Pairwise-nonrecurrent sequences

Thomas R. Bewley
Dept of MAE, UC San Diego
La Jolla, CA 92093-0411, USA
bewley@ucsd.edu

Abstract

A new class of integer sequences is introduced in which any pair of identical symbols in the sequence, separated a certain number of digits apart, appears in the sequence, with that separation, at most once. For any given sequence length n and number of distinct symbols used, m, such a sequence, if it exists, is not unique; our attention is thus focused on the lexographically first such sequence available in each case. Three variations of this class of sequence are considered: infinite, finite, and periodic, and two open questions in Ramsey theory are identified. Application of such sequences to the matrix coloring problem is also explored briefly.

1 Introduction

Consider the following new class of sequences:

Definition: A **pairwise-nonrecurrent sequence** is a sequence of symbols (taken in this work to be the non-negative integers) in which any pair of identical symbols, separated a certain number of digits apart, appears in the sequence, with that separation, at most once.

This paper explores and extends this class of sequences, focusing in particular, in each case, on the *leading* pairwise-nonrecurrent sequence available—that is, lexographically (i.e., in dictionary order), on the *first* pairwise-nonrecurrent sequence available for that value of m (the number of distinct symbols used) and n (the length of the sequence). We also identify an immediate application of such sequences to the matrix coloring problem; other potentially fruitful areas for application of such sequences include cryptography and coding theory.

1.1 Background: some related classes of integer sequences

The existing integer sequences that are most closely related to the pairwise-nonrecurrent sequences defined and explored in this work are Skolem and Langford sequences (see Skolem 1957 and Langford 1958, respectively), which have received a lot of attention in the literature; see, e.g., Colbourn and Dinitz (1996; page 457) and Shalaby & Stuckless (2000), and the references contained therein.

A Skolem sequence of order n, denoted $S = \{s_1, s_2, \ldots, s_{2n}\}$, is a permutation of the $2n$ integers $\{1,1,2,2,\ldots,n,n\}$ such that, if $s_i = s_j = k$ with $i < j$, then $j - i = k$; an example with $n = 5$ is given by $\{4,5,1,1,4,3,5,2,3,2\}$.

As a slight generalization, a Langford sequence of order n and defect d, denoted $L = \{l_1, l_2, \ldots, l_{2n}\}$, is a permutation of the $2n$ integers $\{d,d,d+1,d+1,\ldots,n+d-1,n+d-1\}$ such that, again, if $l_i = l_j = k$ with $i < j$, then $j - i = k$; an example with $n = 5$ and $d = 3$ is given by $\{7,5,3,6,4,3,5,7,4,6\}$. The Skolem sequence as defined above is a Langford sequence with $d = 1$. Note that the term "Langford sequence" is sometimes, less generally, used to denote a Langford sequence, as defined here, with $d = 2$.

A related concept is known as a Golomb ruler (Sidon 1932 and Babcock 1953). An n'th-order Golomb ruler is a set of n marks at integer positions along an imaginary ruler such that no two pairs of marks are the same distance apart. If a Golomb ruler is able to measure all integer distances up to its length, it is said to be perfect. A Golomb ruler is said to be optimal if no shorter Golomb ruler of the same order exists. An example of a Golomb ruler of order 4 that is both optimal and perfect is $\{0,1,4,6\}$; an optimal Golomb ruler of order 8 is $\{0,1,4,9,15,22,32,34\}$.

1.2 The leading pairwise-nonrecurrent infinite sequence

We consider first the leading pairwise-nonrecurrent *infinite* sequence, the first 150 terms of which, composed of 15 distinct symbols, are:

$$\begin{aligned}&0,0,1,0,1,1,2,0,2,1,3,2,0,3,3,1,4,2,2,3,0,3,4,1,4,4,5,5,6,4,0,2,1,3,5,2,5,6,6,4,5,6,7,3,\\&0,7,7,1,8,6,2,4,5,7,8,5,6,8,3,7,9,8,8,1,9,0,2,4,7,9,9,3,5,6,10,10,11,8,6,8,0,9,10,1,7,4,\\&7,2,9,3,11,10,11,10,5,11,0,10,6,5,12,9,11,8,4,3,12,12,13,1,11,12,6,2,7,4,13,9,10,11,12,\\&10,0,11,8,13,13,14,12,5,12,3,13,7,14,4,8,13,1,9,14,2,6,7,14,14,12,0,3,2,\ldots\end{aligned} \tag{1}$$

