The de Bruijn–Erdös–Hanani theorem

Nikolai V. Ivanov

Contents

Preface

The present paper is devoted to a somewhat idiosyncratic account of the theorem of de Bruijn–Erdös and Hanani from the combinatorics of finite geometries and its various proofs. Among the proofs discussed are the original proofs by de Bruijn–Erdös and by Hanani (the latter seems to be largely forgotten, being published in a hard to access journal) and few others. Each of these proofs sheds new light on the theorem, illustrating the maxim that proofs are more important than the theorems proved. Some proofs and arguments in this paper seem to be new. I explain how one of the proofs was discovered, and how another one could have been discovered. See Sections 4 and 8.

1. N. Bourbaki and the *"Kvant"* magazine

Problem. *Let* E *be a set of* n *elements. Suppose that* m *different subsets of* E *(not equal to* E *itself) are selected in such a way that for every two elements of* E *there is exactly one selected subset containing both these elements. Prove that* $m \geqslant n$.

When an equality is possible?

In 1970 this problem was included as the Problem M5 in the very first issue of the Soviet *"Kvant"* magazine and attributed to N. Bourbaki [B-70]. The intended audience of the *"Kvant"* magazine (its name means *"Quantum"*) was the school students in the USSR of the last two-three grades. Nowadays the audacity of the editorial board inspires awe: Problem M5 was offered to this audience exactly as it is cited above, as an abstract problem about finite sets without any motivation and any hints. The readers were expected to be interested in this problem and to appreciate its beauty without any crutches.

In 1970 I was among the intended audience of *"Kvant"*, but I was more interested in the foundations of mathematics and in the set theory than in the combinatorics of finite sets. I easily found this problem in the Russian translation [B-65] of the *"Théorie des ensembles"* by N. Bourbaki. It turned out to be the Exercise 12 to the section *"Calcul sur les entiers"*. In all editions this exercise is marked as one of the most difficult.

The editors of *"Kvant"* were faithful to N. Bourbaki in not offering any motivation. But, in contrast with *"Kvant"*, N. Bourbaki split the result into few steps, offered a hint to the key one, and stated the expected result in the case of the equality. The first two steps were rather easy, but the hint to key step turned out to be incomprehensible for me.

According to the authors of the solution [T] published in *"Kvant"* a few months later, they followed *"the hints of the author of the problem, N. Bourbaki himself"* and referred to [B-65]. The habit of N. Bourbaki to include in his tract recent results without attribution as exercises is well known, and was well known in 1970 in the Soviet Union. But it seems that neither the editors of the *"Kvant"* magazine, nor the authors of the solution [T] were aware that this result is due to N.G. de Bruijn and P. Erdös [dB-E] and H. Hanani [H1, H2].

Neither was I before by an accident I returned to this problem in 2016. By this time I was able to immediately recognize that this exercise from [B-65] is about points and lines in a geometry, and this realization quickly lead me to the de Bruijn–Erdös paper [dB-E]. The exercise turned out to be a quite faithful summary of the de Bruijn–Erdös proof, and the key part of the proof, summarized by N. Bourbaki as a hint, turned out to be nearly as obscure as this hint. Here is my translation of this exercise based on the reprint [B-06] of the 1970 edition (where it appears as the Exercise 14 to § 5). It is slightly different from the translation in [B-04].

Exercise. *Let* E *be a finite set of* n *elements,* $(a_j)_{1 \leqslant j \leqslant n}$ *be the sequence of elements of* E *arranged in some order,* $(A_i)_{1 \leqslant i \leqslant m}$ *be a sequence of parts of* E.

(a) *For each index* j, *let* k_j *be the number of indices* i *such that* $a_j \in A_i$; *for each index* i *let* $s_i = \mathrm{Card}(A_i)$. *Show that*

$$\sum_{j=1}^{n} k_j = \sum_{i=1}^{m} s_i \ .$$

(b) *Suppose that for each subset* $\{x, y\}$ *of two elements of* E, *there exists one and only one index* i *such that* x *and* y *are contained in* A_i. *Show that if* $a_j \notin A_i$, *then* $s_i \leqslant k_j$.

(c) *Under the assumptions of* (b), *show that* $m \geqslant n$. *(Let* k_n *be the least of the numbers* k_j; *show that one may assume that, whenever* $i \leqslant k_n$, $j \leqslant k_n$ *and* $i \neq j$, *one has* $a_j \notin A_i$, *and* $a_n \notin A_j$ *for all* $j \geqslant k_n$.)

(d) *Under the assumptions of* (b), *show that in order for* $m = n$ *to hold, it is necessary and sufficient that one of the following two cases occurs:*

(i) $A_1 = \{a_1, a_2, \ldots, a_{n-1}\}$, $A_i = \{a_{i-1}, a_n\}$ *for* $i = 2, \ldots, n$;

(ii) $n = k(k-1) + 1$, *each* A_i *is a set of* k *elements, and each element of* E *belongs to exactly* k *sets* A_i.

Remarks. Two aspects of this exercise need to be clarified. First, the parts A_i are implicitly assumed to be different from E. Second, the case (i) of the part (d) is expected to hold only up to renumbering of elements a_i and parts A_j.

The troubles with the hint. The parts (a) and (b) of this exercise are rather easy, and there is a hint for the part (c). But for me this hint turned out to be more of a riddle than of a help.

It would be quite easy to accept and follow the suggestion to consider the least of the numbers k_j. But why it should be k_n? The phrase *"Let k_n be the least of the numbers k_j"* is fairly hard to interpret (the expressions used in the French original and in the Russian translation have the same meaning). The standard usage of *"Let"* (and of *"Soit"* in French) in mathematics is to introduce new notations. But k_n is already defined.

The authors of the solution [T] found a clever way out. They introduce the number k_n *before* introducing other numbers k_j! This trick helps only partially: the question *"Why k_n?"* remains. The de Bruijn–Erdös exposition [dB-E] is better. They write *"Assume now that k_n is the smallest k_i …"*. This is less obscure, and amounts to renumbering elements of E, but leaves the question *"Why k_n?"* unanswered.

If one manages to put this question aside, there is another riddle: how the subscripts i, j, which are merely marking the points (and do not even need to be numbers) may be compared with k_n, which is a genuine characteristic of the point marked by n? Perhaps, this difficulty is encountered only by the categorically minded mathematicians; analysts appear to be quite comfortable with using the values of a function in its domain of definition.

Here de Bruijn and Erdös [dB-E] are again doing better. They write *"Assume … that* A_1, A_2, *…, A_{k_n} are lines through a_n"* (they call the parts A_i lines). This amounts to renumbering the parts A_i, and one may wonder why renumbering is treated as an assumption. The trick of the authors of [T] saves the day here for them. They simply denote the k_n lines through a_n by A_1, A_2, …, A_{k_n} and other lines by A_{k_n+1}, A_{k_n+2}, …, A_m.

There is one more riddle in the store. How one uses the assumption that k_n is the least of the numbers k_j in the proof of the claim in the hint? One does not, this claim is true without it.

Partially decrypting the hint. Even if one encounters all these troubles and is not aware of the de Bruijn–Erdös paper (like me in 1970), the hint still may be of some help. The first message is that it is important to know when an element a_j is not in the part A_i. Together with the part (b) this suggest that the inequalities $s_i \leqslant k_j$, which hold for $a_j \notin A_i$, should play a key role.

Another message is that the least of the numbers k_j should play some role. After wasting some time assuming that for a given u the number k_u is minimal among all numbers k_j and trying to use this minimality to prove something like stated in the hint, it is only natural to abandon this assumption and consider an arbitrary subscript u such that $1 \leqslant u \leqslant n$.

The 1970 proof of $m \geqslant n$. With no more than this limited help from this exercise (in 1970 I definitely understood less than in 2016) I managed to prove in the early 1970 the inequality $m \geqslant n$. Among my schoolmates this qualified as a solution of the Problem M5. This solution was lost long time ago. In April of 2016 and another time one year later I attempted to reconstruct this proof. In these attempts I encountered the same difficulties as in 1970, and it is likely that I dealt with them in the same manner. At the very least, the resulting proof does not use any tools not known to me at the time, and does not involve any tricks (such as the cyclic ordering of some parts A_i by de Bruijn–Erdös) which I was unlikely to discover at the time. It is presented in Section 2 below.

The question *"When an equality is possible?"* was considered by my classmates as too vague to be addressed seriously, and this was indirectly admitted by the authors of the solution [T]. If $m = n$, then (d) easily implies that $A_i \cap A_j \neq \emptyset$ if $i \neq j$. In fact, proving this property is an almost inevitable part of the proof of (d). This property means that the set E together with the parts A_i is a *finite projective plane*, possibly degenerate in the case (i) of the part (d). Therefore, this question amounts to the classification of finite projective planes and, to the best of my knowledge, it remains largely open. See the paper by Ch. Weibel [W] for a survey of the state of the art as of 2007, and [I1] for an introduction (not focusing on the finite case).

