UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2019:9

The Happy Ending Problem and its connection to Ramsey theory

Sara Freyland

Examensarbete i matematik, 15 hp
Handledare: Anders Öberg
Examinator: Martin Herschend
Mars 2019

Department of Mathematics
Uppsala University

Contents

1 Introduction **2**
1.1 Notation 2

2 The Happy Ending Problem **2**
2.1 Background 3

3 Ramsey Theory **4**
3.1 Ramsey's theorem 5
3.2 Schur's and van der Waerden's theorems 7

4 The Happy Ending Problem in terms of Ramsey theory **8**

5 Erdős and Szekeres **9**
5.1 Proof 1 9
5.2 Proof 2 11
5.3 The upper bound of $N(n)$ 14

6 Later improvements **15**

1 Introduction

This thesis aims to give a brief history of Ramsey theory while also giving a deeper understanding of the Happy Ending Problem and its place in that history. The name of the problem was suggested by mathematician Paul Erdős when two of his colleges married and he thought that the problem was what brought them together, [8].

Erdős was together with George Szekeres and Ester Klein the first to study the problem which is a geometrical problem about how many points that are sufficient to construct a convex polygon of a particular size, [3]. A detailed review of the first article written about the problem in 1935 ([3]) will be given, but before that one should have a better understanding of the field of Ramsey. This due to the fact that even though it is a geometrical problem, it has strong ties to combinatorics and the history of Ramsey theory.

1.1 Notation

Let $n, k \in \mathbb{N}$
$[n]$ - Arbitrary set of cardinality n, n-combination
$[n]^k = \{X \subseteq [n] \mid |X| = k\}$
General position - No three on a line

2 The Happy Ending Problem

The Happy Ending Problem was formulated by mathematician Ester Klein. It is a geometrical problem about different polygons in the plane, and how many points (in general position) are sufficient to construct a convex polygon of a particular size. In this case, a convex constellation of points in the plane, is defined so that following the edge of the figure counterclockwise, one would have to turn left at every corner. The opposite, concave, is defined as not convex. [3] For the duration of this thesis, all sets of points will be assumed to be in general position.

Klein first considered the special case about constructing quadrilaterals, but later generalized the problem to cover all polygons with four or more vertices. She suggested this generalized version to her coworkers Paul Erdős and George Szekeres, who would later publish the first article on the problem.

> *"Can we find for a given n a number $N(n)$ such that from any set containing at least N points it is possible to select n points forming a convex polygon?"* ([3], p. 1)

A quadrilateral is the smallest polygon for which this problem is non trivial (no concave triangles exist, see figure 2) and if the plane has sufficiently many points, it is quite simple to find four in convex position. But four points are obviously not sufficient, as shown in figure 1. The Happy Ending Problem is therefore to establish the least number of points that make it possible to construct a convex n-combination.

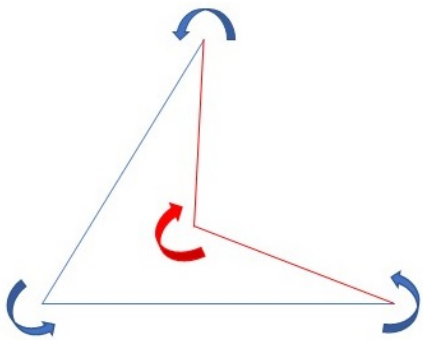

Figure 1: A concave quadrilateral.

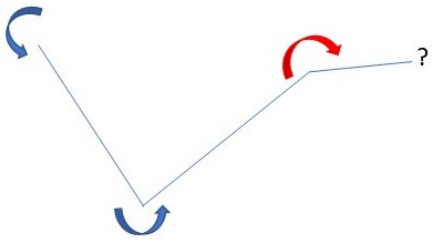

Figure 2: No concave triangles exist.

2.1 Background

Although Klein was the first to suggest the problem, Erdős and Szekeres were the first to publish an article about the Happy Ending Problem. This was in their 1935 paper *A combinatorial problem in geometry* ([3]). They there stated Klein's general version of the problem as follows:

> *"Can we find for a given n a number $N(n)$ such that from any set containing at least N points it is possible to select n points forming a convex polygon?"* [3] p.1

In the article Erdős and Szekeres gave credits to Klein for suggesting the general problem, and supplied a quick demonstration of her proof for $n = 4$.

Theorem 1 (Erdős and Szekeres. 1935. [3])**.** *N(4)=5*

Proof. Assume there are five points in general position in the plane. If the smallest convex polygon encircling the rest of the points is a 4-combination or a 5-combination, then all is as desired. If it is instead a 3-combination (ABC), with two points (D and E) in its interior, then study the line $\overline{DE}$. Due to the general position, $\overline{DE}$ will then intersect two of the sides of ABC, see figure 3. Thus $\overline{DE}$ divide ABC into a smaller 3-combination (BDE) and the desired 4-combination ($ACED$). □

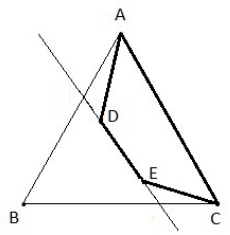

Figure 3: Visualization of the proof of theorem 1.

To prove the general version of the problem they divided it into two questions: (1) does $N(n)$ exist for all n and if so, (2) how is the least value of $N(n)$ expressed or determined as a function of n. For $n \geq 5$, Erdős and Szekeres were unable to determine the exact value of $N(n)$ and tried instead to find an upper bound.