Note that the pairs $\{0,0\}$, $\{1,1\}$, $\{2,2\}$, $\{3,3\}$, etc., appear at most once in this sequence; denoting $*$ as any symbol, the pairs $\{0,*,0\}$, $\{0,*,*,0\}$, $\{0,*,*,*,0\}$, $\{1,*,1\}$, $\{1,*,*,1\}$, etc., appear at most once as well.

The unique sequence given in (1) is easy to generate numerically, and can be built from left to right, at each step adding to the existing pairwise-nonrecurrent finite sequence the lexographically smallest symbol that maintains the pairwise-nonrecurrence property in the extended sequence. A very simple single-threaded code, available at

`http://renaissance.ucsd.edu/pubs/PNinfinite.f90`,

generates the first 1,000 terms of this sequence (composed of 48 distinct symbols) in 22 seconds on a 3GHz Intel Xeon desktop computer, and generates the first 10,000 terms of this sequence (composed of 193 distinct symbols) in 39 hours. These first 10,000 terms of the leading pairwise-nonrecurrent infinite sequence so generated are available at:

`http://renaissance.ucsd.edu/pubs/PNinfinite.txt`

1.3 Leading pairwise-nonrecurrent finite sequences

A finite pairwise-nonrecurrent sequence composed of a given number of distinct symbols m may be generated by simple truncation of the sequence given in (1) immediately before the first symbol outside the set of symbols under consideration. However, as m is increased, much longer pairwise-nonrecurrent finite sequences using the same number of distinct symbols may, in fact, be generated, though finding such sequences is computationally expensive. The logic involved in the maximally efficient generation of such sequences for increasing values of m is quite intricate; a brief outline of the two algorithms developed in this work to identify such sequences is given below.

1.3.1 Algorithm A

We start by developing a maximally efficient exhaustive search in lexographic order. There are two ways to proceed: either specifying the number of each symbol to be used in the sequence in advance, or allowing the number of each symbol used to be determined by the search algorithm itself. We follow here the former approach; this has the numerical benefit of examining considerably fewer possibilities when searching for the solution. The algorithm developed applies the conditions of pairwise nonrecurrence from left to right to maximally accelerate the efficient nonrecursive algorithm for permuting symbols given as Algorithm L of Knuth (2005), §7.2.1.2, eliminating from consideration large sets of cases at a time wherever possible. The resulting exhaustive search algorithm (for both this problem and that discussed in §1.4) is available at:

`http://renaissance.ucsd.edu/pubs/PNfiniteA.f90`

1.3.2 Algorithm B

An alternative approach for computing finite pairwise-nonrecurrent sequences splits the problem in half: in cases for which n is an even multiple of m, we first determine *all* groups of $r = n/m$ locations in a sequence of length n such that a single symbol may be put into all r of these locations without violating the conditions of pairwise-nonrecurrence, then attempt to fit m of these groups together in a nonoverlapping fashion. For example, taking $n = 24$, $m = 4$, and

$r = 6$, there are $g = 1910$ such groups of locations, four of which are:

$$\mathbf{x},\mathbf{x},*,*,*,*,*,*,\mathbf{x},*,*,\mathbf{x},*,\mathbf{x},*,*,*,\mathbf{x},*,*,*,*,*,*;$$
$$*,*,\mathbf{x},*,\mathbf{x},*,*,*,*,*,*,*,\mathbf{x},*,*,\mathbf{x},\mathbf{x},*,*,*,*,\mathbf{x},*,*;$$
$$*,*,*,\mathbf{x},*,\mathbf{x},*,*,*,\mathbf{x},\mathbf{x},*,*,*,*,*,*,*,*,\mathbf{x},*,*,\mathbf{x},*;$$
$$*,*,*,*,*,*,\mathbf{x},\mathbf{x},*,*,*,*,*,*,\mathbf{x},*,*,*,\mathbf{x},*,\mathbf{x},*,*,\mathbf{x}.$$

