On the density of Lyndon roots in factors

Maxime Crochemore[*] Robert Mercaş[†]

December 24, 2015

1 Introduction

The concept of a run coined by Iliopoulos et al. [11] when analysing repetitions in Fibonacci words, has been introduced to represent in a succinct manner all occurrences of repetitions in a word. It is known that there are only $\mathcal{O}(n)$ many of them in a word of length n from Kolpakov and Kucherov [12] who proved it in a non-constructive manner. The first explicit bound was later on provided by Rytter [15]. Several improvements on the upper bound can be found in [16, 4, 14, 5, 8]. Kolpakov and Kucherov conjectured that this number is in fact smaller than n, which has been proved by Bannai et al. [1, 2]. Recently, Holub [10] and Fischer et al. [9] gave a tighter upper bound reaching $22n/23$.

In this note we provide a proof of the result, slightly different than the short and elegant proof in [2]. Then we provide a relation between the border-free root conjugates of a square and the critical positions [13, Chapter 8] occurring in it. Finally, counting runs extends naturally to the question of their highest density, that is, to the question of the type of factors in which there is a large accumulation of runs. This is treated in the last section.

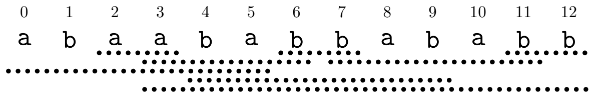

Figure 1: Dotted lines show the 8 runs in abaababbababb. For example, $[7\,..\,11]$ is the run of period 2 and length 5 associated with factor babab.

Formally, a *run* in a word w is an interval $[i\,..\,j]$ of positions, $0 \leq i < j < |w|$, for which both the associated factor $w[i\,..\,j]$ is periodic (i.e. its smallest period p satisfies $p \leq (j-i+1)/2$), and the periodicity cannot be extended to the right nor to the left: $w[i-1\,..\,j]$ and $w[i\,..\,j+1]$ have larger periods when these words are defined (see Figure 1).

[*] King's College London and Université Paris-Est. Maxime.Crochemore@kcl.ac.uk

[†] Kiel University and King's College London. robertmercas@gmail.com

2 Fewer runs than length

We consider an ordering $<$ on the word alphabet and the corresponding lexicographic ordering denoted $<$ as well. We also consider the lexicographic ordering $\tilde{<}$, called the reverse ordering, inferred by the inverse alphabet ordering $<^{-1}$. The main element in the proof of the theorem is to assign to each run its greatest suffix according to one of the two orderings.

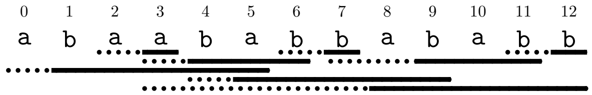

Figure 2: Plain lines show the 8 greatest proper suffixes assigned to runs of `abaababbababb` from Figure 1 in the proof of the theorem. Note that no two suffixes start at the same position.

Theorem 1 *The number of runs in a word of length n is less than n.*

Proof. Let w be a word of length n. Let $[i\mathinner{..}j]$ $(0 \leq i < j < n)$ be a run of smallest period p in w. If $j+1 < n$ and $w[j+1] > w[j-p+1]$ we assign to the run the position k for which $w[k\mathinner{..}j]$ is the greatest proper suffix of $w[i\mathinner{..}j]$. Else, k is the position of the greatest proper suffix of $w[i\mathinner{..}j]$ according to $\tilde{<}$.

Note that if $k > i$ then $k > 0$, and that $w[k\mathinner{..}j]$ contains a full period of the run factor, i.e. $j-k+1 \geq p$. Also note that $w[k\mathinner{..}k+p-1]$ is a greatest conjugate of the period root $w[i\mathinner{..}i+p-1]$ according to one of the two orderings. Therefore, it is border-free, known property of Lyndon words.

