Words with more periods

Let $w[i]$, $0 \leq i < |w|$, be the $(i-1)$th letter of the word w, so that $w = w[0]w[1]\cdots w[n-1]$ with $n = |w|$. We say that $p \geq 1$ is *a period* of w if $w[i] = w[i+p]$ holds for all $0 \leq i < n-p$. The least period of w is called *the period* of w.

By the definition, any word has infinitely many periods. Namely, each $p \geq |w|$ is a period of w. Also, each multiple of a period is again a period. This are trivial examples of multiple periods. However, there are also nontrivial cases. For example, the word *abaababaaba* of length 11 has periods 5 and 8. When this can happen explains the Theorem of Fine and Wilf, also called the *Periodicity lemma.*

Theorem. If a word of length at least $p + q - \gcd(p,q)$ has periods p and q, then they have also a period $\gcd(p,q)$.

On the other hand, for each $p < q$ such that $p \nmid q$, there is a word of length $p + q - \gcd(p,q) - 1$ having periods p and q, and not a period $\gcd(p,q)$.

We give several proofs of this fundamental result. The first one is by induction.

Proof. Let WLOG $p \leq q$. To proof the first claim, proceed by induction on $p+q$. The first part of the claim holds if $p = q$. Let $p < q$, and set $d := \gcd(p,q)$. Let w have periods p and q with $|w| \geq p + q - d$. Consider the prefix v of w of length $|w| - p$. The word v has a period p. We show that it also has a period $q - p$. If $i < |v| - (q-p)$, then $i + q < |w|$, and we have

$$v[i+q-p] = w[i+q-p] = w[i+q] = w[i] = v[i]\,.$$

Since $d = \gcd(p,q) = \gcd(p, q-p)$, we have $|v| \geq p + (q-p) - \gcd(p, q-p)$, and v has a period d by the induction assumption. Note that $|v| \geq q - d \geq p$. Thus

$$w[i] = w[i \bmod p] = v[i \bmod p] = v[(i+d) \bmod p] = w[(i+d) \bmod p)] = w[i+d]$$

holds for each $i < |w| - d$ since p is a period of w. The equality

$$v[i \bmod p] = v[(i+d) \bmod p]$$

above holds for any i, since $d \mid p$ which implies that the difference between $i \bmod p$ and $i + d \bmod p$ is divisible by d.

In order to show optimality, assume $p < q$ and $p \nmid q$ with $d = \gcd(p,q)$. Consider first the word

$$(a^{d-1}b)^{k-1}a^{d-1}c(a^{d-1}b)^{k-1}a^{d-1}$$

which has periods kd and $(k+1)d$, but not the period d. This shows the optimality if $d = q - p$. Assume that $d < q - p$. Then, by induction, there is a word v of length $q - d - 1$ having periods p and $q - p$ but not having the period d. Extend v to the word w of length $p + q - d - 1$ so that it has a period p. Certainly, the number d is not a period of w, and it remains to show that w has a period q. Let $i < (p + q - d - 1) - q = p - d - 1$. Then

$$w[i] = v[i] = v[i+q-p] = w[i+q-p] = w[i+q].$$

Crucial fact here is that $i + q - p < q - d - 1 = |v|$. □

The following proof makes the modular computation more explicit.

Proof. First, suppose that p and q are coprime. We $\approx$ be the smallest equivalence on the set $I = \{0, 1, \ldots, p+q-2\}$ (that is, on letter indeces of w) satisfying $i \approx (i \bmod p)$ and $i \approx (i \bmod q)$. The definition implies that $w[i] = w[j]$ if $i \approx j$. We want to show that all elements of I are equivalent. Obviously, it is enough to show that for $\{0, 1, \ldots, p-1\}$.

Set $i_k := (kq \bmod p)$. Because p and q are coprime, the number q is a generator of the cyclic group $\mathbb{Z}_p$, that is, $\{i_0, i_1, \ldots, i_{p-1}\} = \{0, 1, \ldots, p-1\}$. If $i_k < p-1$, then $i_k + q < p - 1 + q$, which implies $i_k \approx i_k + q \approx i_{k+1 \bmod p}$. We deduce that all elements of $\{0, 1, \ldots, p-1\}$ are equivalent.

Let now $\gcd(p, q) = d$. For $r \in \{0, 1, \ldots, d-1\}$ dlet

$$w_r := w[r]w[r+d]w[r+2d] \cdots w[r + (p' + q' - 2)d],$$

with $p' = p/d$ a $q' = q/d$. It is easy to see that w has periods p a q if and only if w_r has periods p' a q' for each $r = 0, 1, \ldots, d-1$. Since the words w_r are of length $p' + q' - 1$, they are by the first part of the proof powers of the same letter. Thus w has a period d.

Let us show optimality. Let w be of length $p + q - \gcd(p, q) - 1$ with $p < q$ and $p \nmid q$. Again, we first suppose that p a q are coprime. If the equivalence $\approx$ has at least two classes, then we can identify each $w[i]$ with the equivalence class $[i]_\approx$, in order to obtain a word with required properties. Consider again only words $\{0, 1, \ldots, p-1\}$ and view that as vertices of a directed graph G, in which $i \longrightarrow j$ holds if and only if $i + q < p + q - 2$ a $i + q \equiv j \mod p$. It is easy to see that classes of $\approx$ restricted to $\{0, 1, \ldots, p-1\}$ are (weakly connected) components of G. Each vertex has obviously outgoing and incoming degree at most one. Also, $p-1$ a $p-2$ have the outcoming degree zero. Similarly $q-1 \mod p$ a $q-2 \mod p$ have indegree zero. This implies that G has more than one component and the word w is nontrivial. Let now $\gcd(p, q) = d$. The word w_{d-1} defined as above has length $p' + q' - 2$, since $(d-1) + (p' - q' - 2)d = p + q - d - 1$. Therefore, it can contain two different letters and w does not have a period d. □

The following proof uses the Fourier transform. For this proof, it is convenient to reformulate the claim in terms of sequences.

