Theoretical Computer Science 295 (2003) 107–121
Theoretical Computer Science
www.elsevier.com/locate/tcs

Periodicity, morphisms, and matrices

Sabin Cautis[a], Filippo Mignosi[b,1], Jeffrey Shallit[c,*,2], Ming-wei Wang[c], Soroosh Yazdani[d]

[a] *Department of Mathematics, Harvard University, Cambridge, MA 02138, USA*
[b] *Dipartimento di Matematica ed Applicazioni, Università di Palermo, Via Archirafi 34, 90123 Palermo, Italy*
[c] *Department of Computer Science, University of Waterloo, Waterloo, Ont., Canada N2L 3G1*
[d] *Department of Mathematics, University of Illinois, Urbana, IL 618101, USA*

Received 24 September 2001; received in revised form 4 December 2001; accepted 5 December 2001

Abstract

In 1965, Fine and Wilf proved the following theorem: if $(f_n)_{n\geqslant 0}$ and $(g_n)_{n\geqslant 0}$ are periodic sequences of real numbers, of period lengths h and k, respectively, and $f_n = g_n$ for $0 \leqslant n < h + k - \gcd(h,k)$, then $f_n = g_n$ for all $n \geqslant 0$. Furthermore, the constant $h + k - \gcd(h,k)$ is best possible. In this paper, we consider some variations on this theorem. In particular, we study the case where $f_n \leqslant g_n$ instead of $f_n = g_n$. We also obtain generalizations to more than two periods.

We apply our methods to a previously unsolved conjecture on iterated morphisms, the decreasing length conjecture: if $h : \Sigma^* \to \Sigma^*$ is a morphism with $|\Sigma| = n$, and w is a word with $|w| > |h(w)| > |h^2(w)| > \cdots > |h^k(w)|$, then $k \leqslant n$.

Keywords: Periodicity; Iterated morphism; Sturmian word

1. Introduction

In this paper we explore several related topics: some generalizations of a classical theorem of Fine and Wilf, the solution of a conjecture on the length sequence obtained

* Corresponding author.
E-mail addresses: scautis@math.harvard.edu (S. Cautis), mignosi@math.unipa.it (F. Mignosi), shallit@math.uwaterloo.ca (J. Shallit), m2wang@math.uwaterloo.ca (M.-w. Wang), syazdani@math.uiuc.edu (S. Yazdani).
URLs: http://dipinfo.math.unipa.it/∼mignosi, http://www.math.uwaterloo.ca/∼shallit
[1] Research supported in part by CNR-NATO fellowship no. 215.31.
[2] Research supported in part by a grant from NSERC.

PII: S0304-3975(02)00398-5

by iterating a morphism, and an equivalent problem about non-negative matrices. The single thread uniting these different topics is periodicity.

Periodicity is an important property of words that has applications in various domains. For instance, it has applications in string searching algorithms (cf. [5]), in formal languages (cf. for instance the pumping lemmas in [14]), and it is an important part of combinatorics on words (cf. [4,1]).

We say a sequence $(f_n)_{n\geqslant 0}$ is periodic with period length $h\geqslant 1$ if $f_n = f_{n+h}$ for all $n\geqslant 0$. The following is a classical "folk theorem":

Theorem 1.1. *If $(f_n)_{n\geqslant 0}$ is a sequence of real numbers which is periodic with period lengths h and k, then it is periodic with period length* $\gcd(h,k)$.

Proof. By the extended Euclidean algorithm, there exist integers $r,s\geqslant 0$ such that $rh - sk = \gcd(h,k)$. Then we have

$$f_n = f_{n+rh} = f_{n+rh-sk} = f_{n+\gcd(h,k)}$$

for all $n\geqslant 0$. □

The 1965 theorem of Fine and Wilf [7] is the following:

Theorem 1.2. *Let $(f_n)_{n\geqslant 0}$, $(g_n)_{n\geqslant 0}$ be two periodic sequences of real numbers, of period lengths h and k, respectively.*

(a) *If $f_n = g_n$ for $0\leqslant n < h+k-\gcd(h,k)$, then $f_n = g_n$ for all $n\geqslant 0$.*

(b) *The conclusion in* (a) *would be false if $h+k-\gcd(h,k)$ were replaced by any smaller number.*

We first consider some variations on the theorem of Fine and Wilf in which equality is replaced by inequality.

2. First variation

We begin with a bit of notation and a lemma. Let $\mathbf{a} = (a_i)_{i\geqslant 0}$ be a sequence of real numbers, and let $p = (p_0, p_1, \ldots, p_{h-1})$ be a vector of real numbers of dimension $h\geqslant 1$. We will frequently need the new sequence $p\circ\mathbf{a}$ resulting from taking successive "windows" of length h of $\mathbf{a}$ and forming their dot product with p. More formally, we define $p\circ\mathbf{a} := (\sum_{0\leqslant i<h} p_i a_{n+i})_{n\geqslant 0}$.

Lemma 2.1. *Let $p = (p_0, p_1, \ldots, p_{h-1})$ be a vector of $h\geqslant 1$ real numbers and $q = (q_0, q_1, \ldots, q_{k-1})$ be a vector of $k\geqslant 1$ real numbers. Then $q\circ(p\circ\mathbf{a}) = (qp)\circ\mathbf{a}$, where by qp we mean the vector $(r_0, r_1, \ldots, r_{h+k-2})$ defined by*

$$r_n = \sum_{\substack{0\leqslant i<h\\ 0\leqslant j<k\\ i+j=n}} p_i q_j.$$

Proof. Define $P(z)=\sum_{0\leqslant i<h} p_i z^i$, $Q(z)=\sum_{0\leqslant i<k} q_i z^i$, and $A(z)=\sum_{i\geqslant 0} a_i z^{-i}$. If $p\circ \mathbf{a}=(t_i)_{i\geqslant 0}$ then it is easy to see that $P(z)A(z)=(\sum_{i\geqslant 0} t_i z^{-i})+W(z)$, where W is a polynomial of degree $\deg P$ such that $W(0)=0$. If $q\circ(p\circ\mathbf{a})=(u_i)_{i\geqslant 0}$ it follows that $Q(z)P(z)A(z)=(\sum_{i\geqslant 0} u_i z^{-i})+S(z)$ where $S(0)=0$. Hence $q\circ(p\circ\mathbf{a})=(qp)\circ\mathbf{a}$. □

For the rest of this paper, we abuse notation slightly by writing $P\circ\mathbf{a}$ for $p\circ\mathbf{a}$, where $p=(p_0,p_1,\ldots,p_{h-1})$ and $P(z)=\sum_{0\leqslant i<h} p_i z^i$.

