Combinatorics, Probability and Computing (2012) 21, 159–177.
doi:10.1017/S0963548311000563

2-Cancellative Hypergraphs and Codes

ZOLTÁN FÜREDI†

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
and
Rényi Institute of Mathematics of the Hungarian Academy of Sciences, Budapest, PO Box 127, Hungary-1364
(e-mail: z-furedi@illinois.edu, furedi@renyi.hu)

Received 10 April 2011; revised 11 October 2011; first published online 2 February 2012

A family of sets $\mathcal{F}$ (and the corresponding family of 0–1 vectors) is called *t-cancellative* if, for all distinct $t+2$ members $A_1, \dots, A_t$ and $B, C \in \mathcal{F}$,

$$A_1 \cup \cdots \cup A_t \cup B \neq A_1 \cup \cdots \cup A_t \cup C.$$

Let $c_t(n)$ be the size of the largest t-cancellative family on n elements, and let $c_t(n,r)$ denote the largest r-uniform family. We improve the previous upper bounds, *e.g.*, we show $c_2(n) \leqslant 2^{0.322n}$ (for $n > n_0$). Using an algebraic construction we show that $c_2(n, 2k) = \Theta(n^k)$ for each k when $n \to \infty$.

1. Introduction, definitions

There are many instances in coding theory when codewords must be restored from partial information, such as defective data (error correcting codes) or some superposition of the strings (these can lead to Sidon sets, sum-free sets, *etc.*). A family of sets $\mathcal{F}$ (and the corresponding family of 0–1 vectors) is called *cancellative* if A and $A \cup B$ determine B (in the case of $A, B \in \mathcal{F}$ and $A \neq A \cup B$). For a precise definition we require that for all $A, B, C \in \mathcal{F}$, $A \neq B$, $A \neq C$,

$$A \cup B = A \cup C \Longrightarrow B = C.$$

Let $c(n)$ be the size of the largest cancellative family on n elements, and let $c(n,r)$ denote the size of the largest r-uniform family on n elements. This definition can be extended (as in the abstract). In this paper we focus on 2-cancellative r-uniform hypergraphs, *i.e.*, families of r-sets, and 2-cancellative *codes*, where there is no restriction on the sizes of the hyperedges.

† Research supported in part by the Hungarian National Science Foundation OTKA, and by the National Science Foundation under grant NFS DMS 09-01276.

Speaking about a hypergraph $\mathbb{F} = (V, \mathcal{F})$, we frequently identify the vertex set $V = V(\mathbb{F})$ with the set of integers $[n] := \{1, 2, \dots, n\}$, or some elements of a q-element finite field $\mathbf{F}_q$. To abbreviate notation we say 'hypergraph $\mathcal{F}$' (or set system $\mathcal{F}$), thus identifying $\mathbb{F}$ with its edge set $\mathcal{F}$. The *degree*, $\deg_{\mathbb{F}}(x)$, of an element $x \in V$ is the number of hyperedges of $\mathcal{F}$ containing x. The hypergraph $\mathcal{F}$ is *uniform* if every edge has the same number of elements; r-uniform means $|F| = r$ for all $F \in \mathcal{F}$. An r-uniform hypergraph $(V, \mathcal{F})$ is called *r-partite* if there exists an r-partition of V, $V = V_1 \cup \cdots \cup V_r$, such that $|F \cap V_i| = 1$ for all $F \in \mathcal{F}$, $i \in [r]$.

Let $f(n, P_1, P_2, \dots)$ denote the maximum number of subsets which can be selected from $\{1, \dots, n\}$ satisfying all the properties $P_1, P_2, \dots$. With this notation $c_t(n) := f(n, t\text{-CANC})$, where t-CANC stands for t-cancellativeness.

A hypergraph is *linear* if $|E \cap F| \leqslant 1$ holds for every pair of edges. An $(n, r, 2)$-*packing* is a linear r-uniform hypergraph $\mathcal{P}$ on n vertices. Obviously, $|\mathcal{P}| \leqslant \binom{n}{2}/\binom{r}{2}$. If equality holds, then $\mathcal{P}$ is called an $S(n, r, 2)$ *Steiner system.*

2. Cancellative and locally thin families

The asymptotics of the maximum size of a cancellative family was given by Tolhuizen [45] (construction) and in [19] (upper bound), showing that there exists a $\gamma > 0$ such that

$$\frac{\gamma}{\sqrt{n}} 1.5^n < c(n) < 1.5^n.$$

The problem was proposed by Erdős and Katona [26], who conjectured that $c(n) = \Theta(3^{n/3})$, which was disproved by an elegant construction by Shearer showing that $c(3k) \geqslant k3^{k-2}$, leading to $c(n) > 1.46^n$ for $n > n_0$. Since a product of two cancellative families is again cancellative, we have $c(n+m) \geqslant c(n)c(m)$. Thus $\lim c(n)^{1/n}$ exists. This is not known for 2-cancellative hypergraphs, so Körner and Sinaimeri [32] introduced

$$t(4) := \limsup_{n \to \infty} \frac{1}{n} \log_2 c_2(n)$$

and proved $0.11 < t(4) \leqslant 0.42$. As usual all logarithms are to base two. The lower bound follows from a standard probabilistic argument. We will show that $t(4) \leqslant \log_2 5 - 2 = 0.3219\dots$.

Theorem 2.1. $c_2(n) < 9\sqrt{n}\left(\frac{5}{4}\right)^n$.

The proof is postponed to Section 6. Without loss of generality we can suppose that the n-element underlying set of $\mathcal{F}$ is $[n]$. We associate to every subset $A \in \mathcal{F}$ its characteristic binary vector, $\mathbf{x} := \mathbf{x}(A) = (x_1, \dots, x_n)$, with $x_i = 1$ if $i \in A$ and $x_i = 0$ otherwise. One can immediately see that requiring the family $\mathcal{F}$ to be t-cancellative is equivalent to its representation set of binary vectors, satisfying the following: for every $(t+2)$-tuple $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(t+2)})$ of distinct vectors in the set (considered in an arbitrary but fixed order), there exist at least $t+1$ different values of $k \in [n]$, such that the corresponding ordered $(t+2)$-tuples $(x_k^{(1)}, x_k^{(2)}, \dots, x_k^{(t+2)})$ are all different, while for each of them we have the sum $x_k^{(1)} + x_k^{(2)} + \cdots + x_k^{(t+2)} = 1$. In hypergraph language, at least $t+1$ of the sets among the $t+2$ have degree-one vertices. This problem can be seen in a more general context. We can require that for every ordered a-tuple of vectors in the set, there exist at least b different columns, which sum up to 1. Such a family is called *locally* (a,b)-*thin*. Let

$$t(a,b) := \limsup_{n\to\infty} \frac{1}{n} \log_2 f(n, \text{ locally } (a,b)\text{-thin}).$$

We then have $t(a,1) \geqslant t(a,2) \geqslant \cdots \geqslant t(a,a)$.

This problem was investigated by Alon, Fachini, Körner and Monti [2, 3, 17]. They proved that $t(4,1) < 0.4561\ldots$ and $t(a,1) < 2/a$ for all even a, and

$$\Omega\left(\frac{1}{a}\right) \leqslant t(a,1) \leqslant O\left(\frac{\log a}{a}\right) < 0.793$$

for all a. This is a notoriously hard problem. In particular, we do not even know whether $t(3,1) < 1$ (see Erdős and Szemerédi [16]).

Concerning one of the most interesting cases, the case $a=4$, a locally $(4,1)$-thin family is also *weakly union-free* ($A\cup B = C\cup D$ implies $\{A,B\}=\{C,D\}$). The best upper bound,

$$\log_2 f(n, \text{ weakly union-free}) < (0.4998\cdots + o(1))n,$$

is due to Coppersmith and Shearer [8]. Nothing non-trivial is known about $t(4,2)$. Our Theorem 2.1 implies $t(4,3) \leqslant \log_2 5 - 2 = 0.3219\ldots$. One can find more similar problems in the survey article by Körner [30] and in the more recent paper by Körner and Monti [31].

3. Cancellative and cover-free families

A family $\mathcal{F} \subseteq 2^{[n]}$ is g-*cover-free* if it is locally $(g+1, g+1)$-thin. In other words, for arbitrary distinct members $A_0, A_1, \ldots, A_g \in \mathcal{F}$,

$$A_0 \not\subseteq \bigcup_{i=1}^{g} A_i.$$

Let $C_g(n)$ ($C_g(n,r)$) be the maximum size of a g-cover-free n vertex code (r-uniform hypergraph, respectively). Clearly, $C_g(n) \leqslant C_{g-1}(n) \leqslant \cdots \leqslant C_1(n)$ and $C_g(n,r) \leqslant C_{g-1}(n,r) \leqslant \cdots \leqslant C_2(n,r)$. Note that $C_{t+1}(n) \leqslant c_t(n)$ (and $C_{t+1}(n,r) \leqslant c_t(n,r)$) since a $t+1$-cover-free family is t-cancellative as well.