We may fit together these $m = 4$ sets of $r = 6$ locations by assigning $\mathbf{x} = 0$ to all slots in the first set of locations, $\mathbf{x} = 1$ to all slots in the second set of locations, etc., resulting in a sequence of length $n = m * r = 24$ given by:

$$0,0,1,2,1,2,3,3,0,2,2,0,1,0,3,1,1,0,3,2,3,1,2,3.$$

Implementation of this algorithm is given in:
http://renaissance.ucsd.edu/pubs/PNfiniteBeven.f90

This approach doesn't necessary find sequences in lexographic order, and thus must be run exhaustively in order to find the first such sequence in lexographic order. However, for a given value of m, there are ultimately a decreasing number of pairwise nonrecurrent groups as n gets large; for example, taking $m = 4$, there are:

- $g = 2520$ pairwise nonrecurrent groupings of $r = n/m = 5$ symbols for $n = 20$,
- $g = 1910$ pairwise nonrecurrent groupings of $r = n/m = 6$ symbols for $n = 24$,
- $g = 98$ pairwise nonrecurrent groupings of $r = n/m = 7$ symbols for $n = 28$, and
- $g = 0$ pairwise nonrecurrent groupings of $r = n/m = 8$ symbols for $n = 32$

As g descreases, the problem of finding all possible ways of fitting the groupings together without overlap becomes computationally easier to solve.

In cases for which n is not an even multiple of m, we proceed by rounding $r = n/m$ both up and down, obtaining r_1 and $r_2 = r_1 - 1$, respectively, then find all pairwise nonrecurrent groups of both r_1 and r_2 locations in a sequence of length n; we then attempt to fit together m_1 groups of length r_1 with m_2 groups of length r_2 in a nonoverlapping fashion, where $n = m_1 r_1 + m_2 r_2$. [We note that it's possible, though perhaps somewhat unlikely, that some symbol may appear $r_1 + 1$ or $r_1 - 2$ times in a pairwise nonrecurrent sequence with the largest value of n possible for a given value of m; we do not consider this possibility in the present work, focusing instead on more "balanced" pairwise nonrecurrent sequence built solely from groups of length r_1 and r_2.] This approach is a straightforward generalization of the approach described above; its implementation is available here:
http://renaissance.ucsd.edu/pubs/PNfiniteBuneven.f90.

1.3.3 Results

Six examples of leading pairwise-nonrecurrent finite sequences generated using Algorithms A and B, with $m = \{2,3,4,5,6,7\}$ and length $n = \{7,14,24,35,49,62\}$, respectively, are:

$$0,0,1,1,0,1,0; \tag{2a}$$
$$0,0,1,1,2,2,1,0,2,0,2,1,0,1; \tag{2b}$$
$$0,0,1,2,1,2,3,3,0,2,2,0,1,0,3,1,1,0,3,2,3,1,2,3; \tag{2c}$$
$$0,0,1,1,2,2,3,4,1,4,3,2,3,4,4,2,0,3,0,1,3,1,0,4,2,0,4,2,1,2,0,1,3,3,4; \tag{2d}$$
$$0,0,1,1,2,3,4,5,2,1,0,5,5,1,4,3,4,3,5,4,4,3,3,2,0,1,5,0,1,2,3,4,0,1,2,5,2,2,5,0,4,3,1,0,3,0,2,4,5; \tag{2e}$$
$$\left.\begin{array}{l}0,0,1,0,2,3,2,0,4,3,5,5,0,1,2,4,1,6,4,6,0,5,1,1,6,3,5,3,2,5,4,2,6,6,3,0,6,3,4,4,2,1,3,5,4,1,2,2,\\ \quad 3,0,5,2,5,1,6,4,5,4,1,0,6,3.\end{array}\right\} \tag{2f}$$

By applying the conditions of pairwise nonrecurrence in a maximally efficient fashion, as described in §1.3.1, Algorithm A needed to check only 6, 406, 1.00×10^6, and 1.08×10^{10} cases, respectively, in order to find the sequences listed in (2a)-(2d). Regardless of the significant care that has been taken in accelerating Algorithm A, as its search is exhaustive, the scaling of the problem difficulty with increasing m is seen to be poor, and cases with higher values of

m are exceedingly difficult to determine following this approach, and might take several years of computation time to complete. Algorithm B, on the other hand, scales to larger values of m and n much more efficiently, and was used to find the sequences listed in (2e)-(2f).

