On the maximum number of integer colourings with forbidden monochromatic sums

Hong Liu[*]

Mathematics Institute
University of Warwick
Coventry, U.K.

h.liu.9@warwick.ac.uk

Maryam Sharifzadeh[†]

Department of Mathematics and Mathematical Statistics
Umeå University
Umeå, Sweden

maryam.sharifzadeh@umu.se

Katherine Staden

Mathematical Institute
University of Oxford
Oxford, U.K.

staden@maths.ox.ac.uk

Abstract

Let $f(n, r)$ denote the maximum number of colourings of $A \subseteq \{1, \ldots, n\}$ with r colours such that each colour class is sum-free. Here, a sum is a subset $\{x, y, z\}$ such that $x + y = z$. We show that $f(n, 2) = 2^{\lceil n/2 \rceil}$, and describe the extremal subsets. Further, using linear optimisation, we asymptotically determine the logarithm of $f(n, r)$ for $r \leqslant 5$. Similar results were obtained by Hán and Jiménez in the setting of finite abelian groups.

Mathematics Subject Classifications: 11B75,11B13

1 Introduction and results

A recent trend in combinatorial number theory has been to consider versions of classical problems from extremal graph theory in the sum-free setting. We state some examples. The famous theorem of Mantel from 1907 [30] states that every n-vertex graph with more than $\lfloor n^2/4 \rfloor$ edges necessarily contains a triangle. On the other hand, it is not hard to

show that every subset A of $[n] := \{1, \ldots, n\}$ of size more than $\lceil n/2 \rceil$ necessarily contains a *Schur triple*, or *sum*; that is, a triple $\{x, y, z\}$ of not necessarily distinct elements such that $x + y = z$. The name dates back to a result of Schur from 1916 which states that every r-colouring of $[n]$ contains yields a monochromatic Schur triple when n is sufficiently large. Its graph-theoretic counterpart is Ramsey's theorem from 1928 which guarantees a monochromatic clique in any r-edge-colouring of a sufficiently large complete graph. The triangle removal lemma of Ruzsa and Szemerédi [33] states that every n-vertex graph containing $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges. In the sum-free setting, the removal lemma of Green [15], and Král', Serra and Vena [27] states that every $A \subseteq [n]$ containing $o(n^2)$ Schur triples can be made sum-free by removing $o(n)$ elements. Erdős, Kleitman and Rothschild [12] proved that the number of n-vertex triangle-free graphs is $2^{n^2/4+o(n^2)}$, that is, the obvious lower bound of taking every subgraph of a maximal triangle-free graph is, in a sense, tight. In the sum-free setting, resolving a conjecture of Cameron and Erdős [7], Green [14] and independently Sapozhenko [34] proved that, for every $i = 0, 1$, there exists a constant C_i, depending only on the parity of n, such that $[n]$ contains $(C_i + o(1))2^{n/2}$ sum-free sets. So again, the obvious lower bound is tight.

1.1 The Erdős-Rothschild problem for cliques in graphs

Taking inspiration from the extremal graph theory literature, in this paper we consider another classical graph problem in the sum-free setting: the problem of Erdős and Rothschild [10, 11], which is stated as follows. Given an n-vertex graph G and positive integers r, k, say that a colouring of its edges with r colours (an *r-edge-colouring*) is *valid* if there are no monochromatic copies of K_k. Among all such graphs G, what is the maximum number $F(n, r, k)$ of valid colourings? Ramsey's theorem implies that any graph containing a sufficiently large clique has no valid colourings. Clearly, any colouring of a K_k-free graph is valid. Turán's theorem implies that the largest such graph is $T_{k-1}(n)$, the complete balanced $(k-1)$-partite graph. Thus we obtain the bound

$$F(n, r, k) \geqslant r^{t_{k-1}(n)}, \tag{1.1}$$

where $t_{k-1}(n)$ is the number of edges in $T_{k-1}(n)$. Erdős and Rothschild conjectured that this trivial lower bound is tight when n is large and $(r, k) = (2, 3)$ (i.e. one uses two colours and forbids monochromatic triangles), and further, that $T_2(n)$ is the unique extremal graph. This was confirmed for all $n \geqslant 6$ by Yuster [38], who himself extended the conjecture to larger cliques in the case $r = 2$. This was in turn verified by Alon, Balogh, Keevash and Sudakov [1], showing that for $r \in \{2, 3\}$, equality holds in (1.1), and $T_{k-1}(n)$ is the unique extremal graph. They further showed that for all $r, k \in \mathbb{N}$ such that $k \geqslant 3$ and $r \geqslant 2$, the limit $\lim_{n \to \infty} \log_r F(n, r, k)/\binom{n}{2}$ exists. In other words, there is some $h(r, k)$ such that

$$F(n, r, k) = r^{h(r,k)\binom{n}{2}+o(n^2)}. \tag{1.2}$$

However, for every other choice of (r, k), there exists a constant $c = c(r, k)$ such that $h(r, k) > (k-2)/(k-1) + c$. Note that the edge density of $T_{k-1}(n)$ tends to $(k-2)/(k-1)$

with n, so this says that the trivial lower bound in (1.1) is *not* correct for all other choices of (r, k). Observe that this implies that extremal graphs must therefore contain many copies of the forbidden K_k. The authors were, however, able to determine $h(3, 4)$ and $h(4, 4)$. The exact results in these cases were obtained by Pikhurko and Yilma [32], who showed that the unique extremal graphs for $F(n, 3, 4)$ and $F(n, 4, 4)$ are $T_4(n)$ and $T_9(n)$ respectively.

The problem remains unresolved, even asymptotically in the logarithmic, in all other cases. In particular, there is no (approximate) solution when the number of colours r is at least five, for any $k \geqslant 3$. A generalisation of the Erdős-Rothschild problem was considered by Pikhurko, Yilma and the third author of this paper [31], wherein one may forbid differently-sized cliques for different colours. A special case of the main result is that there is a certain finite optimisation problem whose maximum is equal to $h(r, k)$. Roughly speaking, they showed that to determine $F(n, r, k)$ up to an error in the exponent, one should maximise over all possible 'layerings' of 'small' K_k-free graphs.

1.2 The Erdős-Rothschild problem in the sum-free setting: our results

As stated above, the purpose of this paper is to initiate the study of the Erdős-Rothschild problem in the case when our underlying discrete structure is not K_n but $[n]$, and the forbidden substructure is not a triangle but a Schur triple. Let us formulate the problem precisely.

Problem 1.1. *Given positive integers* n, r, *determine* $f(n, r)$, *defined as follows. For each* $A \subseteq [n]$, *say that a colouring* $\sigma : A \to [r]$ *of* A *with* r *colours is* valid *if it contains no monochromatic sums. In other words,* $\sigma^{-1}(c)$ *is a sum-free set for each colour* $c \in [r]$. *Let* $f(A, r)$ *be the number of valid colourings of* A *and let*

$$f(n, r) := \max_{A \subseteq [n]} f(A, r).$$

Notice that Schur's theorem says that $f([n], r) = 0$ whenever n is sufficiently large (as a function of r). Before stating our results, let us see what we might conjecture via analogy with the graph setting. Here, for $r \in \{2, 3\}$, the unique extremal graph was the largest triangle-free graph. It is not hard to see that every largest sum-free sets in $[n]$ has size at most $\lceil n/2 \rceil$, and the only subsets attaining this bound are

(i) $O := \{1, 3, 5, \ldots, 2\lceil n/2 \rceil - 1\}$ and

(ii) $I_2 := \{\lfloor n/2 \rfloor + 1, \lfloor n/2 \rfloor + 2, \ldots, n\}$.

Additionally, if n is even, the set $\{n/2, \ldots, n-1\}$ is also a largest sum-free subset.

Our first theorem is an exact result for two colours, which states that the largest sum-free subsets of $[n]$ are also extremal subsets for Problem 1.1.

Theorem 1.2. *There exists $n_0 > 0$ such that for all integers $n \geqslant n_0$, we have*

$$f(n, 2) = 2^{\lceil n/2 \rceil}.$$

Moreover, the only extremal subsets are O, I_2*; and if n is even, we additionally have* $\{n/2, \ldots, n-1\}, \{n/2, \ldots, n\}$.

Note that in the sum-free setting with two colours, there is an extremal subset which is *not* sum-free, namely $\{n/2, \ldots, n\}$, containing at most one sum. In the graph setting with two colours, there is a unique extremal graph, namely $T_2(n)$, which is K_3-free.

As in the graph setting, our results decrease in strength as the number of colours increases. When $r = 3$ we can obtain the following stability theorem.

Theorem 1.3. *For all positive integers n, we have*

$$f(n, 3) = 3^{n/2 + o(n)}.$$

Moreover, the following holds. For all $\varepsilon > 0$ there exist $\delta, n_0 > 0$ such that for all integers $n \geqslant n_0$, whenever $A \subseteq [n]$ satisfies $f(A, 3) \geqslant f(n, 3) \cdot 2^{-\delta n}$, we have that either $|A \triangle \mathrm{O}| \leqslant \varepsilon n$; or $|A \triangle I_2| \leqslant \varepsilon n$.

(Here, $\triangle$ denotes symmetric difference; see Section 2.1.) Finally, we asymptotically determine the logarithm of $f(n, r)$ when $r \in \{4, 5\}$.

Theorem 1.4. *For $r \in \{4, 5\}$ and all positive integers n, we have*

$$f(n, r) = \left(r \left\lfloor \frac{r^2}{4} \right\rfloor \right)^{\frac{n}{4} + o(n)}.$$

In particular, we are able to asymptotically solve the 5-colour case of the Erdős-Rothschild problem in the sum-free setting, in contrast to the graph setting in which it is wide open for triangles and larger cliques. (This turns out to be a consequence of the rigid structure of large maximal sum-free sets.) The (asymptotic) lower bounds in Theorems 1.2–1.4 come from the facts that

$$f(\mathrm{O}, r) = f(I_2, r) = r^{\lceil n/2 \rceil} \quad \text{and} \quad f(\mathrm{O} \cup I_2, r) \geqslant \left(r \left\lfloor \frac{r^2}{4} \right\rfloor \right)^{\left\lfloor \frac{n}{4} \right\rfloor}.$$

The first assertion follows from the fact that every r-colouring of a sum-free set A is valid, so $f(A, r) = r^{|A|}$. For the second, note that any colouring $\sigma : \mathrm{O} \cup I_2 \to [r]$ with $\sigma(x) \in \{1, \ldots, \lfloor r/2 \rfloor\}$ whenever $x \in \mathrm{O} \setminus I_2$, and $\sigma(y) \in \{\lfloor r/2 \rfloor + 1, \ldots, r\}$ whenever $y \in I_2 \setminus \mathrm{O}$ (and $\mathrm{O} \cap I_2$ coloured arbitrarily), is valid, giving the claimed bound.

Note that $r^{1/2} > (r \lfloor r^2/4 \rfloor)^{1/4}$ for $r \in \{2, 3\}$, while this inequality becomes an equality for $r = 4$, and reverses for $r \geqslant 5$. It is tempting to believe that the bound in Theorem 1.4 holds for all $r \geqslant 4$. However, this is *not* true. Indeed, when r is large, the set $A :=$ $\mathrm{O} \cup I_2 \cup \{x \in [n] : x \equiv 1, 4 \mod 5\}$ contains exponentially more valid r-colourings than

$O \cup I_2$. To see this, suppose for simplicity that $3|r$, and partition the colours $[r]$ into three sets $\mathcal{C}_O, \mathcal{C}_{I_2}, \mathcal{C}_{1,4}$ of equal size. Let any colour in $\mathcal{C}_O$ be *allowed* for any $x \in O$, and do the same for $\mathcal{C}_{I_2}$ and I_2, and for $\mathcal{C}_{1,4}$ and $\{x \equiv 1, 4 \mod 5\}$. So, for example the largest $10t + 1$ which is less than n has all r colours allowed since it lies in all three sets. Any colouring which only uses allowed colours for each vertex is valid, since each of the three sets is sum-free. Calculating the sizes of the intersections, we see that

$$f(n, r) \geqslant \left(\frac{r}{3}\right)^{\frac{8n}{20}} \left(\frac{2r}{3}\right)^{\frac{7n}{20}} r^{\frac{2n}{20}} \cdot r^{o(n)} \geqslant r^{\frac{17n}{20}+o(n)}.$$

It would also be interesting to see if for every fixed integer $r \geqslant 6$, the limit

$$\lim_{n \to \infty} \log f(n, r)/n$$

exists.

1.3 Some remarks on the methods and proofs

An important tool in our proof is the Green's container theorem for finite abelian groups (Theorem 2.2). The special case that we need states that, for every positive integer n, there is a small family $\mathcal{F}$ of subsets of $[n]$, called *containers*, each of which is almost sum-free, and such that every sum-free subset of $[n]$ lies in some member of $\mathcal{F}$.

In the proof of the main result in [1] and other Erdős-Rothschild-type results for graphs, Szemerédi's regularity lemma [36] is used to approximate a large graph G by another graph of bounded size (the *reduced graph*). Then, for each valid r-edge-colouring σ of G, for each $i \in [r]$ one can approximate the K_k-free subgraph $\sigma^{-1}(i)$ of G by a K_k-free graph of bounded size. In our proofs, for each valid r-colouring of $[n]$, we approximate the sum-free subset $\sigma^{-1}(i)$ of $[n]$ by a container.

We use Theorem 2.2 to reduce the problem of determining $f(n, r)$ to solving an optimisation problem (Problem 2) whose maximum approaches $\log f(n, r)/n$ as n tends to infinity (Theorem 3.3). Roughly speaking, Problem 2 involves layering sum-free subsets $A_1, \ldots, A_r$ of $[n]$ and measuring a weighted overlap $g(A_1, \ldots, A_r)$.

To attack Problem 2, we require a second important tool, namely a very strong stability theorem of Deshouillers, Freiman, Sós, and Temkin [9], which was recently strengthened by Tran [37]. This states that every sum-free subset of $[n]$ is either 'small', or has a very rigid structure: either it contains only odd elements, or it somehow resembles the interval I_2. Now it turns out that, if r is small and $(A_1, \ldots, A_r)$ is a maximiser for Problem 2, then at most one of the A_i can be small. The rigid structure of the others means that the feasible set for Problem 2 is not too large. In fact, for $r \in \{2, 3\}$, it can be easily solved at this stage, and we find that either all of $A_1, \ldots, A_r$ are close to O; or they are all close to I_2. This proves Theorem 1.3 and completes the first step of the proof of Theorem 1.2.

To complete the proof of Theorem 1.2, we have two cases to consider. The solution to Problem 2 when $r = 2$ implies that any subset $A \subseteq [n]$ with $f(A, r) = f(n, r)$ satisfies either (1) $|A \bigtriangleup O| = o(n)$; or (2) $|A \bigtriangleup I_2| = o(n)$. We use stability arguments, together with techniques from [5], to obtain the exact structure of A.

For $r \in \{4, 5\}$, we find a reduction of Problem 2 to a *linear* optimisation problem. First, for each $i \in [r]$, and any feasible $(A_1, \ldots, A_r) \subseteq [n]^r$, we obtain $d_i \in [0, 1]$ which are each functions of $A_1, \ldots, A_r$ and n and such that $g(A_1, \ldots, A_r)$ is linear in $d_1, \ldots, d_r$. Now, using the structural information returned from stability, we obtain constraints, linear in $d_1, \ldots, d_r$, which every maximiser $(A_1, \ldots, A_r)$ must satisfy. This gives rise to a linear program in the variables $d_1, \ldots, d_r$. Now, this linear program is a relaxation of Problem 2, so its maximiser may not correspond to a feasible solution $(A_1, \ldots, A_r)$ of Problem 2. But, if we can exhibit a feasible tuple $(A_1, \ldots, A_r)$ such that the maximum M of this program satisfies $M = g(A_1, \ldots, A_r)$, then $(A_1, \ldots, A_r)$ *is* a maximiser of Problem 2. Thus our task is to find enough constraints (of sufficient strength) so that this is possible. In so doing, we will prove Theorem 1.4.