We claim that each position $k > 0$ on w is the starting position of at most one greatest proper suffix of a run factor. Let us consider two distinct runs $[i\mathinner{..}j]$ and $[\bar{i}\mathinner{..}\bar{j}]$ of respective periods p and q, and which are called respectively the p-run and the q-run. Assume $p \neq q$ since the runs cannot be distinct and have the same period. For the sake of contradiction, we assume that their greatest suffixes share the same starting position k.

First case, $j = \bar{j}$, which implies $w[k\mathinner{..}j] = w[k\mathinner{..}\bar{j}]$. Assume for example that $p < q$. Then, $w[k\mathinner{..}k+q-1]$ has period p and thus is not border-free, which is a contradiction.

Second case, assume without loss of generality that $j < \bar{j}$ and that both suffixes are the greatest in their runs according to the same ordering, say $<$. Let $d = w[j+1]$, the letter following the p-run. By definition we have $w[j-p+1] < d$ and then $w[i\mathinner{..}j-p+1] < w[i\mathinner{..}j-p]d$. But since $w[i+p\mathinner{..}j]d$ is a factor of the q-run this contradicts the maximality of $w[k\mathinner{..}\bar{j}-1]$.

Third case, $j \neq \bar{j}$ and the suffixes are greatest according to different orderings. Assume without loss of generality that $p < q$ and the suffix of the p-run factor is greatest according to $<$. Since $q > 1$ we have both $w[k+q-1]\tilde{>}w[k]$ and $w[k+q-1] = w[k-1]$, then $w[k-1] < w[k]$. We cannot have $p > 1$ because this implies $w[k-1] > w[k]$. And we cannot have either $p = 1$ because this implies $w[k-1] = w[k]$. Therefore we get again a contradiction.

This ends the proof of the claim and shows that the number of runs is no more than the number $n-1$ of potential values for k, as stated. ■

3 Lyndon roots

The proof of Theorem 1 by Bannai et al. [2] relies on the notion of a Lyndon root. Recall that, for a fixed ordering on the alphabet, a Lyndon word is a primitive word that is not larger than any of its conjugates (rotations). Equivalently, it is smaller than all its proper suffixes. The root of a run $[i \mathinner{.\,.} j]$ of period p in w is the factor $w[i \mathinner{.\,.} i+p-1]$. Henceforth, the Lyndon root of a run is the Lyndon conjugate of its root. Therefore, since a run has length at least twice as long as its root, the first occurrence of its Lyndon root is followed by its first letter. This notion of Lyndon root is the basis of the proof of the $0.5n$ upper bound on the number of cubic runs given in [6]. Recall that a run is said to be cubic if its length is at least three times larger than its period.

Lyndon roots considered in [2] are defined according to the two orderings $<$ and $\tilde{<}$. However, these Lyndon roots can be defined as smallest or greatest conjugates of the run root according to only one ordering.

The proof of Theorem 1 is inspired by the proof in [2] but does not use explicitly the notion of Lyndon roots. The link between the two proofs is as follows: when the suffix $w[k \mathinner{.\,.} j]$ is greatest according to $<$ in the run factor, then its prefix of period length, $w[k \mathinner{.\,.} k+p-1]$, is a Lyndon word according to $\tilde{<}$. As a consequence, the assignment of positions to runs is almost the same whatever greatest suffixes or Lyndon roots are considered.

The use of Lyndon roots leaves more flexibility to assign positions to runs. Indeed, a run factor may contain several occurrences of the run Lyndon root. Furthermore, any two consecutive occurrences of this root do not overlap and are adjacent. The multiplicity of these occurrences can be transposed to greatest suffixes by considering their borders. Doing so, what is essential in the proof of Theorem 1 is that the suffixes and borders so defined are at least as long as the period of the run. Consequently, consecutive such marked positions can be assigned to the same run. As a consequence, since every cubic run is associated to at least two positions, this yields the following corollaries.

Corollary 2 *If a word of length n contains c cubic runs, it contains less than $n-c$ runs.*

Corollary 3 *A word of length n contains less than $0.5n$ cubic runs.*

The last statement is proved in [6] employing the notion of Critical position, which is discussed in the next section.

