Perfect sequence covering arrays

Raphael Yuster *

Abstract

An (n,k) *sequence covering array* is a set of permutations of $[n]$ such that each sequence of k distinct elements of $[n]$ is a subsequence of at least one of the permutations. An (n,k) sequence covering array is *perfect* if there is a positive integer λ such that each sequence of k distinct elements of $[n]$ is a subsequence of precisely λ of the permutations.

While relatively close upper and lower bounds for the minimum size of a sequence covering array are known, this is not the case for perfect sequence covering arrays. Here we present new nontrivial bounds for the latter. In particular, for $k=3$ we obtain a linear lower bound and an almost linear upper bound.

AMS subject classifications: 05B40, 05B30, 05B15, 05A05
Keywords: covering array; sequence covering array; completely scrambling set of permutations; directed t-design

1 Introduction

Let $2 \le k \le n$ be positive integers. Let S_n denote the set of permutations of $[n]=\{1,\ldots,n\}$ and let $S_{n,k}$ denote the set of all sequences of k distinct elements of $[n]$. An (n,k) *sequence covering array* denoted by SCA(n,k), is a set $X \subseteq S_n$ such that each $\kappa \in S_{n,k}$ is a subsequence of some element of X. Naturally, one is interested in constructing an SCA(n,k) which is as small as possible. Thus, let $f(n,k)$ denote the minimum size of an SCA(n,k).

Sequence covering arrays have been extensively studied, see [2, 8] and the references therein which also provide some important applications of sequence covering arrays to the area of event sequence testing. Observe first that $f(n,2)=2$ as can be seen by taking any permutation and its reverse. A-priori, for a constant k, it is not entirely obvious that $f(n,k)$ grows with n, as each permutation covers $\binom{n}{k}$ sequences while $|S_{n,k}| = k!\binom{n}{k}$. However, more is known. The first to provide nontrivial bounds for $f(n,k)$ was Spencer [12] and various improvements on the upper and lower bounds were sequentially obtained by Ishigami [6, 7], Füredi [4], Radhakrishnan [11], and Tarui [13]. The (asymptotic) state of the art regarding $f(n,3)$ is the upper bound by Tarui [13] and the lower bound of Füredi [4]:

$$\frac{2}{\log e}\log n \le f(n,3) \le (1+o_n(1))2\log n\ .^1 \tag{1}$$

*Department of Mathematics, University of Haifa, Haifa 31905, Israel. Email: raphy@math.haifa.ac.il

[1] Unless stated otherwise, all logarithms are in base 2.

We note that the limit $f(n,3)/\log n$ exists [4, 13], but apparently its value is not known. For general fixed k, the best asymptotic upper and lower bounds are that of Spencer [12] and Radhakrishnan [11], respectively:

$$(1-o_n(1))\frac{(k-1)!}{\log e}\log n \leq f(n,k) \leq \frac{k}{\log(\frac{k!}{k!-1})}\log n \ . \tag{2}$$

We see that (2) provides logarithmic upper and lower bounds for $f(n,k)$, so the order of magnitude of $f(n,k)$ for fixed k, is known.

A natural design-theoretic question that arises when studying sequence covering arrays is that of perfectness. Let X be an SCA(n,k). We call X *perfect* if there exists an integer λ such that each $\kappa \in S_{n,k}$ is a subsequence of precisely λ elements of X. We call λ the *multiplicity* and denote (n,k) perfect sequence covering arrays by PSCA(n,k) allowing them to be multisets. In design-theoretic terms, a PSCA(n,k) with multiplicity λ is a $k-(n,n,\lambda)$ directed design, see [3] for the chapter on directed designs by Bennett and Mahmoodi. Notice that a PSCA(n,k) exists for every $2 \leq k \leq n$ since S_n is such. Let, therefore, $g^*(n,k)$ denote the minimum size of a PSCA(n,k) and observe the trivial bounds $k! \leq f(n,k) \leq g^*(n,k) \leq n!$.