1.4 The Erdős-Rothschild problem in other settings

Erdős and Rothschild also considered the problem of counting monochromatic H-free colourings, for an arbitrary fixed graph H. In [1], it is shown that the analogue of their main result for cliques in fact holds when H is colour-critical. Further cases including matchings, stars, paths, trees were investigated in [20, 21]. Other works have considered a fixed forbidden colour pattern of H, see [3, 6, 22, 24].

An analogous problem for directed graphs was solved by Alon and Yuster [2], who determined, for each k-vertex tournament T, the maximum number of T-free orientations of an n-vertex graph, when n is sufficiently large. The hypergraph analogue was addressed in [19, 28, 29].

The authors of [8] and [19] considered the problem of counting the number of colourings of families of r-sets such that every colour class is ℓ-intersecting. A related result in the context of vector spaces over a finite field $GF(q)$ is proved in [23]. These results are Erdős-Rothschild versions of the classical Erdős-Ko-Rado theorem.

During the preparation of this paper, we became aware of the results of Hàn and Jiménez [18] who recently studied similar questions in the setting of finite abelian groups. Given a finite abelian group $(\Gamma, +)$, define an r-colouring of $A \subseteq \Gamma$ to be *valid* if it has no monochromatic sum. Let $f(\Gamma, r)$ be the maximum number of valid r-colourings among all subsets A of Γ. The results of Hàn and Jiménez show a close relationship between $f(\Gamma, r)$ and the largest sum-free sets of Γ, and characterise for $r \leqslant 5$ the extremal sets. As in [18], our proof begins with an application of the container method. But, in the setting of abelian groups, there is much more structure than in the integer setting, and thus one can obtain much stronger stability, leading to precise results for larger values of r than those obtained here. Most of the work in this paper goes into dealing with 'interval-like' sets, (an analogue of which) does not appear in the abelian group setting. We remark that Hàn and Jiménez's result and ours do not imply one another.

1.5 Organisation of the paper

Section 2 sets up the notation we will use and contains the statements of results on sum-free sets necessary for the proof. In Section 3 we define Problem 2, the optimisation

problem whose maximum is a parameter $g(n, r)$ which is closely related to $f(n, r)$. Then in Sections 4, 5 and 6 we prove Theorems 1.3, 1.2 and 1.4 respectively. We make some concluding remarks in Section 7.

2 Notation and Preliminaries

In this section we define the notation that we will use, and some results on sum-free subsets which are needed in our proofs.

2.1 Notation

Given integers m, n such that $m \leqslant n$, we write $[m, n]$ to denote the set $\{m, \ldots, n\}$, and write $[n] := [1, n]$. For a set $A \subseteq [n]$, we define $d(A) = |A|/n$ and $\min(A)$ to be the *density* and the minimum element of A, respectively. We also define E and O to be respectively the set of all even and odd integers in $[n]$. As we defined earlier, $I_2 := [\lfloor n/2 \rfloor + 1, n]$ and $I_1 := [\lfloor n/2 \rfloor]$ (we suppress the dependence on n in the notation). Given $A, B \subseteq \mathbb{Z}$, we write $A + B := \{a + b : a \in A, b \in B\}$, and $A \triangle B := (A \setminus B) \cup (B \setminus A)$ denotes the *symmetric difference* of A and B. For any $x \in \mathbb{Z}$, we also write $x \cdot A := \{xa : a \in A\}$. Logarithms will always be taken to the base 2.

2.2 Tools for sum-free subsets

The first result we state is a very strong stability theorem for sum-free subsets due to Deshouillers, Freiman, Temkin and Sós [9]. It states that every large sum-free set $S \subseteq [n]$ either contains no even number, or is, in a certain sense, close to the interval I_2.

Theorem 2.1 ([9])**.** *Every sum-free set S in $[n]$ satisfies at least one of the following conditions:*

(a) $|S| \leqslant 2n/5 + 1$;

(b) S consists of odd numbers;

(c) $|S| \leqslant \min(S)$.

Throughout the rest of the paper, we refer to such sum-free sets as *type (a)*, *type (b)*, and *type (c)* respectively.

We use the following *container theorem* of Green [15] (see also Green-Ruzsa [16, 17]), which, for large n, guarantees a small collection of subsets of $[n]$ which somehow approximates the collection of sum-free sets. We should also mention that (hyper)graph containers have been used successfully in many contexts, see [4, 25, 26, 35].

Theorem 2.2 (Proposition 6 in [14])**.** *For all $\varepsilon > 0$ there exists $n_0 > 0$ such that, for all integers $n \geqslant n_0$, there exists a family $\mathcal{F}$ of subsets of $[n]$ with the following properties:*

(i) Every $F \in \mathcal{F}$ contains at most εn^2 Schur triples;

(ii) If $S \subseteq [n]$ is sum-free, then $S \subseteq F$ for some $F \in \mathcal{F}$;

(iii) $|\mathcal{F}| \leqslant 2^{\varepsilon n}$;

(iv) $|F| \leqslant (1/2 + \varepsilon)n$ for all $F \in \mathcal{F}$.

It is convenient for us to include property (iv) although it is not explicitly stated in [14], but follows immediately from (i), Theorem 2.1 and Theorem 2.3. Given a sum-free set $S \subseteq [n]$, the set $F \in \mathcal{F}$ guaranteed by (ii) is called a *container* for S. We also need the following removal lemma of Green [14, 15], and Král', Serra and Vena [27], which guarantees that a subset of $[n]$ containing $o(n^2)$ sums can be made sum-free by removing $o(n)$ elements.

Theorem 2.3 ([14, 15, 27]). *For all $\varepsilon > 0$, there exists $\delta, n_0 > 0$ such that the following holds for all integers $n \geqslant n_0$. Suppose that $A \subseteq [n]$ is a set containing at most δn^2 Schur triples. Then there exist $B, C \subseteq [n]$ such that $A = B \cup C$ where B is sum-free and $|C| \leqslant \varepsilon n$.*

This result comes from applying a regularity lemma and consequently the dependence of δ on ε is rather poor: $1/\delta$ is at most a tower of twos of height logarithmic in $1/\varepsilon$, which follows from [27] and [13]. Finally, we will use the following folklore inequality which bounds the size of the set $A + B$:

Proposition 2.4. *For all finite non-empty subsets A, B of $\mathbb{Z}$, we have that $|A + B| \geqslant |A| + |B| - 1$.*

Proof. Write $A := \{a_1 < \cdots < a_k\}$ and $B := \{b_1 < \cdots < b_\ell\}$. Then $A + B \supseteq \{a_1 + b_1 < a_1 + b_2 < \cdots < a_1 + b_\ell < a_2 + b_\ell < a_3 + b_\ell < \cdots < a_k + b_\ell\}$. $\square$

3   An equivalent covering problem

In this section, we define a new maximisation problem whose value $g(n, r)$ is closely related to $f(n, r)$. Then, for the rest of the paper, it suffices to consider this new problem. To motivate the problem, consider the following procedure for finding a subset $A \subseteq [n]$ with many valid colourings. Let $r \in \mathbb{N}$ be the number of colours, as usual, and choose sum-free subsets $A_1, \ldots, A_r$ of $[n]$. Then the number of valid colourings of $\bigcup_{i \in [r]} A_i$ is at least the number of colourings σ which colours x with some i such that $x \in A_i$. If x lies in many A_i then the number of choices for $\sigma(x)$ is large. So a choice of $A_1, \ldots, A_r$ with a large appropriately weighted overlap generates many valid colourings. We now make this precise.

Problem 3.1. *Given $n, r \in \mathbb{N}$, determine $g(n, r)$, defined as follows. Given a tuple $(A_1, \ldots, A_r)$ of sum-free subsets of $[n]$, for each $I \in 2^{[r]}$, let $E_I := \bigcap_{i \in I} A_i \setminus \bigcup_{j \notin I} A_j$ be the set of $x \in [n]$ which lie in A_i if and only if $i \in I$. Define*

$$g(A_1, \ldots, A_r) := \frac{1}{n} \sum_{I \in 2^{[r]} \setminus \{\emptyset\}} |E_I| \log |I|.$$

Equivalently, for each $i \in [r]$ let $D_i := \bigcup_{I \in 2^{[r]} : |I| = i} E_I$; that is, the set of all elements that are in exactly i different A_j's. Let $d_i := |D_i|/n$ and define $(d_1, \ldots, d_r)$ to be the *intersection vector* of $(A_1, \ldots, A_r)$. Let $d_0 := 1 - \sum_{i \in [r]} d_i$. Then

$$g(A_1, \ldots, A_r) = \sum_{i \in [r]} d_i \log i.$$

Define

$$g(n, r) := \max \left\{ g(A_1, \ldots, A_r) : A_i \subseteq [n] \text{ is sum-free for all } i \in [r] \right\}.$$

Define also $g(A, r)$ to be the maximum of $g(A_1, \ldots, A_r)$ over all tuples of sum-free subsets of $[n]$ such that $\bigcup_{i \in [r]} A_i = A$.

Remark 3.2. It is not hard to see that, for every n and r, there is always some tuple $(A_1, \ldots, A_r)$ of sum-free subsets of $[n]$ which is *extremal* (that is, $g(A_1, \ldots, A_r) = g(n, r)$), and A_i is a *maximal* sum-free subset for all $i \in [r]$. It will be useful to choose such an extremal tuple later, since if we know e.g. that A_i contains no even element, then by Theorem 2.1 we can assume that $A_i = O$.

The first step in the proofs of Theorems 1.2–1.4 is to show that these problems are, in a sense, equivalent.

Theorem 3.3. *For all $\varepsilon > 0$ and $r \in \mathbb{N}$, there exists $n_0 > 0$ such that the following holds for all integers $n \geqslant n_0$. Let $A \subseteq [n]$. Then there exists $A' \subseteq A$ with $|A'| \geqslant |A| - \varepsilon n$ for which*

$$2^{g(A,r)n} \leqslant f(A, r) \leqslant 2^{(g(A',r)+\varepsilon)n}. \tag{3.1}$$

Therefore

$$g(n, r) \leqslant \frac{\log f(n, r)}{n} \leqslant g(n, r) + \varepsilon. \tag{3.2}$$

Proof. We first prove the lower bound (for all n and r). Fix integers n, r and let $A \subseteq [n]$. Choose a tuple $(A_1, \ldots, A_r)$ of sum-free subsets of $[n]$ whose intersection vector $(d_1, \ldots, d_r)$ is extremal, i.e. satisfies

$$\sum_{i \in [r]} d_i \log i = g(A_1, \ldots, A_r) = g(A, r). \tag{3.3}$$

For each $I \in 2^{[r]} \setminus \{\emptyset\}$, define E_I as in the statement of Problem 3.1. Consider any colouring $\sigma : A \to [r]$ such that, for each $I \in 2^{[r]} \setminus \{\emptyset\}$ and $x \in E_I$, we have $\sigma(x) \in I$. Then $\sigma^{-1}(i) \subseteq A_i$ for all $i \in [r]$, so the fact that A_i is sum-free for all $i \in [r]$ implies that σ is valid. Thus the number of such σ is a lower bound for the total number of valid colourings, and so

$$f(A, r) \geqslant \prod_{I \in 2^{[r]} \setminus \{\emptyset\}} |I|^{|E_I|} = \prod_{i \in [r]} i^{d_i n} \overset{(3.3)}{=} 2^{g(A,r)n},$$

as required.

For the remainder of the proof we focus on the upper bound. Fix an integer r and let $\varepsilon > 0$. We may assume that $\varepsilon \ll 1/r$. Choose η such that $0 < \eta \ll \varepsilon$. Apply Theorem 2.3 to obtain $\gamma, n_0 > 0$ such that, for all integers $n \geqslant n_0$, every $A \subseteq [n]$ which contains at most γn^2 Schur triples may be made sum-free by removing at most ηn elements. Without loss of generality we may assume that $\gamma \ll \eta$. Theorem 2.2 implies that, by increasing n_0 if necessary, for all integers $n \geqslant n_0$, there exists a family $\mathcal{F} = \mathcal{F}_n$ of containers such that

(i) every $F \in \mathcal{F}$ contains at most γn^2 Schur triples;

(ii) every sum-free subset of $[n]$ lies in at least one $F \in \mathcal{F}$;

(iii) $|\mathcal{F}| \leqslant 2^{\gamma n}$; and

(iv) $|F| \leqslant (1/2 + \gamma)n$ for all $F \in \mathcal{F}$.

Without loss of generality, we may assume that $1/n_0 \ll \gamma$ and $\varepsilon \ll 1$. We have the hierarchy

$$1/n_0 \ll \gamma \ll \eta \ll \varepsilon \ll 1.$$

Given any $n \geqslant n_0$ and the family $\mathcal{F}$ of containers, for each $F \in \mathcal{F}$, fix a largest sum-free subset F^* of F (note that there may be more than one choice of F^*, but we just pick one of these). Then (i) together with Theorem 2.3 implies that $|F| - \eta n \leqslant |F^*| \leqslant |F|$.

Now let $n \geqslant n_0$ be an integer, and $A \subseteq [n]$ be arbitrary. Consider any fixed valid r-colouring σ of A. Then $\sigma^{-1}(i)$ is sum-free for all $i \in [r]$. By (ii), we may choose a tuple $(F_1, \ldots, F_r) \in \mathcal{F}^r$ of containers such that $\sigma^{-1}(i) \subseteq F_i$. By (i), F_i contains at most γn^2 Schur triples for all $i \in [r]$. For each $i \in [r]$, let us write $F_i^* := (F_i)^*$ for the largest sum-free subset of F_i we fixed earlier. Then $|F_i^*| \geqslant |F_i| - \eta n$. Thus, for each valid colouring $\sigma : A \to [r]$, we obtain a tuple $(F_1^*, \ldots, F_r^*)$. Observe that

$$|\sigma^{-1}(i) \setminus F_i^*| \leqslant \eta n, \tag{3.4}$$

but F_i^* may contain many elements which do not lie in $\sigma^{-1}(i)$.

We now claim that the following procedure generates every valid colouring σ of A, and therefore the number of choices in this procedure is an upper bound on $f(A, r)$. Each choice will generate a colouring τ:

1. For all $i \in [r]$, choose a container $G_i \in \mathcal{F}$, and let G_i^* be the largest sum-free subset of G_i we fixed earlier.

2. For each $I \in 2^{[r]} \setminus \{\emptyset\}$, let

$$E_I' := \left(\bigcap_{i \in I} G_i^* \setminus \bigcup_{j \notin I} G_j^* \right) \cap A.$$

Let also $D_i' := \bigcup_{I \subseteq [r] : |I| = i} E_I'$ for each $i \in [r]$. So D_i' is the set of those elements in A which lie in exactly i of the G_j^*. Let $d_i' := |D_i'|/n$ for all $i \in [r]$.