"Kvant" publishes a solution. *"Kvant"* published a solution [T] of the Problem M5 in the August or September of 1970, close to the beginning of the school year in the USSR (always September 1). The editors of the problem section wrote (see [T], p. 49):

> The letters to editors indicate that this problem is extremely difficult, but interesting. As a matter of fact, here we have two problems: 1) prove that $m \geqslant n$, 2) when an equality is possible?
>
> The first problem was completely solved only by *A. Suslin* from the city of Leningrad. His proof is based on a basic theorem of the linear algebra: if the number of n-vectors is greater than n, then they are linearly dependent.
>
> Looking for such a proof will be interesting for whose who are familiar with these notions. Nobody solved completely the second problem. Of course, this is not surprising, since, as it will be explained below, it can be reduced to a well known, but unsolved problem in mathematics.

Among my schoolmates, these remarks stirred a renewed interest in the problem. A. Suslin was known as a very strong problem solver and as a winner of the gold medal at 1967 International Mathematical Olympiad. Since only he submitted a complete solution, the problem had to be really difficult. Since he used tools going beyond the school level, the problem had to be even more difficult. And this caused a real interest in my unsubmitted to the *"Kvant"* solution. I had an outline as a sparsely filled with formulas sheet of paper. One of my schoolmates borrowed this sheet for few days, and I have not seen it anymore.

But I am not aware of any serious attempt to study the published solution [T]. For me it was almost as condensed and obscure as the N. Bourbaki hint. The role of the numbering of elements and parts is overemphasized:

> Let us pay attention once again to the way we numbered elements and sets.
>
> First of all, k_n is the least of the numbers $k_1, k_2, \ldots, k_{n-1}$ (sic! – *N.I.*). …

See [T], p. 51. And I always disliked random numerical examples, which are supposed to help the reader and are extensively used in [T]. I must admit that I did not even look at the last two pages of [T] before writing these comments, and, in particular, before writing down the proof in the next section. Surprisingly, it turned out that the proof [T] contains a gap: it is mentioned that $k_n = 2$ in the situation described in the case (i) of the Bourbaki exercise, but no proof that this is the only possibility is even attempted.

2. A solution of the N. Bourbaki exercise

The terminology and notations. In contrast with N. Bourbaki and with the *"Kvant"*, I have no reasons to hide the geometric content of this result. Following de Bruijn and Erdös, I will call the elements of E *points* and the sets A_i *lines*. Since the lines are assumed to be proper subsets of E, every point is contained in at least 2 lines. Indeed, if a point is contained in only one line, then all points are contained in this line, i.e. it is not a proper subset.

It is convenient to explicitly introduce a counterpart to the set E of points, namely the set of lines $\mathscr{L} = \{A_1, A_2, \ldots, A_m\}$. If the case (i) of the part (d) of the Bourbaki exercise occurs, up to renumbering of points and lines, then the pair $(E, \mathscr{L})$ is called a *near-pencil*. If the case (ii) of the part (d) occurs, then $(E, \mathscr{L})$ is called a *projective plane*.

I also do not see any reason to follow the outdated fashion of using *numerical* indices (i.e. subscripts), which amounts to ordering objects even when their order is irrelevant. Instead of this, for every point z we will denote by k_z the number of lines containing z, and for every line l we will denote by s_l the number of points in l, i.e. the number of elements of the set l.

The part (a) of the Bourbaki exercise. With the above notations the part (a) takes the form

$$(2.1) \qquad \sum_{l \in \mathscr{L}} s_l \;=\; \sum_{z \in E} k_z.$$

after interchanging the sides. This immediately follows from counting in two different ways the pairs $(z, l) \in E \times \mathscr{L}$ such that $z \in l$.

The part (b) of the Bourbaki exercise. For the rest of the paper we will assume that the assumption of the part (b) holds, i.e. that for every pair of distinct points there is exactly one line containing both of them. If a line l contains $\leqslant 1$ points, then removing l from the set of lines does not affects this assumption, and at the same time decreases number of lines by 1. Hence we may assume for the rest of the paper that every line contains at least 2 points.

With the above notations the part (b) takes the form

$$\text{If} \quad z \notin l, \quad \text{then} \quad s_l \leqslant k_z.$$

We will call these inequalities the *de Bruijn–Erdös inequalities*.

In order to prove the de Bruijn–Erdös inequalities, suppose that $z \notin l$. Then for every $y \in l$ there is a unique line containing $\{z, y\}$ and it is different from l because $z \notin l$. These lines are pairwise distinct because if $y, y' \in l$ and $y \neq y'$, then l is the only line containing $\{y, y'\}$. There are is s_l such lines and all of them contain z; therefore $s_l \leqslant k_z$.

Lines through an arbitrary point. Let $u \in E$ be an arbitrary point, let $p = k_u$ be the number of lines containing u, and let $\mathcal{U}$ be the set of these lines.

By the definition of $\mathcal{U}$, if a line l is not in $\mathcal{U}$, then $u \notin l$. For every $l \notin \mathcal{U}$ we have the de Bruijn–Erdös inequality $s_l \leqslant k_u$. By summing all these inequalities and taking into account that there are $m - p$ lines not belonging to $\mathcal{U}$, we see that

$$(2.2) \qquad \sum_{l \notin \mathcal{U}} s_l \leqslant (m - p) k_u .$$

Since every set of the form $\{u, y\}$ with $y \neq u$ is contained in one and only one line, the sets $l \smallsetminus \{u\}$ with $l \in \mathcal{U}$ are pairwise disjoint and form a partition of $E \smallsetminus \{u\}$. Since we assumed that $s_l \geqslant 2$ for all lines l, all these sets are non-empty. Let U be a set of representatives of these sets. In other terms, U is contained in $E \smallsetminus \{u\}$ and intersects every set $l \smallsetminus \{u\}$ with $l \in \mathcal{U}$ in exactly 1 point. In particular, U consists of exactly p points.

If $(l, z) \in \mathcal{U} \times U$ and $z \notin l$, then the de Bruijn–Erdös inequality $s_l \leqslant k_z$ holds. There are $p(p - 1)$ of such pairs (l, z) and hence $p(p - 1)$ of such inequalities. For each $l \in \mathcal{U}$ the number s_l occurs $p - 1$ times in the left hand sides of them, and for each $z \in U$ the number k_z occurs $p - 1$ times in the right hand sides. Hence the sum of all these inequalities is

$$(2.3) \qquad \sum_{l \in \mathcal{U}} (p - 1) s_l \leqslant \sum_{z \in U} (p - 1) k_z .$$

After dividing (2.3) by $p - 1$ we get

$$(2.4) \qquad \sum_{l \in \mathcal{U}} s_l \leqslant \sum_{z \in U} k_z .$$

Now it is only natural to take the sum of the inequalities (2.2) and (2.4) and conclude that

$$(2.5) \qquad \sum_{l \in \mathcal{L}} s_l \leqslant (m - p) k_u + \sum_{z \in U} k_z .$$

The left hand side of the inequality (2.5) is the same as the left hand side of the equality (2.1). The right hand side of (2.5) can be compared with the right hand side of the equality (2.1) if k_u is the least among the numbers k_z and $m \leqslant n$.

Proof of the inequality $m \geqslant n$. Now we are ready to do the part (c) of the Bourbaki exercise. Let $u \in E$ be a point such that k_u is the least of the numbers k_z over all points $z \in E$. Then

$$(2.6) \qquad (m - p) k_u \leqslant \sum_{z \in Y} k_z .$$

for every subset $Y \subset E$ consisting of $m - p$ points.

Suppose that $m \leqslant n$. Then the subset Y can be chosen to be disjoint from U (because U consists of p points). Let us choose an arbitrary Y disjoint from U and let $Z = Y \cup U$. Then Z is a subset of E consisting of m points and the inequalities (2.5) and (2.6) imply that

$$(2.7) \qquad \sum_{l \in \mathscr{L}} s_l \;\leqslant\; \sum_{z \in Z} k_z \;\leqslant\; \sum_{z \in E} k_z \,,$$

where the last inequality is strict unless $Z = E$. In view of (2.1) this inequality cannot be strict and hence $Z = E$ and $m = n$. Since $m \leqslant n$ implies that $m = n$, we see that $m \geqslant n$.

The case $m = n$. After the work done in the proofs of (a), (b), and (c), the part (d) nearly proves itself. As we will see, in this case all inequalities (2.2) – (2.7) are, in fact, equalities.