We are now ready to state and prove our first variation on the theorem of Fine and Wilf.

Theorem 2.2. *Let $\mathbf{f}=(f_n)_{n\geqslant 0}$, $\mathbf{g}=(g_n)_{n\geqslant 0}$ be two periodic sequences of real numbers, of period lengths h and k, respectively, such that*

$$\sum_{0\leqslant i<h} f_i \geqslant 0 \tag{1}$$

and

$$\sum_{0\leqslant j<k} g_j \leqslant 0. \tag{2}$$

Let $d=\gcd(h,k)$.
(a) *If*

$$f_n \leqslant g_n \quad \text{for } 0 \leqslant n < h+k-d \tag{3}$$

then
(i) *$f_n=g_n$ for all $n\geqslant 0$; and*
(ii) *$\sum_{j\leqslant i<j+d} f_i=\sum_{j\leqslant i<j+d} g_i=0$ for all integers $j\geqslant 0$.*
(b) *Conclusion* (a)(i) *would be false if in the hypothesis $h+k-d$ were replaced by any smaller integer.*

Proof. (a)(i) Let $d=\gcd(h,k)$, and define

$$P(z)=1+z+\cdots+z^{h-1}=(z^h-1)/(z-1),$$
$$Q(z)=1+z+\cdots+z^{k-1}=(z^k-1)/(z-1).$$

Define

$$R(z)=(z^k-1)/(z^d-1),$$
$$S(z)=(z^h-1)/(z^d-1).$$

Then none of P,Q,R,S is identically zero, but all have non-negative coefficients. By hypothesis (1) we have $P\circ\mathbf{f}\geqslant 0$. Hence $R\circ(P\circ\mathbf{f})\geqslant 0$. But by Lemma 2.1 this means

$$RP\circ\mathbf{f}\geqslant 0. \tag{4}$$

Similarly by hypothesis (2) we have $Q \circ (-\mathbf{g}) \geqslant 0$; hence

$$SQ \circ (-\mathbf{g}) = S \circ (Q \circ (-\mathbf{g})) \geqslant 0. \tag{5}$$

Note that $RP = SQ$, and RP is a polynomial of degree $h + k - d - 1$. Define the coefficients e_i by $R(z)P(z) = \sum_{0 \leqslant i < h+k-d} e_i z^i$. By (4) and (5) we have

$$\sum_{0 \leqslant i < h+k-d} e_i (f_i - g_i) \geqslant 0. \tag{6}$$

Now we claim that all the coefficients e_i are strictly positive. To see this, note that

$$\begin{aligned} R(z)P(z) &= \frac{z^h - 1}{z^d - 1} \frac{z^k - 1}{z - 1} \\ &= (1 + z^d + z^{2d} + \cdots + z^{h-d})(1 + z + z^2 + \cdots + z^{k-1}). \end{aligned}$$

If $i < h$, write $i = qd + r$ where $0 \leqslant r < d$, and choose the term z^{qd} from the left factor and z^r from the right factor to see $e_i > 0$. If $h \leqslant i < h + k - d$, choose z^{h-d} from the left factor and z^{i-h+d} from the right factor to see $e_i > 0$.

Since the e_i are all strictly positive, combining inequality (6) with hypothesis (3) that $f_n \leqslant g_n$ for $0 \leqslant n < h + k - d$ gives $f_n = g_n$ for $0 \leqslant n < h + k - d$. But then, by the Fine and Wilf theorem, $f_n = g_n$ for all $n \geqslant 0$. This proves (a)(i).

Next we prove (a)(ii). Since $f_n = g_n$ for all $n \geqslant 0$, it follows that f is periodic of period length h and k, and hence by Theorem 1.1, of period d. The sum over the terms of this period must be 0, since if it were less than 0 this would contradict hypothesis (1), while if it were greater than 0 this would contradict hypothesis (2).

Then $f_j + f_{j+1} + \cdots + f_{j+d-1}$ is just a cyclic permutation of $f_0 + f_1 + \cdots + f_{d-1}$, which equals 0. A similar argument applies to g.

We now turn to the proof of part (b). Actually, we provide two different proofs, one where inequality (3) is actually an equality for as long as possible and the terms are over an alphabet of minimal size, and one where inequality (3) is strict.

If z is a finite sequence, then by z^ω we mean the infinite sequence $zzz\cdots$.

We first prove:

Theorem 2.3. *Let h, k be integers with $0 < h \leqslant k$, $(h, k) \neq (1, 1)$. Define $d = \gcd(h, k)$. There exist finite sequences $v_{h,k}$ and $w_{h,k}$ over $\{-1, 0, 1\}$, each summing to 0, such that if $v_{h,k}^\omega = (f_0, f_1, f_2, \ldots)$ and $w_{h,k}^\omega = (g_0, g_1, g_2, \ldots)$, then*
(a) $f_i = g_i$ *for* $0 \leqslant i < h + k - d - 2$;
(b) $f_{h+k-d-2} < g_{h+k-d-2}$;
(c) $f_{h+k-d-1} > g_{h+k-d-1}$.

