# Tutorial Week 2

**Definition 1.** *For a string $w$, an integer $p$ with $0 < p \leq |w|$ is a* **period** *of $w$ if for all defined positions $i$ and $i + p$ in $w$, we have $w[i] = w[i + p]$ (here by $w[i]$ we refer to the symbol in position $i$ of $w$, thus $i < |w|$). A string $u$ is a* **border** *of $w$ if $u$ is both a prefix and a suffix of $w$, but $u \neq w$.*

**Definition 2.** *For two strings $u$ and $v$, we say that $u$ is a* **conjugate** *of $v$, if there exist two strings $x$ and $y$ such that $u = xy$ and $v = yx$.*

**Exercise 1.** *For the following list of strings, give the list of all their conjugates, and the full lists of their borders and their periods.*

| | conjugates | borders | periods |
|---|---|---|---|
| $ababab$ | | | |
| $aaaaaa$ | | | |
| $abcacb$ | | | |
| $abaaba$ | | | |

*Solution:*

| | conjugates | borders | periods |
|---|---|---|---|
| $ababab$ | $\{ababab,$ $bababa\}$ | $\{\varepsilon, ab, abab\}$ | $\{2, 4, 6\}$ |
| $aaaaaa$ | $\{aaaaaa\}$ | $\{\varepsilon, a, aa, aaa, aaaa,$ $aaaaa\}$ | $\{1, 2, 3, 4, 5, 6\}$ |
| $abcacb$ | $\{abcacb$ $bcacba$ $cacbab$ $acbabc$ $cbabca$ $babcac\}$ | $\{\varepsilon\}$ | $\{6\}$ |
| $abaaba$ | $\{abaaba$ $baabaa$ $aabaab\}$ | $\{\varepsilon, a, aba\}$ | $\{3, 5, 6\}$ |

■

**Proposition 2.** *Two strings $u$ and $v$ are conjugate if and only if there exists a string $z$ such that $uz = zv$*

*Proof.* Let us first assume that $u$ and $v$ are conjugate. Then there exist two strings $x$ and $y$ such that $u = xy$ and $v = yx$. It is straightforward that taking $z = x$, the result follows, $uz = xyx = zv$.

For the other direction, we assume that $uz = zv$ and we want to prove that $u$ and $v$ are conjugate. For this, observe first that $u$ and $v$ have the same length. Furthermore, by comparing the lengths of $v$ and $z$, we note that there must exist an integer $k > 0$ such that $|z| < |u^k| \leq |u| + |z|$.

Since $uz = zv$, by considering the string $u^{k+1}z$ we have the following situation:

$$u^{k+1}z = u^k uz = u^k zv = u^{k-1}uzv = u^{k-1}zv^2 = \cdots = zv^{k+1}$$

But, since $|u^{k+1}| > |z|+|u| < |v^{k+1}|$, we conclude from the above equality that $u^{k+1}$ and $v^{k+1}$ have an overlap of at least $|v|$ positions, or, in other words, $u$ is a factor of $vv$ and $v$ is a factor of $uu$. However, any of these implies that there exist some strings $x$ and $y$ such that $u = xy$ and $v = yx$, and our conclusion follows. $\square$

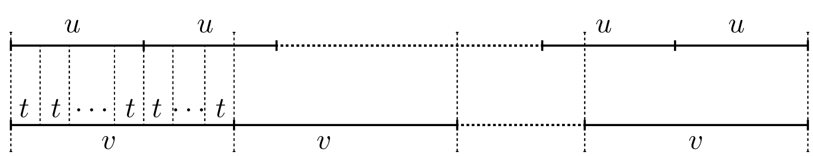


Figure 1: $u\{u, v\}^k$ and $v\{u, v\}^\ell$ are in a perfect alignment, and $u = t^k$

**Lemma 3** (Fine and Wilf – Periodicity Lemma)**.** *If a string can be written as either $u\{u, v\}^k$ and $v\{u, v\}^\ell$, respectively, for some positive integers $k, \ell > 0$, such that its length is at least $|u|+|v|$, then the string is* $\gcd(|u|, |v|)$*-periodic.*

*Proof.* A string is of the form $u\{u, v\}^k$ (or form $v\{u, v\}^\ell$) if it starts with $u$ (resp. $v$) and it consists only of concatenations of occurrences of $u$ or $v$.

Let us denote such a string by $w$. We shall prove our result by induction on the length of $|u| + |v|$.

If $|u| = |v|$ since $w$ has both $u$ and $v$ as prefixes, it follows that $u = v$, and the conclusion follows. This also includes our base step when $|u| = |v| = 1$.

Therefore, let us assume without loss of generality that $1 \leq |u| < |v|$ and that the condition is true for any string having a decomposition as ours involving some strings whose total length is less than $|u| + |v|$. We immediately have that $u$ is a prefix of $v$. Furthermore, since $w = v\{u, v\}^{\ell}$, its prefix must be $vu$ (see Figure 1).

