応用数学特論 II (集中講義)

Day 5 Cyclic Designs and Their Applications

盧 暁南 (山梨大学)
Xiao-Nan LU (University of Yamanashi)

Aug. 31, 2021
Kobe University

Outline

1. Binary sequences with optimal autocorrelation and cyclic designs
2. Statistical designs for fMRI experiments and cyclic almost orthogonal arrays (CAOA)
3. CAOAs, optimal sequences, and ADSs

Binary periodic sequences

Definition (binary periodic sequences with period n)

A binary sequence (系列) $\mathbf{s} = (s_0, s_1, \ldots, s_{n-1}, s_n, s_{n+1}, \ldots)$ is said to be periodic with period (周期) n if

$$s_i = s_{i+n} \qquad \text{for any } i \geq 0.$$

- A binary periodic sequence $\mathbf{s}$ with period n can be seen as a binary string of length n, say, $\mathbf{s} = (s_t) \in \{0, 1\}^n$.
- In this lecture, we say $\mathbf{s}$ is a binary sequence of length n.

Autocorrelation magnitude of binary sequences

Definition (Autocorrelation of binary sequences)

(n=4 の例)

The periodic autocorrelation (周期的自己相関関数) of a binary sequence $\mathbf{s} = (s_t) \in \{0,1\}^n$ at shift w is defined by

$$\rho_{\mathbf{s}}(w) = \sum_{t=0}^{n-1} (-1)^{s_{t+w} - s_t},$$

where the subscript of s_{t+w} is reduced by modulo n.

1 1 1 0 | 1 1 1 0 ....

1 | 1 1 0 1 | 1 1 0

0 0 1 -1

1 1 -1 -1

- The maximum absolute value of the off-peak autocorrelation $\max_{w \neq 0} |\rho_{\mathbf{s}}(w)|$ is called the autocorrelation magnitude (自己相関のマグニチュード) of $\mathbf{s}$.
- Sequences with low autocorrelation magnitudes are desired.

Example: Jupyper Notebook

Perfect sequence

Definition (Perfect sequences)

A binary sequence $\mathbf{s} = (s_t) \in \{0,1\}^n$ with autocorrelation magnitude 0 is called a perfect sequence (完全系列).

Proposition

A binary sequence of length n is a perfect sequence $\implies n \equiv 0 \pmod 4$.

Example ($n = 4$)

0001

$\mathbf{s} = (1110)$ is a perfect sequence. This is the only known binary perfect sequence (up to equivalence of cyclic shift).

Binary sequences with optimal autocorrelation magnitude

1. For $n \equiv 0 \pmod 4$, $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| \geq 0$. There is only one example (perfect sequence) with $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| = 0$.
   So, it is natural to consider $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| = 4 \iff \rho_{\mathbf{s}}(w) \in \{0, 4\}$ or $\{0, -4\}$ for all $1 \leq w \leq n-1$; in this case, the sequence is said to have optimal autocorrelation.
2. For $n \equiv 3 \pmod 4$, $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| \geq 1$. Moreover, $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| = 1 \iff \rho_{\mathbf{s}}(w) = -1$ for all $1 \leq w \leq n-1$; in this case, the sequence is said to have optimal autocorrelation.
3. For $n \equiv 1 \pmod 4$, $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| \geq 1$. But there is some evidence that there is no binary sequence of length $n > 13$ with $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| = 1$.
   So, it is natural to consider $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| = 3 \iff \rho_{\mathbf{s}}(w) \in \{1, -3\}$ for all $1 \leq w \leq n-1$; in this case, the sequence is said to have optimal autocorrelation.
4. For $n \equiv 2 \pmod 4$, $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| \geq 2$.
   Moreover, $\max_{w\neq 0} |\rho_{\mathbf{s}}(w)| = 2 \iff \rho_{\mathbf{s}}(w) \in \{2, -2\}$ for all $1 \leq w \leq n-1$; in this case, the sequence is said to have optimal autocorrelation.

Example of sequences with optimal autocorrelation

Example (A bad example)

- $\mathbf{s} = (1101000101)$
- For $w = 5$, we have $(s_{t+w} - s_t)_t = (-1, -1, 1, -1, 1, 1, 1, -1, 1, -1)$.

$$\rho_{\mathbf{s}}(w) = \sum_{t=0}^{n-1} (-1)^{s_{t+w} - s_t} = -10 \notin \{2, -2\}$$

Example (A good example)

- $\mathbf{s} = (1000101101)$
- For $w = 5$, we have $(s_{t+w} - s_t)_t = (-1, 1, 1, 0, 0, 1, -1, -1, 0, 0)$.

Optimal binary sequences with one valued autocorrelation

1. For $n \equiv 0 \pmod 4$, there are only two examples ($n = 8, 40$) with $\rho_{\mathbf{s}}(w) = 4$ for all $1 \leq w \leq n-1$. It is proved that no other exists. [1]
2. For $n \equiv 3 \pmod 4$, there are infinite many examples with $\rho_{\mathbf{s}}(w) = -1$ for all $1 \leq w \leq n-1$ (see the next slide).
3. For $n \equiv 1 \pmod 4$, there are only two examples ($n = 5, 13$) with $\rho_{\mathbf{s}}(w) = 1$ for all $1 \leq w \leq n-1$. It is conjectured that no other exists (verified true for $13 < n < 101701$ except possibly for $n \in \{29525, 30013, 34061\}$).
4. For $n \equiv 2 \pmod 4$, there is only one examples ($n = 6$) with $\rho_{\mathbf{s}}(w) = 2$ for all $1 \leq w \leq n-1$. It is conjectured that no other exists (verified true for $6 < n < 33895686$).

