Perfect Sequence Covering Arrays and related concepts

Enrico Iurlano

13. December 2022

Notation

General assumption: nonneg. integers $n \geq k \geq 3$.

$$S_{n,k} := \{(j_1, \ldots, j_k) \in \{1, \ldots, n\}^k : i \neq \ell \Rightarrow j_i \neq j_\ell\}.$$

Under consideration: Families $\mathcal{F} = \{\pi_1, \ldots, \pi_d\} \subseteq S_n$ $(i \neq \ell \nRightarrow \pi_i \neq \pi_\ell)$.

Definition (Kuhn et al. 2012; Spencer 1971: "completely k-scrambling family")

$\mathcal{A} \subseteq S_n$ is called *Sequence Covering Array (SCA) of strength* k, if

$$\forall x = (x_1, \ldots, x_k) \in S_{n,k}\ \exists \pi \in \mathcal{A}\ \ \pi^{-1}(x_1) < \ldots < \pi^{-1}(x_k). \qquad (*)$$

π in $(*)$ covers x.

E.g. $n = 4, k = 3$:

$$\mathcal{A} = \begin{array}{|cccc|}\hline 3 & 1 & 2 & 4 \\ 1 & 3 & 4 & 2 \\ 2 & 3 & 4 & 1 \\ 4 & 1 & 2 & 3 \\ 2 & \mathbf{1} & \mathbf{4} & \mathbf{3} \\ 4 & 3 & 2 & 1 \\ \hline \end{array} \qquad (x_1, x_2, x_3) = (1, 4, 3)$$

- $N^*(n,k)$ defined as min. $d \in \mathbb{N}$ s.t. $\exists$ strength-k SCA over n symbols consisting of d permutations
- Bounds for $N^*(n,k)$ employed for
  - Order theory, combinatorial geometry, graph theory, etc.
  - Combinatorial Software Testing with SCAs (Kuhn et al. 2012)
  - Min-wise Hashing: Rapid document similarity estimation in huge collections (Google, AltaVista)

Example (SCA with $n = 7$, $k = 3$)

`EIUSB`: Establish internet connection via USB port;
`CWLAN`: Connect to a WLAN;
`CWCAM`: Connect peripheral device webcam;
`ENBLU`: Enable Bluetooth connectivity;
`CMOUS`: Connect peripheral device mouse;
`LOGOI`: Logout and successively login;
`SUSWU`: Suspend and wakeup

Performed test							Error?
EIUSB	CWLAN	CWCAM	ENBLU	CMOUS	LOGOI	SUSWU	/
ENBLU	CWCAM	CWLAN	EIUSB	SUSWU	LOGOI	CMOUS	/
EIUSB	CMOUS	LOGOI	SUSWU	CWLAN	CWCAM	ENBLU	/
SUSWU	LOGOI	CMOUS	EIUSB	ENBLU	CWCAM	CWLAN	/
CWLAN	CMOUS	EIUSB	CWCAM	ENBLU	LOGOI	SUSWU	/
CMOUS	CWLAN	SUSWU	LOGOI	ENBLU	CWCAM	EIUSB	Yes
CWCAM	LOGOI	EIUSB	CWLAN	ENBLU	CMOUS	SUSWU	/
LOGOI	CWCAM	SUSWU	CMOUS	ENBLU	CWLAN	EIUSB	Yes
ENBLU	SUSWU	EIUSB	CWLAN	CWCAM	CMOUS	LOGOI	/
SUSWU	ENBLU	LOGOI	CMOUS	CWCAM	CWLAN	EIUSB	Yes

Theorem (Spencer–Hajnal; Radhakrishnan–Spencer)

$$N^*(n,k) \quad = \Theta(\log_2 n) \qquad (k \geq 3,\ \textit{fixed}).$$

- $N^*(n,k) \geq (k-1)!(1+o_n(1))\log_2 n$ (Entropy method).
- $N^*(n,k) \leq \frac{k}{\log_2 \frac{k!}{k!-1}} \log_2 n + 1$ (Probabilistic method).

- Fix $(j_1,\ldots,j_k) \in S_{n,k}$. $\Pr[\neg(\pi(j_1) < \ldots < \pi(j_k)) \mid \pi \xleftarrow{u} S_n] = 1 - \frac{1}{k!}$
- Sample indep. and unif. at random $\mathcal{F}_d = \{\pi_1,\ldots,\pi_d\} \leftarrow S_n$
- $\Pr[\forall \pi \in \mathcal{F}_d : \neg(\pi(j_1) < \ldots < \pi(j_k))\ ] = (1-\frac{1}{k!})^d$
- Hence,

$$\Pr\left[\bigcup_{(j_1,\ldots,j_k)\in S_{n,k}} \forall\pi\ \neg(\pi(j_1) < \ldots < \pi(j_k))\right] \leq \sum_{(j_1,\ldots,j_k)\in S_{n,k}} \left(1-\frac{1}{k!}\right)^d$$

$$= \binom{n}{k} k! \left(1-\frac{1}{k!}\right)^d \overset{!}{<} 1 \qquad (*)$$

- $\Rightarrow$ smallest d ensuring $(*)$ guarantees existence of SCA on d rows

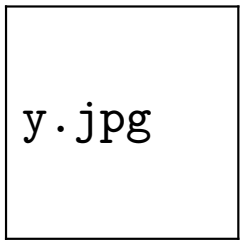
Definition (Yuster 2020)

