Growth of the perfect sequence covering array number

Enrico Iurlano[1] 📧

Abstract

In this note we answer positively an open question posed by Yuster in 2020 [14] on the polynomial boundedness of the *perfect sequence covering array number* $g(n, k)$ (PSCA number). The latter determines the (renormalized) minimum row-count that *perfect sequence covering arrays* (PSCAs) can possess. PSCAs are matrices with permutations in S_n as rows, such that each ordered k-sequence of distinct elements of $[n]$ is covered by the same number of rows. We obtain the result after illuminating an isomorphism between this structure from design theory and a special case of min-wise independent permutations. Afterwards, we point out that asymptotic bounds and constructions can be transferred between these two structures. Moreover, we sharpen asymptotic lower bounds for $g(n, k)$ and improve upper bounds for $g(n, 4)$ and $g(n, 3)$, for some concrete values of n. We conclude with some open questions and propose a new matrix class being potentially advantageous for searching PSCAs.

1 Introduction

Set $[n] = \{1, \ldots, n\}$. Let S_n denote the symmetric group on $[n]$, and define $S_{n,k}$ as the set of all ordered tuples $(x_1, \ldots, x_k) \in [n]^k$ whose entries x_j are distinct. Following [14], a *PSCA of strength k and multiplicity $\lambda \geq 1$ on n symbols* is a matrix $A \in [n]^{(\lambda k!) \times n}$ with distinct symbols per row such that each sequence $s \in S_{n,k}$ is contained as a subsequence in exactly λ rows (s is said to be *covered λ times by A*). Denote the class of all such matrices A by PSCA(n, k, λ). Let $g(n, k)$ be the minimum multiplicity λ such that PSCA(n, k, λ) is non-empty, and call $g(n, k)$ the *perfect sequence covering array number* (in analogy to [2]). A PSCA can be seen as a (k, λ)-directed design of blocksize n and order n (cf. [4]). PSCAs

are particular *sequence covering arrays* (SCAs), i.e., matrices with permutations as rows that cover each $s \in S_{n,k}$ at least once [2]. SCAs were introduced in [8] for the purpose of event sequence testing (cf. also [2, 4]).

Recently, the following two results were obtained by Yuster.

Theorem 1 [14] *If $k/2$ is a prime, then for all $n \geq k$ we have*

$$g(n, k)k! \geq \binom{n}{k/2} - \binom{n}{k/2 - 1}. \tag{1}$$

For arbitrary k, provided $n \gg k$, the bound $g(n, k) > n^{k/2 - o_k(1)}$ applies.

Theorem 2 [14] *Let $n \geq 3$. There exists a constant $C > 0$ such that*

$$n/6 \leq g(n, 3) \leq Cn \log_2 (n)^{\log_2 (7)}.$$

We now point to an isomorphic structure being analyzed in the setting of min-wise independent permutations [3]. We index families $\mathscr{F} \subset S_n$ by an index set $[d]$, $d \in \mathbb{N}$. The elements are allowed to occur repeatedly, and hence the cardinality of $\mathscr{F}$ refers to the cardinality of the indexing set. We need the concept of the *rank* of an element before we get to the actual definition of interest.

Definition 1 For a subset $X \subset [n]$ and arbitrary $x \in X$, we denote the *rank* of x in X as $\text{rank}(x, X) := |\{y \in X : y < x\}|$.

Definition 2 A non-empty family $\mathscr{F} \subset S_n$ is called *k-rankwise independent* [6, 13], if for each set $X = \{x_1, \ldots, x_k\}$ of k distinct elements of $[n]$ and each choice of k distinct values $r_1, \ldots, r_k \in \{0, \ldots, k - 1\}$, we have

$$\Pr\left[\bigwedge_{i=1}^{k} \text{rank}(\pi(x_i), \{\pi(x_1), \ldots, \pi(x_k)\}) = r_i\right] = \frac{1}{k!}, \tag{2}$$

when π is drawn uniformly at random from $\mathscr{F}$.

