2011

Self-matching bands in the paperfolding sequence

Bruce Bates  
*University of Wollongong, bbates@uow.edu.au*

Martin Bunder  
*University of Wollongong, mbunder@uow.edu.au*

Keith Tognetti  
*University of Wollongong, tognetti@uow.edu.au*

Publication Details

Self-matching bands in the paperfolding sequence

Abstract
We compare term by term the paperfolding sequence with a copy displaced by d terms to obtain the matching fraction M(d).

Keywords
matching, self, bands, sequence, paperfolding

Disciplines
Physical Sciences and Mathematics

Publication Details
SELF-MATCHING BANDS IN THE PAPERFOLDING SEQUENCE

BRUCE BATES, MARTIN BUNDER, AND KEITH TOGNETTI

Abstract. We compare term by term the paperfolding sequence with a copy displaced by \( d \) terms to obtain the matching fraction \( M(d) \). It is shown that \( M(d) \) has an interesting structure in that if \( d = 2^b (1 + 2 \alpha) \), then \( M(d) = \left| \frac{1 - \frac{3}{2^{b+1}}}{} \right| \) thereby generating horizontal bands for each value of \( b \). That is, \( M(d) \) depends only on \( b \).

1. Introduction

Consider two binary sequences: \( S = f_1 f_2 f_3 \ldots \) and \( S \) displaced by \( d \), that is, the sequence \( f_{d+1} f_{d+2} f_{d+3} \ldots \). As the terms can differ only by a unit, we look at the expression \( |f_{d+i} - f_i| \) for \( i \in \mathbb{N} \). If this is zero we have a match at the \( i \)-th term; otherwise it is unity and we have a mismatch.

Example 1. Let \( S = 1101100111 \ldots \) be displaced by 3 terms. Then \( |f_{3+i} - f_i| \) can be represented pictorially as follows.

\[
\begin{array}{cccccccc}
1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
\end{array}
\]

\( |f_{3+i} - f_i| : \) 0 0 0 1 0 1 1 0

This suggests the following definition.

Definition 1. (The self-matching function). Let \( S \) be an infinite binary sequence. The proportion of matches for \( S \), with \( S \) displaced by \( d \), is given by:

\[
M(d) = \lim_{m \to \infty} \left( m - \sum_{i=1}^{m} |f_{d+i} - f_i| \right) / m.
\]

Recently, Tognetti [4] described a surprisingly simple matching pattern for Bernoulli sequences for which \( f_i = \lfloor \frac{i}{\alpha} \rfloor - \lfloor \frac{i+1}{\alpha} \rfloor \). This represents the difference sequence for the integer parts sequence. It was shown that the graph of \( M(d) \) against \( d \) exhibited a Moiré pattern and that unexpectedly this pattern was obtained by simply folding the fractional parts graph about its middle.

This paper examines the self-similarity within the paperfolding sequence and reveals yet another interesting pattern within the graph of a paperfolding \( M(d) \) against \( d \). We show that the graph forms horizontal bands.

2. The Paperfolding Operation

There have been many studies on the paperfolding sequence, \( S = 11011001110 \ldots \), since the seminal paper by Davis and Knuth [2]. It is based on the following simple operation: repeatedly fold a piece of paper, right over left, \( i \) times. When unfolded,
the paper contains v-shaped and inverted v-shaped creases. If we represent a v-shape by a 1 and an inverted v-shape by a 0, we obtain the following paperfolding subsequence after i folds (containing $2^i - 1$ creases):

$$S_i = f_1 f_2 f_3 \ldots f_{2^i-1} = 110 \ldots 100.$$ 

For example,

$$S_1 = 1,$$
$$S_2 = 110,$$
$$S_3 = 1101100.$$ 

As i becomes unbounded we have the infinite sequence, $S$. A comprehensive treatment of various paperfolding properties as well as a survey of the development of the paperfolding sequence can be found in Bates et al [1]. There it was shown that $S$ can be represented by the interleaving of two sequences, as follows.

**Definition 2. (Interleave operator).** The interleave operator $\#$ acting on the two sequences $U = u_1 u_2 \ldots u_k$ and $V = v_1 v_2 \ldots v_n$, where $k > n$, generates the following interleaved sequence:

$$U \# V = u_1 \ldots u_p v_1 u_{p+1} \ldots u_{2p} v_2 u_{2p+1} \ldots u_{np} v_n u_{np+1} \ldots u_k,$$

where $p = \left\lfloor \frac{k}{n+1} \right\rfloor$.

