On 0-1 polytopes and the Erdős-Szekeres theorem

Attila Pór

Eötvös Lóránd Tudományegyetem Természettudományi Kar

Matematika doktori program

Programvezető: Laczkovich Miklós

Témavezető: Bárány Imre
Tudományos fokozata: A Matematika Doktora

Készült: Rényi Alfréd Matematikai Kutatóintézet

Contents

1 Bevezetés magyarul

A doktori disszertáció három cikkre épül, egy a véletlen 0-1 politópokról szól [BP01] és kettő az Erdős-Szekeres tétel témakörében mozog [PV01], [P02]. Az első ránézésre egymástól távol álló témaköröket a kombinatorika és a geometria közös határa köti össze.

Az első részben véletlen 0-1 politópokkal foglalkozunk, amely néhány n-dimenziós 0-1 vektor konvex burka. Ezen politópok tulajdonságai, főleg a struktúráltak, igen fontos szerepet kapnak a kombinatorikus optimalizációban, ahol gyakran a politóp lapjainak egy teljes, illetve rövid leírása a cél. A 0-1 oplitópok több osztályára is kiderült, hogy ez a feladat igen nehéz, mint például az utazó ügynök politóp [GP], [ABCC] és a vágáspolitóp [DL] esetében. Még arra az ártatlannak tűnő kérdésre sem tudunk felelni, hogy "Hány lapja van az utazóügynök politópnak? vagy a vágáspolitópnak?". K. Fukuda és G. M. Ziegler több előadáson és cikkben hívták fel a figyelmet erre a szép és fontos problémára és kérdeztek rá az n-dimenziós 0-1 politópok lapszámának egy jó felső becslésére. Jelölje $g(n)$ ezt a maximumot. Majdhogynem elemi, miszerint $2n!$ a $g(n)$ egy felső becslése. A felső becslések javítása mellett az alsó becslések terén az exponenciális, hasonló kaptafára gyártott, becslések voltak a legjobbak. Bárány Imrével közös eredményben beláttuk hogy $g(n)$ szuperexponenciálisan nő

Tétel *Létezik egy pozitív c konstans, hogy*

$$g(n) > \left(\frac{cn}{\log n}\right)^{\frac{n}{4}}.$$

Az alsó becslést adó konstrukció véletlen. Az eredmény alapötlete Dyer, Füredi és McDiarmid [DFM] egy gyönyörű eredményéből való. Ők azt vizsgálják, hogy az N csúcsú véletlen politóp, ami az egyszerűbb számolhatóság kedvéért egy ± 1 politóp, mely N érték esetén tölti ki a ± 1 kockát, azaz lesz a térfogata lényegében a teljes ± 1 kocka térfogata. Itt kiderül, hogy a véletlen politópok egy jól leírható felület közelében vannak. Kiderül, hogy tetszőleges $c_4 \leq 1$ és $c_5 \geq 1$ mellett a következő tartományban tudjuk jól kezelni a véletlen politópokat

$$(*) \qquad \exp\{c_4(\log n)^2\} < N < \exp\left\{c_5\frac{n}{\log n}\right\}.$$

Egy kis kitérővel a $(*)$ tartományban egy a korábbi felső becslést javító becslést adunk

Tétel *Bármely N csúcsú n-dimenziós 0-1 politóp lapjainak száma legfeljebb*

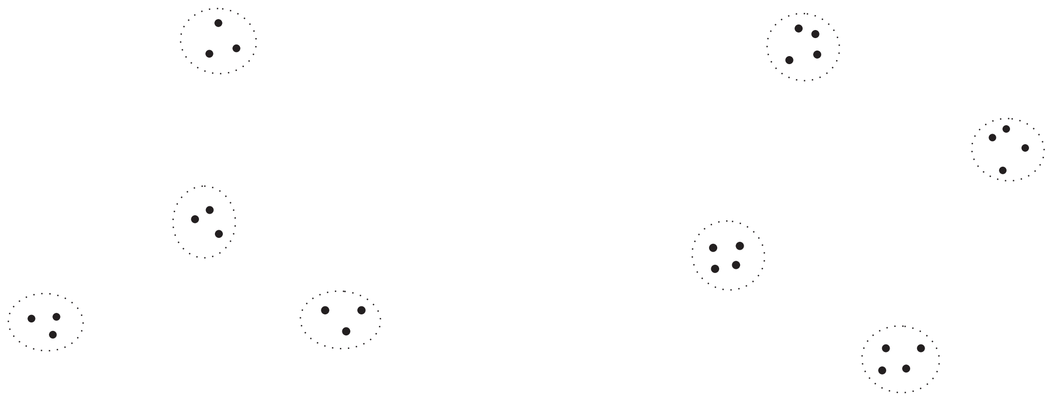
$$\left(c_8 n \log \frac{N}{n}\right)^{\frac{n}{2}}$$

A doktori disszertáció második részében az Erdős - Szekeres tétellel és annak szerteágazó módosításaival illetve általánosításaival foglalkozunk. Ez a fejezet két cikkre épül, az egyik egy önálló eredmény, a másik Pavel Valtr-rel közös munka.

A sík egy ponthalmazára azt mondjuk hogy *általános helyzetben* vannak, ha semmelyik három pont nem esik egy egyenesre. Egy véges ponthalmazra a síkon azt mondjuk hogy *konvex helyzetben* vannak ha egy konvex sokszög csúcsainak halmaza. A rész alaptétele Erdős és Szekeres [ES35] híres tétele:

Tétel (Erdős–Szekeres) *Bármely $k \geq 3$ esetén létezik egy legkisebb egész szám $f(k)$, hogy a sík bármely legalább $f(k)$ általános helyzetű pontja közt van k, melyek konvex helyzetben vannak.*

Egy (a, b, c) ponthármas orientációja a síkon pozitív, ha az általuk alkotott háromszög csúcsai az óramutató járásával megegyezően körbejárva a sorrend abc, ellenkezőleg negatív. Két rendezett t elemű ponthalmaz $(p_1, \dots, p_t)$ és $(p'_1, \dots, p'_t)$ azonos elrendezésűek, ha bármely i, j, l választás esetén a (p_i, p_j, p_l) és (p'_i, p'_j, p'_l) ponthármas orientációja azonos. A sík egy X ponthalmaza egy *k-osztály*, ha az $X_1, \dots, X_k$ azonos méretű halmazok diszjunkt uniója és az $(x_1, \dots, x_k), x_1 \in X_1, \dots, x_k \in X_k$ pont k-asok mindegyike azonos elrendezésű. Ha még az is teljesül, hogy $x_1, \dots, x_k$ konvex helyzetben vannak akkor X egy *konvek k-osztály* (1. ábra).



(non-convex) 4-clustering
(nem konvex) 4-osztály

convex 4-clustering
konvex 4-osztály

Fig.1

A második rész egyik fő eredménye a particiós Erdős–Szekeres tétel.

Tétel (A particiós Erdős–Szekeres tétel) *Bármely $k \geq 3$ esetén léteznek $c = c(k), c' = c'(k)$ konstansok melyekre teljesül a következő. Ha X a sík egy véges általános helyzetű ponthalmaza, akkor létezik egy legfeljebb c' méretű X' részhalmaza, hogy az $X \setminus X'$ ponthalmaz felbontható legfeljebb c darab k-osztályra.*

Az előző tételre a $k = 4$ esetben független eredményt igazolunk:

Tétel *Ha X a sík egy általános helyzetű ponthalmaza, akkor létezik egy legfeljebb c' méretű X' részhalmaza, hogy $X \setminus X'$ ponthalmaz felbontható legfeljebb 26 konvex k-osztályra.*

Nézzük meg mi történik, ha particiós tételből elhagyjuk a "konvex" feltételt, azaz már csak egy nagy k-osztályt keresünk. Ebben az esetben nem csak síkra, de tetszőleges dimenzióban I. Bárány és P. Valtr [BV98] igazolták az állítás pozitív százalékos változatát, "Same Type Lemma" néven.

Tétel (Same Type Lemma) *Bármely $m, d \geq 1$ természetes számokhoz létezik egy $c(m, d)$ konstans, hogy ha $X_1, \dots, X_m \subset \mathbb{R}^d$ véges ponthalmazok és $\bigcup_i X_i$ általános helyzetű, akkor vannak $Y_i \subset X_i, |Y_i| \geq c(m, d)|X_i|$ részhalmazok, melyekre az $Y_1, \dots, Y_m$ halmazok bármely transzverzálisa azonos elrendezésű.*

Az, hogy k azonos méretű halmaz azonos elrendezésű, éppen azt jelenti, hogy az uniójuk egy k-osztály. Ez igazából egy szeparáltságot jelent, amit a dolgozatban három különböző módon is definiálunk. Bevezetünk egy új fogalmat a mátrix-particiót (μ-partició). Egy μ-particiót szeparáltnak nevezünk, ha tetszőleges hipersík az egy oszlopban lévő elemek közül legfeljebb d darabot metsz, továbbá kiegyensúlyozottnak nevezzük a μ-particiót ha az egy oszlopban lévő ponthalmazok mérete azonos (pontos definiciót lásd később). Ezek segítségével tudjuk megfogalmazni a Same Type Lemma particiós változatát.

Tétel *Bármely $m, d \geq 1$ természetes számokhoz létezik egy legkisebb $n_{m,d}$ természetes szám, hogy ha $X_1, \dots, X_m \subset \mathbb{R}^d$ azonos méretű halmazok és az uniójuk általános helyzetű, akkor létezik a halmazok egy kiegyensúlyozott szeparált μ-particiója legfeljebb $n_{m,d}$ oszloppal.*

A fenti tétel bizonyítása során az egy dimenziós esetben, $n_{m,1}$-re egy exponenciális felső becslést kapunk. A még oly egyszerűnek tűnő probléma ellenére sem sikerült a pontos értéket meghatározni. Azonban adunk egy kvadratikus alsó és felső becslést az $n_{m,1}$ értékére:

Tétel

$$\left\lfloor \frac{m-1}{2} \right\rfloor \cdot \left\lfloor \frac{m+1}{2} \right\rfloor + 1 \leq n_{m,1} \leq (m-1)^2 + 1$$

2 On 0-1 polytopes with many facets

2.1 Introduction

A 0-1 polytope is, by definition, the convex hull of some 0-1 vectors from n-space. Properties of 0-1 polytopes, especially structured ones, play an important role in combinatorial optimization where the target is, quite often, a complete or concise description of the facets of the polytope. This task turned out to be difficult for several classes of 0-1 polytopes, most notably for the traveling salesman polytope [GP], [ABCC] and for the cut polytope [DL]. We don't know for instance the answer to the innocent question "How many facets has the traveling salesman polytope? or the cut polytope?" It was K. Fukuda and G. M. Ziegler who, in several lectures and papers [F], [KRSZ], [Z] have drawn attention to this attractive and important problem, and asked for good estimates for the maximum number of facets an n-dimensional 0-1 polytope can have. Write $g(n)$ for this maximum. It is almost elementary to see that $2n!$ is an upper bound for $g(n)$. Stronger is the result of Fleiner, Kaibel, and Rote [FKR]:

$$g(n) \leq 30(n-2)!$$

for large enough n. Using the blowing up technique [KRSZ] and T. Christof's construction of a 13-dimensional 0-1 polytope with more than 3.6^{13} facets, it can be shown that, again for large enough n,

$$g(n) > 3.6^n.$$

Earlier, Fukuda gave a similar example with 3.26^n facets (see [KRSZ]), based on the computational experience [F] concerning the behavior of the number of facets in random 0-1 polytopes, as the number of vertices changes. Imre Bárány and me proved that $g(n)$ grows superexponentially:

Theorem 2.1. *There is a positive constant c such that*

$$g(n) > \left(\frac{c\,n}{\log n}\right)^{\frac{n}{4}}.$$

The construction giving this lower bound is random. It is perhaps instructive to see here how the number of facets of random polytopes behave. The best analogy comes from the random polytope $P_N = P_N^n$ whose vertices $v_1, \dots, v_N$ are chosen randomly, independently and uniformly from the sphere S^{n-1}. The expected number of facets, $E[f_{n-1}(P_N)]$, of P_N is asymptotically const$N$ when n is fixed and $N \to \infty$. But here we are interested in the case when $n \to \infty$ and $N < 2^n$. There is a simple formula in Buchta et al. [BMT] which can be used to show that in the range $2n < N < 1.5^n$, say,

$$\left(c_1 \log \frac{N}{n}\right)^{\frac{n}{2}} < E[f_{n-1}(P_N)] < \left(c_2 \log \frac{N}{n}\right)^{\frac{n}{2}}$$

with suitable positive constants c_1 and c_2.

Note that a 0-1 polytope has all of its vertices on a sphere. It is tempting to believe that a random 0-1 polytope, K_N, on N vertices behaves similarly. This may be even true for arbitrary 0-1 polytopes as well. In particular, it seems likely that

$$g(n) < (c_3 n)^{\frac{n}{2}}.$$

Here and throughout the first part $c, c_1, \dots$ and $b, b_1, \dots$ denote positive constants that are independent of n and N.

2.2 The model, the result, the idea

Write $C = C^n = [-1,1]^n$ for the n-dimensional ± 1 cube. (This is more convenient to work with.) Let Z be a random variable distributed uniformly over $\{-1,1\}$, and let $Z_1, \dots, Z_n$ be independent random variables each distributed like Z. Set $\underline{Z} = (Z_1, \dots, Z_n)$. Thus $\underline{Z}$ is uniformly distributed over the 2^n vertices of C. Take N independent copies of $\underline{Z}$, namely $\underline{Z}_1, \dots, \underline{Z}_N$ and define

$$K_N = \operatorname{conv}\{\underline{Z}_1, \dots, \underline{Z}_N\} \ ,$$

the convex hull of the vectors $\underline{Z}_1, \dots, \underline{Z}_N$. This is going to be our random 0-1, or rather ± 1, polytope on N vertices. Note, however, that some vertices may be repeated. (K_N is one of the usual models of random ± 1 polytopes.)

We can state our main result now. Assume

$$(*) \qquad \exp\{c_4 (\log n)^2\} < N < \exp\left\{c_5 \frac{n}{\log n}\right\}.$$

Here one can take any constants $c_4 \le 1$ and $c_5 \ge 1$.

Theorem 2.2. *Under condition* $(*)$

$$E[f_{n-1}(K_N)] > (c_6 \log N)^{\frac{n}{4}}.$$

If the expected number of facets is large, then, of course, there has to be an example where the number of facets is large. We will, in fact, prove this stronger statement in a form that implies Theorem 2.2.

Theorem 2.3. *Under condition $(*)$, there exists a polytope K_N with*

$$f_{n-1}(K_N) > (c_7 \log N)^{\frac{n}{4}}.$$

The proof of this result is based on several lemmas, some of them quite involved. So we first present the basic idea, which is simple, rather informally. Assume $x \in C$ and define

$$p(x, N) = \text{Prob}[x \in K_N].$$

General principles would tell that, for most $x \in C$, $p(x, N)$ is either close to one or close to zero. To be more specific, set

$$P(t) = \{x \in C : p(x, N) \geq t\}.$$

Our approach is based on the fact that for all small $\varepsilon > 0$ and large n $P(1-\varepsilon) \subset P(\varepsilon)$, of course, but the drop from $1-\varepsilon$ to ε is very abrupt: $P(\varepsilon)$ is in a small neighbourhood of $P(1-\varepsilon)$. This shows that $P(1-\varepsilon) \subset K_N$ with high probability. But only a tiny fraction of K_N lies outside $P(\varepsilon)$: most of the boundary of $P(\varepsilon)$ is outside K_N. Thus most of the boundary of $P(\varepsilon)$ is cut off by facets of K_N. These facets lie outside $P(1-\varepsilon)$. Comparing the surface area of $P(\varepsilon)$ with the amount a facet can cut off from it gives the lower bound.

But how to find the sets $P(1-\varepsilon)$ and $P(\varepsilon)$? This is the point where we extensively use a beautiful result of Dyer, Füredi, and McDiarmid [DFM]. Their target was to determine the threshold $N = N(n)$ such that K_N contains most of the volume of C. As they prove, this happens at $N = (\frac{2}{\sqrt{e}})^n$. Their method describes where $p(x, N)$ drops from one to zero as $n \to \infty$ and $N = e^{\alpha n}$. The analysis carries over for other values of N. In our case higher precision is required as we need a good estimate on how fast $p(x, N)$ drops from one to zero. We were able to control this only where the curvature of the boundary of $P(\varepsilon)$ behaves nicely. This is perhaps the spot where the exponent $n/2$ for P_N (the random spherical polytope) is lost and we only get $n/4$ for K_N.

In the next section we give another upper bound on the number of facets of a 0-1 polytope. Then we state four lemmas related to $p(x, N)$. After some geometric background, together with the proof of Theorems 2.2 and 2.3 we prove the probabilistic lemmas and geometric lemmas.

2.3 Another upper bound on the numbers of facets of a 0-1 polytope

In the range given by condition $(*)$ we can improve the bound of Fleiner, Kaibel, and Rote [FKR]. In fact, the bound below is better as long as N is less than exponential in n.

Theorem 2.4. *Every n-dimensional 0-1 polytope with N vertices has at most*

$$\left(c_8 n \log \frac{N}{n}\right)^{\frac{n}{2}}$$

facets.

Proof. We are going to use the following volume estimate from [BáF] and [CP]. Given points $x_1, \ldots, x_N$ from B^n, the euclidean unit ball of R^n,

$$\frac{\text{vol conv}\{x_1, \ldots, x_N\}}{\text{vol} B^n} \leq \left(\frac{c_0}{n} \log \frac{N}{n}\right)^{\frac{n}{2}},$$

where c_0 is a universal constant.