Erdős and Szekeres were able to answer the first question by using theorem 1 as the base case in an induction argument and, eventually, found an upper bound to $N(n)$, $n \geq 5$. Their proof and reasoning for this will be explained later in this thesis.

As they were only able to find an upper bound and not the exact value of $N(n)$, the second question was unanswered in their article. Question (2) is in fact still unanswered today ([9]) and the focus on the Happy Ending Problem has therefore been more on reducing the upper bound of $N(n)$ than on finding the exact value. Erdős and Szekeres, however, made the conjecture that the exact value should be $N(n) = 2^{n-2} + 1$, [3].

3 Ramsey Theory

To prove their earlier stated questions Erdős and Szekeres used Frank Ramsey's 1930 paper *On a Problem of Formal Logic* ([3]), where he introduced and proved a theorem which would later be known as Ramsey's theorem. Ramsey was a British mathematician mainly interested in mathematical economics and mathematical logic in the late 1920s, [7]. Although he died young, at only 27 in 1930, he managed to contribute both to the field of economics and logic [7].

In this particular paper he focused on logic, but Ramsey did see that his theorem could be of "independent interest" ([13]). However it was not until 1935 after Erdős and Szekeres paper *A combinatorial problem in geometry* that Ramsey's theorem was seen to be of value to combinatorical analysis [7]. A later contribution of Ramsey's theorem was Ramsey numbers which are a big part of Ramsey theory today, [13]. These will be discussed more later on.

When Erdős and Szekeres wrote their paper Ramsey theory was obviously not an established field. It would be much later that Ramsey's contributions to combinatorial analysis would gain recognition to the point where an entire field within it would be named after him. According to Graham *et al.*, combinatorial analysis was for a long time considered "the slums of topology" ([7], p. xi). It was not until the 1980s that Ramsey theory emerged and have since then been formed into what it is today, [7].

What today is called Ramsey theory does however have a pre-history before Ramsey's

theorem. Both I. Schur in 1916 and B. L. van der Waerden in 1927 contributed with fundamental theorems on which the field has later been built on, [7].

3.1 Ramsey's theorem

In his 1930 article, Ramsey introduced and proved both an infinite and a finite version of what would later be called Ramsey's theorem, [13]. He proved them separately, but today the finite version follows relatively directly from the infinite version, [7]. His interest in them was to help solve what he called "one of the leading problems of mathematical logic" ([13] p. 264), namely to be able to determine whether a logical formula is true or false [13].

To be able to use the combinatorial theorems later in his paper he divided the paper into four parts. The first stated and proved both versions of Ramsey's theorem and then the paper went on to the logical argument in the second to forth part, [13]. Ramsey's entire paper contributed to the field of logic, but it would be those first two theorems that would become one of the cornerstones of Ramsey Theory, [7].

Ramsey used different notation in 1930 than what is used in Ramsey theory today. The finite version of Ramsey's theorem stated in Ramsey's paper ([13]) is still comprehensible.

> *"Given any r, n, and μ we can find an m_0 such that, if $m \geq m_0$ and the r-combinations of any Γ_m are divided in any manner into μ mutually exclusive classes C_i $(i = l, 2, ..., \mu)$, then Γ_m must contain a sub-class Δ_n such that all the r-combinations of members of Δ_n belong to the same C_i".* ([13], p. 267)

A more modern version of the finite Ramsey theorem can be stated as below.

> *"$\forall r, n, k, n + k \geq r, \exists m_0$ so that, for $m \geq m_0$, if $[m]^r$ is 2-colored there exist $S, T \subseteq [m], |S| = n, |T| = k, S \cap T = \emptyset$, so that all r-subsets of $S \cup T$ containing at least one $x \in S$ are the same colour."* ([7], p. 21)

In both versions of the theorem, one can see the importance of what is denoted as m_0. This is the divisor such that the statements that follow hold true for all numbers higher than m_0 but not for those that are lower. The value of m_0 of course differ depending on the starting criteria (in the modern version of the theorem: r, n, k). What Ramsey's theorem proves is the existence of such divisors for all different starting criteria. These divisors have formally been named Ramsey numbers, [7]. They show how big a structure needs to be, before a more ordered substructure will emerge, [7].

Example 1. *([7])* Imagine 6 people who either mutually know or mutually do not know each other. When choosing one person, here called A, this person will either know 3 or more of the rest or do not know 3 or more of the rest of the group.

If A do not know 3 people from the group, then either 2 of these 3 do not know each other (together with A this is 3 who do not know each other) or all 3 of them mutually know each other, see figure 4.

If A know 3 people from the group, then either 2 of them mutually know each other (together with A this is 3 who know each other) or all 3 of them do not know each other.

Instead imagine 5 people who either mutually know or mutually do not know each other. Then there exist an example when neither 3 mutually know each other nor 3 mutually do not know each other, see figure 5.

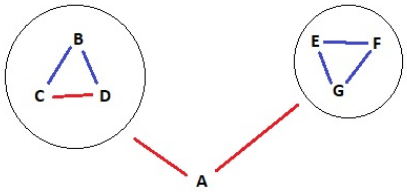

Figure 4: **A** either do not know (red) 3 people whom all know each other (blue) or 3 people out of which at least 2 do not know each other.