Theorem. Let $f = (f_n)_{n \in \mathbb{N}}$ and $g = (g_n)_{n \in \mathbb{N}}$ be sequences with periods p and q respectively. If $f_n = g_n$ for $0 \le n < p + q - \gcd(p, q)$, then $f = g$, and it has a period $\gcd(p, q)$.

On the other hand, for each p and q, there are two distinct f and g, with periods p and q respectively, such that $f_n = g_n$ for $0 \le n < p + q - \gcd(p, q) - 1$.

Proof. Let the alphabet be from $\mathbb{C}$. Let $d = \gcd(p, q)$. Let $\varphi_{m,n}$ denote the sequence with jth coefficient

$$\varphi_{m,n}(j) = e^{2\pi i \frac{nj}{m}}.$$

Since f has a period p, it is generated by the set

$$\Phi_p = \{\varphi_{p,k} \mid k = 0, 1, \ldots, p-1\}$$

of p sequences with the period p. In the same way, since g has a period q, it is generated by the set

$$\Phi_q = \{\varphi_{q,k} \mid k = 0, 1, \ldots, q-1\}.$$

The set $\Phi = \Phi_p \cup \Phi_q$ contains exactly $p + q - \gcd(p, q)$ (distinct) elements.

Consider now the common $p + q - d$ first values of f and g as the element $h \in \mathbb{C}^{p+q-d}$, and let $\Phi' \subset \mathbb{C}^{p+q-d}$ be the initial parts of elements of Φ. The key observation is that Φ' is linearly independent. That follows from the fact that Φ' forms a Vandermonde matrix, or in other terms, the vectors are values of $p+q-\gcd(p,q)$ distinct polynomials of degree less than $p+q-\gcd(p,q)$ in $p+q-\gcd(p,q)$ distinct points. This implies that h is given uniquely as linear combination of elements of Φ', hence also the two expressions in terms of Φ_p and of Φ_q must be the same. Therefore, $f=g$ and it is generated by elements of the set $\Phi_p \cap \Phi_q = \Phi_d = \{\varphi_{d,k} \mid k=0,\dots,d-1\}$ of sequences with period d.

On the other hand, Φ' generates the vector $e_{p+q-d-1} = (0,0,\dots,0,1)$ which can be therefore written as the difference of two sequences generated by Φ_p and $\Phi \setminus \Phi_p \subset \Phi_q$ respectively. Such sequences have periods p and q respectively, they agree on first $p+q-d-1$ positions but differ on the next position. □

The last proof uses formal series.

Proof. Let the sequences be represented by formal series $f=\sum_{n\in\mathbb{N}} f_n x^n$ and $g=\sum_{n\in\mathbb{N}} g_n x^n$. Due to their periods, the sequences can be written as

$$f = \frac{P}{(1-x^p)}, \qquad\qquad g = \frac{Q}{(1-x^q)},$$

where P and Q are polynomials with degree less that p and q respectively.Note that $\gcd(1-x^p, 1-x^q) = 1-x^d$ with $d=\gcd(p,q)$. We have

$$\begin{aligned} f-g &= \frac{P}{1-x^p} - \frac{Q}{1-x^q} = \frac{(1-x^d)}{(1-x^p)(1-x^q)} \left(P\frac{(1-x^q)}{(1-x^d)} - Q\frac{(1-x^p)}{(1-x^d)} \right) \\ &= \frac{(1-x^d)}{(1-x^p)(1-x^q)}\, R, \end{aligned}$$

which is a product of a formal series with the absolute coefficient 1, and a polynomial R of degree less than $p+q-d$. This implies that if R is not zero, then the least non-zero coefficient of $f-g$ has index less than $p+q-d$. In other words, if f and g agree on first $p+q-d$ positions, then $R=0$, and $f=g$. Then also

$$P\,\frac{(1-x^q)}{(1-x^d)} = Q\,\frac{(1-x^p)}{(1-x^d)}\,.$$

Since $(1-x^q)/(1-x^d)$ and $(1-x^p)/(1-x^d)$ are coprime, polynomials P and Q are divisible by $(1-x^p)/(1-x^d)$ and $(1-x^q)/(1-x^d)$ respectively. Therefore

$$f = g = \frac{D}{(1-x^d)},$$

where

$$D = \frac{P(1-x^d)}{1-x^p} = \frac{Q(1-x^d)}{1-x^q}$$

is a polynomial of degree less than d, and $f=g$ has a period d.

On the other hand, if we put $R = x^{p+q-d-1}$, then there are polynomials P and Q of degree less than p and q respectively satisfying

$$P\,(1-x^q) - Q\,(1-x^p) = (1-x^d)\,x^{p+q-d-1}.$$

The corresponding f and g then agree on first $p+q-d-1$ positions but disagree on the next position. □

Remark: The polynomials P and Q from the last proof are obtained as follows. The extended Euclidean algorithm yields P' and Q' such that

$$P'\,(1-x^q) - Q'\,(1-x^p) = \gcd(1-x^p, 1-x^q) = 1-x^d.$$

We now set

$$P = P' \, x^{p+q-d-1} \mod (1 - x^p), \qquad Q = Q' \, x^{p+q-d-1} \mod (1 - x^q).$$