Low Autocorrelation Binary Sequences, LABS

Ilias S. Kotsireas

Wilfrid Laurier University
Waterloo ON, Canada

`ikotsire@wlu.ca`

`Director, CARGO Lab, http://www.cargo.wlu.ca/`

Outline of the talk

- Autocorrelation (periodic and aperiodic)
- Unified description of combinatorial objects
- LABS, Merit Factor
- Algorithms
- Computational Results
- Future work, open probelm

Autocorrelation (periodic and aperiodic)

- The **periodic autocorrelation function** associated to a finite sequence $A = [a_0, \ldots, a_{n-1}]$ of length n is defined as

  $$P_A(s) = \sum_{k=0}^{n-1} a_k a_{k+s}, \;\; s = 0, \ldots, n-1,$$

  where $k + s$ is taken modulo n, when $k + s > n$.

- The **aperiodic autocorrelation function** associated to a finite sequence $A = [a_0, \ldots, a_{n-1}]$ of length n is defined as

  $$N_A(s) = \sum_{k=0}^{n-1-s} a_k a_{k+s}, \;\; s = 0, \ldots, n-1,$$

We are mostly concerned with binary $\{-1, +1\}$, ternary $\{-1, 0, +1\}$ and quaternionic $\{\pm 1, \pm i\}$ sequences.

Note that for sequences with complex number elements, a_{k+s} is replaced by $\overline{a_{k+s}}$.

Example: $n = 7$, $A = [a_1, \ldots, a_7]$

$$\begin{aligned}
P_A(0) &= {a_1}^2 + {a_2}^2 + {a_3}^2 + {a_4}^2 + {a_5}^2 + {a_6}^2 + {a_7}^2 \\
P_A(1) &= a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_6 + a_6a_7 + a_7a_1 \\
P_A(2) &= a_1a_3 + a_2a_4 + a_3a_5 + a_4a_6 + a_5a_7 + a_6a_1 + a_7a_2 \\
P_A(3) &= a_1a_4 + a_2a_5 + a_3a_6 + a_4a_7 + a_5a_1 + a_6a_2 + a_7a_3 \\
P_A(4) &= a_1a_4 + a_2a_5 + a_3a_6 + a_4a_7 + a_5a_1 + a_6a_2 + a_7a_3 \\
P_A(5) &= a_1a_3 + a_2a_4 + a_3a_5 + a_4a_6 + a_5a_7 + a_6a_1 + a_7a_2 \\
P_A(6) &= a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_6 + a_6a_7 + a_7a_1
\end{aligned}$$

$$\begin{aligned}
N_A(0) &= {a_1}^2 + {a_2}^2 + {a_3}^2 + {a_4}^2 + {a_5}^2 + {a_6}^2 + {a_7}^2 \\
N_A(1) &= a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_6 + a_6a_7 \\
N_A(2) &= a_1a_3 + a_2a_4 + a_3a_5 + a_4a_6 + a_5a_7 \\
N_A(3) &= a_1a_4 + a_2a_5 + a_3a_6 + a_4a_7 \\
N_A(4) &= a_1a_5 + a_2a_6 + a_3a_7 \\
N_A(5) &= a_1a_6 + a_2a_7 \\
N_A(6) &= a_1a_7
\end{aligned}$$

Unified description of combinatorial objects

number/type of sequences	defining property	name
2 binary	aper. autoc. 0	Golay sequences
2 binary	per. autoc. 0	Hadamard matrices
2 binary	per. autoc. 2	D-optimal matrices
2 binary	per. autoc. -2	Hadamard matrices
2 ternary	aper. autoc. 0	TCP
2 ternary	per. autoc. 0	Weighing matrices
3 binary	aper. autoc. const.	Normal sequences
4 binary	aper. autoc. 0	Base sequences
4 binary	aper. autoc. 0	Turyn type sequences
4 ternary	aper. autoc. 0	T-sequences
$2 \ldots 12$ binary	per. autoc. zero	PCS

Complementary Sequences

Definition:

Let $\{A_i\}_{i=1,\dots,t}$ be t sequences of length v with complex elements. The sequences $\{A_i\}_{i=1,\dots,t}$ are called complementary, if

$$\sum_{i=1}^{t} PAF_{A_i} = [\alpha_0,\ \underbrace{\alpha,\dots,\alpha}_{v-1 \text{ terms}}\ ]$$

with the convention:

$$PAF_{A_i} = [PAF_{A_i}(0), PAF_{A_i}(1), \dots, PAF_{A_i}(v-1)].$$

Optimization formalism

The search for complementary sequences can be formulated as an optimization problem, via the concept of the PAF.

