|$$

which is a contradiction with the existence of those factors inside $w$. In conclusion $w\overline{w} = w_1\gamma\tau'$ where $|\tau'| < |\alpha\beta|$ 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. This section will be devoted to the proof of Theorem 1.2. The main idea is to find a subcovering of the natural covering of $\mathcal{M} \cap (-\infty, t) \subseteq (\mathbb{N}^* \cap [1, [t]]) + K_t + K_t$, for a suitable Cantor set $K_1$, and then use classical techniques to estimate the Hausdorff dimension of $K_t$, and a fortiori, the Hausdorff dimension of $\mathcal{M} \cap (-\infty, t)$: 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 such 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. Moreover, we will need the following combinatorial lemma.

Lemma 4.5. Given a positive integer $U$ and a real number $m$, the number of natural solutions $(\ell, x_1, x_2, \ldots, x_\ell)$ of $x_1 + x_2 + \cdots + x_\ell \leq (U - \ell)m$ is at most

$$U \left( \frac{em^e m}{1 - \epsilon_m} \right)^{(1 - \epsilon_m)(U + 1)} = U e^{(1 - \epsilon_m)(U + 1)/\epsilon_m},$$

where $\epsilon_m$ is the solution in $(0, 1)$ of the equation $\log \left( \frac{em^e m}{1 - \epsilon} \right) = \frac{1}{\epsilon}$. In particular, if $U = m^{o(1)}$, then this number is at most $e^{(\log m - \log \log m + o(1))(U + 1)}$.

Proof. We should have $1 \leq \ell \leq U$. Given such $\ell$, the number of solutions of this inequality is the number of natural solutions of $x_0 + x_1 + \cdots + x_\ell = \lfloor (U - \ell)m \rfloor$, which is equal to $\left( \lfloor (U - \ell)m \rfloor \right)$, and using the inequalities $\left( \frac{n}{\ell} \right) \leq \left( \frac{m}{\ell} \right)^{\ell}$, which hold for $1 \leq k \leq n$, this number of solutions is at most

$$\left( \frac{e\left( \lfloor (U - \ell)m \rfloor + \ell \right)}{\ell} \right)^{\ell} \leq \frac{e\left( (U - \ell)m + \ell \right)}{\ell} \leq \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \right)^{\ell},$$

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

$$\log \left( \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \right)^{\ell} = \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \log \left( \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \right) = \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \left( \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \right)^{\ell}$$

$$\log \left( \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \right)^{\ell} = \frac{\widetilde{U}}{\ell} - \frac{\ell}{\ell} = \log \left( \frac{e\left( \widetilde{U} - \ell \right)m}{\ell} \right)^{\ell} - \frac{1}{\epsilon},$$
Moreover, since there are $U$ possible values of $\ell$, the number of solutions we are estimating is at most

$$U : e^{e((1-\epsilon_n)\tilde{U})} = U e^{(1-\epsilon_n)(U+1)/\epsilon_n},$$

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

If $U = m^{o(1)}$, for $1 \leq \ell \leq U$, then $\log \left(\frac{e((\ell-1)m)}{\ell} \right) = (1+o(1)) \log m$. Thus, the maximum of $g(\ell)$ is attained for $\ell = \tilde{U} \left(1 - \frac{1+o(1)}{\log m}\right)$, and, hence, the maximum of $g(\ell)$ is $\left(\frac{1+o(1)}{\log m}\right) \tilde{U} \left(1 - \frac{1+o(1)}{\log m}\right)$. Since the number of possible choices for $\ell$ is $U$, the total number of solutions is

$$O \left( U \left(\frac{1+o(1)}{\log m}\right) \tilde{U} \left(1 - \frac{1+o(1)}{\log m}\right) \right) = O \left( e^{\log m - \log \log m + o(1)} \left(1 - \frac{1+o(1)}{\log m}\right) \right),$$

where the first equality uses the fact that

$$U = \left(\frac{1+o(1)}{\log m}\right) \log U / \log m \leq \left(\frac{1+o(1)}{\log m}\right) \tilde{U} / \log m \leq \left(\frac{1+o(1)}{\log m}\right) \log \tilde{U} / \log m.$$

\[ \square \]

**Proof of Theorem 1.2** Let us recall how the covering of the spectra is constructed. Define the sets of words

$$C(t, r) = \{ \alpha = (a_1, \ldots, a_n) \in P_r \mid K_t \cap I(\alpha) \neq \emptyset \} = \{ \alpha \in P_r \mid \alpha \text{ subword of a word } \omega \in (\mathbb{N}^*)^N \text{ with } m(\omega) \leq t \}$$

Here $K_t = \{ [0; \gamma] \mid \gamma \in \pi_{\infty}(\Sigma_t) \}$ where $\pi_{\infty} : \Sigma \to \Sigma^*$ is the projection associated to the decomposition $\Sigma = \Sigma^* \times \Sigma^* = (\mathbb{N}^*)^N \times (\mathbb{N}^*)^N$. That is,

$$\pi_{\infty}(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots) = (a_0, a_1, a_2, \ldots).$$

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

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

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 the set $C(t, s)$, in the sense for any interval $I(\alpha)$ with $\alpha = (a_1, \ldots, a_n) \in C(t, s)$ there is $m \leq n$ such that $\tilde{\alpha} = (a_1, \ldots, a_m) \in C(t, r)$ and $I(\tilde{\alpha}) \subseteq I(\alpha)$. 

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

$$U : e^{e((1-\epsilon_n)\tilde{U})} = U e^{(1-\epsilon_n)(U+1)/\epsilon_n},$$
To prove that the Hausdorff dimension of $K_t$ is at most $d$, we replace an interval $I$ (corresponding to $C(t, r)$) with several intervals $I_j$ contained in it, but smaller and of different sizes, each one corresponding to a word in $C(t, Tr)$, where $T \in \{10, \lfloor \log^2 r \rfloor, \lfloor r/3 \rfloor\}$, whose union still contains the intersection of $K_t$ with $I$ and that satisfy $\sum_j |I_j|^d < |I|^d$.

By the renormalization algorithm (see Lemma 3.20), if $w \in \Sigma^{(r+2)}(3 + e^{-r-1})$, then there is a sequence of alphabets $(\alpha_j, \beta_j)$ such that, for all $0 \leq j \leq m$, $w$ is $(\alpha_j, \beta_j)$-renormalizable, with

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

for each $0 \leq j < m$ and $|\alpha_m \beta_m| \geq r/6$.

We consider such a renormalization $(\alpha_t, \beta_t)$ with

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

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 $\tilde{w} \in \Sigma^{(r)}(3 + e^{-r-4})$ for some $T \in \{10, \lfloor \log^2 r \rfloor, \lfloor r/3 \rfloor\}$ such that $w \tilde{w} \in \Sigma(3 + e^{-r-4}, |w\tilde{w}|)$.

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 $\tilde{w} \in P_{r+2}$ of $w\tilde{w}$ starting with this factor $\alpha_t \beta_t$; it should be $(\alpha_t, \beta_t)$-renormalizable. We will attempt to use this argument several times in order to cover the whole word $w\tilde{w}$ by $(\alpha_t, \beta_t)$-renormalizable words.

For each $\tilde{w}$, we only need to estimate the number of words in $(\alpha_t, \beta_t)$ after the last factor equal to $\alpha_t \beta_t$ in $w$. For this sake, we consider several cases according to the size of $\alpha_t^i, \beta_t^i$ as a factor of $w$, for some integer $s$.

Case 1: Suppose first that, for every factor of the form $\alpha_t^i \beta_t^j$ of $w$, we have $|\alpha_t^i| < |w|/3$ and $|\beta_t^j| < |w|/3$.

Then there is a factor $\alpha_t \beta_t$ in the first half of $w$ and until the next appearence 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 $\alpha_t \beta_t \alpha_t^i \beta_t$ or $\alpha_t \beta_t \alpha_t \beta_t$ 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 $\overline{w} \in \Sigma^{(10r)}(3 + e^{-r-4})$, then Lemma 4.4 gives that $\tilde{w} \overline{w} \tilde{w} \overline{w} = \tau \gamma \tau'$ with $\gamma \in (\alpha_t, \beta_t)$, $\tau$ a suffix of $\alpha_t \beta_t$ and $\tau'$ 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 $\tilde{w} \overline{w}$ is a concatenation of factors of the form $\alpha_t^i \beta_t$ or $\alpha_t \beta_t^j$. The number of such factors is at most $20\sqrt{\log r}$. Moreover, if we have to consecutive such factors $\alpha_t^i \beta_t$ and $\alpha_t^j \beta_t$ (or $\alpha_t \beta_t^i$ and $\alpha_t \beta_t^j$), then $|j_1 - j_2| \leq 1$, and if we have two consecutive factors $\beta_t \alpha_t^i \beta_t$ and $\alpha_t \beta_t \alpha_t \beta_t$, then $2 \leq |j_1| + |j_2| \leq 8$. This implies that each of these factors of the form $\alpha_t^j \beta_t$ or $\alpha_t \beta_t^i$ has at most 3 continuations of this form, and so the number of such continuations $\tilde{w} \overline{w}$ of $w$ is at most $3^{20\sqrt{\log r}} < r$. Since the number of possible $w \in \Sigma^{(r)}(3 + e^{-r-4})$ is $O(r^3)$ then the number of possible continuations $\tilde{w} \overline{w}$ is this case is $O(r^4)$. 
Case 2: Suppose now that \( w \) has a factor \( \alpha_j^t \) with \( |\alpha_j^t| \geq |w|/3 \) (the case of \( w \) having a factor \( \beta_j^t \) with \( |\beta_j^t| \geq |w|/3 \) will be analogous).