3. For each $I \in 2^{[r]} \setminus \{\emptyset\}$ and $x \in E_I'$, choose $i \in I$ and set $\tau(x) := i$.

4. For each uncoloured $y \in A$, let $\tau(y) \in [r]$ be arbitrary.

We need to show that there is a choice in (1)–(4) which will yield $\tau = \sigma$. In (1), for each $i \in [r]$, (ii) and the fact that σ is valid implies that we can choose $G_i := F_i \in \mathcal{F}$ such that $\sigma^{-1}(i) \subseteq F_i$. Note that $G_i^* = F_i^*$ for all $i \in [r]$. The choice in (2) is fixed by our choices in (1). In (3), by construction, for every $x \in (\bigcup_{i\in[r]} F_i^*) \cap A$, we have that $x \in E_I'$ for some $I \ni \sigma(x)$. Thus for every $x \in (\bigcup_{i\in[r]} F_i^*) \cap A$ we can choose $\tau(x) := \sigma(x)$. In (4) we are free to colour the uncoloured elements of A with σ. Since σ was an arbitrary valid colouring of A, we have proved the claim.

Thus it remains to count the number of colourings generated by (1)–(4). Given a tuple $(G_1, \ldots, G_r) \in \mathcal{F}^r$ of containers, let $\mathcal{C}$ be the set of colourings $\tau : A \to [r]$ generated by it, i.e. the set of τ which arise from the procedure after fixing the choice $(G_1, \ldots, G_r)$ in (1). Observe that $(G_1, \ldots, G_r)$ gives rise to a unique tuple $(G_1^*, \ldots, G_r^*)$. Then, since the only choices are in (3) and (4), we have

$$|\mathcal{C}| \leqslant \prod_{j\in[r]} j^{|D_j'|} \cdot r^{\left|A \setminus \bigcup_{i\in[r]} G_i^*\right|} \overset{(3.4)}{\leqslant} \prod_{j\in[r]} j^{d_j' n} \cdot r^{r\eta n}.$$

Taking logarithms, we have, bounding very loosely, that

$$\frac{\log|\mathcal{C}|}{n} \leqslant \sum_{j\in[r]} d_j' \log j + \sqrt{\eta} = g(G_1^* \cap A, \ldots, G_r^* \cap A) + \sqrt{\eta} \leqslant g(A', r) + \sqrt{\eta}, \qquad (3.5)$$

where $A' := A \cap \bigcup_{i\in[r]} G_i^*$. So $|A'| \geqslant |A| - r\eta n \geqslant |A| - \varepsilon n$. But, by (iii), the number of choices of $(G_1, \ldots, G_r) \in \mathcal{F}^r$ is at most $|\mathcal{F}|^r \leqslant 2^{r\gamma n}$, so

$$f(A, r) \leqslant 2^{r\gamma n} \cdot 2^{g(A',r)n} \cdot 2^{\sqrt{\eta} n} \leqslant 2^{(g(A',r)+2\sqrt{\eta})n} \leqslant 2^{(g(A',r)+\varepsilon)n},$$

completing the proof of the upper bound. The second assertion is an obvious consequence of the first. $\qquad \square$

The discussion about lower bounds after the statement of Theorem 1.4 amounts to the following inequalities: For all integers $n \geqslant r \geqslant 2$, we have

$$\frac{\log f(n,r)}{n} \geqslant g(n,r) \geqslant g(\underbrace{I_2, \ldots, I_2}_{r}) = \frac{\lceil n/2 \rceil}{n} \cdot \log r; \quad \text{and} \qquad (3.6)$$

$$\frac{\log f(n,r)}{n} \geqslant g(n,r) \geqslant g(\underbrace{I_2, \ldots, I_2}_{\lfloor r/2 \rfloor}, \underbrace{O, \ldots, O}_{\lceil r/2 \rceil}) \geqslant \frac{\lfloor n/4 \rfloor}{n} \log\left(r \left\lfloor \frac{r^2}{4} \right\rfloor\right). \qquad (3.7)$$

Figure 1 shows these three constructions in the case when $r = 4$, when they each give rise to roughly the same lower bound. Theorem 1.4) implies that each one is in fact an approximate optimal solution of Problem 1.
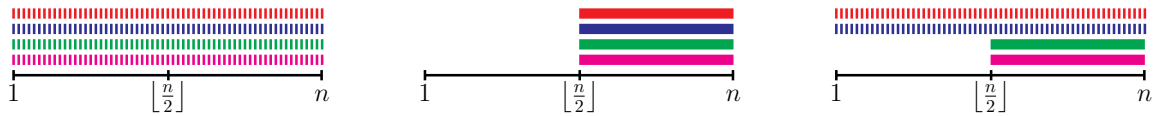
Figure 1: Three approximate solutions to Problem 2 for $r = 4$.

Observe the following easy correspondence between feasible solutions of Problems 1 and 2. Given a feasible solution $A \subseteq [n]$ of Problem 1 and a valid r-colouring σ of A, we have that $(\sigma^{-1}(1), \ldots, \sigma^{-1}(r))$ is a feasible solution of Problem 2. Given a feasible solution $(A_1, \ldots, A_r) \subseteq [n]^r$ of Problem 3.1, we have that $A_1 \cup \ldots \cup A_r$ is a feasible solution of Problem 1.1.

Theorem 3.3 is essentially an analogue of the main result of [31]. Recall the parameters $h(r, k)$ and $F(n, r, k)$ related to the graph problem defined in the introduction, related by (1.2). Informally speaking, determining $g(n, r)$ involves layering r sum-free subsets of $[n]$ so that an appropriately weighted overlap is as large as possible, whereas determining $h(r, k)$ involves layering r finite K_k-free graphs so that their weighted overlap is as large as possible. Importantly and unfortunately, $g(n, r)$ does of course depend on n. However, the cases in which $F(n, r, k)$ has been determined (when r is small) give us some valuable intuition for determining $g(n, r)$ (and hence approximately determining $f(n, r)$): namely that for an extremal tuple $(A_1, \ldots, A_r)$ of sum-free sets, each A_i should perhaps be a *largest* sum-free set: either O or I_2. Unlike in the case of graphs, our ground set $[n]$ comes with a fixed labelling. So there is only one way to layer, say, O and I_2, whereas there are many ways to layer any two r-vertex graphs G and H.

4 The proof of Theorem 1.3

Given Theorem 3.3, it is now a fairly simple task to obtain stability in the case when $r = 2, 3$. Indeed, it suffices to prove stability for Problem 2.

Lemma 4.1. *Let $r \in \{2, 3\}$. For all $\varepsilon > 0$, there exists $n_0 > 0$ such that the following holds for all integers $n \geqslant n_0$. Let $A \subseteq [n]$ be such that $g(A, r) \geqslant g(n, r) - \varepsilon/(50r)$. Then either $|A \triangle \mathrm{O}| \leqslant \varepsilon n$ or $|A \triangle I_2| \leqslant \varepsilon n$.*

Proof. Let $\varepsilon > 0$ and assume without loss of generality that $\varepsilon < 1/100$. Let also n be a sufficiently large integer compared to ε. Choose a tuple $(A_1, \ldots, A_r)$ of sum-free subsets of $[n]$ such that $g(A_1, \ldots, A_r) = g(A, r)$. Let $(d_1, \ldots, d_r)$ be its intersection vector. Recall from (3.6) that $g(n, r) \geqslant (1/2) \cdot \log r$. So $g(A, r) \geqslant (1/2) \cdot \log r - \varepsilon/(50r)$. We will need the following claim.

Claim 4.2. *It suffices to show that $d_r \geqslant 1/2 - \varepsilon/(3r)$.*

<u>Proof:</u> Let $A' := A_1 \cap \ldots \cap A_r$. Assume that $|A'| = d_r n \geqslant (1/2 - \varepsilon/(3r))n$. Clearly A' is a sum-free subset of $[n]$. Suppose that A' contains at least one even element.

Then Theorem 2.1 implies that A' is a sum-free set of type (c), so $\min(A') \geqslant |A'| \geqslant (1/2 - \varepsilon/(3r))n$, and so $|A' \triangle I_2| \leqslant 2\varepsilon n/(3r)$. Otherwise (if every element of A' is odd) we have $|A' \triangle \mathrm{O}| \leqslant \varepsilon n/(3r)$. Finally, since every sum-free subset of $[n]$ has size at most $\lceil n/2 \rceil$, we have

$$|A \triangle A'| \leqslant \sum_{i \in [r]} |A_i \triangle A'| \leqslant r(\lceil n/2 \rceil - d_r n) \leqslant \varepsilon n/3 + r.$$

So, in the first case (if $|A' \triangle I_2| \leqslant 2\varepsilon n/(3r)$), we have $|A \triangle I_2| \leqslant |A \triangle A'| + |A' \triangle I_2| \leqslant \varepsilon n$. Similarly, in the second case, we have $|A \triangle \mathrm{O}| \leqslant \varepsilon n$. ∎

First consider the case when $r = 2$. Then

$$\frac{1}{2} - \frac{\varepsilon}{50r} \leqslant g(A, 2) = g(A_1, A_2) = d_2,$$

as required. Now let $r = 3$. Then

$$\frac{1}{2} \cdot \log 3 - \frac{\varepsilon}{50r} \leqslant g(A, 3) = g(A_1, A_2, A_3) = d_2 + d_3 \cdot \log 3. \tag{4.1}$$

Recall that each A_i is sum-free with size at most $\lceil n/2 \rceil$, so

$$d_1 + 2d_2 + 3d_3 = \frac{1}{n}\left(|A_1| + |A_2| + |A_3|\right) \leqslant \frac{3}{n}\left\lceil \frac{n}{2} \right\rceil \leqslant \frac{3}{2} + \frac{\varepsilon}{100r},$$

which implies that $d_2 + 3d_3/2 \leqslant 3/4 + \varepsilon/(200r)$. Therefore, if $d_3 < 1/2 - \varepsilon/(3r)$, we have

$$d_2 + d_3 \cdot \log 3 = d_2 + \frac{3}{2}d_3 + \left(\log 3 - \frac{3}{2}\right) \cdot d_3 < \frac{3}{4} + \frac{\varepsilon}{200r} + \left(\log 3 - \frac{3}{2}\right) \cdot \left(\frac{1}{2} - \frac{\varepsilon}{3r}\right)$$

$$= \frac{1}{2}\log 3 - \left(\frac{\log 3}{3} - \frac{1}{200} - \frac{1}{2}\right) \cdot \frac{\varepsilon}{r} < \frac{1}{2}\log 3 - \frac{\varepsilon}{50r},$$

a contradiction to (4.1). □

Note that, in the case $r = 3$, the proof proceeds by solving a linear program in variables d_1, d_2, d_3 (and similarly for $r = 2$, but the program is trivial here). The very same linear programs (approximately) yield $F(n, r, 3)$ for $r \in \{2, 3\}$ in the proof of the main result in [1], where the variables d_i correspond to densities of overlapping triangle-free graphs.

Corollary 4.3. *Let $r \in \{2, 3\}$. For all $\varepsilon > 0$, there exists $n_0 > 0$ such that the following holds. Let $n > n_0$ be an integer and let $A \subseteq [n]$ be such that $f(A, r) \geqslant f(n, r) \cdot 2^{-\varepsilon n/(200r)}$. Then either $|A \triangle \mathrm{O}| \leqslant \varepsilon n$ or $|A \triangle I_2| \leqslant \varepsilon n$.*

Proof. Choose $n_0 > 0$ sufficiently large so that the conclusion of Theorem 3.3 holds with parameter $\varepsilon/(200r)$ and the conclusion of Lemma 4.1 holds with parameter $\varepsilon/2$. Let $n \geqslant n_0$ be an integer and let $A \subseteq [n]$ be such that $f(A, r) \geqslant f(n, r) \cdot 2^{-\varepsilon n/(200r)}$. From Theorem 3.3 obtain $A' \subseteq A$ with $|A'| \geqslant |A| - \varepsilon n/(200r)$. Then

$$g(A', r) \overset{(3.1)}{\geqslant} \frac{\log f(A, r)}{n} - \frac{\varepsilon}{200r} \geqslant \frac{\log f(n, r)}{n} - \frac{\varepsilon}{100r} \overset{(3.2)}{\geqslant} g(n, r) - \frac{\varepsilon}{100r}.$$

Lemma 4.1 implies that either $|A' \triangle \mathrm{O}| \leqslant \varepsilon n/2$ or $|A' \triangle I_2| \leqslant \varepsilon n/2$. The result follows. □

The first assertion $f(n, 3) = 3^{n/2+o(n)}$ of Theorem 1.3 follows immediately from Corollary 4.3.

5 The proof of Theorem 1.2

In this section we use Corollary 4.3 to prove Theorem 1.2. Before starting the proof, we need the following useful notion.

Definition 5.1. Given sets $S, B \subseteq [n]$, define the *link graph of B generated by S*, denoted $L_S[B]$, as follows. We set $V(L_S[B]) := B$, and given $x, y \in B$, we have $xy \in E(L_S[B])$ if and only if there is some $z \in S$ such that $\{x, y, z\}$ is a Schur triple, and x, y, z are distinct. Note that $L_S[B]$ is a simple graph (i.e. it does not contain loops). If $S = \{v\}$, then we use the shorthand $L_v[B] := L_{\{v\}}[B]$.

This notion is useful since each edge in a link graph represents some restriction of valid colourings, and so a large set of independent edges in a link graph limits the number of valid colourings.

Lemma 5.2. *Let $n, r \in \mathbb{N}$ and let $A \subseteq [n]$. Let $x \in A$ and suppose that the link graph $L_x[A \setminus \{x\}]$ contains a matching M_x of size m. Then*

$$f(A, r) \leqslant (r^2 - 1)^m \cdot r^{|A|-2m}.$$

Proof. We will bound $f(A, r)$ by re-constructing valid colourings using the following procedure, in which every choice yields a colouring $\tau : A \to [r]$ (which may or may not be valid).

    (1) Let $\tau(x) \in [r]$ be arbitrary.

    (2) For each edge uv in M_x, choose $(\tau(u), \tau(v)) \in [r]^2 \setminus \{(\tau(x), \tau(x))\}$.

    (3) For each uncoloured $y \in A$, let $\tau(y) \in [r]$ be arbitrary.

To see that the procedure generates every valid r-colouring of A, we just need to check that every valid r-colouring σ has the property in Step 2 (since all other choices were arbitrary). That is, for all $uv \in M_x$, σ does not assign u, v, x the same colour. But this is clear since $\{u, v, x\}$ is a Schur triple.

Therefore the number of colourings generated by the procedure is an upper bound for $f(A, r)$. Thus, using the fact that $|A \setminus (\{x\} \cup V(M_x))| = |A| - 2m - 1$, we have

$$f(A, r) \leqslant r \cdot (r^2 - 1)^m \cdot r^{|A|-2m-1},$$

as required. $\qquad\square$

We are now ready to prove Theorem 1.2.

Proof of Theorem 1.2. First recall that $2^{\lceil n/2 \rceil}$ is a lower bound for $f(n,2)$ for all positive integers n, since we can always find a sum-free subset of $[n]$ of size $\lceil n/2 \rceil$, namely O or I_2, and colour it arbitrarily. Let $0 < \varepsilon < 1/200$ and apply Corollary 4.3 to obtain n_0. Now fix an integer $n > n_0$, and let $A \subseteq [n]$ be such that

$$ f(A,2) = f(n,2) \geqslant 2^{\lceil n/2 \rceil}. $$

By Corollary 4.3, we have that either $|A \bigtriangleup \mathrm{O}| \leqslant \varepsilon n$; or $|A \bigtriangleup I_2| \leqslant \varepsilon n$. Thus $|A| \leqslant \lceil n/2 \rceil + \varepsilon n$. We will use the following claim throughout the proof.