There are optimization algorithms that deal with problems with $20K$ (discrete) variables.

We need **symmetric matrices** and certain vector/matrix products

$$\min_{x \in \{0,1\}^n} x^T A x$$

Let $a = [a_1, a_2, \ldots, a_n]^T$ be a column $n \times 1$ vector, where $a_1, a_2, \ldots, a_n \in \{-1, +1\}$ and consider the elements of the PAF vector $P_A(1), \ldots, P_A(m)$. Define the following $m = [n/2]$ symmetric matrices (which are independent of the sequence a)

$$M_i = (m_{jk}), \text{ s.t. } \begin{cases} m_{jk} = m_{kj} = \frac{1}{2}, & \text{when } a_j a_k \in P_A(i),\ j, k \in \{1, \ldots, n\} \\ 0, & \text{otherwise} \end{cases}, i = 1, \ldots, m$$

Lemma

The matrices M_i can be used to write the PAF equations in a matrix form:

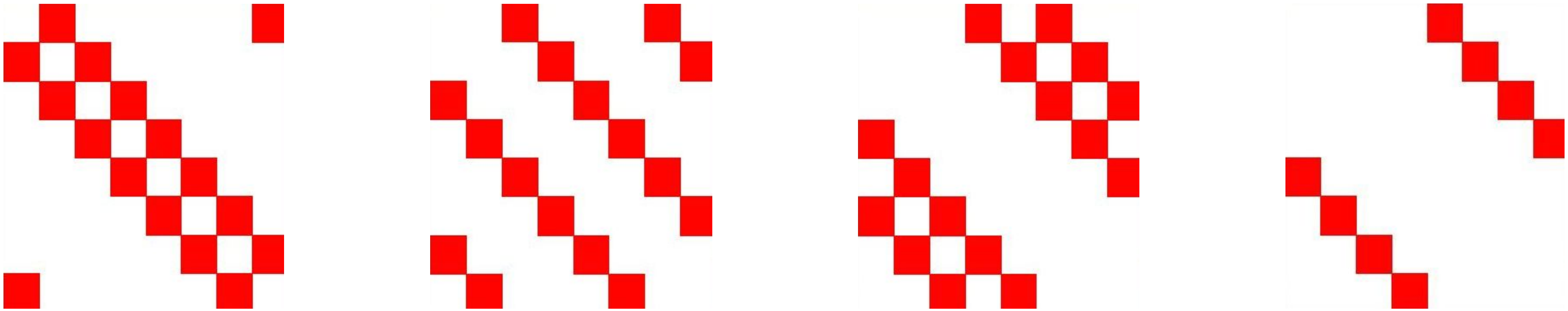
- for n odd: $a^T M_i a = P_A(i),\ \ i = 1, \ldots, m.$
- for n even: $a^T M_i a = P_A(i),\ \ i = 1, \ldots, m-1$ and $a^T M_m a = \frac{1}{2} P_A(m).$

Example

Let $n = 8$, $a = [a_1, \ldots, a_8]$. Then we have that $m = 4$ and

$$a^T M_i a = P_A(i),\ \ i = 1, 2, 3 \text{ and } a^T M_4 a = \frac{1}{2} P_A(4)$$

(1) Ilias S. Kotsireas, Panos M. Pardalos, Oleg V. Shylo et al.
Periodic complementary binary sequences and combinatorial optimization algorithms. J. Comb. Optim. 20 (2010), no. 1, 63-75.
(2) Ilias S. Kotsireas, Panos M. Pardalos
D-optimal matrices via quadratic integer optimization.
J. Heuristics 19(4), (2013) 617--627.