In order to prove this theorem, we use the so-called standard Sturmian words [1,11], which can be defined as follows for integers $0 \leqslant h \leqslant k$ with $\gcd(h, k) = 1$:

$$\sigma(h,k) := \begin{cases} 0 & \text{if } (h,k) = (0,1), \\ 0^{k-1}1 & \text{if } h = 1, \\ \sigma(r,h)^q \ \sigma(r',r) & \text{if } h > 1 \text{ and } k = qh + r, h = q'r + r'. \end{cases} \tag{7}$$

Actually, we need the following slight generalization of these words, which removes the restriction $\gcd(h,k)=1$:

$$\sigma'(h,k) := \begin{cases} 0^k & \text{if } h=0, \\ 0^{k-1}1 & \text{if } h|k, \\ \sigma'(r,h)^q\ \sigma'(r',r) & \text{if } h>1 \text{ and } k=qh+r, h=q'r+r'. \end{cases} \tag{8}$$

Define the morphism φ_t by $\varphi_t(0)=0^t$ and $\varphi_t(1)=0^{t-1}1$. Then it is easy to see that for $\gcd(h,k)=1$ and $t\geqslant 1$ we have $\sigma'(ht,kt)=\varphi_t(\sigma(h,k))$.

We also need some very basic facts about finite continued fractions, as found, for example, in [9]. A finite continued fraction is an expression of the form

$$u = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2+\cdots+\frac{1}{a_n}}},$$

where a_0 is an integer and the other a_i are positive integers. We use the well-known fact that every rational number u has a unique expansion as a continued fraction under the additional restriction that $a_n\neq 1$ if $u\neq 1$. We define the length of u, $\ell(u)$, to be n. Note that $\ell(u)=0$ if u is an integer and further that $\ell(u/v)=\ell((v \bmod u)/u)+1$ for integers $0<u<v$.

We need the following fact about $\sigma'(h,k)$, as found, for example, in [11]. (Strictly speaking this result was proved for $\sigma(h,k)$ and $\gcd(h,k)=1$, but the generalization to arbitrary h,k is straightforward.)

Theorem 2.4. *Let h,k be integers with $0<h<k$. Write $k=qh+r$ with $0\leqslant r<h$. Then $\sigma'(h,k)\in\{0,1\}^k$. Let* $\gcd(h,k)$. *Further, let $\sigma'(r,h)^\omega=(a_0,a_1,a_2,\ldots)$ and $\sigma'(h,k)^\omega=(b_0,b_1,b_2,\ldots)$. Then*
(a) *$a_i=b_i$ for $0\leqslant i<h+k-d-1$;*
(b) *if $\ell(h/k)$ is even, then $a_{h+k-d-1}=1$, $b_{h+k-d-1}=0$;*
(c) *if $\ell(h/k)$ is odd, then $a_{h+k-d-1}=0$, $b_{h+k-d-1}=1$.*

Given a finite sequence $x=(x_0,x_1,\ldots,x_{t-1})$ we define its cyclic first difference by $\Delta x=(x_1-x_0,x_2-x_1,\ldots,x_{t-1}-x_{t-2},x_0-x_{t-1})$.

We can now prove Theorem 2.3, thus completing the proof of Theorem 2.2.

Proof of Theorem 2.3. The case $h=k\geqslant 2$ is left to the reader. Assume $0<h<k$, and let $r=k \bmod h$. If $\ell(h/k)$ is even, define $v_{h,k}=\Delta\sigma'(h,k)$ and $w_{h,k}=\Delta\sigma'(r,h)$. If $\ell(h/k)$ is odd, define $v_{h,k}=\Delta\sigma'(r,h)$ and $w_{h,k}=\Delta\sigma'(h,k)$. By definition, the terms of any cyclic first difference sequence sum to 0. Parts (a) and (b) now follow immediately from Theorem 2.4. For part (c), note that if $f_i\leqslant g_i$ for $0\leqslant i<h+k-d$, then $f_i=g_i$ for all $i\geqslant 0$ by Theorem 2.2(a)(i). But then $\sigma'(h,k)^\infty$ and $\sigma'(r,h)^\infty$ differ by a constant, and since they are over the alphabet $\{0,1\}$, must be equal if $(h,k)\neq 1$. This, however, contradicts Theorem 2.4. □

Fig. 1 gives some examples of $v_{h,k}$ and $w_{h,k}$, where -1 is represented by $\bar{1}$.

h	k	$\sigma'(h,k)$	$v_{h,k}$	$w_{h,k}$
1	2	01	0	$1\bar{1}$
2	2	01	00	$1\bar{1}$
2	3	010	$1\bar{1}$	$1\bar{1}0$
2	4	0001	00	$001\bar{1}$
3	3	001	000	$01\bar{1}$
3	5	01001	$1\bar{1}0$	$1\bar{1}01\bar{1}$
4	6	000100	$001\bar{1}$	$001\bar{1}00$
5	8	01001010	$1\bar{1}01\bar{1}$	$1\bar{1}01\bar{1}1\bar{1}0$
5	12	010100101001	$1\bar{1}1\bar{1}0$	$1\bar{1}1\bar{1}01\bar{1}1\bar{1}01\bar{1}$
8	13	0100101001001	$1\bar{1}01\bar{1}1\bar{1}0$	$1\bar{1}01\bar{1}1\bar{1}01\bar{1}01\bar{1}$

Fig. 1. Some sample values of $v_{h,k}$ and $w_{h,k}$.

Example. Suppose $h=8$ and $k=13$. Then from Fig. 1 we have

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
f_n	1	−1	0	1	−1	1	−1	0	1	−1	0	1	−1	1	−1	0	1	−1	0	1
g_n	1	−1	0	1	−1	1	−1	0	1	−1	0	1	−1	1	−1	0	1	−1	1	−1

so we find $f_n \leqslant g_n$ for $0 \leqslant n < 19$.

There is another construction for proving Theorem 2.2(b), under the additional restriction that $h \neq k$. This construction gives strict inequality for $h+k-\gcd(h,k)-1$ consecutive terms. We describe it next.

