Minimum Number of Affine Simplexes of Given Dimension

István Szalkai *,† Zsolt Tuza ‡,§

Latest update on 2013–9–23

Abstract

In this paper we formulate and solve extremal problems in the Euclidean space $\mathbb{R}^d$ and further in hypergraphs, originating from problems in stoichiometry and elementary linear algebra. The notion of affine simplex is the bridge between the original problems and the presented extremal theorem on set systems. As a sample corollary, it follows that if no triple is collinear in a set S of n points in $\mathbb{R}^3$, then S contains at least $\binom{n}{4} - cn^3$ affine simplexes for some constant c. A function related to Sperner's theorem and the YBLM inequality is also considered and its relation to hypergraph Turán problems is discussed.

Keywords: linear hypergraph, extremal set theory, Euclidean affine simplex, minimal linear dependency, stoichiometry.

AMS Subject Classification: 05C65, 05D05, 15A03.

* corresponding author

† Department of Mathematics, University of Pannonia, Veszprém, Hungary; e-mail: szalkai@almos.uni-pannon.hu

‡ Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, and Department of Computer Science and Systems Technology, University of Pannonia, Veszprém, Hungary; e-mail: tuza@dcs.uni-pannon.hu

§ Research supported in part by the Hungarian Scientific Research Fund, OTKA grant T-81493, and co-financed by the European Social Fund through the project TÁMOP-4.2.2.C-11/1/KONV-2012-0004 – National Research Center for Development and Market Introduction of Advanced Information and Communication Technologies.

1 Introduction

The roots of the present study date back to the mid-1980s, to the paper by S. Kumar and Á. Pethő [6], concerning an application of linear algebra in stoichiometry. From the algebraic point of view, their very natural question asks about the number of those subsets of a set of vectors which are linearly dependent but each of whose proper subsets is independent. Here we give an asymptotically tight solution for the minimum in terms of dimension and the number of vectors when low-dimensional dependencies are excluded. Our method is to prove a more general result in extremal set theory (Theorem 6 below), hence without assuming anything about the structure of algebraic dependencies.

1.1 Motivation in chemistry

Restricting attention to a "universe" of D kinds of atoms (or atomic parts), each molecule (species) can be represented with a vector in $\mathbb{R}^D$ whose ith coordinate means the number of atoms of ith type in the molecule in question.[1] Then a chemical reaction naturally corresponds to a zero-sum linear combination of these vectors (using the law of mass balance).

The reaction is called minimal if none of the molecules, taking role in it, can be omitted so that the remaining ones could form still a(nother) reaction. In the language of linear algebra this assumption is equivalent to the property that the corresponding set of vectors is linearly dependent but each proper subset of it is independent, that is the defining condition of *linear algebraic simplex*. Both from practical and theoretical purposes the following problem was raised:

Problem 1 *What is the minimum and maximum number of linear algebraic simplexes $S \subsetneqq \mathcal{V}$ in a set $\mathcal{V}$ of vectors in $\mathbb{R}^D$ if only the size $|\mathcal{V}|$ is given and $\mathcal{V}$ spans $\mathbb{R}^D$? What are the structures of sets $\mathcal{V}$ which contain extremal number of simplexes?*

The answer was given in [7]. Moreover, Problem 1 was generalized for matroids in [3]; actually its authors solved it a decade earlier than published, see [2].

[1] The types of atoms are supposed to be in a fixed order. E.g., if $D = 3$ and the universe is $[\mathrm{C}, \mathrm{H}, \mathrm{O}]$, then we have the vector $(0, 2, 1)$ for H_2O and $(2, 4, 2)$ for CH_3COOH.

Concerning minimum, the results in [7] show that almost all vectors must be *parallel*, i.e. almost all molecules (species) are isomer molecules or multiple doses. The problem where parallel vectors are excluded is still unsolved in general:

Problem 2 *What is the minimum number of linear algebraic simplexes $S \subseteq \mathcal{V}$ if only the size $|\mathcal{V}|$ is given, $\mathcal{V}$ does not contain parallel vectors and $\mathcal{V}$ spans $\mathbb{R}^D$? What are the structures of sets $\mathcal{V}$ which contain the minimum number of simplexes?*

A conjecture on both the minimum number and the structure attaining it is stated in [8]. The cases $D = 3$ and $D = 4$ were solved in [8] and [15], respectively.

1.2 Geometric formulation

In the framework of linear algebra the problem is somewhat non-symmetric because the zero vector plays a special role. This asymmetry can be eliminated if we translate the problem to the language of geometry. Moreover, restricting attention to sets $\mathcal{V} \subset \mathbb{R}^D$ containing neither the zero vector nor a pair of parallel vectors, dimension can be reduced from D to $d = D - 1$: first associate each element $\underline{v} \in \mathcal{V}$ with its direction $\lambda \cdot \underline{v}$ ($\lambda \in \mathbb{R}$), and then intersect this system $\Lambda\mathcal{V}$ with a $(D-1)$-dimensional hyperplane $\mathcal{P}$ which does not contain the origin and is not parallel to any element of $\mathcal{V}$.

The mapping from $\mathcal{V}$ to the set $\mathcal{V}^{\mathcal{P}} := \Lambda\mathcal{V} \cap \mathcal{P}$ is a bijection under which linear algebraic simplexes $S \subset \mathcal{V}$ correspond to *affine simplexes* $S^{\mathcal{P}} \subset \mathbb{R}^{D-1}$, where a set S of $k \geq 3$ points in the Euclidean d-space is called an *affine simplex* if S is contained in some $(k-2)$-dimensional hyperplane but no proper subset $S' \subsetneqq S$ is contained in a hyperplane of dimension $|S'| - 2$. For instance, in $\mathbb{R}^3$ the following three types of affine simplexes occur:

- three collinear points;
- four coplanar points, no three of which are collinear;
- five points, no four of which are coplanar.