An easy observation is that $g^*(n,k)$ is a multiple of $k!$. Indeed, if each k-sequence is covered precisely λ times, then the size of the corresponding PSCA(n,k) is $\lambda k!$ since each permutation covers precisely $\binom{n}{k}$ sequences and there are $k!\binom{n}{k}$ sequences to cover. So, we define the integer $g(n,k) = g^*(n,k)/k!$. Stated otherwise, $g(n,k)$ is the smallest λ such that a $k-(n,n,\lambda)$ directed design exists. Observe the trivial bounds $1 \leq f(n,k)/k! \leq g(n,k) \leq n!/k!$. We will also use the simple bounds $g(n,k) \geq g(n-1,k)$ and $g(n,k) \geq g(n,k-1)/k$. Indeed, the former can be seen by taking any PSCA(n,k) and removing element n from each permutation while the latter can be seen by taking the union of k repeated copies of any PSCA(n,k).

Determining when $g(n,k)=1$ or, equivalently, when $f(n,k)=k!$, is an open problem. While clearly $g(k,k)=1$ and $g(n,2)=1$ it is a result of Levenshtein [9] that $g(k+1,k)=1$. It is also known that $g(6,4)=1$ [10] and it is conjectured that $g(n,k)=1$ only if $n \leq k+1$ except for $k=2,4$ [9, 10]. The conjecture is known to hold for $k=3,5,6$ and it is also known that $g(7,4)>1$ [10]. For general $k \geq 3$, a result of Chee et al. [2] shows that $g(2k,k)>1$.

In Section 3 we determine the first (hence presently the only) exact bound of $g(n,k)$ which is not 1 as we prove that $g(5,3)=2$. However, our first main result is a lower bound for $g(n,k)$ which is much larger than the logarithmic lower bound for $f(n,k)$.

Theorem 1 *For all $n \gg k$, $g(n,k) > n^{k/2-o_k(1)}$. Furthermore, if $k/2$ is a prime then for all $n \geq k$ we have*

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

Notice that $g(n,4) \geq n(n-3)/48$ so coupled with the fact that $g(n,k) \geq g(n,k-1)/k$ we obtain, for every fixed $k \geq 4$, a polynomial in n lower bound for $g(n,k)$ while $f(n,k)$ is only logarithmic in n. Yet, Theorem 1 does not give valuable input for the smallest nontrivial case $k=3$. This is done in the next theorem, where we prove that $g(n,3)$ is at least linear in n and at most quasi-linear in n.

Theorem 2 *For all $n \geq 3$, $n/6 \leq g(n,3) \leq Cn(\log n)^{\log 7}$ for some absolute constant C.*

We note that the $\log 7 < 2.81$ can slightly be improved to any value strictly larger than $\log 6$ at the price of increasing C, but we cannot eliminate it completely.

In the next section we prove our general lower bound, Theorem 1. The case $k = 3$ and the proof of Theorem 2 appear in Section 3. The final section lists some open problems.

2 Lower bounds

Here we prove Theorem 1. Let X be a multiset of elements of S_n and let t be a positive integer. We define the binary incidence matrix $A = A_{X,t}$ as follows. The rows of A are indexed by the elements of $S_{n,t}$, (all sequences of t distinct elements of $[n]$) and the columns of A are indexed by X. For $\sigma \in X$ and $\kappa \in S_{n,t}$ we have $A[\kappa, \sigma] = 1$ if κ is a subsequence of σ. Otherwise, $A[\kappa, \sigma] = 0$. We trivially have $\text{rank}(A_{X,t}) \leq |X|$.

We will prove Theorem 1 for even values of k such that $k/2$ is a prime. We will then show that the result for other k follows as a consequence. Suppose now that X is a PSCA(n,k) with multiplicity λ. Let $t = k/2$ and consider $A = A_{X,t}$. Thus, we have $\text{rank}(A_{X,t}) \leq |X| = k!\lambda$. We next consider the matrix $B = B_{X,k} = AA^T$. So clearly, $\text{rank}(B) \leq \text{rank}(A) \leq k!\lambda$. Our goal is to obtain a lower bound for $\text{rank}(B)$ which will imply a lower bound for λ. Consider for example the case of $n = 5$, $k = 4$, $t = 2$ and $\lambda = 1$ where a corresponding PSCA$(5,4)$, which is also a $4-(5,5,1)$ directed design proving that $g(5,4) = 1$, is given in Figure 1. Figure 2 shows the line of B which corresponds to the sequence 12.