Union-free and cover-free families were introduced by Kautz and Singleton [27]. They studied binary codes with the property that the disjunctions (bitwise ORs) of distinct at most g-tuples of codewords are all different. In information theory these codes are usually called *superimposed*, and they have been investigated in several papers on multiple access communication (see, *e.g.*, Nguyen Quang A and Zeisel [34], D'yachkov and Rykov [10] and Johnson [25]). The same problem has been posed – in different terms – by Erdős, Frankl and Füredi [12, 13] in combinatorics, by Sós [44] in combinatorial number theory, and by Hwang and Sós [24] in group testing. For recent generalizations see, *e.g.*, Alon and Asodi [1], and De Bonis and Vaccaro [9]. D'yachkov and Rykov [10] proved that

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

there are positive constants α_1 and α_2 such that

$$\alpha_1 \frac{1}{g^2} < \frac{\log C_g(n)}{n} < \alpha_2 \frac{\log g}{g^2} \tag{3.1}$$

for every g and $n > n_0(g)$. One can find short proofs of this upper bound in [22] and in Ruszinkó [36].

Using induction on t we extend Theorem 2.1 for all $t \geqslant 2$.

Theorem 3.1. *There exists an absolute constant $\alpha > 0$ such that*

$$c_t(n) < \alpha n^{(t-1)/2} \left(\frac{t+3}{t+2}\right)^n$$

holds for all $n, t \geqslant 2$.

The proof is postponed to Section 7. This might give a decent upper bound for small t but the true order of magnitude of $c_t(n)$ for large t is much smaller.

Theorem 3.2. *There exist positive constants β_1 and β_2 and a bound $n_0(t)$ depending only on t such that the following bounds hold for all $n > n_0(t)$, $t \geqslant 2$:*

$$\beta_1 \frac{1}{t^2} < \frac{\log c_t(n)}{n} < \beta_2 \frac{\log t}{t^2}. \tag{3.2}$$

Proof. The lower bound follows from (3.1) since $C_{t+1}(n) \leqslant c_t(n)$. It can be proved by a standard random choice. For the upper bound we observe that

$$c_t(n) \leqslant 1 + \left\lfloor \frac{t}{2} \right\rfloor + C_{\lfloor t/2 \rfloor}(n). \tag{3.3}$$

Indeed, if $\mathcal{F} \subset 2^{[n]}$, where $|\mathcal{F}|$ exceeds the right-hand side, then one can find $h+1$ distinct members $A_0, A_1, \ldots, A_h \in \mathcal{F}$, where $h = \lfloor t/2 \rfloor$, such that $A_0 \subset A_1 \cup \cdots \cup A_h$. Then, the size of the family $\mathcal{F}' := \mathcal{F} \setminus \{A_0, A_1, \ldots, A_h\}$ still exceeds $C_h(n)$, so there is another set of distinct members $B_0, \ldots, B_h \in \mathcal{F}'$ with $B_0 \subset B_1 \cup \cdots \cup B_h$. Taking another set $D \in \mathcal{F}'$ if t is odd, we have selected $t+2$ distinct members of $\mathcal{F}$ such that the union of t of them, namely $A_1, \ldots, A_h$ and $B_1, \ldots, B_h$ and possibly D, covers the other two, namely A_0 and B_0. Hence $\mathcal{F}$ cannot be 2-cancellative.

Finally, the upper bound (3.2) is implied by (3.3) and (3.1). □

4. Three-uniform cancellative families and sparse hypergraphs

The rest of our results concern r-uniform cancellative families. We are especially interested in the case when n is large with respect to r.

Frankl and the present author [20] determined asymptotically the maximum size of an r-uniform g-cover-free family by showing that there exists a positive constant $\gamma := \gamma(r,t)$ such that

$$C_g(n,r) = (\gamma + o(1)) n^{\lceil r/g \rceil}, \tag{4.1}$$

as r and g are fixed and n tends to infinity. A way to determine $\gamma(r,t)$ was also described. This and the r-uniform version of (3.3),

$$C_{t+1}(n,r) \leqslant c_t(n,r) \leqslant 1 + \left\lfloor \frac{t}{2} \right\rfloor + C_{\lfloor t/2 \rfloor}(n,r),$$

imply

$$(\gamma(r,t+1) - o(1))n^{\delta_1} \leqslant c_t(n,r) \leqslant (\gamma(r,\lfloor t/2 \rfloor) + o(1))n^{\delta_2}, \tag{4.2}$$

where the exponents are $\delta_1 := \lceil r/(t+1) \rceil$ and $\delta_2 := \lceil r/\lfloor t/2 \rfloor \rceil$. The next theorem shows that to obtain the true asymptotic for $c_t(n,r)$, like the one in (4.1) for $C_g(n,r)$, is probably a very difficult problem even in the case $r = 3$.

Brown, Erdős and Sós [11, 7, 6] introduced the function $f_r(n,v,e)$ to denote the maximum number of edges in an r-uniform hypergraph on n vertices which does not contain e edges spanned by v vertices. Such hypergraphs are called $\mathbb{G}(v,e)$-*sparse* (more precisely $\mathbb{G}_r(v,e)$-*sparse*). They showed that $f_r(n, e(r-k)+k, e) = \Theta(n^k)$ for every $2 \leqslant k < r$ and $e \geqslant 2$. The upper bound $(e-1)\binom{n}{k}$ is easy: no k-set can be contained in e hyperedges. If we forbid e edges spanned by one more vertex then the problem becomes much more difficult. Brown, Erdős and Sós conjectured that

$$f_r(n, e(r-k)+k+1, e) = o(n^k).$$

The $(6,3)$-theorem of Ruzsa and Szemerédi [37] deals with the case $(e,k,r) = (3,2,3)$, *i.e.*, when no six points contain three triples. They showed that there exists an $\alpha > 0$ such that

$$n^2 e^{-\alpha\sqrt{\log n}} = n^{2-o(1)} < f_3(n,6,3) = o(n^2). \tag{4.3}$$

Since a $\mathbb{G}(6,3)$-sparse system is $\mathbb{G}(7,4)$-sparse, we have

$$f_3(n,6,3) \leqslant f_3(n,7,4), \tag{4.4}$$

and Erdős *conjectured* that $f_3(n,7,4) = o(n^2)$.

Theorem 4.1.

$$f_3(n,7,4) - \frac{2}{5}n \leqslant c_2(n,3) \leqslant \frac{9}{2} f_3(n,7,4) + n. \tag{4.5}$$

The proof is presented in Section 8. The $(6,3)$ theorem was extended by Erdős, Frankl and Rödl [14] for arbitrary fixed $r \geqslant 3$,

$$n^{2-o(1)} < f_r(n, 3(r-2)+3, 3) = o(n^2), \tag{4.6}$$

and then by Alon and Shapira [4], $n^{k-o(1)} < f_r(n, 3(r-k)+k+1, 3) = o(n^k)$. Even the case $k = 2$, $f_r(n, e(r-2)+3, e) = o(n^2)$, is still open for general e. Nearly tight upper bounds were established by Sárközy and Selkow [38, 39]: $f_r(n, e(r-k)+k+\lfloor \log_2 e \rfloor, e) = o(n^k)$ for $r > k \geqslant 2$ and $e \geqslant 3$, and $f_r(n, 4(r-k)+k+1, 4) = o(n^k)$ for the case $e = 4$, $r > k \geqslant 3$.

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

5. An upper bound for uniform families

Theorem 5.1. *For every k and n we have*

$$c_2(n,2k) \leqslant \frac{\binom{n}{k}}{\frac{1}{2}\binom{2k}{k}}.$$

Proof of Theorem 5.1. Suppose that $\mathcal{F}$ is a $2k$-uniform, 2-cancellative family with the underlying set $[n]$. We may suppose that $|\mathcal{F}| > 3$, so $\mathcal{F}$ is 1-cancellative too.