Graphical representations of the four symmetric matrices M_1, M_2, M_3, M_4

Problem I Now suppose that we are looking for two $\{-1, +1\}$ sequences A and B of lengths n, such that

$$P_A(i) + P_B(i) = 2, \;\; i = 1, \ldots, m.$$

Via the previous lemma we can reformulate this problem as follows:

Problem II Find two binary sequences a, b, (viewed as $n \times 1$ column vectors) such that

$$a^T M_i a + b^T M_i b = 2, \;\; i = 1, \ldots, m.$$

The LABS problem I

- J. Phys. A, 29 (1996), S. Mertens
- J. Phys. A, 49 (2016), T. Packebusch & S. Mertens

communication engineering problems:

Binary $\{\pm 1\}$ sequences $S = \{s_1, \ldots, s_N\}$ with low off-peak autocorrelations

$$C_k(S) = \sum_{i=1}^{N-k} s_i s_{i+k}, \quad k = 1, \ldots, N-1$$

Physics

Consider binary sequences as one-dimensional systems of **Ising-spins**. In this context, low-autocorrelation binary sequences appear as minima of the **energy**

$$E(S) = \sum_{k=1}^{N-1} C_k^2(S) \qquad \text{Bernasconi model}$$

J. Bernasconi, J. Physique, 48, (1987)
Low autocorrelation binary sequences: statistical mechanics and configuration space analysis

The LABS problem II

With periodic autocorrelation, instead of aperiodic, the construction of ground states is possible for special values of N:

for $N = 4n + 3$ prime, the **modified Legendre sequence** yields $C_k^2 = 1$, the minimum possible value for odd N

Other ground states can be constructed from **linear shift register sequences** based on **primitive polynomials** over **Galois fields**.

For the ground states of the Bernasconi model, no construction is known, even for special values of N. Ground states are highly disordered!

The Legendre sequences are far from the true ground states

M. J. E. Golay, 1983 The merit factor of Legendre sequences IEEE Trans. Inf. Theory

The only exact results have been provided by exhaustive enumerations

Restricted to systems: $N \leq 32$, (1996) $N \leq 66$, (2016) **exponential complexity**

Partial enumerations allow larger values of N but are **not guaranteed** to yield true ground states.

Merit Factor

Finding the ground states of the Bernasconi model has turned out to be a hard mathematical problem.

Golay has conjectured that the maximal **merit factor** $F = \frac{N^2}{2E}$ should obey $F < 12.32$ for large N

However, heuristic searches (SA) among **skew-symmetric** (odd N, $s_{n+l} = (-1)^l s_{n-l}, l = 1, \ldots, n-1$) sequences up to $N \leq 199$ suggest $F < 6$ for long sequences

Beenker G F M, Claasen T A C M and Hermens P W C, 1985 Binary sequences with a maximally flat amplitude spectrum Philips J. Res. 40 289-304

This large discrepancy indicates that the ground states, i.e. the sequences with high merit factors $6 < F < 12$, must be **extremely isolated energy minima** in configuration space (boolean cube).

Exhaustive search seems to be the only approach!

Algorithms

- Any algorithm that performs an exhaustive search for the ground state of the Bernasconi model, has to cope with the enormous size 2^N of the configuration space
- Methods to restrict the search to smaller subspaces without missing the true ground state
- Symmetries can reduce the size of the search space by a factor of about an eighth
- combinatorial optimization: branch and bound
- parallelization

Symmetries

- The correlations $C_k(1)$ are unchanged when the sequence is complemented or reversed
- When alternate elements of the sequence are complemented, the even-indexed correlations are not affected, the odd-indexed correlations only change sign
- Hence, with the exception of a small number of symmetric sequences, the 2^N sequences will come in classes of eight which are equivalent.
- The total number of non-equivalent sequences is slightly larger than 2^{N-3}
- The m left-most and m right-most elements of the sequence can be used to parametrize the symmetry-classes

Branch and bound

- Branch and bound methods are commonly used in combinatorial optimization
- They solve a discrete optimization problem by breaking up its feasible set into successively smaller subsets (branch), calculating bounds on the objective function value over each subset and using these to discard certain subsets from further consideration (bound)
- The procedure ends when each subset has either produced a feasible solution, or been shown to contain no better solution that already in hand
- The best solution found during this procedure is a global optimum
- The idea is of course to discard many subsets as early as possible during the branching process, i.e. to discard most of the feasible solutions before actually evaluating them
- The success of this approach, which depends upon the branching rule and (very heavily) upon the quality of the bounds, can be quite dramatic