By (2.1) the leftmost and the rightmost sums in (2.7) are equal. It follows that $Z = E$ and hence $Y = E \smallsetminus U$. Moreover, the sides of each of the inequalities (2.5) and (2.6) are equal. Since k_u is the least of the numbers k_z, the equality of the sides of (2.6) implies that

$$(2.8) \qquad k_u = k_z \quad \text{for all} \quad z \in Y = E \smallsetminus U \,.$$

The fact that the sides of (2.5) are equal implies that the sides of each of the inequalities (2.2) and (2.4) are equal also. The equality of the sides of (2.2) implies that

$$(2.9) \qquad s_l = k_u \quad \text{for all} \quad l \notin \mathscr{U} \,.$$

Since the sides of (2.4) are equal, the sides of (2.3) are also equal. Since (2.3) is the sum of the inequalities $s_l \leqslant k_z$ over all pairs $(l, z) \in \mathscr{U} \times U$ such that $z \notin l$, the equality of the sides of (2.3) implies that $s_l = k_z$ for all $(l, z) \in \mathscr{U} \times U$ such that $z \notin l$. Equivalently,

$$(2.10) \qquad s_l = k_z \quad \text{if} \quad l \in \mathscr{U} \quad \text{and} \quad z \in U \smallsetminus l \,.$$

The rest of the proof splits into two subcases depending on if $p = 2$ or $p \geqslant 3$.

The subcase $p = 2$. In this case $\mathscr{U} = \{l, l'\}$ for some l, l' and hence $E = l \cup l'$. It follows that every line different from l, l' contains only 2 points, namely the points of its intersection with the lines l, l'. If $s_l, s_{l'} \geqslant 3$, then there are at least 4 points $z \neq u$ and the part (b) implies that $k_z \geqslant 3$ for every $z \neq u$. On the other hand, (2.8) implies that

$$k_z = k_u = p = 2$$

for every $z \notin U$. But U consists of only two points and hence $k_z \geqslant 3$ for no more than two points z. The contradiction shows that either $s_l = 2$ or $s_{l'} = 2$. We may assume that $s_l = 2$. Then $l = \{u, a\}$ for some $a \in E$ and every line different from l, l' has the form $\{a, z\}$ with $z \in l' \smallsetminus \{u\}$. It follows that $(E, \mathscr{L})$ is a near-pencil.

The subcase $p \geqslant 3$**.** The set U is a set of representatives of the sets $l \smallsetminus \{u\}$ with $l \in \mathscr{U}$. For any two lines l, $l' \in \mathscr{U}$ the assumption $p \geqslant 3$ implies that there exists a point $z \in$ U such that $z \notin l, l'$. If z is such a point, then (2.10) implies that

$$s_l = k_z = s_{l'}.$$

Similarly, if z, $z' \in$ U, then there exists a line $l \in \mathscr{U}$ such that z, $z' \notin l$ and hence

$$k_z = s_l = k_{z'}.$$

It follows that in the subcase $p \geqslant 3$ all numbers s_l, k_z with $l \in \mathscr{U}$ and $z \in$ U are equal. Since $k_u = p \geqslant 3$ is the smallest of the numbers k_z over all $z \in$ E, it follows that

$$s_l = k_z \geqslant 3 \quad \text{for all} \quad l \in \mathscr{U}, \quad z \in \text{U}.$$

Let $l \in \mathscr{U}$, and let y be the unique element of U contained in l. Since $s_l \geqslant 3$, there exists a point $x \in l$ not equal to u, y. We can replace in U the point y by the point x and get a new set of representatives U$'$. Then all previous results apply to U$'$ in the role of U. In particular, $k_x = k_z$ for all $z \in \text{U} \smallsetminus l = \text{U}' \smallsetminus l$ and hence

$$k_x = k_z \quad \text{for all} \quad z \in \text{U}.$$

On the other hand, $x \notin$ U and hence $k_x = k_u$ by (2.8) applied to the original set U. At the same time (2.8) implies that $k_u = k_z$ for all $z \notin$ U and hence

$$k_x = k_z \quad \text{for all} \quad z \notin \text{U}.$$

It follows that all numbers k_z are equal. At the same time by (2.9) and (2.10) every s_l is equal to some k_z. It follows that all numbers s_l, k_z with $l \in \mathscr{L}$ and $z \in$ E are equal. It remains to apply the following lemma.

Lemma 1. *If all the numbers* s_l, k_z *are equal, then* (E, $\mathscr{L}$) *is a projective plane.*

Proof. Let k be the common value of the numbers s_l, k_z, and let $y \in$ E. The sets $l \smallsetminus \{y\}$ with $y \in l$ are pairwise disjoint and form a partition of E $\smallsetminus \{y\}$. Each of them consists of

$$s_l - 1 = k - 1$$

points, and there are $k_y = k$ such sets. It follows that the number n of elements of E is equal to $k(k - 1) + 1$. Therefore, (E, $\mathscr{L}$) is a projective plane. $\blacksquare$

Remarks. A key step of this solution and the solution [T] differ from the de-Bruijn–Erdös paper in the same way: the cyclic order argument of de Bruijn–Erdös (see Section 3) is replaced by the inequalities (2.3) and (2.4).

3. The de Bruijn–Erdös proof

Since the proof presented in Section 2 grow out of a summary of the de Bruijn–Erdös proof, albeit not quite understood, it is not surprising that the two proofs have a lot in common. In the following exposition of the de Bruijn–Erdös proof we will use the notations of Section 2 and will refer to Section 2 for the arguments which differ from [dB-E] only in the notations and the amount of details. The Bruijn–Erdös paper is concise on the border of being cryptic.

The de Bruijn–Erdös proof begins with the parts (a) and (b) of the Bourbaki exercise. After this de Bruijn–Erdös introduce k_u as the smallest among all numbers k_z (and denote it by k_n). Then de Bruijn–Erdös observe that it can assumed that every line contains at least two points. Following the notations of Section 2, let us denote by $\mathscr{U}$ the set of all lines containing u. By the de Bruijn–Erdös inequalities $s_l \leqslant k_u$ for every $l \notin \mathscr{U}$. The inequalities (2.2) and (2.6) follows. The following argument plays a role similar to the role of the inequality (2.3).

The cyclic order argument. Let $l_1, l_2, \ldots, l_p$ be a cyclically ordered list of elements of $\mathscr{U}$. We treat the subscripts $1, 2, \ldots, p$ as integers mod p. For each $i = 1, 2, \ldots, p$ let us choose some point $a_j \in l_j \smallsetminus \{u\}$ and let U be the set of these points. Let

$$s_i = s_{l_i} \quad \text{and} \quad k_j = k_{a_j}.$$

Since $a_{i+1} \notin l_i$, by the de Bruijn–Erdös inequalities $s_i \leqslant k_{i+1}$ for all $i = 1, 2, \ldots, p$, i.e.

(3.1) $s_1 \leqslant k_2, \quad s_2 \leqslant k_3, \quad \ldots, \quad s_p \leqslant k_1.$

By summing the inequalities (3.1) one concludes that

(3.2) $$\sum_{j=1}^{p} s_j \leqslant \sum_{j=1}^{p} k_j.$$

The inequality (3.2) is nothing else but another form of (2.4). The arguments of Section 2 show that (3.2) implies that $m \geqslant n$. In fact, de Bruijn and Erdös do not bother to write down even the inequality (3.2), to say nothing about other details presented in Section 2.

The case $m = n$. In view of the equality (2.1) in this case the left hand and the right hand sides of the inequality (3.2) are equal. Together with (3.1) this implies that

(3.3) $s_1 = k_2, \quad s_2 = k_3, \quad \ldots, \quad s_p = k_1.$

Similarly, in this case the left hand and the right hand sides of the inequality (2.6) are equal. Since $m = n$, one can take $Y = E \smallsetminus U$ in (2.6). It follows that $k_u = k_z$ for all $z \in E \smallsetminus U$. Finally, the left hand and the right hand sides of the inequality (2.2) are equal. It follows that

$s_l = k_u$ for all $l \notin \mathcal{U}$. By combining the last two observations, we see that

$$s_l = k_z \quad \text{for all} \quad l \in \mathcal{L} \smallsetminus \mathcal{U}, \quad z \in E \smallsetminus U.$$

Since $m = n$, both sets $\mathcal{L} \smallsetminus \mathcal{U}$ and $E \smallsetminus U$ consist of $n - p$ elements. It follows that one can number the points and lines in such a way that (in the notation of the Bourbaki exercise)

$$s_1 = k_1, \quad s_2 = k_2, \quad \ldots, \quad s_n = k_n.$$

As the next step, let us renumber the points and lines once more and assume that

(3.4) $k_1 \geqslant k_2 \geqslant \ldots \geqslant k_n$.

The rest of the proof splits into two subcases depending on if $k_1 > k_2$ or not.