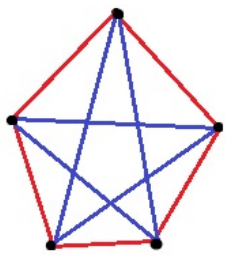

Figure 5: 5 people who either know (blue) or do not know (red) each other.

From example 1 one can see that in a group of 6 people one can always find 3 who mutually either do or do not know each other, but from 5 people that might be impossible. Since it is true for 6 people it will also be true for all larger groups, [7]. This is a good example of a divide where something is true for all numbers bigger or equal to 6 but not for numbers smaller than 6. This is therefore the Ramsey number for finding either a subgroup of size three colored red or a subgroup of size 3 colored blue, [7].

To instead give a formal definition of Ramsey's theorem one must first define what is meant by coloring and introduce a notation for Ramsey numbers. This notation will be used for the duration of this thesis.

Definition 1 ([7])**.** *Given a set S and $r \in \mathbb{N}$. An r-coloring of S is then a map $\chi : S \to [r]$. For $s \in S$, $\chi(s)$ is referred to as the color of s and if χ is constant on $T \subseteq S$, then T is called monochromatic under χ.*

Definition 2 ([7])**.** *Let $n, i, l_i, k \in \mathbb{N}$. If for every r-coloring of $[n]^k$, there exists i $(1 \leq i \leq r)$ and $T \subseteq S$ $(|T| = l_i)$ such that $[T]^k$ is monochromatically colored i. Then*

$$R_k(l_1, ..., l_r) \leq n$$

$R_k(l_1, ..., l_r)$ (Ramsey number for $(l_1, ..., l_r)$) denotes the smallest such n. If $l_1 = ... = l_r$, then $R_k(l_1, ..., l_r) = R_k(l)^r$.

Example 1 showed first that $R_2(3,3) \leq 6$, and then that $R_2(3,3) \nleq 5$. Since $R_2(3,3)$ is the smallest number from which one can find a monochromatic substructure of size three, one has proven that $R_2(3,3) = 6$.

Theorem 2 (*Ramsey's theorem*, [7])**.** *$R_k(l_1, ..., l_r)$ is well defined.*

3.2 Schur's and van der Waerden's theorems

The history of Ramsey theory can be traced back to before Ramsey's theorem. What might be considered as its first theorem was proved by Schur's in 1916, [7]. Schur was born in Russia in 1875 but had his entire academic career in Germany where he was appointed professor in 1919 and eventually became a member of the Prussian Academy of Science, [12].

In his 1916 paper *Über die kongruenz $x^m + y^m \equiv z^m$ (mod p)* ([14]), Schur was interested in Fermat's last theorem and was able to prove it wrong for finite fields of sufficiently large prime characteristic, [14]. This result would later be built on to what today is referred to as Schur's theorem.

Theorem 3 (Schur's theorem. [7]. p. 69)**.** *If $N \subset \mathbb{N}$ is finitely colored, then there exist $x, y, z \in N$, having the same color such that $x + y = z$.*

Another important result from before Ramsey's theorem was van der Waerden's theorem, which was proven in 1927. Van der Waerden seems to have first encountered the theorem as a conjecture made by either Schur or Dutch mathemathician Baudet. [5] The teorem is connected to arithmetic progressions within the natural numbers. Arithmetic progressions are series of numbers where the difference between consecutive terms is constant, [17]. The theorem which van der Waerden proved showed that if the natural numbers are partitioned into classes (finitely coloured), then at least one of them will always contain an arithmetic progression of arbitrary length, [5].

Theorem 4 (van der Waerden's theorem. [7]. p. 29)**.** *If the positive integers are partitioned into two classes, then at least one of the classes must contain arbitrarily long arithmetic progressions.*

Later on van der Waerden's theorem has been modified to work for not only two colours but arbitrary many. Mathematicians have also been trying to determine so called van der Waerden-numbers, $W(r, k)$. They are the lowest natural number such that an r-coloring of $[W(r, k)]$ assure an monochromatic, arithmetic progression of length k. [5]

Van der Waerden's theorem states that an arbitrary long progression must exist in one of the classes, but not in which one. It was first in 1974 that Szemerédi were able to generalize van der Waerdens theorem and proved that the progressions are connected to the upper density of the class, [5].

The density of a subset of the natural numbers is a way to measure its size compared to the entire set of natural numbers. All natural numbers and all even natural numbers are both sets that are infinitely countable, [17]. But intuitively the first feels bigger than the second. The density of all even natural numbers as a subset of all natural numbers is therefore a way of to express this intuitive feeling. The density is the limit of

the probability of picking an even number from a subset of all natural numbers, as the subsets cardinality goes toward infinity, [2].

Definition 3 ([2]). *The natural density, $d(A)$, of a set $A \subset \mathbb{N}$ is defined as:*

$d(A) := \lim_{N\to\infty} \frac{|A\cap[1,N]|}{N},\ A \subset \mathbb{N}, N \in \mathbb{N}$

The upper density of a set $A \subset \mathbb{N}$ is defined as:

$\overline{d(A)} := \limsup_{N\to\infty} \frac{|A\cap[1,N]|}{N},\ A \subset \mathbb{N}, N \in \mathbb{N}$

The lower density is defined in a similar manner.

The name density is quite fitting as it measures how dense the subset is within the natural numbers. For example is the density of all even natural numbers $0,5$ while it is 0 for the set of primes, [2].