Define a graph $G = (V,E)$ with vertex set $V := \binom{[n]}{k}$, *i.e.*, the family of k-subsets of $[n]$. A pair $A,B \in V$ forms an edge of G if $A \cup B \in \mathcal{F}$. Such a pair necessarily contains disjoint sets. Since every $F \in \mathcal{F}$ has $\frac{1}{2}\binom{2k}{k}$ partitions into k-sets, we have

$$|\mathcal{E}(G)| = \frac{1}{2}\binom{2k}{k}|\mathcal{F}|.$$

We claim that $|\mathcal{E}(G)| \leqslant |V| = \binom{n}{k}$.

Consider four adjacent edges in G on five (not necessarily distinct) vertices $V_1,\dots,V_5 \in V(G)$ (in fact these are k-sets of $[n]$) such that $\{V_i, V_{i+1}\} \in \mathcal{E}(G)$ $(1 \leqslant i \leqslant 4)$ and $V_i \neq V_{i+2}$. If these four edges determine four distinct sets $V_i \cup V_{i+1} \in \mathcal{F}$, then the identity

$$(V_1 \cup V_2) \cup (V_4 \cup V_5) \cup (V_2 \cup V_3) = (V_1 \cup V_2) \cup (V_4 \cup V_5) \cup (V_3 \cup V_4)$$

yields a contradiction, since $\mathcal{F}$ is 2-cancellative. By definition we have $(V_1 \cup V_2) \neq (V_2 \cup V_3) \neq (V_3 \cup V_4) \neq (V_4 \cup V_5)$. We also have $V_1 \cup V_2 \neq V_3 \cup V_4$ (and by symmetry $V_2 \cup V_3 \neq V_4 \cup V_5$). Indeed, $V_3 \subset (V_1 \cup V_2)$, $V_2 \cap V_3 = \emptyset$ leads to $V_1 = V_3$, which we have excluded. The last case to investigate is when $V_1 \cup V_2 = V_4 \cup V_5$, and the four edges determine exactly three sets. This leads to the contradiction

$$\begin{aligned}(V_1 \cup V_2) \cup (V_2 \cup V_3) &= (V_1 \cup V_2) \cup V_3 = (V_4 \cup V_5) \cup V_3\\ &= (V_4 \cup V_5) \cup (V_3 \cup V_4) = (V_1 \cup V_2) \cup (V_3 \cup V_4).\end{aligned}$$

We conclude that G does not have such a sequence of four edges. Therefore G contains no cycles, implying $|\mathcal{E}(G)| < |V|$. □

To estimate $c_2(n,2k+1)$, let us consider a $(2k+1)$-uniform family on $[n]$ and join the element $(n+1)$ to each hyperedge. If the original family is t-cancellative, then so is the extended family. We can apply Theorem 5.1 to get

$$c_2(n,2k+1) \leqslant \frac{\binom{n+1}{k+1}}{\frac{1}{2}\binom{2k+2}{k+1}}. \tag{5.1}$$

6. The non-uniform case, the proof of Theorem 2.1

Suppose that $\mathcal{F}$ is a 2-cancellative family of maximal size on the underlying set $[n]$. Split $\mathcal{F}$ according to the sizes of its edges: $\mathcal{F}_r := \{F \in \mathcal{F} : |F| = r\}$.

The sequence $\sqrt{2k-1}\binom{2k}{k}4^{-k}$ is monotone increasing for $k = 1, 2, 3\ldots$, so we obtain that $\binom{2k}{k}^{-1} \leqslant 2\sqrt{2k-1} \times 4^{-k}$ for all $k \geqslant 1$. Using this upper bound in Theorem 5.1, we obtain

$$c_2(n, 2k) \leqslant \frac{\binom{n}{k}}{\frac{1}{2}\binom{2k}{k}} \leqslant \binom{n}{k}4^{-k}4\sqrt{2k-1} < 4\sqrt{n} \times \binom{n}{k}4^{-k}.$$

The same inequality and (5.1) give

$$c_2(n, 2k+1) \leqslant \frac{\binom{n+1}{k+1}}{\frac{1}{2}\binom{2k+2}{k+1}} \leqslant \binom{n+1}{k+1}4^{-k-1}4\sqrt{2k+1} \leqslant 4\sqrt{n} \times \binom{n+1}{k+1}4^{-k-1}.$$

Finally,

$$\begin{aligned} c_2(n) = |\mathcal{F}| = \sum_r |\mathcal{F}_r| \leqslant \sum_r c_2(n,r) &= \Big(\sum_{k\geqslant 0} c_2(n,2k)\Big) + \Big(\sum_{k\geqslant 0} c_2(n,2k+1)\Big) \\ &< \Big(\sum_{k\geqslant 0} 4\sqrt{n} \times \binom{n}{k}4^{-k}\Big) + \Big(\sum_{k\geqslant 0} 4\sqrt{n} \times \binom{n+1}{k+1}4^{-k-1}\Big) \\ &= 4\sqrt{n}\bigg(\Big(1+\frac{1}{4}\Big)^n + \Big(1+\frac{1}{4}\Big)^{n+1}\bigg) = 9\sqrt{n}\Big(\frac{5}{4}\Big)^n. \end{aligned}$$

7. The case of t-cancellative codes, the proof of Theorem 3.1

We will define a monotone sequence $0 < \alpha_2 \leqslant \alpha_3 \leqslant \cdots \alpha_t \leqslant \cdots$, which is bounded above (by α), such that we have

$$c_t(n) < \alpha_t n^{(t-1)/2}\left(\frac{t+3}{t+2}\right)^n \tag{7.1}$$

for every $n, t \geqslant 2$. By Theorem 2.1 this holds for $t = 2$ with $\alpha_2 := 9$. Suppose that $t \geqslant 3$ and (7.1) holds for $t-1$. We use the upper bound

$$c_t(n,r) \leqslant c_{t-1}(n-r). \tag{7.2}$$

Indeed, if $\mathcal{F}$ is an r-uniform t-cancellative family on $[n]$, then for any $F_0 \in \mathcal{F}$ the family $\{F \setminus F_0 : F \in \mathcal{F}, F \neq F_0\}$ is $(t-1)$-cancellative. Use the inequality

$$\binom{r(t+2)}{r} > \frac{1}{3\sqrt{r}}\left(\frac{(t+2)^{t+2}}{(t+1)^{t+1}}\right)^r,$$

which holds for all integers $r \geqslant 1$, $t \geqslant 0$, and substitute $n = r(t+2)$ into (7.2). We obtain

$$\frac{c_t(r(t+2),r)}{\binom{r(t+2)}{r}} < \alpha_{t-1}n^{(t-2)/2}\left(\frac{t+2}{t+1}\right)^{r(t+1)} \times 3\sqrt{r}\frac{(t+1)^{r(t+1)}}{(t+2)^{r(t+2)}} = \frac{3\alpha_{t-1}}{\sqrt{t+2}}n^{(t-1)/2}(t+2)^{-r}.$$

For $n \geqslant m$ we have $c_t(n,r)\binom{n}{r}^{-1} \leqslant c_t(m,r)\binom{m}{r}^{-1}$. We obtain that the right-hand side is an upper bound for $c_t(n,r)/\binom{n}{r}$ for every $n \geqslant r(t+2)$. For any given n and t this gives

$$\sum_{r\leqslant n/(t+2)} c_t(n,r) \leqslant \sum \binom{n}{r}\frac{3\alpha_{t-1}}{\sqrt{t+2}}n^{(t-1)/2}(t+2)^{-r} \leqslant \frac{3\alpha_{t-1}}{\sqrt{t+2}}n^{(t-1)/2}\left(1+\frac{1}{t+2}\right)^n.$$

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

We estimate the case $n < r(t+2)$ using (7.2) again:

$$\sum_{n/(t+2)<r\leqslant n} c_t(n,r) \leqslant \sum_{r>n/(t+2)} c_{t-1}(n-r) < \sum_{n-r<(t+1)n/(t+2)} \alpha_{t-1} n^{(t-2)/2}\left(\frac{t+2}{t+1}\right)^{n-r}$$
$$< \alpha_{t-1} n^{(t-2)/2}(t+2)\left(\frac{t+2}{t+1}\right)^{n(t+1)/(t+2)}.$$

Here $(\frac{t+2}{t+1})^{(t+1)/(t+2)} < (t+3)/(t+2)$, so the sum of the above two displayed formulas gives

$$c_t(n) \leqslant \alpha_{t-1}\left(\frac{3}{\sqrt{t+2}} + \frac{t+2}{\sqrt{n}}\right) n^{(t-1)/2}\left(\frac{t+3}{t+2}\right)^n.$$

The rest is a little calculation (*e.g.*, we may suppose that $n > 2(t+2)^2$, otherwise our upper bound (7.1) for $c_t(n)$ exceeds the same upper bound for $c_{t-1}(n)$). □

8. Three-partite hypergraphs

In this section we prove Theorem 4.1 on 3-uniform 2-cancellative families.