4 Critical positions

The consideration of the two above orderings appears in the simple proof of the Critical Factorisation Theorem [7] (for another proof see [13, Chapter 8]).

Let us recall that the local period at position $|u|$ in uv is the length of the shortest non-empty word z for which z^2 is a repetition centred at position $|u|$. Equivalently, in simpler words, z is the shortest non-empty word that satisfies both conditions: either z is a suffix of u or u is a suffix of z, and either z is a prefix of v or v is a prefix of z. Note that vu satisfies the conditions but is not necessarily the shortest word to do it. The Critical Factorisation Theorem states that a word x of period p admits a factorisation $x = uv$ whose local period at position $|u|$ is p. Such a factorisation uv of x is called a critical factorisation and the position $|u|$ on x a critical position.

When considering the starting positions of greatest suffixes defined above according to $<$ and to $\tilde{<}$, the shorter of the two is known to provide a critical position following [7]. Thus, it does not come as a surprise to us that the simple proof of Theorem 1 relies on alphabet orderings. Nevertheless, as the initial question does not involve any ordering on the alphabet, we could expect a proof using, for example, only the notion of critical positions. The next lemma may be a step on this way.

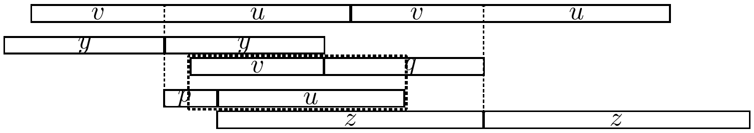

Figure 3: If uv is a border-free factor of $(vu)^2$, then at least one of its local period words y or z have length $|uv|$. Otherwise, the common part in the dash-box has length equal to the sum of its periods p and q generating a contradiction.

Lemma 4 *Let $x^2 = (vu)^2$ be a square whose root conjugate uv is border-free. Then, at least $|v|$ or $|vuv|$ are critical positions on x^2.*

Proof. Let y be the local period word at position $|v|$ on x^2. Since uv is border-free, v is a proper suffix of y. Similarly, for the local period word z at position $|vuv|$, the border-freeness of uv implies that u is a proper prefix of z. The situation is displayed in Figure 3.

For the sake of contradiction we assume the conclusion does not hold, i.e., both y and z are shorter than uv (note that they cannot be longer than uv).

Let $|p|$ be the induced period of pu and $|q|$ the induced period of vq. The overlap between the two words p and q admits period lengths $|p|$ and $|q|$ and has length $|pu| - (|uv| - |vq|) = |p| + |q|$. Thus, by the Periodicity Lemma, p and q are powers of the same word r. But then r is a nonempty prefix of u and a suffix of v contradicting the border-freeness of uv. ■

Example. Consider the square `baba` of period 2. The occurrence of its border-free factor `ab` induces the two critical positions 1 and 3. On the contrary, the first occurrence of its border-free factor `ba` induces only one critical position, namely 2, while the local period at 0 has length $1 < 2$.

In the square abaaba of period 3, the occurrence of the border-free factor aab produces the critical position 2. However, its position 5 is not critical since the local period 2 is smaller than the whole period of the square.

5 Lyndon roots density

In this section we consider a generalisation of the problem of counting the maximal number of runs in a word. In particular, we are interested in the following problem concerning first occurrences of Lyndon roots within a run factor. Let us call the interval corresponding to the first such occurrence the Lroot associated with the run. Then, we are dealing with the following conjecture:

Conjecture 1 ([3]) *For any two positions i and j on a word x, $0 \leq i \leq j < |x|$, the maximal number of run* Lroot*s included in the interval $[i\,..\,j]$ is not more than the interval length $j - i + 1$.*

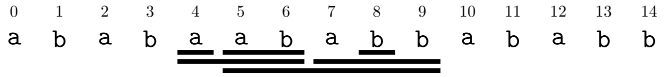

Figure 4: Lines show the 6 run Lroots inside the interval $[4\,..\,9]$ corresponding to the factor aababb.