**Definition 3. (Alternating sequence).** The alternating sequence of length $2r$ is given by $A_{2r} = 1010 \ldots 10$.

**Definition 4. (Interleaving expression for paperfolding).** For $i \geq 2$, the paperfolding sequence of length $2^i - 1$, $S_i$, is defined as

$$S_i = A_{2^i-1} \# S_{i-1}$$

where $S_1 = 1$.

$S$ can also be represented through mirroring.

**Definition 5. (Mirror paperfolding sequence).** The mirror paperfolding sequence of length $2^i - 1$, $S^R_i$, is defined as the reversal of $S_i$ combined with each 1 being replaced by 0 and each 0 being replaced by 1.

The following results are found in Bates et al [1].

**Theorem 1.** $S_{i+1} = S_i \ 1 \ S^R_i$ and $S^R_{i+1} = S_i \ 0 \ S^R_i$ where $S_1 = 1$.

**Corollary 1.** $S_i = A_{2^i-1} \# A_{2^i-2} \# \ldots \# A_2 \# 1$ and $S^R_i = A_{2^i-1} \# A_{2^i-2} \# \ldots \# A_2 \# 0$.

Corollary 1 tells us that the paperfolding sequence is equivalent to a series of successive interleaves of alternating sequences applied to the term $S_1 = 1$; and the mirror paperfolding sequence is equivalent to a series of successive interleaves of alternating sequences applied to the term $S^R_1 = 0$.

**Theorem 2.** $S_i$ contains $2^{i-1} - 1$ instances of 0 and $2^{i-1}$ instances of 1.

We now demonstrate a more general result: the paperfolding sequence is an interleave of smaller paperfolding sequences.
Definition 6. (Alternating paperfolding sequence). The alternating paperfolding sequence of length \(2^n - 2^m\), \(0 < n < i\), is given by
\[
A_{i,n} = S_{i-n} \overline{S}_{i-n} S_{i-n} \overline{S}_{i-n} \cdots S_{i-n} \overline{S}_{i-n},
\]
where the right hand side consists of \(2^{n-1}\) copies of \(S_{i-n} \overline{S}_{i-n}\).

Theorem 3. \(S_i = A_{i,n} \# S_n\) and \(S_i^R = A_{i,n} \# S_n^R\).

Note that particular values of \(n\) yield familiar expressions for \(S_i\). That is,
\begin{enumerate}
  \item For \(n = 1\), \(S_1 = A_{1,1} \# S_1 = S_{i-1} 1 \overline{S}_{i-1}\), and
  \item For \(n = i - 1\), \(S_i = A_{i,i-1} \# S_{i-1} = A_{i-1} \# S_{i-1}\).
\end{enumerate}

In order to evaluate \(f_i\), we represent \(i\) as \(2^k (2r + 1)\) where \(k, r \geq 0\). This representation is characteristic of many folding structures apart from paperfolding, such as with the stickbreaking sequence, the Stern-Brocot tree and the Sarkovskiy ordering of cycles in chaos (See Devaney [3]). It follows that \(i\) in binary is the binary number \(r\), followed by a 1 and then \(k\) 0s.

The following two results for \(f_i\) are found in Bates et al [1].

Theorem 4. For \(i = 2^k (2r + 1)\), \(f_i = 1 + r \mod 2\).

We use the fact that \(2r + 1\) can be partitioned into \(4h + 1\), for \(r = 2h\); and \(4h + 3\), for \(r = 2h + 1\) in the formulation of the following result.

Theorem 5. For \(k, h \geq 0\),
\[
f_i = \begin{cases} 
  1, & \text{if } i = 2^k (4h + 1) \\
  0, & \text{if } i = 2^k (4h + 3)
\end{cases}
\]

Corollary 2. For \(i = 2^k (4h + a)\) and \(s = 2^k (4l + t)\) where \(a, t \in \{1, 3\}\),
\begin{enumerate}
  \item \(f_i = \frac{1}{2} (3 - a)\)
  \item \(f_i = f_s\), if and only if \(a = t\).
\end{enumerate}

