Aperiodic correlation and the merit factor

Aina Johansen

02.11.2009
Correlation

The periodic correlation between two binary sequences \( \{ x_t \} \) and \( \{ y_t \} \) of length \( n \) at shift \( \tau \) is defined as

\[
\theta_{x,y}(\tau) = \sum_{t=0}^{n-1} (-1)^{x_{t+\tau} - y_t}
\]

\[
x_\tau \quad x_{\tau+1} \quad \cdots \quad x_{n-1} \quad x_0 \quad x_1 \quad \cdots \quad x_{\tau-1}
\]
\[
y_0 \quad y_1 \quad \cdots \quad y_{n-1-\tau} \quad y_{n-\tau} \quad y_{n-\tau+1} \quad \cdots \quad y_{n-1}
\]

The aperiodic correlation between two binary sequences \( \{ x_t \} \) and \( \{ y_t \} \) of length \( n \) at shift \( \tau \) is defined as

\[
C_{x,y} = \sum_{t=0}^{n-\tau-1} (-1)^{x_{t+\tau} - y_t}
\]

\[
x_0 \quad x_1 \quad \cdots \quad x_{\tau-1} \quad x_\tau \quad \cdots \quad x_{n-1}
\]
\[
y_0 \quad \cdots \quad y_{n-1-\tau} \quad y_{n-\tau} \quad \cdots \quad y_{n-1}
\]
Autocorrelation

When \( \{x_t\} = \{y_t\} \) we call it the autocorrelation, and we denote it \( C_x(\tau) \).

\[
\begin{array}{cccccccc}
  x_{\tau} & x_{\tau+1} & \cdots & x_{n-1} & x_0 & x_1 & \cdots & x_{\tau-1} \\
  x_0 & x_1 & \cdots & x_{n-1-\tau} & x_{n-\tau} & x_{n-\tau+1} & \cdots & x_{n-1} \\
\end{array}
\]

\[
\theta_x(\tau) = C_x(\tau) + C_x(n - \tau) \quad \text{when } \tau \neq 0
\]

Example: Let \( s_t = 01011 \).
CDMA

We have a set of sequences where all pairs of sequences from the set are orthogonal, i.e. the inner product is 0.

Each user \( i \) is given a sequence \( s_i \). If the user wants to send a 1 he transmits his sequence. If the user wants to send a 0 he transmits the complement of his sequence.

If two users \( i \) and \( j \) transmit a sequence simultaneously, the received signal will be the sum of the sequences, \( s_i + s_j \). To determine what was sent by user \( i \), we can compute the inner product \((s_i + s_j) \cdot s_i\). As \( s_i \cdot s_j = 0 \), we are left with the data sent by user \( i \).

When user \( i \) starts to transmit a sequence, and user \( j \) starts to transmit a sequence after \( i \) has started, but before \( i \) has finished, the received signal will be a partial overlap between the two sequences. In order to ease synchronization we need the sequences to have low correlation.
Other applications

Binary sequences with low correlation have a number of uses:

- spread-spectrum communication systems
- stream ciphers
- pseudorandom number generation
- pulse compression
- synchronization
- theoretical physics and chemistry
- etc
The merit factor

We want binary sequences whose aperiodic autocorrelations are collectively small. The merit factor can be used as a measure of this, and is defined as follows:

**Definition**

Let \( \{s_t\} \) be a binary sequence of length \( n \). The *merit factor* of \( \{s_t\} \) is given as

\[
F(s) = \frac{n^2}{2 \sum_{u=1}^{n-1} [C_s(u)]^2}
\]

If the aperiodic autocorrelation of \( \{s_t\} \) is small for all shifts, the merit factor will be high. So the higher the merit factor, the better the sequence.

**Example:** Let \( s = 01011 \). Then we get the merit factor

\[
F(s) = \frac{5^2}{2 \times ((-2)^2 + 1^2 + 0^2 + (-1)^2)} = \frac{25}{2 \times 6} = 2.08
\]
The merit factor problem

Let $S_n$ be the set of all binary sequences of length $n$. We define $F_n$ to be the optimal merit factor value for sequences of length $n$, that is

$$F_n = \max_{s \in S_n}(F(s))$$

The merit factor problem
Determine the value of $\limsup_{n \to \infty} F_n$.

Proposition

The mean value of $1/F$, taken over all sequences of length $n$, is $\frac{n-1}{n}$.