If all the natural numbers are partitioned into two classes, as is the case in van der Waerden's theorem, then Szemerédi proved that the classes only contain an arithmetic progression if they have a positive upper density, [5]. I.e. the distance between the numbers does not become arbitrary large as is the case with the primes.

Theorem 5 (Szemerédi's theorem. [5]. p. 27). *If S is a set of positive integers with positive upper density,*

$$\limsup_{n\to\infty} \frac{|S \cap [n]|}{n} > 0 \tag{1}$$

then S contains arbitrarily long arithmetic progressions.

4 The Happy Ending Problem in terms of Ramsey theory

As mentioned in section 3, Ramsey saw that his theorem could be of interest on its own, but it was not until Erdős and Szekeres sought to prove the Happy Ending Problem that its usefulness and connection to combinatorics was shown, [7]. Erdős and Szekeres wanted to prove that for all $n \in \mathbb{N}$, $N(n)$ exist, [3]. To do so they used that $N(4) = 5$ (theorem 1) to show that:

$$N(n) = R_4(5, n). \text{ [3]} \tag{2}$$

If one colors all 4-combinations of a set of points, after whether they are convex or concave. Then $R_4(5, n)$ is the smallest number, such that there either exists (a) a 5-combination with all 4-combinations of those five elements in concave position or (b) a convex n-combination with all 4-combinations of those elements in a convex position. (a) is impossible due to the fact that $N(4) = 5$. One can always find a convex 4-combination from 5 points. Therefore (a) can never be fulfilled since it requires all 4-combinations of the five points to be concave. Since (a) can not be true (b) must be, and $R_4(5, n)$ is thus the smallest number of points from which one can find a convex n-combination. [3]

This way Erdős and Szekeres were able to reduce the first of their questions, does N exist for all n, to whether or not Ramsey's theorem is well defined. Since Ramsey had already done that, the first question was answered. The version of Ramsey's theorem stated in Erdős and Szekeres' paper and the proof of it however, differed from the original version, [3]. This change was due to fact that Erdős and Szekeres was trying to find an answer to their second question. They were looking for a better and easier way to estimate the upper bound of $N(n)$, than that which could be obtained from Ramsey's original proof, [7]. This, to try to get as close as possible to their conjectured value.

Although Erdős and Szekeres were able to get an upper bound with their proof of Ramsey's theorem. It was still far from the conjectured value. They therefore gave a second more geometrical proof to reduce the upper bound of $N(n)$.

> "Mr. E. Makai proved that $N(5) = 9$, and from our second demonstration, we obtain $N(5) = 21$ (from the first a number of order 2^{10000})."
> ([3], p. 1)

By presenting these two proofs, Erdős and Szekeres were able to answer the first of their questions: N does exist for all n. For the second question the upper bounds found were expressed as functions of n, but as mentioned, they were far from their conjectured value.

5 Erdős and Szekeres

The two proofs in Erdős and Szekeres article both give an upper bound to $N(n)$, but have different approaches to the theorem. The first proof is more combinatorial with a strong connection to Ramsey theory as it uses Ramsey's theorem. While the second proof is a more geometrical proof that uses a different definition of convex.

5.1 Proof 1

Theorem 6 (*Erdős and Szekeres. 1935. [3]*)**.** *Let $k, l, i \in \mathbb{N}$, $k \geq i$ and $l \geq i$. Suppose that there exist two classes, α and β, of i-combinations of N points such that each k-combination shall contain at least one combination from class α and each l-combination shall contain at least one combination from class β. Then for sufficiently large $N \geq R_i(k, l)$ this is impossible.*

Proof. ([3]) To prove the existence of such an N one uses induction over i, k and l on the N points in general position together with a proof of contradiction.

Base Cases:

1. $i = 1$

   Assume that the theorem is false. Let the i-combinations belonging to α be such that each k-combination contains at least one of them. There are then at most $k - 1$ i-combinations belonging to β. If $l \leq |\alpha| = (N - (k - 1))$, then there exist a

l-combination of only elements belonging to α. So for the theorem to hold:

$$l > (N - (k-1)) \Leftrightarrow N < k + l - 2 \tag{3}$$

This is false for sufficiently large N.

2. $k = i$ or $l = i$

   For $k = i$, then $N = l$. If there exist one i-combination belonging to β then there exist a k-combination only containing i-combinations from β. Otherwise all i-combinations belong to α and therefore there exist a l-combination (all elements) only containing i-combinations from α. The reasoning for $l = i$ is symmetrical.

Induction hypothesis: $R_{i-1}(k,l)$, $R_i(k-1,l)$ and $R_i(k,l-1)$ are well defined.

Claim: $R_i(k,l) \leq R_{i-1}(R_i(k-1,l), R_i(k,l-1)) + 1$

Set $s := R_i(k-1,l)$, $t := R_i(k,l-1)$ and $N = R_{i-1}(s,t) + 1$. Assuming one have N points, let $S = \{s'_1, ..., s'_s\}$ and $T = \{t'_1, ..., t'_t\}$ be an arbitrary s-combination respectively t-combination of $N-1$ of the points. Denote the remaining N^{th} point n.

Assuming that $N < R_i(k,l)$ the following can be derived from S and T.

