SOME RESULTS ON FIXED POINTS OF THE BURROWS-WHEELER TRANSFORM

FRANCESCO DOLCE AND CHRISTIAN B. HUGHES

In this contribution, we present some results concerning fixed points of the Burrows-Wheeler Transform (BWT). In particular, we characterize all BWT fixed points which are Lyndon words, as well as all non-primitive BWT fixed points. We also give explicit formulæ for a few families of BWT fixed points according to their number of runs.

1. Preliminaries

Given an (ordered) alphabet $\mathcal{A}$, a *language* is a factorial set $\mathcal{L} \subset \mathcal{A}^*$. Given a (finite or infinite) word $w \in A^*$, we denote by $\mathcal{L}(w)$ the set containing all its factors. The set of letters appearing in the word is denoted by $\mathrm{alph}\,(w)$. A word $w \in \mathcal{A}^+$ is *pangrammatic* if $\mathrm{alph}\,(w) = \mathcal{A}$. It is *primitive* if $w = u^p$ for some $u \in \mathcal{A}^*$ implies $p = 1$.

Two words u, v are *conjugate* if $u = xy$ and $v = yx$ for some words $x, y \in \mathcal{A}^*$. We denote by $[w]$ the equivalence class of w for word conjugation. If w is primitive, then every word in $[w]$ is also primitive. A *Lyndon word* is the smallest element (for the lexicographic order on $\mathcal{A}$) of the conjugacy class of a primitive word.

A factor $w_i w_{i+1} \cdots w_{i+n}$ of a word w is a *run* if $w_i = w_{i+1} = \ldots = w_{i+n}$ and $w_{i-1} \neq w_i \neq w_{i+n+1}$, when w_{i-1}, w_{i+n+1} are defined. We denote by $\rho(w)$ the number of runs in a word w, and we define $\Delta(w) = \rho(w) - |\mathrm{alph}\,(w)\,|$.

Given a language $\mathcal{L}$ and a word $w \in \mathcal{L}$, the set $\mathcal{R}_{\mathcal{L}}(w)$ of *return words* to w in $\mathcal{L}$ is the set of words $u \in \mathcal{L}$ such that $uw \in \mathcal{L}$ with w appearing exactly twice in uw: once as a prefix and once as a suffix.

2. The Burrows-Wheeler Transform

The *Burrows-Wheeler transform* (BWT) is the map sending a word $w \in \mathcal{A}^*$ to the word $\mathrm{bwt}_{\mathcal{A}}\,(w)$ obtained by concatenating the last (not necessarily distinct) letters of the $|w|$ conjugates of w sorted lexicographically w.r.t. $\mathcal{A}$.

Given an alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_n\}$ and a permutation $\pi \in S_{\mathcal{A}}$, a word $w \in \mathcal{A}^*$ is π-clustering for $\mathcal{A}$ if $\mathrm{bwt}_{\mathcal{A}}\,(w) = a_{\pi(a_1)}^{k_1} a_{\pi(a_2)}^{k_2} \cdots a_{\pi(a_n)}^{k_n}$, for $k_i \geq 0$, $1 \leq i \leq n$. It is *perfectly clustering* if π is the symmetric permutation $\pi(i) = n-i+1$. When π is unimportant or clear from the context, we simply call such a word *clustering*.

Example 1. *Let* $w = \texttt{patata}$ *over the standard ordered Italian alphabet* $\mathcal{I} = \{\texttt{a} < \texttt{b} < \ldots < \texttt{z}\}$. *Arranging the conjugates of* w *from smallest to* *largest in the lexicographic order, we obtain the matrix*

$$\begin{bmatrix} \mathtt{a} & \mathtt{p} & \mathtt{a} & \mathtt{t} & \mathtt{a} & \mathtt{t} \\ \mathtt{a} & \mathtt{t} & \mathtt{a} & \mathtt{p} & \mathtt{a} & \mathtt{t} \\ \mathtt{a} & \mathtt{t} & \mathtt{a} & \mathtt{t} & \mathtt{a} & \mathtt{p} \\ \mathtt{p} & \mathtt{a} & \mathtt{t} & \mathtt{a} & \mathtt{t} & \mathtt{a} \\ \mathtt{t} & \mathtt{a} & \mathtt{p} & \mathtt{a} & \mathtt{t} & \mathtt{a} \\ \mathtt{t} & \mathtt{a} & \mathtt{t} & \mathtt{a} & \mathtt{p} & \mathtt{a} \end{bmatrix}.$$