The subcase $k_1 > k_2$. In this case $s_1 = k_1 > k_i$ for all $i \geqslant 2$. By the de Bruijn–Erdös inequalities this implies that $a_i \in A_1$ for all $i \geqslant 2$. It follows that $(E, \mathcal{L})$ is a near-pencil.

The subcase $k_1 = k_2$. Suppose that $k_j < k_1 = k_2$ for some j. By the de Bruijn–Erdös inequalities a_j belongs to the both lines A_1 and A_2. This is possible for only one point, namely the point of the intersection of the lines A_1 and A_2. In view of (3.4), this may happen only if

$$k_1 = k_2 = \ldots = k_{n-1} > k_n$$

and hence $s_j = k_j > k_n \geqslant 2$ for all $j \neq n$. It follows that $s_j \geqslant 3$ if $j < n$. In particular, all k_n lines containing a_n consist of $\geqslant 2$ points and all except, perhaps, the line A_n, consist of $\geqslant 3$ points. Therefore one can choose 2 points $x, y \neq a_n$ on one of these lines, and a point $z \neq a_n$ on some other line. Let $l_j, l_{j'}$ be the lines containing the pairs $\{x, z\}$ and $\{y, z\}$ respectively. Then $j \neq j'$ and $a_n \notin l_j, l_{j'}$. Hence the de Bruijn–Erdös inequalities imply that $s_j, s_{j'} \leqslant k_n$, contrary to the fact that $s_j > k_n$ if $j \neq n$. The contradiction shows that all numbers k_j are equal, and hence all numbers s_i, k_j are equal. Now the observation at the end of Section 2 implies that $(E, \mathcal{L})$ is a projective plane.

Intersection of lines. After the proof is completed, de Bruijn–Erdös point out that in the subcase $k_1 = k_2$ of the case $m = n$ every two lines intersect. Indeed, if l', l'' are two disjoint lines and $a \in l''$, then there are $s_{l'}$ lines containing a and intersecting l', and still one more line, namely l'', containing a. Therefore $k_a \geqslant s_{l'} + 1$, contrary to the fact all numbers k_z, s_l are equal. In fact, every two lines obviously intersect in the subcase $k_1 > k_2$ also.

Why k_n? Now it is clear why the smallest of the numbers k_z is denoted by k_n. The number k_n is indeed the smallest if the points are ordered in such a way that (3.4) holds. At the same time (3.4) plays almost no role in the proof. One may speculate that (3.4) and notation k_n for the smallest of the numbers k_z are remnants of an earlier approach to the theorem.

4. From de Bruijn–Erdös to systems of distinct representatives

The cyclic order argument and systems of distinct representatives. The key step of the de Bruijn–Erdös proof is the cyclic order argument used to prove the inequality (3.2) and the equalities (3.3) in the case $m = n$. Ultimately, the cyclic order argument is based on the fact that $a_{i+1} \notin l_i$ for all $i = 1, 2, \ldots, p$, i.e. on the fact that $i \longmapsto a_{i+1}$ is a system of distinct representatives for the family $i \longmapsto E \smallsetminus l_i$ of subsets of E, where $i = 1, 2, \ldots, p$.

Once this is realized, it is only natural to look for a system of distinct representatives of the full family $l \longmapsto E \smallsetminus l$ of the complements of lines, i.e. for an injective map $l \longmapsto a(l)$ from $\mathscr{L}$ to E such that $a(l) \in E \smallsetminus l$ for all $l \in \mathscr{L}$.

By the well known Ph. Hall's marriage theorem, such a system of distinct representatives exists if and only if for every subset $\mathscr{K} \subset \mathscr{L}$ the union

$$(4.1) \qquad \bigcup_{l \in \mathscr{K}} E \smallsetminus l \;=\; E \smallsetminus \bigcap_{l \in \mathscr{K}} l$$

contains $\geqslant |\mathscr{K}|$ elements, where $|X|$ denotes the number of elements of a set X. But the intersection of $\geqslant 2$ lines consists of $\leqslant 1$ points, and, almost obviously, this condition holds.

The message. All this emerged in my mind in one instant as an irreducible revelation. My first thought after this revelation was that it cannot be true, because if it is true, then everybody writing about this topic would use systems of distinct representatives. Perhaps, the right question is not how I came up with this idea, but why experts missed it. The rest of this section is devoted to the proof [I2] based on this revelation.

Proof of $m \geqslant n$**.** We may assume that $m \leqslant n$. Let $\mathscr{K}$ be a subset of $\mathscr{L}$. If $|\mathscr{K}| = 1$, then (4.1) is the complement of a line and hence contains $\geqslant 1$ elements. If $2 \leqslant |\mathscr{K}| \leqslant m - 1$, then (4.1) is a complement in E of $\leqslant 1$ point and hence contains

$$\geqslant n - 1 \geqslant m - 1 \geqslant |\mathscr{K}|$$

elements. If $|\mathscr{K}| = m$, then (4.1) contains $n \geqslant m = |\mathscr{K}|$ elements. Therefore there exists a system of distinct representatives for the family $l \longmapsto E \smallsetminus l$, i.e. there exists an injective map $l \longmapsto a(l)$ such that $a(l) \notin l$ for every l. By the de Bruijn–Erdös inequalities

$$(4.2) \qquad s_l \leqslant k_{a(l)} \quad \text{for every} \quad l \in \mathscr{L}.$$

By summing all these inequalities and using the injectivity of $l \longmapsto a(l)$ we see that

$$(4.3) \qquad \sum_{l \in \mathscr{L}} s_l \;\leqslant\; \sum_{l \in \mathscr{L}} k_{a(l)} \;\leqslant\; \sum_{z \in E} k_z \,.$$

Moreover, the second inequality is strict unless $m = n$ (otherwise the last sum has more positive summands than the previous one). But (2.1) implies that both inequalities in (4.3) should be actually equalities. It follows that $m = n$. Moreover, in view of the inequalities (4.2), it follows that $s_l = k_{a(l)}$ for every $l \in \mathcal{L}$ (under the assumption $m \leqslant n$).

The case $m = n$**.** Suppose that a point z is contained in $\geqslant m - 1$ lines. Each of these lines contains at least one point in addition to z. Since $m = n$, there are no other points and z is contained in exactly $m - 1$ lines. Since there are exactly m lines, only one line does not contain z. This line should contain all points $\neq z$. It follows that $(E, \mathcal{L})$ is a near-pencil.

Suppose now that no point is contained in $\geqslant m - 1$ lines. Let $\mathcal{K}$ be a proper subset of $\mathcal{L}$. If $|\mathcal{K}| = 1$, then (4.1) is equal to $E \smallsetminus l$ for some line l. If $E \smallsetminus l$ consists of only one point z, then by the de Bruijn–Erdös inequalities z is contained in $\geqslant s_l = m - 1$ lines, contrary to the assumption. Therefore, (4.1) contains $\geqslant 2 = |\mathcal{K}| + 1$ points. If $|\mathcal{K}| \leqslant m - 2$, then (4.1) contains $\geqslant n - 1 = m - 1 \geqslant |\mathcal{K}| + 1$ points. Finally, if $|\mathcal{K}| = m - 1$, then (4.1) contains all $n = m = |\mathcal{K}| + 1$ points because no point is contained in $\geqslant m - 1$ lines.

We see that (4.1) contains $\geqslant |\mathcal{K}| + 1$ elements for every proper subset $\mathcal{K} \subset \mathcal{L}$. This allows to get from the marriage theorem more than just the existence of a system of distinct representatives. Let $\lambda \in \mathcal{L}$ and $z \in E \smallsetminus \lambda$. Then there exists a system of distinct representatives $l \longmapsto a(l)$ such that $a(\lambda) = z$. This immediately follows from an application of the marriage theorem to the family of sets $(E \smallsetminus \{z\}) \smallsetminus l$ with $l \in \mathcal{L} \smallsetminus \{\lambda\}$.

Since $m \leqslant n$, the existence of a system of distinct representatives $l \longmapsto a(l)$ such that $a(\lambda) = z$ implies that $s_\lambda = k_{a(\lambda)} = k_z$. Therefore, $z \notin l$ implies that $s_l = k_z$ and hence every line containing z intersects l. It follows that every two lines intersect.

If E cannot be obtained as the union of two lines, then for every two lines l, l' there exists a point z such that $z \notin l, l'$ and hence $s_l = k_z = s_{l'}$. In this case all the numbers s_l, k_z are equal and hence $(E, \mathcal{L})$ is a projective plane by Lemma 1 at the end of Section 2. If there exist two lines l, l' such that $E = l \cup l'$, then $k_y = 2$, where y is the point of intersection of l and l', and the proof is completed by applying the following lemma.