Let us check that we can apply Lemmas 4.2 and 4.3 to this factor. Observe that \( r/18 \leq s(\alpha_i) \) and that \( s \geq r/(18|\alpha_i|) > \sqrt{\log r}/36 \), so the condition \((T + 2)r \leq s^2|\alpha_i|/8 \) holds for \( T = 10 \). In particular \( \ell \leq 2Tr/|\alpha_i^t| + 1 < 70 \). Hence one has that \((s - \ell/2)|\alpha_i| > r/4 \) for sufficiently large \( r \). Going back to the inequality \( Tr \geq (\ell - 1)(s - \ell/2)|\alpha_i| \) we get a stronger bound \( \ell < Tr/((s - \ell/2)|\alpha_i|)+1 \leq 41 \).

We consider two subcases depending on the length of \( \alpha_i \).

Case 2.1: Suppose that \( |\alpha_i| > r^{7/8} \) and \( \alpha_i = uv \) with \( (u, v) \in \mathcal{P} \).

Then, we claim that after \( \alpha_j^t = (uv)^t \) the word can be renormalizable with alphabet \{\( u \), \( v \)\}, and the first appearance of \( uv \) or \( vu \) determines the new alphabet \{\( \{u, uv\} \) or \( \{v, vu\} \}\}. To prove that, let \( \tilde{w} \in \Sigma((T+2)|v|)(3 + e^{-r^{-4}}) \) be the factor of \( w\tilde{w}w \) starting at that factor \( \alpha_j^t \). Therefore Lemma 4.2 gives that \( \tilde{w} = \hat{w} \tilde{w}_2 \), where \( \hat{w} = (uv)^t \theta_1 (uv)^s \theta_2 \cdots (uv)^t \), and each \( \theta_j \) belongs to \{\( uv, uu \)\} and moreover \( \ell \leq 40 \).

In particular given a continuation \( \tilde{w}w \) of \( w \) with \( \tilde{w} \in \Sigma(r)(3 + e^{-r^{-4}}) \) and \( \bar{w} \in \Sigma(10r)(3 + e^{-r^{-4}}) \), from the first such factor \( (uv)^t \), the sequence should be \( (uv)^t \theta_1 (uv)^s \theta_2 \cdots (uv)^t \), with \( \theta_j \in \{uv, uu\} \), \( \ell \leq 40 \) and \( s_1 + \cdots + s_j \leq 20r^{1/8} \), so we have in total at most \( 2^{40} \) choices for the \( \theta_j \), and, given \( \ell \leq 40 \), the number of choices for the \( s_j \) is at most the number of natural solutions of \( x_1 + x_2 + \cdots + x_{\ell+1} = 20r^{1/8} \) = \( M \), which is \( (\ell M^{\ell}) < (21r^{1/8})^{40} = 21^{10}r^{5} \), and so the total number of such words \( w\tilde{w}w \) is \( O(r^{3} \cdot 2^{40} \cdot 40 \cdot 21^{10}r^{5}) = O(r^{8}) \).

Case 2.2: Suppose \( |\alpha_i| \leq r^{7/8} \) and that the largest factor \( \alpha_j^t \) of \( w\tilde{w} \) satisfies \( r(\alpha_j^t) \leq r - 80|\alpha_i| \).

We claim there is \( \beta \) such that \( (\alpha_i, \beta) \in \mathcal{P} \), \( \beta(\alpha_j^t) \beta \) is a subword of \( w\tilde{w} \), and the continuation \( w\tilde{w}w \) has the form \( \beta(\alpha_i^t) \beta(\alpha_i^t) \cdots (\alpha_i^t) \) with \( |s_j - s_{j+1}| < 1 \) and for all \( 1 \leq j \leq \ell \leq 40 \).

If \( \alpha_i = uv \) with \( (u, v) \in \mathcal{P} \) then we use Lemma 4.2 to obtain a continuation \( \hat{w} \) with \( \hat{w} = (uv)^t \theta_1 (uv)^s \theta_2 \cdots \theta_j \in \{uv, uu\} \) and \( \hat{w} = \hat{w}_{\ell+2} \). Moreover \( \ell \leq 40 \). Since \( r(\alpha_i^t) \leq r - 80|\alpha_i| \), by induction \( r(\alpha_j^t) \leq r - (82 - 2j)|\alpha_i| \) and since \( \ell \leq 40 \) by the proof of that Lemma 4.2 we get that all \( \theta_j \) are equal and that \( |s_j - s_{j+1}| < 1 \) for all \( j \geq 1 \). We chose \( \beta \) such that \( \alpha_j \beta = \theta_j \).

When \( \alpha_i = a \), we use Lemma 4.3 to find a continuation \( \bar{w} \) such that \( \bar{w} = 2^{2}b2^{2}b \cdots \). But observe that \( 2^{2} = \alpha_j^2/2 \) is inside \( w \), so by hypothesis \( r(2^{2}) < r - 80|\alpha_i| \) and there is \( b \) before \( 2^{2} \), so \( e_2 = 2 \) is even and \( |e_1 - e_2| = \{-2, 0, 2\} \). By induction we obtain \( r(2^{2}) \leq r - (82 - 2j)|\alpha_i| \), which forces all \( e_j \) to be even and \( |e_{j+1} - e_j| = \{-2, 0, 2\} \). In this case we set \( \beta = b \), so we get \( \bar{w} = 2^{2}b2^{2}b \cdots = \alpha_j^t \beta \alpha_j^t \beta \cdots \) with \( |s_j - s_{j+1}| < 1 \) for all \( j \geq 1 \). Therefore, given a continuation \( w\tilde{w} \) of \( w \) with \( \tilde{w} \in \Sigma(r)(3 + e^{-r^{-4}}) \) and \( \bar{w} \in \Sigma(10r)(3 + e^{-r^{-4}}) \), there is \( \beta \) with \( (\alpha_i, \beta) \in \mathcal{P} \) such that \( w\tilde{w} \) has a factor \( \beta(\alpha_i^t) \beta \), after which the continuation of \( w\tilde{w} \) is a concatenation of at most 39 sequences...
of the type $(\alpha_t)^{\beta_t} \beta_t \leq j \leq 40$ with $|s_{j+1} - s_j| \leq 1$ for every $j \geq 1$. This gives at most $3^{40}$ continuations of $\beta(\alpha_t)^{\beta_t} \beta_t$, and so, since we have at most $O((r^3)^2) = O(r^6)$ choices for $w \bar{w}$, we have in total, $O(2^{40} \cdot r^6) = O(r^6)$ such words $w \bar{w} \bar{w}$.

In both cases, Case 2.1 and Case 2.2, if $d = \frac{\log r - \log \log r}{r}$, then

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

and we have $O(r^8)$ possible such words $w \bar{w} \bar{w}$. Notice that

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

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

**Case 3:** On the same conditions of case 2, suppose that $|\alpha_t| \leq r^{7/8}$ and that $w \bar{w}$ has a factor $\alpha_t^{s_t}$ satisfying $r(\alpha_t^{s_t}) \geq r - 80|\alpha_t|$.

We will consider in this case continuations $\bar{w} \in \Sigma(Tr) (3 + e^{-r-4})$ for $T = \lfloor \log_2 r \rfloor$ such that $w \bar{w} \bar{w} \bar{w} \bar{w} \in \Sigma(3 + e^{-r-4}, |w \bar{w} \bar{w} \bar{w} \bar{w}|)$. Again, consider two subcases depending on the length of $\alpha_t$.