It is clear that (2) corresponds to the following condition: For each sequence $(x_1, \ldots, x_k) \in S_{n,k}$ and for each permutation π randomly drawn (uniform probability) from $\mathscr{F}$, we have

$$\Pr[\pi(x_1) < \pi(x_2) < \cdots < \pi(x_k)] = \frac{1}{k!}. \tag{3}$$

Next, we give a proposition that illuminates the isomorphy of PSCAs and rankwise independent families. Before that, we need an appropriate mapping and an auxiliary observation from which the proposition follows. For the sake of completeness we include a proof of the following Lemma 3 which was already noticed in [4, Lemma 1.1].

In the following, consider for a family $\mathscr{F} = (\pi_1, \ldots, \pi_d) \subset S_n$ the mapping

$$A : \mathscr{F} \mapsto A(\mathscr{F}) := \begin{bmatrix} \pi_1^{-1}(1), & \ldots, & \pi_1^{-1}(n) \\ \vdots & \ldots, & \vdots \\ \pi_d^{-1}(1), & \ldots, & \pi_d^{-1}(n) \end{bmatrix} \in [n]^{d \times n}. \tag{4}$$

Lemma 3 *Let $\mathscr{F} \subset S_n$ be a family of cardinality $d \geq 1$. Consider for fixed $i \in [d]$ the i-th element $\pi := \pi_i \in \mathscr{F}$ and the i-th row $r \in S_n$ of $A(\mathscr{F})$. Let $(x_1, \ldots, x_k) \in S_{n,k}$. We have $\pi(x_1) < \pi(x_2) < \cdots < \pi(x_k)$ if and only if $(x_1, \ldots, x_k)$ is a subsequence of r.*

Proof First, we show that monotonicity implies containment as subsequence: If $\pi(x_1) < \cdots < \pi(x_k)$, then the row $r = (\pi^{-1}(1), \ldots, \pi^{-1}(n))$ in particular contains $(\pi^{-1}(\pi(x_1)), \pi^{-1}(\pi(x_2)), \ldots, \pi^{-1}(\pi(x_k))) = (x_1, \ldots, x_k)$ as subsequence.

For the other proof direction, if π does not satisfy $\pi(x_1) < \pi(x_2) < \cdots < \pi(x_k)$, then there must be a permutation $\psi \in S_k \setminus \{\text{id}\}$ such that $\pi(x_{\psi(1)}) < \cdots < \pi(x_{\psi(k)})$, which implies (as before) that r contains $(x_{\psi(1)}, \ldots, x_{\psi(k)})$ as subsequence. Consequently $(x_1, \ldots, x_k)$ is not a subsequence of r. □

Proposition 4 *Let $n \geq k$. $\mathscr{F} \subset S_n$ is k-rankwise independent if and only if there exists $\lambda \in \mathbb{N} \setminus \{0\}$ such that $A(\mathscr{F}) \in \text{PSCA}(n, k, \lambda)$. Therefore, $g(n, k)k!$ determines the minimum cardinality of a k-rankwise independent family of permutations of $[n]$.*

Proof Define $C_x := \{\pi \in S_n : \pi(x_1) < \cdots < \pi(x_k)\}$, for $x = (x_1, \ldots, x_k) \in S_{n,k}$.

If $\mathscr{F}$ is k-rankwise independent, then for $x \in S_{n,k}$ we can find $|\mathscr{F}|/k!$ permutations in $\mathscr{F}$ which lie in C_x. By Lemma 3, this means that x is covered by precisely $|\mathscr{F}|/k!$ rows of $A(\mathscr{F})$. As a consequence, by choosing $\lambda := |\mathscr{F}|/k!$ we obtain $A(\mathscr{F}) \in \text{PSCA}(n, k, \lambda)$.

Conversely, if there is $\lambda \geq 1$ such that $A(\mathscr{F}) \in \text{PSCA}(n, k, \lambda)$, then for $x \in S_{n,k}$ there are λ rows in $A(\mathscr{F})$ that cover x. By Lemma 3, $\mathscr{F}$ must contain λ permutations which all are in C_x (here permutations are counted with respect to multiplicity). Therefore, a permutation π chosen uniformly at random from $\mathscr{F}$ satisfies $\pi(x_1) < \cdots < \pi(x_k)$ with probability $1/k!$. □

The inverse operation of $A(\cdot)$ in Proposition 4, i.e., the conversion of a PSCA to a rankwise independent family, is determined by interpreting the rows of the PSCA as permutations and successively storing their inverses one by one in a family of permutations.