Lemma 8.1 (Erdős and Kleitman [15]). *Let $\mathcal{F}$ be an r-uniform hypergraph. Then there exists an r-partite $\mathcal{F}^* \subset \mathcal{F}$ with $|\mathcal{F}^*| \geqslant \frac{r!}{r^r}|\mathcal{F}|$.*

Proof of Theorem 4.1. Suppose that $\mathcal{H}$ is a 3-uniform $\mathbb{G}(7,4)$-sparse family with vertex set $[n]$. We claim that there exists a subfamily $\mathcal{H}'' \subset \mathcal{H}$ such that

$$|\mathcal{H}''| \geqslant |\mathcal{H}| - \frac{2}{5}n \quad \text{and} \quad \mathcal{H}'' \text{ is linear.}$$

First, note that if the hyperedge $F \in \mathcal{H}$ has two other edges F_1, F_2 with $|F \cap F_i| = 2$, then these three edges form a separate connected component of $\mathcal{H}$ on 5 vertices. Let $\mathcal{H}'$ be the hypergraph obtained from $\mathcal{H}$ after deleting two edges from each such 5-vertex component. If $F_1 \in \mathcal{H}'$ and there exists an $F_2 \in \mathcal{H}'$ with $|F_1 \cap F_2| = 2$ then this F_2 is unique. Moreover, if $|F_1 \cap F_2| = 2$ and $|F_3 \cap F_4| = 2$ hold for four distinct sets, then $F_1 \cup F_2$ is disjoint to $F_3 \cup F_4$. Remove an edge from each such pair to obtain $\mathcal{H}''$, which is a linear hypergraph by definition, and we have left out at most $(2/5)n$ edges of $\mathcal{H}$.

Second, observe that a linear, 3-uniform, $\mathbb{G}(7,4)$-sparse family $\mathcal{H}''$ is 2-cancellative. Indeed, if $X := C \setminus (A \cup B) = D \setminus (A \cup B)$ for four distinct members $\{A, B, C, D\} \subset \mathcal{H}''$, then $X \subset C \cap D$, so $|X| \leqslant 1$, and $|A \cup B| \leqslant 6$, so they form a $\mathbb{G}(7,4)$ family, a contradiction. If we take $|\mathcal{H}|$ as large as possible, then we complete the proof of the first inequality of (4.5) as follows:

$$c_2(n,3) \geqslant |\mathcal{H}''| \geqslant |\mathcal{H}| - \frac{2}{5}n = f_3(n,7,4) - \frac{2}{5}n.$$

Next, let $\mathcal{F}$ be a 3-uniform, 2-cancellative family on n vertices. We claim that there exists a subfamily $\mathcal{F}' \subset \mathcal{F}$ such that

$$|\mathcal{F}'| \geqslant |\mathcal{F}| - n \quad \text{and} \quad \mathcal{F}' \text{ is linear.}$$

Indeed, leave out a hyperedge from $\mathcal{F}$ if it has a vertex of degree one. Repeat this until we get $\mathcal{F}' \subset \mathcal{F}$ for which every degree is either 0 or at least 2. We claim that $\mathcal{F}'$ is

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

linear (in the case of $|\mathcal{F}'| \geqslant 4$). Suppose not, that is, $|F_1 \cap F_2| = 2$, $F_1, F_2 \in \mathcal{F}'$, x_i is the unique element of $F_i \setminus (F_1 \cap F_2)$ when $i = 1, 2$. By our degree condition there exist $A_i \in \mathcal{F}'$, $x_i \in A_i$, $A_i \neq F_1$ and $A_i \neq F_2$. This leads to the contradiction $A_1 \cup A_2 \cup F_1 = A_1 \cup A_2 \cup F_2$. (The case $|\mathcal{F}'| = 3$ is left to the reader.)

Apply Lemma 8.1 to $\mathcal{F}'$ to obtain a 3-partite $\mathcal{F}^*$ of size $|\mathcal{F}^*| \geqslant (2/9)|\mathcal{F}'|$. We claim it is $\mathbb{G}(7,4)$-sparse, because every 3-partite, 2-cancellative, linear family $\mathcal{F}^*$ is $\mathbb{G}(7,4)$-sparse. Indeed, take any four distinct members $\{A, B, C, D\} \subset \mathcal{F}^*$. If $A \cap B = \emptyset$ then $C \setminus (A \cup B) \neq \emptyset$ (by linearity) and it is not equal to $D \setminus (A \cup B)$ (by 2-cancellativeness), so the union of the four of them has at least 8 vertices. Otherwise, pairwise they have a one-element intersection. If there is a degree-three vertex, say $A \cap B \cap C = \{x\}$, then linearity and 3-partiteness imply that D is not covered by $A \cup B \cup C$, so again their union has eight vertices (at least). If these four triples meet pairwise but their maximum degree is two, then $C \cup D$ covers $(A \cup B) \setminus (A \cap B)$, and they have a single common vertex outside $A \cup B$, yielding the contradiction $A \cup B \cup C = A \cup B \cup D$.

Finally, if we take $|\mathcal{F}|$ as large as possible, then we complete the proof of the second inequality of (4.5) as follows:

$$f_3(n, 7, 4) \geqslant |\mathcal{F}^*| \geqslant \frac{2}{9}|\mathcal{F}'| \geqslant \frac{2}{9}(|\mathcal{F}| - n) = \frac{2}{9}(c_2(n, 3) - n). \qquad \square$$

Define the hypergraphs $\mathbb{G}_6$ and $\mathbb{G}_7$ as follows on 6 and 7 vertices:

$$\mathcal{E}(\mathbb{G}_6) := \{123, 156, 426, 453\},$$
$$\mathcal{E}(\mathbb{G}_7) := \{123, 456, 726, 753\}.$$

Note that both are three-partite and the 3-partition of their vertices is unique.

Proposition 8.2. *Suppose that $\mathcal{F}$ is a three-partite, linear hypergraph. It is 2-cancellative if and only if it avoids $\mathbb{G}_6$ and $\mathbb{G}_7$. It is $\mathbb{G}(7,4)$-sparse if and only if it avoids $\mathbb{G}_6$ and $\mathbb{G}_7$.* $\square$

9. A construction by induced packings

According to the upper bounds in Theorem 5.1 we have

$$c_2(n, 2) \leqslant n, \quad c_2(n, 3) \leqslant \frac{1}{6}n(n+1), \quad c_2(n, 4) \leqslant \frac{1}{6}n(n-1).$$

Obviously, $c_2(n, 2) = n - 1$ for $n > 3$. The second inequality, although it is close to the true order of magnitude, is not sharp if Erdős's conjecture is true: see (4.3), (4.4) and (4.5). Any 4-uniform Steiner system $S(n, 4, 2)$ is 2-cancellative, yielding the lower bound $c_2(n, 4) \geqslant \frac{1}{12}n(n-1) - O(n)$ for all n.

Theorem 9.1. $c_2(n, 4) = \frac{1}{6}n^2 - o(n^2)$.

Theorem 9.2. $c_2(n, 2k) \geqslant \frac{n^k}{(2k)^k} - o(n^k)$.

The proof of Theorem 9.2 is postponed to Section 10.3. For the construction giving Theorem 9.1 we use induced packings of graphs.

A set of graphs $\mathcal{P} := \{G_1 = (V_1, \mathcal{E}_1),\ G_2 = (V_2, \mathcal{E}_2)\ldots\}$ is called a *packing* if they are edge-disjoint subgraphs of $G = (V, \mathcal{E})$ (by definition $V_i \subset V$ for each i). The packing $\mathcal{P}$ is called an *induced packing* if G restricted to V_i is exactly G_i (for all i). The induced packing $\mathcal{P}$ is called an *almost disjoint induced packing* into the graph G if $|V_i \cap V_j| \leqslant 2$ (for all $i \neq j$). It follows that whenever $F = V_i \cap V_j$, $|F| = 2$, then F is not an edge of G. In other words, any two induced graphs $G[V_i]$ and $G[V_j]$ are either vertex-disjoint, or share one vertex, or meet in a non-edge. For example, if V is an n-element set, n is even, $V = A_1 \cup A_2 \cup \cdots \cup A_{n/2}$ where each $|A_i| = 2$, and G is the complete graph on V minus the $n/2$ edges of the perfect matching $\{A_1, A_2, \ldots\}$, then $\mathcal{E}(G)$ can be decomposed into $n(n-2)/8$ almost disjoint, induced four-cycles, namely those induced by $A_i \cup A_j$.

