

# The critical factorization theorem

Dominique Perrin

October 1, 2007

- The critical factorization theorem
- consequences
- A proof by Maxime
- conclusion

# Two-Way String-Matching

MAXIME CROCHEMORE AND DOMINIQUE PERRIN

*L.I.T.P., Institut Blaise Pascal, Université Paris 7, 2, Place Jussieu, Paris*

**Abstract.** A new string-matching algorithm is presented, which can be viewed as a middle ground between the classical algorithms of Knuth, Morris, and Pratt on the one hand and Seiferas on the other hand. The algorithm is linear in time and uses constant space as the latter. It presents the advantage of being remarkably simple which consequently makes its implementation as easy as possible. The algorithm relies on a previously known result in combinatorics, the *Critical Factorization Theorem*, which relates the global period of a word to the lengths of its blocks.

# The critical factorization theorem

The **repetition** of  $(u, v)$  is the minimal length of a word left compatible with  $u$  and right compatible with  $v$ . It is denoted  $r(u, v)$ .

## Theorem (Cesari, Duval)

*For each word  $x$ , there is at least one factorization  $(u, v)$  of  $x$  such that*

$$r(u, v) = p(x).$$

*Moreover  $u$  can be chosen with  $|u| < p(x)$ .*

$(u, v)$  is a **Critical factorization** of  $x$ .

# Example

Let  $x = \mathit{abaababa}$  which has period 5 (Fibonacci). Then  $\mathit{abaababa}$  is a critical factorization with  $w = \mathit{babaa}$ . Repetitions:

$a$	$b$	$a$	$a$	$b$	$a$	$b$	$a$
2	3	1	5	5	2	2	

- A word having more than  $\text{Card}(X) + 1$  disjoint factorizations in words of a finite set  $X$  has period at most  $\max_{x \in X} |x|$ .
- The two-way string-matching algorithm (MC & DP, 1991).
- Simpler proofs of results in combinatorics on words (e.g. the solutions of  $x^n y^m = z^p$  (Lyndon and Schützenberger, 1962)

- 1 A simple proof in the particular case  $3p(x) \leq |x|$  consists in writing  $x = \ell y r$  with  $y$  Lyndon.
- 2 The original proof by Cesari, Vincent, 1978 (with a slightly different definition of a critical factorization)
- 3 The proof by Duval, 1979 (based on the property that if  $x = a y b$  with  $a, b$  letters, either  $a y$  or  $y b$  give a critical factorization of  $x$ ).
- 4 a proof by Maxime, 1991.

## Theorem

Let  $\leq$  and  $\subseteq$  be the alphabetic orders induced by two inverse total orders on the alphabet  $A$ . Let  $v$  (resp.  $v'$ ) be the maximal suffix of a nonempty word  $x$  for  $\leq$  and  $\subseteq$ . Let  $x = uv = u'v'$ .

If  $|v| \leq |v'|$ , then  $(u, v)$  is a critical factorization, otherwise  $(u'v')$  is.

Example:

$$\begin{aligned} \text{abaababaabaab} &= \text{abaababaabaab} \\ &= \text{abaababaabaab} \end{aligned}$$

The second factorization is critical.

- Further consequences of the factorization theorem
- Relation between the proofs?