Let A and B be two integers to be specified later, and define two codings τ_0 and τ_1 as follows: $\tau_0(0)=A$, $\tau_0(1)=B$, $\tau_1(0)=A+1$, $\tau_1(1)=B+1$.

Theorem 2.5. *There exist choices for A and B such that if $k=qh+r$, and $0<h<k$, then $f := \tau_0(\sigma'(r,h))$ and $g := \tau_1(\sigma'(h,k))$ both have periods that sum to 0 and satisfy $f_i < g_i$ for $0 \leqslant i < h+k-\gcd(h,k)-1$, but not at $i = h+k-\gcd(h,k)-1$.*

Proof. First, we observe that the number of 1's in $\sigma(h,k)$ is equal to u, where $hu/d \equiv (-1)^{\ell(h/k)+1} \pmod{k/d}$, $0 \leqslant u < k/d$, and $d=\gcd(h,k)$. See, for example, [1]. (Strictly speaking this was proved only for the case $d=1$ but the proof in the more general case is not difficult.) Similarly, the number of 1's in $\sigma'(r,h)$ is equal to t, where $th/d \equiv (-1)^{\ell(r/h)+1} \pmod{h/d}$ and $0 \leqslant t < h/d$. It follows that the sum of the period for $\tau_0(\sigma'(r,h))$ is $A(h-t)+Bt$, and the sum of the period for $\tau_1(\sigma'(h,k))$ is $(A+1)(k-u)+(B+1)u$. We want both these sums to be 0. Such A and B exist provided

$$\det \begin{bmatrix} h-t & t \\ k-u & u \end{bmatrix} \neq 0.$$

In other words, we want $(h-t)u-(k-u)t = hu-kt \neq 0$. It suffices to show that $(h/d)u-(k/d)t$ is non-zero. Consider this expression modulo k/d; we get $(h/d)u \,(\text{mod}\, k/d)$. But from above this is ± 1, and hence $\neq 0$ provided $k/d > 1$. But since $0<h<k$, we have $k=d$ if and only if $h=k$. Since $h<k$, the required A and B exist.

Now, by the construction, f and g have periods that sum to 0, and $f_i < g_i$ for $0 \leqslant i < h+k-\gcd(h,k)-1$, but $f_i > g_i$ for $i = h+k-\gcd(h,k)-1$. □

Example. Take $h=5$, $k=8$. Then we find $\sigma'(r,h) = 01001$ and $\sigma'(h,k) = 01001010$. Solving the linear system $3A+2B=0$, $5(A+1)+3(B+1)=0$, we find $A=-16$, $B=24$. Thus if f is the periodic sequence $(-16, 24, -16, -16, 24)^{\omega}$ and g is the periodic sequence $(-15, 25, -15, -15, 25, -15, 25, -15)^{\omega}$, we see from the following table:

n	0	1	2	3	4	5	6	7	8	9	10	11
f_n	−16	24	−16	−16	24	−16	24	−16	−16	24	−16	24
g_n	−15	25	−15	−15	25	−15	25	−15	−15	25	−15	−15

that $f_n < g_n$ for $0 \leqslant n \leqslant 10$, but not at $n = 11$.

Remark. Theorem 2.2 is reminiscent of some classical theorems on trigonometric polynomials. For example, Fejér [6] proved that a real trigonometric polynomial with 0 constant term

$$\lambda_1 \cos\theta + \mu_1 \sin\theta + \lambda_2 \cos(2\theta) + \mu_2 \sin(2\theta) + \cdots + \lambda_r \cos(r\theta) + \mu_r \sin(r\theta)$$

cannot have the same sign for all real θ unless it is identically zero. Also see [13, pp. 80, 263], [8].

An earlier version of this paper [12] contained a slightly different proof of Theorem 2.2.

3. Second variation: more than two periods

In this section, we consider some variations on the Fine and Wilf theorem for more than two periods. (For other generalizations of Fine and Wilf to more than two periods, see [3,10].)

For our first theorem, we need a little notation. For integers $p \geqslant 1$ let ω_p denote a primitive p'th root of unity, i.e., $\omega_p := e^{2\pi\sqrt{-1}/p}$. Define

$$R_p := \{\omega_p^i\colon\ 0 \leqslant i < p\} = \{\omega \in \mathbb{C}\colon\ \omega^p = 1\}.$$

Finally, for integers $h_1, h_2, \ldots, h_r \geqslant 1$ define

$$\gamma(h_1, h_2, \ldots, h_r) = |R_{h_1} \cup R_{h_2} \cup \cdots \cup R_{h_r}|,$$

the number of distinct roots of unity among the h_1th, h_2th, etc., roots of unity.

By the principle of inclusion–exclusion, it follows that

$$\gamma(h_1, h_2, \ldots, h_r) = \sum_{\substack{S \subseteq \{h_1, h_2, \ldots, h_r\} \\ S \neq \emptyset}} \gcd(S)(-1)^{|S|+1},$$

where by $\gcd(S)$ for S a non-empty set we mean the greatest common divisor of all elements of S. For example,

$$\gamma(6,10,15) = 6 + 10 + 15 - \gcd(6,10) - \gcd(6,15) - \gcd(10,15) + \gcd(6,10,15) = 22.$$

Theorem 3.1. *Let* $(f_i(n))_{n \geqslant 0}$, $1 \leqslant i \leqslant r$, *be* r *periodic complex-valued sequences with period lengths* $h_1, h_2, \ldots, h_r$, *respectively. Suppose* $\sum_{1 \leqslant i \leqslant r} f_i(n) = 0$ *for* $0 \leqslant n < \gamma(h_1, h_2, \ldots, h_r)$. *Then* $\sum_{1 \leqslant i \leqslant r} f_i(n) = 0$ *for all* $n \geqslant 0$.