Claim 5.3. *For all $x \in A$, every matching in $L_x[A \setminus \{x\}]$ has size less than $3\varepsilon n$.*

Indeed, if not, then there exists a matching of size exactly $3\varepsilon n$. Lemma 5.2 implies that

$$ f(A,2) \leqslant 3^{3\varepsilon n} \cdot 2^{|A| - 6\varepsilon n} \leqslant 3^{3\varepsilon n} \cdot 2^{\left\lceil \frac{n}{2} \right\rceil + \varepsilon n - 6\varepsilon n} \leqslant 2^{\left\lceil \frac{n}{2} \right\rceil - \frac{\varepsilon n}{5}}, $$

a contradiction.

To find large matchings in link graphs in the next two claims, we will use the fact that a graph G with e edges and maximum degree Δ contains a matching of size at least $e/(\Delta+1)$. This is an immediate consequence of Vizing's theorem on edge-colourings.

Case 1. $|A \bigtriangleup \mathrm{O}| \leqslant \varepsilon n$.

In this case, we will prove that $A = \mathrm{O}$. It suffices to show that $A \cap E = \emptyset$, since for any such set we have $f(A,2) = 2^{|A|}$, and thus $A = \mathrm{O}$ is clearly the unique extremal subset among such sets. Arguing by contradiction, the following claim together with Claim 5.3 will complete the proof of this case.

Claim 5.4. *Suppose that there exists $x \in A \cap E$. Then $L_x[A \setminus \{x\}]$ contains a matching M_x of size at least $9\varepsilon n$.*

<u>Proof:</u> Since x is even, $A \cap \mathrm{O} \subseteq A \setminus \{x\}$, and thus $L_x[A \cap \mathrm{O}] \subseteq L_x[A \setminus \{x\}]$. Also, since $|A \bigtriangleup \mathrm{O}| \leqslant \varepsilon n$, it suffices to show that $L_x[\mathrm{O}]$ contains a matching of size at least $10\varepsilon n$. First, we will assume that $x \leqslant n - 60\varepsilon n$. In $L_x[\mathrm{O}]$, every odd number y has degree at least one, since either $x + y \leqslant n$, or $y - x \geqslant 1$ (or both). Moreover, y has degree two if both hold, i.e. when $x + 1 \leqslant y \leqslant n - x$, and there are at least $\lceil (n - 2x)/2 \rceil$ such y (note that there could be vertices less than x which have degree two). Therefore $e(L_x[\mathrm{O}]) \geqslant \frac{1}{2}(x + (n - 2x))$. By Vizing's theorem, $L_x[\mathrm{O}]$ contains a matching of size at least

$$ \frac{e\left(L_x[\mathrm{O}]\right)}{\Delta\left(L_x[\mathrm{O}]\right) + 1} \geqslant \frac{n - x}{6} \geqslant 10\varepsilon n, $$

as required. Now, if $x > n - 60\varepsilon n$, then $M = \{\{1, x - 1\}, \{3, x - 3\}, \dots, \{x/2 - 1, x - (x/2 - 1)\}\} \subseteq L_x[\mathrm{O}]$ is a matching of size $\lfloor x/4 \rfloor \geqslant n/4 - 20\varepsilon n \geqslant 10\varepsilon n$. ∎

Case 2. $|A \triangle I_2| \leqslant \varepsilon n$.

Now A has few elements from I_1. We first show that the smallest element in A cannot be far from $n/2$. For all $i \in \mathbb{N}$, denote by δ_i the ith smallest element in A (denoted so in analogy with the minimum degree of a graph). We may suppose that $\delta_1 \in I_1$ since otherwise $A \subseteq I_2$ is sum-free and we are done.

Claim 5.5. $\delta_1 > n/2 - 12\varepsilon n$.

<u>Proof:</u> Suppose not. Similar to Case 1, since $\delta_1 \in I_1$, we have that $A \cap I_2 \subseteq A \setminus \{\delta_1\}$. Therefore, it suffices to show that the link graph $L_{\delta_1}[A \cap I_2]$ contains a matching of size at least $3\varepsilon n$ which contradicts Claim 5.3. Since $|A \triangle I_2| \leqslant \varepsilon n$, we only need to show that $L_{\delta_1}[I_2]$ contains a matching of size at least $4\varepsilon n$. In $L_{\delta_1}[I_2]$, every element in $[\lfloor n/2 \rfloor + 1, \lfloor n/2 \rfloor + \delta_1] \cup [n - \delta_1 + 1, n]$ has degree one, and all the other $\lceil n/2 \rceil - 2\delta_1$ elements have degree two. Therefore $e(L_{\delta_1}[I_2]) \geqslant \frac{1}{2}(2\delta_1 + 2(\lceil n/2 \rceil - 2\delta_1)) = \lceil n/2 \rceil - \delta_1$. Therefore, by Vizing's theorem, $L_{\delta_1}[I_2]$ contains a matching of size at least

$$\frac{e\left(L_{\delta_1}[I_2]\right)}{\Delta\left(L_{\delta_1}[I_2]\right) + 1} \geqslant \frac{\left\lceil \frac{n}{2} \right\rceil - \delta_1}{3} \geqslant \frac{n}{6} - \frac{\delta_1}{3} \geqslant 4\varepsilon n,$$

a contradiction. ∎

Define k such that $A \cap I_1 = \{\delta_1, \ldots, \delta_k\}$. Then $B := 2 \cdot (A \cap I_1)$ is a subset of I_2 by Claim 5.5. There are two ways of colouring each of the k pairs $\{\delta_i, 2\delta_i\}$, so

$$f(A, 2) \leqslant 2^k \cdot 2^{|(A \cap I_2) \setminus B|} \leqslant 2^k \cdot 2^{|I_2| - k} = 2^{\lceil n/2 \rceil}$$

with equality in the second inequality if and only if $|(A \cap I_2) \setminus B| = |I_2| - k$, i.e. if and only if $I_2 \setminus A \subseteq B$.

Claim 5.6. $\delta_1 + \delta_2 \geqslant n + 1$.

<u>Proof:</u> Suppose that $\delta_1 + \delta_2 \leqslant n$. Then $\delta_1 \in I_1$ (but δ_2 may or may not be). Since $\delta_1 + \delta_2 \notin B$, we have $\delta_1 + \delta_2 \in A$. Let $C := \{\delta_1, \delta_2, 2\delta_1, \delta_1 + \delta_2, 2\delta_2\}$. Then $\{\delta_1, \delta_2, \delta_1 + \delta_2\} \subseteq A \cap C \subseteq C$, and one can easily check that there are at most $6 = 2^4 - 2$ valid colourings of $A \cap C$, the missing colourings being the two monochromatic ones.

Suppose first that $\delta_2 \in I_1$. Then $I_2 \cap C = \{2\delta_1, \delta_1 + \delta_2, 2\delta_2\}$. So

$$f(A, 2) \leqslant 2^{k-2} \cdot 2^{|I_2 \setminus C| - (k-2)} \cdot 6 = 6 \cdot 2^{|I_2| - 3} = \frac{3}{4} \cdot 2^{\lceil n/2 \rceil},$$

a contradiction. Suppose instead that $\delta_2 \in I_2$. Then $k = 1$ and $I_2 \cap C = \{\delta_2, 2\delta_1, \delta_1 + \delta_2\}$, and

$$f(A, 2) \leqslant 2^{|I_2 \setminus C|} \cdot 6 = 6 \cdot 2^{|I_2| - 3} = \frac{3}{4} \cdot 2^{\lceil n/2 \rceil}.$$

Thus $\delta_1 + \delta_2 \geqslant n + 1$, as required. ∎

So certainly $|A \cap I_1| \leqslant 1$, and $I_2 \setminus A \subseteq \{2\delta_1\}$, but we can say more. If $A \cap I_1 = \emptyset$ then $A = I_2$ and we are done. So we may assume that $A \cap I_1 = \{\delta_1\}$. Suppose $\delta_2 \neq \lfloor n/2 \rfloor + 1$.

Then $\lfloor n/2 \rfloor + 1$ is missing from A and hence $2\delta_1 = \lfloor n/2 \rfloor + 1$, a contradiction to Claim 5.5. Thus $\delta_2 = \lfloor n/2 \rfloor + 1$ and therefore

$$n + 1 \leqslant \delta_1 + \delta_2 \leqslant 2\lfloor n/2 \rfloor + 1.$$

So n is even and we have equality if and only if $(\delta_1, \delta_2) = (n/2, n/2+1)$. Since $I_2 \backslash A \subseteq \{n\}$, there are two candidates for extremal sets: $[n/2, n-1]$ and $[n/2, n]$. It is easy to see that both have $2^{n/2}$ valid 2-colourings. This completes the proof of Theorem 1.2. $\qquad\square$

6 The proof of Theorem 1.4

To prove Theorem 1.4, by Theorem 3.3, it suffices to determine $g(n, r)$ asymptotically. Our aim is to show that

$$g(n, 4) = 1 + o(1) \quad \text{and} \quad g(n, 5) = \frac{1}{4} \log 30 + o(1).$$

Recall from the proof of Lemma 4.1 that we reduced the problem of asymptotically determining $g(n, 3)$ to solving a *linear* program. Indeed, we let (A_1, A_2, A_3) be a tuple of sum-free subsets with intersection vector (d_1, d_2, d_3), and found a reduction of Problem 2 into a linear program in variables d_1, d_2, d_3. Our task was then to maximise $\sum_{i \in [3]} d_i \log i$ subject to $d_1, d_2, d_3 \geqslant 0$ and $\sum_{i \in [3]} i d_i \leqslant \frac{3}{n} \lceil \frac{n}{2} \rceil$.

So to prove Theorem 1.4, we will again reduce Problem 2 to a linear program in variables $d_1, \ldots, d_r$ and an additional slack variable a (defined below) for $r \in \{4, 5\}$.

Given $n \in \mathbb{N}$, let $A_1, \ldots, A_r \in [n]$ be maximal sum-free sets. Throughout the rest of this section, define $D_1, \ldots, D_r, d_1, \ldots, d_r$ as in Problem 2. So D_i is the set of all elements that are in exactly i different A_j's, and $d_i := |D_i|/n$. Further, we let C be the subset of $[r]$ such that A_i is a type (a) set for every $i \in C$, and a type (b) or (c) set, for every $i \in [r] \backslash C$. Recall that a type (a) set S has $|S| \leqslant 2n/5 + 1$, a type (b) set consists of odd numbers, and a type (c) set S has $|S| \leqslant \min(S)$. We define

$$a := \frac{1}{n} \sum_{i \in [r] \backslash C} \left(\left\lceil \frac{n}{2} \right\rceil - |A_i| \right) = \frac{(r - |C|)}{n} \cdot \left\lceil \frac{n}{2} \right\rceil - \frac{1}{n} \sum_{i \in [r] \backslash C} |A_i|. \tag{6.1}$$

An important observation is that, if A is a maximal sum-free set of type (b), then A is precisely O, the set of odd integers in $[n]$. Thus

$$a = \frac{1}{n} \sum_{i : A_i \text{ of type (c)}} \left(\left\lceil \frac{n}{2} \right\rceil - |A_i| \right).$$

Definition 6.1. Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Let $0 \leqslant f_1, \ldots, f_r \leqslant 1$ be variables. Suppose that

$$\alpha_{i1} f_1 + \ldots + \alpha_{ir} f_r + \alpha_i a \leqslant \beta_i \quad \text{for all} \quad i \leqslant N(r),$$

where $N(r)$ is a positive integer depending only on r; and $\alpha_{ij}, \beta_i, \alpha_i \in \mathbb{R}$ for all $i \leqslant N(r)$ and $j \leqslant r$. We say that this set $\mathcal{C}$ of inequalities is (ε, r)-*sufficient* if

$$\max \sum_{i \in [r]} f_i \log i \quad \text{subject to} \quad \mathcal{C} \quad \text{is at most} \quad \max\left\{\frac{1}{2}\log r, \frac{1}{4}\log(r\lfloor r^2/4\rfloor)\right\} + \varepsilon.$$

When $(d_1, \ldots, d_r)$ playing the roles of $(f_1, \ldots, f_r)$ satisfies $\mathcal{C}$, we say that $\mathcal{C}$ is a *family of constraints for* $A_1, \ldots, A_r$.

We will always use d_i instead of f_i since we are only interested in $\mathcal{C}$ which are a family of constraints for $A_1, \ldots, A_r$. However, we would like these constraints to be independent of n, which of course $A_1, \ldots, A_r$ are not. As an example, when $r = 3$, we showed that, for every $\varepsilon > 0$, sufficiently large $n \in \mathbb{N}$ and sum-free subsets A_1, A_2, A_3 of $[n]$, the family

$$\left\{d_1 \geqslant 0; d_2 \geqslant 0; d_3 \geqslant 0; d_1 + 2d_2 + 3d_3 \leqslant \frac{3}{2} + \frac{\varepsilon}{300}\right\}$$

of constraints is $(\varepsilon, 3)$-sufficient. Indeed, the final constraint we actually derived was $d_1 + 2d_2 + 3d_3 \leqslant 3/n \cdot \lceil n/2 \rceil$, but we can replace the right-hand side by the weaker $3/2 + \varepsilon/300$ to obtain a family which is still a family of constraints for A_1, A_2, A_3, but is independent of n.

The proof of Theorem 1.4 follows from the next lemma.

Lemma 6.2. *Let* $r \in \{4, 5\}$. *For all* $\varepsilon > 0$, *there exists an* $n_0 > 0$ *such that for all integers* $n \geqslant n_0$, *every choice of maximal sum-free subsets* $A_1, \ldots, A_r$ *of* $[n]$ *has a family of* (ε, r)-*sufficient constraints.*

Indeed, suppose that the lemma holds. The construction after the statement of Theorem 1.4 shows that, whenever n is a sufficiently large integer, we have $f(n, 4) \geqslant 16^{\lfloor n/4 \rfloor} > 2^{(1-\varepsilon)n}$. Thus it suffices to find $n_0 > 0$ such that $f(n, 4) \leqslant 2^{(1+\varepsilon)n}$ whenever $n \geqslant n_0$ is an integer. Choose n_0 so that the conclusions of Theorem 3.3 (with $r = 4$) and Lemma 6.2 hold for parameter $\varepsilon/2$ and all $n \geqslant n_0$. Now let $n \geqslant n_0$ be an arbitrary integer. By Remark 3.2, there are maximal sum-free subsets $A_1, \ldots, A_4$ of $[n]$ such that $g(A_1, \ldots, A_4) = g(n, 4)$. By Lemma 6.2, $A_1, \ldots, A_4$ has a family of $(\varepsilon/2, 4)$-sufficient constraints. Thus $g(n, 4) = g(A_1, \ldots, A_4) = \sum_{i \in [4]} d_i \log i \leqslant 1 + \varepsilon/2$. Then

$$f(n, 4) \overset{(3.2)}{\leqslant} 2^{(g(n,4)+\varepsilon/2)n} \leqslant 2^{(1+\varepsilon)n},$$

as required. The case $r = 5$ is almost identical.

For each $r = 4, 5$, we split the proof of Lemma 6.2 into cases depending on the structure of $A_1, \ldots, A_r$ (obtained from Theorem 2.1). Then, in each case, we find a family of constraints which is (ε, r)-sufficient. Given a family $\mathcal{C}$ of inequalities, we must

1. show that it is a family of constraints for $A_1, \ldots, A_r$, i.e. that each inequality holds; then

2. show that it is (ε, r)-sufficient, i.e. consider the linear program $\max \sum_{i \in [r]} d_i \log i$ subject to $\mathcal{C}$, and show that its optimal solution is at most the required value.

Since it is only a series of tedious calculations, we defer the details of (2) to the appendix, and limit ourselves to some remarks here.

