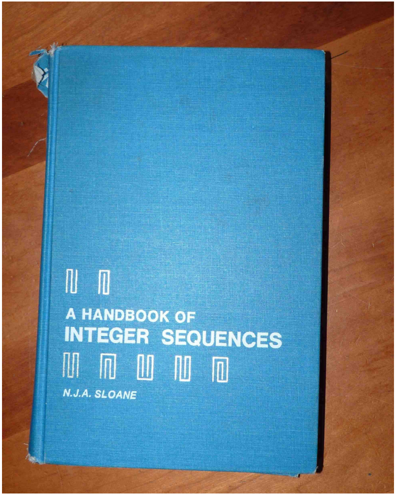
40 Years with Sloane's Integer Sequences

Jeffrey Shallit
School of Computer Science, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
shallit@cs.uwaterloo.ca
http://www.cs.uwaterloo.ca/~shallit

My Christmas Present in 1974



Sloane's First Letter to Me

Bell Laboratories

600 Mountain Avenue
Murray Hill, New Jersey 07974
Phone (201) 582-3000

December 31, 1974

Mr. Jeffrey Shallit
619 Greythorne Road
Wynnewood, Pennsylvania 19096

Dear Mr. Shallit:

Thank you for your very kind words about the book. That is the sort of letter an author likes to receive.

A copy of the first Supplement is enclosed. I deliberately omitted the continued fraction for e (for as a sequence it is very dull) but in response to public demand it will go into the next Supplement!

Best regards,

Neil Sloane

MH-1216-NJAS-mv

N. J. A. Sloane

Enc.
As above

Modified Engel expansions

Kalpazidou, Knopfmacher, and Knopfmacher (1990) studied expansions of the form

$$\alpha = a_0 + \frac{1}{a_1} - \frac{1}{a_1+1} \cdot \frac{1}{a_2} + \frac{1}{a_1+1} \cdot \frac{1}{a_2+1} \cdot \frac{1}{a_3} - \cdots$$

for real numbers α, where the a_i are non-negative integers with a_i positive for $i \geq 1$.

This expansion is essentially unique provided $a_{i+1} \geq a_i$.

I tried expanding $\alpha = 2/5$ and found doubly-exponential growth

$$(a_0, a_1, \ldots) = (0, 2, 3, 7, 13, 97, 193, 18817, 37633, 708158977, \ldots).$$

Although this particular sequence was not in the OEIS at the time, its even-indexed subsequence

$$(2, 7, 97, 18817, 708158977, \ldots)$$

certainly was (as A002812) and led to the discovery that ...

Modified Engel expansions

$$a_{2i+1} = \frac{1}{2}(2+\sqrt{3})^{2^i} + (2-\sqrt{3})^{2^i}$$

and

$$a_{2i+2} = (2+\sqrt{3})^{2^i} + (2-\sqrt{3})^{2^i} - 1.$$

From this it was easy to deduce the general form of the expansion for $2/(2r+1)$.

Compare Aho and Sloane's 1973 paper, "Some Doubly Exponential Sequences".

Open Problem: find a simple closed form for the expansion of $3/7$.

Some confusing quotients

Suppose we set $c_0 = a$, $c_1 = b$, and for $n \geq 2$ define c_n to be the smallest integer such that

$$\frac{c_n}{c_{n-1}} > \frac{c_{n-1}}{c_{n-2}}.$$

The resulting sequence $(c_n^{a,b})_{n\geq 0} = (c_n)_{n\geq 0}$ is often easy to describe:

If $a = 1$, $b = 2$, then $c_n = F_{2n+1}$, the $(2n+1)$'th Fibonacci number.

If $a = 2$, $b = 3$, then $c_n = 2^n + 1$.

But for other choices, some weird behavior results...

Some confusing quotients

For example, for $a = 8$ and $b = 55$, the resulting sequence

$$8, 55, 379, 2612, 18002, 124071, 855106, 5893451, 40618081, 279942687, \ldots$$

appears to satisfy the linear recurrence

$$c_n = 6c_{n-1} + 7c_{n-2} - 5c_{n-3} - 6c_{n-4}.$$

In fact, this is true up to c_{11055}, but fails for c_{11056} !

Some confusing quotients

The reason why was explained by David Boyd.