Now let $z_1, \ldots, z_N$ be some vertices of the cube. Define the polytope P as $P = \text{conv}\{z_1, \ldots, z_N\}$. Let π_i stand for the projection onto the subspace $x_i = 0$. Note that all the vertices of $\pi_i(P)$ lie in an $(n-1)$-dimensional ball of radius $\sqrt{n-1}$ (actually, on its boundary). The above estimate gives then

$$\frac{\text{vol}_{n-1} \pi_i(P)}{\text{vol}_{n-1} \sqrt{n-1} B^{n-1}} \leq \left(\frac{c_0}{n-1} \log \frac{N}{n-1}\right)^{\frac{n-1}{2}}.$$

Let $L_1, \ldots L_m$ be the facets of P. Note that $\text{vol}_{n-1} \pi_i(L_j)$ cannot be zero for all i, and it is at least $1/(n-1)!$ if it is nonzero. So summing the equalities $\sum_1^m \text{vol}_{n-1} \pi_i(L_j) = 2\text{vol}_{n-1} \pi_i(P)$ for all i we get

$$\begin{aligned}
\frac{m}{(n-1)!} &\leq \sum_{i=1}^{n} \sum_{j=1}^{m} \text{vol}_{n-1} \pi_i(L_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{vol}_{n-1} \pi_i(L_j) \\
&= 2 \sum_{1}^{n} \text{vol}_{n-1} \pi_i(P) \\
&\leq 2n \text{vol}_{n-1} \sqrt{n-1} B^{n-1} \left(\frac{c_0}{(n-1)} \log \frac{N}{n-1}\right)^{\frac{n-1}{2}}.
\end{aligned}$$

The estimate in the theorem follows now readily. □

2.4 Results of Dyer, Füredi and McDiarmid

From now on we will denote vectors (or points) by underlining in order to distinguish them from scalars. (We actually used this notation for the random vertex $\underline{Z}$.) So given a vector $\underline{x} \in C$ define

$$q(\underline{x}) = \inf\{\text{Prob}[\underline{Z} \in H] : \underline{x} \in H, H \text{ a halfspace}\}$$

and for $\beta > 0$ define

$$Q^\beta = \{\underline{x} \in C : q(\underline{x}) \geq \exp\{-\beta n\}\}.$$

Note that Q^β is a convex polytope. In fact, it is the k-core of the vertices of C (with $k = 2^n e^{-\beta n}$), see [BP] and [E]. We introduce the function

$$f(x) = \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x),$$

defined for $x \in (-1,1)$; at $x = \pm 1$ the limit exists and equals $\log 2$. For $\underline{x} = (x_1, \dots, x_n)$ we set

$$F(\underline{x}) = \frac{1}{n}\sum_1^n f(x_i).$$

Again, for positive β we define

$$F^\beta = \{\underline{x} \in C : F(\underline{x}) \leq \beta\}.$$

f and consequently F are nicely behaving, strictly convex functions whose connection to K_N will become clear soon. To explain how q and F are related we are going to show, following [DFM],

Lemma 2.1. *For $\underline{x} \in (-1,1)^n$, we have $q(\underline{x}) \leq \exp\{-nF(\underline{x})\}$.*

Proof. (from [DFM]) Check, first, that $K(t)$, the so-called cumulant generating function equals

$$K(t) = \log E[\exp\{tZ\}] = \log\cosh t.$$

Then $K'(t) = \tanh t$ and for each $x \in (-1,1)$ there is a unique t with $x = K'(t) = \tanh t$, and

$$t = h(x) = \frac{1}{2}\log\frac{1+x}{1-x}.$$

Note, further, that

$$f(x) = -K(h(x)) + xh(x) \text{ and } h(x) = f'(x).$$

Assume now that $F(\underline{x}) = \beta$ ($\beta > 0$). Then $\underline{x}$ is on the boundary of F^β. In order to estimate $q(\underline{x})$ we need to find a halfspace H of the form $\{\underline{z} : \underline{t}(\underline{z} - \underline{x}) \geq 0\}$ with $\text{Prob}[\underline{Z} \in H]$ as small as possible. Consider the halfspace $H(\underline{x})$ (with $\underline{0} \notin H(\underline{x})$) whose bounding hyperplane is tangent to F^β at $\underline{x}$. So

$$H(\underline{x}) = \{\underline{z} : \underline{t}(\underline{z} - \underline{x}) \geq 0\}$$

with $t_j = f'(x_j)$ $j = 1, \dots, n$. Markov's inequality says $\text{Prob}[X \geq 0] \leq E[e^X]$. Using this

$$\begin{aligned} q(\underline{x}) &\leq \text{Prob}[\underline{Z} \in H(\underline{x})] = \text{Prob}\left[\sum_{j=1}^{n} t_j(Z_j - x_j) \geq 0\right] \\ &\leq E\left[\exp\left\{\sum_{j=1}^{n} t_j(Z_j - x_j)\right\}\right] = \prod_{j=1}^{n} E[\exp\{t_j(Z_j - x_j)\}] \\ &= \prod_{j=1}^{n} \exp\{K(t_j) - x_j t_j\} = \exp\left\{-\sum_{j=1}^{n}(x_j t_j - K(t_j))\right\} \\ &= \exp\{-nF(\underline{x})\}. \end{aligned}$$

□

It is surprising that this trivial estimate is sharp. Dyer, Füredi, and McDiarmid show, for certain values of $\underline{x}$, that

$$q(\underline{x}) \geq \exp\{-n(F(\underline{x}) + \Delta)\}$$

with Δ "small". We will make this statement quantitative in Lemma 2.3.

We have to set a few parameters next. Let α be defined as

$$\alpha = \frac{\log N}{n} \quad \text{or} \quad N = e^{\alpha n}.$$

Then condition $(*)$ reads as

$$c_4 \frac{(\log n)^2}{n} < \alpha < c_5 \frac{1}{\log n}.$$

We will need several small ε_i that are all of the form (with constant $b_i > 0$)

$$\varepsilon_i = b_i \sqrt{\frac{\alpha}{n}} = b_i \frac{\sqrt{\log N}}{n}.$$

The main discovery of Dyer, Füredi, and McDiarmid is that Q^α and F^α are close to each other and both of them approximate K_N quite well as $N = e^{\beta n}$ (β a constant) and $n \to \infty$. We will use several results from [DFM]. The next one is essentially part (b) of Lemma 2.1 of [DFM].

Lemma 2.2. *For large enough n*

$$\text{Prob}[Q^{\alpha - \varepsilon_1} \subset K_N] > 0.99.$$

Define $C^* = \frac{1}{10}C$, this is a shrunk copy of C. Dyer, Füredi and McDiarmid prove (the proof is hard) that $F^\beta \subseteq Q^\alpha$ for every $\alpha > \beta$ if n is large. We make this statement quantitative within C^*.

Lemma 2.3. *For large enough n*

$$F^{\alpha-\varepsilon_2} \cap C^* \subseteq Q^{\alpha-\varepsilon_1}.$$

The next result is simple and is related to part (a) of Lemma 2.1 [DFM]

Lemma 2.4. *For large enough n, at least half of the surface area of $F^{\alpha+\varepsilon_3}$ lying in C^* is missed by K_N with probability at least 0.99.*

One of our targets will be achieved once the last three Lemmas have been proved. Namely, the part of K_N lying in C^* is weakly sandwiched between $F^{\alpha-\varepsilon_2}$ and $F^{\alpha+\varepsilon_3}$ with high probability. Here "weakly sandwiched" means that

$$F^{\alpha-\varepsilon_2} \cap C^* \subseteq K_N$$

and K_N misses half of $C^* \cap \partial F^{\alpha+\varepsilon_3}$.

2.5 Geometric lemmas, proof of Theorem 2.3

We will need some geometric properties of

$$C^* \cap \partial F^\beta$$

where $\beta = \alpha \pm \varepsilon_i$.

Lemma 2.5. $\mathrm{vol}_{n-1}(C^* \cap \partial F^\beta) \geq \frac{1}{2}(0.99\sqrt{2n\beta})^n \mathrm{vol}_{n-1} S^{n-1}$

Lemma 2.6. *Let H be a closed halfspace which is disjoint from $C^* \cap F^{\alpha-\varepsilon_2}$. Then H contains at most*

$$(3n(\varepsilon_2+\varepsilon_3))^{\frac{n-1}{2}} \mathrm{vol}_{n-1} S^{n-1}$$

of the surface area of $C^* \cap \partial F^{\alpha+\varepsilon_3}$.

Using Lemmas 2.2, 2.4, 2.5, 2.6 we can now give the proof of Theorem 2.3.

Proof. As we have seen, K_N is weakly sandwiched between $F^{\alpha-\varepsilon_2}$ and $F^{\alpha+\varepsilon_3}$ with probability 0.98. Let K_N be such a ± 1 polytope. Each facet of K_N cuts off at most

$$(3n(\varepsilon_2+\varepsilon_3))^{\frac{n-1}{2}} \mathrm{vol}_{n-1} S^{n-1}$$

of the surface area of $C^* \cap \partial F^{\alpha+\varepsilon_3}$. In view of weak sandwiching, at least half of the surface area is cut off. Thus there are at least

$$\frac{0.5(0.99\sqrt{2n(\alpha+\varepsilon_3)})^n}{(3n(\varepsilon_2+\varepsilon_3))^{\frac{n-1}{2}}} \geq (c_7 \log N)^{\frac{n}{4}}$$

facets. □

Of course this proves Theorem 2.2 as well : The random ± 1 polytope K_N is weakly sandwiched with high probability so

$$E[f_{n-1}(K_N)] \geq 0.98(c_7 \log N)^{\frac{n}{4}}.$$

2.6 Auxiliary lemmas

We fix the one-to-one correspondence between $x \in (-1,1)$ and $t \in R$ via

$$t = f'(x) = h(x) = \frac{1}{2}\log\frac{1+x}{1-x} \text{ and } x = K'(t) = \tanh t$$

throughout the paper. This induces a one-to-one correspondence between $\underline{x} \in \mathrm{int}C$ and $\underline{t} \in R^n$ with

$$\underline{t} = n\,\mathrm{grad}F(\underline{x}).$$

Lemma 2.7. *The function*

$$g(t) = \frac{f(\tanh t)}{t^2} = -\frac{1}{t^2}\log\cosh t + \frac{\tanh t}{t}$$

is strictly decreasing on $[0,\infty)$. *Its limit at* $t = 0$ *is* $1/2$.

The value of $g(t)$ is 0.497... when $\tanh t = 0.1$ implying

$$\frac{1}{2.02}t^2 \leq f(\tanh t) \leq \frac{1}{2}t^2, \text{ for } t \in [-0.1, 0.1].$$

The last inequality shows that, when $\underline{x} = (x_1, \ldots, x_n) \in \mathrm{int}\ C^*$, and $\underline{t} = (t_1, \ldots, t_n)$ with $x_j = \tanh t_j$,

$$\frac{1}{2.02n}|\underline{t}|^2 \leq F(\underline{x}) \leq \frac{1}{2n}|\underline{t}|^2.$$

Let $\underline{\omega} \in S^{n-1}$ be a unit vector and define $\underline{x}(\underline{\omega}, \beta)$ as the unique point (if exists) on the boundary of F^β where

$$\underline{t}(\underline{\omega}, \beta) = n\,\mathrm{grad}F(\underline{x}(\underline{\omega}, \beta)) = \vartheta(\underline{\omega}, \beta)\underline{\omega}$$

for some positive $\vartheta(\underline{\omega}, \beta)$. Define $\mathrm{supp}\,\underline{\omega} = \{i \in \{1, \ldots, n\} : \omega_i \neq 0\}$. We have

Lemma 2.8. *$\underline{x}(\underline{\omega}, \beta)$ is welldefined when*

$$0 < \beta < \frac{|\operatorname{supp} \underline{\omega}|}{n} \log 2$$

and $\vartheta(\underline{\omega}, \beta)$ is strictly increasing in β.

Define

$$\Omega = \{\underline{\omega} \in S^{n-1} : \sqrt{\frac{n}{3 \log n}}\, \underline{\omega} \in C\}.$$

It is simple consequence of Dvoretzky's theorem [D] that for large enough n,

$$\operatorname{Prob}[\underline{\omega} \in \Omega | \underline{\omega} \in S^{n-1}] > 0.99.$$

We will use this in the proof of

Lemma 2.9. *Assume $\underline{\omega} \in \Omega$, and $\beta < \frac{1}{606 \log n}$. Then $\underline{x}(\underline{\omega}, \beta) \in C^*$.*

2.7 Proof of the lemmas

Proof of lemmas 2.2 and 2.4

Proof. (Lemma 2.2) This is a copy of the proof of Lemma 2.1(b) from [DFM] with the parameters adjusted properly. Suppose K_N is full-dimensional and there exists a point $\underline{x} \in Q^\beta \backslash K_N$ (where $\beta = \alpha - \varepsilon_1$). Then there is a facet of K_N, spanned by $\underline{Z}_{i_1}, \dots, \underline{Z}_{i_n}$, such that the corresponding halfspace contains K_N but excludes $\underline{x}$. Let $J = \{j_1, \dots, j_n\}$ and define the event E_J:

The points $\underline{Z}_{j_1}, \dots, \underline{Z}_{j_n}$ span a hyperplane and for one of the two corresponding halfspaces H both $\operatorname{Prob}[\underline{Z} \notin H] \geq e^{-\beta n}$ and the event $\{\underline{Z}_j : j \notin J\} \subset H$ occurs.

It is clear that in our case the event E_I with $I = \{i_1, \dots, i_n\}$ occurs. Let E denote the event that K_N is not full dimensional. Then

$$\{Q^\beta \not\subseteq K_N\} \subset E \cup \bigcup_{\text{all } J} E_J.$$

Thus, with notation $D = \{1, \dots, n\}$,

$$\begin{aligned} \operatorname{Prob}[Q^\beta \not\subseteq K_N] &\leq \operatorname{Prob}[E] + \sum_{\text{all } J} \operatorname{Prob}[E_J] \\ &= \operatorname{Prob}[E] + \binom{N}{n} \operatorname{Prob}[E_D]. \end{aligned}$$

For any fixed set S of dimension less than n, $\text{Prob}[\underline{Z} \in S] \leq \frac{1}{2}$, so $\text{Prob}[E] \leq \binom{N}{n} 2^{-(N-n)} < 0.001$ if n is large enough.

To bound $\text{Prob}[E_D]$ suppose $\underline{Z}_1, \dots, \underline{Z}_n$ are affinely independent. Let H_1 and H_2 be the two halfspaces they determine. If $\text{Prob}[\underline{Z} \notin H_1] \geq e^{-\beta n}$, then

$$\text{Prob}[\underline{Z}_j \in H_1 : j = n+1, \dots, N] \leq (1 - e^{-\beta n})^{N-n}$$

and similarly for H_2. Hence

$$\text{Prob}[E_D | \underline{Z}_1, \dots, \underline{Z}_n \text{ aff. indep.}] \leq 2(1 - e^{-\beta n})^{N-n} < 2\exp\{-(N-n)e^{-\beta n}\}.$$

By removing the conditioning we get the same bound on $\text{Prob}(E_D)$. Hence

$$\begin{aligned}
\text{Prob}[Q^\beta \not\subseteq K_N] &\leq \text{Prob}(E) + \binom{N}{n} \text{Prob}[E_D] \\
&< 0.001 + 2\exp\{n \log N - (N-n)e^{-\beta n}\} \\
&< 0.001 + 2\exp\left\{n(e^{-\beta n} + \log N) - N e^{-\beta n}\right\}.
\end{aligned}$$

Here $\beta = \alpha - \varepsilon_1$, $N = e^{\alpha n}$, $\varepsilon_1 = b_1 \frac{\sqrt{\log N}}{n}$, and consequently $N e^{-\beta n} = e^{\varepsilon_1 n} = \exp\{b_1 \sqrt{\log N}\}$. By condition $(*)$ this is much larger than the other term $n(e^{-\beta n} + \log N)$ in the exponent if b_1 is chosen large enough. For instance, with $b_1 \geq 3/\sqrt{c_4}$ and large enough n

$$\text{Prob}[Q^\beta \not\subseteq K_N] < 0.001 + 2\exp\{-n^2\} < 0.01.$$

□

Proof. (Lemma 2.4) Let $\underline{x}$ be any point of the boundary of F^β (where $\beta = \alpha + \varepsilon_3$). Then, using 2.1

$$\begin{aligned}
\text{Prob}[\underline{x} \in K_N] &\leq N q(\underline{x}) \leq N \exp\{-nF(\underline{x})\} \leq \\
&\leq \exp\{\alpha n - \beta n\} = \exp\{-\varepsilon_3 n\} < 0.001
\end{aligned}$$

if n is large enough. Then the expectation of the surface area of $C^* \cap \partial F^\beta$ contained in K_N is at most

$$\int_{C^* \cap \partial F^\beta} \text{Prob}[\underline{x} \in K_N] d\underline{x} \leq 0.001 \text{vol}_{n-1}(C^* \cap \partial F^\beta).$$

So the probability that half of $C^* \cap \partial F^\beta$ is missed by K_N is at least 0.998. □

The proof of lemma 2.3

The target is to show that the inequality $q(\underline{x}) \leq \exp\{-nF(\underline{x})\}$ in Lemma 2.1 is rather sharp. First we need a quantitative version of Lemma 4.4 of [DFM]. We assume $\beta = \alpha \pm \varepsilon$.