6.1 Achieving (2): Solving linear programs

Since there are many cases (depending on the structure of $A_1, \ldots, A_r$), and sometimes rather a lot of inequalities in each family of constraints $\mathcal{C}$, where possible, we use Mathematica to solve the resulting linear program $\max_{i \in [r]} d_i \log i$ subject to $\mathcal{C}$. Suppose that $r = 4$ (when $r = 5$ the situation is similar). Given $A_1, \ldots, A_4$, an $(\varepsilon, 4)$-sufficient family $\mathcal{C}$ is such that $\sum_{i \in [4]} d_i \log i \leqslant 1 + \varepsilon$.

There are two cases, depending on whether the 4-tuple (A_1, A_2, A_3, A_4) is close to extremal or not, namely if $g(A_1, A_2, A_3, A_4)$ is close to the maximum possible value or not. Suppose that there is some specific value of ε, say $\varepsilon = 1/1000$, and a family $\mathcal{C}$ of constraints which is $(1/1000, 4)$-sufficient, for which Mathematica shows the output $\sum_{i \in [4]} d_i \log i \leqslant 0.999$. The level of accuracy of the program is enough for us to know that certainly $g(A_1, \ldots, A_4) = \sum_{i \in [4]} d_i \log i < 1 - 1/2000$. So in this case, we are done, and in fact since this number is less than our lower bound (3.6) by some absolute constant, we see that $A_1, \ldots, A_4$ cannot be close to extremal.

If instead, given input $\varepsilon = 1/1000$, Mathematica shows an output

$$0.999 < \sum_{i \in [4]} d_i \log i \leqslant 1.001,$$

say, we need to be more careful. In this case, we will write out the dual program of $\max \sum_{i \in [4]} d_i \log i$ subject to $\mathcal{C}$, which is a minimisation problem. We then exhibit a feasible solution to the dual which is at most $1 + \varepsilon$. By the weak duality theorem, we see that $\max \sum_{i \in [4]} d_i \log i \leqslant 1 + \varepsilon$, as required.

6.2 Linear constraints for general r

To achieve (1), we will first derive a set of linear constraints which apply for any number r of colours.

Lemma 6.3. *For all $\varepsilon > 0$ and integers $r \geqslant 4$, there exists an $n_0 > 0$ such that the following holds. Let $n, s, t \in \mathbb{N}$ be such that $n \geqslant n_0$ and $s + t \leqslant r$. Also, let $A_1, \ldots, A_r$ be maximal sum-free subsets of $[n]$ such that s of them are of type (b), t of them are of type (c), and $(d_1, \ldots, d_r)$ is their intersection vector. Then*

$$\sum_{i \in [r]} i d_i \leqslant \frac{r}{2} - \frac{r - s - t}{10} - a + \varepsilon \quad and \quad d_{r-1} + d_r \leqslant \frac{1}{2} + \varepsilon.$$

Proof. Let $n_0 := r/\varepsilon$ and let $n \geqslant n_0$ be an integer. For the first inequality, by (6.1) and the definition of type (a) sets, we have

$$\sum_{i \in [r]} i d_i \leqslant \frac{s+t}{n} \cdot \left\lceil \frac{n}{2} \right\rceil + \left(\frac{2}{5} + \frac{1}{n} \right)(r - s - t) - a \leqslant \frac{r}{2} - \frac{r - s - t}{10} - a + \varepsilon,$$

where the last inequality follows from $n \geqslant r/\varepsilon$. To prove the second part of the lemma, since $\varepsilon \geqslant r/n \geqslant 1/n$, it suffices to show that the set $D_r \cup D_{r-1}$ is sum-free. Assume to the contrary that there exist $x_1, x_2, x_3 \in D_r \cup D_{r-1}$ such that $x_1 + x_2 = x_3$. For every $i \in [3]$, define $I_i := \{j : j \in [r] \text{ and } x_i \in A_j\}$ and $\overline{I_i} = [r] \setminus I_i$. Since $x_i \in D_r \cup D_{r-1}$, we have $|\overline{I_i}| \leqslant 1$ for every $i \in [3]$, and therefore $\left| \bigcap_{i \in [3]} I_i \right| \geqslant r - \sum_{i \in [3]} |\overline{I_i}| \geqslant r - 3 \geqslant 1$, where the last inequality follows from $r \geqslant 4$. Therefore, there exists an $i \in [r]$ such that $x_1, x_2, x_3 \in A_i$, which contradicts A_i being sum-free. $\qquad\square$

We will use the next two simple facts repeatedly. Recall that $d(A) = |A|/n$ is the density of a set A.

Observation 6.4. *Let $k \in [r]$ and suppose that $A_1, \ldots, A_k$ are of type (c). Then*

(i) $\min(A_i) \geqslant \lceil n/2 \rceil - an$ *for all $i \in [k]$.*

(ii) $\sum_{i \in [k]} d(A_i) \geqslant k \cdot \frac{\lceil n/2 \rceil}{n} - a$.

Proof. The second equality for parameter a just after equation (6.1) states that an is the sum of $\lceil n/2 \rceil - |A_i|$ over $i \in [k]$. Thus, for (i), using the definition of type (c) sets, every $i \in [k]$ satisfies $\min(A_i) \geqslant |A_i| \geqslant \lceil n/2 \rceil - an$, and for (ii), $|A_1| + \ldots + |A_k| = k\lceil n/2 \rceil - an$. $\qquad\square$

The next lemma concerns the size of the intersection of type (c) sets.

Lemma 6.5. *For all $\varepsilon > 0$ and positive integers r, there exists an $n_0 > 0$, such that the following holds. For every integer $n \geqslant n_0$, let $A_1, \ldots, A_r$ be maximal sum-free subsets of $[n]$ such that there is some $k \leqslant r$ for which $A_1, \ldots, A_k \subseteq [n]$ are type (c) sets. Then*

$$d(\cap_{i \in [k]} A_i) \geqslant \frac{1}{2} - ka - \varepsilon.$$

Proof. Let $n_0 := k/\varepsilon$ and let $n > n_0$ be an integer and $A_1, \ldots, A_r$ be subsets of $[n]$ as in the statement. Let $A^* = \cap_{i \in [k]} A_i$. It suffices to show that $|A^*| \geqslant \lceil n/2 \rceil - (k-1) - kan$. By Observation 6.4(i), for all $i \in [k]$, $\min(A_i) \geqslant \lceil n/2 \rceil - an$, and therefore $A_i \subseteq I_a := [\lceil n/2 \rceil - an, n]$. Every integer x in the subset $L := I_a \setminus A^*$ lies in at most $k - 1$ of the A_i's. Assume to the contrary that $|A^*| < \lceil n/2 \rceil - (k-1) - kan$. Then

$$\sum_{i \in [k]} |A_i| \leqslant k \cdot |A^*| + (k-1)|L| = (k-1)|I_a| + |A^*|$$

$$< (k-1)(n - (\lceil n/2 \rceil - an) + 1) + \lceil n/2 \rceil - (k-1) - kan$$

$$= (k-1)n - (k-2)\lceil n/2 \rceil - an. \tag{6.2}$$

But Observation 6.4(ii) implies that $\sum_{i \in [k]} |A_i| \geqslant k \lceil n/2 \rceil - an$. Together with (6.2), we get $2(k-1) \lceil n/2 \rceil < (k-1)n$, a contradiction. $\qquad \square$

Throughout the rest of the paper, given a defined by (6.1), we will let

$$J_1 := \left[1, \left\lfloor \frac{n}{2} \right\rfloor - an \right], \quad J_2 := \left[\left\lfloor \frac{n}{2} \right\rfloor - an + 1, n \right], \quad \text{and} \quad J_3 := \left[\left\lfloor \frac{n}{2} \right\rfloor - an + 1, \left\lfloor \frac{n}{2} \right\rfloor \right]$$
(6.3)

and will refer to J_1, J_2, and J_3 as the *first*, *second*, and *middle interval* respectively. Note that by definition an is an integer, and the set $J_2 \backslash J_3 = I_2$ is a sum-free set of maximum size $\lceil n/2 \rceil$. The following observation is a straightforward consequence of Observation 6.4(i) and the fact that the unique maximal sum-free subset of type (b) is O.

Observation 6.6. (i) *If A_i is a maximal sum-free set of type (c), then $|A_i \setminus J_2| \leqslant 1$.*

(ii) *For all $\varepsilon > 0$ and positive integers r, there exists $n_0 > 0$ such that the following holds. Let $A_1, \ldots, A_r$ be maximal sum-free subsets of $[n]$ such that s of them are of type (b), t of them are of type (c), and $r - s - t$ of them are of type (a). Then*

$$d\left((D_0 \cup \ldots \cup D_{r-s-t}) \cap J_1 \cap E \right), d\left((D_s \cup \ldots \cup D_{r-t}) \cap J_1 \cap O \right) \geqslant 1/4 - a/2 - \varepsilon.$$

The final result in this subsection states some constraints involving the intervals J_2 and J_3.

Lemma 6.7. *For all $\varepsilon > 0$ and positive integers r, there exists $n_0 > 0$ such that the following holds. Let $k \leqslant r$ and $n \geqslant n_0$ be positive integers. Let $A_1, \ldots, A_k \subseteq [n]$ be type (c) maximal sum-free sets and define $q_i := d(A_i \cap J_3)$ for all $i \in [k]$. Then*

$$\sum_{i \in [k]} q_i \leqslant a + \varepsilon \quad \text{and} \quad d\left((\cap_{i \in [k]} A_i) \cap I_2 \right) \geqslant \frac{1}{2} - \sum_{i \in [k]} q_i - a - \varepsilon \geqslant \frac{1}{2} - 2a - 2\varepsilon.$$

Proof. Let $n_0 := r/\varepsilon$ and let $n \geqslant n_0$ be an integer. Since A_i has $q_i n$ elements in J_3, we have that $\min(A_i) \leqslant \left\lfloor \frac{n}{2} \right\rfloor + 1 - q_i n$ for all $i \in [k]$. Using Observation 6.4(ii) and the definition of type (c) sets, we have

$$\frac{k}{n} \left\lceil \frac{n}{2} \right\rceil - a \leqslant \sum_{i \in [k]} d(A_i) = \sum_{i \in [k]} \frac{|A_i|}{n} \leqslant \sum_{i \in [k]} \frac{\min(A_i)}{n} \leqslant \frac{k}{n} \left(\left\lceil \frac{n}{2} \right\rceil + 1 \right) - \sum_{i \in [k]} q_i. \qquad (6.4)$$

Thus $\sum_{i \in [k]} q_i \leqslant k/n + a \leqslant a + \varepsilon$, proving the first inequality. To prove (ii), let $B := I_2 \setminus \cap_{i \in [k]} A_i$. In other words, B is the set of all elements in $I_2 = J_2 \setminus J_3$ that are missing from at least one of the $A_1, \ldots, A_k$. By Observation 6.6(i),

$$\sum_{i \in [k]} d(A_i) \leqslant \sum_{i \in [k]} q_i + \sum_{i \in [k]} d(A_i \cap (J_2 \setminus J_3)) + \frac{k}{n} \leqslant \sum_{i \in [k]} q_i + \frac{k}{n} \lceil n/2 \rceil - d(B) + \varepsilon. \qquad (6.5)$$

Thus

$$d(B) \leqslant \sum_{i \in [k]} q_i + \frac{k}{n} \lceil n/2 \rceil - \sum_{i \in [k]} d(A_i) + \varepsilon \overset{(6.4)}{\leqslant} \sum_{i \in [k]} q_i + a + \varepsilon,$$

and the second required inequality follows. $\qquad \square$

6.3 The 4 colour case

Proof of Lemma 6.2 when $r = 4$. Let $\varepsilon > 0$ and choose $\varepsilon', n_0 > 0$ such that $1/n_0 \ll \varepsilon' \ll \varepsilon \leqslant 1/100$. (we may assume the last inequality without loss of generality). Let $n \geqslant n_0$ be an integer and let $A_1, \ldots, A_4 \subseteq [n]$ be maximal sum-free sets with intersection vector $(d_1, \ldots, d_4)$ as defined in Problem 2. We need to obtain a family of $(\varepsilon, 4)$-sufficient constraints. We have the following set of basic constraints which will be used throughout the proof:

$$\begin{cases} d_i \geqslant 0 \text{ for all } i \in \{0, \ldots, 4\}, \\ \displaystyle\sum_{i \in \{0,\ldots,4\}} d_i \leqslant 1, \\ d_3 + d_4 \leqslant \left\lceil \frac{n}{2} \right\rceil / n \leqslant \frac{1}{2} + \varepsilon', \end{cases} \tag{C1}$$

where the last inequality follows from Lemma 6.3.

Suppose first that there is an $i \in [4]$ for which A_i is of type (a). By Lemma 6.3, we have that

$$\sum_{i \in [4]} i d_i \leqslant 19/10 + \varepsilon'. \tag{C0*}$$

(The * denotes the fact that this inequality does not hold in Case 1 onwards.) The family $\{(\mathrm{C0}^*), (\mathrm{C1})\}$ is $(\varepsilon, 4)$-sufficient (for justification here and from now on, see the appendix).

Thus we can assume that all of $A_1, \ldots, A_4$ are of type (b) or (c). Let s be the number of A_i of type (b) and t the number of A_i of type (c) (so $s + t = 4$). Define a as in (6.1). By Lemma 6.3 (with ε' playing the role of ε), we have

$$\sum_{i \in [4]} i d_i \leqslant 2 - a + \varepsilon'. \tag{C2}$$

Suppose that $a \geqslant 1/10$. Then (C2) implies that (C0*) holds. But, as we have seen, $\{(\mathrm{C0}^*), (\mathrm{C1})\}$ is an $(\varepsilon, 4)$-sufficient family. Hence, throughout the rest of the proof, we can assume that

$$a < 1/10. \tag{C3}$$

Given a, define J_1, J_2, J_3 as in (6.3). By Observation 6.6(ii),

$$\left. \begin{array}{l} d_0 \geqslant d(D_0 \cap J_1) \\ d_s \geqslant d(D_s \cap J_1) \end{array} \right\} \geqslant \frac{1}{4} - \frac{a}{2} - \varepsilon'. \tag{C4}$$

Case 1. $s \in \{0, 1\}$.

Suppose first that $s = 0$. Then $A_1, \ldots, A_4$ are all type (c), so Lemma 6.5 (with ε' playing the role of ε) implies that

$$d_4 \geqslant \frac{1}{2} - 4a - \varepsilon'. \tag{C5}$$

The family $\{(C1), (C2), (C4), (C5)\}$ is $(\varepsilon, 4)$-sufficient. (We could also include (C3) but it is not needed.)

Suppose instead that $s = 1$. Without loss of generality, let A_4 be the only type (b) set. By Lemma 6.5 (with ε' playing the role of ε), $d(\cap_{i \in [3]} A_i) \geqslant 1/2 - 3a - \varepsilon'$. By Observation 6.6(i), $|\cap_{i \in [3]} A_i \setminus J_2| \leqslant 1$. Also, every element $x \in \cap_{i \in [3]} A_i$ is in D_3 if it is even, and D_4 if it is odd. Thus

$$d_3 \geqslant d((\cap_{i \in [3]} A_i) \cap J_2) - d(O \cap J_2) \geqslant d(\cap_{i \in [3]} A_i) - \frac{1}{n} - \frac{1}{n} \cdot \left\lceil (n - \left\lfloor \frac{n}{2} \right\rfloor + an)/2 \right\rceil \quad \text{(C6)}$$

$$\geqslant \frac{1}{2} - 3a - \varepsilon' - \frac{1}{n} - \frac{1}{4} - \frac{a}{2} - \frac{1}{2n} \geqslant \frac{1}{4} - \frac{7a}{2} - 2\varepsilon'.$$

The family $\{(C1), (C2), (C3), (C4), (C6)\}$ is $(\varepsilon, 4)$-sufficient. This completes the proof of Case 1.