In [6] it is already remarked that k-rankwise independence implies ℓ-rankwise independence, for all $\ell \in [k]$. In other words, a nesting property analogous to the one for PSCAs holds (if $A \in \text{PSCA}(n, k, \lambda)$, then also $A \in \text{PSCA}(n, \ell, \lambda k!/\ell!)$) [14]. We highlight that k-rankwise independent families can be considered as *completely scrambling families*[1] (introduced by Spencer [12]) with an additional requirement of regularity. Furthermore, k-rankwise independence is a special case of k-restricted min-wise independence (cf. Definition 3). It is known (cf. [6, p. 139]) that these two notions coincide for $k = 3$ and that k-rankwise independence is strictly more specific when $k > 3$. In [3], k-restricted min-wise independent families were introduced to efficiently estimate the resemblance of two documents. The latter is motivated by practice, where the aim is to reduce the computational cost of searching for near-duplicate documents on the World Wide Web [3].

Definition 3 [3] A non-empty family $\mathscr{F}$ is k-restricted min-wise independent if for any set $X \subset [n]$ with $|X| \leq k$ and arbitrary $x \in X$, any permutation π drawn with uniform probability randomly from $\mathscr{F}$ satisfies

$$\Pr\left[\pi(x) = \min_{y \in X} \pi(y)\right] = \frac{1}{|X|}.$$

Remark 1 Given a matrix in $\text{PSCA}(n, k, \lambda)$ with $n \geq k + 1$, we can utilize it to construct a matrix in $\text{PSCA}(n - 1, k, \lambda)$ by dropping the symbol n in each row (see [14]). Consequently, we also know how to construct a k-rankwise independent family $\mathscr{F}' \subset S_{n-1}$ (with $|\mathscr{F}'| = |\mathscr{F}|$) from a k-rankwise independent family $\mathscr{F} \subset S_n$ (this was already noticed and applied in [6] without referring to PSCAs).

We can now transfer bounds from rankwise independent families to PSCAs (and vice versa).

---

[1] A family $\mathscr{F} \subset S_n$ is completely k-scrambling if for any $(x_1, \ldots, x_k) \in S_{n,k}$ there exists an element $\pi \in \mathscr{F}$ such that $\pi(x_1) < \cdots < \pi(x_k)$.

2 Lower and upper bounds

In this section asymptotic results for rankwise independent permutations will be stated jointly with their implications (cf. Proposition 4) for the number $g(n, k)$.

The following result can in particular be analyzed when h is a prime; we can therefore compare it with Theorem 1 and conclude that it improves the lower bound by the factor $!h$, the *subfactorial* of h,[2] being superexponential in h, but independent of n.

Theorem 5 [1] *Let $\mathscr{F} \subset S_n$ be a k-rankwise independent family. If $k = 2h$ ($h \in \mathbb{N}$), then we have*

$$|\mathscr{F}| \geq g(n, k)k! \geq \sum_{i=0}^{h} !i \binom{n}{i} = \frac{!h}{h!} n^h (1 + o(1)). \tag{5}$$

Otherwise, if $k = 2h + 1$ ($h \in \mathbb{N}$), then

$$|\mathscr{F}| \geq g(n, k)k! \geq \sum_{i=0}^{h} !i \binom{n}{i} + !(h+1)\binom{n-1}{h} = \frac{!h + !(h+1)}{h!} n^h (1 + o(1)). \tag{6}$$

The next Theorem 6 follows from the proof of [6, Theorem 3.2]. It positively answers the question of Yuster on the existence of non-trivial upper bounds for $g(n, k)$ for fixed k. In fact, the estimate is polynomial in n.