Let H be a graph of e edges and let $i(n, H)$ denote the maximum number of almost disjoint induced copies of H that can be packed into any n-vertex graph. It was proved by Frankl and the present author that

$$i(n, H) = \frac{1}{e(H)}\binom{n}{2} - o(n^2).$$

In other words we have the following.

Lemma 9.3 ([20]). *For any fixed graph H with e edges, one can delete $o(n^2)$ edges of the graph K_n such that the rest of the edges, the graph $L_n = L_n(H)$, can be decomposed into $(1 - o(1))\binom{n}{2}/e$ almost disjoint induced copies of H.*

Proof of Theorem 9.1. The graph H_k, for $k \geqslant 3$, is defined as a complete graph K_k and $\binom{k}{3}$ vertices of degree three, each of those connected to a different triple of $V(K_k)$. We have that H_k has $k + \binom{k}{3}$ vertices, $\binom{k}{2} + 3\binom{k}{3}$ edges, and it contains $\binom{k}{3}$ *special* K_4s, those having a vertex of degree three in H_k. Take any almost disjoint packing of copies of H_k, $\mathcal{P} := \{H_k^1, H_k^2, \ldots\}$, and define a 4-uniform family $\mathcal{F}(\mathcal{P})$ as the vertex sets of the special K_4s. Obviously $|\mathcal{F}(\mathcal{P})| = \binom{k}{3}|\mathcal{P}|$.

We claim that $\mathcal{F}(\mathcal{P})$ is a 2-cancellative family.

Suppose, on the contrary, that there are four distinct members $A, B, C, D \in \mathcal{F}$ with $A \cup B \cup C = A \cup B \cup D$. Note that $|F \cap F'| \leqslant 2$ for $F, F' \in \mathcal{F}$. Furthermore, in the case of equality F and F' are generated by the same H_k^i.

Consider first the case when C and D are generated by the same copy of H_k^i, that is, $C = \{c, x_1, x_2, x_3\}$ and $D = \{d, y_1, y_2, y_3\}$, where c and d are distinct degree-three vertices of H_k^i. The element c is covered by $A \cup B$, say $c \in A$. By definition, the pairs cx_1, cx_2, cx_3 are only covered by C among the members of $\mathcal{F}$, and since there is no edge (of any H_k^j) from c to $D \setminus C$, those pairs are not covered by any member of $F \in \mathcal{F}$. Hence $A \cap (C \cup D) = \{c\}$. Similarly, $B \cap (C \cup D) = \{d\}$. Hence $(A \cup B) \cap (C \cup D) = \{c, d\}$. Since the symmetric difference $C \Delta D$ is contained in $A \cup B$, we obtain that it is $\{c, d\}$. This leads to the contradiction $|C \cap D| = 3$.

Now consider the other case, that $i \neq j$ and $C = \{c, x_1, x_2, x_3\}$ is generated by H_k^i, where c is a degree-three vertex of H_k^i, and $D = \{d, y_1, y_2, y_3\}$ is generated by H_k^j, where d is a degree-three vertex of H_k^j. Since they come from different copies of H_k, we have $|C \cap D| \leqslant 1$. This implies that either A or B meets C in two vertices, say $|A \cap C| = 2$. Then A is generated by H_k^i as C is, say $A = \{a, x_2, x_3, x_4\}$. It follows that $|A \cap D| \leqslant 1$, so

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

$|B \cap D| = 2$, and D is generated by H_k^j as well. The pair $\{x_1, c\} = C \setminus A$ is covered by $B \cup D$, thus it is covered by $V(H_k^j)$. This leads to the contradiction that $V(H_k^i) \cap V(H_k^j)$ contains an edge of H_k^i. This completes the proof that $\mathcal{F}$ is 2-cancellative.

For given n, taking a large induced packing of H_ks Lemma 9.3 implies that

$$c_2(n,4) \geqslant i(n, H_k)\binom{k}{3} \geqslant (1 - o(1))\frac{\binom{k}{3}}{\binom{k}{2} + 3\binom{k}{3}}\binom{n}{2} \tag{9.1}$$

when k is fixed and $n \to \infty$. Let $\pi_4 := \liminf_{n\to\infty}\{c_2(n,4)/\binom{n}{2}\}$. The lower bound (9.1) gives that

$$\pi_4 \geqslant \frac{1}{\binom{k}{2}/\binom{k}{3} + 3}.$$

Since this holds for each k we obtain $\pi_4 \geqslant 1/3$. Finally, $\pi_4 \leqslant 1/3$ was proved in Theorem 5.1, completing the proof of $\pi_4 = 1/3$. □

10. The lower bound for the 2k-uniform case

10.1. Non-vanishing polynomials

For a set of variables $X = \{x_1, \dots, x_s\}$ and $0 \leqslant i \leqslant s$, the symmetric polynomial $\sigma_i(X)$ is defined as $\sum_{I \subset X, |I|=i} \prod_{\alpha \in I} x_\alpha$, $\sigma_0(X) = 1$. For convenience, $\sigma_i(X)$ is defined to be 0 for $|X| < i$ (and for $i < 0$). Suppose that $X_1, X_2, \dots, X_\ell$ are disjoint sets of variables, $|X_j| = t_j$, $0 < t_j < k$, $\sum_j (k - t_j) = k$. The entries of a row of the $k \times k$ matrix $M(X_1, \dots, X_\ell)$ consists of a block with the symmetric polynomials $\{\sigma_0(X_j), \sigma_1(X_j), \dots, \sigma_{t_j}(X_j)\}$ and zeros otherwise. The rows are distinct, so these blocks are shifted in all possible $k - t_j$ ways, that is,

$$M(X_1,\dots,X_\ell) :=$$

$$\begin{pmatrix}
1 & \sigma_1(X_1) & \sigma_2(X_1) & \cdots & \cdots & \sigma_{t_1}(X_1) & 0 & \cdots & 0 \\
0 & 1 & \sigma_1(X_1) & \sigma_2(X_1) & \cdots & \cdots & \sigma_{t_1}(X_1) & 0 & \vdots \\
\vdots & \vdots & \ddots & \ddots & & & & \ddots & 0 \\
0 & 0 & 0 & 1 & \sigma_1(X_1) & \sigma_2(X_1) & \cdots & \cdots & \sigma_{t_1}(X_1) \\
1 & \sigma_1(X_2) & \sigma_2(X_2) & \cdots & \cdots & \sigma_{t_2}(X_2) & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & & & & \ddots & 0 \\
 & & & & & & & & 0 \\
0 & 0 & 0 & 1 & \sigma_1(X_2) & \sigma_2(X_2) & \cdots & \cdots & \sigma_{t_2}(X_2) \\
\vdots & \vdots & & & & & & & \vdots \\
1 & \sigma_1(X_\ell) & \sigma_2(X_\ell) & \cdots & & \cdots & \sigma_{t_\ell}(X_\ell) & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & & & & \ddots & 0 \\
0 & \cdots & 1 & \sigma_1(X_\ell) & \sigma_2(X_\ell) & & & \cdots & \sigma_{t_\ell}(X_\ell)
\end{pmatrix}.$$

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

Fact 10.1. *The polynomial* $\det M(X_1,\ldots,X_\ell)$ *of* $\sum|X_j|$ *variables is non-vanishing.*

Proof. Over any field we can substitute only ones and zeros such that the matrix M becomes a lower triangular matrix, having only ones in the main diagonal and zeros above. Namely, let $x=0$ for each $x\in X_1$. In the second block of M, in rows $(k-t_1)+1$ to $(k-t_1)+(k-t_2)$ only $\sigma_i(X_2)$ stands in the main diagonal, where $i=k-t_1$. Define $k-t_1$ variables of X_2 to be 1, the rest 0.

In general, in the jth block, in rows $(k-t_1)+\cdots+(k-t_{j-1})+1$ to $\sum_{1\leqslant s\leqslant j}(k-t_s)$ we define the variables of X_j in such a way that $(k-t_1)+\cdots+(k-t_{j-1})$ of them are ones and the rest are zeros. This can be done, since $(k-t_1)+\cdots+(k-t_{j-1})\leqslant t_j$. □

One can define the matrix in a more general setting when the blocks consist of rows of the form $(\sigma_{m+1}(X_j),\sigma_{m+2}(X_j),\ldots,\sigma_{m+k}(X_j))$. We can obtain, for example, that the determinant of the $k\times k$ matrix M with $M_{i,j}:=\sigma_{m+i+j-2}(X)$ is non-vanishing if $m\geqslant 0$, $|X|\geqslant m+k-1$.

