Journal of Combinatorics
Volume 4, Number 4, 449–456, 2013

A Sidon-type condition on set systems

Peter J. Dukes* and Jane Wodlinger*

Consider families of k-subsets (or blocks) on a ground set of size v. Recall that if all t-subsets occur with the same frequency λ, one obtains a t-design with index λ. On the other hand, if all t-subsets occur with different frequencies, such a family has been called (by Sarvate and others) a t-adesign. An elementary observation shows that such families always exist for $v > k \geq t$. Here, we study the smallest possible maximum frequency $\mu = \mu(t, k, v)$.

The exact value of μ is noted for $t = 1$ and an upper bound (best possible up to a constant multiple) is obtained for $t = 2$ using PBD closure. Weaker, yet still reasonable, asymptotic bounds on μ for higher t follow from a probabilistic argument. Some connections are made with the famous Sidon problem of additive number theory.

1. Introduction

Given a family (which may contain repetition) $\mathcal{A}$ of subsets of a ground set X, the *frequency* of a set $T \subset X$ is the number of elements of $\mathcal{A}$ (counting multiplicity) which contain T.

Let $v \geq k \geq t$ be nonnegative integers. A *t-design*, or $S_\lambda(t, k, v)$, is a pair $(V, \mathcal{B})$ where $\mathcal{B}$ is a family of k-subsets of V such that every t-subset has the same frequency λ. Typically, V is called a set of *points*, $\mathcal{B}$ are the *blocks*, t is the *strength* (reflecting that t-designs are also i-designs for $i \leq t$) and λ is the *index*. Repeated blocks are normally permitted in the definition.

There are 'divisibility' restrictions on the parameters v, k, t, λ and beyond that very little is known in general about the existence of $S_\lambda(t, k, v)$. There are some trivial cases, such as $t = 0$, $t = k$, or $k = v$, and some mildly interesting ones: $\lambda = t = 1$ leads to uniform partitions; $\lambda = \binom{v-t}{k-t}$ is realized via the complete k-uniform hypergraph of order v. For $t = 2$ and fixed k there is a rich and deep asymptotic existence theory due to R.M. Wilson; see [5]. Spherical geometries and Hadamard matrices lead to some examples for $t = 3$.

*This research is supported by NSERC.

In [4], Sarvate and Beam consider an interesting twist on the definition. A t-*adesign* is defined as a pair $(V, \mathcal{A})$, where V is a ground set of v points and $\mathcal{A}$ is a collection of blocks of size k, satisfying the condition that every t-subset of points has a **different** frequency.

Here, we abbreviate a t-adesign with $A(t, k, v)$. It is easy to see that such families always exist for integers $v > k \geq t \geq 1$: simply assign multiplicities to $\binom{V}{k}$ according to different powers of two.

This begs a more intricate question. Let $\mu(t, k, v)$ denote the smallest maximum frequency, taken over all adesigns $A(t, k, v)$. The main question motivating this article is the following.

Problem 1.1. Given t, k, v, determine (or bound) $\mu(t, k, v)$.

In most of the previous investigations on adesigns, the cases of interest have been for $t = 2$ and when the different pairwise frequencies are $1, 2, \ldots, \binom{v}{2}$. It should be noted that here we allow zero as a frequency, although, if desired, it is not hard to bump up all frequencies to be positive.

From the definitions and easy observations above, we have

$$\binom{v}{t} - 1 \leq \mu(t, k, v) < 2^{\binom{v}{k}}. \tag{1}$$

However, the basic upper bound in (1) is unsatisfactory, at least asymptotically in v. Our main goal is a substantial reduction of the upper bound (to something independent of k).

Theorem 1.2. *For $k > 2t + 2$ and sufficiently large v,*

$$\mu(t, k, v) \leq 16tv^{2t+2} \log v.$$

The constant is surely not best possible; however, we are content until more is known about the exponent.