Theorem 6 (variant of [6, Theorem 3.2]) *Let $p \geq n \geq k$ such that p is a prime and $n \geq (k-1)!$. Furthermore, let $p_1 < \cdots < p_m$ be the sequence of all primes not exceeding $k-1$, i.e., $p_m \leq k - 1$. Let $(e_1, \ldots, e_m) \in (\mathbb{N} \setminus \{0\})^m$ be a minimizer of $Q = \prod_{i=1}^{m} p_i^{e_i}$ under the side constraints that $(k-1)!$ divides Q, and $p_i^{e_i} > p$, for $i = 1, \ldots, m$. Then, there exists a k-rankwise independent family $\mathscr{P}$ of permutations of $[p]$ such that $|\mathscr{P}| \leq (p^k - p)Q^{\lfloor k/2 \rfloor}$. Consequently, cf. Remark 1, there exists a family $\mathscr{P}' \subset S_n$ satisfying*

$$g(n, k)k! \leq |\mathscr{P}'| = |\mathscr{P}| = n^{O(k^2 / \ln k)}. \tag{7}$$

Remark 2 For the case that $(k-1)! > n$ (not covered by Theorem 6), it is established in [6] that a k-rankwise independent family with cardinality of order $e^{O(k^3)}$ exists.

Remark 3 The proof of Theorem 6 presented in [6] is constructive: For a prime p and the finite field $\mathbb{F}_p$ of p elements ($\mathbb{F}_s$ is to be understood accordingly, when s is a prime power), the set of univariate polynomials

$$\mathbb{F}_p(\xi, [a, b]) := \{r \in \mathbb{F}_p[\xi] : a \leq \deg r \leq b\}$$

is used as base for the construction. Under the assumptions of Theorem 6, setting

$$E := \mathbb{F}_{p_1^{e_1}}(\xi, [0, \lfloor k/2 \rfloor - 1]) \times \cdots \times \mathbb{F}_{p_m^{e_m}}(\xi, [0, \lfloor k/2 \rfloor - 1]),$$

a tuple $(b_y)_{y \in \mathbb{F}_p} \in \{0, \ldots, Q - 1\}^p$ is generated by a manipulation which depends on a fixed parameter $h = (h_1, \ldots, h_m) \in E$. The tuple $(b_y)_y$ is then used to carry out the so-called "$\lfloor k/2 \rfloor$-tie breaking scheme" [6] which serves to identify each of the $p^k - p$ elements of $\mathbb{F}_p(\xi, [1, k])$ with a permutation of $\{0, \ldots, p - 1\}$. Deriving from each tuple in E (having cardinality $|E| = (p_1^{e_1})^{\lfloor k/2 \rfloor} \cdots (p_m^{e_m})^{\lfloor k/2 \rfloor} = Q^{\lfloor k/2 \rfloor}$) such a collection of permutations, yields a total of $(p^k - p)Q^{\lfloor k/2 \rfloor}$ permutations. Collectively these permutations fulfill k-rankwise independence. For a more detailed discussion of the proof of Theorem 6 (and of the aforementioned scheme) we point to [6].

---

[2] By definition, $!m := (m!) \cdot \sum_{j=0}^{m} (-1)^j / j!$. By power series representation, $\lim_{m \to \infty} \frac{!m}{m!} = 1/e$.

Table 1 Upper bounds for $g(n, 4)$

n	Bound in [5] (best known)	Bound by optimization in Remark 4
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

Entries constituting new best-known values are in bold print. Asterisks indicate optimal bounds

We state more specific results for $k \in \{3, 4\}$. Originally, the subsequent result concerning 3-rankwise independence was established for the equivalent property of 3-restricted min-wise independence.