Lemma 2. *If* $m = n$ *and* $k_y = 2$ *for some point* y, *then* $(E, \mathcal{L})$ *is a near-pencil.*

Proof. Let l, l' be the lines containing y. Then $E = l \cup l'$ and there are $n = s_l + s_{l'} - 1$ points. In addition to the lines l, l' there are $(s_l - 1)(s_{l'} - 1)$ lines consisting of a point in $l \smallsetminus \{y\}$ and a point in $l' \smallsetminus \{y\}$. If $s_l \geqslant s_{l'} \geqslant 3$, then the number m of lines is

$$\geqslant 2 + (s_l - 1)(s_{l'} - 1) \geqslant 2 + 2(s_l - 1) = 2s_l \geqslant s_l + s_{l'} = n + 1,$$

contrary to the assumption $m = n$. Therefore one of the lines l, l' consists of 2 points and hence $(E, \mathcal{L})$ is a near-pencil. ∎

5. Linear algebra and the inequality $m \geqslant n$

A proof of the inequality $m \geqslant n$ based on the linear independence. This proof was communicated to me by F. Petrov [P]. I believe that this is essentially the proof found by A. Suslin.

Let $\mathbf{R}^{\mathscr{L}}$ be the vector space of maps $\mathscr{L} \longrightarrow \mathbf{R}$ with the scalar product

$$(v, \, w) \;\; = \;\; \sum_{l \, \in \, \mathscr{L}} v(l) \, w(l).$$

Every $z \in E$ defines a map $v_z \colon \mathscr{L} \longrightarrow \mathbf{R}$ by the rule $v_z(l) = 1$ if $z \in l$ and $v_z(l) = 0$ otherwise. There are n maps v_z. Since the dimension of $\mathbf{R}^{\mathscr{L}}$ is equal to m, it is sufficient to prove that the maps v_z are independent as vectors of $\mathbf{R}^{\mathscr{L}}$.

The scalar product $(v_z, \, v_z)$ is equal to the number of lines containing the point z, and hence $(v_z, \, v_z) \geqslant 2$ for all $z \in E$. If $z \neq y$, then $(v_z, \, v_y)$ is equal to the number of lines containing both z and y, and hence $(v_z, \, v_y) = 1$. If the vectors v_z are linearly dependent, then

$$\sum_{z \, \in \, E} c_z v_z \;\; = \;\; 0$$

for some real numbers c_z, $z \in E$, such that not all c_z are equal to 0. For every $y \in E$ taking the scalar product of this equality with the vector v_y results in the equality

$$\sum_{z \, \in \, E} c_z (v_z, \, v_y) \;\; = \;\; 0.$$

Since $(v_z, \, v_y) = 1$ for all $z \neq y$, this equality implies that

$$c_y ((v_y, \, v_y) - 1) \; + \; \sum_{z \, \in \, E} c_z \;\; = \;\; 0.$$

Since $(v_y, \, v_y) \geqslant 2$, it follows that the coefficient c_y and the sum

$$\sum_{z \, \in \, E} c_z$$

have opposite signs. But since not all c_y are equal to 0, this cannot be true for all $y \in E$. The contradiction shows that vectors v_z are linearly independent and hence $m \geqslant n$. $\blacksquare$

Standard linear algebra proofs of the inequality $m \geqslant n$. In order to present standard proofs it is convenient to return to the notations of the N. Bourbaki exercise. Let M be the *incidence matrix* of the points a_j and sets A_i. Namely, M is an $n \times m$ matrix with entries $m_{ji} = 1$ if $a_j \in A_i$ and $m_{ji} = 0$ otherwise. Let us consider the product MM^{T}, where M^{T} is the matrix transposed to M. It is an $n \times n$ matrix with all non-diagonal entries equal to 1 and

with diagonal entries $k_1,\ k_2,\ \dots,\ k_n$. The most classical linear algebra proofs, going back to the paper [Bo] by R.C. Bose, proceed with the computation of the determinant of MM^T. It is rarely presented in details; apparently, it is expected that the readers enjoy computations of determinants. Curious readers may find a computation of $\det MM^T$ at the end of this section; in particular, the computation shows that this determinant is non-zero. The non-vanishing of $\det MM^T$ means that the rank of the matrix MM^T is equal to n, and this implies that the rank of M is $\geqslant n$. Since M is an $n \times m$ matrix, this may happen only if $m \geqslant n$.

More modern expositions avoid computation of the determinant $\det MM^T$ by observing that MM^T is equal to the sum of the diagonal matrix with the diagonal entries

$$k_1 - 1,\ k_2 - 1,\ \dots,\ k_n - 1$$

and the $n \times n$ matrix J with all entries equal to 1. Since $k_j \geqslant 2$ and hence $k_j - 1 \geqslant 1$ for all j, the first matrix is positive definite. The matrix J is positive semi-definite, although is not definite. In fact, the associated quadratic form $\mathbf{x} J \mathbf{x}^T$, where $\mathbf{x} = (x_1,\ x_2,\ \dots, x_n)$ is a row vector, is equal to $(x_1 + x_2 + \dots + x_n)^2$. It follows that the sum MM^T of these matrices is positive definite and hence has the rank n. As above, this implies that $m \geqslant n$.

Comparing the proofs. The standard proofs do not fit the *"Kvant"* description of the proof by A. Suslin: they use more advanced tools than the theorem about the linear dependence of more than n vectors in an n-dimensional vector space. One can find a proof based only on this theorem in the unpublished book draft [BF] by L. Babai and P. Frankl. But even in this remarkable book it is hidden in the exercises. See Exercise 4.1.5 and its solution on p. 184. The preference for using the matrix MM^T seems to be a part of a dominating culture. On the other hand, all proofs based on the linear algebra more or less explicitly reduce the inequality $m \geqslant n$ to the following lemma and then prove it.

Lemma. *Let* V *be an* m-*dimensional vector space over* $\mathbf{R}$ *equipped with a scalar product* $(\bullet,\ \bullet)$. *Let* P *be a set of* n *vectors in* V. *Suppose that there exists* $\lambda \in \mathbf{R}$, $\lambda > 0$, *such that*

$$(u,\ u) > \lambda \quad and \quad (v,\ w) = \lambda$$

for every $u \in$ P *and every two distinct vectors* $v,\ w \in$ P. *Then* $m \geqslant n$. ■

A generalization. The linear algebra proofs apply with only trivial changes to a more general situation. Namely, it is sufficient to assume that there exist a natural number $\lambda \geqslant 1$ such that every two distinct points are contained in exactly λ lines and every point is contained in $> \lambda$ lines. Then the conclusion $m \geqslant n$ still holds. This is due to H.J. Ryser [R]. Apparently, no combinatorial proof of Ryser's theorem is known. Ryser [R] also used linear algebra to provide a description of the case $m = n$ similar to de Bruijn–Erdös description in the case $\lambda = 1$.

The determinant of MM^T**.** For the benefit of the readers who do not like to compute the determinants themselves, here is a computation of $\det MM^T$ following the textbook [HP].

Let $m_j = k_j - 1$ for all j. Then

$$
MM^T = \begin{bmatrix}
m_1 + 1 & 1 & 1 & \ldots\ldots & 1 & 1 \\
1 & m_2 + 1 & 1 & \ldots\ldots & 1 & 1 \\
1 & 1 & m_3 + 1 & \ldots\ldots & 1 & 1 \\
& & & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \\
1 & 1 & 1 & \ldots\ldots & 1 & m_n + 1
\end{bmatrix} .
$$

Let us subtract the first row from every other one and get the matrix

$$
\begin{bmatrix}
m_1 + 1 & 1 & 1 & \ldots\ldots & 1 & 1 \\
-m_1 & m_2 & 0 & \ldots\ldots & 0 & 0 \\
-m_1 & 0 & m_3 & \ldots\ldots & 0 & 0 \\
& & & \ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \\
-m_1 & 0 & 0 & \ldots\ldots & 0 & m_n
\end{bmatrix} .
$$

For $j = 2, 3, \ldots, n$, let us multiply the j-th column by m_1/m_j (recall that $k_j \geqslant 2$ and hence $m_j \geqslant 1 > 0$) and add the result to the first column. We get the matrix

$$
\begin{bmatrix}
D & 1 & 1 & \ldots\ldots & 1 & 1 \\
0 & m_2 & 0 & \ldots\ldots & 0 & 0 \\
0 & 0 & m_3 & \ldots\ldots & 0 & 0 \\
& & & \ldots\ldots\ldots\ldots\ldots\ldots & & \\
0 & 0 & 0 & \ldots\ldots & 0 & m_n
\end{bmatrix} ,
$$

where $\quad D = m_1 + 1 + \displaystyle\sum_{j=2}^{n} \frac{m_1}{m_j} = m_1 + \sum_{j=1}^{n} \frac{m_1}{m_j} .$

It follows that $\quad \det MM^T = D \cdot \displaystyle\prod_{j=2}^{n} m_j = \prod_{j=1}^{n} m_j \cdot \left(1 + \sum_{j=1}^{n} \frac{1}{m_j} \right) \neq 0 .$

6. Hanani's theorem

Two papers of H. Hanani. According to the Th. Motzkin [M], the first proof of the inequality $m \geqslant n$ and, it seems, of the full de Bruijn–Erdös theorem, was given in 1938 by H. Hanani. He published an outline of his proof [H1] only in 1951. Later on H. Hanani published a detailed exposition [H2] of a simplified proof. In fact, in [H2] he proved (at no extra cost) a stronger version of the de Bruijn–Erdös theorem. He also used his methods to prove a 3-dimensional analogue dealing with points, lines, and planes.