Lemma 2.10. *For every positive integer n the following holds. If $0 \leq x_i \leq 0.1$, $t_i = h(x_i)$ $(i = 1, \dots, n)$ and $nF(\underline{x}) \geq 10$, then*

$$\text{Prob}\left[\sum_{i=1}^{n} t_i(Z_i - x_i) \geq 0\right] \geq \exp\left\{-nF(\underline{x}) - 3\sqrt{nF(\underline{x})}\right\}.$$

Proof. (It goes via exponential centering and a Berry-Esséen type theorem, just like in [DFM].) Let $X_1, \dots, X_n$ be independent discrete random variables and set $X = \sum X_i$. Define new random variables W_i with distribution

$$\text{Prob}[W_i = y] = e^y \frac{\text{Prob}[X_i = y]}{E[e^{X_i}]}.$$

Set $W = \sum W_i$ and observe

$$\begin{aligned}\text{Prob}[W = y] &= \sum_{y_i : \sum y_i = y} \prod_{i=1}^{n} e^{y_i} \frac{\text{Prob}[X_i = y_i]}{E[e^{X_i}]} \\ &= \left(\prod_{1}^{n} E[e^{X_i}]\right)^{-1} e^y \,\text{Prob}[X = y].\end{aligned}$$

Apply this with $X_i = t_i Z_i$ where, as usual, Z_i is uniform over $\{-1, 1\}$

$$\text{Prob}\left[\sum_{1}^{n} t_i Z_i = w\right] = \exp\left\{\sum_{1}^{n} K(t_i)\right\} e^{-w}\,\text{Prob}[W = w].$$

It is easy to check that $E[W_i] = t_i \tanh t_i = t_i x_i$. Let $Y = W - E[W] = W - \sum t_i x_i$. With this notation

$$\begin{aligned}\text{Prob}\left[\sum_{1}^{n} t_i(Z_i - x_i) \geq 0\right] &= \exp\left\{\sum_{1}^{n} K(t_i)\right\} \sum_{w \geq \sum t_j x_j} e^{-w}\,\text{Prob}[W = w] \\ &= \exp\left\{\sum_{1}^{n}(K(t_i) - t_i x_i)\right\} \sum_{y \geq 0} e^{-y}\,\text{Prob}[Y = y].\end{aligned}$$

Here $\sum(K(t_i) - t_i x_i) = -\sum f(x_i)$ so we have to show that $\sum_{y \geq 0} e^{-y}\text{Prob}[Y = y]$ is not too small. Setting $Y_j = W_j - E(W_j)$ we have $Y = \sum Y_j$ and $E(Y_j) = 0$. Easy calculations give

$$\sigma_j^2 = E[Y_j^2] = \frac{t_j^2}{\cosh^2 t_j} \text{ and } E[|Y_j^3|] = \left(2\cosh t_j - \frac{1}{\cosh t_j}\right)\sigma_j^3.$$

We need a few simple estimates: when $0 \leq x_i \leq 0.1$,

$$0 \leq t_i \leq h(0.1) = 0.1003353\ldots \text{ and } 1 \leq \cosh t_i \leq 1.00503\ldots.$$

It is easy to check that

$$\frac{E[|Y_j|^3]}{E[Y_j^2]} = t_j(1 + \tanh^2 t_j)$$

is an increasing function in t_j. Thus, in the given range,

$$M = \max_j \frac{E[|Y_j|^3]}{E[Y_j^2]} \leq 0.102\ldots.$$

Define $\sigma = \sqrt{\sum \sigma_j^2}$. Now Berry's theorem (see [Fe]) says that, under the present conditions, for all n, the distribution of $\frac{1}{\sqrt{\sigma}} \sum_1^n Y_j$ differs from that of the standard normal by at most

$$\frac{33}{4}\frac{M}{\sigma}.$$

Now

$$\sigma = \sqrt{\sum_1^n \sigma_j^2} = \sqrt{\sum_1^n \frac{t_j^2}{\cosh^2 t_j}} \geq \frac{1}{1.00503\ldots}\sqrt{\sum_1^n t_j^2} \geq$$
$$\geq 0.99\sqrt{2nF(\underline{x})} \geq 0.99\sqrt{20} > 4.427.$$

Since the standard normal between 0 and $\sqrt{\sigma} > \sqrt{4.427} > 2.1$ is larger than 0.49, Berry's theorem implies that,

$$\text{Prob}\left[0 \leq \frac{\sum_j^n Y_j}{\sqrt{\sigma}} \leq \sqrt{\sigma}\right] > 0.49 - 2\frac{33}{4}\frac{M}{\sigma} > \frac{1}{10}.$$

With this

$$\sum_{y \geq 0} e^{-y}\,\text{Prob}[Y = y] \geq \sum_{0 \leq y \leq \sigma} e^{-y}\,\text{Prob}[Y = y]$$
$$\geq e^{-\sigma}\,\text{Prob}[0 \leq \sum_1^n Y_j \leq \sigma] \geq \frac{1}{10}e^{-\sigma} \geq e^{-3\sqrt{nF(\underline{x})}},$$

since $\sigma = \sqrt{\sum \sigma_j^2} \leq \sqrt{\sum t_j^2} \leq \sqrt{2.02nF(\underline{x})} < 3\sqrt{nF(\underline{x})} - \log 10$, because $nF(\underline{x}) \geq 10$. □

Lemma 2.11. *Assume $\alpha_i > 0$ for $i = 1, \dots, m$. Then*

$$\text{Prob}\left[\sum_1^m \alpha_i(Z_i - 0.1) \geq 0\right] \geq \frac{1}{m^2}\exp\{-mf(0.1)\}.$$

Proof. Let H^*, H respectively be the halfspaces

$$H^* = \left\{\underline{x} \in R^m : \sum_1^m \alpha_i(x_i - 0.1) \geq 0\right\},$$
$$H = \left\{\underline{x} \in R^m : \sum_1^m (x_i - 0.1) \geq 0\right\}.$$

Define $\sigma : R^m \to R^m$ to be the cyclic shift of the components of x, that is, $\sigma(x_1, \dots, x_m) = (x_m, x_1, \dots, x_{m-1})$. The orbit of $\underline{x}$ under σ is, by definition, $\{\underline{x}, \sigma(\underline{x}), \sigma^2(\underline{x}), \dots\}$. As $\sigma^m(\underline{x}) = \underline{x}$, any orbit has at most m elements. If $\underline{x} \in H$ then so is $\sigma(\underline{x})$. At least one element of each orbit with $\underline{x} \in H$ is in H^* as otherwise

$$\sum_{j=1}^m \alpha_j(\sigma^k(\underline{x})_j - 0.1) < 0 \quad \text{for all } k = 0, 1, \dots, m-1.$$

Summing these inequalities for all k we get

$$\sum_{j=1}^m (\alpha_1 + \ldots + \alpha_m)(x_j - 0.1) < 0,$$

a contradiction. Now we see that

$$\begin{aligned}\text{Prob}\left[\sum_1^m \alpha_i(Z_i - 0.1) \geq 0\right] &= \text{Prob}[\underline{Z} \in H^*] \\ &\geq \frac{1}{m}\text{Prob}\left[|\{i : Z_i = 1\}| \geq 0.55m\right] \\ &= \frac{1}{m}\frac{1}{2^m}\sum_{k=0.55m}^m \binom{m}{k} \geq \frac{1}{m^2}\exp\{-mf(0.1)\}\end{aligned}$$

as a simple calculation using Stirling formula reveals. □

Proof. (Lemma 2.3) We have to show that $q(\underline{x}) \geq \exp\{-(\alpha - \varepsilon_1)n\}$ for each $\underline{x} \in F^{\alpha-\varepsilon_2} \cap C^*$, or, in other words, every halfspace H intersecting $F^{\alpha-\varepsilon_2} \cap C^*$ contains at least $2^n \exp\{-(\alpha - \varepsilon_1)n\}$ vertices of C. It suffices to show this for halfspaces H whose bounding hyperplane H^o is tangent to $F^{\alpha-\varepsilon_2} \cap C^*$.

We show first, that H contains a point $\underline{x}$ with $F(\underline{x}) = \alpha - \varepsilon_2$ on its boundary. If H^o touches $F^{\alpha-\varepsilon_2}$ then the point of tangency satisfies this condition. If not, then H contains a point $\underline{y}$ with $F(\underline{y}) < \alpha - \varepsilon_2$ and there is a smallest face of C^* containing $\underline{y}$. Since the vertices of C^* are not contained in $F^{\alpha-\varepsilon_2}$ there is a point $\underline{x}$ on this face with $F(\underline{x}) = \alpha - \varepsilon_2$.

By symmetry we can suppose that all components of $\underline{x}$ are nonnegative and in increasing order. Let $n_1 \in \{1, \dots, n\}$ be such that $x_{n_1} < 0.1$ and $x_{n_1+1} = 0.1$. Set $\underline{t} = n\,\mathrm{grad}F(\underline{x})$ and let $\underline{t}^*$ be the normal to H^o. We will prove the lemma assuming that $\underline{t}^*$ is in the relative interior of the normal cone to $F^{\alpha-\varepsilon_2} \cap C^*$ at the point $\underline{x}$; this assumption means that

$$\begin{aligned} t_i^* &= t_i \qquad && \text{for } i = 1, \dots, n_1 \\ t_i^* &> t_i \qquad && \text{for } i = n_1 + 1, \dots, n. \end{aligned}$$

The statement of the lemma for general $\underline{t}^*$ follows from this easily.

Next we have to consider cases according to where the terms of the sum $\sum_1^n f(x_i) = n(\alpha - \varepsilon_2)$ are concentrated. If $n_1 \geq n - 2000$, then Lemma 2.10 applies: check that $\sum_1^{n_1} f(x_i) \geq n(\alpha - \varepsilon_2) - 2000 f(0.1) > 10$ if n is large. Choose the last, at most 2000, Z_i to be 1 ($i = n_1 + 1, \dots, n$). We get

$$\begin{aligned} \mathrm{Prob}\left[\sum_{i=1}^{n} t_i^*(Z_i - x_i) \geq 0\right] &\geq 2^{-2000}\mathrm{Prob}\left[\sum_{i=1}^{n_1} t_i(Z_i - x_i) \geq 0\right] \\ &\geq 2^{-2000}\exp\left\{-\sum_{i=1}^{n_1} f(x_i) - 3\sqrt{\sum_{i=1}^{n_1} f(x_i)}\right\} \\ &\geq \exp\left\{-n(\alpha - \varepsilon_2) - 3\sqrt{n(\alpha - \varepsilon_2)} - 2000\log 2\right\} \\ &\geq \exp\left\{-n(\alpha - \varepsilon_1)\right\}. \end{aligned}$$

The last inequality follows when n and thus N is large enough if one chooses here, with $\varepsilon_i = b_i \frac{\sqrt{\log N}}{n}$, $i = 1, 2$,

$$b_2 \geq 2b_1 + 3.$$

If $n_1 < n - 2000$, then set $n_2 = n_1 + 2000$ and write

$$\begin{aligned} \mathrm{Prob}\left[\sum_{i=1}^{n} t_i^*(Z_i - x_i) \geq 0\right] &\geq \mathrm{Prob}\left[\sum_{i=1}^{n_2} t_i(Z_i - x_i) \geq 0\right] \times \\ \times \mathrm{Prob}&\left[\sum_{i=n_1+1}^{n_2} (t_i^* - t_i)(Z_i - x_i) + \sum_{i=n_2+1}^{n} t_i^*(Z_i - x_i) \geq 0\right]. \end{aligned}$$

Lemma 2.10 applies to the first probability since $\sum_{i=1}^{n_2} f(x_i) \geq 2000 f(0.1) > 10$. It gives

$$\text{Prob}\left[\sum_{i=1}^{n_2} t_i(Z_i - x_i) \geq 0\right] \geq \exp\left\{-\sum_{i=1}^{n_2} f(x_i) - 3\sqrt{\sum_{i=1}^{n_2} f(x_i)}\right\}.$$

Lemma 2.11 works for the second factor and shows that it is at least

$$\frac{1}{(n-n_1)^2}\exp\{-(n-n_1)f(0.1)\}.$$

The last two inequalities combine to

$$\text{Prob}\left[\sum_{i=1}^{n} t_i^*(Z_i - x_i) \geq 0\right]$$
$$\geq \exp\left\{-nF(\underline{x}) - 3\sqrt{nF(\underline{x})} - 2000f(0.1) - 2\log(n-n_1)\right\}.$$

The exponent here is $-n(\alpha-\varepsilon_2) - 3\sqrt{n(\alpha-\varepsilon_2)} - 2000f(0.1) - 2\log(n-n_1)$ which is larger than $-n(\alpha-\varepsilon_1)$, if, in the definition of ε_2, the constant b_2 is chosen large enough. □

Proof of the geometric lemmas

Proof. (Lemma 2.5) A routine argument shows how to compute the product curvature $\kappa(\underline{x})$ of the surface given implicitly by $F(\underline{x}) = \beta$: it gives, at the point $\underline{x}$

$$\frac{1}{\kappa(\underline{x})} = \frac{|\text{grad}F(\underline{x})|^n}{\det F''} = \frac{|\underline{t}|^n}{\prod_{i=1}^n \frac{1}{1-x_i^2}} \geq \frac{(2n\beta)^{\frac{n}{2}}}{(0.99)^{-n}} \geq (0.99\sqrt{2n\beta})^n$$

since $\underline{x} \in C^*$ implies $x_i^2 \leq 0.01$. We use this in the well-known formula [BF] giving the surface area as the integral of $\frac{1}{\kappa(\underline{x})}$ on S^{n-1}. Now with $\beta = \alpha + \varepsilon_3$

$$\text{vol}_{n-1}(\partial F^\beta \cap C^*) = \int_{\underline{\omega}\in S^{n-1}} \frac{1}{\kappa(\underline{x})} d\underline{\omega}$$

$$\geq \int_{\underline{\omega}\in\Omega} (0.99\sqrt{2n\beta})^n d\underline{\omega} \geq \frac{1}{2}(0.99\sqrt{2n\beta})^n \text{vol}_{n-1} S^{n-1}.$$

□

Proof. (Lemma 2.6) We can assume that the touching hyperplane H^o of the halfspace H is tangent to $F^{\alpha-\varepsilon_2} \cap C^*$ at the point $\underline{x}$ with $F(\underline{x}) = \alpha - \varepsilon_2$.

We assume, by symmetry, that all $x_i \geq 0$. If $\underline{x}$ is in $\operatorname{int} C^*$ then H is welldefined with normal $\underline{t} = n \operatorname{grad} F(\underline{x})$.

If $\underline{x}$ is not in $\operatorname{int} C^*$ then we can assume (as in the proof of Lemma 2.3) that the outer normal $\underline{t}^*$ of H is in the relative interior of the normal cone to $C^* \cap F^{\alpha-\varepsilon_2}$ at $\underline{x}$. Then $\underline{t}^*$ can be chosen so that

$$\begin{aligned} t_i^* = t_i = f'(x_i) &\qquad \text{for } \ 0 \leq x_i \leq 0.1, \text{and} \\ t_i^* \geq t_i = f'(0.1) &\qquad \text{for } \ x_i = 0.1 \end{aligned}$$

Assume $\underline{y} \in H \cap C^*$. Then, as $y_i - x_i \leq 0$ if $x_i = 0.1$,

$$\sum_{i=1}^{n} t_i(y_i - x_i) \geq \sum_{i=1}^{n} t_i^*(y_i - x_i) \geq 0,$$

showing that $H \cap C^* \subset \{\underline{z} : \underline{t}(\underline{z} - \underline{x}) \geq 0\}$. So we may assume that the normal vector of H is just $\underline{t} = n \operatorname{grad} F(\underline{x})$.

Now let $\underline{w} \in H \cap F^{\alpha+\varepsilon_3} \cap C^*$. Set $\underline{u} = \underline{w} - \underline{x}$ which is clearly in $2C^*$. Then with suitable $\zeta_i \in [0, x_i]$ we have

$$\begin{aligned} F(\underline{w}) &= F(\underline{x}) + (\underline{w} - \underline{x}) \operatorname{grad} F(\underline{x}) + \frac{1}{2}(\underline{w} - \underline{x})^{\mathrm{T}} F''(\underline{x})(\underline{w} - \underline{x}) \\ &+ \frac{1}{6} \sum_{i=1}^{n} \frac{f'''(\zeta_i)}{n}(w_i - x_i)^3 \\ &\geq \alpha - \varepsilon_2 + \frac{1}{2n} \sum_{i=1}^{n} \frac{1}{1 - x_i^2} u_i^2 + \frac{1}{6n} \sum_{i=1}^{n} \frac{2\zeta_i}{(1 - \zeta_i^2)^2} u_i^3 \\ &\geq \alpha - \varepsilon_2 + \frac{1}{2n} \sum_{i=1}^{n} u_i^2 \left(\frac{1}{1 - x_i^2} - \frac{2x_i}{3(1 - x_i^2)^2} u_i \right) \geq \alpha - \varepsilon_2 + \frac{|\underline{u}|^2}{3n}. \end{aligned}$$

On the other hand $\underline{w} \in F^{\alpha+\varepsilon_3}$, so

$$\alpha + \varepsilon_3 \geq F(\underline{w}) \geq \alpha - \varepsilon_2 + \frac{|\underline{u}|^2}{3n}$$

implying

$$|\underline{u}| \leq \sqrt{3n(\varepsilon_2 + \varepsilon_3)}.$$

This shows that the cut-off from $\partial F^{\alpha+\varepsilon_3}$ by the halfspace H is contained in a ball of radius $\sqrt{3n(\varepsilon_2 + \varepsilon_3)}$ so its surface area is at most

$$(3n(\varepsilon_2 + \varepsilon_3))^{\frac{n-1}{2}} \operatorname{vol}_{n-1} S^{n-1}.$$

$\square$

Proof of the auxiliary results

Proof. (Lemma 2.7) It is elementary to see that $\lim_{t\to 0} g(t) = 1/2$. We have to show that, for all $t \in (0,\infty)$, $g'(t) \le 0$, or, what is the same, $t^3 g'(t) \le 0$. Direct computation gives

$$h(t) = t^3 g'(t) = 2\log\cosh t - 2t\tanh t + \frac{t^2}{\cosh^2 t}.$$

As $\lim_{t\to 0} h(t) = 0$, it is enough to see that $h'(t)$ is nonpositive:

$$h'(t) = -\frac{2t^2 \sinh t}{\cosh^3 t} \le 0.$$

□

Proof. (Lemma 2.8) Fix $\underline{\omega}$. Let $\vartheta \in (0,\infty)$ and define

$$x_i = \tanh \vartheta\omega_i \quad i = 1,\dots,n.$$

This gives a point $\underline{x} \in C$ with $n\mathrm{grad}F(\underline{x}) = \vartheta\underline{\omega}$. For fixed $\underline{\omega} \in S^{n-1}$, the mapping $\vartheta \longrightarrow F(\underline{x}) = \frac{1}{n}\sum_{i=1}^n f(\tanh\vartheta\omega_i)$ is strictly increasing and continuous, it is 0 at $\vartheta = 0$ and its limit at $\vartheta \to \infty$ is $\frac{1}{n}|\mathrm{supp}\,\underline{\omega}|\log 2$. This proves the first part of the statement. The second part follows from the monotonicity of $\vartheta \longrightarrow \frac{1}{n}\sum_{i=1}^n f(\tanh\vartheta\omega_i)$. □

Proof. (Lemma 2.9) Define $\lambda = \sqrt{\frac{n}{3\log n}}$. As $\underline{\omega} \in \Omega$, $\lambda\underline{\omega} \in C$ and $\frac{\lambda}{10}\underline{\omega} \in C^*$, so there is a point $\underline{x} \in C$ with $n\,\mathrm{grad}F(\underline{x}) = \frac{\lambda}{10}\underline{\omega}$. We may assume, by symmetry, that all $\omega_i \ge 0$. Of course $x_i = \tanh\frac{\lambda}{10}\omega_i$. At this point

$$F(\underline{x}) \ge \frac{1}{2.02n}|\frac{\lambda}{10}\underline{\omega}|^2 > \frac{1}{606\log n}.$$

Monotonicity implies then, that

$$\begin{aligned}0 \le x(\underline{\omega},\beta)_i = \tanh\vartheta(\underline{\omega},\beta)\omega_i <\\ < \tanh\frac{\lambda}{10}\omega_i \le \tanh 0.1 < 0.1\end{aligned}$$

for all $i = 1,\dots,n$. □

3 The Erdős and Szekeres theorem

3.1 Introduction

A set of points in the plane is said to be *in general position* if it contains no three points on a line. A finite set of points in the plane is *in convex position*, if it is the vertex set of a convex polygon. In this part of the work we investigate generalizations of the following famous result of Erdős and Szekeres [ES35]:

Theorem 3.1 (Erdős and Szekeres). *For any $k \geq 3$ there is a least integer $f(k)$ such that any set of at least $f(k)$ points in general position in the plane contains k points in convex position.*

For a triple (a, b, c) of points in general position in the plane, we say that its *orientation* is "+", if the clockwise order of the vertices of the triangle abc is a, b, c. Otherwise we say that the *orientation* of (a, b, c) is "−". We say that two t-tuples of points in general position $(p_1, \ldots, p_t)$ and $(p'_1, \ldots, p'_t)$ have the same *order type*, if the orientations of the two triples (p_i, p_j, p_l) and (p'_i, p'_j, p'_l) are equal for each choice of distinct indices i, j, l. A finite planar point set X is called a *k-clustering*, if it is a disjoint union of k sets $X_1, \ldots, X_k$ of equal sizes such that all k-tuples $(x_1, \ldots, x_k)$, $x_1 \in X_1, \ldots, x_k \in X_k$, have the same order type. If, moreover, $x_1, x_2, \ldots, x_k$ are in convex position, then X is called a *convex k-clustering* (see Fig.1).