Theorem 7 ([13]) *Let $n \geq 4$. Then, there exists a 3-rankwise independent family $\mathcal{E} \subset S_n$ and a 4-rankwise independent family $\mathcal{F} \subset S_n$, with cardinalities*

$$g(n, 3) \cdot 3! \leq |\mathcal{E}| \leq 12\sqrt{e}(1 + o(1))n \log_2 (n)^2 , \tag{8}$$

$$g(n, 4) \cdot 4! \leq |\mathcal{F}| \leq 15e(1 + o(1))n^3 \log_2 (n)^6 . \tag{9}$$

Remark 4 The latter result is obtained by methods from affine/projective finite geometry by a recursive construction. The bound (8) improves on the bound in Theorem 2 by a factor of approximately $\log_2 (n)^{0.81}$. The bound (9) is obtained by estimating the members of a recursive sequence which describes cardinalities of 4-rankwise independent families of permutations of $[n]$, for $n = 2^{2q} + 1$. The involved recursion is homogeneous in the cardinality of the family employed as base case. For the construction it is required to start with a 4-rankwise independent family of $\mathcal{F} \subset S_5$. In [13], the authors choose $\mathcal{F} = S_5$, such that $|\mathcal{F}| = 120$.

We can improve this base case by making use of a construction due to Levenshtein [9] allowing to construct a PSCA$(n + 1, n, 1)$. It provides us with a PSCA$(5, 4, 1)$, which we can map via the isomorphism in Proposition 4 to a 4-rankwise independent family $\tilde{\mathcal{F}}$ satisfying $|\tilde{\mathcal{F}}| = 24$. Hence, we obtain an improvement of the bound (9), which reads

$$g(n, 4) \cdot 4! \leq |\mathcal{F}| \leq 3e(1 + o(1))n^3 \log_2 (n)^6 . \tag{10}$$

The optimized recursion leading to (10) permits us to improve on some recently obtained results in [5] (by taking the exact values of the recursion). We state the hereby found bounds in Table 1.

Remark 5 Recently in [7], analogously to Remark 4, instead of the entire symmetric group S_4 chosen as base case for a similar recursion designed for the case $k = 3$ [13, Section 3.4],

Table 2 Upper bounds for $g(n, 3)$

n	Best known bound (with reference)		Bound by optimization in Remark 5
4		*1	1
5–7	[14];[5, 11]	*2	10
8	[5]	*3	–
9	[11]	4	–
10–12	[5]	6	–
13–14	[5]	7	–
15–16	[5]	16	**10**
17–19	[5]	19	180
20–32	[5]	96	180
33–64	–		180
65–256	–		340

Entries constituting new best-known values are in bold print. Asterisks indicate optimal bounds

a 3-rankwise independent family of cardinality 6 has been used (such a family exists due to PSCA$(4, 3, 1) \neq \emptyset$, cf. [9]). The bound

$$g(n, 3) \cdot 3! \leq 3\sqrt{e}(1 + o(1))n \log_2 (n)^2 \,, \tag{11}$$

slightly improving upon (8), is obtained. By evaluation of the respective recursion, bounds for $g(15, 3)$ and $g(16, 3)$ have been improved (see Table 2).

3 Conclusion and open questions

In our discussion, we have found that the theory of PSCAs and the theory of min-wise independent permutations can benefit considerably from each other. Despite the isomorphy of many concepts in these theories, strong interconnections seemed not to be exploited in the past (perhaps because one theory consistently uses probabilistic language). By combining the latter theories, we have improved several bounds and established polynomial boundedness for $g(n, k)$, which Yuster asked for in [14]. Furthermore, we achieved progress in another question appearing in [14] which asks to find the right order of magnitude of $g(n, 3)$: The quasi-linear upper bound (established in [14]) is tightened by a factor of $\log_2 (n)^{0.81}$ (cf. (8)). It remains open by how much lower and upper bounds for $g(n, 3)$ can still be improved.

The great difficulty, already for $k \in \{3, 4\}$ and small n, to determine existence of PSCAs (on a certain number of rows) manifests itself in some works, which come up with computationally intensive search procedures [5, 11]. It would be highly interesting to find out whether the regularity, that PSCAs have compared to SCAs, facilitates the determination of the complexity class of the problem of calculating $g(n, k)k!$, the minimum number of rows able to host a PSCA. For SCAs the respective problem concerning the minimal row count is still open (even if in [2] NP-completeness has been established for an altered problem).