Reading the last column of the above matrix gives $\mathrm{bwt}_{\mathcal{I}}(w) = \mathtt{ttpaaa}$*. The word* w *is thus (perfectly) clustering for* $\mathcal{I}$*.*

The Burrows-Wheeler transform is used in data pre-compression and is particularly effective when the language is somehow structurally repetitive (see [4] for a recent survey on the topic). In particular, we have the following bounds on the number of runs when applying the transform.

Theorem 1 ([8])**.** *Let* $w \in \mathcal{A}^*$*. Then,* $|\mathrm{alph}(w)| \leq \rho(\mathrm{bwt}_{\mathcal{A}}(w)) \leq 2\rho(w)$*.*

Example 2. *The two bounds in Theorem 1 are strict. An example of the lower bound is given in Example 1. On the other hand, the word* $w' = \mathtt{aacbbccc} \in \mathcal{A}^*$*, with* $\mathcal{A} = \{\mathtt{a} < \mathtt{b} < \mathtt{c}\}$*, is such that* $\rho(w') = 4$*, while* $\rho(\mathrm{bwt}_{\mathcal{A}}(w')) = 8$*.*

Note that changing the order of an alphabet $\mathcal{A}$ changes the behavior of the BWT when applied to words in $\mathcal{A}^*$.

Example 3. *Let us consider the word* w *of Example 1 over the three alphabets* $\mathcal{A} = \{\mathtt{a} < \mathtt{p} < \mathtt{t}\}$*,* $\mathcal{A}' = \{\mathtt{p} < \mathtt{t} < \mathtt{a}\}$*, and* $\mathcal{A}'' = \{\mathtt{a} < \mathtt{t} < \mathtt{p}\}$*. We have* $\mathrm{bwt}_{\mathcal{A}}(w) = \mathtt{ttpaaa}$*,* $\mathrm{bwt}_{\mathcal{A}'}(w) = \mathtt{aaattp}$*, and* $\mathrm{bwt}_{\mathcal{A}''}(w) = \mathtt{tptaaa}$*.*

The BWT is not an injective function on $\mathcal{A}^*$. Indeed, $\mathrm{bwt}_{\mathcal{A}}(u) = \mathrm{bwt}_{\mathcal{A}}(v)$ if and only if $[u] = [v]$. The map is not surjective (see [7] for an extended version that is surjective on $\mathcal{A}^*$). A description of the words obtainable as an image of the BWT is given in [5]. In the same paper, the authors describe which run sequences can be an image of the BWT. A sequence $(b_1, b_2, \ldots b_r)$ has a *global ascent* if there is an $1 \leq h < r$ such that $\max_{1 \leq i \leq h} b_i \leq \min_{h < j \leq r} b_j$.

Theorem 2 ([5])**.** *There exists a word of the form* $b_1^{k_1} b_2^{k_2} \cdots b_r^{k_r}$*, for some* $k_i > 0$*, as an image of the BWT if and only if the sequence* $(b_1, b_2, \ldots b_r)$ *has no global ascent.*

Example 4. *Let* $\mathcal{A}$ *be as in Example 3. The sequence* $(\mathtt{t}, \mathtt{p}, \mathtt{a})$ *has no global ascent, and we have shown that* $\mathrm{bwt}_{\mathcal{A}}(\mathtt{patata}) = \mathtt{t}^2\mathtt{p}^1\mathtt{a}^3$*.*

Corollary 1. *If* $w = w_1 w_2 \ldots w_m$ *is an image of the BWT over the alphabet* $\{a_1 < \cdots < a_n\}$*, with* $|\mathrm{alph}(w)| > 1$*. Then,* $w_1 \neq a_1$ *and* $w_n \neq a_n$*.*

Corollary 2. *Let* $w \in \mathcal{A}^*$ *be a word that is an image of the BWT. If* w *is a Lyndon word, then* $w \in \mathcal{A}$*.*

3. Fixed points of the BWT

A word w over an ordered alphabet $\mathcal{A}$ is a *fixed point* of the Burrows-Wheeler transform if $\mathrm{bwt}_{\mathcal{A}}(w) = w$.