The above proposition means that the expected asymptotic value of $F$ for a randomly-chosen binary sequence is 1.
Some results so far

The best asymptotic results so far are given by explicitly constructed families of sequences with a provable asymptotic merit factor of 6.

In recent years new constructions have given sequences of very long length (millions of elements) with asymptotic merit factor $> 6.34$.

Exhaustive computer search has shown that for $1 \leq n \leq 40, \ n \neq 11, 13, 3.3 \leq F_n \leq 9.58$. For lengths up to 117 the highest known merit factor is between 8 and 9.56.

$F_n$ has been computed for all $n \leq 60$, and for larger values of $n$ in the range $61 \leq n \leq 217$. Some large values of $F_n$ has occurred.

The only known sequences with merit factor $\geq 10$ are the Barker sequences of length 11 and 13. These sequences have merit factor $F_{11} = 12.1$ and $F_{13} = 14.08$. 
Golay’s postulate

The aperiodic autocorrelations $C(1), C(2), \ldots, C(n-1)$ of a sequence that is chosen at random from the $2^n$ binary sequences of length $n$ are clearly dependent random variables.

In 1977 Golay made a postulate that said that the correct value of $\limsup_{n \to \infty} F_n$ can be found by treating $C(1), C(2), \ldots, C(n-1)$ as independent random variables for large $n$. From this, he made the following conjecture:

Conjecture

$$\limsup_{n \to \infty} F_n = 12.32 \ldots$$
Barker sequences

A Barker sequence is a binary sequence \( \{s_t\} \) of length \( n \) where

\[
|C_s(\tau)| \leq 1 \quad \text{for all } 0 < \tau < n
\]

Only the following Barker sequences are known so far:

\[
\begin{align*}
\text{n} = 2 & \quad 11 & \quad \text{n} = 7 & \quad 1110010 \\
\text{n} = 3 & \quad 110 & \quad \text{n} = 11 & \quad 11100010010 \\
\text{n} = 4 & \quad 1110 & \quad \text{n} = 13 & \quad 1111100110101 \\
\text{n} = 5 & \quad 11101
\end{align*}
\]

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Acyclic shift</th>
<th>( C(\tau) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1110010</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>111001</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>11100</td>
<td>−1</td>
</tr>
<tr>
<td>3</td>
<td>1110</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>111</td>
<td>−1</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>−1</td>
</tr>
</tbody>
</table>
The Barker property is preserved under the transformations

\[ s_t \rightarrow \overline{s}_t \]

\[ s_t \rightarrow \begin{cases} 
  s_t & \text{if } t \text{ is even} \\
  \overline{s}_t & \text{if } t \text{ is odd}
\end{cases} \]

\[ s_t \rightarrow s_{n-1-t} \]

Example

Let \( \{s_t\} = 1110010 \) be the Barker sequence of length 7. \( s_t \rightarrow \overline{s}_t \) gives us 0001101.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Acyclic shift</th>
<th>( C(\tau) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0001101</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>000110</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>00011</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>0001</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>000</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>-1</td>
</tr>
</tbody>
</table>
Example

In the second transformation we complement the elements in all odd positions of the sequence. Hence $1110010 \rightarrow 1011000$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Acyclic shift</th>
<th>$C(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1011000</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>101100</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>10110</td>
<td>$-1$</td>
</tr>
<tr>
<td>3</td>
<td>1011</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>$-1$</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>$-1$</td>
</tr>
</tbody>
</table>
Example

In the last transformation, we simply reverse the order of the sequence. Hence $1110010 \rightarrow 0100111$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Acyclic shift</th>
<th>$C(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0100111</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>010011</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>01001</td>
<td>−1</td>
</tr>
<tr>
<td>3</td>
<td>0100</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>010</td>
<td>−1</td>
</tr>
<tr>
<td>5</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>−1</td>
</tr>
</tbody>
</table>
There are no Barker sequences of odd length greater than 13. If \( n \) is even, it is necessary that \( n \equiv 0 \pmod{4} \).

It has been shown that there are no Barker sequences for even \( n \), \( 4 < n \leq 1\,898\,884 \).