Let $q>1$ be a power of a prime, $\mathbf{F}:=\mathbf{F}_q$, the finite field of size q. For a polynomial $p(x_1,\ldots,x_s)$ over this field the zero set $Z(p)$ is defined by

$$Z(p)=\{(x_1,\ldots,x_s)\in\mathbf{F}_q^s : p(x_1,\ldots,x_s)=0\}.$$

The following fact is well known [33] and easy to prove by induction on $s+h$.

Fact 10.2. *If the degree of* $p(x_1,\ldots,x_s)$ *is* $h>0$*, then*

$$|Z(p(x_1,\ldots,x_s))|\leqslant hq^{s-1}.$$

10.2. A lemma on independent polynomials

Let k be a positive integer, let $\mathcal{P}:=\mathcal{P}_{<k}[\mathbf{F},x]$ be the ring of polynomials of degree at most $k-1$:

$$\mathcal{P}_{<k}:=\{a_0+a_1x+\cdots+a_{k-1}x^{k-1} : a_i\in\mathbf{F}\}.$$

The number of such polynomials is q^k and they form a linear space of dimension k over $\mathbf{F}$. A set of polynomials $p_1(x),\ldots,p_\ell(x)\in\mathcal{P}$ is called $(k_1,\ldots,k_\ell)$-*independent*, where $k_1,\ldots,k_\ell$ are positive integers if

$$f_1(x)p_1(x)+\cdots+f_\ell(x)p_\ell(x)\equiv 0,$$

and $\deg(f_i)<k_i$ for all i imply that each $f_i(x)$ is the 0 polynomial. Equivalently, all the $q^{\sum k_i}$ polynomials of the form $\sum f_ip_i$ (with $\deg(f_i)<k_i$) are distinct. The case when every $k_i=1$ corresponds to the usual linear independence. To stay in the space $\mathcal{P}_{<k}$ we also suppose that $\deg(p_i)+k_i<k$. Then necessarily $\sum_i k_i\leqslant k$.

For $Z\subset\mathbf{F}$ there is a unique polynomial with leading coefficient 1 and roots Z, namely

$$p_Z(x):=\prod_{z\in Z}(x-z).$$

Suppose that $\ell\geqslant 2$, $k_1,\ldots,k_\ell$ are positive integers with $k_1+\cdots+k_\ell=k$, and let $x_1,\ldots,$ $x_{(\ell-1)k}$ be a sequence of elements of $\mathbf{F}_q$. Define the (multi)sets X_i of size $k-k_i$ as intervals

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

of this sequence, $X_1 := \{x_s : 1 \leqslant s \leqslant k - k_1\}$, in general

$$X_j := \Big\{x_s : \sum_{i<j}(k-k_i) < s \leqslant \sum_{i\leqslant j}(k-k_i)\Big\}.$$

Lemma 10.3. *The polynomials $p_{X_1}(x), \ldots, p_{X_\ell}(x)$ are $(k_1, \ldots, k_\ell)$-independent for all but at most*

$$\binom{\ell k}{2} q^{(\ell-1)k-1}$$

sequences.

Proof. There are at most $\binom{(\ell-1)k}{2} q^{(\ell-1)k-1}$ sequences with repeated entries. Let $\mathcal{B}$ be the set of sequences $x_1, \ldots, x_{(\ell-1)k}$ with distinct elements such that $p_{X_1}(x), \ldots, p_{X_\ell}(x)$ are $(k_1, \ldots, k_\ell)$-dependent. Next we give an upper bound for $|\mathcal{B}|$.

The polynomials $p_1, \ldots, p_\ell$ are $(k_1, \ldots, k_\ell)$-dependent if and only if the set of polynomials $\{x^i p_j(x) : 0 \leqslant j < k_j\}$ are linearly dependent. These k polynomials are linearly dependent if and only if their coefficient matrix is singular. The coefficients of p_{X_j} are the values of the symmetric polynomials $\sigma_s(-X_j)$, so the coefficient matrix is exactly $M(-X_1, \ldots, -X_\ell)$ defined in the previous subsection. According to Fact 10.1 its determinant is a non-zero polynomial $f(x_1, \ldots, x_{(\ell-1)k})$. We have $\mathcal{B} \subseteq Z(f)$.

Each entry of the $k \times k$ matrix M is a symmetric polynomial of degree at most $k-1$, and thus the degree of the polynomial f is at most $k(k-1)$. Then Fact 10.2 gives an upper bound $k(k-1)q^{(\ell-1)k-1}$ for $|Z(f)|$. So the number of sequences $x_1, \ldots, x_{(\ell-1)k}$ such that $p_{X_1}(x), \ldots, p_{X_\ell}(x)$ are $(k_1, \ldots, k_\ell)$-dependent is at most $(k(k-1) + \binom{(\ell-1)k}{2})q^{(\ell-1)k-1}$. □

Corollary 10.4. *For every k there exists a $q_0(k)$ such that, if $q > q_0(k)$, then there exists a $2k$-element set $S \subset \mathbf{F}_q$ such that the polynomials*

$$p_X(x), p_Y(x), p_W(x) \text{ are } (k-|X|, k-|Y|, k-|W|)\text{-independent}$$

for every partition of $S = X \cup Y \cup W$, $|X| + |Y| + |W| = 2k$, $1 \leqslant |X|, |Y|, |W| < k$.

In fact, applying the previous lemma with $\ell = 3$, we can see that almost all $2k$-sets, all but at most $O(q^{2k-1})$ of them, have this total independence property. □

10.3. The algebraic construction yielding Theorem 9.2

Let q be the largest prime power not exceeding $n/(2k)$. Since there are no large gaps among primes we have $q > n/(2k) - O(n^{5/8})$. We also suppose that $q > q_0(k)$, as used in Corollary 10.4. We are going to define a 2-cancellative, $2k$-uniform family $\mathcal{F}$ of size q^k.

Take a set $S \subset \mathbf{F}_q$ of size $2k$ satisfying the conclusion of Corollary 10.4. Our hypergraph $\mathcal{F} := \mathcal{F}(q, S)$ consists of the graphs of the polynomials $\mathcal{P}_{<k}$ restricted to S. $V(\mathcal{F}) := S \times \mathbf{F} = \{(s, y) : s \in S, y \in \mathbf{F}\}$, every $p \in \mathcal{P}$ defines a set $F(p) := \{(s, p(s)) : s \in S\}$ and let $\mathcal{F} := \{F(p) : p \in \mathcal{P}\}$.

https://doi.org/10.1017/S0963548311000563 Published online by Cambridge University Press

To show that $\mathcal{F}$ is 2-cancellative suppose, on the contrary, that A, B, C and D are four distinct members of $\mathcal{F}$ with $A \cup B \cup C = A \cup B \cup D$. There are four distinct polynomials $a(x), b(x), c(x)$ and $d(x) \in \mathcal{P}$ generating these sets, $A = F(a)$, $B = F(b)$, *etc.*

Let $W \subset S$ be the set of coordinates where C and D meet, $W := \{s \in S : c(s) = d(s)\}$. Let $X := \{x \in S \setminus W : c(x) = a(x)\}$, and let $Y := S \setminus (X \cup W)$. For $x \in X$ $(x, d(x))$ is not covered by C or A, so it must belong to B, $b(x) = d(x)$. For $y \in Y$ we have $c(y) \neq d(y)$, $c(y) \neq a(y)$, so $(y, c(y))$ must be in B, $c(y) = b(y)$. Considering the same $y \in Y$, the element $(y, d(y))$ is not covered by C or B so it must belong to A, $a(y) = d(y)$. Let us summarize: there exists a partition of $S = W \cup X \cup Y$ such that

$$c(w) = d(w) \quad \text{for } w \in W, \tag{10.1}$$

$$c(x) \neq d(x) \quad \text{for } x \in X, \quad \text{but } c(x) = a(x) \text{ and } d(x) = b(x), \tag{10.2}$$

$$c(y) \neq d(y) \quad \text{for } y \in Y, \quad \text{but } d(y) = a(y) \text{ and } c(y) = b(y). \tag{10.3}$$

Since c and d are distinct polynomials of degree at most $k-1$, we have $|W| < k$. Similarly $|X|, |Y| < k$. These also imply that $|X|, |Y|, |W| \geqslant 2$ (and thus $k \geqslant 3$).