Proof. As Fine and Wilf observed [7], any periodic complex-valued sequence $(f(n))_{n \geqslant 0}$ of period length p can be written in the form

$$f(n) = \sum_{0 \leqslant i < p} c_i \omega_p^{in}$$

for some coefficients $c_0, c_1, \ldots, c_{p-1}$.

It follows that there exist coefficients $c_{i,j}$, $1 \leqslant i \leqslant r$ and $0 \leqslant j < h_i$ such that

$$f_i(n) = \sum_{0 \leqslant j < h_i} c_{i,j} \omega_{h_i}^{jn}.$$

Define

$$\begin{aligned} s &= [s_1, s_2, \ldots, s_m] \\ &= [1, \omega_{h_1}, \omega_{h_1}^2, \ldots, \omega_{h_1}^{h_1 - 1}, 1, \omega_{h_2}, \omega_{h_2}^2, \ldots, \omega_{h_2}^{h_2-1}, \ldots, 1, \omega_{h_r}, \omega_{h_r}^2, \ldots, \omega_{h_r}^{h_r-1}], \end{aligned}$$

where $m = h_1 + h_2 + \cdots + h_r$. Let $B := \gamma(h_1, h_2, \ldots, h_r)$ and define the $B \times m$ matrix $M = (t_{i,j})_{0 \leqslant i < B, 1 \leqslant j \leqslant m}$ by $t_{i,j} := s_j^i$. Define the column vector

$$v := [c_{1,0}, c_{1,1}, \ldots, c_{1,h_1-1}, c_{2,0}, c_{2,1}, \ldots, c_{2,h_2-1}, \ldots, c_{r,0}, c_{r,1}, \ldots, c_{r,h_r-1}]^{\mathrm{T}}.$$

Then the hypothesis of the theorem is $Mv = 0$. Some of the columns of M are identical because some of the entries in the vector s coincide. We may delete the repeated columns of M and sum the corresponding entries of v to get $M'v' = 0$, where M' is a $B \times B$ matrix and v' is a column vector with B entries. Now M' is a Vandermonde matrix and hence invertible, so $v' = 0$. It follows that $\sum_{1 \leqslant i \leqslant r} f_i(n) = 0$ for all n. □

We next turn to another variation on Fine and Wilf for more than two periods. This generalization is more in the spirit of Theorem 2.2.

Theorem 3.2. *Let* $\mathbf{f}_1=(f_1(n))_{n\geqslant 0}$, $\mathbf{f}_2=(f_2(n))_{n\geqslant 0},\ldots,\mathbf{f}_r=(f_r(n))_{n\geqslant 0}$ *be* r *periodic real-valued sequences of periods* $h_1,h_2,\ldots,h_r$, *respectively. Suppose that for all* i *with* $1\leqslant i\leqslant r$, *we have*

$$\sum_{0\leqslant n<h_i} f_i(n) \geqslant 0.$$

If

$$\sum_{1\leqslant i\leqslant r} f_i(n) \leqslant 0$$

for $0\leqslant n<h_1+h_2+\cdots+h_r-r+1$, *then*

$$\sum_{1\leqslant i\leqslant r} f_i(n) = 0$$

for all $n\geqslant 0$.

Proof. The proof is very similar to the proof of Theorem 2.2, and we indicate only what needs to be changed. First, we need the following easy generalization of Lemma 2.1.

Lemma 3.3. *If* $P_1,P_2,\ldots,P_r$ *are polynomials with real coefficients and* $\mathbf{a}=(a_n)_{n\geqslant 0}$ *then* $P_1\circ(P_2\circ\cdots\circ(P_r\circ\mathbf{a})\cdots)=(P_1\cdots P_r)\circ\mathbf{a}$.

For $1\leqslant i\leqslant r$, define $P_i(z)=1+z+\cdots+z^{h_i-1}=(z^{h_i}-1)/(z-1)$. Then by hypothesis $P_i\circ\mathbf{f}_i$ is a sequence of non-negative real numbers for each i, $1\leqslant i\leqslant r$. It follows using Lemma 3.3 that if $P:=P_1P_2\cdots P_r$, then $P\circ\mathbf{f}_i$ is a sequence of non-negative real numbers for $1\leqslant i\leqslant r$. But P has degree $h_1+h_2+\cdots+h_r-r$ and hence has $h_1+h_2+\cdots+h_r-r+1$ coefficients. Furthermore, all the coefficients of P are strictly positive. Hence if $\sum_{1\leqslant i\leqslant r} f_i(n)\leqslant 0$ for $0\leqslant n<h_1+h_2+\cdots+h_r-r+1$, it follows that $\sum_{1\leqslant i\leqslant r} f_i(n)=0$ for $0\leqslant n<h_1+h_2+\cdots+h_r-r+1$. Now $h_1+h_2+\cdots+h_r-r+1\geqslant\gamma(h_1,h_2,\ldots,h_r)$, since the left-hand side counts the total number of roots of unity among $R_{h_1},\ldots,R_{h_r}$ without double-counting occurrences of 1, while the right-hand side counts the number of distinct roots of unity. But then $\sum_{1\leqslant i\leqslant r} f_i(n)=0$ for all $n\geqslant 0$ by Theorem 3.1. □

We note that the bound $h_1+h_2+\cdots+h_r-r+1$ is not, in general, optimal, although the bound is optimal if the period lengths $h_1,h_2,\ldots,h_r$ are relatively prime.