The reciprocal of the roots of the characteristic polynomial in the denominator are the real numbers

$$6.892\cdots, .95484560059\cdots$$

and two complex numbers with magnitude

$$.9548478767\cdots.$$

Note that the quotient of the the smaller real root with the magnitude of the imaginary roots is $1.00000238378\cdots$, very close to 1.

This is A010918 in the OEIS.

Open Problem: It is still not known exactly which of these sequences satisfy linear recurrences.

Counting monomials

Question:
how many distinct monomials are there in the (expanded) product

$$X_1(X_1+X_2)(X_1+X_2+X_3)\cdots(X_1+X_2+\cdots+X_n)$$

?

$$\begin{aligned} X_1 &: 1 \\ X_1(X_1+X_2) &: 2 \\ X_1(X_1+X_2)(X_1+X_2+X_3) &: 5 \\ X_1(X_1+X_2)(X_1+X_2+X_3)(X_1+X_2+X_3+X_4) &: 14 \end{aligned}$$

Answer: $C_n = \frac{\binom{2n}{n}}{n+1}$, the n'th Catalan number.

Lexicographically least sequences avoiding patterns

Consider generating a sequence avoiding "squares" (two consecutive identical blocks, like the word `hotshots`) in a greedy way, starting with 0.

We first write down 0.

Now we can't follow it with 0, because that would form 00, a square.

So the next symbol is 1, giving 01.

Now the next symbol can be 0, giving 010.

But the next one can't be 0 or 1, so choose 2, giving 0102.

Lexicographically least sequences avoiding patterns

It is easy to see, and not hard to prove, that the resulting infinite sequence over $\mathbb{N}$ is

$$01020103\cdots = (\nu_2(n))_{n\geq 1},$$

where $\nu_2(n)$ is the exponent of the largest power of 2 dividing n.

This is OEIS sequence A007814.

It is also the fixed point of the map that sends each integer i to the pair $(0, i+1)$.

Lexicographically least sequences avoiding patterns

Now change the problem very slightly. Instead of trying to avoid squares, try to avoid overlaps. An *overlap* is a word of the form $axaxa$, where a is a single letter and x is a (possibly empty) word, like the English word `alfalfa`.

The resulting sequence is $M =$

$$0010011001002001001100100210010020010011001002001001200100110010020010011001003 \cdots$$

which is OEIS A161371.

Lexicographically least sequences avoiding patterns

Mathieu Guay-Paquet and I studied this sequence and proved a number of properties of it:

For example,

- the first occurrence of $i = 0, 1, 2$ is at position $P_i = 1, 3, 13, 79, \ldots$ where $P_n = \lfloor 2^n \cdot n! \cdot \sqrt{e} \rfloor$. This is OEIS A010844. Once we knew the formula for P_n, we were able to prove it rather easily.

Lexicographically least sequences avoiding patterns

- M is the fixed point of the map where $0 \to 001$, $1 \to 1001002$, $2 \to 200100110010020010011001003$, etc., where the length of the image of i is $Q_i = 3, 7, 27, 159 \ldots$ This is OEIS A161370 and $Q_i = 2P_i + 1$.

Open problem: describe the lexicographically least sequence over $\mathbb{N}$ avoiding $\frac{5}{2}$-powers. Is it over the alphabet $\{0, 1, 2\}$?

Counting abelian squares

A word x is an *abelian square* if the first half is a permutation of the second half, like the English word `reappear`.

Consider counting the number $A_k(n)$ of length-$2n$ abelian squares over an alphabet of size k, say $\Sigma = \{1, 2, \ldots, k\}$

Suppose there are i 1's in the first half of the string. Choose the position of the 1's in the first and last halves of the string. This can be done in $\binom{n}{i}^2$ ways. Now fill in the remaining $n - 2i$ positions of the string with $k - 1$ symbols in $A_{k-1}(n - i)$ ways. Thus

$$\begin{aligned} A_k(n) &= \sum_{0 \le i \le n} \binom{n}{i}^2 A_{k-1}(n-i) = \sum_{0 \le i \le n} \binom{n}{n-i}^2 A_{k-1}(n-i) \\ &= \sum_{0 \le j \le n} \binom{n}{j}^2 A_{k-1}(j). \end{aligned}$$

Counting abelian squares

Since $A_1(j) = 1$ we immediately get

$$A_2(n) = \sum_{0 \le i \le n} \binom{n}{i}^2 .$$

From this we can find an easy proof of the famous identity

$$\sum_{0 \le i \le n} \binom{n}{i}^2 = \binom{2n}{n} :$$

Consider a string of length $2n$, and choose n positions in it. If a position falls in the first half of the string, make it 1; if a position falls in the last half of the string, make it 2. Of the remaining unchosen positions, make them 2 if they fall in the first half and 1 if they fall in the last half.