Every l-combination in S contain a i-combination belonging to β. The definition of s then implies that there exist a $(k-1)$-combination, $S_{k-1} = \{s'_{a_1}, ..., s'_{a_{k-1}}\} \subset S$, only containing i-combinations belonging to β. Similarly there then exist a k-combination, $S_{k-1} \cup \ n$, containing at least one i-combination belonging to α due to the fact that $|S|+1 \leq N < R_i(k,l)$. This i-combination contain n since S_{k-1} only have i-combinations belonging to β.

$$n \ \cup \{s'_{b_1}, ..., s'_{b_{i-1}}\} \subset S_{k-1} \cup n \tag{4}$$

Denote all such possible $(i-1)$-combinations $(\{s'_{b_1}, ..., s'_{b_{i-1}}\})$ α'.

Similarly for T, every k-combination in T contain an i-combination belonging to α. The definition of t then implies that there exist a $(l-1)$-combination, $T_{l-1} = \{t'_{a_1}, ..., t'_{a_{l-1}}\} \subset T$, only containing i-combinations belonging to α. Similarly there then exist a l-combination, $T_{l-1} \cup \ n$, containing at least one i-combination belonging to β due to the fact that $|T|+1 \leq N < R_i(k,l)$. This i-combination contain n since T_{l-1} only have i-combinations belonging to α.

$$n \cup \{t'_{b_1}, ..., t'_{b_{i-1}}\} \subset T_{l-1} \cup n \tag{5}$$

Denote all such possible $(i-1)$-combinations $(\{t'_{b_1}, ..., t'_{b_{i-1}}\})$ β'.

All $(i-1)$-combinations of N can therefore be divided into two classes, α' and β'. Every s-combination then contain at least one combination from class α' and each t-combination one from class β'. This contradicts $N = R_{i-1}(s,t) + 1$, and the assumption that $N < R_i(k,l)$ must therefore be false.

$\square$

5.2 Proof 2

As mentioned earlier, Erdős and Szekeres changed their definition of convex for the second proof, [3]. From the one used in proof 1 to a definition much closer to what is normally used in geometry. But to do so one must first define *cup/cap* which is done with the earlier definition of convex.

Definition 4 ([15])**.** *Let N be a set of points in general position in the plane. N forms a cup (cap) if the points are in convex position and its hull is bounded above (below) by a single edge. See figure 6.*

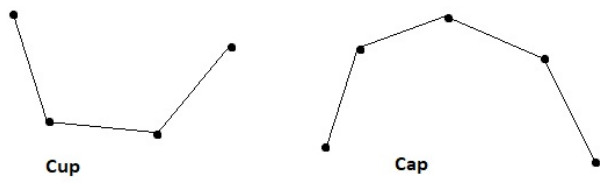

Figure 6: Two convex constellations with the first definitions of convex and one concave and one convex with the second definition.

Definition 5 ([3])**.** *A constellation of points $N = N_1N_2N_3...$ is said to be convex if the gradients of the lines $\overline{N_1N_2}, \overline{N_2N_3}, ...$ monotonously decrease, and concave if they monotonously increase.*

With definition 5 one can see that the constellations defined in definition 4 are no longer convex. Definition 5 defines a cup as concave and a cap as convex. Erdős and Szekeres also needed two lemmas before they could finish their proof.

Lemma 7. *(Erdős and Szekeres. 1935. [3])*

From $n^2 + 1$ distinct points in general position in the plane, it is always possible to select either $n + 1$ points with increasing x-coordinates and monotonously increasing y-coordinates or $n + 1$ points with increasing x-coordinates and monotonously decreasing y-coordinates.

Proof. ([3]) To prove this one uses induction over n.

Let $f(n, n)$ be the minimum number of points out of which one can select either n points with monotonously increasing y-coordinates or n points with monotonously decreasing y-coordinates. Two points with the same y-coordinate can be considered both increasing and decreasing.

Base case: $n = 2$. $f(2, 2) = 2$.

Induction hypothesis: $f(n, n) \leq (n - 1)^2 + 1$

Claim: $f(n + 1, n + 1) \leq n^2 + 1$.

Assume one have $(n^2 + 1)$ points. From the first (ordered by x-coordinates) $(n - 1)^2 + 1$ points, it is possible to choose either n monotonously increasing or n monotonously

decreasing points by the induction assumption. Remove the last point (greatest x-coordinate) in the monotonous sequence, in exchange for one of the other $2n-1$ points $((n^2+1)-((n-1)^2+1)=2n-1)$. Again one has $(n-1)^2+1$ points from which a monotonous sequence can be found. Following the same procedure one will get $2n$ points, that are all the endpoint of a monotonous sequence.

Now two cases emerge from those $2n$ points:

1. $(n+1)$ points are the endpoints of increasing sets or $(n+1)$ points are the endpoints of decreasing sets

   Assume that the first statement is true. Then those $(n+1)$ points are the endpoints of monotonously increasing sequences of length n. Denote two arbitrary sequences from those $(n+1)$ a and b. If the endpoint of a has x-coordinate and y-coordinate greater than those of all points in b, then the endpoint of a together with the n points in b create a monotonously increasing set of length $(n+1)$. If no such a and b exists, then the $(n+1)$ endpoints create a decreasing set.
   The proof is symetrical for the second statement.

2. n points are the endpoints of increasing sets and n points are the endpoints of decreasing sets

   By the same reasoning as in case 1. one get either an increasing (decreasing) sequence of length $(n+1)$, or all the endpoints create a decreasing (increasing) sequence. If the first is true then all is as desired, therefore assume that the second case is true for both sets of endpoints.