Bárány and Valtr [BV98] proved the following generalization of the Erdős–Szekeres theorem:

Theorem 3.2 (positive fraction Erdős–Szekeres theorem). *For any $k \geq 3$ there is an $\varepsilon_k > 0$ such that if X is a finite set of points in general position in the plane with $|X| \geq f(k)$ then it contains a convex k-clustering of size at least $\varepsilon_k|X|$. ($f(k)$ is the function from the Erdős–Szekeres theorem.)*

For $k = 4$ this was proved earlier by Nielsen [N95]. Solymosi [SO88] proved in his master thesis a closely related result, that among n points in the plane there is a sequence of $c_r n$ elements such that any r consecutive are in convex position.

Repeated applications of the positive fraction Erdős–Szekeres theorem show that any set of n points in general position in the plane can be partitioned into at most $c_k \log n$ convex k-clusterings and a remaining set of size at most c'_k. Gil Kalai asked if $c_k \log n$ can be replaced by c_k, which would give a strengthening of the positive fraction Erdős–Szekeres theorem. It is easy to see for $k = 3$. P. Valtr and me proved it for arbitrary k:

Theorem 3.3 (partitioned Erdős–Szekeres theorem). *For every $k \geq 3$ there are two constants $c = c(k)$, $c' = c'(k)$ such that the following holds. If X is a finite set of points in general position in the plane then it has a subset X' of size at most c' such that $X \setminus X'$ can be partitioned into at most c convex k-clusterings.*

The proof gives Theorem 3.3 with constants $c(k) = k^{O(k^2)}$ and $c'(k) = f(k) - 1$, where $f(k) < 4^k$ is the number from the Erdős–Szekeres theorem. If $|X|$ is sufficiently large, then $c'(k)$ can be lowered to $rr(k) - 1$, where $rr(k)$ is the so-called Ramsey-remainder for convex sets (see Section 3.3 for details). Both bounds on $c'(k)$ are optimal, while the best lower bound on $c(k)$ is exponential in k.

If the special case when $k = 4$ we can give an upper bound:

Theorem 3.4. $c(4) \leq 26$

In the case of $k = 4$ Gyula Károlyi investigated a closely related problem: when can a set of $4n$ points be partitioned into n convex (vertex-disjoint but possibly intersecting) quadrilaterals [KA02].

The proof of Theorem 3.3 relies on the positive fraction Erdős–Szekeres theorem (Theorem 3.2). Pach and Solymosi [PS98] proved that Theorem 3.2 holds with $\varepsilon_k = \alpha^{-k^2}$ for some fixed $\alpha > 1$. We give the following better bound:

Theorem 3.5 (Exponential bound). *Theorem 3.2 holds with $\varepsilon_k = k \cdot 2^{-32k}$.*

Further we show that there is a $\beta > 1$ such that Theorem 3.2 does not hold with the constants $\varepsilon_k = \beta^{-k}$.

The proof of Theorem 3.4 is based on another partitional statement, a generalization of the same type lemma. This beautiful result of Bárány and Valtr holds in any dimension. Given sets $Y_1, \dots, Y_m$ a collection $y_1, \dots, y_m$ where $y_i \in Y_i$ is a transversal of the sets $Y_1, \dots, Y_m$.

Same Type Lemma [BV98] *For any natural numbers $m, d \geq 1$ there exists a constant $c(m, d)$ such that if $X_1, \dots, X_m \subset \mathbb{R}^d$ are finite sets and $\bigcup_i X_i$ is in general position, there are subsets $Y_i \subset X_i$, $|Y_i| \geq c(m, d)|X_i|$ such that every transversal of the sets $Y_1, \dots, Y_m$ has the same geometric (order) type.*

Pach and Solymosi [PS98] generalized the same type lemma for disjoint convex sets.

Each $d+1$ points in general position may have two different orientations, "+" and "-". Here the geometric type means the collection of orientations of the $\binom{m}{d+1}$ $d+1$-tuples, i.e. two point sets of size m have the same geometric type , under a given order of the points, if the $d+1$ points taken from each set from the same $d+1$ positions have the same orientation. We will call sets like $Y_1, \dots, Y_m$ to be separated in "algebraic" sense. We will define two other types of separation: "geometric" and "combinatorial".

The sets $X_1, \dots, X_m$ are separated in "geometric" sense, if each hyperplane H intersects at most d sets from $\mathrm{conv}X_1, \dots, \mathrm{conv}X_m$.

The number $t_{m,d}$ is the number of bipartitions of m points in $\mathbb{R}^d$ in general position by hyperplanes. The sets $X_1, \dots, X_m$ are separated in "combinatorial" sense, if the number of bipartitions of the sets by hyperplanes which do not meet any of the sets $\mathrm{conv}X_i$ is equal to $t_{m,d}$.

We will prove in section 6 that all these definitions are equivalent and therefore call the sets $X_1, \dots, X_m$ with either property separated.

We need some further definitions to formulate a partitioned version of the same type lemma.

We call M a matrix-partition (or μ-partition, for short) with N columns on the sets $X_1, \dots, X_m \subset \mathbb{R}^d$ if

$$
\begin{array}{ll}
(i) & M = (M_{ij})_{m\times N}, M_{ij} \subset X_i \\
(ii) & M_{ij_1} \cap M_{ij_2} = \emptyset, \quad \text{for all } \ i, j_1, j_2 (j_1 \neq j_2) \\
(iii) & \bigcup_j M_{ij} = X_i \quad \text{for all } \ i
\end{array}
$$

Combining this with separability, we call the μ-partition M a *separat ed* μ-partition if the sets in each column are separated. We call the μ-partition *balanced* if the sets in the same column have the same size. Clearly, in this case $|X_1| = \ldots = |X_m|$. We can look at a culomn of a balanced separated μ-partition as an m-cluster.

The "partitioned" same type lemma:

Theorem 3.6. *For all natural numbers m, d there exists a least integer $n_{m,d}$ such that if finite sets $X_1, \dots, X_m \subset \mathbb{R}^d$ have the same size and $\bigcup_i X_i$ is in general position, then there exists a balanced, separated μ-partition with at most $n_{m,d}$ columns on these sets.*

The proof of Theorem 3.6 goes by a double induction, based on two lemmas and gives an exponential bound in m for $n_{m,1}$. Better bounds in lower dimensions would yield better bounds in higher dimensions too.

There are asymptotically tight bounds for $n_{m,1}$, namely

Theorem 3.7.

$$\left\lfloor \frac{m-1}{2} \right\rfloor \cdot \left\lfloor \frac{m+1}{2} \right\rfloor + 1 \leq n_{m,1} \leq (m-1)^2 + 1$$

3.2 The exponential bound, Theorem 3.5

Let X be a set of n points in general position in the plane. We may suppose that $n > 2^{32k}$. It follows from the best known upper bounds [ES35, TV98] on the function $f(k)$ from the Erdős-Szekeres theorem (Theorem 3.1) that among any 4^{4k} points in general position it is always possible to find $4k$ points in convex position. Therefore, if we choose randomly and uniformly a 4^{4k}-point subset X_0 of X and consequently randomly and uniformly a $4k$-point subset X_1 of X_0, then X_1 is in convex position with probability at least $1/\binom{4^{4k}}{4k}$. Clearly, every $4k$-point subset of X is chosen for X_1 with the same probability. It follows that X contains at least $\frac{\binom{n}{4k}}{\binom{4^{4k}}{4k}}$ $4k$-point subsets in convex position.

If X_1 is a $4k$-point subset of X in convex position, then we say that its subset Y of size $2k$ *supports* X_1 if the points of X_1 sorted in the clockwise order alternately belong to Y and to $X_1 \setminus Y$. Clearly, X_1 is supported by two subsets.

Since X has $\binom{n}{2k}$ $2k$-point subsets, there is a $2k$-point subset Y of X which supports at least $2 \cdot \frac{\binom{n}{4k}}{\binom{4^{4k}}{4k}} / \binom{n}{2k} > \frac{(n-4k)^{2k} \cdot \frac{(2k)!}{(4k)!}}{\binom{4^{4k}}{4k}} > \frac{(n-4k)^{2k} \cdot (2k)!}{\left(4^{4k}\right)^{4k}} > \frac{(n-4k)^{2k}}{2^{32k^2}}$ $4k$-point subsets of X in convex position. We fix such a set Y.

Let $y_1, \ldots, y_{2k}$ be the points of Y listed in the clockwise order. For each $i = 1, \ldots, 2k$, let T_i denote the region outside of conv$Y$ bounded by the segment $y_i y_{i+1}$ and by parts of the lines $y_{i-1}y_i, y_{i+1}y_{i+2}$ (indices are counted modulo $2k$) – see Fig.2

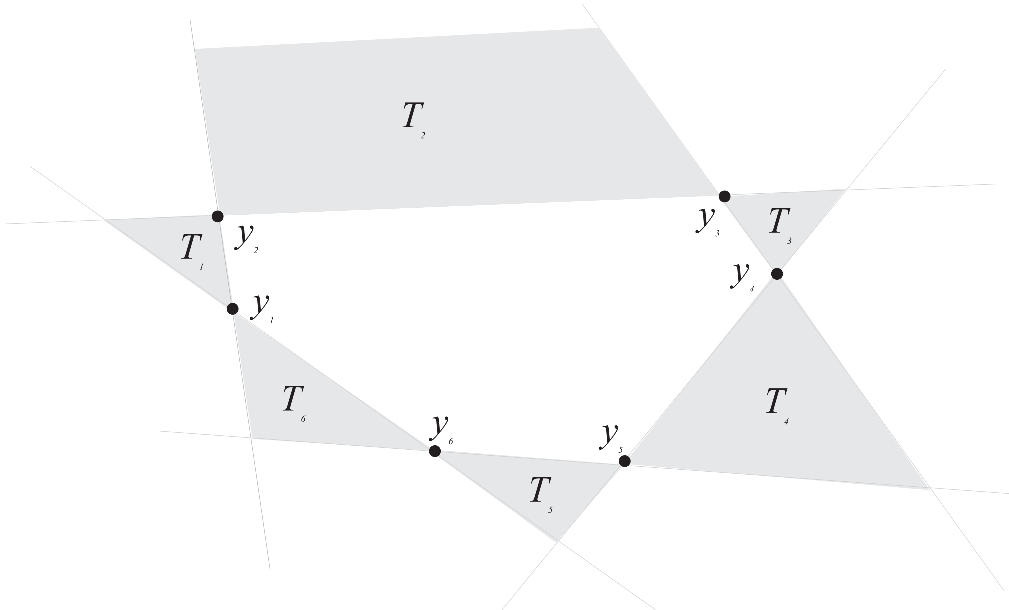


Fig.2

Observation 1. *If* $Y = \{y_1, \dots, y_{2k}\}$ *supports a* $4k$*-point subset* X_1 *of* X*, then* $X_1 = Y \cup \{x_1, \dots, x_{2k}\}$*, where* x_i *lies in the interior of* T_i *for each* $i = 1, \dots, 2k$*.* □

For $i = 1, \dots, 2k$, let t_i denote the number of points of X lying in the interior of T_i. It follows from Observation 1 that Y supports at most $\prod_{i=1}^{2k} t_i$ $4k$-point sets. Thus,

$$\frac{(n-4k)^{2k}}{2^{32k^2}} \leq \prod_{i=1}^{2k} t_i.$$

If α denotes the k-th smallest number t_i, then $\prod_{i=1}^{2k} t_i \leq \alpha^k n^k$. It follows that

$$\frac{(n-4k)^{2k}}{2^{32k^2}} \leq \alpha^k n^k$$

and thus

$$\alpha \geq \frac{(n-4k)^2}{2^{32k} n} > \frac{n-8k}{2^{32k}} > \frac{n}{2^{32k}} - 1.$$

Let $I \subset \{1, \dots, 2k\}$ be a set of k indices i with $t_i \geq \alpha > \frac{n}{2^{32k}} - 1$. For each $i \in I$, we fix a set C_i of $\lceil \frac{n}{2^{32k}} - 1 \rceil$ points of X in the interior of T_i. Then the sets $C_i \cup \{y_i\}, i \in I$, form a convex k-clustering of size $k \cdot \lceil \frac{n}{2^{32k}} \rceil$. □

In the proof of Theorem 3.5 we used that the best known bounds on $f(k)$ give $f(k) \leq 4^k$. If the conjecture of Erdős and Szekeres [ES61] that $f(k) = 2^{k-2} + 1$ is true then the above proof gives still a better constant $\varepsilon_k = k \cdot 2^{-16k}$.

3.3 The partitioned version of the Erdős-Szekeres theorem

In the whole proof, $k \geq 3$ is fixed and $K := 2^{32(k+1)+1}$.

For $t \geq 3$, we say that a family $\{X_1, X_2, \dots, X_t\}$ of t planar point sets has *the same-type property*, if all t-tuples $(x_1, \dots, x_t)$, $x_1 \in X_1, \dots, x_t \in X_t$, have the same order type (and we say that the pair $\{X_1, X_2\}$ of two planar point sets has *the same-type property*, if $\text{conv}X_1 \cap \text{conv}X_2 = \emptyset$). If, moreover, $x_1, x_2, \dots, x_t$ are in convex position, then we say that the sets $X_1, X_2, \dots, X_t$ are *in convex position.*

We say that a sequence $(X_1, X_2, \dots, X_t)$ of planar point sets has *the sequential same-type property*, if the family $\{X_1, X_2, \dots, X_i, X_{i+1} \cup X_{i+2} \cup \dots \cup X_t\}$ has the same-type property for each $i = 1, \dots, t-1$. A crucial notion in our proof is the notion of so-called s-configurations. An *s-configuration* is a finite set X of points in general position in the plane which is partitioned into $s+1$ sets $\mathcal{C}_1, \dots, \mathcal{C}_s, Y$ such that:

(i) $\mathcal{C}_i$ is a convex $(k+1)$-clustering for each $i = 1, \dots, s$,

(ii) for each $i = 1, \dots, s$, one of the $k+1$ clusters of $\mathcal{C}_i$ is denoted C_i so that the sequence $(C_1, \dots, C_s, Y)$ has the sequential same-type property, and

(iii) $|C_1| = |C_2| = \dots = |C_s| = \frac{|X|}{K} = \frac{|Y|}{K - s(k+1)}$, where $K = 2^{32(k+1)+1}$.

Sometimes we mean by an *s-configuration* not the set X but the corresponding $(2s+1)$-tuple $(\mathcal{C}_1, \dots, \mathcal{C}_s; C_1, \dots, C_s; Y)$ of sets. The definition of an s-configuration is somewhat technical. At this point we give two remarks to it:

Remark 1: The convex $(k+1)$-clustering $\mathcal{C}_i$ can be partitioned almost completely into few $(k+1)$ convex k-clusterings (Observation 2 in Section 3.3). Moreover, if we use the points of the cluster $C_i \subset \mathcal{C}_i$ elsewhere in the final partition, then the rest of $\mathcal{C}_i$ is a convex k-clustering. It is the reason why we require $\mathcal{C}_i$ to be a convex $(k+1)$-clustering.

Remark 2: s-configurations are defined so that the following two conditions hold. An s-configuration can be partitioned almost completely into a bounded number of $(s+1)$-configurations (Lemma 1 below), and if s is large enough ($s = 2k^2$ suffices) then an s-configuration can be partitioned almost completely into a bounded number of convex k-clusterings (Lemma 2 below).

Here are the two key lemmas:

Lemma 1. *For $s = 1, \dots, 2k^2$, every s-configuration is a disjoint union of at most $8s^2$ $(s+1)$-configurations and a remaining set of size at most $25s^2K$.*

Lemma 2. *Every $(2k^2)$-configuration is a disjoint union of at most $2K \log K$ convex k-clusterings and a remaining set of size at most 4^k.*

Lemma 1 will be obtained using the separation method of Pór [P02], Lemma 2 relies on Theorem 3.5 and on the Erdős-Szekeres theorem on monotone subsequences (Theorem 3.8 in Section 3.3). We now derive Theorem 3.3 from Lemmas 1 and 2.