Affine simplexes can alternatively be defined by requiring that the vectors $\underline{s}_2 - \underline{s}_1, \underline{s}_3 - \underline{s}_1, \ldots, \underline{s}_k - \underline{s}_1$ be linearly dependent but their proper subsets shouldn't (for every choice of a point to be labeled $\underline{s}_1$).

In cases of low dimension, as solved in [8] and [15], almost all points of the extremal configurations for Problem 2 attaining the minimum number of affine simplexes lie on one or two lines, i.e. mostly contain affine simplexes of three points. In this way the natural question arises to determine the minimum in the other extreme, where no three points are collinear. For this reason our goal is to study point sets which contain no affine simplexes smaller than a given size. The first interesting case is $\mathbb{R}^3$.

Let $S \subset \mathbb{R}^d$ be a set of n points, no d of which lie on a $(d-2)$-dimensional hyperplane. Then two kinds of subsets of S form an affine simplex:

- $d+1$ points on a hyperplane of dimension $d-1$, or
- $d+2$ points, no $d+1$ of which lie on a common hyperplane of dimension $d-1$.

Theorem 3 *For every $d \geq 3$ there is a constant $c = c(d)$ with the following property. If $S \subset \mathbb{R}^d$ is a set of n points, no d of them lying on a hyperplane of dimension $d-2$, then S determines at least $\binom{n}{d+1} - cn^d$ affine simplexes.*

Corollary 4 *For any n points in the 3-space, no three being collinear, the number of coplanar quadruples plus the 5-tuples containing no coplanar quadruples is at least $\binom{n}{4} - O(n^3)$ as $n \to \infty$.*

These results are asymptotically tight, as shown by the obvious example of n coplanar points in $\mathbb{R}^3$ (no three of them being on a line) and also for any $d \geq 3$ by n points of $\mathbb{R}^{d-1}$ in general position when embedded isometrically into $\mathbb{R}^d$. Such a set of points has exactly $\binom{n}{d+1}$ affine simplexes. In fact, configurations with even fewer affine simplexes exist, which in addition span the d-space. For instance, $n-1$ points of $\mathbb{R}^{d-1}$ in general position embedded in a hyperplane of $\mathbb{R}^d$ plus an nth point outside that hyperplane generate just $\binom{n-1}{d+1}$ affine simplexes (as no affine simplex contains the nth point).

In $\mathbb{R}^3$, the two arrangements of points just mentioned yield $\frac{1}{24}n^4 - \frac{1}{4}n^3 + O(n^2)$ and $\frac{1}{24}n^4 - \frac{5}{12}n^3 + O(n^2)$, respectively. Currently we do not know whether or not the latter error term $\frac{5}{12}n^3$ is asymptotically tight. We do know, however, that the construction above is not extremal; an improvement of the order $O(n^2)$ will be proved in Proposition 7.

1.3 Combinatorial formulation

Here we put the problems and results above in a more general setting. Let $\mathcal{H} = (X, \mathcal{E})$ be a hypergraph, where X is the finite vertex set and $\mathcal{E}$ is the edge set consisting of subsets of X. We extend the notion of linear hypergraph (also called "simple" or "almost disjoint" in some parts of the literature) as follows.

Definition 5 *We say that a hypergraph $\mathcal{H} = (X, \mathcal{E})$ is q-*linear *(for some integer $q \geq 1$) if $|E \cap E'| < q$ holds for all $E, E' \in \mathcal{E}$, $E \neq E'$.*

Hence, in a 1-linear hypergraph any two edges are disjoint, and 2-linear coincides with linear hypergraphs in the usual sense, in analogy with Euclidean spaces where any two points uniquely determine a line.

We also introduce some notation. As usual, $\binom{S}{k}$ will stand for the collection of all k-element subsets of set S. For any hypergraph $\mathcal{H} = (X, \mathcal{E})$, let

- $\mathcal{E}_k := \bigcup_{E \in \mathcal{E}} \binom{E}{k}$ — usually $(X, \mathcal{E}_k)$ is called the *k-section hypergraph* of $\mathcal{H}$;
- $\mathcal{E}^0_{k+1} := \{F \in \binom{X}{k+1} \mid \binom{F}{k} \cap \mathcal{E}_k = \emptyset\}$.

Corresponding to $k = d+1$, in analogy with the geometric interpretation, we call the members of $\mathcal{E}_k \cup \mathcal{E}^0_{k+1}$ the $(k-1)$-dimensional *semi-simplexes* in $\mathcal{H}$.

Theorem 6 *For every $k \geq 3$ there is a constant $c = c(k)$ such that*

$$|\mathcal{E}_k| + |\mathcal{E}^0_{k+1}| \geq \binom{n}{k} - cn^{k-1}$$

holds for all $(k-1)$-linear hypergraphs $\mathcal{H} = (X, \mathcal{E})$ on n vertices.

This result implies Theorem 3, by considering the hypergraph whose edges are the sets of points lying on a common hyperplane of dimension $d-1$.

1.4 Sperner families and Turán numbers

For any hypergraph $\mathcal{H} = (X, \mathcal{E})$ (not necessarily q-linear for a prescribed value of q) and for any k, the set system $\mathcal{S}_k(\mathcal{H}) := \mathcal{E}_k \cup \mathcal{E}^0_{k+1}$ is a *Sperner family*, which means that none of its members contains any other $S \in \mathcal{S}$.