Let us consider the word $x = (\mathtt{ab})^k\mathtt{a}(\mathtt{ab})^k\mathtt{b}(\mathtt{ab})^k\mathtt{b}$ and the interval of positions $[2k\,..\,4k+1]$ corresponding to the factor $\mathtt{a}(\mathtt{ab})^k\mathtt{b}$. The number of Lroots corresponding to this interval is exactly the length $2(k+1)$ of the interval. Figure 4 shows the situation when $k = 2$. This example gives a lower bound on the maximal number of Lroots contained in an interval of positions.

Proposition 5 *The number of* Lroot*s contained in an interval of positions on a word, can be as large as the length of the interval.*

In addition to the conjecture, we believe that factors associated with intervals of length at least 4 containing the maximal number of Lroots are of the form $a(ab)^+b$ for two different letters a and b. It can be checked that the maximal number of Lroots is respectively 1 and 3 for intervals of lengths 1 and 3 with factors a and aab, but is only 1 for intervals of length 2. All these factors are Lyndon words for the ordering $a < b$. This is due to the fact that such factors contain overlapping Lyndon roots making the whole factor a Lyndon word itself.

Remark 6 *In order to obtain an upper bound on the number of* Lroot*s inside an interval of positions, it is enough to restrict ourselves to counting the maximal number of* Lroot*s within an interval corresponding to some Lyndon word.*

Indeed, each Lroot corresponds to a Lyndon word. Since we want an interval that contains the maximal such number, all the positions of this interval are covered by some Lyndon word. However, since the overlap between every two

Lyndon words produces a Lyndon word, and since every word can be expressed as a concatenation of Lyndon words, our claim follows.

We show that the number of Lroots inside an interval corresponding to a Lyndon word is bounded by 1.5 times the length of the interval. For this we make use of the result from [2] stating that each position of a word is the starting position of at most one specific root associated with the run. The root is chosen according to some order defined by the letter following the run. We denote such a root relative to the order as the Oroot of the run. Formally:

Definition 1 *Let r be a run of period p_r of the word w and let r_L be the* Lroot *associated with r. If r ends at the last position of w, or if the letter at the position following r is smaller than the letter p_r positions before it, then the* Oroot *is the interval corresponding to the first occurrence of a Lyndon root that is not a prefix of r. Otherwise, the* Oroot *is the interval corresponding to the length p_r prefix of the greatest proper suffix of the run factor r.*

Observe that since a run is at least as long as twice its minimal period, this ensures the existence of both its Lroot as well as its Oroot. To see that the Oroot is never a prefix of the run factor it is associated with, observe from its above definition that in the second case this is actually the interval corresponding to the length p_r prefix of the maximal proper suffix of the run, which is different from the run itself (being proper).

Henceforth we fix an interval $[i\,..\,j]$ with its corresponding Lyndon word w of length ℓ. Furthermore, we denote by $r_{\mathrm{L}} = [i_{\mathrm{L}}\,..\,j_{\mathrm{L}}]$ the Lroot of the run $r = [i_r\,..\,j_r]$ and by $r_{\mathrm{O}} = [i_{\mathrm{O}}\,..\,j_{\mathrm{O}}]$ its Oroot. For r, we denote by p_{r} the (smallest) period of the run. Please note that $|r_{\mathrm{L}}| = |r_{\mathrm{O}}| = p_r$, while both must start and end within the run r.

We make the following remarks based on the already known properties of Lroots and Oroots.