By (10.1), $c - d$ is divisible by p_W, and there exists a polynomial $c_1(x) \in \mathcal{P}$ such that

$$c = d + c_1 p_W \text{ where } c_1 \in \mathcal{P}, \text{ and } \deg(c_1) < k - |W|.$$

The first halves of (10.2) and (10.3) similarly imply that

$$a = c + a_1 p_X \text{ where } a_1 \in \mathcal{P}, \text{ and } \deg(a_1) < k - |X|,$$

$$d = a + a_2 p_Y \text{ where } a_2 \in \mathcal{P}, \text{ and } \deg(a_2) < k - |Y|.$$

Adding these three equations we obtain

$$0 = c_1 p_W + a_1 p_X + a_2 p_Y.$$

Then the independence of p_X, p_Y and p_W implies $c_1 = a_1 = a_2 = 0$, a contradiction. □

One might think that if we use the second halves of (10.2) and (10.3) then we have more constraints, and maybe we do not really need independence and Corollary 10.4. In fact, independence is essential. The second halves only imply that $b = d - a_1 p_X = c + a_2 p_Y$, so $\mathcal{F}$ can have many non-2-cancellative fourtuples if S is not chosen properly.

11. A remark on 1-cancellative uniform families

An r-partite hypergraph is cancellative if it contains no three distinct edges with $A \cup B = A \cup C$. Considering the complete r-partite hypergraph on n vertices with almost equal parts, we get

$$c(n, r) \geqslant \left\lfloor \frac{n}{r} \right\rfloor \times \left\lfloor \frac{n+1}{r} \right\rfloor \times \cdots \times \left\lfloor \frac{n+r-1}{r} \right\rfloor =: p(n, r). \tag{11.1}$$

The right-hand side is exactly n^r/r^r when r divides n. An old result of Mantel on the maximum size of triangle-free graphs gives $c(n, 2) = p(n, 2) = \lfloor n^2/4 \rfloor$. Katona [26] conjectured and Bollobás [5] proved that $c(n, 3) = p(n, 3)$. Bollobás also conjectured that equality holds in (11.1) for all $n \geqslant r \geqslant 4$ as well. This was established for $2r \geqslant n \geqslant r$ in [19].

Sidorenko [41] proved Bollobás's conjecture for $r = 4$. (There is a recent refinement of this by Pikhurko [35].) However, Shearer [40] gave a counterexample. His result implies that there exist an $\varepsilon > 0$ and $n_0(r)$ such that $c(n,3) > (1+\varepsilon)^r(n/r)^r$ for $n > n_0(r)$, $r \geqslant 11$. The cases $5 \leqslant r \leqslant 10$ are still undecided.

It was observed in [19] that $c(n,r) = 2^{n-r}$ for $2r \geqslant n \geqslant r$. Moreover, if $\mathcal{F}$ is a cancellative family of r-sets from an n-set and $n \geqslant 2r$, then

$$|\mathcal{F}| \leqslant \frac{2^r}{\binom{2r}{r}}\binom{n}{r}.$$

Here we show an almost matching lower bound.

Theorem 11.1. *For every $n \geqslant r \geqslant 2$,*

$$c(n,r) > \frac{\gamma_0}{2^r}\binom{n}{r},$$

where $\gamma_0 := \prod_{k\geqslant 1}\frac{2^k-1}{2^k} = 0.2887\ldots$

This result immediately follows from a construction of Tolhuizen [45], although he was not interested in r-uniform hypergraphs and wrote that 'the rate of a cancellative code is $\frac{\log 3}{\log 2} - 1 = 0.5849\ldots$.' His publication is not even reviewed in MathSciNet, so we briefly describe his work.

Proof (Tolhuizen [45]). If M is a random $m \times m$ matrix with entries from the two-element field $\mathbf{F}_2 = \{0,1\}$, then

$$\text{Prob}(M \text{ is non-singular}) = \frac{2^m-1}{2^m} \times \frac{2^m-2}{2^m} \times \cdots \times \frac{2^m-2^{m-2}}{2^m} \times \frac{2^m-2^{m-1}}{2^m} > \gamma_0.$$

Considering $(n-r)\times n$ random matrices we obtain an $(n-r)\times n$ matrix A (over $\mathbf{F}_2$) containing at least $\gamma_0\binom{n}{n-r}$ non-singular $(n-r)\times(n-r)$ submatrices. Let $\mathcal{F}$ be the set of those r-sets $F \subset [n]$ where the columns of A labelled by the elements of $[n]\setminus F$ have full rank. We have $|\mathcal{F}| > \gamma_0\binom{n}{r}$.

Let $\mathcal{S}$ be the $(n-r)$-dimensional subspace generated by the rows of A in $\mathbf{F}_2^n$ and let $\mathcal{R}$ be a subspace of dimension r such that $\mathcal{S}+\mathcal{R}$ is the whole space. Decompose the n-dimensional space into 2^r disjoint affine subspaces:

$$\mathbf{F}_2^n = \bigcup_{\mathbf{v}\in\mathcal{R}}(\mathcal{S}+\mathbf{v}).$$

For any set $F \subset [n]$, let $\widehat{F}$ be a 0–1 vector with support F. For each $\mathbf{v}\in\mathcal{R}$ let

$$\mathcal{F}(\mathbf{v}) := \{F : F \in \mathcal{F}, \widehat{F} \in (\mathcal{S}+\mathbf{v})\}.$$

We have partitioned $\mathcal{F}$ into 2^r pairwise disjoint r-uniform families. Given any $F \in \mathcal{F}$, the vectors of $(\mathcal{S}+\mathbf{v})$ truncated to $([n]\setminus F)$ are all distinct. Hence each $\mathcal{F}(\mathbf{v})$ is a cancellative family. □

There are $\Theta(r^2)$ non-isomorphic hypergraphs consisting of three edges $\{A, B, C\}$ with $A \cup B = A \cup C$. The *Turán number* of the class of r-uniform hypergraphs $\mathbb{H} := \{\mathcal{H}_1, \mathcal{H}_2, \dots\}$ is denoted by $\mathrm{ex}(n, \mathbb{H})$. It is the size of the largest r-graph on n vertices avoiding every $\mathcal{H} \in \mathbb{H}$ as a subgraph. The sequence $\mathrm{ex}(n, \mathbb{H})\binom{n}{r}^{-1}$ is monotone decreasing; its limit is denoted by $\pi(\mathbb{H})$. When we consider the determination of $c(n,r)$ as a Turán-type problem, then there is a score of forbidden hypergraphs. Take only one of them, namely $\mathbb{G}_r^3$ defined by three sets on $2r-1$ elements $[r] := \{1, 2, \dots, r\}$, $[r-1] \cup \{r+1\}$ and $\{r, r+1, \dots, 2r-1\}$. It was proved in [21] that

$$\left(\binom{r}{2} e^{1+1/(r-1)}\right)^{-1} \leqslant \pi(\mathbb{G}_r^3) \leqslant \left(e\binom{r-1}{2}\right)^{-1}.$$

Concerning another case, for an even r when $\mathbb{T}_r$ is a blown-up triangle, its three edges are $X \cup Y$, $Y \cup Z$, and $Z \cup X$ where $|X| = |Y| = |Z| = r/2$. Frankl [18] and Sidorenko [42, 43] showed independently that $\pi(\mathbb{T}_r) = 1/2$. For more on this see [29].

12. Conclusion, problems

One of our main results is to give a better upper bound for the size of 2-cancellative codes. We *conjecture* that the upper bounds of Theorems 2.1 and 3.2 are much closer to the truth than the simple probabilistic lower bounds we have. This is probably also true for the uniform case (see (5.1)).

Conjecture 12.1. *$n^{k+1-o(1)} < c_2(n, 2k+1) = o(n^{k+1})$ as $n \to \infty$ and k is fixed.*

Call a code $\mathcal{F}$ *t^*-cancellative* if

$$A_1 \cup \cdots \cup A_t \cup B = A_1 \cup \cdots \cup A_t \cup C \Longrightarrow B = C \text{ or } \{B, C\} \subset \{A_i, \dots, A_t\}$$

for every $t+2$ member *sequence* from $\mathcal{F}$, and let $c_t^*(n)$ be the maximum size of such a code $\mathcal{F} \subset 2^{[n]}$. Obviously $C_t(n) \leqslant c_t^*(n) \leqslant C_{t+1}(n) \leqslant c_t(n)$. One wonders if equality holds in some of these, and what other relations these functions can have.

Using the Erdős, Frankl and Rödl [14] estimate (see (4.6)), we have

$$n^2 e^{-\alpha_r \sqrt{\log n}} \leqslant f_r(n, 3(r-2)+3, 3) \leqslant c_{r-1}(n, r).$$

The general upper bound (4.2) for $c_t(n,r)$ here only gives $O(n^3)$, but in this case, leaving out those r-sets having an own pair, one can easily prove

$$c_{r-1}(n, r) \leqslant \binom{n}{2}.$$

For more of these types of problems, see, *e.g.*, [23].