Branch and bound for LABS

- In accordance with our symmetry classes, we specify a set of feasible solutions by fixing the m left-most and m right-most elements of the sequence
- The N-2m centre elements are not specified, i.e. the set contains 2^{N-2m} feasible solutions
- Given a feasible set specified by the m border elements, four smaller sets are created by fixing the elements s_{m+1} and s_{N-m} to $\{\pm 1\}$
- This is the branching rule. It is applied recursively until all elements have been fixed
- The energy of the resulting sequence is compared with the minimum energy found so far
- If it is lower, the sequence is kept as the potential ground state
- After all $c(m)2^{N-2m}$ sequences have been considered, the potential ground state has turned into a true one

Relaxation for LABS

- Lower bounds are usually obtained by replacing the original problem over a given subset with an easier (relaxed) problem such that the solution value of the latter bounds that of the former
- A good relaxation is one that (i) is easy and fast to solve and (ii) yields strong lower bounds. (Most often these are conflicting goals)
- A relaxation for the LABS problem is given by adjusting the free elements (i.e. the centre elements $s_{m+1}, \ldots, s_{N-m}$) to minimize all values C_k^2 **separately**
- i.e. we replace the original problem

$$E_{min} = \min_{subset} \left(\sum_{k=1}^{N-1} C_k^2 \right)$$

by the relaxed version

$$E^{\star}_{min} = \sum_{k=1}^{N-1} \min_{subset} (C_k^2) \leq E_{min}$$

- $O(2^N) \rightsquigarrow O(1.85^N)$

```
# Ground states of the Bernasconi model with open boundary conditions
# ===================================================================
#
# (C) Copyright 1996-2001 S. Mertens, ITP, University of Magdeburg
#     stephan.mertens@physik.uni-magdeburg.de
#
# See J.Phys.A 29 L473 (1996) for the algorithm used.
# The values for N>48 have been found with an improved implementation due
# to Heiko Bauke (heiko.bauke@physik.uni-magdeburg.de)
#
# Configurations are given in run length notation, i.e. each figure indicates
# the number of consecutive elements with the same sign:
#
#          2 5 2 2 1 1 1 2 1 = --+++++--++-+-++-
#
# N   Emin      configuration
  3       1     2 1
  4       2     2 1 1
  5       2     3 1 1
  6       7     1 1 1 3
  7       3     1 1 2 3
  8       8     1 2 1 1 3
  9      12     4 2 1 1 1
 10      13     2 2 1 1 4
 11       5     1 1 2 1 3 3
 12      10     1 2 2 1 1 1 4
 13       6     5 2 2 1 1 1 1
 14      19     2 2 2 1 1 1 5
 15      15     5 2 2 2 1 1 1 1
 16      24     2 2 5 1 1 1 1 2 1
 17      32     2 5 2 2 1 1 1 2 1
 18      25     4 4 1 1 1 2 2 2 1
 19      29     4 1 1 1 1 4 2 2 1 2
 20      26     5 1 1 3 1 1 2 3 2 1
 21      26     2 7 2 2 1 1 1 1 1 2 1
 22      39     5 1 2 2 1 1 1 1 2 3 3