We can do much better when $t \leq 2$. For $t = 1$, an elementary argument gives the exact value of μ. And for $t = 2$, Wilson's theory of PBD closure reduces the upper bound on μ to a constant multiple of its lower bound.

Theorem 1.3. *For positive integers $v > k$,*

$$\mu(1, k, v) = \begin{cases} v - 1 & \text{if } 2k \leq v \text{ and } \binom{v}{2} \equiv 0 \pmod{k}, \\ v & \text{if } 2k \leq v \text{ and } \binom{v}{2} \not\equiv 0 \pmod{k}, \\ \left\lceil \frac{1}{v-k} \binom{v}{2} \right\rceil & \text{if } 2k > v. \end{cases}$$

Theorem 1.4. *There is a constant $C = C(k)$ such that $\mu(2, k, v) \leq Cv^2$.*

The proof of Theorem 1.2 follows a probabilistic argument and occurs in Section 2. The proofs of Theorems 1.3 and 1.4 are given in Section 3.

Before beginning our detailed investigations, we should mention some connections with a central topic in additive combinatorics. Briefly, a *Sidon sequence* (or *Golomb ruler*) is a list of positive integers whose pairwise sums are all distinct, up to swapping summands. More generally, a B_r*-sequence* or *Sidon sequence of order* r has the property that all its r-wise sums are distinct. It is known (see [3]) that the largest cardinality $F_r(n)$ of a Sidon sequence of order r contained in $[n]$ satisfies

$$n^{1/r}(1 - o(1)) \leq F_r(n) \leq C(r)n^{1/r}. \tag{2}$$

Now consider an adesign $A(t, v-1, v)$, where V is the ground set of size $v = k+1$. Assign multiplicity $f(x)$, chosen from a Sidon sequence of order t, to the 'co-singleton' set $V \setminus \{x\}$, $x \in V$. The inherited weight on a t-subset T is $\sum_{x \notin T} f(x)$. By construction, this takes distinct values on all t-subsets. From this and (2), we see that $\mu(t, v-1, v) \leq Cv^t$, which is best possible up to a constant multiple. However, it is also clear that the exact determination of μ, even in the case $v = k+1$, is as difficult as the Sidon problem.

2. The general bound

We prove Theorem 1.2 by employing B_r-sequences along the lines of the discussion concluding Section 1. But here, a probabilistic selection is needed to control the upper bound on μ.

Proof of Theorem 1.2. Assume $t > 1$, appealing to Theorem 1.3. Suppose first that v is a prime power. Bose and Chowla [1] construct a B_t-sequence of size v in $[v^t]$. Let V be such a set of integers and consider the family $\mathcal{B}$ of all k-subsets of V, where a k-set K is taken with multiplicity

$$f(K) = \sum_{m \in V \setminus K} m.$$

Then the frequency of a t-subset T in $\mathcal{B}$ is

$$f(T) = \sum_{K \supseteq T, |K|=k} f(K) = \binom{v-t-1}{k-t} \sum_{m \in V \setminus T} m. \tag{3}$$

By choice of V, these are all distinct frequencies. Observe that $\sum_{m \in V \setminus T} m < v^t(v-t)$, so that $f(T)$ is at most a polynomial of order v^{k+1}.

Consider next a family $\mathcal{A} = \mathcal{A}(p)$ consisting of each element of $\mathcal{B}$ chosen independently with probability p. We claim there is some p guaranteeing an adesign $A(t, k, v)$ of the required form.

Let $f_{\mathcal{A}}(T)$ denote the frequency of T in $\mathcal{A}$. This is a sum of $f(T)$ independent binomial random variables X_i, one for each k-set in $\mathcal{B}$ containing T. So $f_{\mathcal{A}}(T)$ has expected value $\mu = pf(T)$ by linearity.