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

So, $\alpha_t = uv$ with $(u, v) = \overrightarrow{P}$. Now let $T = \lfloor (\log r)^2 \rfloor$. The condition on $\alpha_t^{s_t}$ implies that $s_t \geq (r - 80|\alpha_t|)/|\alpha_t| \geq r^{1/8} - 80$, so we have that $s_t^2/|\alpha_t|/2 \geq r^{1/8}(r - 80r^{1/8})/2 \geq (\log r)^2 r \geq Tr$. Let $\bar{w} \in \Sigma((T + 2)r) (3 + e^{-r-4})$ be the factor of $w \bar{w} \bar{w} \bar{w} \bar{w}$ starting at that factor $\alpha_t^{s_t}$. Lemma 4.2 guarantees that

$$\bar{w} = (uv)^{s_t} \ldots \cdot (uv)^{t} \bar{w}_2$$

where each $\theta_t \in \{u, v, uv\}$ and $\ell \leq 2Tr/|\alpha_t^{s_t}| + 1 \leq 2T + 2$ for $r$ big enough. Putting this again into the inequality gives a better bound

$$\ell \leq Tr/((s_1 - \ell/2)|\alpha_t|) \leq T + 3$$

for large enough $r$.

Therefore, from this factor $\alpha_t^{s_t}$, the continuation of $w \bar{w} \bar{w} \bar{w} \bar{w}$ is an initial factor of a word of the form $(uv)^{s_t} \ldots \cdot (uv)^{t} \bar{w}_2$ and

$$Tr \leq l((uv)^{s_t} \ldots \cdot (uv)^{t} \bar{w}_2) < (T + 3)r,$$

with $\theta_t \in \{u, v, uv\}$, $\ell \leq T + 3$ (and such that $(uv)^{s_t} \ldots \cdot (uv)^{t} \bar{w}_2$ is an initial factor of this word beginning in this factor $\alpha_t^{s_t}$ and going till the end of $w \bar{w} \bar{w} \bar{w} \bar{w}$) with $r(\alpha_t^{s_t}) \geq r - (83 + T)|\alpha_t| > r - [2\log r \cdot r^{7/8}] =: M$ (notice that if $r((uv)^{s_t}) < r - 2|\alpha_t|$ then $|s_{j+1} - s_j| \leq 1$). Let $s_0$ be the smallest integer satisfying $r(\alpha_t^{s_0}) \geq M$. Then, $s_j = s_0 + \bar{s}_j$ with $\bar{s}_j \geq 0$ for each $1 \leq j \leq \ell$.

Since $q_{2, |\alpha_t|}(\alpha_t^{s_t}) \geq q_{2, |\alpha_t|}(\alpha_t)^r$ and $q_{2, |\alpha_t|}(\alpha_t^{s_t}) \geq q_{4}(1122) = 12$, we have $r(\alpha_t^{s_t}) \geq [\log(q_{2, |\alpha_t|}(\alpha_t^{s_t})^2)] \geq [\log((12)^2)] = [\log(144)] > 4s$. Since

$$(T + 3)r > r((uv)^{s_t} \ldots \cdot (uv)^{t} \bar{w}_2) \geq \ell r(\alpha_t^{s_0}) + (\bar{s}_1 + \bar{s}_2 + \cdots + \bar{s}_\ell) r(\alpha_t^{s_t}) \geq \ell M + 4(\bar{s}_1 + \bar{s}_2 + \cdots + \bar{s}_\ell).$$
Since \((T + 4)M = (T + 4)(r - [2 \log^2 r \cdot r^{7/8}]) > (T + 3)r\), it follows that, given \(1 \leq \ell \leq T + 3\), the number of choices of the \(s_j, j \leq \ell\) is at most the number of natural solutions of
\[
\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_{\ell} \leq (T + 4 - \ell)M/4,
\]
which is \((\lfloor (T + 4 - \ell)M/4 + \ell \rfloor)^\ell < \left(\frac{e((T - \ell)M/4)}{\ell}\right)^\ell\), where \(\tilde{T} := T + 5\) (here we used the inequalities \(\frac{\ell}{\tilde{T}} \leq \frac{1}{\ell + 1} \leq \frac{1}{\ell + 1}\), which hold for \(1 \leq k \leq n\)). We have at most \(2^\ell\) choices for the \(\theta_j, j \leq \ell\); let us estimate the maximum of \(f(\ell) = 2^\ell \left(\frac{e((\tilde{T} - \ell)M/2)}{\ell}\right)^\ell = \left(\frac{e((T - \ell)M/2)}{\ell}\right)^\ell\) for \(1 \leq \ell \leq \tilde{T} - 1\). The derivative of \(\log f(\ell)\) is \(\log \left(\frac{e((T - \ell)M/2)}{\ell}\right) = (1 + o(1)) \log M = (1 + o(1)) \log r\)
and \(\tilde{T} = T + 5 = \log^2 r + O(1)\), we have the maximum attained for \(\ell = \tilde{T}(1 - \frac{\log(1 + o(1))}{\log r}) = \log^2 r - (1 + o(1)) \log r < T\), and, for such value of \(\ell\),
\[
f(\ell) = \left(\frac{e((\tilde{T} - \ell)M/2)}{\ell}\right)^\ell = \left(\frac{e((1 + o(1))M/2)}{\log r}\right)^\ell < \left(\frac{3M}{2 \log r}\right)^{T - (1 + o(1)) \log r}.
\]
We have at most \(\tilde{T} = \log^2 r + O(1)\) choices for \(\ell\), and we have at most \(O(r^6)\) choices for \(w\tilde{w}\tilde{w}\), so we have at most
\[
O\left(r^6 \log^2 r \left(\frac{3M}{2 \log r}\right)^{T - (1 + o(1)) \log r}\right) < \left(\frac{2r}{\log r}\right)^{T - (1 + o(1)) \log r}
\]
such words \(w\tilde{w}\tilde{w}\).

Notice that, for \(d = \frac{\log r - \log \log r}{r}\), we have \((e^{-Tr})^d = e^{-Tr(\log r - \log \log r)}\), 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
\[
\left(\frac{2r}{\log r}\right)^{T - (1 + o(1)) \log r} (e^{-Tr})^d \leq \left(\frac{2r}{\log r}\right)^{- (1 + o(1)) \log r} \cdot 2^{\log^2 r}
\]
\[
= e^{-(1 + o(1)) \log^2 r} \cdot e^{\log 2 \log^2 r}
\]
\[
= e^{(1 - \log 2 \log^2 r)}
\]
\[
< e^{-\frac{1}{4}\log^2 r} \ll 1.
\]

**Case 3.2:** Suppose that \(|\alpha| = 2\).

Then, for some \(c \in \{1, 2\}\), \(w\tilde{w}\) has a factor \(c^\ell\) satisfying \(r(c^\ell) \geq r - 80\). Let
\[ d = 3 - c \in \{1, 2\} \text{ and } \theta = dd. \]

Observe that \((s + 1) \log x \geq r(c^i) \geq r - 80\), so \((T + 2)r \leq (s - 2)^2 (\log x)/4 \) holds. Using Lemma 4.3 from this factor \(c^1\), the continuation of \(w\bar{w}w\) has the form \(c^1 \theta c^2 \theta \cdots \theta c^\ell\) with \(\ell < 2T r/(s - 4 \log x) < 2T r/(r - 80) < 3T\) for large \(r\). Using this information in the inequality \((s - \ell - 3) \log x < Tr\), gives that \(\ell \leq T + 1\) for large \(r\). Therefore, from this factor \(c^1\), the continuation of \(w\bar{w}w\) is an initial factor of a word of the form \(c^1 \theta c^2 \theta \cdots c^\ell\), \(T r \leq r(c^1 \theta c^2 \theta \cdots c^\ell) \leq (T + 1)r, \ell \leq T + 1\) (and such that \(c^1 \theta c^2 \theta \cdots c^{\ell-1} \theta\)

is an initial factor of this word beginning in this factor \(c^1\) and going till the end of \(w\bar{w}w\)) with \(r(c^i) \geq r - (81 + 2T) > r - [3 \log^2 r] =: N\), notice that if \(r(c^i) < r - 1\) then \(s_j\) is even and \(|s_{j+1} - s_j| \in \{-2, 0, 2\}\). Let \(s_0\) minimum such that \(r(c^{\ell}) \geq N\). Then \(s_j = s_0 + \tilde{s}_j\) with \(\tilde{s}_j \geq 0\) for each \(1 \leq j \leq \ell\). Notice that, given \(c, w\), is determined by the choice of \((\ell, s_1, s_2, \ldots, s_\ell)\).