Hanani's Theorem. *Under the previous assumptions, let* $\mathrm{L} \in \mathscr{L}$ *be a line containing the maximal number of points among all lines, let* $\mathscr{P}$ *be the set of all lines intersecting* L *(in particular,* $\mathrm{L} \in \mathscr{P}$ *), and let* p *be the number of elements of* $\mathscr{P}$*. Then* $p \geqslant n$*, and if* $p = n$*, then* $\mathscr{P} = \mathscr{L}$ *and* $(\mathrm{E}, \mathscr{L})$ *is either a near-pencil, or a projective plane.*

Suppose that $n \geqslant p$ and $(\mathrm{E}, \mathscr{L})$ is not a near-pencil. As usual, we assume that every line contains $\geqslant 2$ points. Let $a = s_{\mathrm{L}}$. Let K be the line with the maximal number of points among the lines different from L, and let $b = s_{\mathrm{K}}$. Then $a \geqslant b$. The strategy is to estimate n, or, what is the same, $n - 1$ in terms of a and b both from the below and from the above.

Hanani's Lemma. *If* $x \in \mathrm{L}$*, then*

(6.1) $k_x - 1 \;\geqslant\; \dfrac{n - a}{b - 1}\,.$

Proof. Let us consider pairs (l, y) such that l is a line containing x and y is a point in $l \smallsetminus \mathrm{L}$. Such a pair is uniquely determined by the point y and hence there are $n - a$ such pairs. A line l occurs in such a pair if and only if $x \in l$ and $l \neq \mathrm{L}$. It follows that there are $k_x - 1$ choices of l. Given a line l, there are $\leqslant b - 1$ choices for the point y. Therefore the number $n - a$ of such pairs is $\leqslant (k_x - 1)(b - 1)$. The lemma follows. ∎

An upper estimate of $n - 1$. By summing the inequalities (6.1) over all $x \in \mathrm{L}$ and adding 1 in order to account for the line L itself, we can estimate p from below and conclude that

(6.2) $\quad n \;\geqslant\; p \;\geqslant\; 1 + a\,\dfrac{n - a}{b - 1} \;=\; 1 + a\,\dfrac{(n - 1) - (a - 1)}{b - 1}$

or, what is the same,

(6.3) $\quad a(a - 1) \;\geqslant\; (n - 1)(a - b + 1)\,.$

The inequality (6.3) provides an estimate of $n - 1$ from the above.

A lower estimate of $n - 1$. There is another way to estimate p from below. By a miracle, this other estimate of p from the same side leads to an estimate of $n - 1$ from the other side. Let z be a point in $L \cap K$ if $L \cap K \neq \emptyset$, and an arbitrary point of L otherwise. For every $x \in L \smallsetminus \{z\}$, $y \in K \smallsetminus \{z\}$ there is a unique line l containing $\{x, y\}$. All these lines are distinct, not equal to L, and do not contain z. Clearly, there are $(a - 1)(b - 1)$ of such lines. A lower estimate of number k_z of lines containing z is provided by (6.1). It follows that

$$n \geqslant p \geqslant 1 + \frac{n - a}{b - 1} + (a - 1)(b - 1)$$

$$= 1 + \frac{(n - 1) - (a - 1)}{b - 1} + (a - 1)(b - 1)$$

and hence $(n - 1)(b - 1) \geqslant (n - 1) - (a - 1) + (a - 1)(b - 1)^2$ and

$$(n - 1)(b - 2) \geqslant (a - 1)(b^2 - 2b) = (a - 1)b(b - 2).$$

Since $b \geqslant 2$, it follows that either $b = 2$, or $n - 1 \geqslant (a - 1)b$. If $b = 2$, then all lines except L consist of 2 points and the inequality (6.3) implies that $a \geqslant n - 1$. But $L \neq E$ and hence $a \leqslant n - 1$. It follows that $a = n - 1$ and hence L contains all points of E except one and $(E, \mathscr{L})$ is a near-pencil, contrary to the assumption. Therefore

(6.4) $n - 1 \geqslant (a - 1)b$.

The inequality (6.4) provides an estimate of $n - 1$ from the below.

Combining the two estimates. After multiplying the inequality (6.4) by $(a - b + 1)$ and combining the result with the inequality (6.3), we see that

$$a(a - 1) \geqslant (a - 1)b(a - b + 1)$$

and hence $a \geqslant b(a - b + 1) = b(a - b) + b$, or, what is the same

$$0 \geqslant (b - 1)(a - b).$$

Since $b > 1$, this implies that $b \geqslant a$. On the other hand, $b \leqslant a$ by the definition of a, b. It follows that $a = b$. By combining $a = b$ with the inequalities (6.3) and (6.4) we conclude, respectively, that $a(a - 1) \geqslant n - 1$ and $n - 1 \geqslant a(a - 1)$. It follows that $n - 1 = a(a - 1)$ and hence $n = a(a - 1) + 1$. By combining this with (6.2) we see that

$$a(a - 1) + 1 = n \geqslant p \geqslant 1 + a \frac{a(a - 1) + 1 - a}{a - 1} = 1 + a(a - 1).$$

It follows that $n = p$, and therefore $p \geqslant n$ if the inequality $p \leqslant n$ is not assumed.

18

The case $p = n$**.** As we just saw, in this case $a = b$ and $n = a(a - 1) + 1$.

Let us prove first that every line belonging to $\mathscr{P}$ consists of exactly a points. Consider all pairs (l, y) such that $l \in \mathscr{P}$ and $y \in E \smallsetminus L$. The line l is uniquely determined by its point of intersection with L (which can be any point of L) and the point y. Therefore there are $a(n - a) = an - a^2$ such pairs. On the other hand, for every line $l \in \mathscr{P} \smallsetminus \{L\}$ there are $\leqslant a - 1$ choices of the point y and hence the number of such pairs is $\leqslant (p - 1)(a - 1)$. Moreover, if at least one line $l \in \mathscr{P} \smallsetminus \{L\}$ has $< a$ points, then the number of such pairs is $< (p - 1)(a - 1)$. But $(p - 1)(a - 1) = (n - 1)(a - 1) = na - a^2$. It follows that every line belonging to $\mathscr{P} \smallsetminus \{L\}$, and hence every line belonging to $\mathscr{P}$, consists of exactly a points.

Now we are ready to prove that $\mathscr{L} = \mathscr{P}$. By the definition, every line containing a point of L belongs to $\mathscr{P}$. Let $y \in E \smallsetminus L$. For every $x \in L$ there is a unique line containing $\{x, y\}$. These lines are pairwise distinct, intersect only at y, and belong to $\mathscr{P}$. Moreover, every line containing y and belonging to $\mathscr{P}$ is equal to one of these a lines. Each of these lines contains $a - 1$ points different form y. It follows that the total number of points on these lines is equal to $a(a - 1) + 1$, i.e. to the number n of points in E. Therefore for every point $z \neq y$ there is a line belonging to $\mathscr{P}$ and containing $\{z, y\}$. Since there is only one line containing any two given points, it follows that all lines containing a point $y \in E \smallsetminus L$ belong to $\mathscr{P}$. It follows that $\mathscr{L} = \mathscr{P}$ and every point in $E \smallsetminus L$ belongs to exactly a lines. In view of the previous paragraph, $\mathscr{L} = \mathscr{P}$ implies that every line consists of exactly a points.

By the previous paragraph $k_y = a$ if $y \in E \smallsetminus L$. If $y \in L$, then by (6.1)

$$k_y \geqslant 1 + \frac{n - a}{b - 1} = 1 + \frac{a(a - 1) + 1 - a}{a - 1} = a.$$

If $k_y > a$, then the arguments of the previous paragraph show that $n > a(a - 1) + 1$, contrary to $n = a(a - 1) + 1$. The contradiction shows that $k_y = a$ also for $y \in L$. It follows that $(E, \mathscr{L})$ is a projective plane. This completes the proof of Hanani's theorem.