The well known YBLM inequality[2] states that

$$\sum_{S\in\mathcal{S}} \binom{n}{|S|}^{-1} \leq 1 \tag{1}$$

holds for every Sperner family $\mathcal{S}$ (where n is the number of vertices). In particular, (1) is valid for the family $\mathcal{S} = \mathcal{S}_k(\mathcal{H})$ of any $\mathcal{H}$, too. In connection with the main problem studied here, one may also consider the values

$$s(n,k) := \min_{\mathcal{H}\text{ is }(k-1)\text{-linear},\ |X|=n} \sum_{S\in\mathcal{S}_k(\mathcal{H})} \binom{n}{|S|}^{-1}$$

and analogously, without assuming $(k-1)$-linearity,

$$s'(n,k) := \min_{\mathcal{H}=(X,\mathcal{E}),\ |X|=n} \sum_{S\in\mathcal{S}_k(\mathcal{H})} \binom{n}{|S|}^{-1}.$$

Since there exist only finitely many hypergraphs on any given number n of vertices, both $s(n,k)$ and $s'(n,k)$ are well-defined and are at most 1 by the YBLM inequality, for all n and k. In Theorem 8 we prove that for every fixed k, the values of $s(n,k)$ and $s'(n,k)$ tend to constants larger than 0 and smaller than 1 as n gets large. We also consider their relation to the Turán problem on graphs and uniform hypergraphs.

2 Proof of the general lower bound

Here we prove Theorem 6. By the free choice of $c = c(k)$, we may restrict ourselves to n sufficiently large, say $n > k^3$. Moreover, due to the nature

[2] For several decades, it was called LYM inequality, stated and proved in exactly that form independently by Yamamoto [18], Meshalkin [10] and Lubell [9] (in this order of chronology). Bollobás [1] proved a more general result, however, from which the inequality follows immediately. Inequalities of this kind have lots of applications in extremal problems in various areas of mathematics; cf. the two-part survey [16, 17]. The current acronym YBLM coincides (apart from punctuation) with the abbreviated name of famous Hungarian architect Miklós Ybl (1814–1891).

of the problem, we may also assume without loss of generality that $|E| \geq k$ holds for all $E \in \mathcal{E}$. Let $H \in \binom{X}{k}$ be any k-tuple. If it is contained in some $E \in \mathcal{E}$, then H is counted in $\mathcal{E}_k$ precisely once. We will prove that, with possibly few exceptions, also the other k-tuples H generate at least one member of $\mathcal{E}_{k+1}^0$ on the average. More explicitly, it will turn out that most of those sets H can be completed to a member of $\mathcal{E}_{k+1}^0$ in more than k different ways.

From now on we assume that $H \in \binom{X}{k} \setminus \mathcal{E}_k$. Let $x \in H$ be any vertex. If the subset $H \setminus \{x\}$ is contained in an edge of $\mathcal{H}$, we denote that edge by E_x; and otherwise we define $E_x := H \setminus \{x\}$. A more precise and unambiguous notation would be $E_x(H)$, but for simplicity we write E_x as long as just one H is considered. Note that E_x is unique for each $x \in H$ (once H is understood), since $\mathcal{H}$ is $(k-1)$-linear. It also follows for any two distinct $x, x' \in H$ that E_x and $E_{x'}$ share no vertex outside H. We set

$$H^* := \bigcup_{x \in H} E_x.$$

Then we have the implication

$$H \in \binom{X}{k} \setminus \mathcal{E}_k \ \wedge \ z \in X \setminus H^* \quad \Longrightarrow \quad H \cup \{z\} \in \mathcal{E}_{k+1}^0 \tag{2}$$

because the containment relation $(H \cup \{z\}) \setminus \{x\} \subset E'$ for some $x \in H$ and $E' \in \mathcal{E}$ would contradict the assumption $|E_x \cap E'| < k-1$.

We say that the k-tuple H is a *near-cover* of $\mathcal{H}$ if $|H^*| \geq |X| - k$. The proof now splits into two situations, whether $\mathcal{H}$ has, or does not have, a near-cover.

Suppose first that no $H \in \binom{X}{k} \setminus \mathcal{E}_k$ is a near-cover of $\mathcal{H}$. Then by (2) we obtain that each H can be extended to a member F of $\mathcal{E}_{k+1}^0$ in at least $k+1$ different ways. On the other hand, each $F \in \mathcal{E}_{k+1}^0$ can be obtained from exactly $k+1$ sets $H \in \binom{X}{k} \setminus \mathcal{E}_k$, namely from its k-element subsets. Thus, in this case we have

$$|\mathcal{E}_{k+1}^0| \geq \left| \binom{X}{k} \setminus \mathcal{E}_k \right|$$

and the inequality stated in the theorem holds even without the error term $O(n^{k-1})$.

Suppose now that some $H \in \binom{X}{k} \setminus \mathcal{E}_k$ is a near-cover of $\mathcal{H}$. Then the cardinality of the set

$$X' := H^* \setminus H$$

is at least $n-2k$, and X' is partitioned into sets of type $E'_x := E_x \setminus H$ $(x \in H)$. Say, $X' = E'_{x_1} \cup \cdots \cup E'_{x_\ell}$ where $\ell \le k$.