Theorem 6. For \(i = 2^k (4h + a)\) and \(s = 2^k (4l + t)\) where \(a, t \in \{1, 3\}\),
\begin{enumerate}
  \item if \(b < k - 1\),
    \begin{enumerate}
      \item \(f_{i+s} = f_s\),
      \item \(f_{i+s} = f_t\), if and only if \(a = t\),
    \end{enumerate}
  \item if \(b = k - 1\),
    \begin{enumerate}
      \item \(f_{i+s} \neq f_s\),
      \item \(f_{i+s} = f_t\), if and only if \(a \neq t\),
    \end{enumerate}
  \item if \(b = k\),
    \begin{enumerate}
      \item \(f_{i+s} = f_t\), if and only if \(a = t\) and \(2 \mid (h + 1)\); or,
      \[ a \neq t \text{ and } h + l + 1 = 2^u (4v + a) \text{ for some } u, v \geq 0, \]
      \item \(f_{i+s} = f_s\), if and only if \(a = t\) and \(2 \mid (h + l)\); or,
      \[ a \neq t \text{ and } h + l + 1 = 2^u (4v + t) \text{ for some } u, v \geq 0. \]
    \end{enumerate}
\end{enumerate}

Proof. We have \(i = 2^k (4h + a)\) and \(s = 2^k (4l + t)\) where \(a, t \in \{1, 3\}\). We examine each case.
\begin{enumerate}
  \item If \(i + s = 2^k (4 (2h+b) + l + 2^k - b - a)\), (a) and (b) follow from Corollary 2 ii),
  \item If \(i + s = 2^{k-1} (4 (l+2h) + (2a + t))\), as \(t \neq (2a + t) \mod 4\) for any \(a\) and \(t\) and \(a = (2a + t) \mod 4\), if and only if \(a \neq t\), (a) and (b) follow by Corollary 2 ii),
  \item (a) For \(a = t\), \(i + s = 2^{k+1} (2 (h + l) + a)\). Also by Corollary 2 ii),
\end{enumerate}
\* if \( 2 \mid (h + l) \), then \( i + s = 2^{k+1} \left( 4 \left( \frac{h+1}{2} \right) + a \right) \) so \( f_{i+s} = f_i = f_s \).
\* if \( 2 \nmid (h + l) \), and \( a = 3 \), then \( i + s = 2^{k+1} \left( 4 \left( \frac{h+1}{2} \right) + 1 \right) \) so \( f_{i+s} \neq f_i, f_s \).
\* if \( 2 \mid (h + l) \), and \( a = 1 \), then \( i + s = 2^{k+1} \left( 4 \left( \frac{h+1}{2} \right) \right) + 3 \) so \( f_{i+s} \neq f_i, f_s \).

(b) For \( a \neq t \), \( i + s = 2^{k+2} (h + l + 1) \). Accordingly,
\* if \( h + l + 1 = 2^v (4v + a) \) for some \( u, v \geq 0 \), \( f_{i+s} = f_i \).
\* if \( h + l + 1 = 2^v (4v + t) \) for some \( u, v \geq 0 \), \( f_{i+s} = f_s \).

In the special case where \( s = 1 \), by i), ii) and iii) for \( u \geq 0 \) and \( h' = 4 - h \),

\[
f_{i+1} = f_i \quad \text{if and only if} \quad i = \begin{cases} 
2^{u+2} (4h + 1), & \text{or} \\
2 (4h + 3), & \text{or} \\
8h' + 1, & \text{or} \\
2^{u+2} (4v + 3) - 1. & \text{or}
\end{cases}
\]

3. The Graph of the Self-Matching Function, \( M (d) \)

We now state our main result.

**Theorem 7.** Let \( d = 2^k (2r + 1) \). Then \( M (d) = \left| 1 - \frac{3}{2d + 1} \right| \).

**Proof.** There are two cases to consider:

i) \( d \) is odd, that is, \( b = 0 \). There are two sub-cases:

(a) \( d = 4l + 1 \).