 23      47     2 1 2 1 2 1 1 1 1 6 3 2
 24      36     2 2 3 6 1 1 1 1 1 2 1 2 1
 25      36     3 3 7 1 1 1 1 2 1 2 2 1
 26      45     2 1 2 1 2 1 1 1 1 1 6 3 2 2
 27      37     3 4 3 1 3 1 3 1 2 1 1 2 1 1
 28      50     3 4 3 1 3 1 3 1 2 1 1 2 1 2
 29      62     2 1 2 1 1 2 1 3 1 3 1 3 4 3 1
 30      59     5 5 1 2 1 2 1 1 1 1 1 3 2 3 1
 31      67     7 3 3 2 2 1 2 2 1 1 1 1 2 1 1 1
 32      64     7 1 1 1 2 1 1 1 1 3 3 2 2 1 2 2 1
 33      64     7 4 2 1 1 2 1 1 1 1 1 1 1 2 2 2 2 1
 34      65     8 4 2 1 1 2 1 1 1 1 1 1 1 2 2 2 2 1
 35      73     7 1 2 2 1 2 2 1 1 1 1 2 1 1 1 1 3 3 2
 36      82     3 6 3 2 3 1 1 1 3 1 2 1 2 1 1 1 2 1 1
 37      86     8 4 4 2 1 1 2 1 1 1 1 1 1 1 2 2 2 2 1
 38      87     8 4 4 2 1 1 2 1 1 1 1 1 1 1 1 2 2 2 2 1
 39      99     8 2 1 2 1 1 2 1 2 3 4 3 2 1 1 1 1 1 1 1
 40     108     4 4 4 1 2 1 1 2 1 3 1 1 2 1 3 1 3 1 3 1
 41     108     3 4 3 1 1 1 1 1 1 2 2 2 2 8 1 2 1 1 2 1 1
 42     101     3 1 3 1 3 1 3 4 1 3 4 3 1 1 2 1 1 2 1 1 2
 43     109     1 1 3 2 4 3 2 1 1 1 1 1 7 2 1 2 1 1 2 2 1 3
 44     122     5 2 5 3 1 3 1 1 3 1 1 1 2 2 2 1 1 1 2 1 1 1 2 1
 45     118     8 2 1 2 1 1 2 1 2 3 1 2 3 4 3 2 1 1 1 1 1 1 1
 46     131     8 2 3 4 3 1 2 3 1 2 1 1 2 1 2 2 1 1 1 1 1 1 1 1
```

```
  47     135      9 2 3 4 3 1 2 3 1 2 1 1 2 1 2 2 1 1 1 1 1 1 1 1
  48     140      3 1 1 1 1 1 1 8 3 2 1 4 3 2 1 2 2 2 1 1 2 1 1 2 1
  49     136      2 1 5 1 3 1 3 1 1 2 2 4 1 1 2 2 4 1 1 4 1 1 4 1
  50     153      2 1 5 1 3 1 3 1 1 2 2 4 1 1 2 2 4 1 1 4 1 1 4 2
  51     153      2 3 4 3 2 1 1 1 1 4 1 3 1 3 1 1 6 2 1 2 1 1 2 1 2 1
  52     166      5 1 1 6 1 2 1 2 1 2 1 1 1 1 1 3 1 2 2 3 1 2 3 3 3 2
  53     170      4 5 1 1 3 1 1 1 3 3 2 5 1 3 1 2 2 2 1 1 1 2 1 1 1 1 2 1
  54     175      3 5 6 2 2 5 1 4 1 2 1 2 1 1 2 2 2 2 1 1 1 1 1 1 1 2 1
  55     171      9 2 1 2 1 2 3 2 1 2 1 1 4 3 2 1 2 3 3 2 1 1 1 1 1 1 1 1
  56     192      7 6 1 2 2 3 1 1 2 3 2 4 1 1 1 1 1 3 2 1 1 2 1 2 2 1 1 1
  57     188      3 3 2 3 2 6 3 1 1 1 1 1 2 7 1 2 1 1 1 1 2 2 1 2 2 1 2 1 1
  58     197      1 1 1 1 1 3 1 2 3 2 1 3 8 1 4 2 1 2 1 1 3 2 4 3 2 1 1 2
  59     205      7 7 2 4 1 2 2 4 2 1 1 2 2 3 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1
  60     218      7 6 1 1 1 2 1 4 1 1 1 1 1 3 1 1 2 4 2 1 1 3 2 2 2 1 1 2 2 2
```

Future work, open problems

- find the ground states for $N = 67, 68, 69, 70, \ldots$
- improve parallelization (symmetry classes)
- improve and further optimize algorithms implementations
- exploit/adapt the comb. optim. formulation of Pardalos et al.
- convert LABS to linear 0-1 and use cplex/gurobi, joint work with Pardalos

Is this now the limit of what we can do? it may very well be, but certainly advances will not be made by people who think they cannot succeed.

– Carl Pomerance