It is believed that there exist no further Barker sequences other than those already listed, but so far no one has been able to prove this.
Rudin-Shapiro sequences

The Rudin-Shapiro sequence pair \( X(m), Y(m) \) is defined recursively by

\[
X(m) = X(m-1) : Y(m-1) \\
Y(m) = X(m-1) : \overline{Y(m-1)}
\]

where \( X(0) = Y(0) = 1 \), \( a : b \) means appending \( b \) to \( a \), and \( \overline{x} \) is the complement of \( x \).

Example

<table>
<thead>
<tr>
<th>( m )</th>
<th>( X(m) )</th>
<th>( Y(m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1110</td>
<td>1101</td>
</tr>
<tr>
<td>3</td>
<td>11101101</td>
<td>11100010</td>
</tr>
</tbody>
</table>
The merit factor of Rudin-Shapiro sequences

**Theorem**

The merit factor of both sequences $X^{(m)}$ and $Y^{(m)}$ of a Rudin-Shapiro pair of length $2^m$ is

$$F = \frac{3}{1 - (-1/2)^m}$$

Hence the asymptotic merit factor of both sequences is 3.

**Example**

Look at $X^{(3)} = 11101101$. The aperiodic correlation values are $(-1, 0, 3, 0, 1, 0, 1)$.

$$F = \frac{n^2}{2 \sum_{u=1}^{n-1}[C(u)]^2} = \frac{8^2}{2 \times ((-1)^2 + 3^2 + 1^2 + 1^2)} = \frac{8}{3}$$

$$F = \frac{3}{1 - (-1/2)^m} = \frac{3}{1 - (-1/2)^3} = \frac{8}{3}$$
Difference sets

Among the families of sequences whose merit factor has been determined we find several sequences corresponding to difference sets.

Definition
Let $G$ be a group of order $v$. Let $D$ be a subset of $G$ of size $k$. We say that $D$ is a $(v, k, \lambda)$ difference set if every nonzero element of $G$ has exactly $\lambda$ representations as a difference $d_1 - d_2$, where $d_1, d_2 \in D$. The order of the difference set is defined as $n = k - \lambda$.

Example
Let $G$ be the group $\mathbb{Z}_{11}$. $D = \{1, 3, 4, 5, 9\}$ is a subset of $G$ of size 5. $D$ is a $(11, 5, 2)$ difference set of order $n = 11 - 5 = 4$.

<table>
<thead>
<tr>
<th></th>
<th>4 − 3</th>
<th>5 − 4</th>
<th>6</th>
<th>4 − 9</th>
<th>9 − 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 − 1</td>
<td>5 − 3</td>
<td>7</td>
<td>1 − 5</td>
<td>5 − 9</td>
</tr>
<tr>
<td>2</td>
<td>1 − 9</td>
<td>4 − 1</td>
<td>8</td>
<td>1 − 4</td>
<td>9 − 1</td>
</tr>
<tr>
<td>3</td>
<td>5 − 1</td>
<td>9 − 5</td>
<td>9</td>
<td>1 − 3</td>
<td>3 − 5</td>
</tr>
<tr>
<td>4</td>
<td>3 − 9</td>
<td>9 − 4</td>
<td>10</td>
<td>3 − 4</td>
<td>4 − 5</td>
</tr>
</tbody>
</table>
Characteristic function
The characteristic function of a difference set $D$ is the binary sequence $\{s_t\}$ of length $v$ where

$$s_t = \begin{cases} 1 & t \in D \\ 0 & t \notin D \end{cases}$$

Example
$G = \mathbb{Z}_{11}$, $D = \{1, 3, 4, 5, 9\}$, $\{s_t\} = 01011100010$

Hadamard difference sets
Difference sets with parameters of the form

$$(v = 4t - 1, k = 2t - 1, \lambda = t - 1)$$

are called Hadamard difference sets. Only difference sets with Hadamard parameters give rise to sequences with nonzero asymptotic merit factor.
m-sequences

Let \( f(x) = \sum_{i=0}^{m} f_i x^i \) where \( f_i \in 0, 1 \) be a primitive polynomial. From this polynomial we get the recurrence relation
\[
s_{t+m} = \sum_{i=0}^{m-1} f_i s_{t+i}.
\]
This recurrence relation generates a binary sequence of period \( 2^m - 1 \), called an m-sequence.