(1) Consider \( l = 0 \), that is, \( d = 1 \). From Definition 4, \( S \) is the interleave of the sequences in 3.1.1 and 3.1.2 while \( S \) displaced by 1, is the interleave of 3.1.3 and 3.1.4. Corresponding matched or mismatched entries in the overlay are shown by:

\[
\begin{align*}
(3.1.1) & \lim_{i \to +\infty} A_{2i-1} : \quad 1 & 0 & 1 & 1 & 1 & \ldots \\
(3.1.2) & S : \quad 1 & : & 1 & : & 0 & : & 1 & : & 1 & \ldots \\
(3.1.3) & \lim_{i \to +\infty} A_{2i-1} : \quad 1 & : & 0 & : & 1 & : & 0 & : & 1 & \ldots \\
(3.1.4) & S : \quad 1 & 1 & 0 & 1 & \ldots
\end{align*}
\]

Consider (3.1.3):

- Every odd entry is a 1. Each is aligned with odd entries in \( S \) in (3.1.2) which by Definition 4 are consecutive values of an infinite alternating sequence. Thus half of these alignments match.
- Every even entry is a 0. Each is aligned with even entries in \( S \) in (3.1.2) which by Definition 4 are consecutive values of \( S \). By Theorem 2, the ratio of matching 0s in (3.1.3) is \( \lim_{i \to +\infty} \frac{2^i - 1}{2^i + 1} = \frac{1}{2} \).

Thus half of these alignments match.

Consider (3.1.4):

- Consecutive odd entries form an infinite alternating sequence. Each is aligned to even entries in (3.1.1) which are all 0s. Thus half of these alignments match.
Consecutive even entries form $S$. Each is aligned with a 1 from (3.1.1). By Theorem 2, the ratio of matching 1s in (3.1.4) is \[ \lim_{i \to \infty} \frac{2^{i-1}}{2^{i-1} - 1} = \frac{1}{2}. \] Thus half of these alignments match.

It follows that $M(1) = \frac{1}{2}$.

(II) Consider $l > 0$. Each entry in 3.1.3 and 3.1.4 moves $4l$ spaces to the right. Despite this move, each entry in 3.1.1 and 3.2.1 (except the leftmost $d$ entries which are now unaligned) is aligned to a value identical to that found in the case for $l = 0$. Thus $M(4l + 1) = M(1) = \frac{1}{2}$, $l \in \mathbb{N}$.

(b) $d = 4l + 3$.

(I) Consider $l = 0$, that is, $d = 3$. As with (a), $S$ overlaid with itself, with displacement 3, can be broken down into the following four subsequences:

\[
\begin{align*}
(3.2.1) \quad \lim_{i \to \infty} A_{2i-1} : & \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad \cdots \\
(3.2.2) \quad S : & \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad \cdots \\
(3.2.3) \quad \lim_{i \to \infty} A_{2i-1} : & \quad 1 \quad 0 \quad 1 \quad 0 \quad \cdots \\
(3.2.4) \quad S : & \quad 1 \quad 1 \quad 0 \quad \cdots 
\end{align*}
\]

Consider (3.2.3):

- Every odd entry is a 1. Each is aligned with even entries in $S$ in (3.2.2) which by Definition 4 are consecutive values of $S$. By Theorem 2, the ratio of matching 1s in (3.2.3) is \[ \lim_{i \to \infty} \frac{2^{i-1}}{2^{i-1} - 1} = \frac{1}{2}. \] Thus half of these alignments match.

- Every even entry is a 0. Each is aligned with odd entries in (3.2.2) which form an infinite alternating sequence. Thus half of these alignments match.

Consider (3.2.4):

- Consecutive odd entries form an infinite alternating sequence. Each is aligned to odd entries in (3.2.1) which are all 1s. Thus half of these alignments match.

- Consecutive even entries form $S$. Each is aligned with a 0 from (3.2.1). By Theorem 2, the ratio of matching 0s in (3.2.4) is \[ \lim_{i \to \infty} \frac{2^{i-1} - 1}{2^{i-1}} = \frac{1}{2}. \] Thus half of these alignments match.

It follows that $M(3) = \frac{1}{2}$.

(II) Consider $l > 0$. Each entry in 3.2.3 and 3.2.4 moves $4l$ spaces to the right. Despite this move, each entry in 3.2.1 and 3.2.2 (except the leftmost $d$ entries which are now unaligned) is aligned to a value identical to that found in the case for $l = 0$. Thus $M(4l + 3) = \frac{1}{2}$, $l \in \mathbb{N}$.