Exhaustive searches using Algorithms A and B readily establish that (2a)-(2e) are in fact the longest pairwise-nonrecurrent finite sequences possible[1] for their corresponding values of m. Note that the maximum length of these sequences are

- $n = m_1 r_1 + m_2 r_2$ where $\{m_1, r_1\} = \{m-1, m+2\}$ and $\{m_2, r_2\} = \{1, m+1\}$ for $m = 2$ and 3,
- $n = m\,r$ where $r = m+2$ for $m = 4$ and 5, and
- $n = m_1 r_1 + m_2 r_2$ where $\{m_1, r_1\} = \{1, m+3\}$ and $\{m_2, r_2\} = \{m-1, m+2\}$ for $m = 6$.

A natural question remains open:
• What is the maximum length n of a pairwise-nonrecurrent finite sequence composed of $m > 6$ distinct symbols?
If this question can not be answered directly, what are the sharpest possible upper and lower bounds on its value?

1.4 Leading pairwise-nonrecurrent periodic sequences

An important subset of the class pairwise-nonrecurrent finite sequences considered in §1.3 may be identified by requiring that the conditions of pairwise nonrecurrence apply not only over the entire sequence iteself, but also over its periodic connection. That is, the sequence is assumed to be connected in a ring, and we now require that any pair of identical symbols, separated a certain number of digits apart, appears *on the ring*, with that separation, at most once.

Seven examples of leading pairwise-nonrecurrent periodic sequences generated using Algorithms A and B, with $m = \{2,3,4,5,6,7,8\}$ and of length $n = \{4,9,16,25,36,49,62\}$, respectively, are:

$$0,0,1,1; \tag{3a}$$
$$0,0,1,0,1,2,2,1,2; \tag{3b}$$
$$0,0,1,0,1,1,2,0,2,3,1,3,2,2,3,3; \tag{3c}$$
$$0,0,1,0,1,1,2,0,2,1,3,2,2,4,4,0,3,4,3,3,2,1,4,3,4; \tag{3d}$$
$$0,0,1,0,1,1,2,0,2,1,3,2,2,3,4,4,5,5,4,5,1,0,2,3,5,4,0,4,5,2,3,3,1,4,5,3; \tag{3e}$$
$$0,0,1,0,1,2,3,4,3,1,5,4,2,2,3,2,4,6,3,1,1,3,4,1,6,6,5,0,5,4,4,0,4,5,5,6,0,5,6,2,6,3,0,2,6,1,3,5,2; \tag{3f}$$
$$\left.\begin{array}{l} 0,0,1,0,1,1,2,2,0,3,2,4,3,5,0,2,6,6,0,4,5,6,4,6,3,7,4,2,1,7,0,4,4,3,3,6,5,1,3,0,3,7,5,2,5,2,7,5,\\ \quad 4,1,4,3,6,1,7,6,5,5,2,1,7,7. \end{array}\right\} \tag{3g}$$

By applying the conditions of pairwise nonrecurrence in a maximally efficient fashion, Algorithm A needed to check only 1, 4, 28, 1404, and 1.02×10^{11} cases, respectively, in order to find the sequences listed in (3a)-(3e). Again, as this search is exhaustive, the scaling of the problem difficulty with increasing m is poor. Algorithm B was thus used to find the sequences listed in (3f)-(3g).

Exhaustive searches using Algorithms A and B readily establish that (3a)-(3g) are in fact the longest pairwise-nonrecurrent periodic sequences possible for their corresponding values of m. Note that the maximum length of these sequences are

- $n = m\,r$ where $r = m$ for $m = 2$ through 7, and
- $n = m_1 r_1 + m_2 r_2$ where $\{m_1, r_1\} = \{m-2, m\}$ and $\{m_2, r_2\} = \{2, m-1\}$ for $m = 8$.