Now let's invoke a (weak but tidy) two-sided Chernoff bound of the form

$$\mathbb{P}\left[\left|\sum X_i - \mu\right| > 2\sqrt{\mu \log 1/\epsilon}\right] < \epsilon,$$

which holds for $\epsilon > \exp(-\mu/4)$. Taking $\epsilon = \binom{v}{t}^{-1}$, we conclude that there exists (with positive probability) a family $\mathcal{A}$ such that

$$|f_{\mathcal{A}}(T) - pf(T)| < \sigma(T), \tag{4}$$

for **every** t-set T, where $\sigma(T) := 2\sqrt{pf(T)\log\binom{v}{t}}$, and for each p with $pf(T) > 4\log\binom{v}{t}$.

It remains to check that frequencies $f_{\mathcal{A}}(T)$ remain distinct and appropriately bounded for some choice of p. By (3) and (4), we have distinct frequencies provided that

$$2\sigma(T) < p\binom{v-t-1}{k-t}.$$

Using the definition of $\sigma(T)$ and $\sum_{m \notin T} m < v^t(v-t)$, it suffices to have

$$16v^t(v-t)\log\binom{v}{t} < p\binom{v-t-1}{k-t}. \tag{5}$$

The right side of (5) grows faster than the left for $k \geq 2t+2$; hence, for sufficiently large v, we can choose $p = 16v^{t+1}\log\binom{v}{t}/\binom{v-t-1}{k-t} < 1$ (easily permitting application of the Chernoff bound above).

For such a choice, we have

$$\max_T f_{\mathcal{A}}(T) < pf(T) + \sigma(T) < (16v^{2t+2} + 8v^{t+1})\log\binom{v}{t}.$$

The bound $\log\binom{v}{t} \leq t\log v - \log t!$ leaves enough room to eliminate the lower-order term and imply the stated bound.

Finally, if v is not a prime power, we can simply apply the above argument to a prime $v' \leq v + o(v)$ to obtain asymptotically the same result. □

3. The cases $t = 1$ and $t = 2$

When $t = 1$, we simply demand that every point is in a different number of blocks. A complete characterization is possible here, following a technique known to Sarvate and Beam in early investigations. To the best of our knowledge, though, Theorem 1.3 has not been worked out for general v and k.

The proof strategy is as follows. Suppose $f(1) < \cdots < f(v)$ are desired pointwise frequencies whose sum F is divisible by k. Set up $b = F/k$ blocks, and place element '1' in the first $f(1)$ blocks, element '2' in the next $f(2)$ blocks, and so on, with blocks identified modulo b. In other words, the ith block contains those elements x such that

$$\sum_{1 \leq y < x} f(y) < bq + i \leq \sum_{1 \leq y \leq x} f(y)$$

for some integer $q \in \{0, 1, \ldots, k-1\}$. Care must be taken that the maximum frequency $f(v)$ does not exceed b, the number of blocks. Ideally, the frequencies are chosen to be consecutive, or almost consecutive, integers.

Proof of Theorem 1.3. We apply the above construction using a run of (almost) consecutive prescribed frequencies. There is a division into two main cases.

Case 1. $2k \leq v$. Suppose first that $k \mid \binom{v}{2}$. Fill $b = \binom{v}{2}/k$ blocks with pointwise frequencies $0, 1, \ldots, v-1$. Note that $b \geq v-1$ follows from the assumption $2k \leq v$. On the other hand, if $\binom{v}{2} = bk - r$, $0 < r < k$, use b blocks with frequencies $0, 1, \ldots, v-r-1, v-r+1, \ldots, v$. One has sum of frequencies $bk = \binom{v+1}{2} - (v-r) = \binom{v}{2} + r$, as required. In either sub-case, the smallest possible maximum frequency is realized and we have

$$\mu(1,k,v) = \begin{cases} v-1 & \text{if } \binom{v}{2} \equiv 0 \pmod{k}, \\ v & \text{if } \binom{v}{2} \not\equiv 0 \pmod{k}. \end{cases}$$