Example

Let \( f(x) = x^4 + x + 1 \). This gives us the recurrence relation
\[
s_{t+4} = s_{t+1} + s_t.
\]
For initial values \( s_0 = s_1 = s_2 = 0 \), \( s_3 = 1 \) we get the sequence \( \{s_t\} = 000100110101111 \)

Theorem

The mean value of \( 1/F \), taken over all \( n \) rotations of an m-sequence of length \( n = 2^m - 1 \), is
\[
\frac{(n-1)(n+4)}{3n^2}.
\]

Theorem

The asymptotic merit factor of an m-sequence is 3.
Legendre sequences

Let $p$ be an odd prime. We define the “modified” Legendre symbol as

$$\left[ \frac{a}{p} \right] = \begin{cases} 1 & \text{if } a \text{ is a quadratic residue mod } p \\ 0 & \text{if } a \equiv 0 \pmod{p} \text{ or } a \text{ is a quadratic nonresidue mod } p \end{cases}$$

The Legendre sequence of length $p$ is defined as

$$\{s_t\} = \left[ \frac{t}{p} \right]$$

When $p \equiv 3 \pmod{4}$ the Legendre sequence is the characteristic function of a Hadamard difference set, called the Legendre difference set.
Example

Let $v = 11$. We see that

\[
\begin{align*}
1 \times 1 &= 1 \\
2 \times 2 &= 4 \\
3 \times 3 &= 9 \\
4 \times 4 &= 5 \\
5 \times 5 &= 3 \\
6 \times 6 &= 3 \\
7 \times 7 &= 5 \\
8 \times 8 &= 9 \\
9 \times 9 &= 4 \\
10 \times 10 &= 1
\end{align*}
\]

Then $\{1, 3, 4, 5, 9\}$ are the only quadratic residues, and we get the Legendre sequence $\{s_t\} = 01011100010$. 
An “offset” sequence is one in which a fraction \( f \) of the elements of a sequence of length \( n \) are chopped off at one end and appended to the other end. This corresponds to a cyclic shift of \( nf \) places.

**Theorem**

The Legendre sequences, when offset by a fraction \( f \) of their length, have asymptotic merit factor \( F \) satisfying

\[
\frac{1}{F} = \frac{2}{3} - 4|f| + 8f^2, \quad |f| \leq \frac{1}{2}
\]

so that \( F \) reaches a maximum value of 6 when \( |f| = \frac{1}{4} \).
Jacobi sequences

Let $n = p_1 p_2 \cdots p_k$, where the $p_i$’s are primes. The “modified” Jacobi symbol is defined as

$$
\left[ \frac{a}{n} \right] = \bigoplus_{i=1}^{k} \left[ \frac{a}{p_i} \right]
$$

The Jacobi sequence of length $n$ is defined as

$$\{s_t\} = \left[ \frac{t}{n} \right]$$

In the case when $n = p(p + 2)$, there exists a modified version of the Jacobi sequences, called the Twin prime sequences.
Twin prime sequences

Let \( v = p(p + 2) \), where \( p \) and \( p + 2 \) are prime. The Twin prime sequence is defined as

\[
\{s_t\} = \begin{cases} 
0 & \text{if } t \equiv 0 \pmod{p+2} \\
1 & \text{if } t \not\equiv 0, t \equiv 0 \pmod{p} \\
\left[\frac{t}{p}\right] \oplus \left[\frac{t}{p+2}\right] & \text{otherwise}
\end{cases}
\]
Example of Twin prime sets

Let \( v = 3 \times 5 \). \( QR_3 = \{1\} \), \( QR_5 = \{1, 4\} \).

\[
\{s_t\} = \begin{cases} 
0 & \text{if } t \equiv 0 \pmod{5} \\
1 & \text{if } t \not\equiv 0, t \equiv 0 \pmod{3} \\
\left\lfloor \frac{t}{3} \right\rfloor \oplus \left\lfloor \frac{t}{5} \right\rfloor & \text{otherwise}
\end{cases}
\]

\[
\begin{array}{c|c}
 t & s_t \\
0 & 0 \\
1 & 0 \\
2 & 0 \\
3 & 1 \\
4 & 0 \\
5 & 0 \\
6 & 1 \\
7 & 1 \\
\end{array}
\]