Combining (a) and (b), for $b = 0$, $M(d) = \frac{1}{2}$.

ii) $d$ is even, that is, $d = 2^b(4l + t)$ where $t \in \{1, 3\}, b > 0$. 
From Theorem 3, taking limits, $S = S_b f_1 \overline{S_b} f_2 S_b f_3 \overline{S_b} f_4 \ldots$. Since each $S_b f_i$ and $\overline{S_b} f_i$ is of length $2^k$, we also have

$$S = S_b f_2 \overline{S_b} f_2 \overline{S_b} f_3 \overline{S_b} f_3 \overline{S_b} f_4 \overline{S_b} f_4 \overline{S_b} \ldots$$

So $S$ overlaid with itself with displacement $d$ can be viewed as

$$S_b \quad 1 \quad \overline{S_b} \quad 1 \quad \ldots \quad S_b \quad f_d \quad \overline{S_b} \quad f_d \overline{S_b} \quad \ldots$$

where after the $\left\lceil \frac{d+1}{2} \right\rceil$th instance of $S_b$ in the first line, $S_b$ entries are overlaid with $\overline{S_b}$, and $\overline{S_b}$ entries are overlaid with $S_b$. Consider these overlays of $S_b$ and $\overline{S_b}$ entries in (3.3). By Theorem 1, each middle term is mismatched, thereby generating mismatches every $2^k$ spacings in (3.3). Thus for large $m$, $\frac{m}{2^k}$ terms are mismatched. Now consider the overlay of the other entries in (3.3). These occur every $2^k$ spacings and represent $S$ overlaid with itself with odd displacement. By i), half of these entries mismatch and so for large $m$, there are $\frac{m}{2^k + 1}$ of these mismatches. Since these overlays are mutually exclusive, we can add the mismatches. That is, for large $m$, there are $\frac{3m}{2^k + 1}$ mismatches. Thus $M(d) = 1 - \frac{3m}{2^k + 1}$ for $d$ even.

From Theorem 7, as $M(d)$ is a function of $b$ only, then $M(d)$ is constant for constant $b$. Hence the graph of $M(d)$ consists of horizontal bands based on $b$ such that each band has height $\lfloor 1 - \frac{3}{2^k + 1} \rfloor$ as shown in Figure 1. We note that although the matching band for $2(2r + 1)$ is below the band for odd numbers ($b = 0$) all the other bands are above the odd band. That is,

- Band 0, ($b = 0$), $M(d) = \frac{1}{2}$, $d$ is odd,
- Band 1, ($b = 1$), $M(d) = \frac{1}{3}$, $d = 2 + 4s = 2, 6, 10, \ldots$,
- Band 2, ($b = 2$), $M(d) = \frac{1}{6}$, $d = 4 + 8s = 4, 12, 20, \ldots$,
- \vdots
- Band $n$, ($b = n$), $M(d) = \left\lfloor 1 - \frac{3}{2^k + 1} \right\rfloor$, $d = 2^n + 2^n + 1 \cdot s = 2^n, 2^n + 3, 2^n + 5, \ldots$

**Theorem 8.** For $k > 0$, and $1 \leq d < 2^k$,

$$M(d) = M\left(2^k \pm d\right).$$

**Proof.** If $d = 2^k (2r + 1) < 2^k$, then $b < k$, and $2^k \pm d = 2^k (2^{k-b-1} \pm r) \pm 1$. By Theorem 7, $M\left(2^k \pm d\right) = M(d)$.

Theorem 8 tells us that if we have the section of the graph up to $d = 2^k - 1$, we can generate the graph up to $2^k + 1 - d$ by adding the point $(2^k, M(2^k))$ and then translating the earlier section to the right of $2^k$.

### 4. The Expected Value of $M(d)$

The terms associated with band $b$ for $b > 0$ have period $2^k+1$. Hence the proportion of these terms that possess this matching is $\frac{1}{2^k+1}$. Band 0 makes the largest contribution to the expected value of $M(d)$, $E\left(M(d)\right)$, of any band. It contains
half the total number of points, each with value $\frac{1}{2}$, making its total contribution $\frac{1}{4}$. It contributes half of $E(M(d))$ as shown below.

$$E(M(d)) = \sum_{k=0}^{\infty} \left( 1 - \frac{3}{2^{k+1}} \right) \frac{1}{2^{k+1}}$$

$$= \frac{1}{4} + \sum_{k=1}^{\infty} \left( 1 - \frac{3}{2^{k+1}} \right) \frac{1}{2^{k+1}}$$

$$= \frac{1}{2}$$

REFERENCES