One might be tempted to guess that the true bound, as in Theorem 3.1, is not $h_1+h_2+\cdots+h_r-r+1$, but rather $\gamma(h_1,h_2,\ldots,h_r)$. This is not true, however. The following is an example of three periodic sequences of period lengths 6, 10, and 15, respectively, whose periods individually sum to 0 and such that $f_1(n)+f_2(n)+f_3(n)\leqslant 0$ for $0\leqslant n<\gamma(6,10,15)=22$, but not for $n=22$.

$$f_1=(0,0,0,-1,1,0)^{\omega},$$

$$f_2=(0,0,0,0,0,1,-1,0,-1,1)^{\omega},$$

$$f_3=(0,0,0,1,-1,-1,1,0,1,0,-1,0,0,0,0)^{\omega}.$$

4. The decreasing length conjecture

We now turn to a seemingly simple problem about iterated morphisms. Let Σ, Δ be finite alphabets. By a morphism we mean a map $h: \Sigma^* \to \Delta^*$ such that $h(xy)=h(x)h(y)$ for all $x, y \in \Sigma^*$. If $\Sigma = \Delta$ we can iterate h, writing $h^0(x)=x$ and $h^i(x)=h(h^{i-1}(x))$ for $i \geqslant 1$. We can then ask about the sequence of lengths

$$|x|,\ |h(x)|,\ |h^2(x)|, \ldots\ .$$

In particular, for how many consecutive terms can this sequence strictly decrease?

This question arose naturally during the writing of a paper of Shallit and Wang [16] on two-sided infinite fixed points of morphisms, i.e., those two-sided infinite words $\mathbf{w}$ such that $h(w)=\mathbf{w}$. Shallit and Wang made the following conjecture, called the decreasing length conjecture [15]:

Conjecture 4.1. If $h: \Sigma^* \to \Sigma^*$ is a morphism, and Σ has n elements, then

$$|w| > |h(w)| > \cdots > |h^k(w)|$$

implies that $k \leqslant n$.

It is easy to see that the bound of n cannot be decreased, for if we define

$$\Sigma = \{a_1, a_2, \ldots, a_n\},$$

$$w = a_1 a_2 \cdots a_n,$$

$$h(a_i) = a_{i+1} \quad \text{for } 1 \leqslant i \leqslant n-1,$$

$$h(a_n) = \varepsilon,$$

then $h^j(n) = a_{j+1} a_{j+2} \cdots a_n$ for $0 \leqslant j \leqslant n$.

The decreasing length conjecture can be stated in an equivalent fashion that does not involve morphisms. To do so, we recall some basic facts about iterated morphisms.

Let $|x|_a$ denote the number of occurrences of the letter a in the string x. Given a morphism $h: \Sigma^* \to \Sigma^*$ for $\Sigma = \{a_1, a_2, \ldots, a_d\}$, we define the *incidence matrix* $M = M(h)$ as follows:

$$M = (m_{i,j})_{1 \leqslant i,j \leqslant d},$$

where $m_{i,j} = |h(a_j)|_{a_i}$.

The matrix $M(h)$ is useful because of the following proposition.

Proposition 4.2. *We have*

$$\begin{bmatrix} |h(w)|_{a_1} \\ |h(w)|_{a_2} \\ \vdots \\ |h(w)|_{a_d} \end{bmatrix} = M(h) \begin{bmatrix} |w|_{a_1} \\ |w|_{a_2} \\ \vdots \\ |w|_{a_d} \end{bmatrix}.$$

Proof. We have

$$|h(w)|_{a_i} = \sum_{1 \leqslant j \leqslant d} |h(a_j)|_{a_i} |w|_{a_j}. \qquad \square$$

Corollary 4.3.

$$\begin{bmatrix} |h^n(w)|_{a_1} \\ |h^n(w)|_{a_2} \\ \vdots \\ |h^n(w)|_{a_d} \end{bmatrix} = (M(h))^n \begin{bmatrix} |w|_{a_1} \\ |w|_{a_2} \\ \vdots \\ |w|_{a_d} \end{bmatrix}.$$

Corollary 4.4.

$$|h^n(w)| = [1 \;\; 1 \;\; 1 \;\; \cdots \;\; 1] M(h)^n \begin{bmatrix} |w|_{a_1} \\ |w|_{a_2} \\ \vdots \\ |w|_{a_d} \end{bmatrix}.$$

Thus an equivalent way to state the decreasing length conjecture is the following:

Conjecture 4.5. Let M be an $n \times n$ matrix with non-negative integer entries. Let v be a column vector of non-negative integers, and let u be the row vector $[1 \;\; 1 \;\; 1 \;\; \cdots \;\; 1]$. If

$$uv > uMv > uM^2v > \cdots > uM^kv$$

then $k \leqslant n$.

Some partial results on the decreasing length conjecture were already known. If A and B are square matrices of the same dimension, then by $A \leqslant B$ we mean that each entry of A is $\leqslant$ the corresponding entry of B. Wang and Shallit proved [18] that if M is an $n \times n$ matrix of non-negative integers, then there exist integers $0 \leqslant i < j \leqslant 2^n$ such that $M^i \leqslant M^j$. It follows that if

$$uv > uMv > uM^2v > \cdots > uM^kv,$$

then $k < 2^n$. Later, improved results on this related problem were found by Bo [2] and Wang [17].

We begin with the following lemma.

Lemma 4.6. *Let $r \geqslant 1$ be an integer, and suppose there exist r sequences $\mathbf{a}_i = (a_i(n))_{n \geqslant 0}$, $1 \leqslant i \leqslant r$, r positive integers $h_1, h_2, \ldots, h_r$, and positive real numbers $w_{i,j}$, $1 \leqslant i \leqslant r$, $0 \leqslant j < h_i$ such that for $1 \leqslant i \leqslant r$ and all $n \geqslant 0$ we have $\sum_{0 \leqslant j < h_i} w_{i,j} a_i(n+j) \geqslant 0$. Let $B = h_1 + h_2 + \cdots + h_r - r + 1$. Then there exist B positive real numbers $x_0, x_1, \ldots, x_{B-1}$ such that for $1 \leqslant i \leqslant r$ and all $n \geqslant 0$ we have $\sum_{0 \leqslant j < B} x_j a_i(n+j) \geqslant 0$.*

Proof. For $1 \leqslant i \leqslant r$ define $P_i(z) := \sum_{0 \leqslant j < h_i} w_{i,j} z^j$. Then P_i is a polynomial of degree $h_i - 1$. Define $P := P_1 P_2 \cdots P_r$. Then P is a polynomial of degree $B - 1 = h_1 + h_2 + \cdots + h_r - r$. Define $P(z) = \sum_{0 \leqslant j < B} x_j z^j$. By hypothesis each P_i has all positive coefficients, and hence so does P. By hypothesis $P_i \circ \mathbf{a}_i$ consists of all non-negative terms for $1 \leqslant i \leqslant r$. Hence $(P/P_i) \circ (P_i \circ \mathbf{a}_i)$ consists of all non-negative terms, and by Lemma 3.3 this equals $P \circ \mathbf{a}_i$. □

We now prove a lemma similar to Theorem 3.2, with weaker hypotheses and a weaker conclusion.