Deducing the de Bruijn–Erdös theorem. Suppose that $m \leqslant n$. Obviously, $p \leqslant m$ and hence $p \leqslant n$. By Hanani's theorem this implies that $p = n$ and $(E, \mathscr{L})$ is either a near-pencil, or a projective plane. Since $p \leqslant m \leqslant n$ and $p = n$, it follows that $m = n$.

Remarks. In contrast with [dB-E] and many papers written much later, Hanani's proof of his version of the de Bruijn–Erdös theorem in [H2] is quite modern. The points and lines are not enumerated; in fact, there are no subscripts at all. But when he turns to the 3-dimensional case, he returns to the tradition of enumerating almost everything in sight …

Also, in contrast with almost every other proof, Hanani's proof does not use the de Bruijn–Erdös inequalities, at least not directly. But the proof of the fact that $\mathscr{P} = \mathscr{L}$ includes a proof of the de Bruijn–Erdös inequalities $s_L \leqslant k_y$ for $y \notin L$.

7. A simpler proof of Hanani's theorem

This proof follows the outline of the Hanani's one, but brings into the play the smallest number k_u among all k_z. Also, "the second largest" line is chosen not among all lines, but among the lines containing u. This allows to avoid Hanani's Lemma and to replace "miraculous" estimates by rather straightforward ones. The proof was partially inspired by V. Napolitano [N]. If one is interested only in the de Bruijn–Erdös theorem, it can be simplified even further.

Suppose that $n \geqslant p$. Following de Bruijn–Erdös [dB-E], let us consider a point u such that k_u is the smallest number among all numbers k_z. Let $a = s_L$ and $k = k_u$. There are two cases to consider: the case when $k \geqslant a$ and the case when $k < a$. The arguments in both cases are similar and can be unified, but the first case is simpler and we will deal with it first.

The case $k \geqslant a$. Every point is contained in one of the k lines containing u, and each of these lines contains $\leqslant a - 1$ points in addition to u. Therefore the total number of points

$$(7.1) \qquad n \leqslant 1 + k(a - 1).$$

For every point $x \in L$ there are $\geqslant k - 1$ lines containing x and different from L. All these lines belong to $\mathscr{P}$ and are pairwise distinct. Therefore

$$(7.2) \qquad p \geqslant 1 + a(k - 1).$$

If $n \geqslant p$, then the inequalities (7.1) and (7.2) imply that

$$1 + k(a - 1) \geqslant 1 + a(k - 1)$$

and hence $a \geqslant k$. Together with $k \geqslant a$ this implies that $a = k$ and the inequalities (7.1) and (7.2) are actually equalities. It follows that $n = p = 1 + a(a - 1)$, every line containing u consists of exactly a points, and every point belonging to L is contained in exactly k lines. In other terms, $k_y = k$ if $y \in L$. In particular, every point of L can be taken as u and hence every line intersecting L consists of exactly a points. In other terms, $s_l = a$ if $l \in \mathscr{P}$.

Let $y \in E \smallsetminus L$. Then there are a lines containing y and belonging to $\mathscr{P}$, and together they contain $1 + a(a - 1) = n$ points. It follows that for every point $y' \neq y$ there is a line belonging to $\mathscr{P}$ and containing $\{y, y'\}$. Since there is only one line containing $\{y, y'\}$, this implies that $\mathscr{L} = \mathscr{P}$. This implies that $s_l = a$ for all $l \in \mathscr{L}$ and $k_y = a$ for all $y \in E \smallsetminus L$. Since we already proved that $k_y = k = a$ for all $y \in L$, we see that $k_y = a$ for all points y. It follows that $(E, \mathscr{L})$ is a projective plane.

The case $k < a$. By the de Bruijn–Erdös inequalities in this case $u \in L$. Let M be a line containing u and such that s_M is the largest number among all numbers s_l for lines l con-

taining u and different from L. Let $a' = s_M$. Then $a \geqslant a'$. The strategy is to use the fact that $u \in L$ to refine the inequalities (7.1) and (7.2) by using a'.

Every point is contained either in L or in one of the other $k - 1$ lines containing u. Each of these lines contains $\leqslant a' - 1$ points in addition to u. Therefore the total number of points

(7.3) $\quad n \leqslant a + (k - 1)(a' - 1)$.

There are k lines containing u, and for every point $x \in L$ and different from u there are $k_x - 1$ of lines containing x and different from L. All these lines belong to $\mathscr{P}$ and are pairwise distinct. If $x \in L$ and $x \neq u$, then $x \notin M$ and hence $k_x \geqslant s_M = a'$. It follows that

(7.4) $\quad p \geqslant k + (a - 1)(a' - 1)$.

The inequalities (7.3) and (7.4) together with the assumption $n \geqslant p$ imply that

$$a + (k - 1)(a' - 1) \; \geqslant \; k + (a - 1)(a' - 1).$$

By simplifying this inequality we see that $\quad a + k(a' - 1) \geqslant k + a(a' - 1) \quad$ and hence

$$k(a' - 2) \; \geqslant \; a(a' - 2).$$

Since a' is the number of points in a line, $a' \geqslant 2$. It follows that either $k \geqslant a$, or $a' = 2$. But $k \geqslant a$ contradicts to the assumption $k < a$, and hence $a' = 2$.

The equality $a' = 2$ means that M consists of 2 points. By the choice of M, this implies that every line containing u and different from L consists of 2 points. Since L and these other lines contain all points and pairwise intersect only in u, it follows that $n = a + k - 1$.

One of the points of M is u. Let z be the other point. Then $z \notin L$ because M $\neq$ L, and hence there are a lines containing z and belonging to $\mathscr{P}$. Among these lines only M contains u. There are also $k - 1$ lines containing u and not equal to M, and all of them belong to $\mathscr{P}$. It follows that $p \geqslant k + a - 1 = n$. Since $n \geqslant p$, this implies that $p = n$ and every line belonging to $\mathscr{P}$ contains either u or z.

Suppose that there is a line l containing u and different from L, M. Let $y \in l$ and $y \neq u$. Then $y \notin L$ and hence there are a lines containing y and belonging to $\mathscr{P}$. Among these lines only one contains u. By the previous paragraph, the other $a - 1$ lines contain z. Since there is only one line containing $\{y, z\}$, it follows that $a - 1 \leqslant 1$ and hence $a = 2$. Since $k < a$, this implies that $k \leqslant 1$ contrary to the fact that $k_x \geqslant 2$ for all x. The contradiction shows that only the lines L, M contain u.

It follows that $E = L \cup M$ and hence z is the only point not belonging to L. In turn, this implies that the set of lines $\mathscr{L}$ consists of L and the lines of the form $\{x, z\}$, where $x \in L$. Therefore $\mathscr{L} = \mathscr{P}$ and $(E, \mathscr{L})$ is a near-pencil.

8. All the de Bruijn–Erdös inequalities

The Basterfield–Kelly–Conway argument. Suppose that $m < n$. Then

$$(n - s_l)/(m - s_l) > n/m$$

for every $l \in \mathscr{L}$. If $z \notin l$, then $s_l \leqslant k_z$ and hence $m - k_z \leqslant m - s_l$. It follows that

$$n = \sum_{z \in E} \frac{m - k_z}{m - k_z} = \sum_{l \not\ni z} \frac{1}{m - k_z} \geqslant \sum_{l \not\ni z} \frac{1}{m - s_l} = \sum_{l \in \mathscr{L}} \frac{n - s_l}{m - s_l} > \sum_{l \in \mathscr{L}} \frac{n}{m} = n.$$

The contradiction leads to the conclusion that $m \geqslant n$.

This argument is the main part of the proof of Theorem 2.1 (dealing with a more general situation) of the paper [BK] by J.G. Basterfield and L.M. Kelly. Basterfield and Kelly [BK] wrote that they are *"indebted to J. Conway for the simplicity of the present formulation of the proof of Theorem 2.1."* By some reason this acknowledgment led to attributing this argument to J. Conway alone even by some authors referring directly to [BK]. By replacing the strict inequalities $<$ by the non-strict ones $\leqslant$, one can use this argument also to deal with the case $m = n$ along the lines of Sections 2–4. This observation is apparently due to P. de Witte [dW].

This is a sharp-witted, but also the most obscure and puzzling proof. It appears as a rabbit from a hat without any context or explanations and tells nothing about why the theorem is true. In the rest of this section I will explain a natural line of thinking which leads to such a proof. There is no evidence suggesting that it was discovered in this way, but it could have been.

Summing the de Bruijn–Erdös inequalities. Summing de Bruijn–Erdös inequalities and then comparing the result with (2.1) is the key step of both the de Bruijn–Erdös proof and the proof from Section 2. A natural idea is to use all de Bruijn–Erdös inequalities on an equal footing. One way to do this is to use systems of distinct representatives as in Section 4.