Case 2. $2k > v$. We first show that the given value $\lceil \frac{1}{v-k}\binom{v}{2} \rceil$ is a lower bound on $\mu(1,k,v)$. Suppose m is the maximum frequency in an adesign $A(1,k,v)$. Then

$$mk \leq bk \leq (m-v+1) + \cdots + (m-1) + m = mv - \binom{v}{2}.$$

In other words, m is an integer with $m(v-k) \geq \binom{v}{2}$ and the lower bound follows. Conversely, we must realize the given value $\mu := \lceil \frac{1}{v-k}\binom{v}{2} \rceil$ as the maximum frequency in an adesign $A(1,k,v)$. Put $bk = \mu v - \binom{v}{2} - r$, for some positive integer b and $0 \leq r < k$. Again, use the strategy preceding the statement of the theorem, with $b = \frac{1}{k}(\mu v - \binom{v}{2} - r)$ blocks and frequencies

$$\mu - v, \dots, \mu - v - r - 1, \mu - v - r + 1, \dots, \mu.$$

It remains to check that $\mu \leq b$. However, this follows easily since μ is the least integer with $\mu(v-k) \geq \binom{v}{2}$. Therefore, $\mu k \leq \mu v - \binom{v}{2}$. On the other hand, b is the greatest integer so that $bk \leq \mu v - \binom{v}{2}$. □

We turn now to adesigns with $t = 2$. An important tool here is 'PBD closure,' which we briefly outline. Let K be a set of positive integers, each at least two. A *pairwise balanced design* PBD(v,K) is a set of v points, together with a set of blocks whose sizes are in K, having the property that every unordered pair of different points is contained in exactly one block. Wilson's theorem [5] asserts that the necessary 'global' and 'local' divisibility conditions on v given K are asymptotically sufficient for the existence of PBD(v,K).

A key observation for the proof of Wilson's theorem is the 'breaking up blocks' construction: a block, say of size u, of a PBD can be replaced with the family of blocks of a PBD on u points. In particular, if there exists a PBD(v,K) and an $S_\lambda(2,k,u)$ for every $u \in K$, then there exists an $S_\lambda(2,k,v)$.

It was observed in [2] that adesigns actually obey a similar recursion. The basic idea is to place adesigns (instead of designs) on the blocks of a PBD. However, each such adesign needs to be accompanied with a block design on those points with sufficiently large λ so as to 'spread out' the pairwise frequencies. When restated using μ, one obtains the following result.

Lemma 3.1. *Suppose there exists a PBD(v,K) with b blocks having sizes $u_1, u_2, \dots, u_b$. Put $M_0 = 0$ and for $0 < i \leq b$,*

$$M_i = \min\{\lambda \geq M_{i-1} : \exists S_\lambda(2,k,u_i)\} + \mu(2,k,u_i). \tag{6}$$

Then $\mu(2,k,v) \leq M_b$.

Remark. The minimum in (6) is well defined; more generally, $S_\lambda(t,k,v)$ exists for a smallest positive integer $\lambda = \lambda_{\min}(v,k) \leq \binom{v-t}{k-t}$, and such designs can be repeated with arbitrary multiplicity.

We are now ready to prove the quadratic upper bound on $\mu(2,k,v)$.

Proof of Theorem 1.4. For v large and $K = \{k+1, k+2, k+3\}$, apply Lemma 3.1 to a PBD(v, K). Note that such PBDs exist for all sufficiently large integers v. This follows easily from Wilson's theorem since the three consecutive block sizes lead to no 'divisibility' restrictions on v (globally, $\gcd\{k(k+1), (k+1)(k+2), (k+2)(k+3)\} = 2$ always divides $\binom{v}{2}$; locally, $\gcd\{k, k+1, k+2\} = 1$ divides $v-1$).