Lemma 4.7. *Let $r \geqslant 1$ be an integer, and suppose there exist r sequences of real numbers $\mathbf{a}_i = (a_i(n))_{n \geqslant 0}$, $1 \leqslant i \leqslant r$, and r positive integers $h_1, h_2, \ldots, h_r$, such that the following conditions hold*:

(a) $\sum_{0 \leqslant j < h_i} a_i(n+j) \geqslant 0$ *for* $1 \leqslant i \leqslant r$ *and* $n \geqslant 0$;

(b) *There exists an integer* $C \geqslant 1$ *such that* $\sum_{1 \leqslant i \leqslant r} a_i(n) < 0$ *for* $0 \leqslant n < C$.

Then $C \leqslant h_1 + h_2 + \cdots + h_r - r$.

Proof. Assume, contrary to what we want to prove, that $C \geqslant h_1 + h_2 + \cdots + h_r - r + 1$. By Lemma 4.6 there exist positive real numbers $x_0, x_1, \ldots, x_{s-1}$ with $s = h_1 + h_2 + \cdots + h_r - r + 1$ such that $\sum_{0 \leqslant j < s} x_j a_i(j) \geqslant 0$ for $1 \leqslant i \leqslant r$. Then

$$\sum_{1 \leqslant i \leqslant r} \sum_{0 \leqslant j < s} x_j a_i(j) = \sum_{0 \leqslant j < s} x_j \sum_{1 \leqslant i \leqslant r} a_i(j) \geqslant 0,$$

a contradiction. Hence $n \leqslant h_1 + h_2 + \cdots + h_r - r$. □

Remark. When $r = 2$, and $\gcd(h_1, h_2) = 1$, the bound in Lemma 4.7 is tight, as shown by Theorem 2.5.

Lemma 4.8. *Let $r \geqslant 1$ be an integer, and suppose there exist r sequences of real numbers $\mathbf{b}_i = (b_i(n))_{n \geqslant 0}$, $1 \leqslant i \leqslant r$, and r positive integers $h_1, h_2, \ldots, h_r$, such that the following conditions hold*:

(a) $b_i(n + h_i) \geqslant b_i(n)$ *for* $1 \leqslant i \leqslant r$ *and* $n \geqslant 0$;

(b) *there exists an integer* $D \geqslant 1$ *such that* $\sum_{1 \leqslant i \leqslant r} b_i(n) > \sum_{1 \leqslant i \leqslant r} b_i(n+1)$ *for* $0 \leqslant n < D$.

Then $D \leqslant h_1 + h_2 + \cdots + h_r - r$.

Proof. Define $a_i(n) := b_i(n+1) - b_i(n)$.

We find

$$\begin{aligned}\sum_{0 \leqslant j < h_i} a_i(n+j) &= \sum_{0 \leqslant j < h_i} (b_i(n+1+j) - b_i(n+j)) \\ &= b_i(n+h_i) - b_i(n) \\ &\geqslant 0.\end{aligned}$$

Also,

$$\sum_{1\leqslant i\leqslant r} a_i(n) = \sum_{1\leqslant i\leqslant r} (b_i(n+1) - b_i(n))$$
$$= \left(\sum_{1\leqslant i\leqslant r} b_i(n+1)\right) - \left(\sum_{1\leqslant i\leqslant r} b_i(n)\right)$$
$$< 0$$

for $0\leqslant n<D$. Now apply Lemma 4.7. We find $C\leqslant h_1+h_2+\cdots+h_r-r$, so $D\leqslant h_1+h_2+\cdots+h_r-r$. $\square$

Remark. When $r=2$ and $\gcd(h_1,h_2)=1$, then it can be shown that the bound in Lemma 4.8 is tight, as follows: define $b_1(n):=\sum_{0\leqslant i<n} f_i$ and $b_2(n):=-\sum_{0\leqslant i<n} g_i$, where f and g are the periodic sequences in Theorem 2.5. For example, for $h_1=5$, $h_2=8$ we find

n	0	1	2	3	4	5	6	7	8	9	10	11	12
$b_1(n)$	0	−16	8	−8	−24	0	−16	8	−8	−24	0	−16	8
$b_2(n)$	0	15	−10	5	20	−5	10	−15	0	15	−10	5	20
$b_1(n)+b_2(n)$	0	−1	−2	−3	−4	−5	−6	−7	−8	−9	−10	−11	28

We are now ready to prove the decreasing length conjecture. We will prove a conjecture slightly more general than the version given as Conjecture 4.5 above.

Theorem 4.9. *Suppose M is an $n\times n$ matrix with non-negative integer entries. If there exist a row vector u and a column vector v with non-negative real entries such that*

$$uv > uMv > uM^2v > \cdots > uM^kv,$$

then $k\leqslant n$. Also $k=n$ only if $M^n=0$.

Proof. First we recall a fact about path algebra in graphs. Given an $n\times n$ matrix $M=(m_{i,j})_{1\leqslant i\leqslant n, 1\leqslant j\leqslant n}$ with non-negative integer entries, we may construct its associated directed graph $G=G(M)$ on vertices $\{1,2,\ldots,n\}$ as follows: we create $m_{i,j}$ distinct directed edges from vertex i to vertex j. (Note: this may well create self-loops and multiple edges.) Then the i,jth entry of M^s gives the number of distinct walks of length s from vertex i to vertex j in G. (A walk may repeat vertices and edges, and the length of the walk is the number of edges traversed.)