Example: ▸▸ Jupyper Notebook

$n = 10^5$ のとき, $\# seq = 2^{10^5}$

[1] X. Niu, H. Cao, and K. Feng. Binary periodic sequences with 2-level autocorrelation values. Discrete Math. 343(3): 111723 (2020).

Combinatorial designs and optimal binary sequences

- Let $\mathbf{s}$ be a binary sequence of length n.
- Let $\mathrm{supp}(\mathbf{s})$ denote the support of $\mathbf{s}$, i.e., $\mathrm{supp}(\mathbf{s}) = \{1 \leq i \leq n-1 : s_i = 1\}$.

Theorem (Difference sets (DS) & almost difference sets (ADS) $\iff$ optimal sequences)

1. *For* $n \equiv 3 \pmod 4$, $\rho_{\mathbf{s}}(w) = -1$ *for all* $1 \leq w \leq n-1 \iff \mathrm{supp}(\mathbf{s})$ *is an* $(n, (n+1)/2, (n+1)/4)$ *or* $(n, (n-1)/2, (n-3)/4)$ *DS in* $\mathbb{Z}_n$.
2. *For* $n \equiv 1 \pmod 4$, $\rho_{\mathbf{s}}(w) \in \{1, -3\}$ *for all* $1 \leq w \leq n-1 \iff \mathrm{supp}(\mathbf{s})$ *is an* $(n, k, k-(n+3)/4, (n-k)k-(n-1)^2/4)$ *ADS in* $\mathbb{Z}_n$.
3. *For* $n \equiv 0 \pmod 4$, $\rho_{\mathbf{s}}(w) \in \{0, -4\}$ *for all* $1 \leq w \leq n-1 \iff \mathrm{supp}(\mathbf{s})$ *is an* $(n, k, k-(n+4)/4, (n-k)k-n(n-1)/4)$ *ADS in* $\mathbb{Z}_n$.
4. *For* $n \equiv 2 \pmod 4$, $\rho_{\mathbf{s}}(w) \in \{2, -2\}$ *for all* $1 \leq w \leq n-1 \iff \mathrm{supp}(\mathbf{s})$ *is an* $(n, k, k-(n+2)/4, (n-k)k-(n-1)(n-2)/4)$ *ADS in* $\mathbb{Z}_n$.

Cyclic difference sets

$= \{0, 1, \ldots, n-1\} \bmod n$

- For $D = \{b_1, \ldots, b_k\} \subseteq \mathbb{Z}_n$ and nonzero $x \in \mathbb{Z}_n$,

$$\Delta_x(D) := \{(b_i, b_j) \in D \times D : b_i - b_j = x\}.$$

Cyclic difference set (DS) $\longleftrightarrow$ (n, k, λ) Difference family (DF) with 1 block

A k-subset $D \subseteq \mathbb{Z}_n$ is an (n, k, λ) difference set (DS) in $\mathbb{Z}_n$ if $|\Delta_x(D)| = \lambda$ for any nonzero $x \in \mathbb{Z}_n$.

- Paley-type DS: $(4m-1, 2m-1, m-1)$-DS [Hadamard matrix H_{4m}]
- Singer DS: $(\frac{q^{m+1}-1}{q-1}, \frac{q^m-1}{q-1}, \frac{q^{m-1}-1}{q-1})$-DS [finite projective geometry PG$(m, \mathbb{F}_q)$]

Cyclic almost difference sets: Definition

- For $D = \{b_1, \dots, b_k\} \subseteq \mathbb{Z}_n$ and nonzero $x \in \mathbb{Z}_n$,

$$\Delta_x(D) := \{(b_i, b_j) \in D \times D : b_i - b_j = x\}.$$

Cyclic almost difference set (ADS) 準差集合

A k-subset $D \subseteq \mathbb{Z}_n$ is an (n, k, λ, t) almost difference set (ADS) in $\mathbb{Z}_n$ if $|\Delta_x(D)| = \lambda$ for any $x \in X$ and $|\Delta_y(D)| = \lambda + 1$ for any $y \in Y$, where $X \cup Y = \mathbb{Z}_n \setminus \{0\}$ with $|X| = t$.

(n, k, λ)-DS $\Leftrightarrow$ $(n, k, \lambda, n-1)$-ADS

- Paley-type DS: $(4m-1, 2m-1, m-1, 4m-2)$-ADS.
- In this lecture, I will focus on $(4m-2, 2m-1, m-1, 3m-2)$-ADS.
- Motivated by sequences and (recent work on) experimental designs.

Cyclic almost difference sets: Example

Example: $(10, 5, 2, 7)$-ADS in $\mathbb{Z}_{10}$

- $D = \{0, 4, 6, 7, 9\} \subseteq \mathbb{Z}_{10}$.
- $X = \{\pm 1, \pm 2, \pm 4, 5\}$, $Y = \{\pm 3\}$.

$\lambda = 2$ $\quad\quad$ $\lambda + 1 = 3$

$$
\begin{aligned}
\Delta_1(D) &= \{(0,9), (7,6)\}, \\
\Delta_2(D) &= \{(6,4), (9,7)\}, \\
\Delta_3(D) &= \{(7,4), (9,6), (0,7)\}, \\
\Delta_4(D) &= \{(4,0), (0,6)\}, \\
\Delta_5(D) &= \{(9,4), (4,9)\}.
\end{aligned}
$$

$\lambda, \lambda+1$ しかない

$\Updownarrow$

$P_S(w) = 2, -2$ しかない

$\Delta_6(D) = \Delta_{-4}(D)$

Cyclic almost difference sets: Constructions

Known constructions for cyclic ADS with $n \equiv 2 \pmod 4$ and $k = n/2$

There exits an $(n, \frac{n}{2}, \frac{n-2}{4}, \frac{3n-2}{4})$-ADS for the following n:

1. (SLCE[2] [3]) $n = q - 1$, where $q \equiv 3 \pmod 4$ is a prime power;
2. (DHM[4]) $n = 2p$, where $p \equiv 5 \pmod 8$ is a prime and $p - 1$ or $p - 4$ is a perfect square.
3. $n \in \{34, 38, 50\}$ by computer search.