   This gives a decreasing sequence (here called d) of length n, where all the points are endpoints of increasing sequences of length n, and an increasing sequence (here called i) of length n, where all the points are endpoints of decreasing sequences of length n. Denote the points with the greatest x-coordinate in the two sequences e_i (endpoint of an increasing and a decreasing set of length n) and e_d (endpoint of an increasing and a decreasing set of length n) respectively. Now three new cases emerge:

   (a) e_i has a lower x-coordinate than e_d

   Since e_i is the endpoint of both a monotonously increasing and decreasing set of n points, adding e_d will give either a decreasing or an increasing set of $(n+1)$ points.

   (b) e_d has a lower x-coordinate than e_i

   Since e_d is the endpoint of both a monotonously increasing and decreasing set of n points, adding e_i will give either a decreasing or an increasing set of $(n+1)$ points.

   (c) $e_k = e_i$

   Impossible due to the fact that once a point was an endpoint, it was removed and could therefore not be the endpoint of two different monotonous sequences.

$\square$

Lemma 8 ([3])**.** *If $P_1, P_2, P_3, ..., P_n \in \mathbb{R}$ are points on a straight line and $g(i,k)$ denotes the smallest value of n such that one can always select either i points with monotonously increasing distance (measured from left to right) between neighboring points or k points with monotonously decreasing distance. Then $g(i,k) \leq g(i-1,k) + g(i,k-1) - 1$.*

Equal distances can be classified as either increasing or decreasing.

Proof. ([3]) To prove this one uses induction over i and k.

Base case: $g(3,n) = g(n,3) = n$

Consider the case $g(3,n) = n$. Denote the distance between the first two points a. If the distance between the second and third point is bigger than a, then one have 3 with increasing distance. Assume instead that this does not occur for any of the points. This means that the distance between the $(b-1)^{th}$ point and the bth point $(1 \leq b \leq n)$ have to be less than all the previous distances. Therefore, the only way to not have 3 points with monotonously increasing distance, is to have n points with monotonously decreasing distance.

If one instead have $(n-1)$ points one might have all of them with monotonously decreasing distance and thus $g(3,n) = n$. A symmetrical argument can be used for $g(n,3)$.

Induction hypothesis: $g(i-1,k)$ and $g(i,k-1)$ are well defined.

Claim: $g(i,k) \leq g(i-1,k) + g(i,k-1) - 1$

Let $n = g(i-1,k) + g(i,k-1) - 1$ and let P be the bisection of $\overline{P_1P_n}$. Then there are either at least $g(i-1,k)$ points on the first half of the line or at least $g(i,k-1)$ points on the second half.

If the first case is true, one either have k points with monotonously decreasing distance or $(i-1)$ with monotonously increasing distance. In the first case all is as desired and in the second one can add P_n to get i points with monotonously increasing distance.

If the second case is true, one instead have either i points with monotonously increasing distance or $(k-1)$ with monotonously decreasing distance. In the first case all is as desired and in the second one can add P_1 to get k points with monotonously decreasing distance. □

Theorem 9. *([3]) Let $h(i,k)$ be the smallest amount of points such that one can either pick a convex i-combination or a concave k-combination. Then*
$h(i,k) \leq h(i-1,k) + h(i,k-1) - 1$.

Theorem 9 uses the definition of convex found in definition 5. Lemma 7 and 8 are needed to prove the existence of such convex (cap) and concave (cup) constellations. Lemma 7 proves that one can find a constellation with increasing or decreasing y-coordinates and, considering the constellations gradients as distances between points on a line, lemma 8 proves that there exists points with either monotonously increasing or monotonously decreasing distance (gradients).

Proof. ([3]) To prove this one uses induction over i and k.

Base case: $h(3, n) = h(n, 3) = n$

Consider the case $h(n, 3) = n$. Denote the gradient of the line between the first two points a. If the gradient of the line between the second and third point is greater than a, then one have 3 points with increasing gradients (concave). Assume instead that this does not occur for any of the points. This means that the gradient of the line between the $(b-1)^{th}$ point and the bth point $(1 \leq b \leq n)$ have to be less than that of all the previous gradients. Therefore, the only way to not have 3 points with monotonously increasing gradients (concave), is to have n points with monotonously decreasing gradients (convex).

If one instead have $(n-1)$ points one might have all of them with monotonously decreasing gradients and thus $h(3, n) = n$. A symmetrical argument can be used for $h(3, n)$.

Induction hypothesis: $h(i-1, k)$ and $h(i, k-1)$ are well defined.

Claim: $h(i, k) \leq h(i-1, k) + h(i, k-1) - 1$

Let $N = h(i-1, k) + h(i, k-1) - 1$. From the N points, look at the first $h(i-1, k)$. Then there either exists a concave k-combination, and all is as desired, or a convex $(i-1)$-combination. If the later is true, switch the last point in the combination for one of the remaining $h(i, k-1) - 1$ points. Then the same two cases will emerge and either all is as desired or another of the $h(i, k-1) - 2$ points can be chosen. Assume that there never emerges a concave k-combination when this procedure is repeated. Then there will exist $h(i, k-1)$ points, that all are the last (highest x-coordinate) point in a convex $(i-1)$-combination.