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

Case 3.2.1: Assume that \(c = 2\):

Since \(q_s(2^i) \geq 2^i\), we have

\[
|r(2^i)| \geq \left\lfloor \log(q_s(2^i))^2 \right\rfloor \geq \left\lfloor \log((2^i)^2) \right\rfloor = \left\lfloor s \log(4) \right\rfloor \geq 4s/3 - 1.
\]

We have

\[
(T + 1)r > r(2^i 112^2 11 \cdots 2^i) \geq \ell r(2^0) + 4(\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_\ell)/3 - \ell
\]

\[
\geq \ell N + 4(\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_\ell)/3 - \ell.
\]

Since

\[
(T + 2)N = (T + 2)(r - [3 \log^2 r]) > (T + 1) r + (T + 1) \geq (T + 1) r + \ell,
\]

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 \(\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_\ell \leq 3(T + 2 - \ell)N/4\). By Lemma 4.5, it is at most

\[
e^{(log(3N/4 - log log(3N/4) + o(1))(T+3))} = e^{(log N - log log N - log(4/3) + o(1))T}.
\]

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

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

such words \(w\bar{w}w\).

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 \(\log N = \log r + o(1)\),

\[
e^{(log N - log log N - log(4/3) + o(1))T} (e^{-Tr})^d = e^{T(\log r - \log \log r - log(4/3) + o(1))}\]

\[
e^{(log N - log log N - log(4/3) + o(1))T} (e^{-Tr})^d = e^{T(o(1) - log(4/3))} < e^{-\frac{\log^2 r}{2}} < 1.
\]

Case 3.2.2: Assume that \(c = 1\).

Observe that

\[
(T + 1) r > r(1^i 221^2 22 \cdots 221^i) \geq \ell N + (\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_\ell) \log \left(\frac{3 + \sqrt{5}}{2}\right)
\]
by (7.2). 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

\[\tilde{s}_1 + \tilde{s}_2 + \cdots + \tilde{s}_\ell \leq (T + 2 - \ell)N/\log \left(\frac{3 + \sqrt{5}}{2}\right).\]

By Lemma 4.5, it is at most

\[e^{\left(\log \left(N/\log \left(\frac{3 + \sqrt{5}}{2}\right)\right) - \log \log \left(N/\log \left(\frac{3 + \sqrt{5}}{2}\right)\right) + o(1)\right)\cdot T + 3},\]

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

\[O\left(e^{\left(\log \left(N/\log \left(\frac{3 + \sqrt{5}}{2}\right)\right) - \log \log \left(N/\log \left(\frac{3 + \sqrt{5}}{2}\right)\right) + o(1)\right)\cdot T}\right),\]

such words \(w\tilde{w}\tilde{w}\).

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

\[(e^{-Tr})^d = e^{-dT} = e^{-T \left(\log r - \log \log r - \log \log \left(\frac{3 + \sqrt{5}}{2}\right) + \delta\right)},\]

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

\[e^{\left(\log \left(N/\log \left(\frac{3 + \sqrt{5}}{2}\right)\right) - \log \log \left(N/\log \left(\frac{3 + \sqrt{5}}{2}\right)\right) + o(1)\right)\cdot T \cdot e^{-Tr})^d},\]

\[= e^{T \left(\log r - \log \log r - \log \log \left(\frac{3 + \sqrt{5}}{2}\right) + o(1)\right) - \log r + \log \log r + \log \log \left(\frac{3 + \sqrt{5}}{2}\right) - \delta}\]

\[= e^{T (o(1) - \delta)} < e^{-\frac{\delta \log^2 r}{2}} \ll 1.\]

Since \(c_0 := -\log \left(\frac{3 + \sqrt{5}}{2}\right) = 0.03830054\ldots > 0\), 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 \(c = 1\).
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 the case \( c = 1 \).

Observe that it is possible to chose \( T = \lfloor r/5 \rfloor \) in Lemma 4.3 because \( r(1^4) \geq r - 80 \) gives us that \( s \log x \geq r - 80 \) so one has that

\[
\frac{s^2}{4} \geq (r-80)^2/(4 \log x) \geq \frac{r^2}{5}
\]

for large \( r \). Moreover

\[
\ell < 2Tr/(s \log x) \leq 2Tr/(r-80) \leq 5/2T < r/2
\]

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

Let \( T = \lfloor r/5 \rfloor \). We would have a worst lower estimate for \( r(a_i^k) \): for \( i \geq 1 \), we have \( r(a_i^k) \geq r - 2(83 + i) \geq r/3 \). Indeed, \( \frac{r^2}{5} + r \geq r(1^4)221^{s_2}22 \ldots 221^{r_2} \geq \sum_{i=1}^{\min(\ell,r/2)} (r - 2(83 + i)) = \min(\ell,r/2)(r - 165 - \min(\ell,r/2)) \), implies \( \ell < 3r/10 \), and thus \( r(a_i^k) \geq r - 2(83 + i) > r/3 \). We will introduce a parameter \( j \) equal to the number of values of \( i \) for which \( r(a_i^k) < r - 2 \) in \( 1^4 221^{s_2}22 \ldots 221^{r_2} \), 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 \( s_i \geq r - 2 \)); if we consider these \( j \) values \( i_1 < i_2 < \cdots < i_j \) of \( i \), we have \( s_i \geq s_0 - 100 \), so \( s_i > s_0 - 100 - 2t \), \( 1 \leq t \leq j \), and \( \sum_{1 \leq i \leq j} s_i > j \cdot (s_0 - 100) \).

Let \( \ell = \ell - j \) and \( \{s_i, i \in I = \{1, 2, \ldots, \ell\} \setminus \{i_1, 1 \leq t \leq j\} = \{s_{\hat{i}_1}, s_{\hat{i}_2}, \ldots, s_{\hat{i}_\ell}\} \). We have \( \ell < 3r/10 < 2T \). Given \( \ell \) and \( j \) there are at most

\[
\frac{(\ell)}{j} = \left( \frac{\ell + j}{j} \right) < \left( \frac{e\ell}{j} \right)^j < \left( \frac{2eT}{j} \right)^j
\]

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

Let \( \hat{s}_0 \) minimum such that \( r(1^k) \geq r - 2 \). We have \( s_0 > (r - 4)/\log \left( \frac{2 + \sqrt{3}}{2} \right) \).

The number of solutions of the above inequality is at most the number of natural solutions of

\[
\hat{s}_1 + \hat{s}_2 + \cdots + \hat{s}_\ell \leq (T + 2 - \ell)(r - 4)/\log \left( \frac{3 + \sqrt{3}}{2} \right) - j \cdot (s_0 - 100 - j) < (T + 2 - \ell - j \cdot (r - 104 - j)/r)(r - 4)/\log \left( \frac{3 + \sqrt{3}}{2} \right) < (T + 2 - \ell - j/2)(r - 4)/\log \left( \frac{3 + \sqrt{3}}{2} \right).
\]
The number of solutions of
\[
\hat{s}_1 + \hat{s}_2 + \cdots + \hat{s}_\ell \leq (T + 2 - \hat{\ell} - j \cdot (r - 104 - j)/r) (r - 4)/\log \left( \frac{3 + \sqrt{5}}{2} \right)
\]
is at most
\[
(T + 2) \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right)^{(1 - \varepsilon_m) (T + 3 - j (r - 104 - j)/r)}
= (T + 2) e^{(1 - \varepsilon_m)(T + 3 - j (r - 104 - j)/r)/\varepsilon_m},
\]
where \(\varepsilon_m\) is the solution in (0, 1) of the equation \(\log \left( \frac{e m}{1 - \varepsilon_m} \right) = \frac{1}{\varepsilon_m}\), with \(m = (r - 4)/\log \left( \frac{3 + \sqrt{5}}{2} \right)\).