Case 2. $t \in \{0, 1\}$.

Suppose first that $t = 0$. Then all of $A_1, \ldots, A_4$ are of type (b), so D_0 is the set of evens, D_4 is the set of odds, and all the other D_i's are empty. Therefore

$$g(A_1, \ldots, A_4) = 2 \cdot d_4 = \frac{2}{n} \cdot \left\lceil \frac{n}{2} \right\rceil \leqslant 1 + \varepsilon',$$

as required. Suppose instead that $t = 1$. Then $D_0 \cup D_1$ contains every even integer, and therefore

$$d_0 + d_1 \geqslant \frac{1}{n} \cdot \left\lfloor \frac{n}{2} \right\rfloor \geqslant \frac{1}{2} - \varepsilon'. \quad \text{(C7)}$$

The family $\{(C1), (C2), (C3), (C4), (C7)\}$ is $(\varepsilon, 4)$-sufficient. This completes the proof of Case 2. Therefore, the only remaining case is the following.

Case 3. $s = t = 2$.

We will prove that the following constraints hold.

$$d_2 \geqslant \frac{1}{2} - \frac{5a}{2} - 2\varepsilon', \quad \text{(C8)}$$

$$d_3 \leqslant 3a + \varepsilon'. \quad \text{(C9)}$$

Without loss of generality, we can assume that A_1 and A_2 are of type (c). We first prove (C8). By (C4), we have that $d(D_2 \cap J_1) \geqslant 1/4 - a/2 - \varepsilon'$. Therefore, we only need to show that $d(D_2 \cap J_2) \geqslant 1/4 - 2a - \varepsilon'$. Let $B := A_1 \cap A_2 \cap (J_2 \setminus J_3)$. Then $B \cap E \subseteq D_2 \cap J_2$. Lemma 6.7 applied with parameter $\varepsilon'/4$ implies that $d(B) \geqslant 1/2 - 2a - \varepsilon'/2$. Further, $J_2 \setminus J_3$ is an interval of length $\lceil n/2 \rceil$ so contains at most $n/4 + 1$ odd elements. Thus

$$d(D_2 \cap J_2) \geqslant d(B \cap E) = d(B) - d(B \cap O) \geqslant \frac{1}{2} - 2a - \frac{\varepsilon'}{2} - d((J_2 \setminus J_3) \cap O) \geqslant \frac{1}{4} - 2a - \varepsilon'.$$

We now prove (C9). By Observation 6.6(i), we have $|A_1 \cap J_1|, |A_2 \cap J_1| \leqslant 1$. Thus $|D_3 \cap J_1| \leqslant 2$. Further, we have $A_1 \cap A_2 \cap E \subseteq D_2$ and $A_1 \cap A_2 \cap O \subseteq D_4$. In particular, $(A_1 \cap A_2) \cap D_3 = \emptyset$. Therefore

$$d_3 = d(D_3 \cap J_1) + d(D_3 \cap J_2) \leqslant \frac{2}{n} + d(J_2) - d(B) \overset{(6.3)}{\leqslant} \frac{2}{n} + 1 - \frac{1}{n} \cdot \lfloor n/2 \rfloor + a - \frac{1}{2} + 2a + \frac{\varepsilon'}{2} \leqslant 3a + \varepsilon',$$

as required. But $\{(C1), (C2), (C3), (C8), (C9)\}$ is an $(\varepsilon, 4)$-sufficient family. This completes the proof of Case 3, the final case. $\qquad\square$

6.4 The 5 colour case

Proof of Lemma 6.2 in the case $r = 5$. Let $\varepsilon > 0$ and choose $\varepsilon', n_0 > 0$ such that $1/n_0 \ll \varepsilon' \ll \varepsilon \leqslant 1/200$. (we may assume the last inequality without loss of generality). Let $n > n_0$ be an integer and let $A_1, \dots, A_5 \subseteq [n]$ be maximal sum-free sets with intersection vector $(d_1, \dots, d_5)$. We have the following basic constraints which will be used throughout the proof.

$$\begin{cases} d_i \geqslant 0 \text{ for all } i \in \{0, \dots, 5\}, \\ \displaystyle\sum_{i \in \{0,\dots,5\}} d_i \leqslant 1. \end{cases} \tag{D1}$$

Suppose first that A_1, A_2 are of type (a). Now Lemma 6.3 (with ε' playing the role of ε) implies that

$$\sum_{i \in [5]} i d_i \leqslant 23/10 + \varepsilon'. \tag{D0*}$$

(The $*$ denotes the fact that this inequality does not hold in Case 1 onwards.) But $\{(D0^*), (D1)\}$ is an $(\varepsilon, 5)$-sufficient family.

Let s be the number of A_i of type (b) and t the number of A_i of type (c). So we may assume that $s + t \in \{4, 5\}$.

Define a as in (6.1). Lemma 6.3 implies that

$$\sum_{i \in [5]} i d_i \leqslant \begin{cases} 5/2 - a + \varepsilon', & \text{if } s + t = 5, \\ 12/5 - a + \varepsilon', & \text{if } s + t = 4. \end{cases} \tag{D2}$$

Suppose now that $s + t = 5$ and $a \geqslant 1/5$; or $s + t = 4$ and $a \geqslant 1/10$. Then (D2) implies that (D0*) holds. But, as above, $\{(D0^*), (D1)\}$ is an $(\varepsilon, 5)$-sufficient family. Thus we may assume that

$$\text{if } s + t = 5, \text{ then } a < 1/5; \quad \text{and if } s + t = 4, \text{ then } a < 1/10. \tag{D3}$$

By Observation 6.6(ii), we have

$$\left.\begin{array}{l} d_0 \geqslant d(D_0 \cap J_1 \cap E) \\ d_s \geqslant d(D_s \cap J_1 \cap O) \end{array}\right\} \geqslant \frac{1}{4} - \frac{a}{2} - \varepsilon', \quad \text{if } s + t = 5, \tag{6.6}$$

$$\left.\begin{array}{l} d_0 + d_1 \geqslant d((D_0 \cup D_1) \cap J_1 \cap E) \\ d_s + d_{s+1} \geqslant d((D_s \cup D_{s+1}) \cap J_1 \cap O) \end{array}\right\} \geqslant \frac{1}{4} - \frac{a}{2} - \varepsilon', \quad \text{if } s + t = 4, \tag{6.7}$$

We work through some cases depending on the values of (s,t), in increasing order of complexity.

Case 1. $5 - s \leqslant 1$ or $5 - t \leqslant 1$.

In this case, we will see that

$$d_0 + d_1 \geqslant 1/2 - a - \varepsilon'. \tag{D4}$$

Suppose first that $5 - s \leqslant 1$. Now at most one A_i can contain an even number, so $d_0 + d_1 \geqslant |E|/n \geqslant 1/2 - \varepsilon' \geqslant 1/2 - a - \varepsilon'$, as required. Suppose secondly that $5 - t \leqslant 1$. Then summing the inequalities in each of (6.6) and (6.7) implies that $d_0 + d_1 \geqslant 1/2 - a - \varepsilon'$. But $\{(D1), (D2), (D3), (D4)\}$ is an $(\varepsilon, 5)$-sufficient family.

The remaining cases are $(s, t) \in \{(1, 3), (3, 1), (2, 3), (3, 2), (2, 2)\}$.

Case 2. $(s, t) \in \{(1, 3), (3, 1)\}$.

Suppose that A_5 is the only set of type (a). By (6.7), we have that

$$d_0 + d_1 \geqslant 1/4 - a/2 - \varepsilon'. \tag{D5}$$

Suppose first that $(s, t) = (1, 3)$. Then summing the inequalities in (6.7) implies that

$$d_0 + d_1 + d_2 \geqslant 1/2 - a - 2\varepsilon'. \tag{D6}$$

But $\{(D1), (D2), (D3), (D5), (D6)\}$ is an $(\varepsilon, 5)$-sufficient family.

Suppose secondly that $(s, t) = (3, 1)$. Suppose that A_4 is the only set of type (c). We will prove that the following inequalities hold.

$$d_1 + d_2 \geqslant 1/4 - 3a/2 - 3\varepsilon', \tag{D7}$$
$$d_3 \leqslant 1/4 + 3a/2 + 3\varepsilon'. \tag{D8}$$

Note that

$$A_4 \cap E \subseteq D_1 \cup D_2 \quad \text{and} \quad A_4 \cap O \subseteq D_4 \cup D_5. \tag{6.8}$$

Further, using Lemma 6.5 and Observation 6.6(i), we have

$$d(A_4 \cap J_2) = d(A_4) - d(A_4 \setminus J_2) \geqslant (1/2 - a - \varepsilon') - 1/n \geqslant 1/2 - a - 2\varepsilon'. \tag{6.9}$$

Therefore

$$d_1 + d_2 \quad \overset{}{\geqslant} \quad d((D_1 \cup D_2) \cap J_2) \overset{(6.8)}{\geqslant} d(A_4 \cap E \cap J_2) \geqslant d(A_4 \cap J_2) - d(J_2 \cap O)$$
$$\overset{(6.3),(6.9)}{\geqslant} \quad (1/2 - a - 2\varepsilon') - (1/4 + a/2 + \varepsilon') = 1/4 - 3a/2 - 3\varepsilon',$$

proving (D7). To prove (D8),

$$d_3 = d(D_3 \cap J_1) + d(D_3 \cap J_2)$$
$$\leqslant d(J_1) - d((D_0 \cup D_1) \cap J_1) + d(J_2) - d((D_1 \cup D_2 \cup D_4 \cup D_5) \cap J_2)$$
$$\overset{(6.8)}{\leqslant} 1 - d((D_0 \cup D_1) \cap J_1) - d(A_4 \cap J_2) \overset{(6.7),(6.9)}{\leqslant} 1 - (1/4 - a/2 - \varepsilon') - (1/2 - a - 2\varepsilon')$$
$$= 1/4 + 3a/2 + 3\varepsilon',$$

as desired. We have that $\{(D1), (D2), (D3), (D5), (D7), (D8)\}$ is an $(\varepsilon, 5)$-sufficient family.

Case 3. $(s,t) = (2,3)$.

We will assume that A_1 and A_2 are type (b) sets (so $A_1 = A_2 = O$), and A_3, A_4, A_5 are type (c) sets. Let $q_i := d(A_i \cap J_3)$, for every $i \in \{3, 4, 5\}$. First, we will prove that the following constraints hold.

$$d_5 \geqslant 1/4 - 2a - 2\varepsilon', \tag{D9}$$

$$d_3 \leqslant 1/4 + 3a + 4\varepsilon'. \tag{D10}$$

We first prove (D9). By Lemma 6.7, we have

$$\begin{aligned} d_5 &\geqslant d(D_5 \cap I_2) \geqslant d(O \cap (A_3 \cap A_4 \cap A_5) \cap I_2) \\ &\geqslant d((A_3 \cap A_4 \cap A_5) \cap I_2) - d(E \cap I_2) \\ &\geqslant (1/2 - 2a - \varepsilon') - (1/4 + \varepsilon') = 1/4 - 2a - 2\varepsilon', \end{aligned} \tag{6.10}$$

proving (D9). We now prove (D10). Notice that every element of D_3 lies in at least one type (c) set, i.e. $D_3 \subseteq A_3 \cup A_4 \cup A_5$. So $|D_3 \cap J_1| \leqslant \sum_{3 \leqslant i \leqslant 5} |A_3 \cap J_1| = \sum_{3 \leqslant i \leqslant 5} |A_3 \setminus J_2| \leqslant 3$ by Observation 6.6(i). Thus

$$\begin{aligned} d_3 \quad &= \quad d(D_3 \cap J_2) + d(D_3 \cap J_1) \leqslant d(J_2) - d(D_5 \cap J_2) + 3/n \\ &\overset{(6.3),(6.10)}{\leqslant} \quad (1/2 + a + \varepsilon') - (1/4 - 2a - 2\varepsilon') + 3/n \leqslant 1/4 + 3a + 4\varepsilon', \end{aligned}$$

as required.

Since $A_1 = A_2 = O$, we have $E \subseteq D_0 \cup \ldots \cup D_3$. Let $B := J_3 \cap E \cap (D_2 \cup D_3)$. Assume now that $d(B) \leqslant a/4$. We claim that, in this case, the following hold:

$$d_0 + d_1 \geqslant 1/4 - a/4 - 2\varepsilon'. \tag{D11}$$

$$d_2 \geqslant 1/4 - a/2 - \varepsilon'. \tag{D12}$$

Indeed, to see the first inequality, observe that

$$d_0 + d_1 \geqslant d(D_0 \cap J_1) + d(J_3 \cap E) - \frac{a}{4} \overset{(6.6)}{\geqslant} \left(\frac{1}{4} - \frac{a}{2} - \varepsilon'\right) + \frac{\lfloor |J_3|/2 \rfloor}{n} - \frac{a}{4} \geqslant \frac{1}{4} - \frac{a}{4} - 2\varepsilon'.$$

The second is a consequence of (6.6). But $\{(D1), (D2), (D3), (D9), (D10), (D11), (D12)\}$ is an $(\varepsilon, 5)$-sufficient family.

The only case left is when $d(B) > a/4$. We claim that, in this case, the following hold:

$$d_2 \geqslant 1/4 - a/4 - 2\varepsilon'. \tag{D13}$$

$$d_0 \geqslant 1/4 - a/2 - \varepsilon'. \tag{D14}$$

The second inequality is simply (6.6). To see why the first holds, let $m := \min(B)$. Then

$$m < \left\lfloor \frac{n}{2} \right\rfloor - 2 \cdot \left(\frac{an}{4} - 1\right) \leqslant \frac{n}{2} - \frac{an}{2} + 2.$$

Since $m \in E$, we know that $m \notin A_1 \cup A_2$. In addition, since $m \in D_2 \cup D_3$, without loss of generality, we can assume that $m \in A_3 \cap A_4$. Therefore, by the definition of type (c) sets, we see that $|A_3|, |A_4| \leqslant m$. Thus $|A_3| + |A_4| < n - an + 4$. So, recalling that $|A_1| = |A_2| = \lceil n/2 \rceil$, the definition of a implies that $|A_5| = 3\lceil n/2 \rceil - an - |A_3| - |A_4| > n/2 - 4$. Thus $\min(A_5) > n/2 - 4$. So $|A_5 \cap J_3| \leqslant 3$. But $E \cap D_3 \subseteq A_3 \cap A_4 \cap A_5$, so $|J_3 \cap E \cap D_3| \leqslant 3$. Since $d(B) > a/4$, we have $d(J_3 \cap E \cap D_2) > a/4 - 3/n$. Thus

$$d_2 \geqslant d(J_1 \cap D_2) + d(J_3 \cap E \cap D_2) \overset{(6.6)}{\geqslant} 1/4 - a/2 - \varepsilon' + a/4 - 3/n = 1/4 - a/4 - 2\varepsilon',$$

as required. Now $\{(D1), (D2), (D3), (D9), (D10), (D13), (D14)\}$ is an $(\varepsilon, 5)$-sufficient family.

Case 4. $(s, t) = (3, 2)$.