Among these $h(i, k-1)$ points, there will either be a convex i-combination and all is as desired or a concave $(k-1)$-combination. If the last one is true, denote the points in the $(k-1)$-combination $P_1, P_2, ..., P_{k-1}$. P_1 is the last point in a convex $(i-1)$-combination. Denote its neighboring points in this $(i-1)$-combination P.

Studying the gradients of the lines $\overline{P_1P_2}$ and $\overline{PP_1}$, two cases emerge ones again. If the gradient of $\overline{P_1P_2}$ is greater than the gradient of $\overline{PP_1}$ then P together with $P_1P_2...P_{k-1}$ form a concave k-combination. If the opposite is true, then P_2 together with the convex $(i-1)$-combination from which P and P_1 was chosen, form a convex i-combination. □

5.3 The upper bound of $N(n)$

Erdős and Szekeres conjectured that the value was $N(n) = 2^{n-2} + 1$. Both of the proofs from their 1935 paper gave upper bounds to $N(n)$ with differing vicinity to the conjecture. Although the second proof gave a lower bound than the first, they were both far from the conjecture. From the proof of theorem 5, Erdős and Szekeres derived the following system of equations. [3]

$$\begin{cases} (i)\ R_i(k, l) \leq R_{i-1}(R_i(k-1, l), R_i(k, l-1)) + 1 \\ (ii)\ R_1(k, l) = k + l - 1 \\ (iii)\ R_i(i, l) = l,\ R_i(k, i) = k \end{cases} \tag{6}$$

They were thus able to make the deduction that $R_2(k+1, l+1) \leq \binom{k+l}{k}$, [3]. From this one can obtain the bound of $N(n) = R_4(5, n)$ for different values of n. As mentioned before, the upper bound of $N(5)$ is of order 2^{10000} with this first proof. This can be seen from the following calculations.

$$\begin{aligned}
R_4(5,5) &\overbrace{\leq}^{(6)(i)} R_3(R_4(4,5), R_4(5,4)) + 1 \overbrace{=}^{(6)(iii)} R_3(5,5) + 1 \\
&\overbrace{\leq}^{(6)(i)} R_2(R_3(4,5), R_3(5,4)) + 2 \overbrace{=}^{(*)} R_2(10627, 10627) + 2 \\
&\leq \binom{10626+10626}{10626} + 2 = \binom{21252}{10626}
\end{aligned}$$

$$\begin{aligned}
21252! &\approx 6,9 * 10^{82738} \\
10626! &\approx 6,4 * 10^{38171}
\end{aligned}$$

$$\begin{aligned}
(*)\ R_3(4,5) = R_3(5,4) &\overbrace{\leq}^{(6)(i)} R_2(R_3(4,4), R_3(5,3)) + 1 \overbrace{=}^{(6)(iii)} R_2(R_3(4,4), 5) + 1 \\
&\overbrace{\leq}^{(6)(i)} R_2(R_2(R_3(3,4), R_3(4,3)) + 1, 5) + 1 \overbrace{=}^{(6)(iii)} R_2(R_2(4,4) + 1, 5) + 1 \\
&\leq R_2(\binom{6}{3} + 1, 5) + 1 = R_2(21,5) + 1 \leq \binom{20+4}{20} + 1 = 10627
\end{aligned}$$

The second proof gave a better estimate of the upper bound with:

$$\begin{cases} h(k,l) = h(k-1,l) + h(k,l-1) - 1, \\ h(3,n) = h(n,3) = n. \end{cases} \tag{7}$$

From this and the fact that $N(n) = h(n,n)$ they obtained that $N(n) \leq \binom{2n-4}{n-2} + 1$, [3]. $N(n) = h(n,n)$ since the constellations defined as convex and concave in the second proof is both considered convex with the definition used in the first proof and the statement of the Happy Ending Problem.

n	$\binom{2n-4}{n-2}$+1	$2^{n-2}+1$
5	21	9*
6	71	17*
7	253	33
8	925	65

Figure 7: Table 1: Upper bounds of $N(n)$ and the conjectured value.
*Proved to be exact ([10], [16])

6 Later improvements

After Erdős and Szekeres first publication on the problem in 1935 there would be a long time before more publications about the problem came. As mentioned in section 2.1,

proving the exact value of $N(n)$ is still an open problem and the focus has therefore been on improving the upper bound. But this would have to wait until much later in 1998. The lower bound of $N(n)$, however, was proven by Erdős and Szekerez in 1960 ([4]).

$$2^{n-2} + 1 \leq N(n) \leq \binom{2n-4}{n-2} + 1 \text{ [4], [3]} \tag{8}$$

As this lower bound was equal to their conjecture, and the conjecture still stands today, it is believed that this value is correct, [7]. Their upper bound, however, was first improved in 1998 by Chung and Graham ([1]). Their improvement might be slight but it renewed the interest in the Happy Ending Problem. This resulted in two more articles being published during the same year, by Kleitman and Pachter ([11]), and Tóth and Valtr ([18]) respectively. All making small improvements on the upper bound of $N(n)$.

$$N(n) \leq \binom{2n-4}{n-2} \text{ [1]} \tag{9}$$

$$N(n) \leq \binom{2n-4}{n-2} - 2n + 7 \text{ [11]} \tag{10}$$

$$N(n) \leq \binom{2n-5}{n-2} + 2 \text{ [18]} \tag{11}$$

After 1998 their would again be a gap in publications on the Happy Ending Problem. Even if this gap only lasted until 2005. Tóth and Valtr were then able to make a slight improvement on their earlier result.