[2] V. M. Sidelnikov: Some k-valued pseudo-random sequences and nearly equidistant codes. Probl. Peredachi Inf. 5(1):16–22 (1969).

[3] A. Lempel, M. Cohn, W. Eastman. A class of balanced binary sequences with optimal autocorrelation properties. IEEE Trans. Inform. Theory, IT-23 (1), 38–42 (1977).

[4] C. Ding, T. Helleseth, H. Martinsen: New families of binary sequences with optimal three-level autocorrelation. IEEE Trans. Inform. Theory, 47(1): 428–433 (2001).

SLCE sequences

$\mathbb{F}_q^* = \{g^0, g^1, \ldots, g^{q-2}\}$, $C_0^{(2,q)} = \{g^0, g^2, g^4, \ldots, g^{q-3}\}$

$=$ $\mathbb{F}_q^*$上の平方元の集合 (subgrp)

- g: generator of $\mathbb{F}_q^*$
- $C_1^{(2,q)} = \{g^{2t+1} : 0 \le t < (q-1)/2\}$.

$C_1^{(2,q)} = g \cdot C_0^{(2,q)} = \mathbb{F}_q^*$上の非平方元集合

Theorem (SLCE sequence)

Let $q \equiv 3 \pmod 4$ be a prime power and g be a primitive element of $\mathbb{F}_q$. Let D be the subset of $\mathbb{Z}_{q-1}$ defined by $\log_g(C_1^{(2,q)} - 1)$, where $\log_g$ denotes the discrete logarithm in $\mathbb{F}_q$ to the base g. Then D is a $(q-1, \frac{q-1}{2}, \frac{q-3}{4}, \frac{3q-5}{4})$ ADS in $\mathbb{Z}_{q-1}$. In other words, the corresponding binary sequence of D of length $q-1$ is perfect and balanced.

The admissible prime powers $q < 100$ for SLCE sequences are

$$7,\ 11,\ 19,\ 23,\ 27,\ 31,\ 43,\ 47,\ 59,\ 67,\ 71,\ 79,\ 83.$$

An example of SLCE sequences

Table: The cyclic group $\mathbb{F}_{11}^*$ generated by $g = 2$

i	0	1	2	3	4	5	6	7	8	9
2^i mod 11	1	2	4	8	5	10	9	7	3	6

Example (A SLCE type $(10, 5, 2, 7)$ ADS in $\mathbb{Z}_{10}$)

Take $g = 2$ as a generator of $\mathbb{F}_{11}^*$. Then

$$C_1^{(2)} = \{2^1, 2^3, 2^5, 2^7, 2^9\} \equiv \{2, 8, 10, 7, 6\} \pmod{11}.$$

$g=2 \in \mathbb{F}_q^*$ (离散对数)

Let

$$D = \log_2(C_1^{(2)} - 1) = \log_2\{1, 7, 9, 6, 5\} = \{0, 7, 6, 9, 4\},$$

where $\log_2$ denotes the discrete logarithm in $\mathbb{F}_{11}$ to the base 2.

DHM sequences

$C_i^{(4,p)} = \{ g^{4t+i} : 0 \le t < \frac{p-1}{4} \}$

Theorem (DHM sequences)

Let n *be a positive integer such that* $n = 2p$ *with prime* $p \equiv 5 \pmod 8$*. Let* $i, j, l \in \{0, 1, 2, 3\}$ *be three distinct integers, and let*

$$C_0 = C_i^{(4,p)} \cup C_j^{(4,p)} \qquad \text{and} \qquad C_1 = C_j^{(4,p)} \cup C_l^{(4,p)}.$$

Then,

$$D = (\{0\} \times C_0) \cup (\{1\} \times C_1) \cup \{(0,0)\}$$

is an $(n, \frac{n}{2}, \frac{n-2}{4}, \frac{3n-2}{4})$ *ADS in* $\mathbb{Z}_2 \times \mathbb{Z}_p$ *(isomorphic to* $\mathbb{Z}_{2p}$*) if the generator of* $\mathbb{Z}_p^*$ *is properly chosen for the cyclotomic classes and*

1. $p - 4$ *is a perfect square and* $(i, j, l) \in \{(0,1,3), (0,2,3), (1,2,0), (1,3,0)\}$ *or*
2. $p - 1$ *is a perfect square and* $(i, j, l) \in \{(0,1,2), (0,3,2), (1,0,3), (1,2,3)\}$.

An example of DHM sequences

Table: The cyclic group $\mathbb{F}_5^*$ generated by $g = 3$

i	0	1	2	3
3^i mod 5	1	3	4	2

Example (A DHM type $(10, 5, 2, 7)$ ADS in $\mathbb{Z}_{10}$)

Take $g = 3$ as a generator of $\mathbb{F}_5^*$. Let $C_0 = C_0^{(4)} \cup C_1^{(4)} = \{1, 3\}$ and $C_1 = C_1^{(4)} \cup C_2^{(4)} = \{3, 4\}$, and let $D = \big(\{0\} \times C_0\big) \cup \big(\{1\} \times C_1\big) \cup \big\{(0, 0)\big\}$. Then,

$$D = \{(0,1), (0,3), (1,3), (1,4), (0,0)\} \subseteq \mathbb{Z}_2 \times \mathbb{Z}_5 \simeq \mathbb{Z}_{10}.$$

Equivalently, $D = \{6, 8, 3, 9, 0\} \subseteq \mathbb{Z}_{10}$.