Remark 7 *The* Lroot *and the* Oroot *associated with a run r start within the first p_r and $p_r + 1$, respectively, positions of the run, and both have length p_r.*

As a direct consequence of the definition of the Oroot we have the following:

Remark 8 *If the* Oroot *of a run r is a Lyndon word, then the* Oroot *and the corresponding* Lroot *represent the same factor and, either $i_O = i_L$, or the run r starts at position i_L and $i_L + p_r = i_O$.*

In conclusion we have the following:

Remark 9 *To bound the number of* Lroot*s inside the interval $[i\,..\,j]$ corresponding to the word w, it is enough to consider all runs starting within the interval corresponding to a factor $w_{\mathrm{p}}w$, where $|w_{\mathrm{p}}| < |w|$.*

For the rest of this work let us fix the factor preceding a Lyndon word w as w_{p}, while the one following it by w_{s}, such that the interval corresponding to $w_{\mathrm{p}}ww_{\mathrm{s}}$ is the shortest interval that contains all runs with their Lroots in $[i\,..\,j]$.

Now we start looking at the relative positions of the Oroot and Lroot corresponding to the same run.

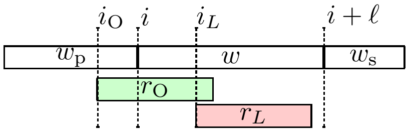

Figure 5: r_{O} starts before the Lyndon word w

Lemma 10 *Fix an* Lroot *occurring within an interval $w = [i \mathinner{.\,.} i+\ell-1]$ of a word. If the* Oroot *corresponding to the same run starts outside w, then either:*

1. *the* Lroot *ends at position $i+\ell-1$, and the* Oroot *starts at position $i+\ell$, or*
2. *the* Lroot *starts at position i, and the* Oroot *starts before position i.*

Proof. Following Remark 6, without loss of generality assume that w_{p} starts at position 0 and has length i, while $w = [i \mathinner{.\,.} i+\ell-1]$ is a Lyndon word. As stated in the hypothesis, $i \leq i_{\mathrm{L}}$ and $|r_{\mathrm{L}}| \leq \ell$.

First let us assume that $i_{\mathrm{O}} \geq i+\ell$, hence the Oroot starts after the end of the Lyndon word w. The result follows immediately from Remark 8.

For the second statement, consider Figure 5 where $i_{\mathrm{O}} < i \leq i_{\mathrm{L}}$. Assume towards a contradiction that $i < i_{\mathrm{L}}$. Since $i_{\mathrm{O}} < i$ and r_{L} corresponds to an interval on w, it must be that the corresponding run starts before or on position i_{O} and it ends after or on position $i+\ell-1$. However, since w is a Lyndon word, it must be that for any word x such that yx is a suffix of w, where y is the factor corresponding to r_{L}, we have $w < yx$. But in this case, unless y is a prefix of w, we get a contradiction with the fact that r_L is a Lyndon root. In the former case, however, we get a contradiction with the definition of r_{L} since $i < i_{\mathrm{L}}$ (r_{L} is the interval corresponding to the first occurrence of a Lyndon root of the run), and the conclusion follows in this case as well. ■

As a consequence of the above lemma, the number of Lroots is bounded by 2ℓ. This is because the Oroots all start inside an interval of length $|w_{\mathrm{p}}|+|w|+1 \leq 2\ell$, and no two share the same starting position [2]. In the following we reduce this bound to 1.5ℓ. The next lemma shows that the situation in Lemma 10.1 is met by at most one Oroot.

Lemma 11 *For any word and any interval $[i \mathinner{.\,.} j]$ on it, there exists at most one run that has its* Oroot *starting after position j while its* Lroot *corresponds to an interval inside $[i \mathinner{.\,.} j]$.*

Proof. According to Lemma 10.1, it must be the case that $i_{\mathrm{O}} = j+1$, while $i_{\mathrm{L}} = j - p_r + 1$, for any appropriate run r. However, having more than one Oroot starting at position $j+1$ with the factor corresponding to its Lroot as a suffix of the Lyndon word, would then imply that the larger of Lroots that corresponds to a Lyndon word is bordered, which is a contradiction. ■

Now we are dealing with Oroots corresponding to Lemma 10.2.