Aiming to calculate further exact values of $g(n, k)$, we experimentally tried, for n being a divisor of $\lambda k!$, to seek small PSCAs by restricting the search space to a class of matrices A (having permutations as rows) and satisfying the following "reflection symmetry" property: For any column indices $j_1 < j_2$, the (j_1, j_2)-submatrix of A satisfies that if h rows coincide with $(a, b) \in [n]^2$, then also h rows coincide with (b, a). The motivation behind this additional

assumption is that recently PSCAs being unions of cosets of S_n have successfully been found (cf. [11]) for several parameter constellations (n, k, λ). We suspect that our proposed kind of symmetry induces a suitable balancing of symbols being advantageous for finding PSCAs, too.

It appears that within the class of matrices having uniform columnwise distribution of symbols (cf. [5]), this latter constraint still allows to find representatives of PSCAs. Indeed, by a recursive search for strength $k = 3$, appending column to column, we find that such special representatives exist for the parameter constellation $(n, k, \lambda) \in \{(3, 3, 1), (6, 3, 2)\}$. Existence for the constellation $(9, 3, 3)$ could no longer be determined due to computational limitations and we would find it interesting to investigate this case with more dedicated computational effort. The representative of PSCA$(6, 4, 1)$ presented in [10, Proposition 2.22] is reflection symmetric, too. For strength $k = 4$, the constellation $(n, k, \lambda) = (8, 4, 3)$ is next, for which it is still open whether it possesses such a special representative (since it was shown in [5] that PSCA$(8, 4, 2) = \emptyset$).

Combinatorial aspects (e.g. enumeration for small n and efficient constructions) of the class of matrices satisfying the latter reflection symmetry might be of independent interest.

Declarations

References

1. Bargachev V.: An improved lower bound on the size of k-rankwise independent families of permutations. Preprint on website of St. Petersburg Department of Steklov Institute of Mathematics. https://www.pdmi.ras.ru/preprint/2004/04-13.html (2004).
2. Brain M., Erdem E., Inoue K., Oetsch J., Pührer J., Tompits H., Yilmaz C.: Event-sequence testing using answer-set programming. Int. J. Adv. Softw. **5**(3&4) (2012).
3. Broder A.Z., Charikar M., Frieze A.M., Mitzenmacher M.: Min-wise independent permutations. J. Comput. Syst. Sci. **60**(3), 630–659 (2000).
4. Chee Y.M., Colbourn C.J., Horsley D., Zhou J.: Sequence covering arrays. SIAM J. Discret. Math. **27**(4), 1844–1861 (2013).
5. Gentle A.R., Wanless I.M.: On perfect sequence covering arrays. Ann. Combinat. (2022).
6. Itoh T., Takei Y., Tarui J.: On permutations with limited independence. In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 137–146 (2000).

7. Iurlano E.: Scrambling permutations and related structures. Master's thesis, TU Wien, Austria (2022). https://repositum.tuwien.at/handle/20.500.12708/101832
8. Kuhn D.R., Higdon J.M., Lawrence J.F., Kacker R.N., Lei Y.: Combinatorial methods for event sequence testing. In: 2012 IEEE Fifth International Conference on Software Testing, Verification and Validation, pp. 601–609. IEEE (2012).
9. Levenshtein V.I.: On perfect codes in deletion and insertion metric. Discret. Math. Appl. **2**, 241–258 (1992).
10. Na J.: Perfect sequence covering arrays. Master's thesis, Simon Fraser University, Canada (2021). https://theses.lib.sfu.ca/6570/show
11. Na J., Jedwab J., Li S.: A group-based structure for perfect sequence covering arrays. Designs Codes Cryptogr. (to appear).
12. Spencer J.: Minimal scrambling sets of simple orders. Acta Math. Hungar. **22**(3–4), 349–353 (1971).
13. Tarui J., Itoh T., Takei Y.: A nearly linear size 4-min-wise independent permutation family by finite geometries. In: Arora S., Jansen K., Rolim J.D.P., Sahai A. (eds.) Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pp. 396–408. Springer, Berlin, Heidelberg (2003).
14. Yuster R.: Perfect sequence covering arrays. Des. Codes Crypt. **88**(3), 585–593 (2020).