Example 5. *The word* $w = \texttt{ccbca}$ *is a fixed point of the BWT over the ordered alphabet* $\mathcal{A} = \{\texttt{a} < \texttt{b} < \texttt{c}\}$*.*

It follows from Corollary 1 that the only fixed points of the BWT over a binary alphabet $\mathcal{A} = \{\texttt{a} < \texttt{b}\}$ are either powers of a letter or in $(\texttt{b}^{+}\texttt{a}^{+})^{+}$. In particular, for such a word w, $\Delta(w)$ is even.

It is well-known that if $\mathrm{bwt}_{\mathcal{A}}(w) = b_1 b_2 \cdots b_m$, then $\mathrm{bwt}_{\mathcal{A}}(w^p) = b_1^p b_2^p \cdots b_m^p$. The following result shows that if a word is both a non-trivial power and a fixed point of the BWT, it must have a very specific structure.

Proposition 1. *Let* $w \in \mathcal{A}^{+}$ *be a fixed point of the BWT. If* $w = u^p$ *with* $p > 1$ *then, either* $w = a^p$ *for some* $a \in \mathcal{A}$*, or* $w = (b^2a^2)^2$ *for some* $a < b \in \mathcal{A}$*.*

In particular, a fixed point of the BWT with at least three different letters is necessarily primitive.

A characterization of fixed points of the BWT in terms of permutations (standard permutation and FL-mapping) associated to a word is given in [6]. In the same paper, the authors state some explicit formulæ in the case of a binary alphabet with at most 8 runs.[1]

All words over a unary alphabet are clearly clustering and fixed points of the BWT. The set of pangrammatic words over the binary alphabet $\{\texttt{a} < \texttt{b}\}$ which are at the same time clustering and fixed points of the BWT is exactly $\texttt{ba}^{+} \cup \texttt{b}^{+}\texttt{a}$. The following result gives a characterization of clustering fixed points of the BWT.

Theorem 3. *Let* $w \in \mathcal{A}^{*}$ *be a fixed point for the BWT, with* $|\mathrm{alph}(w)| > 2$ *and such that* $\Delta(w) = 0$*. Then,* $w = a_m a_1^{k_1} a_2^{k_2} \cdots a_{m-1}^{k_{m-1}}$*, for certain* $a_1 < a_2 < \ldots < a_m \in \mathcal{A}$ *and positive integers* k_i*.*

In particular, each clustering fixed point of the BWT is maximal in its conjugacy class.

Example 6. *The word* $w = \texttt{eaabcdd}$ *over the alphabet* $\{\texttt{a} < \texttt{b} < \texttt{c} < \texttt{d} < \texttt{e}\}$ *is a clustering fixed point of the BWT.*

The situation appears to be more complicated when $\Delta(w) > 0$.

From the observations above, it is clear that over a unary or binary alphabet there is no pangrammatic fixed point w of the BWT with $\Delta(w) = 1$. The following result gives a characterization of fixed points of the BWT over a ternary alphabet with exactly 4 runs.

Theorem 4. *Let* $w \in \{\texttt{a} < \texttt{b} < \texttt{c}\}^{*}$ *be a fixed point for the BWT with* $\Delta(w) = 1$*. Then, there exists an integer* $k > 0$ *such that* w *is one of the following words:* $\texttt{c}^{2k+1}\texttt{ac}^{k}\texttt{b}$, $\texttt{c}^{2k}\texttt{bc}^{k}\texttt{a}$, *or* $\texttt{c}^{2}\texttt{b}^{k}\texttt{ca}$.

For larger alphabets, the structure of runs in a fixed point w of the BWT with $\Delta(w) = 1$ is more restrictive.

Theorem 5. *Let* $w \in \mathcal{A}^{*}$ *be a fixed point for the BWT with* $|\mathrm{alph}(w)| > 2$ *and* $\Delta(w) = 1$*. Then,* $w = a_m^2 a_2^{k_2} a_3^{k_3} \cdots a_{m-1}^{k_{m-1}} a_m a_1$*, for certain* $a_1 < a_2 < \ldots < a_m \in \mathcal{A}$ *and positive integers* k_i*.*

---

[1] The list of conditions corresponding to 4 occurrences of $\texttt{a}$ and 4 occurrences of $\texttt{b}$ in the paper is incomplete, since $\texttt{b}^2\texttt{a}^2\texttt{b}^2\texttt{a}^2$ is a fixed point of the BWT.