$(0, j) \in \mathbb{Z}_2 \times \mathbb{Z}_5 \leftrightarrow \begin{cases} j & (j: \text{even}) \\ j+5 & (j: \text{odd}) \end{cases}$

$(1, j) \in \mathbb{Z}_2 \times \mathbb{Z}_5 \leftrightarrow \begin{cases} j+5 & (j\ \text{even}) \\ j & (j: \text{odd}) \end{cases}$

Existence for $(n = 4m-2, 2m-1, m-1, 3m-2)$-ADS with small n

- (SLCE) $n = q - 1$, where $q \equiv 3 \pmod 4$ is a prime power;
- (DHM) $n = 2p$, where $p \equiv 5 \pmod 8$ is a prime and $p-1$ or $p-4$ is a perfect square.
- $n \in \{34, 38, 50\}$ by computer search.

n	6	10	14	18	22	26	30	34	38	42	46
Construction	S	S,D	$\nexists$	S	S	S,D	S	PC	PC	S	S
n	50	54	58	62	66	70	74	78	82	$[86, 98]$	
Construction	PC	$\nexists$?	S,D	?	S	S	D	S	S	?	

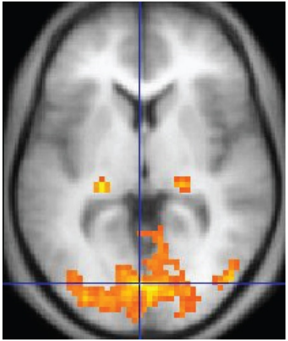
Homework

Outline

Outline

fMRI experiments

- Functional magnetic resonance imaging (fMRI) is a way to study neural correlates of consciousness involving perception, memory, learning, thinking, and affection by measuring hemodynamic response to mental stimuli.
  血液流動反応
- An fMRI experiment measures brain activity by detecting changes associated with blood flow.
- In an fMRI experiment, the experimental subject is asked to participate in mental tasks in response to the stimuli, while the subject's brain is scanned by a magnetic resonance (MR) scanner.

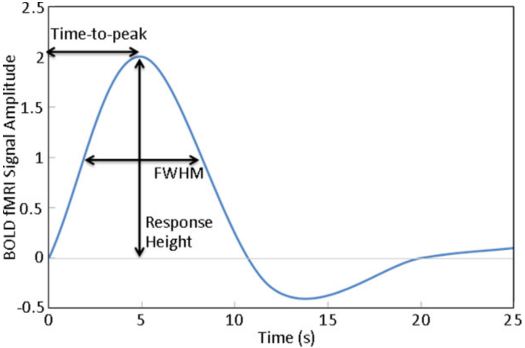


An fMRI image

from https://en.wikipedia.org/wiki/Functional_magnetic_resonance_imaging

HRFs

- The MR signal changes following stimuli are of great interests.
- Hemodynamic response functions (HRF) are typically used for describing the signal changes.



An example of HRF

from : `https://doi.org/10.1002/mrm.27146`

Model assumption for statistical analysis

Time series data

- In an fMRI experiment, each observation at a constant time interval is supposed to be affected by not only the current stimulus but also the preceding stimuli.
- A mental stimulus (e.g., a 1.5-second flickering image) is presented to a subject during n time points in the experiment.
- The HRF completely returns to baseline after k time points.

Next we review the statistical model (linear regression model) for an experiment with a single type of stimulus to estimate hemodynamic response functions (HRFs).

Linear model for estimating an HRF

The linear model for estimating an HRF can be expressed as follows:

$$y_i = \gamma + x_i h_1 + x_{i-1} h_2 + \cdots + x_{i-k+1} h_k + \varepsilon_i, \quad \text{for } i = 1, 2, \ldots, n,$$

- y_i: measurement obtained by an fMRI scanner at the ith time point.
- γ: nuisance parameter.
- h_j: unknown height (magnitude) of the HRF at the $(j-1)$th time point.
- $x_{i-k+1} \in \{0, 1\}$ s.t. $x_{i-k+1} = 1$ if h_j contributes to y_i and $x_{i-k+1} = 0$ otherwise.
- ε_i: Gaussian noise with mean 0 and variance σ^2.

Moreover,

- $x_l = x_{n+l}$ for $l \leq 0$. (cf. [Cheng and Kao, 2015][5]) $(\star)$

[5] C. S. Cheng, M. H. Kao, Optimal experimental designs for fMRI via circulant biased weighing designs, Ann. Stat., 43(6): 2565–2587 (2015).

Linear model for estimating an HRF: matrix form

$$
\begin{aligned}
y_1 &= \gamma + x_1 h_1 + x_0 h_2 + x_{-1} h_3 + \cdots + x_{2-k} h_k + \varepsilon_1,\\
y_2 &= \gamma + x_2 h_1 + x_1 h_2 + x_0 h_3 + \cdots + x_{3-k} h_k + \varepsilon_2,\\
y_3 &= \gamma + x_3 h_1 + x_2 h_3 + x_1 h_3 + \cdots + x_{4-k} h_k + \varepsilon_3,\\
&\vdots\\
y_n &= \gamma + x_n h_1 + x_{n-1} h_2 + x_{n-2} h_3 + \cdots + x_{n-k+1} h_k + \varepsilon_k.
\end{aligned}
$$

Matrix form of the model

$$
\mathbf{y} = \gamma \mathbf{1}_n + \mathbf{X}\mathbf{h} + \boldsymbol{\varepsilon},
$$

- $\mathbf{y} = (y_1, \ldots, y_n)^\top$, $\mathbf{h} = (h_1, \ldots, h_k)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.
- $\mathbf{X} = \left[\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(k)}\right] = (x_{ij}) \in \{0,1\}^{n\times k}$: design matrix.