Proof of Theorem 3.3. Let X be a finite set of points in general position in the plane. We may suppose that $|X|$ is divisible by $2K$ (this can be achieved by removing at most $2K-1$ points from X). By Theorem 3.5, X contains a convex $(k+1)$-clustering $\mathcal{C}_1$ with each cluster of size $\frac{|X|}{K}$. We denote one of the clusters by C_1.

Let l be a line ("ham-sandwich cut") cutting each of the sets $C_1, X \setminus \mathcal{C}_1$ into two equal parts. Let $C_1 = C_1^1 \cup C_1^2, X \setminus \mathcal{C}_1 = Y^1 \cup Y^2$ be the corresponding partitions such that $|C_1^1| = |C_1^2|, |Y^1| = |Y^2| = (K-(k+1))|C_1^1|$, and that l separates $C_1^1 \cup Y^2$ from $C_1^2 \cup Y^1$.

We partition the convex $(k+1)$-clustering $\mathcal{C}_1$ into two convex $(k+1)$-clusterings $\mathcal{C}_1^1, \mathcal{C}_1^2$ of equal sizes by cutting each cluster of $\mathcal{C}_1$ into two equal parts (the cluster C_1 is cut into the above subsets C_1^1, C_1^2).

We obtain a partition of the set X into two 1-configurations $X^j = (\mathcal{C}_1^j; C_1^j; Y^j), j = 1, 2$. Applying first Lemma 1 ($2k^2$-times) and then Lemma 2 on each part, we get a partition of each of the two 1-configurations X^j into at most $\left(\prod_{s=1}^{2k^2-1} 8s^2\right) \cdot 2K \log K = k^{O(k^2)}$ convex k-clusterings and a remaining set of size at most $2 \cdot k^{O(k^2)} \cdot 25(2k^2)^2K + k^{O(k^2)} \cdot 4^k = k^{O(k^2)}$. Theorem 3.3 follows. □

Proof of Lemma 1

Claim 1. *If a family $\{C_1, \dots, C_s, Y\}$ of $s+1$ finite point sets has the same-type property, then there is a closed strip S bounded by two parallel lines such that $Y \subset S$ and $C_i \cap S = \emptyset$ for $i = 1, \dots, s$.*

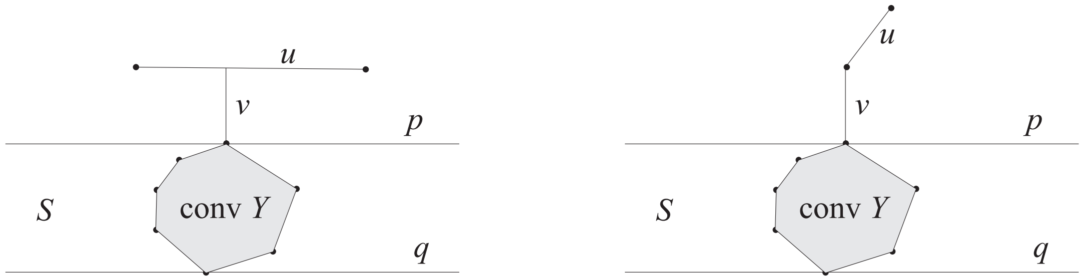


Fig.3

Proof. Let u be a segment connecting a pair of points in $C_1 \cup \cdots \cup C_s$ such that its distance to conv$Y$ is smallest possible. Let v be a segment of length dist$(u, \text{conv}Y)$ connecting a point of u with a point of conv$Y$ (see Fig.3). Further, let p be the line perpendicular to v going through the point $v \cap$ conv$Y$, and let q be the other line of the same direction tangent to conv$Y$.

Then the strip S bounded by the lines p, q contains conv$Y$ and no point of $C_1 \cup \cdots \cup C_s$ (otherwise a segment connecting a point of $S \cap (C_1 \cup \cdots \cup C_s)$ with one of the end-points of u would be closer to conv$Y$ than u – a contradiction). □

Claim 2. *(i) Let $N \geq 1$ be an integer, and let C, Y be two disjoint finite sets of points in the plane such that $|Y| = N \cdot |C|$ and $C \cup Y$ contains no pair of points on a horizontal line. Then there are partitions $Y = Y^1 \cup Y^2 \cup Y^{\text{rem}}$, $C = C^1 \cup C^2 \cup C^{\text{rem}}$ such that $Y^1 \cup C^2$ is separated from $Y^2 \cup C^1$ by a horizontal line, $|Y^x| = N \cdot |C^x|$ for $x \in \{1, 2, \text{rem}\}$, and $|C^{\text{rem}}| \leq 1$.*

(ii) Let C, D, Y be three pairwise disjoint finite sets of points with $|Y| = N \cdot |C| = N \cdot |D|$, where $N \geq 1$ is an integer. Suppose further that $C \cup D \cup Y$ is in general position and any two of the sets C, D, Y are separated from each other by a horizontal line. Then there are partitions $Y = Y^1 \cup Y^2 \cup Y^{\text{rem}}$, $C = C^1 \cup C^2 \cup C^{\text{rem}}$, $D = D^1 \cup D^2 \cup D^{\text{rem}}$ such that:

1. *$|Y^x| = N \cdot |C^x| = N \cdot |D^x|$ for each $x \in \{1, 2, \text{rem}\}$,*
2. *$|C^{\text{rem}}| = |D^{\text{rem}}| \leq 1$, and*
3. *the triple (C^j, D^j, Y^j) has the same-type property, for $j = 1, 2$.*

Proof. **(i)** Let NC be the multiset of points of C, each taken with multiplicity N. Consider the multiset $Y \cup NC$ of size $2N|C|$, and choose a horizontal line p such that at most $N|C|$ points of $Y \cup NC$ lie strictly above p and also at most $N|C|$ points of $Y \cup NC$ lie strictly below p.

If no point of C lies on p, we may suppose that also no point of Y lies on p. The line p determines suitable partitions of Y, C with $Y^{\text{rem}} = C^{\text{rem}} = \emptyset$ in this case.

If a point $c \in C$ lies on p, set $C^{\text{rem}} := \{c\}$ and put N suitable points of Y to Y^{rem}. Then the line p will again determine suitable partitions of $Y \setminus Y^{\text{rem}}, C \setminus C^{\text{rem}}$.

(ii) We distinguish two cases (other cases are obtained by symmetry).

Case 1: the sets C, D, Y appear in the order C, D, Y from top to bottom. (see Fig.4)

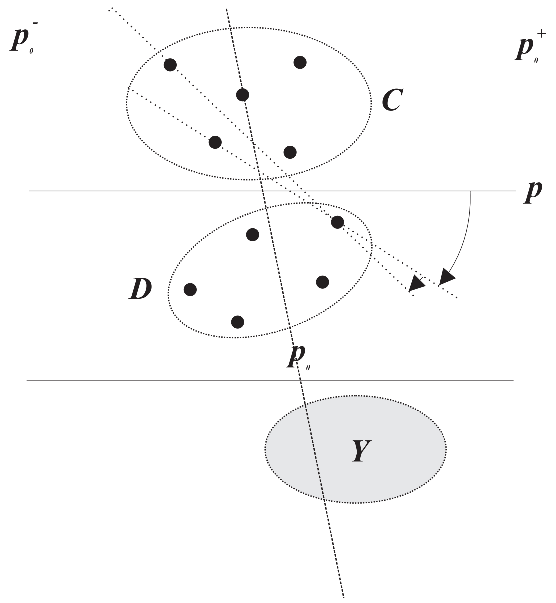


Fig.4

We use a "sweep-line" argument. We take a horizontal line p separating C from D. We rotate p in the clockwise order. During the rotation we are changing the center of rotation so that at each moment p cuts $C \cup D$ into two equal parts (the partition determined by p changes at finitely many moments when two points of $C \cup D$ lie on p) – see Fig.4 (several positions of p during the rotation are shown by dotted lines). Moreover, we can proceed so that p always contains at most one point of Y. Let p^+ be the open halfplane of points to the right of p, and let p^- be the open halfplane opposite to p^+. We rotate p until we achieve $|p^+ \cap Y| \geq N \cdot (|p^+ \cap C| + \frac{1}{2}|p \cap C|)$. Then we stop and denote the final position of p by p_0.

If p passes through a point of Y immediately before it gets to its final position p_0, then p_0 contains no point of $C \cup D \cup Y$ and it gives us the required partitions

$C^1 = C \cap p_0^+$, $C^2 = C \cap p_0^-$,
$D^1 = D \cap p_0^-$, $D^2 = D \cap p_0^+$,
$Y^1 = Y \cap p_0^+$, $Y^2 = Y \cap p_0^-$,

$C^{\mathrm{rem}} = D^{\mathrm{rem}} = Y^{\mathrm{rem}} = \emptyset$.

Otherwise p contains a point $c \in C$ and a point $d \in D$ either in the position p_0 or immediately before it. In these cases we set

$C^{\mathrm{rem}} = \{c\}$, $D^{\mathrm{rem}} = \{d\}$,

$C^1 = C \cap p_0^+$, $C^2 = C \cap p_0^-$,

$D^1 = D \cap p_0^-$, $D^2 = D \cap p_0^+$,

and we appropriately choose a partition $Y = Y^1 \cup Y^2 \cup Y^{\mathrm{rem}}$ such that $|Y^{\mathrm{rem}}| = N, Y^1 \subset p_0^+, Y^2 \subset p_0^-$.

Case 2: the sets C, D, Y appear in the order C, Y, D from top to bottom.

We take a vertical line to the left of $C \cup D \cup Y$ and sweep it to the right so that it cuts C and D in the same proportions at any moment. We define p^+ and p^- as above and stop as soon as $|Y \cap (p \cup p^-)| \geq N \cdot |C \cap p^+|$. Then we continue similarly as in Case 1. Details are left to the reader. $\square$

A *weak* $(s{+}1)$*-configuration* is a $(2s{+}3)$-tuple $(\mathcal{C}_1, \ldots, \mathcal{C}_{s+1}; C_1, \ldots, C_{s+1}; Y)$ such that $(\mathcal{C}_1, \ldots, \mathcal{C}_s; C_1, \ldots, C_s; Y \cup \mathcal{C}_{s+1})$ is an s-configuration and C_{s+1} is one of the clusters of a convex $(k+1)$-clustering $\mathcal{C}_{s+1}$ of size $|\mathcal{C}_{s+1}| = |\mathcal{C}_1|$.

Proof of Lemma 1. Let $X = (\mathcal{C}_1, \ldots, \mathcal{C}_s; C_1, \ldots, C_s; Y)$ be a non-empty s-configuration. By Claim 1, there is a closed strip S such that Y lies in S and $C_1 \cup \cdots \cup C_s$ lies outside of S. We may suppose that the lines bounding S are horizontal. Let S^+ (resp. S^-) be the open halfplane of points above (resp. below) S.

The size of Y is $|X| - s(k+1)\frac{|X|}{K} \geq \frac{|X|}{2}$. Therefore, by Theorem 3.5, Y contains a convex $(k+1)$-clustering $\mathcal{C}_{s+1}$ of size $(k+1)\frac{|X|/2}{K/2} = (k+1)\frac{|X|}{K}$. Fix a cluster C_{s+1} in $\mathcal{C}_{s+1}$.

In the sequel, the weak $(s{+}1)$-configuration $X = (\mathcal{C}_1, \ldots, \mathcal{C}_{s+1}; C_1, \ldots, C_{s+1}; Y \setminus \mathcal{C}_{s+1})$ will be partitioned into a bounded number of $(s+1)$-configurations and a small remaining set.

We apply Claim 2(i) on the sets $C_{s+1}, Y \setminus \mathcal{C}_{s+1}$. We obtain partitions $C_{s+1} = C_{s+1}^1 \cup C_{s+1}^2 \cup C_{s+1}^{\mathrm{rem}}$ and $Y \setminus \mathcal{C}_{s+1} = Y^1 \cup Y^2 \cup Y^{\mathrm{rem}}$ such that $C_{s+1}^1 \cup Y^2$ is separated from $C_{s+1}^2 \cup Y^1$ by a horizontal line, $|C_{s+1}^1| = \frac{|Y^1|}{K-(s+1)(k+1)}$, $|C_{s+1}^2| = \frac{|Y^2|}{K-(s+1)(k+1)}$, $|C_{s+1}^{\mathrm{rem}}| \leq 1$, $|Y^{\mathrm{rem}}| \leq K - (s+1)(k+1)$.

For each $i = 1, \ldots, s$, we take an arbitrary partition $C_i = C_i^1 \cup C_i^2 \cup C_i^{\mathrm{rem}}$ with $|C_i^x| = |C_{s+1}^x|$ for each $x \in \{1, 2, \mathrm{rem}\}$. Finally, for each $i = 1, \ldots, s+1$, we partition the convex $(k+1)$-clustering $\mathcal{C}_i$ into three convex $(k+1)$-clusterings $\mathcal{C}_i^1, \mathcal{C}_i^2, \mathcal{C}_i^{\mathrm{rem}}$ arbitrarily so that, for each $x \in \{1, 2, \mathrm{rem}\}$, C_i^x is one of the clusters of $\mathcal{C}_i^x$.

The two obtained $(2s{+}3)$-tuples $X^i = (\mathcal{C}_1^i, \ldots, \mathcal{C}_{s+1}^i; C_1^i, \ldots, C_{s+1}^i; Y^i)$, $i = 1, 2$, are weak $(s+1)$-configurations forming a partition of $X \setminus X^{\mathrm{rem}}$, where

$|X^{\mathrm{rem}}| \le K$.

Both X^1 and X^2 satisfy the assumptions in the following lemma:

Claim 3. *Let $X = (\mathcal{C}_1, \ldots, \mathcal{C}_{s+1}; C_1, \ldots, C_{s+1}; Y)$ be a weak $(s+1)$-configuration such that the sets C_{s+1}, Y are separated from each other by a horizontal line and also each $C_i, i = 1, \ldots, s$, is separated from $C_{s+1} \cup Y$ by a horizontal line. Further, let $w(X)$ denote the number of indices $i \in \{1, \ldots, s\}$ such that $C_i \subset S^+$ and the triple C_i, C_{s+1}, Y does not have the same-type property. Then X can be partitioned "componentwise" into two weak $(s+1)$-configurations X_j with $w(X_j) \le w(X)/2$ and a remaining set X_{rem} of size at most K. (A partition of $X = (\mathcal{C}_1, \ldots; Y)$ into $X_1, X_2, X_{\mathrm{rem}}$ is* componentwise*, if $X^j = (\mathcal{C}_1^j, \ldots; Y^j)$, $j = 1, 2, \mathrm{rem}$, and $\mathcal{C}_i = \mathcal{C}_i^1 \cup \mathcal{C}_i^2 \cup \mathcal{C}_i^{\mathrm{rem}}$, $C_i = C_i^1 \cup C_i^2 \cup C_i^{\mathrm{rem}}$, $Y = Y^1 \cup Y^2 \cup Y^{\mathrm{rem}}$, $i = 1, \ldots, s+1$.)*

Proof. Suppose that $C_{i_1}, \ldots, C_{i_r}$ are the clusters C_i contributing to $w(X)$, i.e. they are the clusters C_i in S^+ such that the triple C_i, C_{s+1}, Y does not have the same-type property. Moreover, suppose that they are ordered in the order in which they are seen (within an angle smaller than π) from any point in $C_{s+1} \cup Y$, see Fig.Cij. The linear order $C_{i_1}, \ldots, C_{i_r}$ is well-defined because the family $\{C_1, \ldots, C_s, C_{s+1} \cup Y\}$ has the same-type property.

figThe partition $X = X_1 \cup X_2 \cup X_{\mathrm{rem}}$

We apply Claim 2(ii) on the triple of sets $C_{i_{\lceil r/2 \rceil}}, C_{s+1}, Y$. It gives us partitions $C_{i_{\lceil r/2 \rceil}} = C^1_{i_{\lceil r/2 \rceil}} \cup C^2_{i_{\lceil r/2 \rceil}} \cup C^{\mathrm{rem}}_{i_{\lceil r/2 \rceil}}$, $C_{s+1} = C^1_{s+1} \cup C^2_{s+1} \cup C^{\mathrm{rem}}_{s+1}$, $Y = Y^1 \cup Y^2 \cup Y^{\mathrm{rem}}$.

We further partition each set $C_i, i \neq i_{\lceil r/2 \rceil}, s+1$, into three parts $C_i^1, C_i^2, C_i^{\mathrm{rem}}$ arbitrarily so that $|C_i^x| = |C_{s+1}^x|$ for each $x \in \{1, 2, \mathrm{rem}\}$. Finally, any partition of each $(k+1)$-clustering $\mathcal{C}_i$ into three $(k+1)$-clusterings $\mathcal{C}_i^1, \mathcal{C}_i^2, \mathcal{C}_i^{\mathrm{rem}}$ such that C_i^x is a cluster of $\mathcal{C}_i^x$ ($i = 1, \ldots, s+1, x = 1, 2, \mathrm{rem}$) finishes the required partition of X into two weak $(s+1)$-configurations $X_j = (\mathcal{C}_1^j, \ldots, Y^j)$, $j = 1, 2$, and a remaining set $\mathcal{C}_1^{\mathrm{rem}} \cup \cdots \cup \mathcal{C}_{s+1}^{\mathrm{rem}} \cup Y^{\mathrm{rem}}$ of size at most K. (For each $j = 1, 2$ we have $w(X_j) \le r/2 = w(X)/2$ because either only the clusters $C^j_{i_t}, 1 \le t < \lceil r/2 \rceil$, or only the clusters $C^j_{i_t}, \lceil r/2 \rceil < t \le r$, contribute to $w(X_j)$ – see Fig.Cij.) □

Since the partition in Claim 3 is componentwise, each of the subconfigurations X_1, X_2 obtained in Claim 3 satisfies the assumptions in Claim 3 and Claim 3 can be again applied on each of them. We iterate Claim 3 $\lfloor \log_2 s+1 \rfloor$ times. We obtain a partition of a weak $(s+1)$-configuration X into (at most) $2^{\lfloor \log_2 s+1 \rfloor} \le 2s$ weak $(s+1)$-configurations X_j with $w(X_j) = 0$ (since after $\lfloor \log_2 s+1 \rfloor$ iterations of Claim 3 we get $w(X_j) \le w(X)/2^{\lfloor \log_2 s+1 \rfloor} < s/s = 1$) and a remaining set of size at most $(2 \cdot 2^{\lfloor \log_2 s+1 \rfloor} - 1)K < 4sK$. Analogously, we can consequently partition each X_j into at most $2s$ weak

$(s+1)$-configurations X_{jl} and a remaining set of size at most $4sK$ such that $w(X_{jl}) = 0$ holds also if we replace S^+ by S^- in the definition of the parameter w. This means that the configurations X_{jl} are $(s+1)$-configurations.