12345	12543	51423	41523
13524	15342	14325	54132
52134	21453	24135	42513
23514	25341	52431	42315
53124	31452	43512	34125
32154	45321	32451	35421

Figure 1: A construction of a PSCA$(5,4)$ with $\lambda = 1$ showing that $g(5,4) = 1$.

	12	13	14	15	23	24	25	34	35	45	21	31	41	51	32	42	52	43	53	54
12	12	8	8	8	4	4	4	6	6	6	0	4	4	4	8	8	8	6	6	6

Figure 2: The row of $B = AA^T$ corresponding to the sequence 12 for $A = A_{X,2}$ where X is the PSCA from Figure 1.

To obtain a lower bound for $\text{rank}(B)$, let us look first more carefully at the case $k = 4$ (so $t = 2$) but for general n. There are only a few options for the entries of $B[ab, cd]$ where $ab, cd \in S_{n,2}$.

	cd	ab	ad	ba	ca	bd	cb
ab	6λ	12λ	8λ	0	4λ	4λ	8λ

Figure 3: The values of $B[ab, \kappa]$ for the various types of $\kappa \in S_{n,2}$. Here a, b, c, d are distinct.

Indeed, if $\{a, b\} \cap \{c, d\} = \emptyset$, then $B[ab, cd] = 6\lambda$ since the number of elements of $S_{n,4}$ in which a precedes b and c precedes d is precisely 6. If $a = c, b = d$, then $B[ab, ab] = 12\lambda$ since in precisely half of the permutations of X, a precedes b. Similarly one can immediately check that if $a = c, b \neq d$ then $B[ab, ad] = 8\lambda$, and so on. We have listed all possible configurations and their respective values in Figure 3. We notice that each entry of B is (obviously) a multiple of λ. Let C be obtained from B by dividing each element by the gcd of all the entries of B. In particular, this gcd is 2λ and notice that $\text{rank}(C) = \text{rank}(B)$. For a prime p, let $\text{rank}_p(C)$ denote the rank of C over the field $\mathbb{F}_p$. For the case $p = 2$, we see that over $\mathbb{F}_2$, C is now the binary matrix with $C[ab, cd] = 1$ if and only if $\{a, b\} \cap \{c, d\} = \emptyset$.

Consider the sub-matrix C' of C consisting of all rows ab such that $a < b$ and all columns cd such that $c < d$. Then, we can view C' as a matrix whose rows and columns are indexed by the unordered pairs of $[n]$, so over $\mathbb{F}_2$, C' is a binary matrix with $C'[X, Y] = 1$ if and only if $\{a, b\} \cap \{c, d\} = \emptyset$. Thus, $\text{rank}_2(C') \leq \text{rank}_2(C) \leq \text{rank}(C) = \text{rank}(B)$. But one can now observe that C' is precisely the set inclusion matrix of pairs versus subsets of order $n - 2$.

Set inclusion matrices have been introduced by Gottlieb [5] and have been extensively studied. Wilson [14] determined the rank of set inclusion matrices over finite fields - we next state his theorem. For integers $1 \leq t \leq \min\{r, n - r\}$ let $W_{t,r,n}$ denote the following matrix. Its rows are indexed by all t-subsets of $[n]$ and its columns by all r-subsets of $[n]$ and we have $W[T, R] = 1$ if $T \subseteq R$ and $W[T, R] = 0$ otherwise.

Lemma 2.1 *[Wilson [14]] Let p be a prime. Then $\text{rank}_p(W_{t,r,n})$ is*

$$\sum_{i \in D(r,t)} \binom{n}{i} - \binom{n}{i-1}$$

where $D(r, t)$ is the set of all integers i such that $0 \leq i \leq t$ and $\binom{r-i}{t-i} \neq 0 \bmod p$.

Corollary 2.2 *Let p be a prime. Then $\text{rank}_p(W_{t,r,n})$ is at least $\binom{n}{t} - \binom{n}{t-1}$.*

So, in our case above, C' equals $W_{2,n-2,n}$ (recall that the column indices are unordered pairs of $[n]$ but we can just rename them by their complements, which are $(n - 2)$-subsets of $[n]$). So by Corollary 2.2 we obtain that $rank_2(C') \geq n(n - 3)/2$. It follows that $\text{rank}(B) \geq n(n - 3)/2$. Recalling also that $\text{rank}(B) \leq 24\lambda$ we have $\lambda \geq n(n - 3)/48$. Hence, $g(n, 4) \geq n(n - 3)/48$.