It is easy to see that this gives a bijection with the set of abelian squares.

k-regular sequences

Given a sequence $\mathbf{s} = (s(n))_{n\geq 0}$, we can divide it into its odd- and even-indexed subsequences $(s(2n))_{n\geq 0}$ and $(s(2n+1))_{n\geq 0}$, then dividing these sequences up again into *their* odd- and even-indexed subsequences

$$(s(4n))_{n\geq 0}, (s(4n+1))_{n\geq 0}, (s(4n+2))_{n\geq 0}, (s(4n+3))_{n\geq 0},$$

and so forth. The resulting set of sequences

$$D_2(\mathbf{s}) = \{(s(2^i n + j))_{n\geq 0} \ : \ i \geq 0 \ \text{ and } 0 \leq j < 2^i\}$$

is called the 2-*decimation* of $\mathbf{s}$.
In a similar fashion we can form the k-decimation of $\mathbf{s}$.
If the k-decimation of $\mathbf{s}$ is finitely generated (that is, there is a finite set $\mathcal{S}$ of sequences so that $D_k(\mathbf{s})$ is a subset of the $\mathbb{Z}$-span of $\mathcal{S}$), then $\mathbf{s}$ is called *k-regular*.

An example

Consider the divide-and-conquer linear recurrence counting the number of comparisons in mergesort:

$$T(n) = \begin{cases} 0, & \text{if } n = 0, 1; \\ T(\lfloor \frac{n}{2} \rfloor) + T(\lceil \frac{n}{2} \rceil) + n - 1, & \text{if } n > 1. \end{cases}$$

Then

$$\begin{aligned}
T(4n+2) &= -4T(n) + T(2n) + 3T(2n+1) \\
T(4n+3) &= -8T(n) + 2T(2n) + 6T(2n+1) - T(4n+1) \\
T(8n) &= 4T(n) - 7T(2n) - T(2n+1) + 4T(4n) + T(4n+1) \\
T(8n+1) &= 4T(n) - 6T(2n) - 2T(2n+1) + 2T(4n) + 3T(4n+1) \\
T(8n+4) &= -12T(n) + T(2n) + 7T(2n+1) + T(4n+1) \\
T(8n+5) &= -20T(n) + 4T(2n) + 12T(2n+1) - T(4n+1)
\end{aligned}$$

An example

So any sequence in the 2-decimation of $\mathbf{T}$ is a linear combination of the subsequences $(T(n))_{n\geq 0}$, $(T(2n))_{n\geq 0}$, $(T(2n+1))_{n\geq 0}$, $(T(4n))_{n\geq 0}$, and $(T(4n+1))_{n\geq 0}$.

There is a "closed form" for $\mathbf{T}$, as follows:

$$T(n) = n\lceil \log_2 n \rceil - 2^{\lceil \log_2 n \rceil} + 1.$$

k-regular sequences in the OEIS

Some more examples:

The Mallows sequence: the unique monotone sequence $(a(n))_{n\geq 0}$ of non-negative integers such that $a(a(n)) = 2n$ for $n \neq 1$. It is A007378 and is 2-regular.

The Propp sequence: the unique monotone sequence $(b(n))_{n\geq 0}$ of non-negative integers such that $a(a(n)) = 3n$ for $n \geq 0$. It is A007378 and is 3-regular.

We showed that the lexicographically least monotone solution to $a(a(n)) = dn$ is d-regular.

Number of binary overlap-free words of length n: Cassaigne showed that this sequence is 2-regular.

Strauss's sequence

Define

$$r(n) = \sum_{0 \le i < n} \binom{2i}{i},$$

and let $f(n) = \nu_3(r(n+1))$ be the exponent of the highest power of 3 dividing $r(n+1)$. The following table gives the first few values of f:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$f(n)$	0	1	2	0	2	3	1	2	4	0	1	2	0	3	4

Strauss's sequence

A 3-regular sequence recognizer easily produces the following conjectured relations:

$$
\begin{aligned}
f(3n+2) &= f(n)+2; \\
f(9n) &= f(9n+3) = f(3n); \\
f(9n+1) &= f(3n)+1; \\
f(9n+4) &= f(9n+7) = f(3n+1)+1; \\
f(9n+6) &= f(3n+1).
\end{aligned}
$$

With a little more work, one arrives at the conjecture

$$\nu_3(r(n)) = \nu_3(n^2\binom{2n}{n}),$$

which we proved in 1989.