The sequences in (3) are each assumed to be periodically connected (in a ring), whereas the sequences in (2) are not; the conditions of pairwise-nonrecurrence are more restrictive in the periodic case than they are in the nonperiodic case, and thus the sequences in (3) are each a bit shorter than the corresponding sequence in (2).

As in §1.3, a natural question remains open:

[1]Regarding (2f), it is likely that slightly longer pairwise-nonrecurrent finite sequences exist for $m = 7$; this case has not yet been run exhaustively. An updated list of the longest pairwise-nonrecurrent finite sequences found thus far will therefore be maintained at: http://renaissance.ucsd.edu/PairwiseNonrecurrent.html

• What is the maximum length n of a pairwise-nonrecurrent *periodic* sequence composed of $m > 8$ distinct symbols? If this question can not be answered directly, what are the sharpest possible upper and lower bounds on its value?

Curiously, though their definitions are quite simple, no simple patterns have yet been detected in the infinite sequence given in (1), the finite (nonperiodic) sequences given in (2), or the periodic sequences given in (3). Recognizing such patterns would, of course, be immensely valuable in constructing longer pairwise-nonrecurrent sequences, as exhaustively searching for such sequences is numerically prohibitive as m and n are increased, even if the code is written in a maximally efficient manner. The lack of such patterns might lend these sequences well to applications in cryptography.

2 Efficient generation of m-colored matrices

The papers of Cooper, Fenner, and Purewal (2008) and Fenner, Gasarch, Glover, and Purewal (2009) introduce, and put into context with Ramsey theory, the following notion:

Definition: A $p \times q$ matrix A is said to be m**-colored** if each element a_{ik} of the matrix A is one of m symbols [taken here to be the numbers 0 through $(m-1)$] such that there is no set of four integers $\{i, j, k, l\}$ with $a_{ik} = a_{il} = a_{jk} = a_{jl}$; that is, such that there are no **monochromatic rectangles** within A. A matrix of order $p \times q$ for which such an m-colored set of elements exists is said to be m**-colorable**.

For example, the matrix A below is 2-colored, where the matrix B, which contains two monochromatic rectangles ($a_{11} = a_{12} = a_{41} = a_{42}$ and $a_{11} = a_{14} = a_{21} = a_{24}$), is not:

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \tag{4}$$

Note that any submatrix of an m-colored matrix is itself also an m-colored matrix. Note also that the notion of m-colorable matrices (of dimension 2) extends immediately to m-colorable d-dimensional tensors for $d > 2$.

In particular, Cooper, Fenner, and Purewal (2008) explore two natural questions: for which values of $\{n_1, n_2, \ldots, n_d\}$ is an $n_1 \times n_2 \times \ldots \times n_d$ tensor (of dimension $d \geq 2$) m-colorable, and for which values of $\{n_1, n_2, \ldots, n_d\}$ is such a tensor not m-colorable? The paper of Fenner, Gasarch, Glover, and Purewal (2009) focuses on refining the bounds on these two questions specifically in the case of $d = 2$. These two questions are closely related to the new question stated at the end of §1.4 (and solved exhaustively in the present work for $m = 2$ through 8); this new question might in fact be a significantly easier (that is, for values of m larger than those that can be solved exhaustively), as there are now only n unknowns, rather than $n_1 * n_2 * \cdots * n_d$ unknowns. Indeed, these two papers inspired the present focused investigation, which specifically sought a method for the efficient construction of large m-colored matrices. To accomplish such a construction, which is trivial once we know how to generate pairwise-nonrecurrent periodic sequences, we first review the definition of $n \times n$ circulant Toeplitz and Hankel matrices:

Definition: An $n \times n$ circulant Toeplitz matrix is a matrix that is constant along its extended diagonals, and is defined by the symbols in the n positions along its top row. An $n \times n$ circulant Hankel matrix is a matrix that is constant along its extended antidiagonals, and is also defined by the symbols in the n positions along its top row.

For example, in the case where $n = 4$, $n \times n$ circulant Toeplitz and Hankel matrices may be written in the general forms

$$T = \begin{pmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} a & b & c & d \\ b & c & d & a \\ c & d & a & b \\ d & a & b & c \end{pmatrix}. \tag{5}$$

Theorem: An $n \times n$ circulant Toeplitz or Hankel matrix is m-colored if and only if the top row of the matrix is a pairwise-nonrecurrent periodic sequence with m symbols.