Design matrix for the linear model

$$\mathbf{X} = \begin{bmatrix} x_1 & x_0 & x_{-1} & \cdots & x_{2-k} \\ x_2 & x_1 & x_0 & \cdots & x_{3-k} \\ x_3 & x_2 & x_1 & \cdots & x_{4-k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n-1} & x_{n-2} & \cdots & x_{n-k+1} \end{bmatrix} =: \left[\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(k)}\right] \in \{0,1\}^{n\times k}$$

- $x_l = x_{n+l}$ for $l \leq 0$. $(\star)$
- By $(\star)$, $\mathbf{X}$ is circulant, i.e. $\mathbf{X}_{(j)} = \mathbf{C}^{j-1}\mathbf{X}_{(1)}$ for $1 \leq j \leq k$, where $\mathbf{C} = \begin{bmatrix} \mathbf{0}^\top & 1 \\ \mathbf{I}_{n-1} & \mathbf{0} \end{bmatrix}$.

Information matrix for the linear model

- $\mathbf{A}$: $k \times n$ $\{0,1\}$ circulant array. ($\mathbf{X} = \mathbf{A}^\top$: design matrix)
- Let $\tilde{\mathbf{X}} = (\mathbf{J} - 2\mathbf{A})^\top$ ($\leftarrow$ a ± 1 matrix)
- Let $\mathbf{M}(\mathbf{A}) = \tilde{\mathbf{X}}^\top \left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right) \tilde{\mathbf{X}}$. ($\leftarrow$ information matrix for the ± 1 matrix)

The information matrix for estimating $\mathbf{h}$ is given as

$$\mathbf{M}_{\mathbf{X}} = \mathbf{X}^\top \left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right) \mathbf{X} = \frac{1}{4}\mathbf{M}(\mathbf{A}),$$

where $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^\top$.

For convenience, we consider $\mathbf{M}(\mathbf{A})$ instead of $\mathbf{M}_{\mathbf{X}}$.

$$\frac{1}{n}\begin{bmatrix} n-1 & & -1 \\ & \ddots & \\ -1 & & n-1 \end{bmatrix}$$

Design matrix: an example

- $n = 10$, $k = 5$.

$$\mathbf{X}^\top = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \qquad \mathbf{M}_\mathbf{X} = \frac{1}{2}\begin{bmatrix} 5 & -1 & -1 & 1 & -1 \\ -1 & 5 & -1 & -1 & 1 \\ -1 & -1 & 5 & -1 & -1 \\ 1 & -1 & -1 & 5 & -1 \\ -1 & 1 & -1 & -1 & 5 \end{bmatrix}$$

$$\tilde{\mathbf{X}}^\top = \begin{bmatrix} - & + & + & - & + & + & - & + & - & - \\ - & - & + & + & - & + & + & - & + & - \\ - & - & - & + & + & - & + & + & - & + \\ + & - & - & - & + & + & - & + & + & - \\ - & + & - & - & - & + & + & - & + & + \end{bmatrix}, \qquad \mathbf{M}_\mathbf{X} = \begin{bmatrix} 10 & -2 & -2 & 2 & -2 \\ -2 & 10 & -2 & -2 & 2 \\ -2 & -2 & 10 & -2 & -2 \\ 2 & -2 & -2 & 10 & -2 \\ -2 & 2 & -2 & -2 & 10 \end{bmatrix}$$

Definition of CAOAs

Definition (Circulant almost orthogonal arrays; CAOAs)

A binary circulant $k \times n$ array $\mathbf{A}$ is a circulant almost orthogonal array (CAOA) with parameter $(n, k, 2, t, b)$, if in any $t \times n$ subarray of $\mathbf{A}$ it holds that $|\lambda(\mathbf{a}_1) - \lambda(\mathbf{a}_2)| \leq b$ for any distinct $\mathbf{a}_1, \mathbf{a}_2 \in \{0, 1\}^t$, where $\lambda(\mathbf{a})$ is the frequency of $\mathbf{a}$ as column vectors. [Lin, et al., 2017] [6]

Example (CAOA$(10, 5, 2, 2, 1)$)

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

[6] Y. L. Lin, F. K. H. Phoa, M. H. Kao. Optimal design of fMRI experiments using circulant (almost-) orthogonal arrays. Ann. Stat.,45(6): 2483–2510 (2017).

CAOA$(n, k, 2, 2, b)$ for $n \not\equiv 2 \pmod 4$

- When $n \equiv 0 \pmod 4$
  - A CAOA$(n, k, 2, 2, b = 0)$ is equivalent to a circulant partial Hadamard design[7] 0-H $(k \times n)$.
  - Construction via H-sequence, Paley difference sets [8].
  - CAOA$(4u, 14, 2, 2, 0)$ for any $u \geq 9$ [Lin, et al., 2017] (by a recursive construction).
- When $n \equiv 1, 3 \pmod 4$,
  - Construction via extended H-sequences.
  - Construction by difference variance algorithms (DVA).
- When $n \equiv 3 \pmod 4$, a CAOA$(n, n, 2, 2, 1)$ exists if:
  - $n \equiv 3 \pmod 4$ and n is a prime. (via Paley difference sets)
  - $n = p(p + 2)$ where p and $p + 2$ are both odd primes. (via twin prime difference sets)
  - $n = 2^m - 1$ where $m \geq 2$. (via Singer difference sets)

[7] Y. L. Lin, F. K. H. Phoa, M. H. Kao. Circulant partial Hadamard matrices: construction via general difference sets and its application to fMRI experiments. Stat. Sinica, 27(4): 1715–1724 (2017).