A strength-k SCA $\mathcal{A} \subseteq S_n$ is called λ-*perfect*, if each $x \in S_{n,k}$ is covered λ times in $\mathcal{A}$. Notation: $\mathcal{A} \in \mathrm{PSCA}(n, k, \lambda)$.

- $A \in \mathrm{PSCA}(n, k, \lambda) \implies |\mathcal{A}| = \lambda k!$.
- Set $g(n, k) := \min\{\lambda : \mathrm{PSCA}(n, k, \lambda) \neq \emptyset\}$.
- Levenshtein 1992: $g(n, n-1) = 1$ as $\mathcal{A} \in \mathrm{PSCA}(n, n-1, 1)$ constructable

Theorem (Yuster 2020)

Lower bound $g(n, k) = \Omega(n^{k/2})$ *applies.*

- 

  Raphael Yuster: Can we find "small", i.e., polynomial size PSCAs?

- Master's thesis: Yes, always (poly. size: $g(n, k) \leq \frac{1}{k!} n^{O(k^2 / \ln k)}$).

Fascination PSCAs – when optimal packing succeeds...

PSCA$(4,3,1)$	PSCA$(5,3,1)$	PSCA$(5,3,2)$
3 1 2 4	$\nexists$	1 2 4 5 3
1 3 4 2		1 3 4 5 2
2 3 4 1		2 1 5 4 3
4 1 2 3		2 3 4 1 5
2 1 4 3		3 1 5 4 2
4 3 2 1		3 2 5 1 4
		4 1 3 2 5
		4 2 3 5 1
		4 5 2 1 3
		5 1 2 3 4
		5 3 2 4 1
		5 4 3 1 2
$\Rightarrow g(4,3) = 1$	$\Rightarrow g(5,3) > 1$	$\Rightarrow g(5,3) = 2$

PSCAs: Bounds for multiplicities for small n, k

A construction of Tarui et al. 2000 can be slightly optimized:

n ($k = 4$)	A	B
5	1	1
6	1	672
7	2	672
8–12	18	–
13	234	–
14–17	5040	**672**
18–21	5040	3 139 584
22	18 480	–
23	425 040	–
24	10 200 960	**3 139 584**
25-65	/	3 139 584
66-257	/	23 482 368

n ($k = 3$)	A	B
4	1	1
5–7	2	10
8	3	-
9	4	-
10–12	6	-
13–14	7	-
15–16	16	**10**
17–19	19	180
20–32	96	180
33–64	/	180
65–256	/	340

Upper bounds for $g(n, k)$:
A $\rightsquigarrow$ State of the art (Na et al. 2022, Gentle et al. 2022);
B $\rightsquigarrow$ Recursive construction of Tarui et al. 2000, optimized (Iurlano 2022)
Green: optimal bounds

- Mathon et al. 1999: $\lambda = 1$: Backtracking within unions of cosets of S_n
- Na et al. 2022: λ arbitrary: analogous backtracking within unions ...

Example (Na et al. 2022)

Subgroup $H := \{id, h_2, h_3\} \leq S_7$, $\alpha, \beta, \gamma, \delta \in S_7$, such that the union of cosets

$$\mathcal{A} := H\alpha \cup H\beta \cup H\gamma \cup H\delta \in \text{PSCA}(7, 3, 2).$$

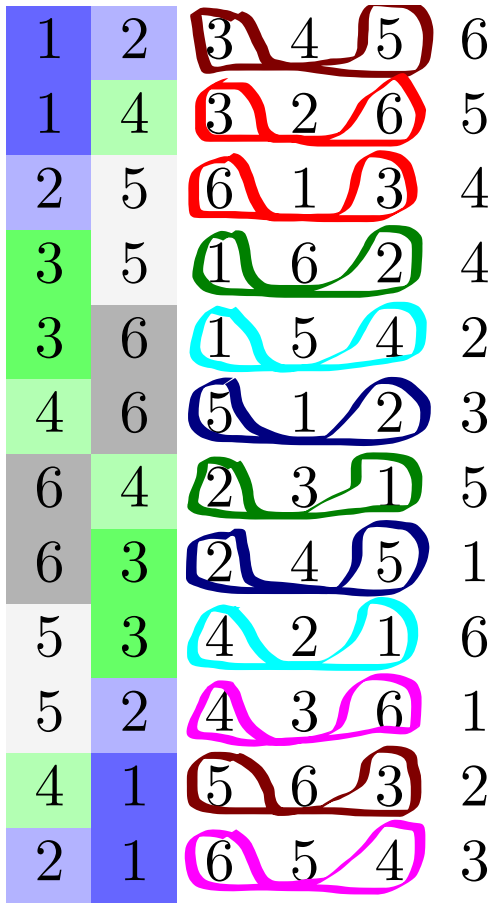
Four left cosets of the order 3 subgroup $\langle 3275641 \rangle$:

$\langle\sigma\rangle$	$1573426\,\langle\sigma\rangle$	$4261735\,\langle\sigma\rangle$	$4756123\,\langle\sigma\rangle$
1234567	1573426	4261735	4756123
3275641	3617524	5243176	5164327
7216453	7431625	6257314	6345721

- Gentle et al. 2022: Analysis of columnwise distribution of symbols
- Master's thesis: Study of feasible gaps between ℓ-subsequences ($\ell < k$)
  - $\rightsquigarrow$ Integer Linear Program formulation for "admissible gap configs"
  - $\rightsquigarrow$ Particular solution of ILP "reflection symmetry"

Reflection symmetry illustrated:

1	2	3	4	5	6
1	4	3	2	6	5
2	5	6	1	3	4
3	5	1	6	2	4
3	6	1	5	4	2
4	6	5	1	2	3
6	4	2	3	1	5
6	3	2	4	5	1
5	3	4	2	1	6
5	2	4	3	6	1
4	1	5	6	3	2
2	1	6	5	4	3



Open questions (selection)

- Complexity class of calculation of $N^*(n,k)$ resp. $g(n,k)$?
- Correctness of the Levenshtein conjecture

$$k \notin \{1,2,4\} \Rightarrow g(k+2,k) > 1?$$

- Does even $\mathrm{PSCA}(n,k,\lambda) \neq \emptyset \Rightarrow \mathrm{PSCA}(n,k,\lambda+1) \neq \emptyset$ apply?
- Calculate next value $g(9,3) \in \{3,4\}$.
- Similar combinatorial structures: Does union-of-cosets-approach succeed?

Related **optimization problems** (broader sense)

Shortest Superpermutation Problem

- Motivation: Watch n episodes of a series in all $n!$ possible ways (contiguously) by a long chain of movie events in order to have "viewed it correctly at least once"
- $n \in \mathbb{N}$. Find shortest length-L string S such that

$$\forall x \in S_{n,n} : x \text{ substring } S.$$

- e.g. $n = 4$: 123412314231243121342132413214321 ($L = 33$, 1432 contained)
- Solving approach: open dedicated pdf

Another related problem

- Problem: Find shortest *Universal Cycle Covering for given k-Subset* (details see Dębski et al. 2016)
- A cyclic sequence of elements of $[n]$ is an (n, k)-*Ucycle covering* if every k-subset of $[n]$ appears in this sequence at at least once as a subsequence of consecutive terms.
- Problem more attackable ... not "doubly exponential" effort (cf. Superpermutation problem)
- Ideas from `https://www.ac.tuwien.ac.at/files/pub/mayerhofer-22.pdf`

A Beam Search for the Shortest Common Supersequence Problem Guided by an Approximate Expected Length Calculation

Jonas Mayerhofer[1], Markus Kirchweger[2], Marc Huber[3], and Günther Raidl[4]

TU Wien, Vienna
e01633065@student.tuwien.ac.at[1],
{mk[2], mhuber[3], raidl[4]}@ac.tuwien.ac.at

exploitable?

Questions?

Nice to find:

- Idea good heuristic/construction for (near-/ optimal) PSCAs / SCAs
- Reduce an NP-hard problem to $\textsc{Calc}(N^*(n,k))$ or $\textsc{Calc}(g(n,k))$

References (Selection)


A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher. Min-wise independent permutations. *J. Comput. Syst. Sci.*, 60(3):630–659, 2000.


Y. M. Chee, C. J. Colbourn, D. Horsley and J. Zhou. Sequence covering arrays. SIAM J. Discrete Math., 27(4), 1844-1861, 2013.


A. R. Gentle and I. M. Wanless. On perfect sequence covering arrays, February 4, 2022. preprint on arXiv:2202.01960.


T. Itoh, Y. Takei, and J. Tarui. On permutations with limited independence. In *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, pages 137–146, 2000.


Iurlano. Growth of the perfect sequence covering array number. *Des. Codes Cryptog., accepted*, 2022.


J. Na, J. Jedwab, and S. Li. A group-based structure for perfect sequence covering arrays, February 4, 2022. preprint on arXiv:2202.01948.


J. Spencer. Minimal scrambling sets of simple orders. *Acta Math. Hung.*, 22(3–4):349–353, 1971.


J. Tarui, T. Itoh, and Y. Takei. A nearly linear size 4-min-wise independent permutation family by finite geometries. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 396–408. Springer, 2003.


R. Yuster. Perfect sequence covering arrays. *Des. Codes Cryptog.*, 88(3):585–593, 2020.