We will assume that A_1, A_2 and A_3 are type (b) sets (and so A_4 and A_5 are type (c) sets). We will prove that the following constraints hold.

$$d_0 \geqslant 1/4 - a/2 - \varepsilon'. \tag{D15}$$
$$d_2 + d_5 \geqslant 1/2 - 2a - \varepsilon', \tag{D16}$$
$$d_1 + d_3 + d_4 \leqslant 1/4 + 5a/2 + 3\varepsilon', \tag{D17}$$
$$d_2 \leqslant 1/4 + a/2 + \varepsilon'. \tag{D18}$$

The first inequality (D15) is simply (6.6). For (D16), observe that $E \cap A_4 \cap A_5 \subseteq D_2$, and $O \cap A_4 \cap A_5 = D_5$. So

$$A_4 \cap A_5 \subseteq D_2 \cup D_5 \quad \text{and} \quad d_2 + d_5 = d(D_2 \cup D_5) \geqslant d(A_4 \cap A_5) \geqslant 1/2 - 2a - \varepsilon', \tag{6.11}$$

where we used Lemma 6.5 for the final inequality. For (D17), we have

$$
\begin{aligned}
d_1 + d_3 + d_4 \quad &\leqslant \quad d((D_1 \cup D_3 \cup D_4) \cap J_1) + d((D_1 \cup D_3 \cup D_4) \cap J_2) \\
&\overset{(6.11)}{\leqslant} \quad d(J_1) - d(D_0 \cap J_1) + d(J_2 \setminus (A_4 \cap A_5)) \\
&= \quad d(J_1) - d(D_0 \cap J_1) + d(J_2) - d(A_4 \cap A_5) + d((A_4 \cap A_5) \setminus J_2) \\
&\leqslant \quad 1 - (1/4 - a/2 - \varepsilon') - (1/2 - 2a - \varepsilon') + 1/n = 1/4 + 5a/2 + 3\varepsilon'.
\end{aligned}
$$

The final inequality follows from (6.6), (6.11) and Observation 6.6(i).

Finally we will prove (D18). Observe that $E \setminus (A_4 \cup A_5) \subseteq D_0$ and $O \setminus (A_4 \cup A_5) \subseteq D_3$. By Observation 6.6(i), we have that $|J_1 \cap (A_4 \cup A_5)| = |(A_4 \cup A_5) \setminus J_2| \leqslant 2$. Therefore $|(J_1 \cap (D_0 \cup D_3)| \geqslant |J_1| - 2$ and so $|D_2 \cap J_1| \leqslant 2$. Further, $O \subseteq D_3 \cup D_4 \cup D_5$. In particular, $D_2 \cap J_2 \subseteq E \cap J_2$. Combining these facts, we see that

$$d_2 = d(D_2 \cap J_1) + d(D_2 \cap J_2) \leqslant 2/n + d(E \cap J_2) \leqslant 2/n + \frac{\lceil |J_2|/2 \rceil}{n} \overset{(6.3)}{\leqslant} 1/4 + a/2 + \varepsilon',$$

as desired. But $\{(D1), (D2), (D3), (D15), (D16), (D17), (D18)\}$ is an $(\varepsilon, 5)$-sufficient family.

Case 5. $(s, t) = (2, 2)$

We will assume that A_1, A_2 are of type (b); A_3, A_4 are of type (c); and A_5 is of type (a). Our immediate aim is to prove that the following inequalities hold.

$$d_0 + d_1 \geqslant 1/4 - a/2 - \varepsilon'. \tag{D19}$$

$$d_3 - d_2 \leqslant 1/4 + 4a + 6\varepsilon'. \tag{D20}$$

The first is a consequence of (6.7). We will now prove (D20). This requires careful analysis of the small unstructured set A_5. Define sets

$$I_1' := \left[1, \left\lceil \frac{n}{4} \right\rceil\right], \quad I_1'' := J_1 \setminus I_1' = \left[\left\lceil \frac{n}{4} \right\rceil + 1, \left\lfloor \frac{n}{2} \right\rfloor - an\right],$$

$$X_1 := A_5 \cap O \cap I_1', \qquad X_2 := A_5 \cap O \cap I_1'', \qquad\qquad x_i := d(X_i) \text{ for } i = 1, 2;$$
$$Y_o := A_5 \cap O \cap J_2, \qquad Y_e := A_5 \cap E \cap J_2, \qquad\qquad y_o := d(Y_o) \text{ and } y_e := d(Y_e),$$

$$S := \{x + y : x, y \in X_2 \cup (Y_o \cap J_3)\}.$$

Clearly, it suffices to show that

$$d_3 \leqslant 3/8 + 7a/2 - x_2 + 5\varepsilon' \quad \text{and} \tag{6.12}$$

$$d_2 \geqslant 1/8 - a/2 - x_2 - \varepsilon'. \tag{6.13}$$

Let $Z := J_1 \cap (A_3 \cup A_4)$. By Observation 6.6(i), we have $|Z| \leqslant 2$. To prove (6.12), we bound $d(D_3 \cap J_1)$ and $d(D_3 \cap J_2)$ separately. If $v \in D_3 \cap E$, then $v \in A_3 \cap A_4 \cap A_5$. So $D_3 \cap J_1 \cap E \subseteq Z$. Similarly, $D_3 \cap J_1 \cap O \subseteq (A_5 \cap J_1) \cup Z$. Thus

$$d(D_3 \cap J_1) = d(D_3 \cap J_1 \cap E) + d(D_3 \cap J_1 \cap O) \leqslant d(Z) + x_1 + x_2 \leqslant x_1 + x_2 + \varepsilon'. \tag{6.14}$$

Now, $A_3 \cap A_4 \cap O \subseteq D_4 \cup D_5$. Further, $D_3 \cap E \subseteq A_5$. Therefore

$$\begin{aligned}
d(D_3 \cap J_2) &= d(D_3 \cap J_2 \cap E) + d(D_3 \cap J_2 \cap O) \leqslant y_e + d(J_2 \cap O) - d(A_3 \cap A_4 \cap J_2 \cap O) \\
&\leqslant y_e + d(J_2 \cap O) - (d(A_3 \cap A_4 \cap J_2) - d(J_2 \cap E)) \\
&\leqslant y_e + d(J_2) - d(A_3 \cap A_4 \cap I_2) \leqslant y_e + (1/2 + a + \varepsilon') - (1/2 - 2a - 2\varepsilon') \\
&= y_e + 3a + 3\varepsilon', \tag{6.15}
\end{aligned}$$

where we used (6.3) and Lemma 6.7 for the final inequality.

For every $s \in S$, we have that s is even and at most n, and additionally $s \geqslant 2(\lceil n/4 \rceil + 1) \geqslant \lfloor n/2 \rfloor - an + 1$. Thus $S \subseteq J_2 \cap E$. Since A_5 is sum-free, we have that $Y_e \cap S = \emptyset$. So $|S| + |Y_e| \leqslant |J_2 \cap E|$. By Proposition 2.4 applied to S, we have $|S| \geqslant 2|X_2 \cup (Y_o \cap J_3)| - 1$. Thus

$$2(|X_2| + |Y_o \cap J_3|) + |Y_e| \leqslant |S| + |Y_e| + 1 \leqslant |J_2 \cap E| + 1 \stackrel{(6.3)}{\leqslant} (1/4 + a/2 + \varepsilon')n. \tag{6.16}$$

Combining this with (6.14) and (6.15) we have

$$\begin{aligned}
d_3 &\leqslant x_1 + x_2 + y_e + 3a + 4\varepsilon' \stackrel{(6.16)}{\leqslant} x_1 + (1/4 + a/2 + \varepsilon' - x_2) + 3a + 4\varepsilon' \\
&\leqslant 3/8 + 7a/2 - x_2 + 5\varepsilon',
\end{aligned}$$

where we used the trival bound $x_1 n \leqslant \lceil |I_1'|/2 \rceil$. This finishes the proof of (6.12).

For (6.13), notice that $\mathrm{O} \setminus (A_3 \cup A_4 \cup A_5) \subseteq D_2$. By definition, $J_1 \setminus Z$ is disjoint from $A_3 \cup A_4$. Therefore

$$
\begin{aligned}
d_2 &\geqslant d(\mathrm{O} \cap I_1'') - d(\mathrm{O} \cap I_1'' \cap (A_3 \cup A_4)) - d(\mathrm{O} \cap I_1'' \cap A_5) \\
&\geqslant d(\mathrm{O} \cap I_1'') - d(Z) - x_2 \geqslant \frac{1}{n} \lfloor |I_1''|/2 \rfloor - 2/n - x_2 \geqslant 1/8 - a/2 - x_2 - \varepsilon',
\end{aligned}
$$

as required. We have proved (6.13) and hence (D20).

The remainder of the proof will be divided into two final subcases. First, suppose that $Y_o \subseteq J_3$. Now (6.16) implies that $2(x_2 + y_o) + y_e \leqslant 1/4 + a/2 + \varepsilon'$. Thus

$$
d(A_5) = x_1 + x_2 + y_o + y_e \leqslant d(\mathrm{O} \cap I_1') + 1/4 + a/2 + \varepsilon' \leqslant 3/8 + a/2 + 2\varepsilon'. \tag{6.17}
$$

Lemma 6.3 implies that $\sum_{i \in [4]} d(A_i) \leqslant 2 - a + \varepsilon'$. Adding these, we see that

$$
\sum_{i \in [5]} i d_i = \sum_{i \in [5]} d(A_i) \leqslant 19/8 - a/2 + 3\varepsilon'. \tag{D21}
$$

Now $\{(D1), (D3), (D19), (D20), (D21)\}$ is an $(\varepsilon, 5)$-sufficient family.

The final subcase is when $Y_o \setminus J_3 \neq \emptyset$. Let $w \in Y_o \setminus J_3$. So $w \geqslant \lfloor n/2 \rfloor + 1$. We claim that the following inequality holds:

$$
d_3 \leqslant 5/16 + 7a/2 + 5\varepsilon'. \tag{D22}
$$

To prove this, define

$$
X_1' := X_1 \setminus \left\{ \left\lceil \frac{n}{4} \right\rceil \right\} \quad \text{and} \quad D := \begin{cases} \{w + x : x \in X_1'\} & \text{if } w \in \left[\left\lfloor \frac{n}{2} \right\rfloor + 1, n - \left\lceil \frac{n}{4} \right\rceil \right], \\[2mm] \{w - x : x \in X_1'\} & \text{if } w \in \left[n - \left\lceil \frac{n}{4} \right\rceil + 1, n \right]. \end{cases}
$$

Now, $w \in \mathrm{O}$ and $X_1' \subseteq \mathrm{O}$, so every element of D is even. In both cases it is easy to check that $D \subseteq E \cap J_2$. Since A_5 is sum-free, $D \cap Y_e = \emptyset$, and so D and Y_e are disjoint subsets of $E \cap J_2$. In particular,

$$
x_1 + y_e \leqslant |D|/n + y_e + 1/n \leqslant d(J_2 \cap E) + 1/n \overset{(6.3)}{\leqslant} 1/4 + a/2 + \varepsilon'. \tag{6.18}
$$

This then implies that

$$
\begin{aligned}
d_3 = d(D_3 \cap J_1) + d(D_3 \cap J_2) &\overset{(6.14),(6.15)}{\leqslant} x_1 + x_2 + y_e + 3a + 4\varepsilon' \\
&= \frac{1}{2}(x_1 + y_e) + \frac{1}{2}(2x_2 + y_e) + \frac{x_1}{2} + 3a + 4\varepsilon' \overset{(6.16),(6.18)}{\leqslant} \frac{1}{4} + \frac{a}{2} + \frac{x_1}{2} + 3a + 5\varepsilon' \\
&\leqslant \frac{5}{16} + \frac{7a}{2} + 5\varepsilon',
\end{aligned}
$$

This proves the claim. But $\{(D1), (D2), (D3), (D19), (D20), (D22)\}$ is an $(\varepsilon, 5)$-sufficient family. This completes the proof of the final case, and hence completes the proof of Lemma 6.2. $\qquad\square$

7 Concluding remarks

We determined $f(n,r)$ exactly when $r = 2$ (Theorem 1.2). It would be interesting to proceed from our stability result (Theorem 1.3) and obtain an exact result for $r = 3$, and characterise the extremal sets. It seems possible to extract a statement about stability from the proof of Theorem 1.4 by more careful analysis of the linear programs. That is, the following may be obtainable. For all $\varepsilon > 0$, as long as n is a sufficiently large integer: if $r = 4$ and $A \subseteq [n]$ is extremal, then one of $A \triangle O$, $A \triangle I_2$ and $A \triangle (O \cup I_2)$ has size at most εn; and if $r = 5$ and $A \subseteq [n]$ is extremal, then $|A \triangle (O \cup I_2)| \leqslant \varepsilon n$.

It is also possible that the method used to prove Theorem 1.4 (namely finding sufficient linear constraints) can prove the analogous result for $r = 6$. The main obstacle is the fact that, among extremal $A_1, \ldots, A_6$, one cannot *a priori* guarantee less than two type (a) sets. This leads to 18 different values of (s,t) to consider. Since the proof for $r = 5$ was already very involved, we did not pursue this further.

Finally, for large r, the value of $f(n,r)$ and the structure of the extremal sets is completely open.

Acknowledgements

References

[1] N. Alon, J. Balogh, P. Keevash and B. Sudakov, The number of edge colorings with no monochromatic cliques, *J. London Math. Soc.* **70** (2004), 273–288.

[2] N. Alon and R. Yuster, The number of orientations having no fixed tournament, *Combinatorica* **26** (2006), 1–16.

[3] J. Balogh, A remark on the number of edge colorings of graphs, *Europ. J. Comb.* **27** (2006), 565–573.

[4] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, *J. Amer. Math. Soc.* **28** (2015), 669–709.

[5] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, The number of maximal sum-free subsets of integers, *Proc. Amer. Math. Soc.*, **143**, (2015) 4713–4721.

[6] F. S. Benevides, C. Hoppen and R. M. Sampaio, Edge-colorings of graphs avoiding complete graphs with a prescibed coloring, **340** (9) (2017), 2143–2160.

[7] P. Cameron and P. Erdős, On the number of sets of integers with various properties, in Number Theory (R.A. Mollin, ed.), 61–79, Walter de Gruyter, Berlin, 1990.

[8] D. Clemens, S. Das and T. Tran, Colourings without monochromatic disjoint pairs, *European J. Combin.* **70** (2018), 99–124.

[9] J. Deshouillers, G. Freiman, V. Sós and M. Temkin, On the structure of sum-free sets II, *Astérisque* **258** (1999), 149–161.

[10] P. Erdős, Some new applications of probability methods to combinatorial analysis and graph theory, Proceedings of the Fifth Southeastern Conference on Combinatorics, Graph Theory and Computing, Congress Numerantium X (1974), 39–51.

[11] P. Erdős, Some of my favorite problems in various branches of combinatorics, *Matematiche (Catania)* **47** (1992), 231–240.

[12] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n-free graphs, in *Colloq. Int. Teorie Combin.* (Rome, 1973), Tomo II, Atti dei Convegni Lincei, No. 17, Accademia Nazionale dei Lincei, Rome, 1976, 19–27.

[13] J. Fox, A new proof of the graph removal lemma, *Ann. of Math.*, **174** (2) (2011), 561–579.

[14] B. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.* **36** (2004), 769–778.

[15] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.* **15** (2005), 340–376.

[16] B. Green and I. Ruzsa, Counting sumsets and sum-free sets modulo a prime, *Studia Sci. Math. Hungarica* **41** (2004), 285–293.

[17] B. Green and I. Ruzsa, Sum-free sets in abelian groups, *Israel J. Math* **147** (2005), 157–189.

[18] H. Hàn and A. Jiménez, Maximum number of sum-free colorings in finite abelian groups, *Israel J. Math* **226** (2018), 505–534.