A case that can directly be settled is when some $\ell-1$ sets from $\{E'_{x_1}, \ldots, E'_{x_\ell}\}$ cover together at most k^3-4k^2+5k vertices. There are at most $2k$ vertices outside X', hence some E'_{x_i} contains at least $n-k^3+4k^2-7k$ elements. Then we obtain that $|\mathcal{E}_k| \ge \binom{n-k^3+4k^2-7k}{k} = \binom{n}{k} - O(n^{k-1})$ is valid[3]. Therefore, we may suppose for the rest of the proof that the union of any $\ell-1$ sets from $\{E'_{x_1}, \ldots, E'_{x_\ell}\}$ has cardinality greater than k^3-4k^2+5k.

Consider any $H' \in \binom{X'}{k} \setminus (\mathcal{E}_k \cup \{H\})$. A coincidence $E_{y_0}(H') = E'_{x_i}$ can happen with only one vertex $y_0 \in H$ and only one index i $(1 \le i \le \ell)$, namely when $|H' \cap E'_{x_i}| = k-1$. If this situation occurs, assume that $E_{y_0}(H') = E'_{x_1}$. Then, since $\mathcal{H}$ is $(k-1)$-linear, for any $y \in H' \setminus \{y_0\}$ and for any $i' \ne i$ with $i' > 1$ we have $|E_y(H') \cap E'_{x_{i'}}| \le k-2$, and so $E_y(H')$ meets $E'_{x_2} \cup \cdots \cup E'_{x_\ell}$ in at most $(\ell-1)(k-2) \le (k-1)(k-2)$ vertices, one of which is y_0. Therefore the $k-1$ choices of $y \ne y_0$ cover at most $(k^3-3k^2+k)-(k^2-3k+1)+1 = k^3-4k^2+4k$ vertices of $E'_{x_2} \cup \cdots \cup E'_{x_\ell}$. Hence, the inequality

$$|\left(E'_{x_2} \cup \cdots \cup E'_{x_\ell}\right) \setminus \bigcup_{y \in H'} E_y(H')| \ge k+1$$

follows by $|X'| \ge n-2k$ and by the assumed lower bound on $|E'_{x_2} \cup \cdots \cup E'_{x_\ell}|$. Thus, in this case, H' can be completed to a member of $\mathcal{E}^0_{k+1}$ in at least $k+1$ different ways. The situation is even better if we have $|E_y(H') \cap E'_{x_i}| \le k-2$ for all $1 \le i \le \ell$. Then summing over all $y \in H'$ and all $1 \le i \le \ell$, we obtain the upper bound

$$|(H'^* \setminus H') \cap (H^* \setminus H)| \le k^2(k-2),$$

so that there are at least $|X' \setminus H'^*| \ge n - (k^3-2k^2+3k) \ge k+1$ ways to extend H' to a member of $\mathcal{E}^0_{k+1}$ whenever $n > k^3$ (and $k \ge 2$). Consequently,

$$|\mathcal{E}^0_{k+1}| \ge \left| \binom{X'}{k} \setminus \mathcal{E}_k \right|$$

holds, and therefore $|\mathcal{E}_k| + |\mathcal{E}^0_{k+1}| \ge \binom{n}{k} - O(n^{k-1})$ is valid as n gets large, because $|X'| \ge n-2k$. □

[3] We may actually write the somewhat larger value $\binom{n-k^3+4k^2-7k+k-1}{k}$, by considering E_{x_i} instead of E'_{x_i}; but this is irrelevant concerning the current proof.

3 Geometric upper bound

As we mentioned in the introduction, n points in $\mathbb{R}^d$ may generate as few as $\binom{n-1}{d+1}$ affine simplexes, each of which has more than d points. Here we show that the number of affine simplexes can be even smaller.

Proposition 7 *There is an arrangement of n points in $\mathbb{R}^3$, such that the number of affine simplexes determined by them is only*

- $\binom{n-1}{4} - \frac{(n-2)(n-5)}{2}$ *if n is even,*
- $\binom{n-1}{4} - \frac{(n-3)(n-5)}{2}$ *if n is odd;*

that is, $\frac{1}{24}n^4 - \frac{5}{12}n^3 + O(n^2)$.

Proof. First, let n be even. Take $n-2$ points $x_1, \ldots, x_{n-2}$ on a plane $P \subset \mathbb{R}^3$, such that no three of them are collinear, moreover all the $n/2-1$ lines $\overline{x_{2i-1}x_{2i}}$ are parallel for $i = 1, 2, \ldots, n/2-1$. Let x_{n-1} and x_n be two points outside P, such that the line $\overline{x_{n-1}x_n}$ is parallel to $\overline{x_1x_2}$ (and hence to the other pairs as well). We have the following types of affine simplexes:

- quadruples of points in P;
- quadruples of the form $\{x_{2i-1}, x_{2i}, x_{n-1}, x_n\}$ $(i = 1, 2, \ldots, n/2-1)$;
- quintuples of the form $\{x_a, x_b, x_c, x_{n-1}, x_n\}$ $(1 \le a < b < c \le n-2)$, where $\{2i-1, 2i\} \not\subset \{a, b, c\}$ for any i.

The number of sets of those three types is $\binom{n-2}{4}$, $\frac{1}{2}(n-2)$, and $\frac{1}{6}(n-2)(n-4)(n-6)$, respectively.