We now generalize the argument to all even $k \geq 6$ such that $t = k/2$ is a prime (thus an odd prime). Consider $B[\kappa, \sigma]$ where $\kappa, \sigma \in S_{n,t}$. Suppose first that $\kappa \cap \sigma = \emptyset$ (meaning that no element appears in both sequences). Then the overall number of elements of $S_{n,k}$ that contain both κ and

σ as subsequences is precisely $\binom{k}{t}$, thus $B[\kappa,\sigma] = \lambda\binom{k}{t}$. But notice that since $t = k/2$ is an odd prime, then $\binom{k}{t}$ is not divisible by t, so $B[\kappa,\sigma]$ is not divisible by λt.

Suppose next that $\kappa \cap \sigma \neq \emptyset$. Let $U = \kappa \cup \sigma$ be the set of symbols used in at least one of them and notice that $t \le |U| \le k-1$. Let Q be the set of permutations of U that is consistent with both κ and σ, so $q \in Q$ if both κ and σ are subsequences of q. For example, suppose $k = 6$, $\kappa = 123$ and $\sigma = 269$, then $Q = \{12369, 12639, 12693\}$. Fix some $S \subset [n]$ with $U \cap S = \emptyset$ and $|U \cup S| = k$. So, in the last example we can take, say, $S = \{4\}$. Let P be the set of permutations of $U \cup S$ that is consistent with both κ and σ. So, each element of P is obtained by taking some $q \in Q$ and placing the elements of S in some locations. We therefore have that $|P| = |Q|\binom{k}{s}s!$ where $s = |S| \ge 1$ and that $B[\kappa,\sigma] = \lambda|P|$. But now notice that $B[\kappa,\sigma] = \lambda|P|$ is divisible by λk, hence by λt.

We have shown that $B[\kappa,\sigma]$ is not divisible by λt if and only if $\kappa \cap \sigma = \emptyset$. Let C be obtained from B by dividing each element by the gcd of all the entries of B (and recall that this gcd is divisible by λ), so $\text{rank}(C) = \text{rank}(B)$. We see that over $\mathbb{F}_t$, C is now a matrix with $C[\kappa,\sigma] \neq 0$ if and only if $\kappa \cap \sigma = \emptyset$ and furthermore, all nonzero entries of C are equal to the same nonzero element of $\mathbb{F}_t$, call it d. Consider the sub-matrix C' of C consisting of all rows κ corresponding to increasing sequences and all columns σ corresponding to increasing sequences. Then, we can view C' as a matrix whose rows and columns are indexed by the unordered t-subsets of $[n]$, so over $\mathbb{F}_t$, C' is a matrix with $C'[X,Y] = d$ if and only if $X \cap Y = \emptyset$. Thus, $\text{rank}_t(C') \le \text{rank}_t(C) \le \text{rank}(C) = \text{rank}(B)$. But now, $d^{-1}C'$ is the set inclusion matrix of t-subsets versus $n-t$ subsets, namely $W_{t,n-t,n}$. So, by Corollary 2.2, $\text{rank}_t(C') \ge \binom{n}{t} - \binom{n}{t-1}$. It follows that $\text{rank}(B) \ge \binom{n}{t} - \binom{n}{t-1}$. Recalling also that $\text{rank}(B) \le k!\lambda$ we have $\lambda \ge (\binom{n}{t} - \binom{n}{t-1})/k!$. Hence, $g(n,k) \ge (\binom{n}{k/2} - \binom{n}{k/2-1})/k!$.

We have thus proved that for all even k such that $k/2$ is a prime and for all $n \ge k$, the statement in Theorem 1 holds. To end the theorem we just recall that $g(n,k) \ge g(n,k-1)/k$ since a PSCA(n,k) with multiplicity λ is also a PSCA$(n,k-1)$ with multiplicity λk and recall the fact that the primes are dense in the sense that for every integer $k \ge 2$ there is always a prime between k and $k - O(k^{21/40}) = k - o(k)$ [1]. Hence we conclude that for all n sufficiently large, $g(n,k) > n^{k/2-o_k(1)}$. ∎

3 $g(n,3)$

The following three lemmas prove Theorem 2.