Since \(\frac{em \varepsilon_m}{1 - \varepsilon_m} > \frac{r}{2 \log r}\), the factor \(\left( \frac{em \varepsilon_m}{1 - \varepsilon_m} \right)^{-j (r - 104 - j)/r}\) is such that
\[
\left( \frac{6 e T}{j} \right)^j \left( \frac{em \varepsilon_m}{1 - \varepsilon_m} \right)^{-j (r - 104 - j)/r} < \left( \frac{6 e T}{j} \left( \frac{r}{2 \log r} \right)^{(r - 104 - j)/r} \right)^j.
\]
This is smaller than \(\left( \frac{6 e T}{j} \left( \frac{r}{2 \log r} \right)^{-1/2} \right)^j\), 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
\[
\left( \frac{6 e T}{j} \left( \frac{r}{2 \log r} \right)^{(r - 104 - j)/r} \right)^j
\]
will be \(o(1)\) since \(- (r - 104 - j)/r < -1 + r^{-1/5}\) and \(\left( \frac{r}{2 \log r} \right)^{-1/5} = 1 + o(1)\), so the estimate becomes \((3 + o(1))e/10)^{10 \log r} = o(1)\). On the other hand, for \(0 \leq j < 10 \log r\), the estimate \(\left( \frac{6 e T}{j} \left( \frac{r}{2 \log r} \right)^{(r - 104 - j)/r} \right)^j\) becomes \(\left( \frac{3 + o(1)}{j} \right)^j \cdot \left( \frac{9 \log r}{j} \right)^j\). The maximum of the function \(v(j) = \left( \frac{9 \log r}{j} \right)^j\) 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 \(O(r^6)\) choices for \(\hat{w}\), in any case we get an upper estimate for the total number of words \(\hat{w}\) which is
\[
O(r^6) \cdot r^4 \cdot (T + 2) \cdot \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right)^{(1 - \varepsilon_m) (T + 3)} = O(r^{14}) \cdot \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right)^{(1 - \varepsilon_m) T}.
\]
As before, this gives an upper estimate for the dimension which is
\[
\frac{(1 - \varepsilon_m) \log \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right) + O(\log r/T)}{r} = \frac{(1 - \varepsilon_m) \log \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right) + O(\log r/r)}{r}
\]
Since \(\varepsilon_m\) is the solution in (0, 1) of the equation \(\log \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right) = \frac{1}{\varepsilon_m}\), with \(m = (r - 4)/\log \left( \frac{3 + \sqrt{5}}{2} \right)\), \((1 - \varepsilon_m) \log \left( \frac{e m \varepsilon_m}{1 - \varepsilon_m} \right) = \frac{1 - \varepsilon_m}{\varepsilon_m}\). Writing \(z = \frac{1 - \varepsilon_m}{\varepsilon_m}\), the equality
log \left( \frac{em}{1-em} \right) = \frac{1}{e_m} \text{ can be written as } \log \left( \frac{em}{z} \right) = z+1, \text{ so } z+\log z = \log m \text{ and } z e^z = m, \text{ so } z = W(m) = W \left( \frac{(r-4) / \log \left( \frac{3 + \sqrt{5}}{2} \right)}{r} \right), \text{ where } W \text{ is Lambert’s function. Since } W'(x) < \frac{1}{x},

W \left( \frac{(r-4) / \log \left( \frac{3 + \sqrt{5}}{2} \right)}{r} \right) = W \left( \frac{r / \log \left( \frac{3 + \sqrt{5}}{2} \right)}{r} \right) + O \left( \frac{\log r}{r^2} \right),

and our upper estimate for the dimension is

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

\square

5. The lower bound

The statements and definitions below are taken from the third author’s work [Mor18].

**Definition 5.1.** Given \( B = \{ \beta_1, \ldots, \beta_l \}, \ l \geq 2, \) a finite alphabet of finite words \( \beta_j \in (\mathbb{N}^+)\), which is primitive (in the sense that \( \beta_i \) doesn’t begin by \( \beta_j \) for all \( i \neq j \)) then the Gauss-Cantor set \( K(B) \subseteq [0, 1] \) associated with \( B \) is defined as

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

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

For each word \( \beta_j \in (\mathbb{N}^+)\), let \( I_j = I(\beta_j) \) be the convex hull of the set \( \{ [0; \beta_j, \gamma_2, \ldots] \mid \gamma_i \in B \} \) and \( \psi|_{I_j} := G|_{I_j} \) where \( G(x) = \{1/x\} = 1/x - \lfloor 1/x \rfloor \) is the Gauss map. This defines an expanding map \( \psi : I(\beta_1) \cup \cdots \cup I(\beta_n) \rightarrow I \). Let \( I = \{ \min K(B), \max K(B) \} \). Then \( I \) is the convex hull of \( I_1 \cup \cdots \cup I_l \) and \( \psi(I_j) = I \) for every \( j \leq l \).

Let us describe how to estimate \( \dim_H(K(B)) \).

According to Palis-Takens [PT93, Chapter 4], let

\[
\lambda_j = \inf |\psi'|_{I_j}, \quad \Lambda_j = \sup |\psi'|_{I_j}
\]

and \( \alpha, \beta \geq 0 \) such that

\[
\sum_{i=1}^{n} \lambda_j^{-\alpha} = 1, \quad \sum_{i=1}^{n} \Lambda_j^{-\beta} = 1.
\]

Therefore

\[
\beta \leq \dim_H(K(B)) \leq \alpha.
\]

Let us discuss how to find estimates for \( \alpha \) and \( \beta \).
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 \( \frac{p_{r_j}^{(j)}}{q_{r_j}^{(j)}} = [0; b_1^{(j)}, \ldots, b_k^{(j)}] \) and \( \beta_j = (b_1^{(j)}, \ldots, b_r^{(j)}) \).

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

**Lemma 5.2.** Let \( x = [a_0, a_1, a_2, \ldots] \) and \( \bar{a} = [a_0, a_1, \ldots, a_n] \). Then
\[
\frac{1}{2q_n q_{n+1}} < \frac{1}{q_n(q_n + q_{n+1})} < |x - \frac{p_n}{q_n}| < \frac{1}{q_n q_{n+1}},
\]
and therefore
\[
\frac{1}{2q_{n+1}} < |q_n x - p_n| < \frac{1}{q_{n+1}}.
\]

Therefore, 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
\[
(q_{r_j}^{(j)})^2 \leq \Lambda_j = \inf |\psi'|_{I_j} | \leq \Lambda_j = \sup |\psi'|_{I_j} | \leq (2q_{r_j}^{(j)})^2.
\]

Let \( a = 22 \), \( s \) minimum such that \( r(1^t) \geq r \), \( k = 2r \), \( \beta_1 = 1^k \) and, for \( 2 \leq j \leq k + 1 \), \( \beta_j = 1^{k+1-j}221^t \), then \( B = \{\beta_1, \beta_2, \ldots, \beta_{k+1}\} \) is primitive.

The alphabet \( B = \{\beta_1, \beta_2, \ldots, \beta_{k+1}\} \) as above induces a subshift
\[
\Sigma(B) = \{ (\gamma_i)_{i \in \mathbb{Z}} \mid \gamma_i \in B \}.
\]

Lemma 3.1 implies that, for any \( \theta \in \Sigma(B) \) and every \( n \in \mathbb{Z} \),
\[
A(\sigma^n(\theta)) < 3 + e^{-r}.
\]

Recall that if \( \alpha = a_1 a_2 \cdots a_m \) and \( \beta = b_1 b_2 \cdots b_n \) are finite words, then
\[
q_m(\alpha)q_n(\beta) < q_{m+n}(\alpha \beta) < 2q_m(\alpha)q_n(\beta).
\]

The above estimates give \( \Lambda_1 = \sup |\psi'|_{I(\beta_1)} | \leq 4 \left( \frac{1 + \sqrt{5}}{2} \right)^{2k} \) and, for \( 2 \leq j \leq k + 1 \), \( \Lambda_j = \sup |\psi'|_{I(\beta_j)} | \leq 8 \left( \frac{1 + \sqrt{5}}{2} \right)^{2k-2(j-2)} \cdot (10^2 \cdot e^{r+1}) \leq \left( \frac{1 + \sqrt{5}}{2} \right)^{2k-2(j-2)} \cdot e^{r+8} \).

Thus, from the above lemma and the third author’s work [Mor18], we conclude that
\[
d(3 + e^{-r}) \geq \dim_{11}(m(\Sigma(B))) = \min\{1, 2 \cdot \dim_{11}(K(B))\} \geq 2d,
\]
where $\tilde{d}$ is the solution of

$$
\left( 4 \left( \frac{1 + \sqrt{3}}{2} \right)^{4r} \right)^{-\tilde{d}} + \sum_{i=0}^{k-1} \left( \frac{1 + \sqrt{3}}{2} \right)^{2i} \cdot e^{-(r+8)d} = 1.
$$

Since $d(3 + e^{-\tilde{r}}) = O\left(\frac{\log \tilde{r}}{\tilde{r}}\right)$, we also have $\tilde{d} = O\left(\frac{\log \tilde{r}}{\tilde{r}}\right) = o(1)$.