Now let M be the matrix in the statement of the theorem and G its associated graph. Let $u=(u_1,u_2,\ldots,u_n)$ and $v=(v_1,v_2,\ldots,v_n)^{\mathrm{T}}$. Let V be the set of vertices in G. Consider some maximal set of vertices forming disjoint cycles $\{C_1,C_2,\ldots,C_r\}$ in G. Then V can be written as the disjoint union

$$V = C_1\cup C_2\cup\cdots\cup C_r\cup W,$$

where W is the set of vertices which do not lie in any of the disjoint cycles; note that W may be empty. Then any directed walk in G of length $|W|$ or greater must intersect some cycle C_i, for otherwise the walk would contain a cycle disjoint from $C_1, C_2, \ldots, C_r$, a contradiction. Associate each walk of length $\geqslant |W|$ to the first cycle C_i it intersects. Define $P^s_{i,j,l}$ to be the number of directed walks of length s from vertex i to vertex j associated with cycle l. Also define

$$T^s_l := \sum_{1 \leqslant i,j \leqslant n} u_i v_j P^s_{ijl}.$$

Then for any $s \geqslant |W|$ we have

$$uM^s v = \sum_{1 \leqslant l \leqslant r} T^s_l. \tag{9}$$

Then

$$T^s_l \leqslant T^{s+|C_l|}_l,$$

since any walk of length s associated with cycle C_l can be extended to a walk of length $s + |C_l|$ by traversing the cycle C_l once. (This construction is a 1–1 map and it maps a walk associated with cycle C_l to another walk also associated with cycle C_l since C_l is the first cycle encountered by both walks.) From the inequality $uM^s v > uM^{s+1} v$ for $0 \leqslant s \leqslant k-1$ and Eq. (9) we have

$$\sum_{1 \leqslant l \leqslant r} T^s_l > \sum_{1 \leqslant l \leqslant r} T^{s+1}_l$$

for $|W| \leqslant s < k$.

For $1 \leqslant i \leqslant r$ and $j \geqslant 0$ define $b_i(j) = T_i^{|W|+j}$ and $h_i = |C_i|$. Then the conditions of Lemma 4.8 are satisfied. We conclude that $k - |W| \leqslant |C_1| + |C_2| + \cdots + |C_r| - r$. Moreover, $|C_1| + |C_2| + \cdots + |C_r| + |W| = |V| = n$ and so $k \leqslant n - r$.

Finally notice that $k = n$ implies that $r = 0$, so G is acyclic and $M^n = 0$. $\square$

Acknowledgements

S. Cautis and S. Yazdani solved the decreasing length conjecture after having heard about it at the talk of J. Shallit [15]. Previously, partial results had been obtained by Mignosi. The first part of this paper is based on an MFCS 2001 conference paper [12]. It is a delight to join forces in writing this paper.

We thank the referees for a very careful reading of our paper.

References

[1] J. Berstel, P. Séébold, Sturmian words. in: M. Lothaire (Ed.), Algebraic Combinatorics on Words. Cambridge University Press, Cambridge, 2002, pp. 45–110.

[2] Z. Bo, Improvements on inequalities for non-negative matrices, Austral. J. Combin. 21 (2000) 251–255.

[3] M.G. Castelli, F. Mignosi, A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words, Theoret. Comput. Sci. 218 (1999) 83–94.
[4] C. Choffrut, J. Karhumäki, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Springer, Berlin, 1997, pp. 329–438.
[5] M. Crochemore, W. Rytter, Text Algorithms, Oxford University Press, Oxford, 1994.
[6] L. Fejér, Sur les polynomes harmoniques quelconques, C. R. Acad. Sci. Paris 157 (1913) 506–509.
[7] N.J. Fine, H.S. Wilf, Uniqueness theorems for periodic functions, Proc. Amer. Math. Soc. 16 (1965) 109–114.
[8] A.D. Gilbert, C.J. Smyth, Zero-mean cosine polynomials which are non-negative for as long as possible, J. London Math. Soc. 62 (2000) 489–504.
[9] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, Oxford, 1985.
[10] J. Justin, On a paper by Castelli, Mignosi, Restivo, Theoret. Inform. Appl. 34 (2000) 373–377.
[11] D.E. Knuth, J. Morris, V. Pratt, Fast pattern matching in strings, SIAM J. Comput. 6 (1977) 323–350.
[12] F. Mignosi, J. Shallit, M.-w. Wang, Variations on a theorem of Fine & Wilf, in: J. Sgall, A. Pultr, P. Kolman (Eds.), Proc. MFCS 2001, Lecture Notes in Computer Science, Vol. 2136, Springer, Berlin, 2001, pp. 512–523.
[13] G. Pólya, G. Szegö, Problems and Theorems in Analysis II, Springer, Berlin, 1976.
[14] A. Salomaa, Formal Languages, Academic Press, New York, 1973.
[15] J. Shallit, Ten problems I can't solve, Talk for University of Waterloo Pure Mathematics Club, July 11, 2000. Slides available at http://www.math.uwaterloo.ca/∼shallit/talks.html.
[16] J. Shallit, M.-w. Wang, On two-sided infinite fixed points of morphisms, in: G. Ciobanu, G. Păun (Eds.), Proc. 12th Internat. Symp. on Fundamentals of Computation Theory (FCT '99), Lecture Notes in Computer Science, Vol. 1684, Springer, Berlin, 1999, pp. 488–499.
[17] M.-w. Wang, An inequality for non-negative matrices II, Linear Algebra Appl. 348 (2002) 259–264.
[18] M.-w. Wang, J. Shallit, An inequality for non-negative matrices, Linear Algebra Appl. 290 (1999) 135–144.