\[
\begin{array}{c|c}
 t & s_t \\
8 & \left\lfloor \frac{2}{3} \right\rfloor \oplus \left\lfloor \frac{3}{5} \right\rfloor = 0 + 0 = 0 \\
9 & 1 \\
10 & 0 \\
11 & \left\lfloor \frac{2}{3} \right\rfloor \oplus \left\lfloor \frac{1}{5} \right\rfloor = 0 + 1 = 1 \\
12 & 1 \\
13 & \left\lfloor \frac{1}{3} \right\rfloor \oplus \left\lfloor \frac{3}{5} \right\rfloor = 1 + 0 = 1 \\
14 & \left\lfloor \frac{1}{3} \right\rfloor \oplus \left\lfloor \frac{4}{5} \right\rfloor = 0 + 1 = 1 \\
\end{array}
\]

\[
\{s_t\} = 000100110101111
\]
The merit factor of Jacobi sequences

Theorem

Let \( \{s_t\} \) be a modified Jacobi sequence of length \( pq \), where \( p \) and \( q \) are distinct primes. If \( \frac{(p+q)^5 \log^4(n)}{n^3} \to 1 \) as \( n \to \infty \)

\[
\frac{1}{F} = \frac{2}{3} - 4|f| + 8f^2, \quad |f| \leq \frac{1}{2}
\]
Skew-symmetric sequences

A common strategy for extending the reach of merit factor computations is to impose restrictions on the structure of the sequence.

A skew-symmetric binary sequence is a binary sequence \((a_0, a_1, \ldots, a_{2m})\) of odd length \(2m + 1\) where

\[
a_{m+i} = \begin{cases} a_{m-i}, & i \text{ even, } 1 \leq i \leq m \\ \overline{a_{m-i}}, & i \text{ odd, } 1 \leq i \leq m \end{cases}
\]

Example

Let \(\{s_t\}\) be the Barker sequence of length 7, \(\{s_t\} = 1110010\). As \(7 = 2 \times 3 + 1\), we have that \(m = 3\). If the sequence is skew-symmetric, we must have that

\[
\begin{array}{cccccccc}
s_{m-3} & s_{m-2} & s_{m-1} & s_m & s_{m+1} & s_{m+2} & s_{m+3} \\
s_0 & s_1 & s_2 & s_3 & \overline{s_2} & s_1 & \overline{s_0} \\
1 & 1 & 1 & 0 & 0 & 1 & 0
\end{array}
\]
All odd-length Barker sequences are skew-symmetric.

Skew-symmetric sequences are known to attain the optimal merit factor value $F_n$ for most odd values of $n < 60$.

**Proposition**
A skew-symmetric binary sequence of odd length has $C(u) = 0$ for all odd $u$.

**Proposition**
The mean value of $1/F$, taken over all skew-symmetric sequences of odd length $n$, is $\frac{(n-1)(n-2)}{n^2}$. 
Periodic appending

Given a sequence $A = (a_0, a_1, \ldots, a_{n-1})$ of length $n$ and a real number $k$ satisfying $0 \leq k \leq 1$, we write $A^k$ for the sequence $(b_0, b_1, \ldots, b_{\lfloor kn \rfloor - 1})$ obtained by truncating $A$ to a fraction $k$ of its length, that is, $b_i = a_i$ for $0 \leq i < \lfloor kn \rfloor$.

Let $X_r$ be a Legendre sequence of prime length $n$, offset by a fraction $r$ of $n$. We have previously seen that $\limsup_{n \to \infty} F(X_{\frac{1}{4}}) = 6$.

Extensive numerical evidence suggests that

- for large $n$, the merit factor of the appended sequence $X_{\frac{1}{4}} : (X_{\frac{1}{4}})^k$ is greater than 6.2 when $k \approx 0.03$
- for large $n$, the merit factor of the appended sequence $X_r : (X_r)^k$ is greater than 6.34 for $r \approx 0.22$ and $r \approx 0.72$ when $k \approx 0.06$. 
Stochastic search algorithms

The basic method of many of the stochastic search algorithms is to move through the search space of sequences by changing only one, or sometimes two, sequence elements at a time. The merit factor of any close neighbour sequence can be computed in $O(n)$ operations from knowledge of the aperiodic autocorrelations of the current sequence. The search algorithm specifies when it is acceptable to move to a neighbour sequence, for example when it has merit factor no smaller than the current sequence, or when it has the largest merit factor amongst all close neighbours not previously visited. The search algorithm must also specify how to choose a new sequence when no acceptable neighbour sequence can be found.

So far no stochastic search algorithm has been found that reliably produces binary sequences with merit factor greater than 6 in reasonable time for large $n$. 
The end