[8] C. S. Cheng, M. H. Kao. Optimal experimental designs for fMRI via circulant biased weighing designs. Ann. Stat., 43(6): 2565–2587 (2015).

CAOA$(n, k, 2, 2, b)$ for $n \equiv 2 \pmod 4$

- When $n \equiv 2 \pmod 4$, it is quite hard to construct! ←We focus on this case!
  - ▸ $\exists$ T_2-CAOA$(2n, n, 2, 2, 1)$ for all odd prime n [Lin, et al., 2017] (using Paley DS).

Definition (T_1-, T_2-, T_3-, T_3^*-CAOA)

1. $\mathbf{A}$ is a T_1-CAOA if $\mathbf{M}(\mathbf{A}) = (n-2)\mathbf{I}_k + 2\mathbf{J}_k$, $\leftarrow \begin{bmatrix} n & & 2 \\ & \ddots & \\ 2 & & n \end{bmatrix}$
2. $\mathbf{A}$ is a T_2-CAOA if $\mathbf{M}(\mathbf{A}) = (n+2)\mathbf{I}_k - 2\mathbf{J}_k$, $\leftarrow \begin{bmatrix} n & & -2 \\ & \ddots & \\ -2 & & n \end{bmatrix}$
3. $\mathbf{A}$ is a T_3-CAOA if $\mathbf{A}$ is neither T_1- nor T_2-CAOA. $\leftarrow \begin{bmatrix} n & & 2 \text{ or } -2 \\ & \ddots & \\ 2 \text{ or } -2 & & n \end{bmatrix}$

$(n, k, 2, 2, 1)$

Problems

1. Which type of CAOA is better (best = optimal)?
2. How large k can be for given $n \equiv 2 \pmod 4$?
3. How do we construct such CAOAs?

Outline

Some well-known optimality criteria

- Roughly speaking, optimality criteria are functionals of the eigenvalues of the information matrix $\mathbf{M}$.
- $\mathbf{M}$: non-negative definite symmetric matrix of rank k with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$.

- D-optimality (determinant): minimize $\det(\mathbf{M}^{-1})$, i.e., maximize $\prod_{i=1}^{k} \lambda_i$.
- A-optimality (trace or "average"): minimize $\mathrm{tr}(\mathbf{M}^{-1})$, i.e., minimize $\sum_{i=1}^{k} \lambda_i^{-1}$.
- Φ_p-optimality: minimize $\sum_{i=1}^{k} \lambda_i^{-p}$ with $0 < p < \infty$.
- E-optimality (eigenvalue): maximize λ_k (the minimum eigenvalue).

Type-1 optimality criteria

- $\mathbf{M}$: non-negative definite symmetric matrix of rank k with eigenvalues $\lambda_1, \ldots, \lambda_k$.

Definition (Type 1 optimality criteria [Cheng, 1978][9])

A type 1 criterion is of the form $\Phi_f(\mathbf{M}) = \sum_{i=1}^{k} f(\lambda_i)$, where $f : \mathbb{R} \to [0, \infty)$ is continuously differentiable in $(0, \infty)$ with $f' < 0, f'' > 0$, and $f''' < 0$, and $\lim_{x \to 0^+} f(x) = f(0) = \infty$.

In particular, the D-optimality criterion is of type 1 for $f(\lambda) = -\log \lambda$.

- $\mathcal{A}$: the set of all $k \times n$ $\{0, 1\}$ circulant arrays.
- $\mathcal{M}_{\mathcal{A}} = \{\mathbf{M}(\mathbf{A}) : \mathbf{A} \in \mathcal{A}\}$.

Definition (Optimality for CAOAs)

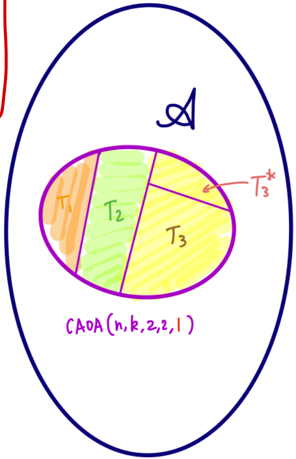
If $\mathbf{M} \in \mathcal{M}_{\mathcal{A}}$ (resp. $\mathbf{A} \in \mathcal{A}$ with $\mathbf{M} = \mathcal{M}_{\mathcal{A}}$) minimizes $\Phi_f(\mathbf{M})$ in the class $\mathcal{M}_{\mathcal{A}}$ for all f, then $\mathbf{M}$ (resp. $\mathbf{A} \in \mathcal{A}$) is optimal over $\mathcal{M}_{\mathcal{A}}$ (resp. $\mathcal{A}$) w.r.t. all type 1 criteria.

[9] C. S. Cheng, Optimality of certain asymmetrical experimental designs, Ann. Stat., 6: 1239–1261 (1978).