Proposition 12 *For a given word, any interval of length ℓ of positions on the word contains at most $3\ell/2$* Lroot*s.*

Proof. Let us denote once more our interval by $w = [i \mathinner{.\,.} i + \ell - 1]$ and the interval preceding it by w_{p}. We know from the definition that an Lroot is the first Lyndon root of a run, and therefore the letter ending every Lroot must be greater than its first letter, while the one on the position right after the end of the Lroot must be the same as the first letter of the Lroot.

Since for any Oroot starting in the interval associated with w_{p}, the Lroot corresponding to it in w starts at position i, we note that these can be bounded by the number of length two factors in w that have the letter on the first position larger than the letter on position i, while the second one identical. If the second letter of such a factor is smaller than the letter on position i, than this situation would make it impossible for a Lroot to start on position i and end before this position (this is because an Lroot is the first Lyndon root occurrence of a run).

Since this number is obviously bounded by $\frac{\ell}{2}$, while the whole length of w_{p} is bounded by $\ell - 1$, (by considering the symmetric situation) we conclude that there are less than $\frac{\ell-1}{2}$ runs that start in w_{p} such that their corresponding Oroots start before position i, while their Lroots start at position i (the Lroots correspond to prefixes of w). Hence, combining this with the fact that within w we have at most ℓ Oroots starting there, see [2], and since according to Lemma 11 there is at most one Oroot starting after position j that has the Lroot in w, we get an upper bound for our problem. ■

The bound given in the above proposition is not really tight. On this point let us complete the conjecture:

Conjecture 2 *For a given word, any interval of length $\ell > 0$ of positions on the word contains at most ℓ* Lroot*s, and the maximum number is obtained only when the factor corresponding to the interval is of the form $a(ab)^{\frac{\ell-2}{2}}b$, where $\ell > 3$, and the letters a and b satisfy $a < b$.*

We end this article with a few more observations regarding the results from [2], when we restrict ourselves to binary words. First we recall a property of Oroots:

Lemma 13 (Bannai et al. [2]) *If two different* Oroot*s obtained considering the same order overlap, then their overlap is the shortest of the* Oroot*s.*

We observe that we can consider Oroots to be obtained according to a certain order based on the letter that these Oroots start with (thus all Oroots starting with `a` are obtained according to the lexicographical order, while the ones starting with `b` are obtained according to the inverse lexicographical one).

Proposition 14 *For a given binary word, any interval of length ℓ of positions on the word contains at most $\frac{\ell-1}{2}$* Oroot*s obtained according to the same order.*

Proof. Without loss of generality we fix an order; let us say lexicographical. Observe first, that for a word to correspond to an Oroot, whenever they are not binary, they must start with a letter $\mathtt{a}$ and end with a $\mathtt{b}$ (as previously mentioned). Furthermore, it must be the case that this interval is preceded by a $\mathtt{b}$ and followed by an $\mathtt{a}$, as otherwise it does not correspond to a Lyndon root (there exists another rotation that has an extra $\mathtt{a}$ in its longest unary prefix).

Finally, observe that considering their relative position, following Lemma 13, two such Oroots are either included one in the other, or they are disjoint.

Now, considering two words corresponding to two Oroots, let us say u and v with u a factor of v, we note that, since their lengths are different, following the initial conditions, they must differ by a length of at least 2, whenever u is not unary (each starts between a $\mathtt{b}$ and an $\mathtt{a}$, and ends between an $\mathtt{a}$ and a $\mathtt{b}$). For the unary case, note that every block of consecutive $\mathtt{a}$'s must be in-between two occurrences of $\mathtt{b}$. Furthermore, we cannot have two unary words corresponding to Oroots overlapping each other. Thus if the position of the second a is an Oroot in the word $\mathtt{ba}^{\ell}\mathtt{b}$, for $\ell > 0$, it is impossible to have a length less than 3 for any word starting with the first a whose interval corresponds to an Oroot.

Given that for any two distinct adjoining Oroots both their lengths and the number of Oroots they contain add up, the result follows in this case as well.