Since $(1 + \sqrt{3})^4 > e^{3/2}$, $(4 \left( \frac{1 + \sqrt{3}}{2} \right)^{2k} \right)^{-\tilde{d}} \leq \left( \frac{1 + \sqrt{3}}{2} \right)^{-4\tilde{d}} \leq e^{-\frac{3}{2} \tilde{d}}$, and we have

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

(5.2)

In particular,

$$
1 \geq e^{-(r+8)d} \cdot \frac{1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-4\tilde{d}}}{1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}}} = e^{-(r+8)d} \cdot \left( 1 + \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} \right)
$$

$$
\geq 2e^{-(r+8)d} \cdot \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} \geq 2e^{-(r+9)d},
$$

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

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

and thus

$$
1 \geq e^{-(r+8)d} \cdot \frac{1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}}}{1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}}} \geq e^{-(r+8)d} \cdot \frac{e^{-(r+8)d}}{2 \left( 1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} \right)}.
$$

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

$$
\left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} = e^{-c_1 \tilde{d}} = 1 - c_1 \tilde{d} + O(\tilde{d}^2),
$$

and therefore $1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} = c_1 \tilde{d} + O(\tilde{d}^2) = (1 + O(\tilde{d}))c_1 \tilde{d}$. It follows that

$$
1 \geq \frac{e^{-(r+8)d}}{2 \left( 1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} \right)} = \frac{e^{-(r+8)d}}{2 \left( 1 - \left( \frac{1 + \sqrt{3}}{2} \right)^{-2\tilde{d}} \right)} \geq \frac{e^{-(r+8)d}}{2\tilde{d}}.
$$
and thus $0 \geq -(r + 8)d - \log 2 - \log\tilde{d}$. It follows that $-r\tilde{d} \leq \log \tilde{d} + O(1)$, and thus

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

From (5.2), we get

$$1 - O(d^{3/2}) = e^{-(r+8)d} \cdot \frac{1 - \left(\frac{1 + \sqrt{2}}{2}\right)^{-2kd}}{1 - \left(\frac{1 + \sqrt{2}}{2}\right)^{-2d}} = e^{-(r+8)d} \cdot \frac{1 - O(d^{3/2})}{(1 + O(d))c_1 d} = (1 + O(d))e^{-r\tilde{d}}/c_1 d ,$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $W: [e^{-1}, +\infty) \to [-1, +\infty)$, which is the inverse function of $f: [-1, +\infty) \to [e^{-1}, +\infty)$, $f(x) = xe^x$ (which is increasing in the domain $[-1, +\infty)$): let $g: (0, +\infty) \to \mathbb{R}$ given by $g(x) = x r + \log x$. We have $g(\tilde{d}) = r \tilde{d} + \log \tilde{d} = c_0 + O(\tilde{d})$. 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

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

and $\tilde{d} = d_0 + O(\log r/r^2) = (1 + O(1/r))d_0$. 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

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

(for a detailed discussion on the function $W$, including its asymptotic expansion, we refer the reader to the work of Corless et al. [Cor+96]).

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 $s_0$, and $s_1, s_2$ are at least $s_0 - 4$. 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

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

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

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

Indeed, we have

$$[0; 1^{2k+1}221^{s_1} \ldots ] = [0; 1^{2k+1}221] + O(\varphi^{-8k})$$

and

$$[0; 1^{2k+2+j}221^{s_2} \ldots ] = [0; 1^{2k+2+j}221] + O(\varphi^{-8k}).$$
Moreover, we have

\[ [0; 1^n 221] = \left[ 0; 1^n, 2 + \frac{1}{2 + \varphi^{-1}} \right] \]

\[ = [0; 1^n, 4 - \varphi] = \frac{(4 - \varphi)F_n + F_{n-1}}{(4 - \varphi)F_{n+1} + F_n} \]

\[ = \frac{F_{n-1}/F_n + (4 - \varphi)}{(4 - \varphi)F_{n-1}/F_n + 5 - \varphi}. \]

On the other hand, the identity \( \frac{a_n+b}{c_n+d} - \frac{a_n+b}{c_n+d} = \frac{(a-b)(u-v)}{(cu+d)(cv+d)} \) applied for \( a = 1, b = 4 - \varphi, c = 4 - \varphi, d = 5 - \varphi, u = F_{2k}/F_{2k+1} \) and \( v = F_{2k+1+j}/F_{2k+2+j} \) together with \((cu+d)(cv+d) = (1 + o(1))(c\varphi^{-1} + d)^2 = (1 + o(1))(3\varphi)^2 \) gives

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

In order to estimate \( v - u \), let us estimate \( F_n/F_{n+1} - \varphi^{-1} \): we have

\[ \frac{F_n}{F_{n+1}} - \frac{1}{\varphi} = \frac{\varphi^n - (-\varphi^{-1})^n - 1}{\varphi^n + (-\varphi^{-1})^n + 1} \]

\[ = (1 + o(1))\frac{(-1)^{n+1}(\varphi + \varphi^{-1})\varphi^{-n}}{\varphi^{n+2}} \]

\[ = \frac{(-1)^{n+1}(3\varphi - 4 + o(1))}{\varphi^{2n}}. \]

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

Case 2: Words of the form \( 1^n 2212^{2k} 221^{2k+3+j} 221^{s_2} \), where \( 2k \) is of the order of \( \tilde{s}_0 \), and \( s_1, s_2 \) are at least \( \tilde{s}_0 - 4 \). In this case the Markov value associated with the cut \( 1^n 2212^{2k} 221^{2k+3+j} 221^{s_2} \) is \( 3 + y \), where

\[ y = \left[ 0; 1^{2k+3+j} 221^{s_2} \ldots \right] - [0; 1^{2k+2} 221^{s_1} \ldots] \]

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

The proof of this estimate is analogous to the previous one, applying the above estimate of \( F_n/F_{n+1} - \varphi^{-1} \) for \( n = 2k + 1, n = 2k + 2 + j \) and subtracting.

Hence, \( y \) belongs to an interval of the type

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

Since

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

and

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

we have
it follows that, for large $k$, none of these Markov values belong to the interval 

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

whose size is comparable to the value of its endpoints, and so there are no sequences of the type $1^n 221^{\tau_1} \ldots 221^{\tau_2} \ldots$ 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 $T = r \log r$, where $r = |\log y_k|$. For each $\hat{T}$ with $T/2 < \hat{T} \leq T$, let $M(\hat{T})$ be the number of elements of the set $B(\hat{T})$ of the sequences $1^n 221^{\tau_1} \ldots 221^{\tau_2} \ldots$ with $r \cdot (\hat{T} - 1) < r(1^n 221^{\tau_1} \ldots 221^{\tau_2} \ldots) \leq r \cdot \hat{T}$, $s_j > 3s/4$ for every $j \leq t$, $s_1, s_t \geq s$ and such that, for each $j \leq t$ with $s_j < s$, $s_j$ is even and $s_j - 1 - s_j, s_{j+1} - s_j \in \{-2, 0, 2\}$. Let $\hat{d} = \max \left\{ \frac{\log M(\hat{T})}{r \hat{T}} \right\}$. Then 

$$d(3 + x_k) \geq 2\hat{d}. \text{ Indeed, } m(\Sigma(B(\hat{T}))) \subset M \cap (-\infty, 3 + x_k).$$

Let us now give upper estimates: suppose that $w\bar{w}$ doesn’t have a factor $1^n$ satisfying $r(1^n) \geq r - 80$, where $r = |\log y_k|$ and consider an infinite continuation $\theta$ of it contained in $\Sigma(3 + e^{r-\tau}) = \Sigma(y_k)$. Then the previous discussion provides a continuation $\bar{w} \in \Sigma^{(Tr)}(3 + e^{r-\tau})$ for some $T \in \{10, |\log^3 r|\}$ depending on $\bar{w}$ such that $w\bar{w} \bar{w}$ is the continuation of $w\bar{w}$ in $\theta$, $w\bar{w} \bar{w} \in \Sigma(3 + e^{r-\tau}, |w\bar{w} \bar{w}|)$, and the number $K$ of these words $w\bar{w} \bar{w}$ satisfies $K \cdot e^{-Tr} < 1/r$ for $d = \frac{\log r - \log \log r}{r}$.

Suppose now that $w\bar{w}$ has a factor $1^n$ satisfying $r(1^n) \geq r - 80$, where 

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

Let us consider continuations $\bar{w} \in \Sigma^{0}(3 + e^{r-\tau})$ for some $r^{3/2} \leq m \leq r|\log r|$ such that $w\bar{w} \bar{w} \bar{w} \in \Sigma(3 + e^{r-\tau}, |w\bar{w} \bar{w} \bar{w}|)$ and the continuation of $1^n$ in $w\bar{w} \bar{w} \bar{w}$ is $1^n 221^{\tau_1} \ldots 221^{\tau_2} \ldots 1^{i_t} 22$ such that there are at least $r^{7/8}$ values of $i \leq t$ with $s_i < s$, and such that $t$ is minimum with this property. Then $s_t > r - 3r^{7/8}$. We will introduce a parameter $j$ equal to the number of values of $i \leq t$ with $s_i < s$; consider these $j$ values $i_1 < i_2 < \ldots < i_j$ of $i$. We have $j \geq r^{7/8}$. There are at most $\binom{j}{3}$ choices for the set $\{i_t, 1 \leq t \leq j\}$. Since for $i \in \{i_v, 1 \leq v \leq j\}$ we have at most $3$ choices for $s_{i+v}$, and the total number of these choices is at most $3^j$. Together with the number of choices for the set $\{i_t, 1 \leq t \leq j\}$, this gives an estimate of $\frac{3^j j!}{j^j}$ for these choices of the set $\{(i_t, s_i), 1 \leq t \leq j\}$. Let $T = t - j$.