Since $u$ is a prefix of $v$, we can denote $v = ux$, for some non-empty string $x$. But, in this case we note that, in fact, one can express the suffix of $w$ of length $|w| - |u|$ in terms of $x$ and $u$; that is this suffix is of both form $u\{u, x\}^r$ and $x\{u, x\}^s$, for some integers $r \geq k$ and $s \geq \ell$. However, according to our induction, we know that this suffix must be $\gcd(|u|, |x|)$-periodic.

Since we know that $\gcd(|u|, |v|) = \gcd(|u|, |x|)$, it follows that both $u$ and $x$, and therefore $v$, are all $\gcd(|u|, |v|)$-periodic. Hence, $w$ which consists only of concatenations of $u$ and $v$ must also be $\gcd(|u|, |v|)$-periodic, and our conclusion follows.

For more details on gcd calculations look at the Euclidian Algorithm (`http://en.wikipedia.org/wiki/Euclidean_algorithm`) □

**Proposition 4.** *If $w$ is a primitive string, then $w$ occurs as a factor of $ww$ only as a prefix or as a suffix (prove using Lemma 3).*

*Proof.* Assume towards a contradiction that $w$ has a third occurrence in $w$. Then there exist non-empty strings $u$ and $v$ such that $ww = uwv$. It immediately follows that $u$ is a prefix of $w$, while $v$ is a suffix of $w$. Furthermore, since $2|w| = |u| + |w| + |v|$, we have that $w = uv$.

Since $uvuv = ww = uwv = uuvv$, by looking at the factor of length $|uv|$ starting at position $|u|$, we have $uv = vu$. By Lemma 3, we get that $u$ and $v$ are both $\gcd(|u|, |v|)$-periodic. Since one is a prefix of the other, it follows that both are powers of the same string, and, moreover, $w$ is also a power of this string. But, since $|w| > |u|, |v|$ we get that $w$ is non-primitive, which is in contradiction with our assumption. The result follows. □

**Proposition 5.** *If for strings $u$ and $v$ we have $u^k = v^{\ell}$, for some integers $k, \ell > 0$, then $u$ and $v$ are powers of the same string (prove using Lemma 3).*

*Proof.* Observe that if $\ell = 1$, then either $u = v$ or $v = u^k$, which follows our statement. The same happens when $k = 1$. If $k, \ell > 1$, then we have that $u^k = v^{\ell}$ is both $|u|$ and $|v|$ periodic, and its length is greater than $\max(|u^2|, |v^2|) \geq |u| + |v|$. Therefore, following Lemma 3, we have that $u^k = v^{\ell}$ is also $\gcd(|u|, |v|)$-periodic. The later implies that both $u$ and $v$ are $\gcd(|u|, |v|)$-periodic, and since they align, the conclusion follows. □

**Exercise 6.** *Use the rolling hash technique to find the representation of all factors of length* 5 *in base 7 modulo* 9*, for each of the following strings:* `1234560123`*,* `2312132132`*, and* `5534555345`*. Finding only the correct value is NOT enough.*

*Solution:* For `1234560123` we have the following list of factors with the corresponding values:

$$
\begin{aligned}
&h(12345) = \\
&(((((((((1\cdot 7+2) \bmod 9)\cdot 7+3) \bmod 9)\cdot 7+4) \bmod 9)\cdot 7+5) \bmod 9)\cdot 7^0 \\
&\qquad = (((((0+3) \bmod 9)\cdot 7+4) \bmod 9)\cdot 7+5) \bmod 9 = \\
&\qquad\qquad = ((7{\cdot}7)+5) \bmod 9 = 0
\end{aligned}
$$

$$
\begin{aligned}
h(23456) &= ((h(12345) - 1\cdot(7^4 \bmod 9))\cdot 7+6) \bmod 9 = \\
&= ((0-7)\cdot 7+6) \bmod 9 = -43 \bmod 9 = 2
\end{aligned}
$$

$$h(34560) = ((h(23456) - 2\cdot 7)\cdot 7+0) \bmod 9 = (-12\cdot 7+0) \bmod 9 = 6$$

$$h(45601) = ((h(34560) - 3\cdot 7)\cdot 7+1) \bmod 9 = (-15\cdot 7+1) \bmod 9 = 4$$

$$h(56012) = ((h(45601) - 4\cdot 7)\cdot 7+2) \bmod 9 = (-24\cdot 7+2) \bmod 9 = 5$$

$$h(60123) = ((h(56012) - 5\cdot 7)\cdot 7+3) \bmod 9 = (-30\cdot 7+3) \bmod 9 = 0$$

For `2312132132` we have the following list of values, in the order of the appearance of each factor: $\{0,4,0,6,1,5\}$

For `5534555345` we have the following list of values, in the order of the appearance of each factor. Here each hash value is numbered once, at its first occurrence: $\{4,4,4,1,4\}$. Please observe that while $h(55345)$ always has the same value, we cannot say that the hash-print of both 55345 and 34555 are the same, although their values are equal. ∎