In order to get the -1, we observe that for any word of length at least 3, for the interval it determines to have the maximum number of Oroots of the same order, according to the previous facts, would imply the word to have the form $(\mathtt{ab})^{+}$. However, now, the Oroots would correspond to words that are just powers of one another, contradicting their property of being Lyndon words. ■

Furthermore, denoting by $|w|_u$ the number of all (possibly overlapping) occurrences of u in w, as consequence of the above we have the following:

Corollary 15 *Every length ℓ interval associated with a factor w of a binary word completely contains at most* $\min\{|w|_{ab}, |w|_{ba}\}$ Oroot*s that correspond to non-unary factors and are obtained according to the same order.*

Corollary 16 *The number of* Oroot*s associated with unary runs within every factor of a binary word is at most one extra than the number of unary maximal blocks within the factor (by a maximal block we refer to a unary factor that cannot be extended either to the left or to the right without losing its periodicity).*

6 Acknowledgement

We would like to warmly thank Gregory Kucherov, Hideo Bannai, and Bill Smyth for helpful discussions on the subject. The work of Robert Mercaş was supported by the P.R.I.M.E. programme of DAAD with funds provided by the Federal Ministry of Education and Research (BMBF) and the European Union's Seventh Framework Programme for research, technological development and demonstration (grant agreement no. 605728).

References

[1] H. Bannai, T. I, S. Inenaga, Y. Nakashima, M. Takeda, and K. Tsuruta. A new characterization of maximal repetitions by Lyndon trees. In *26th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA, pages 562–571, 2015.

[2] H. Bannai, T. I, S. Inenaga, Y. Nakashima, M. Takeda, and K. Tsuruta. The "runs" theorem. *CoRR*, abs/1406.0263v7, 2015.

[3] M. Crochemore. Repeats in strings. Keynote talk at the *25th Annual Symposium on Combinatorial Pattern Matching*, Moscow, June 2014. Personal communication.

[4] M. Crochemore and L. Ilie. Maximal repetitions in strings. *Journal of Computer and System Sciences*, 74(5):796 – 807, 2008.

[5] M. Crochemore, L. Ilie, and L. Tinta. The "runs" conjecture. *Theoretical Computer Science*, 412(27):2931–2941, 2011.

[6] M. Crochemore, C. S. Iliopoulos, M. Kubica, J. Radoszewski, W. Rytter, and T. Waleń. The maximal number of cubic runs in a word. *Journal of Computer and System Sciences*, 78(6):1828–1836, 2012.

[7] M. Crochemore and D. Perrin. Two-way string-matching. *Journal of the ACM*, 38(3):651–675, 1991.

[8] A. Deza and F. Franek. A d-step approach to the maximum number of distinct squares and runs in strings. *Discrete Applied Mathematics*, 163(3):268–274, 2014.

[9] J. Fischer, Š. Holub, T. I, and M. Lewenstein. Beyond the runs theorem. In *Proceedings of the 22nd International Symposium on String Processing and Information Retrieval, SPIRE*, volume 9309 of *LNCS*, pages 272–281, 2015.

[10] Š. Holub. Beyond the runs theorem. *CoRR*, abs/1502.04644v1, 2015.

[11] C. S. Iliopoulos, D. Moore, and W. F. Smyth. A characterization of the squares in a Fibonacci string. *Theoretical Computer Science*, 172(1–2):281–291, 1997.

[12] R. Kolpakov and G. Kucherov. Finding maximal repetitions in a word in linear time. In *40th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 596–604, New York, 1999.

[13] M. Lothaire. *Combinatorics on Words.* Cambridge University Press, second edition, 1997.

[14] S. J. Puglisi, J. Simpson, and W. F. Smyth. How many runs can a string contain? *Theoretical Computer Science*, 401(1-3):165–171, 2008.

[15] W. Rytter. The number of runs in a string: Improved analysis of the linear upper bound. In *23rd Annual Symposium on Theoretical Aspects of Computer Science, STACS*, volume 3884 of *LNCS*, pages 184–195, 2006.

[16] W. Rytter. The number of runs in a string. *Information and Computation*, 205(9):1459–1469, 2007.