Three types of CAOAs (Revisit)

- $n \equiv 2 \pmod 4$ and $k \geq 2$.
- $\mathbf{A}$: CAOA$(n, k, 2, 2, 1)$
- $\mathcal{A}$: the set of $k \times n$ $\{0, 1\}$-circulant arrays.

$$T_3^*: \begin{bmatrix} n & -2 & 2 & -2 & 2 & -2 & \cdots \\ -2 & \ddots & & & & & \\ 2 & & \ddots & & & & \\ \vdots & & & \ddots & & & \\ & & & & n & & \end{bmatrix}$$



Definition (T_1-, T_2-, T_3-, T_3^*-CAOA)

1. $\mathbf{A}$ is a T_1-CAOA if $\mathbf{M}(\mathbf{A}) = (n-2)\mathbf{I}_k + 2\mathbf{J}_k$,
   optimal (by [Cheng, et al., 2017][10])
2. $\mathbf{A}$ is a T_2-CAOA if $\mathbf{M}(\mathbf{A}) = (n+2)\mathbf{I}_k - 2\mathbf{J}_k$,
3. $\mathbf{A}$ is a T_3-CAOA if $\mathbf{A}$ is neither T_1- nor T_2-CAOA.
   In particular, if $\mathbf{M}(\mathbf{A}) = (n-2)\mathbf{I}_k + 2\mathbf{L}_k$ with $\mathbf{L}_k = (\ell_{ij})$ where $\ell_{ij} = (-1)^{i+j}$, then $\mathbf{A}$ is a T_3^*-CAOA. optimal [L.]

[10] C. S. Cheng, M. H. Kao, F. K. H. Phoa. Optimal and efficient designs for functional brain imaging experiments. J Statist. Plann. Inference.181: 71–80 (2017).

A new type of optimal CAOAs

Theorem (L. et. al, 2021)

T_3^*-CAOA $\mathbf{A}$ is optimal over $\mathcal{A}$ w.r.t. any type 1 criterion.

Proof (sketch): T_3^*-CAOA$(n, k, 2, 2, 1)$ has the same eigenvalues with a T_1-CAOA$(n, k, 2, 2, 1)$.

Actually, we proved a stronger theorem.[11]

Theorem (L. et. al, 2021)

If there exists a T_1- or a T_3^*-CAOA$(n, k, 2, 2, 1)$, then, with respect to any type 1 criterion, any optimal array in $\mathcal{A}$ is either a T_1- or a T_3^*-CAOA$(n, k, 2, 2, 1)$.

[11] X.-N. Lu, M. Mishima, N. Miyamoto, M. Jimbo. Optimal and efficient designs for fMRI experiments via two-level circulant almost orthogonal arrays. J Statist. Plann. Inference, 213: 33-49, (2021).

An example of T_3^*-CAOA

- $n = 14$, $k = 6$.
- Generating vector (the first row of a CAOA): 11101100001010

$$\mathbf{M} = \begin{bmatrix} 14 & -2 & 2 & -2 & 2 & -2 \\ -2 & 14 & -2 & 2 & -2 & 2 \\ 2 & -2 & 14 & -2 & 2 & -2 \\ -2 & 2 & -2 & 14 & -2 & 2 \\ 2 & -2 & 2 & -2 & 14 & -2 \\ -2 & 2 & -2 & 2 & -2 & 14 \end{bmatrix}$$

- Eigenvalues of $\mathbf{M}$: $(24, 12, 12, 12, 12, 12)$

Table of T_3^*-CAOAs

Example (T_3^*-CAOA$(n, k_3^*, 2, 2, 1)$ (optimal))

n	k_3^*	(k_1)	Generating vector
10	$\underline{5}$	$(\underline{3})$	1101000101
14	$\underline{6}$	$(\underline{4})$	11101100001010
18	$\underline{6}$	$(\underline{6})$	110101110100001001
22	$\underline{8}$	$(\underline{7})$	0010100100011110111010
26	$\underline{13}$	$(\underline{9})$	11010100000110010101111100
30	$\underline{14}$	$(\underline{10})$	100111111001101010100000110010
34	$\underline{13}$	$(\underline{11})$	0011000000011110011101101101010101
38	14	(13)	11010100000111000100010101101100111110
42	16	(13)	110110101110111100101010000010110011000001
46	17	(14)	1011001001011100001000010011101000111011111010
50	18	(15)	01101010101000001100011001000110100111111110001011

$*$ The best known k for T_1-CAOAs are also listed. The underlined values are best possible (cannot be larger).

Outline

How large k can be? (1/2)

Proposition (Upper bound for k)

1. For $n \geq 6$, any of T_1-, T_2- and T_3^*-CAOA$(n, k, 2, 2, 1)$ satisfies $k \leq n/2$.
2. For $n \geq 10$, any T_1-CAOA$(n, k, 2, 2, 1)$ satisfies $k \leq n/2 - 2$.

Problem

For which kind of T_3-CAOA$(n, k, 2, 2, 1)$, $k = n - 1$ may hold?