Lemma 3.1 $g(n,3) \ge n/6$.

Proof. Suppose that X is a PSCA$(n,3)$, let $A = A_{X,2}$ be the incidence matrix of ordered pairs w.r.t. X as defined in the previous section and let $B = AA^T$. Since A has $|X| = 6\lambda$ columns, $\text{rank}(B) \le 6\lambda$. As we cannot determine all elements of B, we will settle for a sub-matrix of B for which we can. Let C be the sub-matrix of B corresponding to the rows and columns indexed by the ordered pairs (i,n) for $i = 1,\ldots,n-1$ and also by the ordered pair $(n,1)$, which will be the index of the last row and column. (note: there are larger sub-matrices of B with the property

$$\begin{pmatrix} 3 & 2 & 2 & 2 & 2 & 0 \\ 2 & 3 & 2 & 2 & 2 & 1 \\ 2 & 2 & 3 & 2 & 2 & 1 \\ 2 & 2 & 2 & 3 & 2 & 1 \\ 2 & 2 & 2 & 2 & 3 & 1 \\ 0 & 1 & 1 & 1 & 1 & 3 \end{pmatrix}$$

Figure 4: The matrix C^* for $n = 6$.

that all of their elements can be determined, but they do not yield larger rank). So, C is an $n \times n$ matrix. We will prove that C is non-singular.

We observe that each diagonal entry of C is 3λ since there are precisely 3λ elements of S in which i precedes n for $i = 1, \ldots, n-1$ and similarly there are 3λ elements of S in which n precedes 1. Similarly, $C[(i,n),(j,n)] = 2\lambda$ for $i \neq j$ where $1 \leq i,j \leq n-1$, $C[(1,n),(n,1)] = C[(n,1),(1,n)] = 0$ and $C[(i,n),(n,1)] = C[(n,1),(i,n)] = \lambda$ for $i = 2, \ldots, n-1$. For simplicity, we divide all entries by λ and set $C^* := C/\lambda$. Figure 4 is an example of C^* in the case $n = 6$. It is not difficult to see by the matrix determinant lemma that $\det(C^*) = 3(n+1)$ so $\text{rank}(C) = n$, proving that $\text{rank}(B) \geq n$ and that $\lambda \geq n/6$. ■

Lemma 3.2 *Set $\lambda_1 = 1$ and $\lambda_r = 2(3^{\lceil r/2 \rceil}+1)\lambda_{\lceil r/2 \rceil}$ if $r \geq 2$. Then, for $r \geq 1$ we have $g(3^r, 3) \leq \lambda_r$.*

Proof. We prove the lemma by induction on r where the case $r = 1$ holds since $g(3,3) = 1$. Notice that since $g(n,3)$ is monotone non-decreasing in n, we only need to prove $g(3^r, 3) \leq \lambda_r$ for even r, since $\lambda_r = \lambda_{r+1}$ when r is odd. So, let r be even and assume that for $n = 3^{r/2}$ there is a PSCA$(n, 3)$ of multiplicity $\lambda_{r/2}$. We will prove that there is a PSCA$(n^2, 3) =$ PSCA$(3^r, 3)$ of multiplicity λ_r.

Suppose X is a PSCA$(n, 3)$ with multiplicity $\lambda = \lambda_{r/2}$ (hence $|X| = 6\lambda$). We will construct a PSCA$(n^2, 3)$, denoted by Y, such that $|Y| = 2(n+1)|X|$, and hence the lemma will follow by the definition of λ_r.

Our basic building block is a finite affine plane of order n, which exists since n is a prime power. This means, in particular, that there are $n+1$ partitions $P_1, \ldots, P_{n+1}$ of $[n^2]$, such that each P_i consists of n parts of size n each, denoted by $P_{i,j}$ for $j = 1, \ldots, n$ and such that for any pair of distinct elements of $[n^2]$, there is exactly one partition P_i that contains both of them in the same part of P_i.