If n is odd, we take $(n-3)/2$ pairs of points inside P which determine lines $\overline{x_{2i-1}x_{2i}}$ parallel to $\overline{x_{n-1}x_n}$, plus one point x_{n-2} of P which is not collinear with any two of $x_1, \ldots, x_{n-3}$. Then we have $\binom{n-2}{4}$ affine simplexes inside P, further $\frac{1}{2}(n-3)$ ones of the form $\{x_{2i-1}, x_{2i}, x_{n-1}, x_n\}$, moreover $\frac{1}{6}(n-3)(n-5)(n-7)$ of the form $\{x_a, x_b, x_c, x_{n-1}, x_n\}$ $(1 \le a < b < c \le n-3)$ where $\{2i-1, 2i\} \not\subset \{a, b, c\}$, and finally $\frac{1}{2}(n-3)(n-5)$ of the form $\{x_a, x_b, x_{n-2}, x_{n-1}, x_n\}$ not containing any pair $\{x_{2i-1}, x_{2i}\}$. $\square$

4 The YBLM inequality

Recall from Section 1.4 that $s(n,k)$ and $s'(n,k)$ are defined as the minimum of the sum $\sum_{S\in\mathcal{S}_k(\mathcal{H})}\binom{n}{|S|}^{-1}$ where $\mathcal{H}$ runs over all hypergraphs of order n — with or without assuming $(k-1)$-linearity — and $\mathcal{S}_k(\mathcal{H})=\mathcal{E}_k\cup\mathcal{E}^0_{k+1}$. Here we study the asymptotic behavior of these two functions, and point out a relation to Turán numbers.

4.1 The limits of $s(n,k)$ and $s'(n,k)$

Our goal in this subsection is to prove the following result.

Theorem 8 *For every fixed $k\ge 2$, the limits*

$$s_k := \lim_{n\to\infty} s(n,k) \qquad \text{and} \qquad s'_k := \lim_{n\to\infty} s'(n,k)$$

exist and satisfy

$$0 < s'_k \le s_k < 1$$

with strict inequality at both ends.

We state three assertions below which together will immediately imply the validity of the theorem as the middle inequality holds by definition.

Lemma 9 *For every fixed k, the sequences $(s(n,k))_{n=k+1}^{\infty}$ and $(s'(n,k))_{n=k+1}^{\infty}$ are non-decreasing.*

Proof. For any hypergraph $\mathcal{H}=(X,\mathcal{E})$ on n vertices, let us introduce the notation $m_k:=|\mathcal{E}_k|$ and $m_{k+1}:=|\mathcal{E}^0_{k+1}|$. The inequality

$$\frac{m_k}{\binom{n}{k}}+\frac{m_{k+1}}{\binom{n}{k+1}}\ge b$$

is equivalent to

$$m_{k+1}\ge b\cdot\binom{n}{k+1}-\frac{n-k}{k+1}m_k \tag{3}$$

for any $b>0$. We are going to prove that if the analogue of (3) is valid for every hypegraph on $n-1$ vertices, then it is valid for $\mathcal{H}$ as well. In what follows, assume that it is valid for $n-1$.

Let $x \in X$ be any vertex. We derive the hypergraph $\mathcal{H}_{-x}$ from $\mathcal{H}$ by removing x from all edges and deleting the edges which become smaller than k after cutting out x. Note that $\mathcal{H}_{-x}$ is q-linear whenever so is $\mathcal{H}$ (no matter what q we choose), although the inverse implication is not valid in general.

Let us denote by m_k^x and m_{k+1}^x the values corresponding to m_k and m_{k+1} in $\mathcal{H}_{-x}$, and by $\mathcal{E}_{k[-x]}$ and $\mathcal{E}^0_{k+1[-x]}$ the corresponding families of sets, respectively. We then have

$$m_{k+1}^x \geq b \cdot \binom{n-1}{k+1} - \frac{n-k-1}{k+1}\, m_k^x \tag{3_x}$$

by assumption.

A k-element set F occurs in $\mathcal{E}_{k[-x]}$ if and only if it is contained in some edge of $\mathcal{H}_{-x}$; and this happens precisely when $x \notin F$ and some $E \in \mathcal{E}$ contains F as a subset. Thus, each $F \in \mathcal{E}_k$ gives rise to a member of $\mathcal{E}_{k[-x]}$ for exactly $n-k$ choices of x, and no more sets occur in $\bigcup_{x \in X} \mathcal{E}_{k[-x]}$. Similarly, each $F \in \mathcal{E}^0_{k+1}$ yields a member of $\mathcal{E}^0_{k+1[-x]}$ for exactly $n-k-1$ choices of x, and these are all the sets in $\bigcup_{x \in X} \mathcal{E}^0_{k+1[-x]}$. As a consequence, the equalities

$$\sum_{x \in X} m_k^x = (n-k) \cdot m_k \qquad \text{and} \qquad \sum_{x \in X} m_{k+1}^x = (n-k-1) \cdot m_{k+1}$$

hold. Thus, summing up (3_x) for all $x \in X$ we obtain

$$(n-k-1) \cdot m_{k+1} \geq bn \cdot \binom{n-1}{k+1} - \frac{n-k-1}{k+1}\,(n-k) \cdot m_k$$

which is equivalent to (3). This completes the proof. □

Lemma 10 *For every fixed k, we have $s(k+1,k) = s'(k+1,k) = \frac{1}{k+1}$.*

Proof. The minimum for both $s(k+1,k)$ and $s'(k+1,k)$ is attained by the hypergraph with $k+1$ vertices and with precisely one edge of cardinality k. □