How large k can be? (2/2)

Example (T_3^*-CAOA$(n = 10, k = 5, 2, t = 2, b = 1)$ Revisit)

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

How large k can be? (2/2)

Example (Information matrix of T_3^*-CAOA$(n = 10, k = 5, 2, t = 2, b = 1)$)

$$\mathbf{M}(\mathbf{A}) = \begin{bmatrix} 10 & -2 & 2 & -2 & 2 \\ -2 & 10 & -2 & 2 & -2 \\ 2 & -2 & 10 & -2 & 2 \\ -2 & 2 & -2 & 10 & -2 \\ 2 & -2 & 2 & -2 & 10 \end{bmatrix}.$$

How large k can be? (2/2)

Example (Forcing $k = 9$ for a T_3^*-CAOA$(n = 10, k = 5, 2, t = 2, b = 1)$...)

$$
\mathbf{M}(\mathbf{A}) = \begin{bmatrix}
10 & -2 & 2 & -2 & 2 & -10 & 2 & -2 & 2 \\
-2 & 10 & -2 & 2 & -2 & 2 & -10 & 2 & -2 \\
2 & -2 & 10 & -2 & 2 & -2 & 2 & -10 & 2 \\
-2 & 2 & -2 & 10 & -2 & 2 & -2 & 2 & -10 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}.
$$

T_3-CAOAs and sequences

Theorem (Relationship between T_3-CAOAs and sequences)

For $n \equiv 2 \pmod 4$, the following are equivalent:

1. There exists a ~~perfect~~ optimal binary sequence of length n consisting of equal numbers of 0 and 1 (called balanced);
2. A T_3-CAOA$(n, n-1, 2, 2, 1)$ (not necessarily T_3^*) exits;
3. A T_3-CAOA$(n, k, 2, 2, 1)$ (not necessarily T_3^*) exits for any $n/2 < k \leq n-1$.

Binary balanced optimal sequences and almost difference sets (revisit)

- For $\mathbf{s} \in \{0,1\}^n$, regard the indices as in $\mathbb{Z}_n = \{0,1,\ldots,n-1\}$.
- The set of indices of 1's in $\mathbf{s}$ is the support of $\mathbf{s}$, denoted by $\mathrm{supp}(\mathbf{s})$.
- For an optimal $\mathbf{s}$ with $n \equiv 2 \pmod 4$, $\mathrm{supp}(\mathbf{s})$ is an almost difference set in $\mathbb{Z}_n$.

Proposition

A T_3-CAOA$(n,k,2,2,1)$ with generating vector $\mathbf{s}$ $\iff$
An optimal balanced binary sequence $\mathbf{s}$ $\iff$
$D := \mathrm{supp}(\mathbf{s})$ is an $(n, \frac{n}{2}, \frac{n-2}{4}, \frac{3n-2}{4})$ ADS in $\mathbb{Z}_n$

Known constructions for sequences (revisit)

Using known sequence constructions and computer search, we have:

Theorem

There exits a T_3-CAOA$(n, n-1, 2, 2, 1)$ for the following n:

1. *(From SLCE seq.) $n = q - 1$, where $q \equiv 3 \pmod 4$ is a power of a prime;*
2. *(From DHM seq.) $n = 2p$, where $p \equiv 5 \pmod 8$ is a prime such that $p - 1$ or $p - 4$ is a perfect square.*
3. $n \in \{34, 38, 50\}$.

Comparison of D-efficiency(%) between T_3- and T_2-CAOAs

n		$k=n/2$ T_3	$k=n/2$ (T_2)	$k=n-1$ T_3	$k=n-1$ (T_2)	$k=9$ T_3	$k=9$ (T_2)
6	S	92.83	(92.83)	84.41	(84.41)	NA	NA
10	S	97.67	(89.13)	82.74	(69.44)	82.74	(69.44)
	D	97.67	(89.13)	82.74	(69.44)	82.74	(69.44)
18	S	97.01	(89.01)	85.28	(55.47)	97.01	(89.01)
22	S	97.14	(89.49)	86.38	(51.24)	98.05	(95.79)
26	S	98.16	(90.00)	87.37	(47.91)	99.08	(97.78)
	D	98.16	(90.00)	87.37	(47.91)	99.08	(97.78)
30	S	98.02	(90.50)	88.17	(45.18)	99.43	(98.66)
34	PC	99.05	(90.95)	91.26	(42.89)	99.79	(99.12)
		98.61	(90.95)	90.09	(42.89)	99.59	(99.12)
38	PC	98.39	(91.37)	91.79	(40.93)	99.59	(99.39)
42	S	98.67	(91.75)	90.08	(39.23)	99.65	(99.56)
46	S	98.76	(92.10)	90.57	(37.73)	99.87	(99.67)
50	PC	98.88	(92.42)	92.17	(36.39)	99.87	(99.75)

Enumerating $(n, \frac{n}{2}, \frac{n-2}{4}, \frac{3n-2}{4})$ ADS

Problem

For given $n \equiv 2 \pmod 4$, how many different are there?

By further characterizing optimal balanced sequences and employing SUGAR[12], a SAT-based constraint solver, a complete search of all optimal balanced sequences of length $n = 2u$ with $u \in \{3, 5, \ldots, 23\}$ was carried out.

Table: The numbers of equivalence classes of optimal balanced sequences of length n

$n = 2u$	6	10	14	18	22	26	30	34	38	42	46
# eq. classes	1	1	0	8	12	24	30	4	4	16	8
Known constr.	S	S,D	-	S	S	S,D	S	-	-	S	S

[12] https://cspsat.gitlab.io/sugar/

Updated D-efficiency (%) of T_3-CAOA$(n, n-1, 2, 2, 1)$

n		T_3	best T_3
6	S	84.41	-
10	S	82.74	-
	D	82.74	-
18	S	85.28	-
22	S	86.38	86.90
26	S	87.37	89.22
	D	87.37	
30	S	88.17	90.97
34	PC	91.26	-
		90.09	
38	PC	91.79	91.89
42	S	90.08	92.58
46	S	90.57	92.51

Homework assignments (レポート課題) for 5th day

Exercise 1

Construct an SLCE sequence of length $n = 18$.
Hint: take $q = 19$.

Exercise 2

Find all the integers $100 \leq n \leq 200$ such that there exists an SLCE sequence or a DHM sequence of length n.

- You are encouraged to use computer programs.
- Deadline: 6th Sept., 23:59:59