The number of choices of the remaining values of $s_i$ is at most the number of solutions of $s_1 + s_2 + \ldots + s_T \leq m \log \left( \frac{3 + \sqrt{5}}{2} \right) - sT - j(r - 3r^{7/8}) \leq (U - \tilde{U})s$, where 

$$U = m/(r - 2) - j/2, \text{ which is at most } U \left( \frac{e^{sU/(1 - sU)}(1 - \epsilon_3)(U+1)}{1 - \epsilon_3} \right) = U e^{(1 - \epsilon_3)(U+1)/s_1},$$

where $\epsilon_3$ is the solution in $(0, 1)$ of the equation $\log \left( \frac{e^x}{1 - \epsilon} \right) = \frac{1}{\epsilon}$. As before, 

$$(1 - \epsilon_3)/\epsilon_3 = W(s) \text{ and } e^{W(s)} = (1 + o(1))s/\log s, \text{ and the total number } \tilde{K} \text{ of
these sequences is

\[ O\left(r^4 \left(\frac{6em}{y^r}\right) \frac{(m/r)((1 + o(1))s/\log s)^{-j/2}e^{W(s)m/(r-2)}}{r^{5/2}}\right) = O(s^{-j/4}e^{W(s)m/r}), \]

and, since \( j \geq r^{7/8} \), for \( d = \frac{W(r/\log r)}{r^{1/2}} - \frac{1}{r^{5/2}} \) we have \( \tilde{K} : e^{-nd} < e^{-\sqrt{y}}. \)

Consider now the remaining case where there are less than \( r^{7/8} \) values of \( i \leq t \) with \( s_i < s \) and consider the largest continuation of \( 1^{s_i} \) in \( \vec{w} \in \Sigma^{(T)}(3 + e^{-r}) \),

\( T = [r \log r] \) of the form \( 1^{s_i}221^{s_2}22 \cdots 1^s \), \( s_j > s - 3r^{7/8} \) for each \( j \). Taking \( j_1 \) minimum and \( j_2 \) maximum with \( s_{j_1}, s_{j_2} \geq s \) (notice that \( j_1 + t - j_2 \leq r^{7/8} \)), the number \( \tilde{N} \) of such words is at most \( 3^{j_1 + j_2}M < 3^{r^{7/8}}M \), where \( M \) is the number of elements of \( B(\tilde{T}) \), where \( r \cdot (\tilde{T} - 1) < r(1^{s_1}221^{s_2}22 \cdots 1^{s_2}) \leq r \cdot \tilde{T} \).

We have \( \tilde{T} < T - (j_1 + t - j_2)/2 \) and \( M(\tilde{T}) \leq e^{r\tilde{T}} < e^{r(T - (j_1 + t - j_2)/2)} \), so \( \tilde{N} \leq e^{r\tilde{T}}(3e^{-rd/2})^{j_1 + j_2} \). Since, by our lower estimates on \( d(3 + \varepsilon) \), \( \frac{\log \tilde{N}}{\log r} > \frac{\log r - \log \log r}{\log r} \), it follows that \( \tilde{d} \geq \frac{\log M}{rT} > \frac{\log r - \log \log r}{\log r} \), and thus

\[ (3e^{-rd/2})^{j_1 + j_2} \leq (3(\log r/r)^{1/2})^{j_1 + j_2} \leq 1 \]

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

Finally suppose that \( F \) is a twice continuously-differentiable function such that

\[ d(3 + t) = F(t) + o\left(\frac{\log |\log t|}{|\log t|^2}\right). \]

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

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

Let \( c_1 > 1 \) be a constant we will chose later. By Theorem 1.2 we have

\[ F(c_1y_k) - F(y_k) = g_1(c_1y_k) - g_1(y_k) + O\left(\frac{\log |\log y_k|}{|\log y_k|^2}\right) \]

\[ = (2\log(c_1) + o(1))\frac{\log |\log y_k|}{|\log y_k|^2} + O\left(\frac{\log |\log y_k|}{|\log y_k|^2}\right). \]

By choosing \( c_1 > 1 \) large enough and using the mean value theorem, we obtain a \( \xi_k \in (y_k, c_1y_k) \) such that

\[ F'(\xi_k) = C \cdot \frac{\log |\log y_k|}{y_k |\log y_k|^2} \]

Hence for each \( k \), we can find a point in \( (\xi_k, \xi_k) \) where the second derivative of \( F \) is positive and also a point in \( (\xi_c, \xi_k) \) (for \( c \) large enough) where the second derivative of \( F \) is negative.
7. Basic facts and estimates on continued fractions

Lemma 7.1. Let \( w \) be a nonempty finite word in \( 1 \) and \( 2 \) of length \( n \in \mathbb{N}^* \). We have that

\[
(n - 3) \log \left( \frac{3 + \sqrt{5}}{2} \right) \leq r(w) \leq (n + 1) \log (3 + 2\sqrt{2}).
\]

Proof. Given \( \alpha = (a_1, \ldots, a_n) \in (\mathbb{N}^*)^n \), we have that

\[
s(\alpha) = \frac{1}{q_n(q_n + q_{n-1})},
\]
so \( s(\alpha) \) 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 \( q_n = P_n \) (where \( P_n \) is the \( n \)-th Pell number) and where \( q_n = F_n \) (where \( F_n \) is the \( n \)-th Fibonacci number). Hence,

\[
r(1^n) \leq r(w) \leq r(2^n).
\]

Moreover, we have that

\[
s(1^n)^{-1} = F_{n+1}(F_{n+1} + F_n)
\]

and, on the other hand, we have that

\[
s(2^n)^{-1} = P_n(P_n + P_{n-1}) = \frac{(3 + 2\sqrt{2})^n - (3 - 2\sqrt{2})^n}{4\sqrt{2}}
\]

Thus, we obtain that

\[
(n - 1) \log \left( \frac{3 + \sqrt{5}}{2} \right) \leq \log s(w)^{-1} \leq (n + 1) \log (3 + 2\sqrt{2}).
\]

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

\[
(n - 3) \log \left( \frac{3 + \sqrt{5}}{2} \right) \leq r(w) = \lfloor \log s(w)^{-1} \rfloor \leq (n + 1) \log (3 + 2\sqrt{2}).
\]

\( \square \)
Lemma 7.2. Let \( w \) be a finite word and let \( v \) be a factor of \( w \). Then, \( s(w) \leq s(v) \) and \( r(w) \geq r(v) \).

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

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

A property that is useful to simplify some computations is

\[
 r(w_1 k_1 k_2 w_2) \geq r(w_1) + r(w_2)
\]

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

\[
 s(w_1 k_1 k_2 w_2) \leq 4 s(k_1 k_2) s(w_1) s(w_2) \leq s(w_1) s(w_2)/3.
\]

For \( (k_1, k_2) = (1, 1) \) we have that \( r(w_1 b w_2) \geq r(w_1) + r(w_2) - 1 \), since \( r(b) = 1 \).

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

Let \( s_1, \ldots, s_\ell \) be positive integers with \( \ell \geq 2 \). We will show that

\[
 r(1^{s_1} 221^{s_2} 2 \cdots 221^{s_\ell}) \geq (s_1 + \cdots + s_\ell + 3(\ell - 2)) \log \left( \frac{3 + \sqrt{5}}{2} \right).
\]

and

\[
 r(2^{s_1} 112^{s_2} 11 \cdots 112^{s_\ell}) \geq (s_1 + \cdots + s_\ell + \ell - 2) \log(3 + 2\sqrt{2}).
\]

First, we will show inductively that

\[
 q(1^{s_1} 221^{s_2} 2 \cdots 221^{s_\ell}) \geq F_{s_1 + \cdots + s_\ell + 3(\ell - 1) + 1},
\]

and

\[
 q(2^{s_1} 112^{s_2} 11 \cdots 112^{s_\ell}) \geq P_{s_1 + \cdots + s_\ell + \ell}.
\]

Using Euler’ property of continuants

\[
 q(1^{s_1} 221^{s_2}) = q(1^{s_1})q(221^{s_2}) + q(1^{s_1-1})q(21^{s_2}).
\]

Since \( q(1^1) = F_{s+1} \) one has

\[
 q(221^\ell) = 5q(1^\ell) + 2q(1^{\ell-1}) = 5F_{s+1} + 2F_s = 3F_{s+1} + 2F_{s+2},
\]

\[
 q(21^\ell) = 2q(1^\ell) + q(1^{\ell-1}) = 2F_{s+1} + F_s.
\]