Put $m = \max\{\mu(2, k, k+j) : j = 1, 2, 3\}$ and $l = \max\{\lambda_{\min}(k+j, k) : j = 1, 2, 3\}$. Observe that l and m depend only on k. Also, observe that the number of blocks b of a PBD(v, K) satisfies $b \le \binom{v}{2}/\binom{k+1}{2}$, since $k+1$ is the smallest block size. Combining these facts, it follows that

$$\mu(2, k, v) \le lmb \le C(k)\binom{v}{2}. \qquad \square$$

4. Discussion

There is another noteworthy construction of t-adesigns by combining copies of systems which are nearly designs. The general idea to work from a family $\widehat{\mathcal{B}}_T$ of k-subsets such that one preferred t-subset T has frequency λ_1 and all other t-subsets have frequency $\lambda_2 < \lambda_1$. (Such families can be found, for instance, via a linear algebraic argument upon 'clearing denominators.') Then, take copies of $\widehat{\mathcal{B}}_T$ with distinct multiplicities over each $T \in \binom{V}{t}$. The crude bound obtained in this way is $\mu(t, k, v) \le C_1\lambda_1 v^t + C_2\lambda_2 v^{2t}$. However, we presently see no way of keeping λ_2 small enough in general. This would be an interesting problem in its own right. When such families $\widehat{\mathcal{B}}_T$ exist with reasonable λ_2, it is possible to improve Theorem 1.2.

The remaining work for $t = 2$ essentially amounts to a reduction in the multiplicative constant in Theorem 1.4. There are some ideas which seem promising in this direction. For instance, $\mu(2, 3, v)$ was completely determined in [2] using a blend of PBD closure, group divisible designs, and a variation on 'anti-magic cubes.' The latter concerns a neat side-problem: place nonnegative integers in the cells of the cube $[n]^3$ so that the $3n^2$ line sums are distinct and with maximum value as small as possible. Interesting constructions yielding line sums $\{0, 1, \dots, 3n^2 - 1\}$ were found for $n = 2, 3, 5, 7$, and products of these values.

Returning to $\mu(2, 3, v)$, a slightly technical argument shows that the maximum frequency for triples versus pairs in an adesign is best possible.

Theorem 4.1 ([2])**.** *For all $v > 3$,*

$$\mu(2, 3, v) = \begin{cases} \binom{v}{2}, & \text{if } v = 4 \text{ or } v \equiv 2 \pmod 3, \\ \binom{v}{2} - 1, & \text{otherwise.} \end{cases}$$

We omit further analysis of $\mu(2,k,v)$ for $k > 3$ until better general constructions surface for small v relative to k. For fixed k, the complete determination of $\mu(2,k,v)$ can probably be reduced to a finite problem. Quite possibly $\mu(2,k,v) = \binom{v}{2} - 1 + o(v)$ for each k.

Acknowledgement

The authors are indebted to the anonymous referee who suggested the probabilistic method, yielding the present form of Theorem 1.2.

References

[1] R. C. Bose and S. Chowla (1962/1963). Theorems in the additive theory of numbers. *Comment. Math. Helv.* **37**, 141–147. MR0144877

[2] P. J. Dukes and J. A. Short-Gershman (2012). A complete existence theory for Sarvate-Beam triple systems. *Australas. J. Combin.* **54**, 261–272. MR3013255

[3] K. O'Bryant (2004). A complete annotated bibliography of work related to Sidon sequences. *Elec. J. Combin.* **#DS11**, 39pp.

[4] D. Sarvate and W. Beam (2007). A new type of block design. *Bull. Inst. Combin. Appl.* **50**, 26–28. MR2317084

[5] R. M. Wilson (1975). An existence theory for pairwise balanced designs, III: Proof of the existence conjectures. *J. Comb. Theory, Ser. A* **18**, 71–79. MR0366695

Peter J. Dukes
Mathematics and Statistics
University of Victoria
Victoria, BC
Canada
E-mail address: dukes@uvic.ca

Jane Wodlinger
Mathematics and Statistics
University of Victoria
Victoria, BC
Canada
E-mail address: jw@uvic.ca

Received December 5, 2012