Proof: To establish that an $n \times n$ circulant Hankel matrix is m-colored, we need only (due to the symmetry of the matrix) compare the top row to all of the other rows of the matrix and look for monochromatic rectangles. In particular, enumerating the rows and columns from zero, for each $s \in [1, n-1]$, looking at the top (zeroth) and s'th rows of the Hankel matrix, in order for there to be no monochromatic rectangles, no single symbol can simultaneously be in both the i'th and the $\mathrm{mod}(i+s, n)$'th positions as well as both the j'th and the $\mathrm{mod}(j+s, n)$'th positions along the top row for any distinct $i \in [0, n-1]$ and $j \in [0, n-1]$. These are exactly the conditions of pairwise nonrecurrence applied to the top row of the matrix, considered as a periodic sequence. The case of circulant Toeplitz matrices follows immediately, via rearrangement of the rows considered in the circulant Hankel case [see, e.g., (5)]. □

Thus, the circulant Hankel matrices and circulant Toeplitz matrices with top rows given by (3a) through (3g) give immediately: a 2-colored 4×4 matrix, a 3-colored 9×9 matrix, a 4-colored 16×16 matrix, a 5-colored 25×25 matrix, a 6-colored 36×36 matrix, a 7-colored 49×49 matrix, and an 8-colored 62×62 matrix.

Further, as (3a) through (3g) are the longest pairwise-nonrecurrent periodic sequences available at each corresponding value of m (proved via exhaustion in §1), it follows immediately that 2-colored 5×5, 3-colored 10×10, 4-colored 17×17, 5-colored 26×26, 6-colored 37×37, 7-colored 50×50, and 8-colored 63×63 matrices with circulant Hankel or circulant Toeplitz structure do not exist.

It appears that, with some patience, the largest 9-colored and possibly even the largest 10-colored circulant Hankel and circulant Toeplitz matrices might also be found via Algorithm B, as implemented in the single-threaded fortran codes provided herein, if there is sufficient interest. Doing such with the present code would require a fortran compiler with an `integer*16` data type; the present implementation of Algorithm B uses some somewhat sophisticated bitwise arithmetic on `integer*8` variables in order to streamline memory usage and therefore significantly improve execution speed, and is therefore limited to problems with $n \leq 64$ on compilers without `integer*16` data types, such as the present versions of `g95` and `gfortran`. The time-consuming part of Algorithm B is its second half, and should be trivial to parallelize; it also uses memory (and, thus, high speed cache) very efficiently, and incorporates only integer arithmetic and bit-wise comparisons (that is, there is no floating point arithmetic involved). This is thus an attractive algorithm for GPU-based implementation, if there is interest in extending it to larger problems.

References

1. J. Cooper, S. Fenner, and S. Purewal, Monochromatic Boxes in Colored Grids. Preprint arXiv:0810.3019v1. (2008)

2. S. Fenner, W. Gasarch, C. Glover, and S. Purewal, Rectangle Free Coloring of Grids. Preprint arXiv:1005.3750. (2009)

3. D. Knuth, *The Art of Computer Programming, Volume 4: Combinatorial Algorithms*. Preprint. (2005)

4. C. D. Langford, Problem, *Mathematical Gazette* **42** (1958), 228.

5. T. Skolem, On certain distributions of integers in pairs with given differences, *Math. Scand.* **5** (1957), 57-68.

6. C. J. Colbourn, J. H. Dinitz, *The CRC handbook of combinatorial designs*. CRC Press. (1996)

7. N. Shalaby, T. Stuckless, The Existence of Looped Langford Sequences. *CRUX with MAYHEM* **26** (2000), 86-92.

8. S. Sidon, Ein Satz ber trigonometrische Polynome und seine Anwendungen in der Theorie der Fourier-Reihen, *Mathematische Annalen* **106** (1932), 536-539

9. W. C. Babcock, Intermodulation Interference in Radio Systems/Frequency of Occurrence and Control by Channel Selection, *Bell System Technical Journal* **31** (1953), 63-73.