A beautiful proof of this identity using 3-adic analysis was later given by Don Zagier. Zagier showed that if we set

$$F(n) = \frac{\sum_{0 \le k < n} \binom{2k}{k}}{n^2 \binom{2n}{n}},$$

then $F(n)$ extends to a 3-adic analytic function from $\mathbb{Z}_3$ to $-1 + 3\mathbb{Z}_3$, and can be evaluated at the negative integers as follows:

$$F(-n) = -\frac{(2n-1)!}{(n!)^2} \sum_{0 \le k < n} \frac{(k!)^2}{(k-1)!}$$

for $n \geq 0$.

Strauss's sequence

A heuristic k-regular sequence recognizer can produce many interesting conjectures. For example, let

$$a(n) = \sum_{0 \le k \le n} \binom{n}{k} \binom{n+k}{k}.$$

Let $b(n) = \nu_3(a(n))$. Then computer experiments strongly suggest:

$$b(n) = \begin{cases} b(\lfloor n/3 \rfloor) + (\lfloor n/3 \rfloor \bmod 2), & \text{if } n \equiv 0, 2 \pmod 3; \\ b(\lfloor n/9 \rfloor) + 1, & \text{if } n \equiv 1 \pmod 3. \end{cases}$$

Open problem: prove or disprove these relations. (verified for $0 \le n \le 10,000$).

k-regular sequences in the OEIS

It turns out that k-regular sequences form a large class of commonly-occurring sequences that are usually easy to recognize.

Back in the early 90's, when Allouche and I were first looking at this class of sequences, Neil Sloane provided an electronic copy of the EIS and we tested it. Out of about 5000 sequences at the time, approximately 300 (or 6%) of them were k-regular or "probably" k-regular.

It would be nice to add a k-regular recognizer to the OEIS.

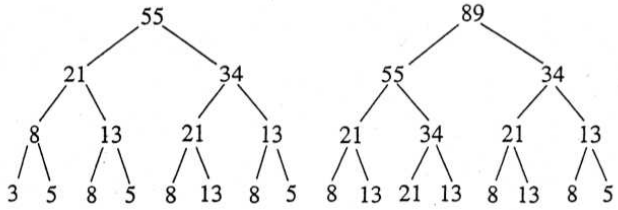
Nørgård's rhythmic infinity series

Start with two consecutive Fibonacci numbers, such as 55 and 89.

Replace each one with the previous two Fibonacci numbers, switching order if position is even.

Keep going until 3 is reached.

The Danish composer Per Nørgård used the resulting sequence to design a series of rhythms in his musical compositions.



Nørgård's rhythmic infinity series

Write out the indices instead of the Fibonacci numbers:

$$4, 5, 6, 5, 6, 7, 6, 5, 6, 7, 8, 7, 6, 7, 6, 5, \dots$$

Subtract 4 from each number to get:

$$0, 1, 2, 1, 2, 3, 2, 1, 2, 3, 4, 3, 2, 3, \dots$$

which the OEIS gives immediately as A005811: number of 1's in Gray code for n.

Counting squares in finite Fibonacci words

The finite Fibonacci words are given by A061107: $X_0 = 0$, $X_1 = 1$, and $X_n = X_{n-1}X_{n-2}$. Here are the first few:

$$0, 1, 10, 101, 10110, 10110101, \ldots$$

Fraenkel and Simpson (1999) came up with an exact formula for $R(n)$, the number of occurrences of squares in the Fibonacci word X_n:

$$R(n) = \frac{4}{5}nF(n+1) - \frac{2}{5}(n+6)F(n) - 4F(n-1) + n + 1,$$

for $n \geq 3$, where $F(n)$ is the n'th Fibonacci number. (Small error in their paper corrected here.) This is A248425, just added!

A decision procedure

Recently we implemented a procedure that can automatically compute formulas of this sort.

Here is a sketch of the idea.