Lemma 11 *For every fixed k, we have $s_k \leq 1 - \frac{k}{2^k}$.*

Proof. To simplify notation, we consider $2n$ vertices instead of n. Consider the hypergraph whose edge set $\mathcal{E}$ consists of just two disjoint sets of cardinality n each. It is linear, of course. Moreover, we clearly have

$$|\mathcal{E}_k| = 2\binom{n}{k} \qquad \text{and} \qquad |\mathcal{E}^0_{k+1}| = \binom{2n}{k+1} - 2\binom{n}{k+1} - 2n\binom{n}{k}$$

because a $(k+1)$-tuple does *not* belong to $\mathcal{E}^0_{k+1}$ if and only if either it is contained in one of the two edges or it meets one of the edges in precisely one vertex and the other edge in k vertices. Thus,

$$\begin{aligned} s(2n,k) &\leq \frac{2\binom{n}{k}}{\binom{2n}{k}} + 1 - \frac{1}{\binom{2n}{k+1}}\left(2\binom{n}{k+1} + 2n\binom{n}{k}\right) \\ &= 1 - \frac{2}{\binom{2n}{k}}\left(\frac{k+1}{2n-k}\left(n\binom{n}{k} + \binom{n}{k+1}\right) - \binom{n}{k}\right) \\ &= 1 - \frac{\binom{n}{k}}{\binom{2n}{k}} \cdot 2 \cdot \left(\frac{n(k+1)}{2n-k} + \frac{n-k}{2n-k} - 1\right) \\ &= 1 - \frac{\binom{n}{k}}{\binom{2n}{k}} \cdot \frac{2nk}{2n-k} \end{aligned}$$

The function in the last line clearly tends to $1 - \frac{k}{2^k}$ as $n \to \infty$, therefore s_k cannot be larger. □

4.2 Turán numbers

For a fixed k-uniform hypergraph $\mathcal{F}$, we use the standard notation $\mathrm{ex}(n, \mathcal{F})$ for its Turán number; that means the maximum number of edges in a k-uniform hypergraph of order n which does not contain any subhypergraph isomorphic to $\mathcal{F}$. Further, let $\mathcal{K}^{(k)}_{k+1}$ denote the hypergraph with $k+1$ vertices and $k+1$ edges of k vertices each (i.e., the complete k-uniform hypergraph of order k). If $k = 2$ then $\mathcal{K}^{(2)}_3$ is just the triangle K_3, the complete graph of order 3. In this very particular case the equality $\mathrm{ex}(n, K_3) = \left\lfloor \frac{n^2}{4} \right\rfloor$ is well known to hold, but for larger k the determination of $\mathrm{ex}(n, \mathcal{K}^{(k)}_{k+1})$ is a famous open problem in extremal hypergraph theory (see, e.g., [14] for a survey and [4] for many further references).

Remark 12 *If $\mathcal{H} = (X, \mathcal{E})$ is a k-uniform hypergraph of order n such that each $(k+1)$-tuple of vertices contains at least one edge of $\mathcal{H}$, then $\mathcal{E}^0_{k+1} = \emptyset$. In particular, taking $\mathcal{H}$ as the complement of a hypergraph extremal for* $\mathrm{ex}(n, \mathcal{K}^{(k)}_{k+1})$*, we obtain:*

$$s'(n,k) \leq 1 - \frac{\mathrm{ex}(n, \mathcal{K}^{(k)}_{k+1})}{\binom{n}{k}}.$$

As a consequence,

$$s'_k \leq 1 - \lim_{n\to\infty} \frac{\mathrm{ex}(n, \mathcal{K}^{(k)}_{k+1})}{\binom{n}{k}}$$

where the limit exists for every fixed k, as proved in [5]. Hence, any lower bound on the Turán density of $\mathcal{K}^{(k)}_{k+1}$ implies an upper bound on s'_k.

Note that an analogous implication in the opposite direction does not work: upper bounds on $\mathrm{ex}(n, \mathcal{K}^{(k)}_{k+1})$ do not imply lower bounds on $s'(n,k)$. On the other hand, applying the results of Sidorenko [13] on $\mathrm{ex}(n, \mathcal{K}^{(k)}_p)$, from the case $p = k+1$ we obtain the following inequality:

Corollary 13 *For every $k \geq 3$ we have $s'_k \leq \left(1 - \frac{1}{k}\right)^{k-1}$.*

It cannot be guaranteed in general that the hypergraphs derived from the extremal ones for the Turán problem lead to constructions of $(k-1)$-linear hypergraphs, hence they cannot automatically imply upper bounds on s_k. But this can be done if $k = 2$, and the following exact formula is valid.

Theorem 14 *For every $n \geq 3$ we have $s(n,2) = s'(n,2) = 1 - \left\lfloor \frac{n^2}{4} \right\rfloor / \binom{n}{2}$, and therefore $s_2 = s'_2 = 1/2$. A hypergraph $\mathcal{H} = (X, \mathcal{E})$ is extremal for $s'(n,2)$ if and only if $\mathcal{E}_2$ is the complementary graph of the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$; and for $s(n,2)$ the extremal hypergraph is unique up to isomorphism.*

Proof. For an upper bound, let $\mathcal{H} = (X, \mathcal{E})$ consist of n vertices and two vertex-disjoint edges E_1, E_2 with $|E_1| = \lfloor n/2 \rfloor$ and $|E_2| = \lceil n/2 \rceil$. Then $|\mathcal{E}_2| = \binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor$ and $\mathcal{E}^0_3 = \emptyset$. Since $\mathcal{H}$ is 1-linear, the upper bound follows for both $s(n,2)$ and $s'(n,2)$. From the argument below, it will also turn out that this is the unique 1-linear hypergraph attaining equality.

To prove the lower bound, let $\mathcal{H} = (X, \mathcal{E})$ be any hypergraph. Note that $\mathcal{E}_2$ is just a graph; we denote its *complement* by $G = (X, E)$, i.e. an unordered vertex pair $x_i x_j$ belongs to the edge set E of G if and only if $\{x_i, x_j\} \notin \mathcal{E}_2$. Then $\mathcal{E}^0_3$ is the family of triangles (K_3-subgraphs) in G. As long as G is triangle-free, we have $|\mathcal{E}_2| \geq \binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor$ and the lower bound follows for $\mathcal{H}$, with equality if and only if $G \simeq K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. Assuming that $|\mathcal{E}_2|$ is smaller, we have $|E| > n^2/4$.