Thus, each of the two weak $(s+1)$-configurations X^1, X^2 obtained above Claim 3 can be partitioned into at most $(2s)^2 = 4s^2$ $(s+1)$-configurations and a remaining set of size at most $(2s+1)4sK \leq 12s^2K$. Since X^1, X^2 form a partition of $X \setminus X^{\text{rem}}$ (where $|X^{\text{rem}}| \leq K$), we obtain a partition of X into $8s^2$ $(s+1)$-configurations and a remaining set of size at most $2 \cdot 12s^2K + K \leq 25s^2K$. This finishes the proof of Lemma 1. □

Proof of Lemma 2

The proof of Lemma 2 relies on another famous result of Erdős and Szekeres [ES35]:

Theorem 3.8 (Erdős–Szekeres lemma). *Any sequences of $(k-1)^2+1$ distinct real numbers contains a monotone subsequence of length k.*

Proof of Lemma 2. Let $X = (\mathcal{C}_1, \ldots, \mathcal{C}_s, C_1, \ldots, C_s, Y)$ be an s-configuration, where $s = 2k^2$. By Claim 3 and by the pigeon-hole principle, there is a line l separating Y from at least k^2 of the clusters C_i. Let $C_{i_1}, \ldots, C_{i_{k^2}}$ be k^2 such clusters listed in the order in which they are seen (within an angle smaller than π) from any point of Y.

By the Erdős-Szekeres theorem on monotone subsequences (Theorem 3.8), the sequence $i_1, \ldots, i_{k^2}$ contains a monotone subsequence $j_1, \ldots, j_{k-1}$ of length $k-1$.

Claim 4. *The k sets $Y, C_{j_1}, \ldots, C_{j_{k-1}}$ are in convex position.*

Proof. Since Y is separated from $\bigcup_{i=1}^{k-1} C_{j_i}$ by a line, it suffices to show that for each $i = 1, \ldots, k-3$ the set $C_{j_{i+1}}$ cannot lie in the convex hull of $C_{j_i} \cup C_{j_{i+2}} \cup Y$. We use the sequential same-type property of the sequence $(C_1, \ldots, C_s, Y)$. If the sequence j_i, j_{i+1}, j_{i+2} is increasing, then $(C_{j_i}, C_{j_{i+1}}, C_{j_{i+2}}, Y)$ has the sequential same-type property. Otherwise the sequence j_i, j_{i+1}, j_{i+2} is decreasing and $(C_{j_{i+2}}, C_{j_{i+1}}, C_{j_i}, Y)$ has the sequential same-type property. In the first case the triple $\{C_{j_i}, C_{j_{i+1}}, C_{j_{i+2}} \cup Y\}$ has the same-type property, in the latter case the triple $\{C_{j_{i+2}}, C_{j_{i+1}}, C_{j_i} \cup Y\}$ has the same-type property. In either case, $C_{j_{i+1}}$ lies outside of the convex hull of $C_{j_i} \cup C_{j_{i+2}} \cup Y$. The lemma is proved. □

We now need the following corollary of Theorem 3.5 and a simple observation.

Corollary 3.1. *The set Y is a disjoint union of at most $K \log K$ convex k-clusterings and two remaining sets Y_0, Y_1 with $|Y_0| = \frac{|X|}{K}, |Y_1| \leq f(k)$.*

Proof. By Theorem 3.5, Y is a disjoint union of a convex k-clustering and a remaining set Y' of size at most $\max\{\left(1 - \frac{1}{K}\right)|Y|, f(k)\}$. Applying Theorem 3.5 on Y', we get a partition of the set Y into two convex k-clusterings and a remaining set of size at most $\max\{\left(1 - \frac{1}{K}\right)^2 |Y|, f(k)\}$.

After $K \log K$ applications of Theorem 3.5, we obtain a partition of Y into $K \log K$ convex k-clusterings and a remaining set Y_{rem} of size at most

$$\max\{\left(1 - \frac{1}{K}\right)^{K \log K} |Y|, f(k)\} \leq \max\{\frac{1}{K}|Y|, f(k)\} \leq \frac{|X|}{K} + f(k).$$

If $|Y_{\text{rem}}| \geq \frac{|X|}{K}$ then we finish the whole partition by partitioning Y_{rem} into two subsets Y_0, Y_1, where $|Y_0| = \frac{|X|}{K}$.

Otherwise we repeatedly enlarge Y_{rem} by choosing an arbitrary of the $K \log K$ k-clusterings and moving one point from each of its clusters to Y_0. Each repetition of this step enlarges $|Y_{\text{rem}}|$ by k, so we can achieve that $\frac{|X|}{K} \leq |Y_{\text{rem}}| \leq \frac{|X|}{K} + (k-1)$, and then we make any partition $Y_{\text{rem}} = Y_0 \cup Y_1$ with $|Y_0| = \frac{|X|}{K}$. □

Observation 2. *Every convex $(k+1)$-clustering is a union of $k+1$ convex k-clusterings and a remaining set of size at most $k-1$.*

Proof. Let $\mathcal{C}$ be a convex $(k+1)$-clustering and let t be the size of a cluster in $\mathcal{C}$. Set $t_0 := t \bmod k$.

We can easily partition $\mathcal{C}$ into t_0 points and $k+1$ convex k-clusterings, each disjoint from one of the clusters of $\mathcal{C}$ and having size $k \cdot \lfloor t/k \rfloor$ or $k \cdot \lfloor t/k + 1 \rfloor$. □

By Corollary 3.1, Y is a disjoint union of at most $K \log K$ convex k-clusterings and two remaining sets Y_0, Y_1 with $|Y_0| = \frac{|X|}{K}, |Y_1| \leq f(k)$. By Claim 4, the sets $Y_0, C_{j_1}, \ldots, C_{j_{k-1}}$ form a convex k-clustering. Further, each of the $k-1$ sets $\mathcal{C}_{j_1} \setminus C_{j_1}, \ldots, \mathcal{C}_{j_{k-1}} \setminus C_{j_{k-1}}$ is a convex k-clustering.

By Observation 2, each of the remaining $2k^2 - (k-1)$ $(k+1)$-clusterings $\mathcal{C}_i$ can be partitioned into $k+1$ convex k-clusterings and a remaining set of at most $k-1$ points.

The s-configuration X is partitioned into at most $K \log K + 1 + (k-1) + (2k^2 - (k-1))(k+1) < 2K \log K$ convex k-clusterings and a remaining set of size at most $f(k) + (2k^2 - (k-1))(k-1) < 4^k$. $\square$

Concluding remarks

1. Upper bounds on the constants $c(k), c'(k)$. Our proof gives Theorem 3.3 with $c(k) = k^{O(k^2)}$, $c'(k) = k^{O(k^2)}$. It is easy to lower $c'(k)$ to $f(k)-1$ (and keep $c(k) = k^{O(k^2)}$ at the same time). As long as the remaining set of points has size bigger than $f(k) - 1$, it contains a convex k-clustering of size k, i.e., k points in convex position, and we may reduce the size of the remaining set by it. The number of convex k-clusterings will increase by at most $k^{O(k^2)} \cdot k^{-1} = k^{O(k^2)}$, thus staying at most $k^{O(k^2)}$. This bound on $c'(k)$ cannot be improved in general, since any set of $f(k) - 1$ points with no k points in convex position contains no non-empty convex k-clustering.

If $|X|$ is sufficiently large, then $c'(k)$ can be lowered to $rr(k)$, the so-called *Ramsey-remainder for convex sets*, defined as the smallest number r such that any set of sufficiently many points in general position in the plane can be partitioned into k-tuples of points in convex position and a remaining set of size at most r (see [ETV96] for estimates on the numbers $rr(k)$).

2. An upper bound on ε_k and a lower bound on $c(k)$. Let $C_{k/2}$ be the set of $2^{k/2-2}$ points with no convex $(k/2)$-gon constructed by Erdős and Szekeres [ES61]. We may suppose that no pair of points in $C_{k/2}$ lies on a horizontal line. We replace each point of C_k by the same number L of points which are very close together, form a so-called L-cup (e.g., see [TV98] for the definition), and any line trough two of them is almost horizontal.

Then any convex k-clustering has all points of at least $\approx k/2$ clusters contained in the same L-cup, and therefore its size is at most ($\approx 2 \cdot 2^{2-k/2}$)-fraction of our set. This gives an upper bound $\approx \left(\sqrt{1/2}\right)^k$ on the constant ε_k in Theorem 3.2 (it can be improved to $\approx (1/2+\varepsilon)^k$ by considering $C_{(1-\varepsilon)k}$ instead of $C_{k/2}$). Consequently, we get a lower bound $\approx \left(\sqrt{2}\right)^k$ on the constant $c(k)$ in Theorem 3.3 (it can be improved to $\approx (2-\varepsilon)^k$).

3. Higher dimensions. For any $d \geq 3$, an analogue of Theorem 3.3 in $\mathbb{R}^d$ holds. We simply project the set $X \subset \mathbb{R}^d$ orthogonally to a plane chosen so that the projection is a set in general position. Then we apply Theorem 3.3. The obtained partition gives us a partition of the original set X into at most $k^{O(k^2)}$ convex k-clusterings and a remaining set of size at most $k^{O(k^2)}$. In fact, if $|X|$ is divisible by k and sufficiently large, then we can partition the whole set X into convex k-clusterings because the Ramsey-remainder for convex sets is 0 in this case (see [KA02]).

3.4 The partitioned "Same Type Lemma"

Denote by $n_{m,d}$ the smallest number (probably infinite) such that for all finite sets $X_1, \dots, X_m \subset \mathbb{R}^d$ a balanced separated μ-partition with at most $n_{m,d}$ columns exists. To prove the theorem, we have to show that $n_{m,d}$ is finite, and the following two lemmas will do this.

Lemma 3.1. *If $n_{d+1,d}$ is finite for a given dimension d, then so are all $n_{m,d}$ numbers, where $m > d+1$ and*

$$n_{m,d} \leq (n_{d+1,d})^{\binom{m}{d+1}}$$

Lemma 3.2. *The number $n_{2,1}$ is equal to 2. If $n_{d+1,d-1}$ is finite, then*

$$n_{d+1,d} \leq 2n_{d+1,d-1}$$

Proof of Theorem 3.6. If $n_{d,d-1}$ is finite, then by the first lemma $n_{d+1,d-1}$ is finite too and by the second lemma $n_{d+1,d}$ is also finite. Since $n_{2,1}$ is finite, we get by induction that $n_{d+1,d}$ is finite for all d. The first lemma proves now our theorem. □

Proof of the lemmas

We will need new notions, the partial μ-partition and the refinement of the μ-partition:

The μ-partition M is partial on the set $I \subset [m] = \{1, \dots, m\}$, if the matrix M has only entries in the rows $i \in I$ and it satisfies the conditions for μ-partitions in the rows $i \in I$. A partial μ-partition is balanced if the cardinality of the sets in one column is equal, of course only for sets which are in a row with index from I. A partial balanced μ-partition can be extended to a balanced μ-partition M. For example partition all other sets X_i $i \notin I$ by taking care only on the cardinality of the entries.

If we have a balanced μ-partition M, we can replace some of its columns by a balanced μ-partition of the sets of these columns. Then we get a new balanced μ-partition. We will call this the refinement of the μ-partition. Note that the refinement of a refinement of a μ-partition is also a refinement of the original one.

Proof of Lemma 3.1. Let $I_1, \ldots, I_r \subset [m]$ be all the subsets of $[m]$ of $d+1$ elements ($r = \binom{m}{d+1}$). Take the "trivial" μ-partition on the sets with 1 column, $M_{i1} = X_i \quad i = 1, \ldots, m$, and refine it in several steps, such that after the l-th step there will be no hyperplane meeting the convex hulls of the $d+1$ sets $M_{l_1 j}, \ldots, M_{l_{d+1} j}$ for all j, where $I_l = \{l_1, \ldots, l_{d+1}\}$.

In the l-th step we will refine all the columns in the following way: We take the sets in a column and produce a balanced separated μ-partition for the sets in the rows corresponding to I_l. This is a partial μ-partition for all the sets and has at most $n_{d+1,d}$ new columns which will replace the one column we took. We extend it to a balanced μ-partition and do this refinement for all the columns. At the end we will have a balanced μ-partition, and there will be no hyperplane intersecting the convex hull of any $d+1$ sets in the same column. This means that the balanced μ-partition is separated and the lemma is proved since in each of the r steps the size of the matrix was multiplied by at most $n_{d+1,d}$. $\square$

In Lemma 3.2 first we prove the starting step ($d = 1$) and then the induction.

Proof of Lemma 3.2. Let $X_1, X_2 \subset \mathbb{R}$ $|X_1| = |X_2| = l$ then there is a point $a \in \mathbb{R}$, such that

$$|\{x \mid x \in X_1 \cup X_2 \quad x \leq a\}| = l$$

and define

$$M_{11} = \{x \in X_1 | x \leq a\}$$
$$M_{12} = \{x \in X_1 | x > a\}$$
$$M_{21} = \{x \in X_2 | x > a\}$$
$$M_{22} = \{x \in X_2 | x \leq a\}$$

Now M is an μ-partition with 2 columns on the sets X_1 and X_2 and

$$|M_{11}| + |M_{12}| = |M_{11}| + |M_{22}| = |M_{21}| + |M_{22}| = l$$

which shows that M is balanced and separated, too.

In the second part of the lemma we have to show that $n_{d+1,d} \leq 2n_{d+1,d-1}$ assuming $n_{d+1,d-1}$ is finite. Let $X_1, \ldots, X_{d+1} \subset \mathbb{R}^d$ be finite sets of size l and suppose that the set $\cup_i X_i$ is in general position. It is well known that there is a hyperplane S, such that the perpendicular projection of $\cup_i X_i$ to S is still in general position (in fact, almost all hyperplanes satisfy this condition). Let $X_i^{'}$ be the image of X_i under this projection, and take a balanced separated μ-partition $M^{'}$ on these sets, with at most $n_{d+1,d-1}$ columns. Let M be the balanced μ-partition on the sets $X_1, \ldots, X_{d+1}$ induced by M' and from now on we examine only one of its columns, and for simplicity of notation we call these sets now $X_1, \ldots, X_{d+1}$. To prove our lemma, we have to divide all the sets $X_1, \ldots, X_{d+1}$ into 2 parts in a balanced way that both parts are separated. The projection of these sets are separated in the hyperplane S. By Radon's theorem there is only one partition of $d+1$ points in $\mathbb{R}^{d-1}$, among all the 2^d, which cannot be separated by a hyperplane, which means, in combinatorial sense, that there are $2^d - 1$ hyperplanes in S which miss $\cup_i \text{conv} X_i'$ and divide them in different ways. Take for each hyperplane in S the hyperplane in $\mathbb{R}^d$ as its inverse image under the perpendicular projection to S. These hyperplanes avoid $\cup_i X_i$ and divide them in $2^d - 1$ different ways. Since there are 2^d different bipartitions of the set $[d+1]$, there is one remaining bipartition ($I \subset [d+1], I^c = [d+1] \setminus I$), for which we need separability.

For technical reasons it will be useful to change the point-set to a measurable set which is close to the point-set in the following way: Replace all points by a ball centered at the point and of radius ε, and denote the new sets by $K_1, \ldots, K_{d+1}$. Since $\cup_i X_i$ is in general position it is possible to choose ε so small that no hyperplane can meet $d+1$ of the balls.

Now we assume that the measure is distributed uniformly on these balls, such that each of them have measure one(i.e. we divide the Lebesgue measure by the volume of the ball of radius ε and denote it by ν). We define a function $f : S^d :\to \mathbb{R}^d$ which is odd. Take $H = (t, \mathbf{h}) \in S^d$, where t is a real number and $\mathbf{h}$ is a d-dimensional vector not equal to zero. Let

$$H_i^+ = \{x | \langle x, \mathbf{h} \rangle \ \geq t\} \cap K_i$$

and

$$H_i^- = \{x | \langle x, \mathbf{h} \rangle \ < t\} \cap K_i = K_i \setminus H_i^+$$

Further let

$$H_i^* = \begin{cases} H_i^+ & \text{if } i \in I \\ H_i^- & \text{if } i \notin I \end{cases}$$

Now we can define f at $(t, \mathbf{h})$ as follows:

$$f((t, \mathbf{h})) = \big(\nu(H_1^*) - \nu(H_{d+1}^*), \ldots, \nu(H_d^*) - \nu(H_{d+1}^*)\big) \in \mathbb{R}^d$$

It's easy to check that $(-H)_i^+ = H_i^-$, and we can extend the definition to $H = (+1, \mathbf{0})$ by choosing

$$(+1, \mathbf{0})_i^* = H_i^* = \begin{cases} \emptyset & \text{if } i \in I \\ K_i & \text{if } i \notin I \end{cases}$$

and for $H = (-1, \mathbf{0})$

$$(-1, \mathbf{0})_i^* = H_i^* = \begin{cases} K_i & \text{if } i \in I \\ \emptyset & \text{if } i \notin I \end{cases}$$

The function f maps S^d into $\mathbb{R}^d$ and it is continuous and odd so Borsuk's theorem applies: there exists an $H \in S^d$ (of course not one of the poles) such that $f(H) = \mathbf{0}$. The hyperplane $S_H = \{x | \langle x, \mathbf{h} \rangle = t\}$ meets at most d balls of radius ε with a center in $\cup_i X_i$, therefore one of the sets K_j has been avoided by S_H and so $\nu(H_j^*) = r$ is an integer number and a hyperplane T (a slight perturbation of the one corresponding to H) can be chosen such that it doesn't meet $\cup_i X_i$ and on a given side it contains r points from X_i if $i \in I$ otherwise it contains $l - r$ points. If T^+ denotes one half-space generated by T, then the μ-partition of two columns are now the subsets of $X_i \cap T^+$ if $i \in I$ and $X_j \backslash T^+$ if $j \notin I$. The properties of H show that this partition is balanced and separated. □

Remark If $d = 2$ the above proof gives the following statement that will be needed in the next section: If three equal sized sets X_1, X_2 and X_3 are separated by two parallel lines in the plane, then there exists a line f which divides the sets into X_{i1} and X_{i2} for $i = 1, 2, 3$, such that X_{11}, X_{22} and X_{31} are on the same side of f and

$$\frac{|X_{11}|}{|X_{12}|} = \frac{|X_{21}|}{|X_{22}|} = \frac{|X_{31}|}{|X_{32}|}$$

are equal (probably infinite).