We start with an infinite word generated by iterated morphism, like the infinite Fibonacci word $\mathbf{f} = 01001010\cdots$, which is a fixed point of the map sending

$$0 \to 01; \quad 1 \to 0.$$

Next, we associate a numeration system with the word. In this case the proper numeration system is the *Zeckendorf* or *Fibonacci* numeration system: every positive integer can be written uniquely in the form $\sum_{i \geq 2} \epsilon_i F(i)$, where $\epsilon_i \in \{0, 1\}$ and $\epsilon_i \epsilon_{i+1} = 0$ for all i.

A decision procedure

Now we need two finite automata that read inputs expressed in this numeration system: one that computes $\mathbf{f}[n]$ and one that adds two Fibonacci representations.

Given all this, we can construct a decision procedure that can decide all first-order queries about the sequence (in this case, $\mathbf{f}$) using operations like $+$, $<$, $=$, and indexing into the sequence.

Even more, we can find a *linear representation* for the number of n satisfying some predicate.

Here, by a *linear representation* we mean square matrices M_0, M_1 and vectors u, v such that

$$f(n) = uM_{a_1}M_{a_2}\cdots M_{a_i}v,$$

where $a_1a_2\cdots a_i$ is the Fibonacci representation of n.

A decision procedure

For example, here is the predicate that there exists a square of length $2m$ beginning at position i in the Fibonacci word:

$$\forall j < m \ \mathbf{f}[i+j] = \mathbf{f}[i+m+j].$$

We can also write a predicate for the assertion that the block $\mathbf{f}[i..i+2m-1]$ never occurred previously:

$$\forall \ell (\forall k < 2m \ \mathbf{f}[i+k] = \mathbf{f}[\ell+k]) \implies \ell \geq i.$$

Finally, we can also write a predicate for the assertion that the square occurs entirely within a prefix of length n of the Fibonacci word: $i + 2m < n$.

A decision procedure

From the automaton we can compute a linear representation.

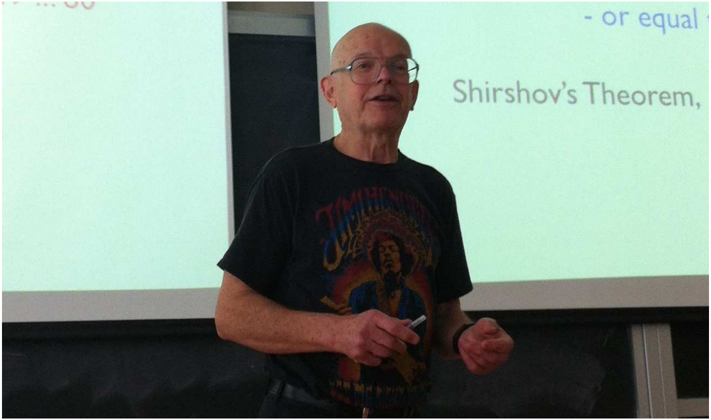
From the linear representation we can compute the minimal polynomial of the matrix.

From the minimal polynomial we can get the roots.

From the fundamental theorem of linear recurrences with constant coefficients we can get an expression in terms of the roots.

This gives a formula like the one above, in a purely mechanical fashion!

Long live Neil Sloane!



Neil Sloane at the Fields Institute, Toronto, April 2013

And long live the Encyclopedia of Integer Sequences!

For Further Reading

J.-P. Allouche and J. Shallit, The ring of k-regular sequences, *Theoret. Comput. Sci.* **98** (1992), 163–197.

David Boyd, Linear recurrence relations for some generalized Pisot sequences, in *Advances in Number Theory*, Oxford, 1993, pp. 333–340.

J. Shallit, Rational numbers with non-terminating, non-periodic modified Engel-type expansions, *Fibonacci Quart.* **31** (1993), 37–40.

Mathieu Guay-Paquet and Jeffrey Shallit, Avoiding squares and overlaps over the natural numbers, *Discrete Math.* **309** (2009), 6245–6254.

L. B. Richmond and J. Shallit, Counting abelian squares, *Electronic J. Combinatorics* **16** (1) (2009), Paper #R72.

C. F. Du, H. Mousavi, L. Schaeffer, and J. Shallit, Decision algorithms for Fibonacci-automatic words, with applications to pattern avoidance, preprint, June 2014, `http://arxiv.org/abs/1406.0670`.