We construct Y as a union of two sets W, Z of S_{n^2}, where $|W| = |Z| = (n+1)|X|$. We describe W and then describe Z. W will further be the union of $n+1$ sets $W_1, \ldots, W_{n+1}$ with $|W_i| = |X|$. We construct W_i using P_i and X. Each element of W_i will correspond to some $\sigma \in X$ as follows. For each $P_{i,j}$, fix some total order of its n elements (for example, the monotone increasing order). For $\sigma \in S_n$, let $\sigma(P_{i,j})$ be the permutation of $P_{i,j}$ corresponding to σ. Formally, if the total order of $P_{i,j}$ is $a_1, \ldots, a_n$ then $\sigma(P_{i,j})$ is the permutation $a_{\sigma(1)}, \ldots, a_{\sigma(n)}$. For $\sigma \in X$ let $\sigma(P_i)$ be the concatenation of $\sigma(P_{i,\sigma(1)}), \ldots, \sigma(P_{i,\sigma(n)})$. We call each part of this concatenation a *block*, so there are n blocks of size n each. We observe that $\sigma(P_i) \in S_{n^2}$ and set $W_i = \{\sigma(P_i) : \sigma \in X\}$. Thus, $W = \bigcup_{i=1}^{n+1} W_i$ is a well-defined subset of S_{n^2}.

Next, define Z to be following "reverse" of W. For a totally ordered set T, its reverse, denoted $rev(T)$ is the the total order which places the last element first, the second to last element second, and so on. Now for $\sigma \in X$ let $q(\sigma(P_i))$ be the concatenation of $rev(\sigma(P_{i,\sigma(1)})), \ldots, rev(\sigma(P_{i,\sigma(n)}))$. Set $Z_i = \{q(\sigma(P_i)) : \sigma \in X\}$ and $Z = \bigcup_{i=1}^{n+1} Z_i$. Finally, let $Y = W \cup Z$ and observe that indeed $|Y| = 2(n+1)|X|$ and $Y \subset S_{n^2}$.

To visualize our construction, consider for example the case $n = 3$ with $X = S_3$ being the trivial PSCA$(3,3)$ (with $\lambda = 1$). We will use the affine space of order 3 formed of $P_1 = \{123, 456, 789\}$, $P_2 = \{147, 258, 369\}$, $P_3 = \{159, 267, 348\}$, $P_4 = \{168, 249, 357\}$. Assume that in this listings, $P_{i,1}$ appears first, then $P_{i,2}$, then $P_{i,3}$ and that the listed order of each $P_{i,j}$ is the fixed total order (we have used here the monotone increasing order). So, for example, for $\sigma = 231 \in X$, we have, say $\sigma(P_{4,2}) = 492$ and $\sigma(P_4)$ is the concatenation of $\sigma(P_{4,2}), \sigma(P_{4,3}), \sigma(P_{4,1})$ so it is 492573681. Similarly, $q(\sigma(P_i))$ is 294375186.

It remains to prove that each element of $S_{n^2,3}$ appears as a subsequence of precisely $2(n+1)\lambda$ elements of Y, thereby proving that Y is a PSCA$(n^2, 3)$ of multiplicity λ_r. So, let $abc \in S_{n^2,3}$. We will distinguish between two cases. Assume first that $\{a, b, c\}$ is contained in some $P_{i,j}$ (in the case $n = 3$ this means that $\{a, b, c\}$ is the whole $P_{i,j}$ but for larger n this is strict containment). Then, since X is a PSCA$(n, 3)$ with multiplicity λ, we have that abc appears precisely λ times in W_i. If $i' \neq i$, then a, b, c appear in distinct blocks of each element of $W_{i'}$ (here we used the property of the affine plane). So, again, since X is a PSCA$(n, 3)$ with multiplicity λ, we have that abc appears precisely λ times in $W_{i'}$. The exact same arguments apply for Z_i and the $Z_{i'}$. Overall, abc appears as a subsequence of precisely $2(n+1)\lambda$ elements of Y.