One may hope for a proof using all de Bruijn–Erdös inequalities in a way closer to the proof of inequalities (2.3) and (2.4) in Section 2 than to the cyclic order argument of de Bruijn–Erdös. Let us sum the inequalities $s_l \leqslant k_z$ over all pairs $(l, z) \in \mathscr{L} \times E$ such that $z \notin l$. Every s_l appears $n - s_l$ times in the left hand side of these inequalities, and every k_z appears $m - k_z$ times in the right hand side. Therefore, taking the sum results in the inequality

$$\sum_{l \in \mathscr{L}} s_l (n - s_l) \leqslant \sum_{z \in E} k_z (m - k_z).$$

This inequality does not lead to a proof of the desired sort, but it admits a natural generalization. Let F be an increasing function. Instead of $s_l \leqslant k_z$, one can sum the inequalities $F(s_l) \leqslant F(k_z)$. In fact, there is no need to apply the same function to s_l and k_z.

Let F, G be a pair of functions such that $s \leqslant k$ implies $F(s) \leqslant G(k)$. Taking the sum of the inequalities $F(s_l) \leqslant G(k_z)$ over all pairs l, z such that $z \notin l$ results in the inequality

$$\sum_{l \in \mathscr{L}} F(s_l)(n - s_l) \quad \leqslant \quad \sum_{z \in E} G(k_z)(m - k_z).$$

It remains to realize that the functions F, G may depend on the numbers m, n and that one can get rid of the factors $n - s_l$ and $m - k_z$ by dividing by these factors.

A proof of the de Bruijn–Erdős-Hanani theorem. As usual, we may assume that $m \leqslant n$. Let

$$F(s) = \frac{s}{n - s} \quad \text{and} \quad G(k) = \frac{k}{m - k}.$$

Then $s \leqslant k$ implies $F(s) \leqslant G(k)$. Indeed, the latter inequality is equivalent to the inequality $s(m - k) \leqslant k(n - s)$, and hence to the inequality $sm \leqslant kn$, which is obviously true if $s \leqslant k$ and $m \leqslant n$. By summing the inequalities $F(s_l) \leqslant G(k_z)$ over all pairs $(l, z) \in \mathscr{L} \times E$ such that $z \notin l$ we get the inequality

$$(8.1) \qquad \sum_{l \in \mathscr{L}} \frac{s_l}{n - s_l}(n - s_l) \quad \leqslant \quad \sum_{z \in E} \frac{k_z}{m - k_z}(m - k_z),$$

which is obviously equivalent to

$$\sum_{l \in \mathscr{L}} s_l \quad \leqslant \quad \sum_{z \in E} k_z.$$

In view of (2.1) the sides of the latter inequality are actually equal, and hence the sides of the inequality (8.1) are also equal. Since the inequality (8.1) is obtained by summing inequalities $F(s_l) \leqslant G(k_z)$, it follows that if $(l, z) \in \mathscr{L} \times E$ and $z \notin l$, then

$$\frac{s_l}{n - s_l} = \frac{k_z}{m - k_z}.$$

and hence $s_l m = k_z n$. Since $m \leqslant n$ and $s_l \leqslant k_z$, the last equality implies that $m = n$ and $s_l = k_z$. In particular, $z \notin l$ implies that $s_l = k_z$ and hence every line containing z intersects l. It remains to repeat the last paragraph of Section 4.

This proof has the advantage of explicitly using the equality (2.1). The Basterfield–Kelley–Conway argument implicitly uses double sums and a change of the order of summation. This change of the order of summation is a stronger tool than the double counting proving (2.1).

A version of this proof. Suppose that $m \leqslant n$. One can take as F, G the functions

$$F(s) = \frac{n}{n - s} \quad \text{and} \quad G(k) = \frac{m}{m - k}.$$

23

They can be obtained by adding 1 to the previous choice of the functions F, G. Therefore $s \leqslant k$ again implies $F(s) \leqslant G(k)$. This can be also verified in the same way as before. By summing the inequalities $F(s_l) \leqslant G(k_z)$ over all pairs l, z such that $z \notin l$ we get

$$(8.2) \qquad \sum_{l \in \mathscr{L}} \frac{n}{n - s_l}\,(n - s_l) \;\; \leqslant \;\; \sum_{z \in \mathrm{E}} \frac{m}{m - k_z}\,(m - k_z),$$

which is obviously equivalent to $mn \leqslant nm$. But the sides of the latter inequality are equal. It follows that if $(l, z) \in \mathscr{L} \times \mathrm{E}$ and $z \notin l$, then

$$\frac{n}{n - s_l} \;=\; \frac{m}{m - k_z}$$

and hence $s_l m = k_z n$. The rest of the proof is exactly the same as above.

Dividing everything in this version of the proof by m, which amounts to taking

$$F(s) \;=\; \frac{n}{m(n - s)} \qquad \text{and} \qquad G(k) \;=\; \frac{1}{m - k},$$

and omitting the explanations turns this version into the Basterfield–Kelly–Conway argument.

References

[BF]   L. Babai, P. Frankl, *Linear algebra methods in combinatorics*, Preliminary Version 2 (September 1992), Department of Computer Science, The University of Chicago.

[BK]   J.G. Basterfield, L.M. Kelly, A characterization of sets of n points which determine n hyperplanes, *Proceedings of the Cambridge Philosophical Society,* V. 64 (1968), pp. 585–588.

[Bo]   R.C. Bose, A note on Fisher's inequality for balanced incomplete block designs, *Ann. Math. Statistics,* V. 20 (1949), pp. 619–620.

[B-65] N. Bourbaki, *Théorie des ensembles, Chapitre III. Ensembles ordonnés – Cardinaux – Nombres entiers,* Hermann, Paris, 1956. Russian translation, "Mir" Publishers, Moscow, 1965.

[B-70] N. Bourbaki, Problem M5, *"Kvant"* collection of problems, *Kvant,* 1970, No. 1, p. 53. "Nauka" Publishing House, Moscow, 1970.

[B-04] N. Bourbaki, *Theory of sets, Chapter III. Ordered Sets, Cardinals, Integers,* English translation of the original 1970 edition. © N. Bourbaki et Springer-Verlag Berlin Heidelberg 2004.

[B-06] N. Bourbaki, *Théorie des ensembles, Chapitre III. Ensembles ordonnés, cardinaux, nombres entiers,* Réimpression inchangée de l'édition originale de 1970, Hermann, Paris, 1970. © N. Bourbaki and Springer-Verlag Berlin Heidelberg 2006.

[dB-E] N.G. de Bruijn, P. Erdös, On a combinatorial problem, *Nederl. Akad. Wetensch., Proc.* V. 51, (1948), 1277–1279 = *Indagationes Math.* V. 10 (1948), 421–423.

[dW] P. de Witte, Combinatorial properties of finite linear spaces II, *Bulletin de la Société Mathématique de Belgique,* T. 25 (1975), pp. 115–155.

[HP] D.R. Hughes, F.C. Piper, *Design theory,* Cambridge University Press, 1988, 240 pp.

[H1] H. Hanani, On the number of straight lines determined by n points (Hebrew, English summary), *Riveon Lematematika,* V. 5 (1951), pp. 10–11.

[H2] H. Hanani, On the number of lines and planes determined by d points, *Scientific Publications, Technion (Israel Institute of Technology, Haifa),* V. 6 (1954-55), pp. 58–63

[I1] N.V. Ivanov, Affine planes, ternary rings, and examples of non-Desarguesian planes, https://arxiv.org/abs/1604.04945, 2016, 25 pp.

[I2] N.V. Ivanov, The de Bruijn-Erdös theorem in incidence geometry via Ph. Hall's marriage theorem, https://arxiv.org/abs/1704.04343, 2017, 4 pp.

[M] Th. Motzkin, The lines and planes connecting the points of a finite set, *Transactions of the AMS,* V. 79 (1951), pp. 451–464.

[N] V. Napolitano, Una dimostrazione unificata dei teoremi di Bridges e de Witte e del teorema fondamentale degli spazi lineari finiti, *Rendiconti del Circolo Matematico di Palermo. Serie II,* T. 50 (2001), pp. 443–454.

[P] F. Petrov, Personal communication, E-mail, April 5, 2017.

[R] H.J. Ryser, An extension of a theorem of de Bruijn and Erdös on combinatorial designs, *Journal of Algebra,* V. 10 (1968), pp. 246–261.

[T] A. Toom, Zh. Rabbot, V. Gutenmakher, Solution of the problem M5, *Kvant,* 1970, No. 8, pp. 49–53. "Nauka" Publishing House, Moscow, 1970.

[W] Ch. Weibel, Survey of Non-Desarguesian planes, *Notices of the AMS,* V. 54, No. 10 (2007), pp. 1294–1303.

May 12, 2017

https://nikolaivivanov.com

E-mail: nikolai.v.ivanov@icloud.com