3.5 Better bound on the $k = 4$ case

First we give an other proof for the case $k = 4$, and after this we improve it 26, as said at the beginning. Observe that among any 5 points in the plane there are 4 in convex position, and therefore it would be enough to bound X_0 by a finite number, since then we can remove 4 points in convex position, and increase N by 1, adding a column with 4 sets of 1 element to our partition.

Proof. Let us divide X in 4 equal sets X_1, X_2, X_3 and X_4 by dropping at most 3 points. By Theorem 3.6 there exists a balanced, separated μ-partition M with at most $n_{4,2}$ columns. Some columns of this matrix are already in convex position, and we can use them as they are in our decomposition. So we can reduce further investigations to a given column, where the sets are not in convex position. From now on we suppose that $X = \cup_i^4 X_i$, where X_1, X_2, X_3, X_4 are separated and of equal size, say $2l$, and for any transversal

$\{x_1, x_2, x_3, x_4\}$, $x_i \in X_i$ the point x_4 is in the convex hull of the other three, $x_4 \in \text{conv}\{x_1, x_2, x_3\}$. The assumption of even size enlarges X_0. Because the sets are separated, there are lines a_1, a_2 and a_3, such that on one side of a_i is X_i, X_4 and all other sets X_j, $j \neq i, 4$ are on the opposite side. Now choose the lines e_1, e_2 and e_3, each parallel to the corresponding line a_i, such that e_i divides the set X_i into two sets of equal size, X_{i1} and X_{i2}, with X_{i2} on the same side of e_i as X_4. Clearly there exists a line l_i parallel to a_i and e_i separating X_i from X_4. Now the two parallel lines e_i and l_i separate X_{i1}, X_{i2} and X_4. Let us denote by $2S$ the set derived from S by doubling and slightly perturbing its points. Now choose the line f_i to divide the equal sized sets $2X_{i1}, 2X_{i2}$, and X_4 as in the remark of the previous section. Since we doubled some of the sets, we will need some correction at the end, but this means only a constant in the size of X_0.

Let X'_{4i}, X''_{4i} the two halves of X_4 divided by f_i. Corresponding to the previous notation, let $X'_{i1}, X''_{i1}, X'_{i2}, X''_{i2}$ the sets derived from the sets X_{i1}, X_{i2} after dividing them by f_i, i.e.:

$$\alpha = \frac{|X'_{11}|}{|X_{11}|} = \frac{|X'_{12}|}{|X_{12}|} = \frac{|X'_{41}|}{|X_4|},$$
$$\beta = \frac{|X'_{21}|}{|X_{21}|} = \frac{|X'_{22}|}{|X_{22}|} = \frac{|X'_{42}|}{|X_4|},$$
$$\gamma = \frac{|X'_{31}|}{|X_{31}|} = \frac{|X'_{32}|}{|X_{32}|} = \frac{|X'_{43}|}{|X_4|}$$

It is easy to see that the following 4-tuples of sets are separated and in convex position (Fig. 1.):

$$\left.\begin{array}{llllr} X'_{11}, & X'_{12}, & X'_{41}, & X_2 & (\lambda_1) \\ X''_{11}, & X''_{12}, & X''_{41}, & X_3 & (\lambda_1) \\ X'_{21}, & X'_{22}, & X'_{42}, & X_3 & (\lambda_2) \\ X''_{21}, & X''_{22}, & X''_{42}, & X_1 & (\lambda_2) \\ X'_{31}, & X'_{32}, & X'_{43}, & X_1 & (\lambda_3) \\ X''_{31}, & X''_{32}, & X''_{43}, & X_2 & (\lambda_3) \end{array}\right. (i)$$

For example choose the first 4-tuple, $X'_{11}, X'_{12}, X'_{41}, X_2$ One can see on Figure 5. that the line e_1 separates X'_{11} from the other 3 sets. Further the lines f_1, a_3, a_1 in order separate X'_{12}, X'_{41} and X_2 from the remaining 3 sets.

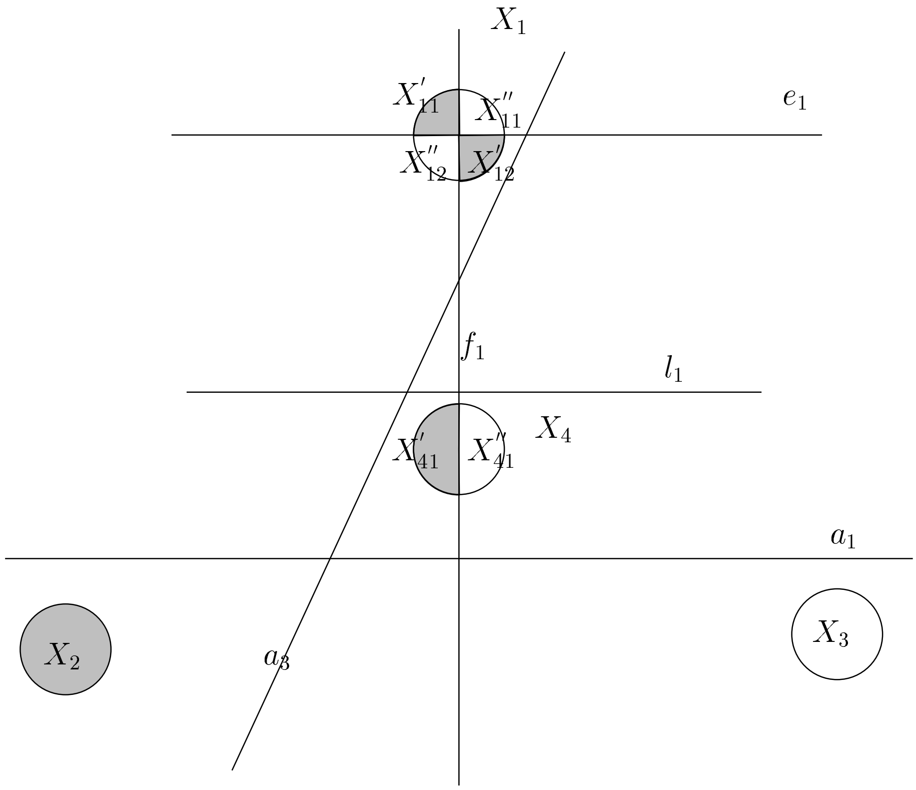


Fig.5

These six 4-tuples cover the set X several times, so we need to reduce them so that they give a partition of X. The 4 sets of one 4-tuple are not equal sized, therefore we prescribe the size of the sets in order by $\lambda_1 X'_{11}, \lambda_1 X''_{11}, \lambda_2 X'_{21}, \lambda_2 X''_{21}, \lambda_3 X'_{31}, \lambda_3 X''_{31}$ To construct the sets themselves, we have to check first that at least the size of the sets X_i are as big as the sum of it's parts. Therefore we will get 4 equalities for the ratios $\lambda_1, \lambda_2, \lambda_3$ which have to be fulfilled in order to continue the construction. Let us denote the subsets by Y_{**} with subscripts like the corresponding X_{**} sets , except Y'_i and Y''_i for $i = 1, 2, 3$ which are the 4th sets of the 4-tuples listed above. First we calculate their sizes, in order $\lambda_1 \alpha l, \lambda_1(1-\alpha)l, \lambda_2 \beta l, \lambda_2(1-\beta)l, \lambda_3 \gamma l, \lambda_3(1-\gamma)l$, where $0 \leq \lambda_i \leq 1$ for all $i = 1, 2, 3$.

$$\begin{aligned}
\lambda_1 \alpha l &= |Y'_{11}| = |Y'_{12}| = |Y'_{41}| = |Y'_2| \\
\lambda_1(1-\alpha) l &= |Y''_{11}| = |Y''_{12}| = |Y''_{41}| = |Y''_3| \\
\lambda_2 \beta l &= |Y'_{21}| = |Y'_{22}| = |Y'_{42}| = |Y'_3| \\
\lambda_2(1-\beta) l &= |Y''_{21}| = |Y''_{22}| = |Y''_{42}| = |Y''_1| \\
\lambda_3 \gamma l &= |Y'_{31}| = |Y'_{32}| = |Y'_{43}| = |Y'_1| \\
\lambda_3(1-\gamma) l &= |Y''_{31}| = |Y''_{32}| = |Y''_{43}| = |Y''_2|
\end{aligned}$$

Since

$$\begin{aligned}
X_1 &= Y'_{11} \cup Y''_{11} \cup Y'_{12} \cup Y''_{12} \cup Y'_1 \cup Y''_1 \\
X_2 &= Y'_{21} \cup Y''_{21} \cup Y'_{22} \cup Y''_{22} \cup Y'_2 \cup Y''_2 \\
X_3 &= Y'_{31} \cup Y''_{31} \cup Y'_{32} \cup Y''_{32} \cup Y'_3 \cup Y''_3 \\
X_4 &= Y'_{41} \cup Y''_{41} \cup Y'_{42} \cup Y''_{42} \cup Y'_{43} \cup Y''_{43}
\end{aligned}$$

where each line is a partition of the corresponding X_i, we get 4 equalities for the sizes (each equation divided by l)

$$\begin{aligned}
&2 = 2\lambda_1 + \lambda_2(1-\beta) + \lambda_3\gamma \\
&2 = 2\lambda_2 + \lambda_3(1-\gamma) + \lambda_1\alpha \\
&2 = 2\lambda_3 + \lambda_1(1-\alpha) + \lambda_2\beta = 4 - \lambda_1(1+\alpha) - \lambda_2(2-\beta) \\
&2 = \lambda_1 + \lambda_2 + \lambda_3
\end{aligned}$$

The solution is

$$\begin{aligned}
\lambda_1 &= \frac{2 + 2\beta\gamma - 2\gamma}{3 + \alpha\beta + \beta\gamma + \gamma\alpha - \alpha - \beta - \gamma} \\
\lambda_2 &= \frac{2 + 2\gamma\alpha - 2\alpha}{3 + \alpha\beta + \beta\gamma + \gamma\alpha - \alpha - \beta - \gamma} \\
\lambda_3 &= \frac{2 + 2\alpha\beta - 2\beta}{3 + \alpha\beta + \beta\gamma + \gamma\alpha - \alpha - \beta - \gamma}
\end{aligned}$$

where we can check that $0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1$.

Now the sizes of our sets $Y'_{11}, \ldots, Y''_{43}$ are known, but how to choose them. Take the set X_1 and choose first the sets $Y'_{11}, Y''_{11}, Y'_{12}, Y''_{12}$ and the remaining points from X_1 can be distributed between Y'_1 and Y''_1 arbitrarily. The distribution of X_2 and X_3 are similar. To distribute the points of X_4 we take the arrangement of the lines f_1, f_2 and f_3 which partitions the set X_4 into at most 7 parts. In each part we distribute the points into 3 parts of size $\lambda_1 : \lambda_2 : \lambda_3$. It is clear that we can put together the sets $Y'_{41}, \ldots, Y''_{43}$ from these parts.

During these steps we don't care about some (finitely many) points which do not satisfy our partition requirements. As we doubled some points for example, at the end we probably have to remove one point from a set. Since these are just finitely many points they may be appended to X_0. $\square$

Remark This immediately gives a bound of $6n_{4,2}$ but we can obtain 26.

Proof of Theorem 3.4. To show, that $N \leq 26$ we need a better separation of the first 4 sets. We change each point of X to a nowhere zero measure distributed on the plane, concentrated on a disc centered at the point (similarly to the proof of Lemma 2.2). One can achieve this by distributing ϵ measure outside a disc of radius ϵ.

Choose an arbitrary directed line l and divide X into 4 equal parts X_1, X_2, X_3 and X_4 by three lines l_1, l_2, l_3 all parallel to l (Figure 6).

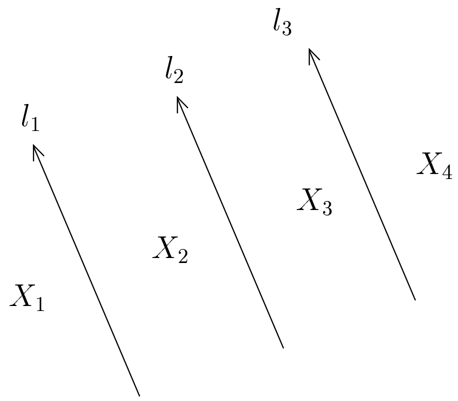


Fig.6

Take the vertical line as l, such that the direction points upwards. For each $0 \leq \varepsilon \leq 1$ there is a well defined halfspace H_ε with boundary line H_ε^0 for which

$$\frac{|H_\varepsilon \cap X_2|}{|X_2|} = 1 - \varepsilon \qquad and$$
$$\frac{|H_\varepsilon \cap X_3|}{|X_3|} = \varepsilon$$

and H_ε is below H_ε^0. Let

$$\beta(\varepsilon) = \frac{|H_\varepsilon \cap X_1|}{|X_1|}$$
$$\gamma(\varepsilon) = 1 - \frac{|H_\varepsilon \cap X_4|}{|X_4|}$$

Clearly $\beta(0) = \gamma(0) = 1$, $\beta(1) = \gamma(1) = 0$ and β, γ are continuous functions in ε. Let

$$\varepsilon(l) = \min_{\varepsilon}\{\min\{\beta(\varepsilon), \gamma(\varepsilon)\} \leq \varepsilon\}$$

We are looking for a line l with

$$\beta(\varepsilon(l)) = \varepsilon(l)$$
$$\gamma(\varepsilon(l)) = \varepsilon(l)$$

At least one of the two equalities holds for every line, so turning l around (by 180^o, at the end the direction of the starting line has been changed) we get that the rolls of β and γ change. Since $\varepsilon(l)$ was a continuous function therefore we had to pass a line l_0 for which

$$\beta(\varepsilon(l_0)) = \varepsilon(l_0) \quad \text{and}$$
$$\gamma(\varepsilon(l_0)) = \varepsilon(l_0)$$

Choosing this line and the corresponding lines l_1, l_2, l_3 and $H^0_\varepsilon(l_0)$ (Figure 7) we get a partition of X into the sets $Z_1, \dots, Z_4, Y_1, \dots, Y_4$, where $X_1 = Z_1 \cup Y_1, \dots, X_4 = Z_4 \cup Y_4$. The shadowed parts are the Z_i and the white parts the Y_i sets by indexing them from left to right.

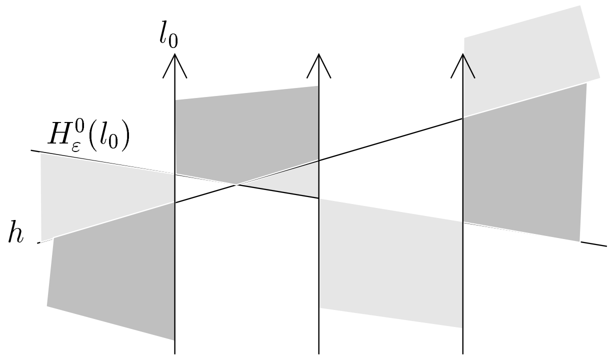


Fig.7

Now separating Z_1, Z_2, Z_4 by h as in the remark after Lemma 2.2 we get two parts. The light shadow part by adding a part of Z_3 to it, is already a separated 4-tuple (nonconvex). Doing the same for Z_1, Z_3, Z_4 we get again a nonconvex separated 4-tuple and the rest is a convex separated 4-tuple. The previous theorem replaces the nonconvex ones by $2 \cdot 6 = 12$ convex parts. The same is true for the sets $Y_1, \dots, Y_4$ and we get at most 26 convex 4-tuples. $\square$

3.6 Asymptotically tight bounds for $n_{m,1}$

From our previous proofs we know that $n_{2,1} = 2$, and that

$$n_{m,1} \leq (n_{2,1})^{\binom{m}{2}} = 2^{m(m-1)/2}$$

which is super-exponential in m, and even for $m = 3$ it gives an upper bound of $n_{3,1} \leq 8$ while the truth should be $n_{3,1} = 5$ (easy as it seems, I could not prove it). We will prove a quadratic upper bound on $n_{m,1}$ and show that it is the truth up to a constant factor.

We prove Theorem 3.7 in two steps.

Let's recall the upper bound of Theorem 3.7 *If $X_1, \ldots, X_m$ are finite sets of equal size on the real line and the set $\bigcup_i^m X_i$ is in general position, then there is a balanced, separated μ-partition with at most $(m-1)^2 + 1$ columns on these sets, i.e.:*

$$n_{m,1} \leq (m-1)^2 + 1$$

Proof. of the upper bound Let $X = \bigcup_i^m X_i$, and $l = |X_i|$ the size of the sets. Take $m - 1$ points, $t_1, \ldots, t_{m-1}$, on the real line which partitions the set X into m equal sets $T_1, \ldots, T_m$:

$$-\infty = t_0 < t_1 < \ldots < t_{m-1} < t_m = \infty$$

and for all $1 \leq i, j \leq m$

$$T_j = \{x \in X | t_{j-1} \leq x < t_j\}$$

Take the intersection of these sets with the sets X_i and denote them for all $1 \leq i, j \leq m$ by

$$A_{i,j} = X_i \cap T_j$$

Define the matrix S of size $m \times m$ with integer entries:

$$S_{i,j} = |A_{i,j}| \quad .$$

The matrix S has the property, that the sum of its entries in a row or a column are equal (constant times a doubly stochastic matrix), and all of its entries are nonnegative. We will call such a nonnegative matrix a "magic matrix". We know (see the Remark below or [MM64]) that S is a positive integer combination of permutation matrices $P_1, \ldots, P_k$, where $k \leq (m-1)^2 + 1$

$$S = \sum_{i=1}^{k} \lambda_i P_i$$

where λ_i is a positive integer. Let σ_i be the permutation of P_i. Now we can construct our μ-partition M of k elements:

$$M_{i,j} \subset A_{i,\sigma_i(j)} \quad \text{and} \quad |M_{i,j}| = \lambda_i$$

and such M exists. (We don't care about how these subsets are chosen, just they have to be disjoint)

It's clear that M is balanced and separated. The points $t_1, \dots, t_{m-1}$ work as separators for each column. □

Remark. We show that a $m \times m$ magic matrix can be written as a positive linear integer combination of at most $m^2 - 2m + 2$ permutation matrices.

