On *On a combinatorial problem* [1]

Misha Lavrov

September 13, 2013

Abstract

The combinatorial version of the theorem [1] of de Bruijn and Erdős is presented in a more accessible way, since it turns out Wikipedia's article [2] only provides the proof in the case that the points are a subset of the plane.

Definition. *An* incidence geometry *consists of a set of points and a collection of lines, which are subsets of the set of points. If a line ℓ contains a point P, we say P lies on ℓ, or ℓ passes through P. The points and lines must satisfy the following axioms:*

`[I1]` *Through any two points, there is exactly one line.*

`[I2]` *Each line contains at least two points.*

`[I3]` *There are three points that do not all lie on one line.*

Theorem (de Bruijn, Erdős, 1948)**.** *Any incidence geometry contains at least as many lines as points.*

Proof. Suppose our geometry consists of points $P_1, \dots, P_n$ and lines $\ell_1, \dots, \ell_m$. We define $\text{lines}(P_i)$ to be the number of lines through point P_i, and $\text{points}(\ell_j)$ to be the number of points on line ℓ_j.

Lemma 1. $\text{lines}(P_1) + \text{lines}(P_2) \cdots + \text{lines}(P_n) = \text{points}(\ell_1) + \text{points}(\ell_2) + \cdots + \text{points}(\ell_m)$.

Proof. Both sides of the equality we want to prove are counting the same thing in two ways: pairs (P_i, ℓ_j) where point P_i lies on line ℓ_j. This can be counted by going through all points, and counting how many lines they lie on (the left-hand side), or by going through all lines, and counting how many points they contain (the right-hand side). □

Lemma 2. *Suppose point P_i does not lie on line ℓ_j. Then* $\text{lines}(P_i) \geq \text{points}(\ell_j)$.

Proof. For each point on ℓ_j, there is a line through that point and P_i, by `[I1]`. If we go through all points on ℓ_j and look at this line, we never see the same line twice: the only line that passes through two points of ℓ_j is ℓ_j itself, also by `[I1]`. Therefore we have found as many lines through P_i as there are points on ℓ_j. □

We are now ready to proceed to the proof of the theorem. We're free to choose the order in which we label our lines and points, so we make the following choices:

- Choose P_n to be the point with the fewest lines through it – or one such point, if there is a tie. Let $k = \text{lines}(P_n)$.
- Choose $\ell_1, \ldots, \ell_k$ to be the k lines through P_n.
- For $1 \le j \le k$, there is at least one point on ℓ_j other than P_n, by `[I2]`; let P_j be one such point.

To avoid really weird cases, we need to rule out the possibilities $k = 0$ and $k = 1$.

1. By `[I3]`, there are at least three points; so there is a point P_i other than P_n. By `[I1]`, there is a line through P_i and P_n, so $k \ge 1$.
2. If we had $k = 1$, then all points would lie on ℓ_1, which also contradicts `[I3]`.

The point P_1 does not lie on ℓ_2; or else there would be two lines (ℓ_1 and ℓ_2) through P_1 and P_n, contradicting `[I1]`. By applying Lemma 2, we get $\text{lines}(P_1) \ge \text{points}(\ell_2)$. For $1 \le i \le k-1$, we can apply the same argument to get $\text{lines}(P_i) \ge \text{points}(\ell_{i+1})$. Finally, this argument gives us $\text{lines}(P_k) \ge \text{points}(\ell_1)$.

If we add all these inequalities, we get:

$$\text{lines}(P_1) + \text{lines}(P_2) + \cdots + \text{lines}(P_k) \ge \text{points}(\ell_1) + \text{points}(\ell_2) + \text{points}(\ell_k). \quad (1)$$

Note that this argument would have failed if $k = 0$ or $k = 1$: it can never get off the ground, because it doesn't even make sense to talk about P_1 and ℓ_2. Proving that $k \ge 2$ is the only time we use `[I3]`: it allows us to rule out the case of a single line through n points, which would otherwise be a counterexample.

Subtract inequality (1) from the equality in Lemma 1, and we get:

$$\text{lines}(P_{k+1}) + \text{lines}(P_{k+2}) + \cdots + \text{lines}(P_n) \le \text{points}(\ell_{k+1}) + \text{points}(\ell_{k+2}) + \cdots + \text{points}(\ell_m). \quad (2)$$

Due to the way we chose P_n, $\text{lines}(P_i) \ge \text{lines}(P_n) = k$ for all i; in particular, on the left-hand side of (2), every term we're adding is at least k. So the left-hand side of (2) is at least $k(n-k)$.

By Lemma 2, for each $j > k$, $\text{lines}(P_n) \ge \text{points}(\ell_j)$, since P_n does not lie on any of $\ell_{k+1}, \ldots, \ell_m$. This means that on the right-hand side of (2), every term we're adding is at most k. So the right-hand side of (2) is at most $k(m-k)$.

Therefore we have $k(n-k) \le k(m-k)$, which we can simplify to $n \le m$. □

References

[1] N.G. de Bruijn and P. Erdős. On a combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen, Proceedings*, 51:1277–1279, 1948.

[2] Wikipedia. De Bruijn–Erdős theorem (incidence geometry) — Wikipedia, the Free Encyclopedia, 2013. [Online; accessed 13-September-2013].