Assume next that $\{a, b, c\}$ is not a subset of any $P_{i,j}$. Let γ be the unique index such that $\{a, b\}$ is a subset of some part of P_γ, let β be the unique index such that $\{a, c\}$ is a subset of some part of P_β and let α be the unique index such that $\{b, c\}$ is a subset of some part of P_α. Note that α, β, γ are indeed unique and distinct as follows from the properties of an affine plane. As in the previous case, we have that if $i \notin \{\alpha, \beta, \gamma\}$ then a, b, c appear in distinct blocks of each element of W_i so we have that abc appears precisely λ times in W_i, and similarly for Z_i. So abc appears $2(n-2)\lambda$ times in $\bigcup_{i \in [n+1] \setminus \{\alpha,\beta,\gamma\}} (W_i \cup Z_i)$. How many times does abc appear as a subsequence in W_β? The answer is 0, since in each element of W_β, a and c appear in the same block while b appears in another block. The same holds for Z_β. How many times does abc appear as a subsequence in $W_\alpha \cup Z_\alpha$? Since bc are in the same block of each element of $W_\alpha \cup Z_\alpha$ and since in precisely half of the elements of each of W_α and Z_α, the block containing a appears before the block containing both b, c (we use here the fact that a PSCA of triples is trivially also a PSCA of pairs), we have that precisely for half of the possible σ precisely one of $\sigma(P_\alpha)$ or $q(\sigma(P_\alpha))$ contains abc as a subsequence. So, overall, abc appears as a subsequence in $W_\alpha \cup Z_\alpha$ precisely $|X|/2$ times. The same argument holds for $W_\gamma \cup Z_\gamma$. In total, abc is a subsequence of

$$2(n-2)\lambda + 0 + |X| = 2(n+1)\lambda$$

where the last equality follows from $|X| = 6\lambda$. We have thus proved that each $abc \in S_{n^2,3}$ is a subsequence of precisely $2(n+1)\lambda$ elements of Y, as required. ■

It is easy to prove by induction that for $r = 2^t$ we have $\lambda_r = 2^{t-1}(3^r - 1)$ hence for n which is of the form 3^{2^t} we obtain from Lemma 3.2 that $g(n,3) \leq \frac{1}{2} n \log_3 n$. The next lemma provides an upper bound that applies to all values of n.

Lemma 3.3 *For all $n \geq 3$ we have $g(n,3) \leq Cn(\log n)^{\log 7}$ for some absolute constant C.*

Proof. We first prove that the lemma holds for $n = 3^r$ where r is an integer. Let $t = \lceil \log r \rceil$. We will prove by induction that

$$\lambda_r \leq 7^t 3^r$$

where λ_r is as defined in Lemma 3.2. Note that this holds for $\lambda_1 = 1$ and for $\lambda_2 = 8$. Since $\lambda_r = \lambda_{r+1}$ when r is odd, it suffices to prove $\lambda_r \leq 7^t 3^r$ when r is odd. Notice that if r is odd, then $\lceil \log((r+1)/2) \rceil \leq t - 1$ so by the definition of λ_r we have for r odd and the induction hypothesis that

$$\begin{aligned}
\lambda_r &= 2(3^{(r+1)/2} + 1)\lambda_{(r+1)/2} \\
&\leq 2(3^{(r+1)/2} + 1)7^{t-1}3^{(r+1)/2} \\
&= 2(3^{(r+1)/2} + 1)(7/3)^{t-1} \cdot 3^{t-1}3^{(r+1)/2} \\
&\leq (7/3) \cdot 3^{(r+1)/2} \cdot (7/3)^{t-1} \cdot 3^{t-1} \cdot 3^{(r+1)/2} \\
&= 7^t 3^r \,.
\end{aligned}$$

So, whenever n is of the form 3^r we have that $g(n,3) \leq 7^t 3^r$, where $t = \lceil \log \log_3 n \rceil$. If n is not of this form, let n' be the unique power of 3 such that $n \leq n_0 < 3n$ and since $g(n,3) \leq g(n',3)$ we have $g(n,3) \leq 7^t 3^r$ where $r = \log_3 n'$ and $t = \lceil \log \log_3 n' \rceil$. Hence, $g(n,3) \leq Cn(\log n)^{\log 7}$ for an absolute constant C. ■

We end this section with a proof that $g(5,3) = 2$, which is currently the only explicitly determined value of $g(n,k)$ which is not one.

Proposition 3.4 $g(5,3) = 2$.