From the identity

\[
 F_n F_m + F_{n-1} F_{m-1} = F_{n+m-1},
\]

we get

\[
 F_{n+1} q(221^m) + F_n q(21^m) = F_{n+1} (2F_{m+2} + 3F_{m+1}) + F_n (2F_{m+1} + F_m)
\]

\[
 = 2F_{n+m+2} + F_{n+m+1} + 2F_{n+1} F_{m+1}
\]

\[
 = F_{n+m+4} + 2F_{n+1} F_{m+1}.
\]
Thus
\[
q(1^{s_1}221^{s_2}) = q(1^{s_1})q(221^{s_2}) + q(1^{s_1-1})q(21^{s_2})
\]
\[
= F_{s_1+1}q(221^{s_2}) + F_{s_1}q(21^{s_2})
\]
\[
= F_{s_1+s_2+4} + 2F_{s_1+1}F_{s_2+1}.
\]
Hence (7.4) is true for \( \ell = 2 \). Assuming it for \( \ell \), we use (7.6) with \( n = s_1 + \cdots + s_\ell + 3(\ell - 1) + 1 \) and \( m = s_{\ell+1} \) to obtain
\[
q(1^{s_1}221^{s_2}22 \cdots 221^{s_{\ell+1}}) = q(1^{s_1}221^{s_2}22 \cdots 221^{s_{\ell-1}})q(221^{s_{\ell+1}}) +
\]
\[
q(1^{s_1}221^{s_2}22 \cdots 221^{s_{\ell+1}})q(21^{s_{\ell+1}}) \geq F_{n}q(221^{s_{\ell+1}}) + F_{n-1}q(21^{s_{\ell+1}})
\]
\[
\geq F_{s_1+\cdots+s_{\ell+1}+3\ell+1}
\]
Finally, using (7.1)
\[
s(1^{s_1}221^{s_2}22 \cdots 221^{s_{\ell+1}})^{-1} \geq F_{s_1+\cdots+s_{\ell}+3(\ell-1)+1}
\]
\[
\left( F_{s_1+\cdots+s_{\ell}+3(\ell-1)+1} + F_{s_1+\cdots+s_{\ell}+3(\ell-1)+1} \right)
\]
\[
\geq \left( \frac{3 + \sqrt{5}}{2} \right)^{s_1+\cdots+s_{\ell}+3(\ell-1)+1}
\]
On the other hand, using that \( F_{n+2} \leq 3F_n \) we get
\[
s(1^n)^{-1} = F_{n+1}F_{n+2} \leq \frac{3}{4}F_{2n+2} \leq \left( \frac{3 + \sqrt{5}}{2} \right)^n
\]
so
\[
(7.7) \quad r(1^n) \leq n \log((3 + \sqrt{5})/2).
\]
Similarly, one has that \( q(2^n) = P_{n+1} \) and \( q(112^n) = P_{n+2} \). The Pell numbers also satisfy the identity
\[
P_nP_m + P_{n-1}P_{m-1} = P_{n+m-1}.
\]
Hence
\[
P_{n+1}q(112^n) + P_nq(12^n) = P_{n+1}P_{m+2} + P_n(P_m + P_{m+1})
\]
\[
= P_{n+m+2} + P_nP_m
\]
Therefore by induction
\[
q(2^{s_1}112^{s_2}11 \cdots 112^{s_{\ell+1}}) \geq P_{s_1+\cdots+s_{\ell}+s_{\ell+1}}q(112^{s_{\ell+1}}) + P_{s_1+\cdots+s_{\ell}+s_{\ell+1}+1}q(12^{s_{\ell+1}})
\]
\[
\geq P_{s_1+\cdots+s_{\ell}+s_{\ell+1}+1}q(12^{s_{\ell+1}})
\]
Finally to show (7.3) we use that
\[
s(2^{s_1}112^{s_2}11 \cdots 112^{s_{\ell}})^{-1} \geq P_n^2 \geq 4(3 + 2\sqrt{2})^n - 2
\]
where \( n = s_1 + \cdots + s_{\ell} + \ell \).
Lemma 7.3. If \( \alpha = (a_1, \ldots, a_n) \in (\mathbb{N}^*)^n \) then
\[
\frac{[1; a_n + 1]}{[1; a_1]} \leq \frac{s(\alpha^*)}{s(\alpha)} \leq \frac{[1; a_n]}{[1; a_1 + 1]}
\]
and, hence,
\[
-\log \left(1 + \frac{1}{a_n + 1}\right) - 1 \leq r(\alpha) - r(\alpha^*) \leq \log \left(1 + \frac{1}{a_n}\right) + 1.
\]

Proof. By Euler’s property of continuants we have
\[
q_n(a_1 \cdots a_n) = q_n(a_n \cdots a_1)
\]
thus
\[
s(\alpha^*)^{-1} = q_n(a_n \cdots a_1)(q_n(a_n \cdots a_1) + q_{n-1}(a_n \cdots a_2))
\geq \left(1 + \frac{1}{a_1 + 1}\right) q_n(a_n \cdots a_1)^2
= \left(1 + \frac{1}{a_1 + 1}\right) q_n(a_1 \cdots a_n)^2,
\]
while
\[
s(\alpha)^{-1} = q_n(a_n \cdots a_n)(q_n(a_1 \cdots a_n) + q_{n-1}(a_1 \cdots a_n-1))
\leq \left(1 + \frac{1}{a_n}\right) q_n(a_1 \cdots a_n)^2.
\]
Hence
\[
\frac{s(\alpha^*)}{s(\alpha)} \leq \frac{[1; a_n]}{[1; a_1 + 1]}
\]
By symmetry we obtain the lower bound. \( \square \)

References

[Bom07] E. Bombieri. “Continued fractions and the Markoff tree”. In: Expo. Math. 25.3 (2007), pp. 187–213. doi: [10.1016/j.exmath.2006.10.002]

[CF89] T. W. Cusick and M. E. Flahive. “The Markoff and Lagrange spectra”. Vol. 30. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 1989, pp. x+97. isbn: 0-8218-1531-8. doi: [10.1090/surv/030].

[Cor+96] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. “On the Lambert W function”. In: Adv. Comput. Math. 5.4 (1996), pp. 329–359. doi: [10.1007/BF02124750].

[Fre68] G. A. Freiman. “Non-coincidence of the spectra of Markov and of Lagrange”. In: Mat. Zametki 3 (1968), pp. 195–200.

[Fre75] G. A. Freiman. “Diophantine approximations and the geometry of numbers (Markov’s problem) [In Russian]”. Kalinin. Gosudarstv. Univ., Kalinin, 1975, p. 144.

[Hal47] M. Hall Jr. “On the sum and product of continued fractions”. In: Ann. of Math. (2) 48 (1947), pp. 966–993. doi: [10.2307/1969389].
REFERENCES

[Lim+21] D. Lima, C. Matheus, C. G. Moreira, and S. Romaña. “Classical and dynamical Markov and Lagrange spectra—dynamical, fractal and arithmetic aspects”. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, [2021] ©2021, pp. xiii+213. isbn: [9789811225284]; [9789811225291]; [9789811225307].

[Mar80] A. Markoff. “Sur les formes quadratiques binaires indéfinies”. In: Math. Ann. 17.3 (1880), pp. 379–399. doi:[10.1007/BF01446234].

[MM20] C. Matheus and C. G. Moreira. “Fractal geometry of the complement of Lagrange spectrum in Markov spectrum”. In: Comment. Math. Helv. 95.3 (2020), pp. 593–633. doi:[10.4171/CMH/498].

[Mor18] C. G. Moreira. “Geometric properties of the Markov and Lagrange spectra”. In: Ann. of Math. (2) 188.1 (2018), pp. 145–170. doi: [10.4007/annals.2018.188.1.3].

[PT93] J. Palis and F. Takens. “Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations”. Vol. 35. Cambridge Studies in Advanced Mathematics. Fractal dimensions and infinitely many attractors. Cambridge University Press, Cambridge, 1993, pp. x+234. isbn: 0-521-39064-8.

(Harold Erazo) IMPA, Estrada Dona Castorina 110, 22460-320, Rio de Janeiro, Brazil.
Email address: harolderaz@gmail.com

(Rodolfo Gutiérrez-Romo) Centro de Modelamiento Matemático, CNRS-IRL 2807, Universidad de Chile, Beauchef 851, Santiago, Chile.
Email address: g-r@rodol.fo
URL: http://rodol.fo

(Carlos Gustavo Moreira) IMPA, Estrada Dona Castorina 110, 22460-320, Rio de Janeiro, Brazil.
Email address: gugu@impa.br

(Sergio Romaña) Universidade Federal do Rio de Janeiro, Av. Athos da Silveira Ramos 149, Centro de Tecnologia - Bloco C - Cidade Universitária, Ilha do Fundão, CEP 21941-909, Rio de Janeiro, Brasil.
Email address: sergiori@im.ufrj.br