4. Interval Exchange Transformations

Let $I = [\ell, r)$ be an interval with $\ell, r \in \mathbb{R}$, $\ell < r$. Let $\mathcal{A}$ be an ordered alphabet and $\pi \in S_{\mathcal{A}}$. An ordered partition $(I_a)_{a\in\mathcal{A}}$ of I is such that I_a is to the left of I_b whenever $a < b$. The n-interval exchange transformation (or n-IET for short) on I associated with π and $(I_a)_{a\in\mathcal{A}}$ is the piecewise translation defined by $T(x) = x + \tau_a$ for $x \in I_a$, with $\tau_a = \sum_{\pi^{-1}(b)<\pi^{-1}(a)} |I_b| - \sum_{b<a} |I_b|$. An IET with the symmetric permutation $\pi : i \mapsto n - i + 1$ is said to be *symmetric*. The *natural coding* of a point $x \in I$ under an IET T is the infinite word $\Omega_T(x) = w_0 w_1 w_2 \ldots$, with $w_k = a$ if $T^k(x) \in I_a$. The *language* of T is $\mathcal{L}(T) = \bigcup_{x\in I} \mathcal{L}(\Omega_T(x))$.

Interval exchange languages have rich clustering properties. Indeed, a word w is clustering if and only if ww appears in the language of an interval exchange transformation ([2]). Moreover, it was shown in both [1] and [3] that, given an IET T with associated permutation π, all return words in $\mathcal{L}(T)$ are π-clustering.

Theorem 6. *Let T be an IET over the alphabet $\mathcal{A} = \{a_1 < \ldots < a_m\}$. If there exists a pangrammatic fixed point of the BWT in the set of return words in $\mathcal{L}(T)$, then the permutation associated to T is $\pi = (a_1, a_m, a_{m-1} \ldots, a_2)$.*

Corollary 3. *Let T be a symmetric interval exchange transformation over the alphabet $\mathcal{A} = \{a_1 < a_2 < \cdots < a_m\}$. Every fixed point of the BWT in the set of return words in $\mathcal{L}(T)$ contains at most two distinct letters.*

References

[1] Francesco Dolce and Christian B. Hughes. Extended branching Rauzy induction. *arXiv preprint arXiv:2511.22588*, 2025.

[2] Sébastien Ferenczi and Luca Q. Zamboni. Clustering words and interval exchanges. *Journal of Integer Sequences*, 16, 2013.

[3] Sébastien Ferenczi and Luca Q. Zamboni. Clustering, order conditions, and languages of interval exchanges. *arXiv preprint arXiv:2507.17370*, 2025.

[4] Gabriele Fici, Sabrina Mantaci, Antonio Restivo, Giuseppe Romana, Giovanna Rosone, and Marinella Sciortino. BWT and combinatorics on words. In *The Expanding World of Compressed Data: A Festschrift for Giovanni Manzini's 60th Birthday (2025)*, pages 1–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2025.

[5] Konstantin M. Likhomanov and Arseny M. Shur. Two combinatorial criteria for BWT images. In *International Computer Science Symposium in Russia*, pages 385–396. Springer, 2011.

[6] Sabrina Mantaci, Antonio Restivo, Giovanna Rosone, Floriana Russo, and Marinella Sciortino. On Fixed Points of the Burrows-Wheeler Transform. *Fundamenta Informaticae*, 154(1-4):277–288, 2017.

[7] Sabrina Mantaci, Antonio Restivo, Giovanna Rosone, and Marinella Sciortino. An extension of the Burrows–Wheeler transform. *Theoretical Computer Science*, 387(3):298–312, 2007.

[8] Sabrina Mantaci, Antonio Restivo, Giovanna Rosone, Marinella Sciortino, and Luca Versari. Measuring the clustering effect of BWT via RLE. *Theoretical Computer Science*, 698:79–87, 2017.

Czech Technical University in Prague (Czech Republic)
Email address: `dolcefra@fit.cvut.cz`

Czech Technical University in Prague (Czech Republic)
Email address: `hughechr@fit.cvut.cz`