In Section 10.3 the $2k$-partite hypergraph $\mathcal{F}$ (with partite sets $V_1, \dots, V_{2k}$) has an interesting property. For every three members A, B, C there exists a class V_i such that $A \cap V_i$, $B \cap V_i$ and $C \cap V_i$ are distinct. It is natural to ask what other small substructures can be avoided this way.

The proof of Theorem 3.1 concerning $c_t(n)$ presented in Section 7 actually gives a slightly better upper bound. A little more calculation yields an explicit bound κ_t for $t \geqslant 3$ such that

$$\limsup_n (c_t(n))^{1/n} \leqslant \kappa_t < \frac{t+3}{t+2}.$$

Many problems remain open.

Acknowledgements

The author is indebted to the referee and to P. Balister and O. Riordan for helpful comments.

References

[1] Alon, N. and Asodi, V. (2006) Tracing a single user. *European J. Combin.* **27** 1227–1234.
[2] Alon, N., Fachini, E. and Körner, J. (2000) Locally thin set families. *Combin. Probab. Comput.* **9** 481–488.
[3] Alon, N., Körner, J. and Monti, A. (2000) String quartets in binary. *Combin. Probab. Comput.* **9** 381–390.
[4] Alon, N. and Shapira, A. (2006) On an extremal hypergraph problem of Brown, Erdős and Sós. *Combinatorica* **26** 627–645.
[5] Bollobás, B. (1974) Three-graphs without two triples whose symmetric difference is contained in a third. *Discrete Math.* **8** 21–24.
[6] Brown, W. G., Erdős, P. and Sós, V. T. (1973) Some extremal problems on r-graphs. In *New Directions in the Theory of Graphs: Proc. Third Ann Arbor Conference, 1971*, Academic Press, 53–63.
[7] Brown, W. G., Erdős, P. and Sós, V. T. (1973) On the existence of triangulated spheres in 3-graphs, and related problems. *Period. Math. Hungar.* **3** 221–228.
[8] Coppersmith, D. and Shearer, J. B. (1998) New bounds for union-free families of sets. *Electron. J. Combin.* **5** #39.
[9] De Bonis, A. and Vaccaro, U. (2006) Optimal algorithms for two group testing problems and new bounds on generalized superimposed codes. *IEEE Trans. Inform. Theory* **52** 4673–4680.
[10] D'yachkov, A. G. and Rykov, V. V. (1982) Bounds on the length of disjunctive codes. *Problemy Peredaci Informacii* **18** 7–13.
[11] Erdős, P. (1964) Extremal problems in graph theory. In *Theory of Graphs and its Applications* (M. Fiedler, ed.), Academic Press, pp. 29–36.
[12] Erdős, P., Frankl, P. and Füredi, Z. (1982) Families of finite sets in which no set is covered by the union of two others. *J. Combin. Theory Ser. A* **33** 158–166.
[13] Erdős, P., Frankl, P. and Füredi, Z. (1985) Families of finite sets in which no set is covered by the union of r others. *Israel J. Math.* **51** 79–89.
[14] Erdős, P., Frankl, P. and Rödl, V. (1986) The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.* **2** 113–121.
[15] Erdős, P. and Kleitman, D. J. (1968) On coloring graphs to maximize the proportion of multicolored k-edges. *J. Combin. Theory* **5** 164–169.
[16] Erdős, P. and Szemerédi, E. (1978) Combinatorial properties of systems of sets. *J. Combin. Theory Ser. A* **24** 308–313.
[17] Fachini, E., and Körner, J. and Monti, A. (2001) A better bound for locally thin set families. *J. Combin. Theory Ser. A* **95** 209–218.

[18] Frankl, P. (1990) Asymptotic solution of a Turán-type problem. *Graphs Combin.* **6** 223–227.
[19] Frankl, P. and Füredi, Z. (1984) Union-free hypergraphs and probability theory. *European J. Combin.* **5** (1984), 127–131. Erratum, *ibid.* **5** 395.
[20] Frankl, P. and Füredi, Z. (1987) Colored packing of sets. In *Combinatorial Design Theory*, *Ann. Discrete Math.* **34** 165–178.
[21] Frankl, P. and Füredi, Z. (1989) Extremal problems whose solutions are the blow-ups of the small Witt-designs. *J. Combin. Theory Ser. A* **52** 129–147.
[22] Füredi, Z. (1996) On r-cover-free families. *J. Combin. Theory Ser. A* **73** 172–173.
[23] Füredi, Z. and Ruszinkó, M. (2011) Uniform hypergraphs containing no grids. Manuscript `arXiv:1103.1691 [math.CO]`
[24] Hwang, F. K. and Sós, V. T. (1987) Non adaptive hypergeometric group testing. *Studia Sci. Math. Hungar.* **22** 257–263.
[25] Johnson, S. M. (1971) On the upper bounds for unrestricted binary error-correcting codes. *IEEE Trans. Inform. Theory* **17** 466–478.
[26] Katona, G. O. H. (1974) Extremal problems for hypergraphs. In *Combinatorics: Proc. NATO Advanced Study Inst., Breukelen 1974*, Part 2, *Graph Theory: Foundations, Partitions and Combinatorial Geometry*, Vol. 56 of *Math. Centre Tracts*, Math. Centrum, pp. 13–42.
[27] Kautz, W. H. and Singleton, R. C. (1964) Nonrandom binary superimposed codes. *IEEE Trans. Inform. Theory* **10** 363–377.
[28] Keevash, P. and Mubayi, D. (2004) Stability theorems for cancellative hypergraphs. *J. Combin. Theory Ser. B* **92** 163–175.
[29] Keevash, P. and Sudakov, B. (2005) On a hypergraph Turán problem of Frankl. *Combinatorica* **25** 673–706.
[30] Körner, J. (1995) On the extremal combinatorics of the Hamming space. *J. Combin. Theory Ser. A* **71** 112–126.
[31] Körner, J. and Monti, A. (2002) Delta-systems and qualitative (in)dependence. *J. Combin. Theory Ser. A* **99** 75–84.
[32] Körner, J. and Sinaimeri, B. (2007) On cancellative set families. *Combin. Probab. Comput.* **16** 767–773.
[33] Lidl, R. and Niederreiter, H. (1994) *Introduction to Finite Fields and their Applications*, Cambridge University Press.
[34] Nguyen Quang A, and Zeisel, T. (1988) Bounds on constant weight binary superimposed codes. *Probl. Control Inform. Theory* **17** 223–230.
[35] Pikhurko, O. (2008) An exact Turán result for the generalized triangle. *Combinatorica* **28** 187–208.
[36] Ruszinkó, M. (1994) On the upper bound of the size of the r-cover-free families. *J. Combin. Theory Ser. A* **66** 302–310.
[37] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In *Combinatorics, Vol. II: Proc. Fifth Hungarian Colloq., Keszthely 1976*, Vol. 18 of *Colloq. Math. Soc. János Bolyai*, North-Holland, pp. 939–945.
[38] Sárközy, G. N. and Selkow, S. (2005) An extension of the Ruzsa–Szemerédi theorem. *Combinatorica* **25** 77–84.
[39] Sárközy, G. N. and Selkow, S. (2005) On a Turán-type hypergraph problem of Brown, Erdős and T. Sós. *Discrete Math.* **297** 190–195.
[40] Shearer, J. B. (1996) A new construction for cancellative families of sets. *Electron. J. Combin.* **3** #15.
[41] Sidorenko, A. F. (1987) The maximal number of edges in a homogeneous hypergraph containing no prohibited subgraphs. *Math. Notes* **41** 247–259. Translated from *Mat. Zametki.*
[42] Sidorenko, A. F. (1989) Extremal combinatorial problems in spaces with continuous measure (in Russian). *Issledovanie Operatchiy i ASU* **34** 34–40.

[43] Sidorenko, A. (1992) Asymptotic solution of the Turán problem for some hypergraphs. *Graphs Combin.* **8** 199–201.

[44] Sós, V. T. (1991) An additive problem in different structures. *Combinatorics, Algorithms, and Applications: Proc. Second International Conference in Graph Theory, San Francisco 1989*, SIAM, pp. 486–510.

[45] Tolhuizen, L. (2000) New rate pairs in the zero-error capacity region of the binary multiplying channel without feedback. *IEEE Trans. Inform. Theory* **46** 1043–1046.