We write the number $|E|$ of edges in G in the form $m = \frac{n^2}{4} + \ell$; hence $\ell \geq 1$ is an integer if n is even, and $\ell + \frac{1}{4} \geq 1$ is an integer if n is odd. It is well known that G has at least $\frac{4m^2 - mn^2}{3n}$ triangles [11, 12]. Thus,

$$|\mathcal{E}_3^0| \geq \frac{4m}{3n}\left(m - \frac{n^2}{4}\right) = \left(\frac{n}{3} + \frac{4\ell}{3n}\right)\ell > \frac{n\ell}{3}$$

and consequently

$$\begin{aligned}
\frac{|\mathcal{E}_2|}{\binom{n}{2}} + \frac{|\mathcal{E}_3^0|}{\binom{n}{3}} &> \frac{\binom{n}{2} - \frac{n^2}{4} - \ell}{\binom{n}{2}} + \frac{\frac{n\ell}{3}}{\binom{n}{3}} \\
&= \frac{\binom{n}{2} - \frac{n^2}{4}}{\binom{n}{2}} + \frac{\frac{n\ell}{3} - \frac{(n-2)\ell}{3}}{\binom{n}{3}} \\
&\geq \frac{\binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor}{\binom{n}{2}} + \frac{2\ell}{3\binom{n}{3}} - \frac{1}{4\binom{n}{2}}.
\end{aligned}$$

This proves the stated inequality for all $\ell \geq (n-2)/8$. Moreover, if n is even, the theorem follows for all $\ell > 0$ because in that case we need not subtract $1/4$ when moving from $\frac{n^2}{4}$ to $\left\lfloor \frac{n^2}{4} \right\rfloor$.

For the rest of the proof, we assume that n is odd and $0 < \ell < (n-2)/8$. Since ℓ is relatively small, G must contain some vertex x of degree at most $\frac{n-1}{2}$, for otherwise the number of edges would be at least $\frac{n(n+1)}{4}$, yielding the contradiction $\ell \geq \frac{n}{4}$. Let now $\ell' = \ell + \frac{1}{4}$, that means $m = \left\lfloor \frac{n^2}{4} \right\rfloor + \ell'$, and consider the graph $G' := G - x$. This G' has $n' = n - 1$ vertices and at least

$$m' := \frac{n^2}{4} + \ell - \frac{n-1}{2} = \frac{(n-1)^2 + 1}{4} + \ell = \frac{(n')^2}{4} + \ell'$$

edges. Therefore, by the theorem cited above, G' contains at least

$$\frac{4m'}{3n'}\left(m' - \frac{(n')^2}{4}\right) = \left(\frac{n-1}{3} + \frac{4\ell'}{3n'}\right)\ell' > \frac{(n-1)\ell'}{3}$$

triangles, which certainly is a lower bound on $|\mathcal{E}_3^0|$, too. Thus, with a slight modification of the computation above, we obtain that

$$\frac{|\mathcal{E}_2|}{\binom{n}{2}} + \frac{|\mathcal{E}_3^0|}{\binom{n}{3}} > \frac{\binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor - \ell'}{\binom{n}{2}} + \frac{\frac{(n-1)\ell'}{3}}{\binom{n}{3}}$$

$$
\begin{aligned}
&= \frac{\binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor}{\binom{n}{2}} + \frac{\frac{(n-1)\ell'}{3} - \frac{(n-2)\ell'}{3}}{\binom{n}{3}} \\
&= \frac{\binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor}{\binom{n}{2}} + \frac{\ell'}{3\binom{n}{3}}.
\end{aligned}
$$

This completes the proof of the theorem. □

5 Concluding remarks

Motivated by a problem arisen in chemistry/stoichiometry, we established asymptotically tight extremal results on geometric point sets and on finite set systems. Below we formulate some problems and conjectures that remain open.

Geometry vs. hypergraph theory. We proved matching asymptotic lower and upper bounds of the form $\binom{n}{d+1} - \Theta(n^d)$ on the minimum number of affine simplexes, for every set of n points in $\mathbb{R}^d$ not containing any affine simplex of fewer than $d+1$ points. Our method was to put the problem in a more general context and to estimate an extremal function for a class of hypergraphs (called q-linear, implying the solution for geometric sets when $q = d$). There remains a gap of order $\Theta(n^d)$, however, between the lower and upper bounds.

Problem 15 *Given the integers n and d, determine the minimum number of affine simplexes generated by n points in $\mathbb{R}^d$, no d of which lie on a $(d-2)$-dimensional hyperplane.*

Problem 16 *Given the integers n and k, determine the minimum value of*

$$|\mathcal{E}_k| + |\mathcal{E}^0_{k+1}|$$

taken over all $(k-1)$-linear hypergraphs $\mathcal{H} = (X, \mathcal{E})$ on n vertices.