Proof. Recall from the introduction that $g(5,3) > 1$ [10], hence we only need to prove $g(5,3) \leq 2$. We construct a PSCA$(5,3)$ with $\lambda = 2$. It is not difficult to compute all sets of six permutations that cover a maximum number of sequences. As it turns out, there are such sets that cover 56 elements of $S_{5,3}$. For example, the following is such:

$$X = \{12345, 43215, 35214, 14523, 25413, 53412\}$$

The only sequences uncovered by X are $132, 231, 154, 451$. On the other hand, the sequences $123, 321, 145, 541$ are each covered twice. For $\sigma \in S_n$ and for $X \subseteq S_n$, let $X_\sigma = \{\pi\sigma \,:\, \pi \in X\}$. Now, consider $\sigma = 13254$. Then, for X above we obtain that

$$X_\sigma = \{13254, 52314, 24315, 15432, 34512, 42513\}$$

The only sequences uncovered by X are $123, 321, 145, 541$. On the other hand, the sequences $132, 231, 154, 451$ are each covered twice. Hence $X \cup X_\sigma$ is a PSCA$(5, 3)$. ■

4 Open problems

Theorem 1 proves that for every fixed $k \geq 3$, $g(n, k)$ is lower bounded by a polynomial in n whose exponent grows with k. While it is not difficult to slightly improve upon the trivial upper bound $g(n, k) \leq n!/k!$, it would be interesting to obtain polynomial upper bounds for $g(n, k)$.

Theorem 2 proves that $g(n, 3)$ is at least linear and not more than quasi-linear in n. It would be interesting to determine the right order of magnitude of $g(n, 3)$.

Proving additional exact values of $g(n, k)$ which are not of unit multiplicity in addition to $g(5, 3)$ also seems challenging.

Acknowledgment

The author thanks the referees for useful comments.

References

[1] R. Baker, G. Harman, and J. Pintz. The difference between consecutive primes, II. *Proceedings of the London Mathematical Society*, 83(03):532–562, 2001.

[2] Y. Chee, C. Colbourn, D. Horsley, and J. Zhou. Sequence covering arrays. *SIAM Journal on Discrete Mathematics*, 27(4):1844–1861, 2013.

[3] C. Colbourn and J. Dinitz. *Handbook of Combinatorial Designs.* CRC press, second edition edition, 2006.

[4] Z. Füredi. Scrambling permutations and entropy of hypergraphs. *Random Structures & Algorithms*, 8(2):97–104, 1996.

[5] D. Gottlieb. A certain class of incidence matrices. *Proceedings of the American Mathematical Society*, 17(6):1233–1237, 1966.

[6] Y. Ishigami. Containment problems in high-dimensional spaces. *Graphs and Combinatorics*, 11(4):327–335, 1995.

[7] Y. Ishigami. An extremal problem of d permutations containing every permutation of every t elements. *Discrete Mathematics*, 159(1-3):279–283, 1996.

[8] D. Kuhn, J. Higdon, J. Lawrence, R. Kacker, and Y. Lei. Combinatorial methods for event sequence testing. In *Fifth International Conference on Software Testing, Verification and Validation (ICST)*, pages 601–609. IEEE, 2012.

[9] V. Levenshtein. Perfect codes in the metric of deletions and insertions. *Diskretnaya Matematika (English translation: Discrete Mathematics and Applications, 1992, 2:3, 241–258)*, 3(1):3–20, 1991.

[10] R. Mathon and T. Van Trung. Directed t-packings and directed t-Steiner systems. *Designs, Codes and Cryptography*, 18(1-3):187–198, 1999.

[11] J. Radhakrishnan. A note on scrambling permutations. *Random Structures & Algorithms*, 22(4):435–439, 2003.

[12] J. Spencer. Minimal scrambling sets of simple orders. *Acta Mathematica Hungarica*, 22(3-4):349–353, 1972.

[13] J. Tarui. On the minimum number of completely 3-scrambling permutations. *Discrete Mathematics*, 308(8):1350–1354, 2008.

[14] R. M. Wilson. A diagonal form for the incidence matrices of t-subsets vs. k-subsets. *European Journal of Combinatorics*, 11(6):609–615, 1990.