All matrices here will have positive integer entries.

First we show that for each magic matrix D, there exists a permutation matrix P, such that $P \leq D$.

Suppose that such a P permutation matrix does not exist. We know (see [MM64]) that D has a $m_1 \times m_2$ submatrix, whose entries are all zero and $m_1 + m_2 > m$. We may suppose that $D_{i,j} = 0$ whenever $i > m - m_1$ and $j > m - m_2$. If T denotes the sum of entries in one row of D, then:

$$T \cdot m_1 = \sum_{i=m-m_1+1}^{m} \sum_{j=1}^{k} D_{i,j} = \sum_{i=m-m_1+1}^{m} \sum_{j=1}^{m-m_2} D_{i,j} \leq \sum_{i=1}^{m} \sum_{j=1}^{m-m_2} D_{i,j} = T \cdot (m - m_2)$$

what yields

$$m_1 + m_2 \leq m$$

This is a contradiction to $m_1 + m_2 > m$.

Now choose one of the permutation matrices $P_1 \leq D$ and choose a_1 such that

$$a_1 P_1 \leq D$$

but

$$(a_1 + 1) P_1 \nleq D$$

.

Let $D_2 = D - a_1 P_1$, and observe that D_2 is a magic matrix again, which has a zero entry, where D has a nonzero (a_1) entry. Do this as long D_i is not the 0-matrix and we get:

$$D = \sum_{i=0}^{k} a_i P_i$$

a positive linear combination of k permutation matrices. Claim: $P_1, \dots, P_k$ are linearly independent. Indeed, if they were not, then some P_j can be written as a linear combination of P_i-s, where $i > j$. When we have choosen P_j we nullified at least one entry of the matrix D_j and therefore all P_i, $i > j$ have a zero value at that entry. This contradiction shows that $P_1, \dots, P_k$ are linear independent permutation matrices, and therefore k is at most the dimension of the space generated by permutation matrices, which is $m^2 - 2m + 2$.

I conjecture that this is the truth in 1-dimension, but could only prove a lower bound which is 4 times smaller. It is enough to work for $m = 2r + 1$.

Recall the lower bound of Theorem 3.7 *For each odd $m = 2r+1$ there are finite sets $X_1, \dots, X_m$ of equal size on the real line, such that their union is in general position and for all balanced separated μ-partitions, the number of columns is at least $r^2 + r + 1$*

In the previous proof we used the term "**magic matrix**". The role of the magic matrix will be taken by a "**nice matrix**". A d matrix is nice by definition, if the sum of each row is equal.

The role of permutation matrices will be taken by "**special matrices**" which are $0 - 1$ matrices with exactly one nonzero entry in each row.

Proof of the lower bound. The general idea of the proof is to fix a structure of the sets, and derive a nonnegative $r+1 \times r+1$ nice matrix for them, i.e. the sum of the entries is the same in each row. In this fixed structure of the sets the columns of the μ-partitions will correspond to an integer multiple of an $r+1 \times r+1$ special matrix, and from the definition we will get that the nice matrix can be written as an integer linear combination of k special matrices, where $k = n_{m,1}$. The next step will be to show that if every nice integer $r+1 \times r+1$-matrix can be written as positive integer linear combination of at most k special matrices, then all nice real $r+1 \times r+1$ matrix can be written as a positive linear combination of at most k special matrices. In the last step we only have to show that, if each nice $r+1 \times r+1$-matrix can be written as a positive linear combination of at most k special matrices, then k cannot be less then $r^2 + r + 1$, the dimension of the space of nice $r+1 \times r+1$-matrices.

In general for $1 \le t \le r$ we choose the sets X_{r+1+t} so that all of its points are in a small ε neighbourhood of t ($[t-\varepsilon, t+\varepsilon]$). The points of the other sets, X_j where $1 \le j \le r+1$ will be at most ε distance from a point $z + \frac{1}{2}$ where z is a nonnegative integer number at most r ($z = 0, \dots, r$, and ε can be chosen as $\frac{1}{6}$) i.e. $X_j \subset \cup_{z=0}^{r}[z + \frac{1}{2} - \varepsilon, z + \frac{1}{2} + \varepsilon]$. For all $1 \le i, j \le r+1$ let

$$A_{i,j} = X_i \cap [j-1, j]$$

and

$$d_{i,j} = |A_{i,j}|.$$

The matrix $d = (d_{ij})$ is the nice integer matrix corresponding to the sets.

Suppose now that M is a μ-partition with k columns of the sets $X_1, \dots, X_m$, and take one of its columns for example $M_{1,1}, \dots, M_{1,m}$. The sets $M_{1,r+1+t}$ where $1 \leq t \leq r$, are in an ε neighborhood of t, $M_{1,r+1+t} \subset X_{r+1+t} \subset [t-\varepsilon, t+\varepsilon]$, and the sets $M_{1,j} \subset X_j \subset \cup_{z=0}^{r}[z+\frac{1}{2}-\varepsilon, z+\frac{1}{2}+\varepsilon]$ cannot have elements in the neighborhood of 2 different points of the form $z+\frac{1}{2}, z \in \mathbb{N}$. This is clear, since otherwise the convex hull of $M_{1,j}$ would contain an integer t' between 1 and r, $t' \in \mathrm{conv} M_{1,j}$, and therefore it contains also its neighborhood $[t'-\frac{1}{2}+\varepsilon, t'+\frac{1}{2}-\varepsilon] \supset [t'-\varepsilon, t'+\varepsilon]$, which contains $M_{1,r+1+t'}$. This contradicts that the sets $M_{1,1}, \dots, M_{1,m}$ are separated.

Denote by i_j the integer for which $M_{1,j} \subset [i_j+\frac{1}{2}-\varepsilon, i_j+\frac{1}{2}+\varepsilon]$, and $1 \leq i_j \leq r+1$.

Now we know that $M_{1,j} \subset A_{j,i_j}$ for $1 \leq j \leq r+1$. Let S_1 be the matrix where the (j,i)-th entry is the cardinality of the set $M_{1,j} \cap [i-1, i]$. As showed above the matrix S_1 is an integer multiple of a special matrix, the one corresponding to the first column. But so are all matrices S_l corresponding to the l-th column, where $1 \leq l \leq k$ and therefore the matrix $d = (d_{i,j})_{r+1 \times r+1}$, is a positive linear combination of at most k special matrices.

Given a $r+1 \times r+1$ nice integer matrix d, we can easily choose m sets so that the nice matrix derived from them is d, and therefore all nice integer matrices can be written as a positive integer combination of at most k special matrices.

Let $D = (D(i,j))$ be a nice positive real matrix and define

$$n_0 = \frac{2r}{\min_{2 \leq i \leq r+1} D(i, r+1)}$$

Let us define a sequence of nice integer matrices as follows:

$$D_N = \lfloor N \cdot D \rfloor$$

where $N \in \mathbb{N}$ and $N \geq n_0$, such that for all $1 \leq j \leq r$ and $i = 1, j = r+1$

$$D_N(i,j) = \lfloor N D(i,j) \rfloor$$

All other entries ($j = r+1, 2 \leq i \leq r+1$) of the matrix D_N are determined since D_N should be a nice matrix. According to rounding

$$|D_N(i, r+1) - \lfloor N D(i, r+1) \rfloor| \leq r+1$$

and so the matrix D_N will be a nice positive integer matrix for every $N \geq n_0$. Further as $N \to \infty$

$$\frac{D_N}{N} \to D$$

and all matrices D_N $(N \geq n_0)$ can be written as a positive linear combination of at most k special matrices.

$$D_N = \sum_i^k a_i^N S_i^N \quad \text{where} \quad a_i^N \in N$$

Since there are only finitely many special matrices (exactly $(r+1)^{r+1}$) we can choose a subsequence of D_N, namely D_{N_j} where $S_i^{N_j} = S_i^{N_{j'}}$ for all $i, j, j' \in N$. Denote the common special matrix by $S_i = S_i^{N_1}$. Further the sequence

$$\frac{a_i^{N_1}}{N_1}, \frac{a_i^{N_2}}{N_2}, \ldots$$

is bounded, and therefore we can even choose a subsequence, that is convergent for all $1 \leq i \leq k$, and has as limit β_i:

$$\lim_j \frac{a_i^{N_j}}{N_j} = \beta_i$$

where β_i is of course nonnegative.

A simple limit argument shows that

$$D = \sum_i^k \beta_i S_i$$

so all positive real nice matrices D can be written as a nonnegative linear combination of at most k special matrices. As there are finitely many k-tuples of special matrices, one of them will cover an open set of the space of positive real nice matrices, and therefore the linear combinations over these k matrices generate the space of the nice matrices, which has dimension $r^2 + r + 1$. This proves our lemma that

$$n_{m,1} \geq k \geq r^2 + r + 1$$

3.7 Equivalence of three definitions about separated sets

Lemma 3.3. *If the sets $X_1, \dots, X_m$ are separated in a "geometric", "algebraic" or "combinatorial" sense, then these sets are separated in all three senses.*

Proof. We will prove that separated in "algebraic" sense implies separated in "geometric", which implies separated in "combinatorial" sense and which implies separated in "algebraic" sense.

"algebraic" $\to$ "geometric"
Let $x_i \in X_i$ and $x_m^{'} \in X_m$. We know that the orientation of $x_1, \dots, x_{m-1}, x_m$ and $x_1, \dots, x_{m-1}, x_m^{'}$ is the same, but then so are the orientations $x_1, \dots,$ x_{m-1}, y_m, where $y_m \in [x_m, x_m^{'}]$. This means that the sets $X_1, \dots, X_d$, convX_{d+1} are also separated in the "algebraic" sense, and therefore the sets conv$X_1, \dots,$ convX_m are separated too in the "algebraic" sense. But now a hyperplane intersecting at least $d+1$ of these convex hulls, contains $d+1$ points having orientation 0 which is a contradiction.

"geometric" $\to$ "combinatorial"
Choose $x_i \in X_i \quad 1 \le i \le m$ and an arbitrary hyperplane H, missing the points x_i. Let $I \subset [m]$ the indices of the points x_i on a given side of H. We need to show that

$$(\text{conv} \bigcup_{i \in I} X_i) \cap (\text{conv} \bigcup_{i \in I^c} X_i) = \emptyset$$

since then a hyperplane $H^{'}$ exists, which has all points of the set X_i on one side for all i, and divide them in the same way as H divide the points x_i. Suppose that

$$(\text{conv} \bigcup_{i \in I} X_i) \cap (\text{conv} \bigcup_{i \in I^c} X_i) \neq \emptyset$$

Due to Radon theorem, there are $d+2$ points $y_1, \dots, y_{d+2}$, we may suppose $y_i \in X_i$, such that

$$\text{conv} \bigcup_{i \in I} y_i \cap \text{conv} \bigcup_{i \in I^c} y_i \neq \emptyset$$

We call this a Radon-partition for the given points. If $d+2$ points are in general position, then the Radon-partition is well defined. Moving y_i continuously toward x_i, the Radon-partition has to be unique and therefore unchanged. But at the end we must have a different one as at the beginning.

This shows that our assumption that the intersection is not empty was wrong, so $H^{'}$ exists.

"combinatorial" → "geometric" Since $d+1$ points can be separated in any bipartition, so can all $d+1$ sets be separated. Therefore a hyperplane H can't meet $d+1$ sets conv$X_1, \dots,$ convX_{d+1} since there would be a Radon-partition according to this hyperplane for which no hyperplane would separate the sets into the parts of the Radon-partition.

□

References

[ABCC] D. Applegate, R. Bixby, V. Chvátal, W. Cook: *On the solution of traveling salesman problems,* in: "International Congress of Mathematics" (Berlin 1998), Documenta Math., Extra Volume ICM 1998, Vol. III, 645–656.

[BP] I. Bárány, M. Perles: *The Carathéodory number for the k-core,* Combinatorica, **10** (1990), 185–194.

[BP01] I. Bárány, A. Pór, *On 0-1 polytopes with many facets, Advances in Mathematics,* **161** No.2 (2001), 209–228.

[BáF] I. Bárány, Z. Füredi: *Computing the volume is difficult,* Discrete Comp. Geom., **2** (1987), 319–326.

[BF] T. Bonnesen, W. Fenchel: *Theorie der konvexen Körper,* Berlin, Springer 1934.

[BMT] C. Buchta, J. Müller, R. F. Tichy: *Stochastical approximation of convex bodies,* Math. Ann. **271** (1985), 225–235.

[BV98] I. Bárány, P. Valtr, *A Positive Fraction Erdős- Szekeres Theorem, Discrete and Computational Geometry,* **19** (1998), 335–342.

[CP] B. Carl, A. Pajor: *Gelfand numbers of operators with values in a Hilbert space,* Invent. Math., **94** (1988), 479–504.

[DL] M. M. Deza, M. Laurent: *Geometry of Cuts and Metrics,* Algorithms and Combinatorics **15**, Springer-Verlag, Berlin Heidelberg 1997.

[D] A. Dvoretzky: *Some near-sphericity results,* Proc. Sympos. Pure Math. Vol VII. 203–210 AMS., Providence, R.I., 1963.

[DFM] M. E. Dyer, Z. Füredi, C. McDiarmid: *Volumes spanned by random points in the hypercube,* Random Structures and Algorithms, **3** (1992), 91–106.

[E] J. Eckhoff: *Helly, Radon, and Carathéodory type theorems,* in: Handbook of Convex Geometry, (d. P.M. Gruber, J. Wills), North Holland, (1993), 389–448.

[EP95] P. Erdős, G. Purdy, *Extremal problems in combinatorial geometry* in: "Handbook of Combinatorics, Chapter 17 (R. Graham, M. Gr otschel and L. Lovász eds.)." Elsevier, New York, pp. 809–874, 1995.

[ES35] P. Erdős, Gy. Szekeres, *A combinatorial problem in geometry, Compositio Math.,* **2** (1935), 463–470.

[ES61] P. Erdős, Gy. Szekeres, *On some extremum problems in elementary geometry, Ann. Univ. Sci. Budapest* **3/4** (1960/61), 53–62.

[ETV96] P. Erdős, Zs. Tuza, and P. Valtr, Ramsey-remainder, *European Journal of Combinatorics* 17 (1996), 519-532.

[Fe] W. Feller: *An Introduction to Probability Theory and Its Applications,* Vol. II, Wiley, New York, 1971.

[FKR] T. Fleiner, V. Kaibel, G. Rote: *Upper bounds on the maximal number of facets of 0/1-polytopes,* European J. Combinatorics, **21** (2000), 121–130.

[F] K. Fukuda: Lecture in Oberwolfach, 1995.

[GP] M. Grötschel, M. Padberg: *Polyhedral Theory/Polyhedral Computations,* in: E.L. Lawler, J.K. Lenstra, A.H.G. Rinnoy Kan, D.B. Schmoys (eds.), "The traveling salesman problem", Wiley 1988, 251–360.

[GP91] J. E. Goodman, R. Pollack, *The complexity of point configurations, Discrete Appl. Math.* **31** (1991), 167–180.

[GP93] J. E. Goodman, R. Pollack, *Allowable sequences and order types in discrete and computational geometry,* in: "New Trends in Discrete and Computational Geometry (J. Pach, ed.)." Springer-Verlag, New York, 1993.

[K] G. Kalai, *Private communication* 1997.

[KA02] Gy. Károlyi, *Ramsey-remainder for convex sets and the Erdős-Szekeres theorem, Discrete Appl. Math.,* **1-2** (2001), 163–175

[KRSZ] U. Kortenkamp, J. Richter-Gebert, A. Sarangarajan, G. M. Ziegler: *Extremal properties of 0/1-polytopes, Discrete Comput. Geometry,* **17** (1997), 439–448.

[MM64] M. Marcus, H. Minc, *Convexity and Matrices* in: "A survey of matrix theory and matrix inequalities" Ally and Bacon, Boston, pp. 121–138, 1964.

[N95] M. J. Nielsen, *Transverse matchings on a finite planar set* (manuscript) University of Idaho, Moscow 1995.

[PS98] J. Pach, J. Solymosi, *Canonical theorems for convex sets, Discrete and Computational Geometry,* **19** (1998), 427–435.

[P02] A. Pór, The partitioned version of the Erdős–Szekeres theorem, submitted.

[PV01] A. Pór, P. Valtr, *Partitioned version of the Erdős-Szekeres theorem, Discrete and Computational Geometry,* submitted

[PV02] A. Pór, P. Valtr, *Partitioned Erdős-Szekeres theorem for convex sets, Discrete and Computational Geometry,* in preparation.

[S93] P. Schmitt, *Problems in discrete and combinatorial geometry,* in: "Handbook of Convex Geometry, Chapter 2.2 (P. M. Gruber and J. M. Willis eds.)." Elsevier, New York, pp. 449–483, 1993.

[SO88] J. Solymosi, *Combinatorial problems in finite Ramsey theory,* Master's thesis, Eötvös Univ. Budapest, 1988.

[TV98] G. Tóth, P. Valtr *Note on the Erdős-Szekeres theorem, Discrete and Computational Geometry,* **19** (1998), 457–459.

[Z] G. M. Ziegler *Lectures on 0/1 polytopes,* in: "Polytopes — Combinatorics and Computation" (G. Kalai and G. M. Ziegler, eds.), DMV-Seminars, Birkhäuser-Verlag Basel, 2000, 1–44, in print.