Problem 17 *For which values of $k = d+1$ is the minimum for hypergraphs in Problem 16 equal to that for point sets in $\mathbb{R}^d$ in Problem 15, for all $n > n_0(d)$?*

For the case of $d = 2$, it was proved in [8] that the minimum number of affine simplexes in $\mathbb{R}^2$ determined by $n \geq 8$ points is attained by placing the points on two lines: one of the lines contains $n - 2$ of the points and the other line contains 3 points (and so their intersection point is also selected). That is, the minimum for Problem 15 with $d = 2$ is

$$\binom{n-2}{3} + \binom{n-3}{2} + 1.$$

The construction implies the same upper bound for Problem 16 with $k = 3$, attained by the linear (that is, 2-linear) hypergraph with n vertices and two edges, one of size $n - 2$ and the other of size 3. Moreover, the proof of the matching lower bound in [8] gets through for linear hypergraphs as well, since it only applies modifications in the incidence structure, without any particular geometric assumptions. Thus, the minimum is the same for Problem 15 with $d = 2$ and Problem 16 with $k = 3$.

It is not clear, however, whether the answer to Problem 17 is positive or negative for $d \geq 3$. We note that the extremal construction of [15] cannot be applied for our problem to derive an upper bound on affine simplexes in $\mathbb{R}^3$, because in [15] the points are arranged in two (equal or nearly equal) collinear sets. Nevertheless, the following conjecture looks easier than the exact determination of minimum.

Conjecture 18 *For every $k \geq 3$ there exists a hypergraph $\mathcal{H}$ of order n which is extremal for Problem 16 and has $O(n^{k-3})$ edges as n gets large.*

We note further that the upper bound on s'_k in Corollary 13 tends to zero as k gets large, and at present we do not have any geometric constructions with the same property for s_k.

Conjecture 19 *There exists an integer k_0 such that, for every $k \geq k_0$, we have $s'_k < s_k$.*

Perhaps the guess $k_0 = 3$ is too brave, but we cannot disprove even that at present.

Stoichiometry. For the original problem originating from [6] in stoichiometry, our Theorem 3 and Corollary 4 imply:

- There are at least $\binom{n}{d+1} - O\left(n^d\right) \approx \frac{n^{d+1}}{(d+1)!}$ minimal reactions among n species if the species are built up from d kinds of atoms (atomic particles), and if the number of species (molecules) forming any minimal reaction must be greater than d.

- For the first case previously unsolved, namely $d = 3$, the asymptotically tight lower bound is $\binom{n}{4} - O\left(n^3\right) \approx \frac{n^4}{24}$, if reactions with three or fewer species are not possible. (Especially parallel species, i.e. multiple doses are also excluded.)

References

[1] B. Bollobás, *On generalized graphs*, Acta Math. Acad. Sci. Hungar. 16 (1965), 447–452.

[2] Gy. Dósa, C. Laflamme and I. Szalkai, *On the maximal and minimal number of bases and simple circuits in matroids and the extremal constructions*, Preprint 046, Dept. Math. Univ. Veszprém, 1997.

[3] Gy. Dósa, I. Szalkai and C. Laflamme, *On the maximal and minimal number of bases and simple circuits in matroids and the extremal constructions*, Pure Math. & Appl. (PUMA) 15 (2006), 383–392.

[4] Z. Füredi and M. Simonovits, *The history of degenerate (bipartite) extremal graph problems*, in: Erdős Centennial (L. Lovász et al., Eds.), Bolyai Society Mathematical Studies 25 (2013), 169–264.

[5] Gy. Katona, T. Nemetz and M. Simonovits, *Újabb bizonyítás a Turán-féle gráftételre és megjegyzések bizonyos általánosításaira*, Mat. Lapok 15 (1964), 228–238. (in Hungarian)

[6] S. Kumar and Á. Pethő, *Note on a combinatorial problem for the stoichiometry of chemical reactions*, Intern. Chem. Eng. 25 (1985), 767–769.

[7] C. Laflamme and I. Szalkai, *Counting simplexes in* $\mathbb{R}^n$, Hung. J. Ind. Chem. 23 (1995), 237–240.

[8] C. Laflamme and I. Szalkai, *Counting simplexes in* $\mathbb{R}^3$, Electron. J. Combin. 5 (1) (1998), #R40, 11 pp. Printed version in: J. Combin. 5 (1998), 597–607.

[9] D. Lubell, *A short proof of Sperner's lemma*, J. Combin. Th. 1 (1966), 299.

[10] L. D. Meshalkin, *A generalization of Sperner's theorem on the number of subsets of a finite set*, Teor. Veroiatn. Primen. 8 (1963), 219–220. (in Russian)

[11] J. W. Moon and L. Moser, *On a problem of Turán*, Magyar Tud. Akad. Mat. Kut. Int. Közl. 7 (1962), 283–286.

[12] E. A. Nordhaus and B. M. Steward, *Triangles in an ordinary graph*, Canad. J. Math. 15 (1963), 33–41.

[13] A. F. Sidorenko, *Systems of sets that have the T-property*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. (1981), 19–22.

[14] A. F. Sidorenko, *What we know and what we do not know about Turán numbers*, Graphs Combin. 11 (1995), 179–199.

[15] B. Szalkai and I. Szalkai, *Counting minimal reactions with specific conditions in* $\mathbb{R}^4$, J. Math. Chem. 49 (2011), 1071–1085.

[16] Zs. Tuza, *Applications of the set-pair method in extremal hypergraph theory*, "Extremal Problems for Finite Sets" (P. Frankl et al., eds.), Bolyai Society Mathematical Studies 3, 1994, 479–514.

[17] Zs. Tuza, *Applications of the set-pair method in extremal problems, II.*, "Combinatorics, Paul Erdős is Eighty" (D. Miklós et al., eds.), Bolyai Society Mathematical Studies 2, 1996, 459–490.

[18] K. Yamamoto, *Logarithmic order of free distributive lattices*, J. Math. Soc. Japan 6 (1954), 343–354.