$$N(n) \leq \binom{2n-5}{n-2} + 1 \text{ [19]} \tag{12}$$

The upper bounds mentioned in equation (8) − (12) could all be approximated by $\frac{4^n}{\sqrt{n}}$, [15]. Although improvements were made on the bound found by Erdös and Szekeres in 1935 it was not enough to get it even to the same base as the conjectured value: $2^{n-2}+1$. It was not until 2017 that this would be achieved by mathematician Andrew Suk. In his article *On the Erdős–Szekeres convex polygon problem* ([15]) Suk was interested in how $N(n)$ behaves for sufficiently large n and were able to reduce the upper bound of $N(n)$ to:

$$N(n) \leq 2^{n+6n^{2/3}\log n} \text{ [15]} \tag{13}$$

Later the same year Holmsen, Mojarrad, Pach and Tardos made an asymptotic improvement of the bound in their article *Two extensions of the Erdős–Szekeres problem* ([9]).

Definition 6. *([6], p. 443) $O(n)$ is defined as a constant when $n \to \infty$.*
$[f(n) = O(n),\ n \to \infty] \Leftrightarrow [\frac{1}{c_1}n \leq f(n) \leq c_2 n]$, *$c_1, c_2$ are real constants.*

$$N(n) \leq 2^{n+O(\sqrt{n \log n})} \text{ [9]} \tag{14}$$

$$\tag{15}$$

The situation of the Happy Ending Problem is therefore, today, as follows.

$$2^{n-2} + 1 \leq N(n) \leq 2^{n+O(\sqrt{n \log n})} \tag{16}$$

References

[1] Chung, F., and Graham, R. Forced convex n-gons in the plane. *Discrete Comput. Geom. vol. 19*, no. 3, Special Issue (1998), pp. 367–371. Dedicated to the memory of Paul Erdős.

[2] Cintra, R., Rêgo, L., de Oliveira, H., and de Souza, R. On a density for sets of integers. *arXiv:1502.02601* (2015).

[3] Erdős, P., and Szekeres, G. A combinatorial problem in geometry. *Compositio Mathematica vol. 2* (1935), pp. 463–470.

[4] Erdős, P., and Szekeres, G. On some extremum problems in elementary geometry. *Mathematical Institute, Eötvös Loránd University, Budapest and University Adelaide vol. 3* (1960), pp. 53–62.

[5] Graham, R., and Butler, S. *Rudiments of Ramsey theory*, 2nd ed., vol. 123 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2015.

[6] Graham, R., Knuth, D., and Patashnik, O. *Concrete mathematics*, 2nd ed. Addison-Wesley Publishing Company, Reading, MA, 1994. A foundation for computer science.

[7] Graham, R., Rothschild, B., and Spencer, J. *Ramsey theory.*, 2nd ed. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2013.

[8] Hoffman, P. *The man who loved only numbers.* Hyperion Books, New York, 1998.

[9] Holmsen, A., Mojarrad, H., Pach, J., and Tardos, G. Two extensions of the erdos–szekeres problem. *arXiv:1710.11415* (2017).

[10] Kalbfleisch, J. D., Kalbfleisch, J. G., and Stanton, R. G. A combinatorial problem on convex n-gons. In *Proc. Louisiana Conf. on Combinatorics, Graph Theory and Computing (Louisiana State Univ., Baton Rouge, La., 1970).* Louisiana State Univ., Baton Rouge, La., 1970, pp. 180–188.

[11] Kleitman, D., and Pachter, L. Finding convex sets among points in the plane. *Discrete Comput. Geom. vol. 19*, no. 3, Special Issue (1998), pp. 405–410. Dedicated to the memory of Paul Erdős.

[12] Ledermann, W. Issai Schur and his school in Berlin. *Bull. London Math. Soc. 15*, no. 2 (1983), pp. 97–106.

[13] Ramsey, F. P. On a problem of formal logic. *Proc. London Math. Soc. vol. 30*, no. 4 (1930), pp. 264–286.

[14] Schur, I. Über die Kongruenz $x^m+y^m \equiv z^m \pmod{p}$. *Jahresber. Dtsch. Math.-Ver. vol. 25* (1916), pp. 114–117.

[15] Suk, A. On the erdos–szekeres convex polygon problem. *J. Amer. Math. Soc. vol. 30*, no. 4 (2017), pp. 1047–1053.

[16] Szekeres, G., and Peters, L. Computer solution to the 17-point Erdos-Szekeres problem. *ANZIAM J. vol. 48*, no. 2 (2006), pp. 151–164.

[17] Thompson, J., Martinsson, T., Martinsson, P.-G., and Thompson, J. *Wahlström & Widstrands matematiklexikon.* Wahlström & Widstrand, 1991.

[18] Tóth, G., and Valtr, P. Note on the Erdos-Szekeres theorem. *Discrete Comput. Geom. vol. 19*, no. 3, Special Issue (1998), pp. 457–459. Dedicated to the memory of Paul Erdős.

[19] Tóth, G., and Valtr, P. The Erdos-Szekeres theorem: upper bounds and related results. In *Combinatorial and computational geometry*, vol. 52 of *Math. Sci. Res. Inst. Publ.* Cambridge Univ. Press, Cambridge, 2005, pp. 557–568.