[19] C. Hoppen, Y. Kohayakawa and H. Lefmann, Kneser colorings of uniform hypergraphs, *Elec. Notes in Disc. Math.* **34** (2009), 219–223.

[20] C. Hoppen, Y. Kohayakawa and H. Lefmann, Edge colourings of graphs avoiding monochromatic matchings of a given size, *Comb. Prob. Comp.* **21** (2012), 203–218.

[21] C. Hoppen, Y. Kohayakawa and H. Lefmann, Edge-colorings of graphs avoiding fixed monochromatic subgraphs with linear Turán number, *Europ. J. Comb.* **35** (2014), 354–373.

[22] C. Hoppen and H. Lefmann, Edge-colorings avoiding a fixed matching with a prescribed color pattern, *Europ. J. Comb.* **47** (2015), 75–94.

[23] C. Hoppen, H. Lefmann and K. Odermann, A coloring problem for intersecting vector spaces, *Disc. Math.* **339** (2016), 2941–2954.

[24] C. Hoppen, H. Lefmann, K. Odermann and J. Sanches, Edge-colorings avoiding fixed rainbow stars, *Elec. Notes in Disc. Math.* **50** (2015), 275–280.

[25] D. Kleitman and K.J. Winston, The asymptotic number of lattices, *Ann. Discrete Math.* **6** (1980), 243–249, Combinatorial mathematics, optimal designs and their applications (Proc. Sympos. Combin. Math. and Optimal Design, Colorado State Univ., Fort Collins, Colo., 1978).

[26] D. Kleitman and K.J. Winston, On the number of graphs without 4-cycles, *Discrete Math.* **6** (1982), 167–172.

[27] D. Král', O. Serra and L. Vena, A combinatorial proof of the Removal Lemma for groups, *J. Combin. Theory A* **116**, (2009), 971–978.

[28] H. Lefmann, Y. Person, V. Rödl and M. Schacht, On colorings of hypergraphs without monochromatic Fano planes, *Comb. Prob. Comp.* **18** (2009), 803–818.

[29] H. Lefmann, Y. Person and M. Schacht, A structural result for hypergraphs with many restricted edge colorings, *J. Comb.* **1** (2010), 441-475.

[30] W. Mantel, Problem 28, *Wiskundige Opgaven* **10** (1907), 60–61.

[31] O. Pikhurko, K. Staden and Z. Yilma, The Erdős-Rothschild problem on edge-colourings with forbidden monochromatic cliques, *Math. Proc. Cambridge Phil. Soc* **163** (2017), 341–356.

[32] O. Pikhurko and Z. Yilma, The maximum number of K_3-free and K_4-free edge 4-colorings, *J. London Math. Soc.* **85** (2012), 593–615.

[33] I.Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics* (Keszthely, 1976), Coll. Math. Soc. J. Bolyai **18**, Volume II, 939–945.

[34] A.A. Sapozhenko, The Cameron-Erdős conjecture, (Russian) *Dokl. Akad. Nauk.*, **393**, (2003), 749–752.

[35] D. Saxton and A. Thomason, Hypergraph containers, *Invent. Math.* **201** (2015), 925–992.

[36] E. Szemerédi, 'Regular partitions of graphs' in *Proc. Colloq. Int. CNRS*, Paris, 1976, pp. 309–401.

[37] T. Tran, On the structure of large sum-free sets in integers, *Israel J. Math.* **228**, (2018), 249–292.

[38] R. Yuster, The number of edge colorings with no monochromatic triangle, *J. Graph Theory* **21** (1996), 441–452.

A   Appendix

It remains to prove that the families obtained in the proof of Lemma 6.2 are indeed (ε, r)-sufficient. Namely, we require that the following lemma holds.

Lemma A.1. *Given $\varepsilon > 0$, for $r \in \{4, 5\}$ there exists $\delta, n_0 > 0$ such that whenever $\delta \leqslant \varepsilon' \leqslant 1/100$ is a real constant and $n \geqslant n_0$ is an integer and $A_1, \ldots, A_r$ are maximal sum-free subsets of $[n]$, we have that: The following families (depending on ε') are $(\varepsilon, 4)$-*

sufficient.

$$
\begin{array}{clll}
1) & \{(C0^*), (C1)\} & \textit{(Case 0)} \\
2) & \{(C1), (C2), (C4), (C5)\} & \textit{(Case 1)} \\
3) & \{(C1), (C2), (C3), (C4), (C6)\} & \textit{(Case 1)} \\
4) & \{(C1), (C2), (C3), (C4), (C7)\} & \textit{(Case 2)} \\
5) & \{(C1), (C2), (C3), (C8), (C9)\} & \textit{(Case 3)}
\end{array}
$$

The following families (depending on ε') are $(\varepsilon, 5)$-sufficient.

$$
\begin{array}{clll}
6) & \{(D0^*), (D1)\} & \textit{(Case 0)} \\
7) & \{(D1), (D2), (D3), (D4)\} & \textit{(Case 1)} \\
8) & \{(D1), (D2), (D3), (D5), (D6)\} & \textit{(Case 2)} \\
9) & \{(D1), (D2), (D3), (D5), (D7), (D8)\} & \textit{(Case 2)} \\
10) & \{(D1), (D2), (D3), (D9), (D10), (D11), (D12)\} & \textit{(Case 3)} \\
11) & \{(D1), (D2), (D3), (D9), (D10), (D13), (D14)\} & \textit{(Case 3)} \\
12) & \{(D1), (D2), (D3), (D15), (D16), (D17), (D18)\} & \textit{(Case 4)} \\
13) & \{(D1), (D3), (D19), (D20), (D21)\} & \textit{(Case 4)} \\
14) & \{(D1), (D2), (D3), (D19), (D20), (D22)\} & \textit{(Case 5)}.
\end{array}
$$

Proof. For 1), 3) and 4), taking $\varepsilon' = 1/100$ in Mathematica yields $\sum_{i\in[4]} d_i \log i < 1 - 1/1000$, so we are done in these cases. Given a linear maximisation (*primal*) program:

Maximise $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{d}$ subject to $A\boldsymbol{d} \leqslant \boldsymbol{b}$ and $\boldsymbol{d} \geqslant \boldsymbol{0}$,

the dual minimisation program is:

Minimise $\boldsymbol{b}^{\mathsf{T}}\boldsymbol{y}$ subject to $A^{\mathsf{T}}\boldsymbol{y} \geqslant \boldsymbol{c}$ and $\boldsymbol{y} \geqslant \boldsymbol{0}$.

Family 2) Taking the program represented by 2) as the primal, we have

$$
\begin{array}{c}
\begin{array}{cccccc}
\text{a} & d_0 & d_1 & d_2 & d_3 & d_4
\end{array} \\
\begin{array}{c}
y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6
\end{array}
\left(\begin{array}{cccccc}
 & 1 & 1 & 1 & 1 & 1 \\
1 & & & & & \\
1 & & 1 & 2 & 3 & 4 \\
 & & & & 1 & 1 \\
-1 & -1 & & & & \\
-4 & & & & & -1
\end{array}\right)
\begin{array}{l}
\leqslant 1 \\
\leqslant \frac{1}{10} \\
\leqslant 2 + \varepsilon' \\
\leqslant \frac{1}{2} + \varepsilon' \\
\leqslant -\frac{1}{2} + \varepsilon' \\
\leqslant -\frac{1}{2} + \varepsilon'.
\end{array} \\
\begin{array}{cccccc}
\geqslant 0 & \geqslant 0 & \geqslant 0 & \geqslant 1 & \geqslant \log 3 & \geqslant 2
\end{array}
\end{array}
\qquad \leftrightarrow \qquad
\begin{array}{c}
\boldsymbol{d}^{\mathsf{T}} \\
\boldsymbol{y} \; (\; A \;) \; \leqslant \boldsymbol{b} \\
\geqslant \boldsymbol{c}^{\mathsf{T}}
\end{array}
$$

A feasible solution to the dual program is $\boldsymbol{y}^* = (0, 0, \frac{\log 3}{3}, 0, 0, \frac{4\log 3}{3} - 2)$. The objective function value of the dual at $\boldsymbol{y}^*$ is

$$
\boldsymbol{b}^{\mathsf{T}}\boldsymbol{y}^* = (2 + \varepsilon')\frac{\log 3}{3} + \left(-\frac{1}{2} + \varepsilon'\right)\left(\frac{4\log 3}{3} - 2\right) = 1 + \left(\frac{5\log 3}{3} - 2\right)\varepsilon' \leqslant 1 + \varepsilon' \leqslant 1 + \varepsilon.
$$

By the weak duality theorem, any feasible solution $\boldsymbol{x}$ to the primal maximisation linear program satisfies $\boldsymbol{c}^\mathsf{T}\boldsymbol{x} \leqslant \boldsymbol{b}^\mathsf{T}\boldsymbol{y}^* \leqslant 1 + \varepsilon$. Thus the family in 2) is $(\varepsilon, 4)$-sufficient.

Family 5) The family yields the following primal and dual linear programs.

	a	d_0	d_1	d_2	d_3	d_4	
y_1		1	1	1	1	1	$\leqslant 1$
y_2					1	1	$\leqslant \frac{1}{2} + \varepsilon'$
y_3	1		1	2	3	4	$\leqslant 2 + \varepsilon'$
y_4	1						$\leqslant \frac{1}{10}$
y_5	$-\frac{5}{2}$			-1			$\leqslant -\frac{1}{2} + 2\varepsilon'$
y_6	-3				1		$\leqslant \varepsilon'.$
	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 1$	$\geqslant \log 3$	$\geqslant 2$	

A feasible solution to the dual program is $\boldsymbol{y}^* = (0, 0, \frac{1}{2}, 0, 0, \log 3 - \frac{3}{2})$. The objective function value of the dual at $\boldsymbol{y}^*$ is

$$\boldsymbol{b}^\mathsf{T}\boldsymbol{y}^* = (2 + \varepsilon') \cdot \frac{1}{2} + \varepsilon'\left(\log 3 - \frac{3}{2}\right) = 1 + (\log 3 - 1)\varepsilon' \leqslant 1 + \varepsilon' \leqslant 1 + \varepsilon.$$

We are again done by the weak duality theorem.

Now we let $r = 5$ and consider families 6)–14). For 6)–9), 13) and 14), taking $\varepsilon' = 1/100$ in Mathematica yields $\sum_{i\in[5]} d_i \log i < \frac{1}{4} \log 30 - 1/10^4$, so we are done in these cases. It remains to consider 10)–12).

Family 10) The family yields the following primal and dual linear programs.

	a	d_0	d_1	d_2	d_3	d_4	d_5	
y_1		1	1	1	1	1	1	$\leqslant 1$
y_2	1							$\leqslant \frac{1}{5}$
y_3	1		1	2	3	4	5	$\leqslant \frac{5}{2} + \varepsilon'$
y_4	-2						-1	$\leqslant -\frac{1}{4} + 2\varepsilon'$
y_5	-3				1			$\leqslant \frac{1}{4} + 4\varepsilon'$
y_6	$-\frac{1}{4}$	-1	-1					$\leqslant -\frac{1}{4} + 2\varepsilon'$
y_7	$-\frac{1}{2}$			-1				$\leqslant -\frac{1}{4} + \varepsilon'.$
	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 1$	$\geqslant \log 3$	$\geqslant 2$	$\geqslant \log 5$	

A feasible solution to the dual program is

$$\boldsymbol{y}^* = \left(4x, 0, \frac{1}{2} - x, \frac{5}{2} - \log 5 - x, \log 3 - \frac{3}{2} - x, 4x, 2x\right)$$

where $x = \frac{3}{2}\log 3 - \log 5$. The objective function value of the dual at $\boldsymbol{y}^*$ is

$$\boldsymbol{b}^\mathsf{T}\boldsymbol{y}^* = \frac{1}{4}\log 30 + \left(-\frac{1}{2} + 22\log 3 - 14\log 5\right)\varepsilon' < \frac{1}{4}\log 30 + 2\varepsilon' \leqslant \frac{1}{4}\log 30 + \varepsilon.$$

So the family in 10) is $(\varepsilon, 5)$-sufficient.

Family 11) The family yields the following primal and dual linear programs.

	a	d_0	d_1	d_2	d_3	d_4	d_5	
y_1		1	1	1	1	1	1	$\leqslant 1$
y_2	1							$\leqslant \frac{1}{5}$
y_3	1		1	2	3	4	5	$\leqslant \frac{5}{2} + \varepsilon'$
y_4	-2						-1	$\leqslant -\frac{1}{4} + 2\varepsilon'$
y_5	-3				1			$\leqslant \frac{1}{4} + 4\varepsilon'$
y_6	$-\frac{1}{4}$			-1				$\leqslant -\frac{1}{4} + 2\varepsilon'$
y_7	$-\frac{1}{2}$	-1						$\leqslant -\frac{1}{4} + \varepsilon'$
	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 1$	$\geqslant \log 3$	$\geqslant 2$	$\geqslant \log 5$	

A feasible solution to the dual is

$$\boldsymbol{y}^* = \left(4x, 0, \frac{1}{2} - x, \frac{5}{2} - \log 5 - x, \log 3 - \frac{3}{2} - x, 4x, 4x\right)$$

where $x = 2\log 3 - \frac{4}{3}\log 5$. The objective function value of the dual at $\boldsymbol{y}^*$ is

$$\boldsymbol{b}^\mathsf{T}\boldsymbol{y}^* = \frac{1}{4}\log 30 + \left(-\frac{1}{2} + 6\log 3 - \frac{10}{3}\log 5\right)\varepsilon' < \frac{1}{4}\log 30 + 2\varepsilon' \leqslant \frac{1}{4}\log 30 + \varepsilon.$$

So the family in 11) is $(\varepsilon, 5)$-sufficient.

Family 12) The family yields the following primal and dual linear programs.

	a	d_0	d_1	d_2	d_3	d_4	d_5	
y_1		1	1	1	1	1	1	$\leqslant 1$
y_2	1		1	2	3	4	5	$\leqslant \frac{5}{2} + \varepsilon'$
y_3	1							$\leqslant \frac{1}{5}$
y_4	$-\frac{1}{2}$	-1						$\leqslant -\frac{1}{4} + \varepsilon'$
y_5	-2			-1			-1	$\leqslant -\frac{1}{2} + \varepsilon'$
y_6	$-\frac{5}{2}$		1		1	1		$\leqslant \frac{1}{4} + 3\varepsilon'$
y_7	$-\frac{1}{2}$		1					$\leqslant \frac{1}{4} + \varepsilon'$
	$\geqslant 0$	$\geqslant 0$	$\geqslant 0$	$\geqslant 1$	$\geqslant \log 3$	$\geqslant 2$	$\geqslant \log 5$	

A feasible solution to the dual is

$$\boldsymbol{y}^* = \left(y_1, \frac{\log 5}{5} - \frac{y_1}{5}, 0, y_1, 0, \log 3 - \frac{3}{5}\log 5 - \frac{2}{5}y_1, 1 - \frac{2}{5}\log 5 - \frac{3}{5}y_1\right)$$

where $y_1 = 1/5$. The objective function value of the dual at $\boldsymbol{y}^*$ is

$$\boldsymbol{b}^\mathsf{T}\boldsymbol{y}^* = \frac{1}{4}\log 30 + \left(3\log 3 - 2\log 5 + 1 - y_1\right)\varepsilon' < \frac{1}{4}\log 30 + \varepsilon' \leqslant \frac{1}{4}\log 30 + \varepsilon.$$

So the family in 12) is $(\varepsilon, 5)$-sufficient. This completes the proof. $\qquad\square$