Formalising Twenty-First Century Mathematics

Lawrence Paulson, Computer Laboratory, University of Cambridge

LMS / BCS-FACS Seminar, 15 January 2024

Prologue

Mathematics is getting formalised!

Cap set problem

Diagonal Ramsey

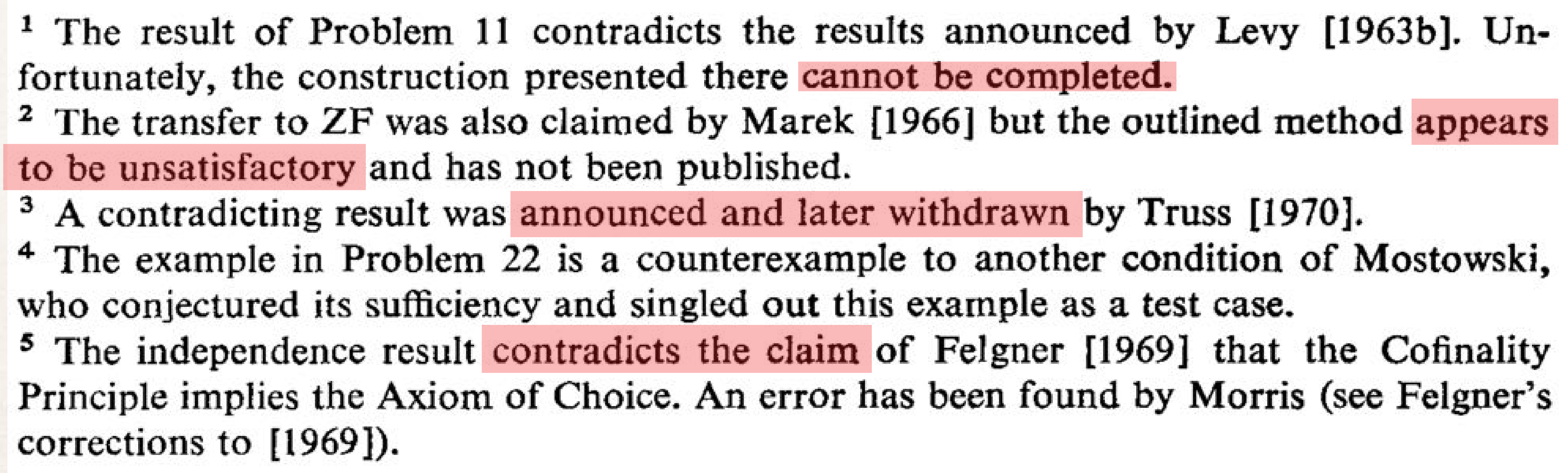
Liquid tensor experiment

Polynomial Freiman–Ruzsa

Brand new results, often formally checked *before* the referees!

But why? Here’s one reason

The footnotes on a **single page** (118)
of Jech's *The Axiom of Choice*



[1] The result of Problem 11 contradicts the results announced by Levy [1963b]. Unfortunately, the construction presented there cannot be completed.

[2] The transfer to ZF was also claimed by Marek [1966] but the outlined method appears to be unsatisfactory and has not been published.

[3] A contradicting result was announced and later withdrawn by Truss [1970].

[4] The example in Problem 22 is a counterexample to another condition of Mostowski, who conjectured its sufficiency and singled out this example as a test case.

[5] The independence result contradicts the claim of Felgner [1969] that the Cofinality Principle implies the Axiom of Choice. An error has been found by Morris (see Felgner’s corrections to [1969]).

Using what? A *proof assistant*

- A *language* for abstract concepts and assertions
- … and another language for expressing proofs.
- An *interface* for doing proofs interactively
- … and for managing a large formal development
- *Simplification* and other proof automation
- *Libraries* of undergraduate level mathematics

Commonly used formalisms

Higher-order logic, also known as *simple type theory*

Relatively simple, allowing good automation

HOL Light, Isabelle/HOL …

Calculus of constructions or other *dependent type theory*

Stronger than ZF, and supports constructive proof

Agda, Coq, Lean …

Formalisation is not new

- *Euclid*: unifying Greek geometry under an axiomatic system
- *Cauchy, Weierstrass*: removing infinitesimals from analysis
- *Dedekind, Cantor, Frege, Zermelo*: set theory and the axiom of choice
- *Whitehead, Russell, Bourbaki*: formal (or super-rigorous) mathematics
- *de Bruijn*: the AUTOMATH type theory and proof checker; also *Trybulec* and Mizar

Now it's widely accepted that <u>all</u> mathematics is formalisable

But is all maths really formalisable?

As to the question what part of mathematics can be written in AUTOMATH, it should first be remarked that we do not possess a workable definition of the word "mathematics".

Quite often a mathematician jumps from his mathematical language into a kind of metalanguage, obtains results there, and uses these results in his original context. It seems to be very hard to create a single language in which such things can be done without any restriction.

– NG de Bruijn (1968)

Formalising Maths in Isabelle/HOL

2017–23: ALEXANDRIA

(ERC Project GA 742178)

Aim: to support working mathematicians

… by developing tools and libraries

What areas of mathematics can we formalise?

What sorts of proofs can we formalise?

Existing Maths in Isabelle (2017)

- Lots formalised already
- But... was it *sophisticated* enough? *Modern* enough?
- We had to **explore our boundaries**, and compare with *dependent type theories*

Matrix theory, e.g. Perron–Frobenius

Analytic number theory, e.g. Hermite–Lindemann

Homology theory

Measure, integration and probability theory

Complex analysis: residue theorem, prime number theorem

Some warmup formalisations

- *Irrational rapidly convergent series*, formalising a 2002 paper by J. Hančl
- *projective geometry* and *quantum computing*
- counting real and complex roots of polynomials; *Budan–Fourier theorem*

Our focus: recent, sophisticated or potentially problematical material

Another early experiment (2019): algebraically closed fields

Every field admits an algebraically closed extension

(Example: adjoining a root of $x^2 + 1$ to $\mathbb{R}$ to get $\mathbb{C}$)

In general, a *limit* of field extensions

$$K = E_0 \to E_1 \to E_2 \to \cdots \to E_n \to \cdots$$

obtained by adjoining roots. We can form this limit using Zorn's lemma

The work of two summer students, Paulo de Vilhena and Martin Baillon, and the first formalisation of this result in any system.

Taking over a special issue of *Experimental Mathematics*

- *Irrationality and transcendence criteria for infinite series*, incorporating Erdős–Straus and Hančl–Rucki

- *Ordinal partition theory*: delicate constructions by Erdős–Milner and Larson on set-theoretic combinatorics

- *Grothendieck schemes*: answering a challenge by Kevin Buzzard (and completed on the first attempt)

These formed 3 of the 6 papers in the special issue

Upping our ambitions

- extremal graph theory
- additive combinatorics
- combinatorial block designs
- graduate-level number theory
- strict ω-categories

Szemerédi's regularity lemma, and Roth on arithmetic progressions

For every $\epsilon > 0$, there exists a constant M such that every graph has an ϵ-regular partition of its vertex set into at most M parts.

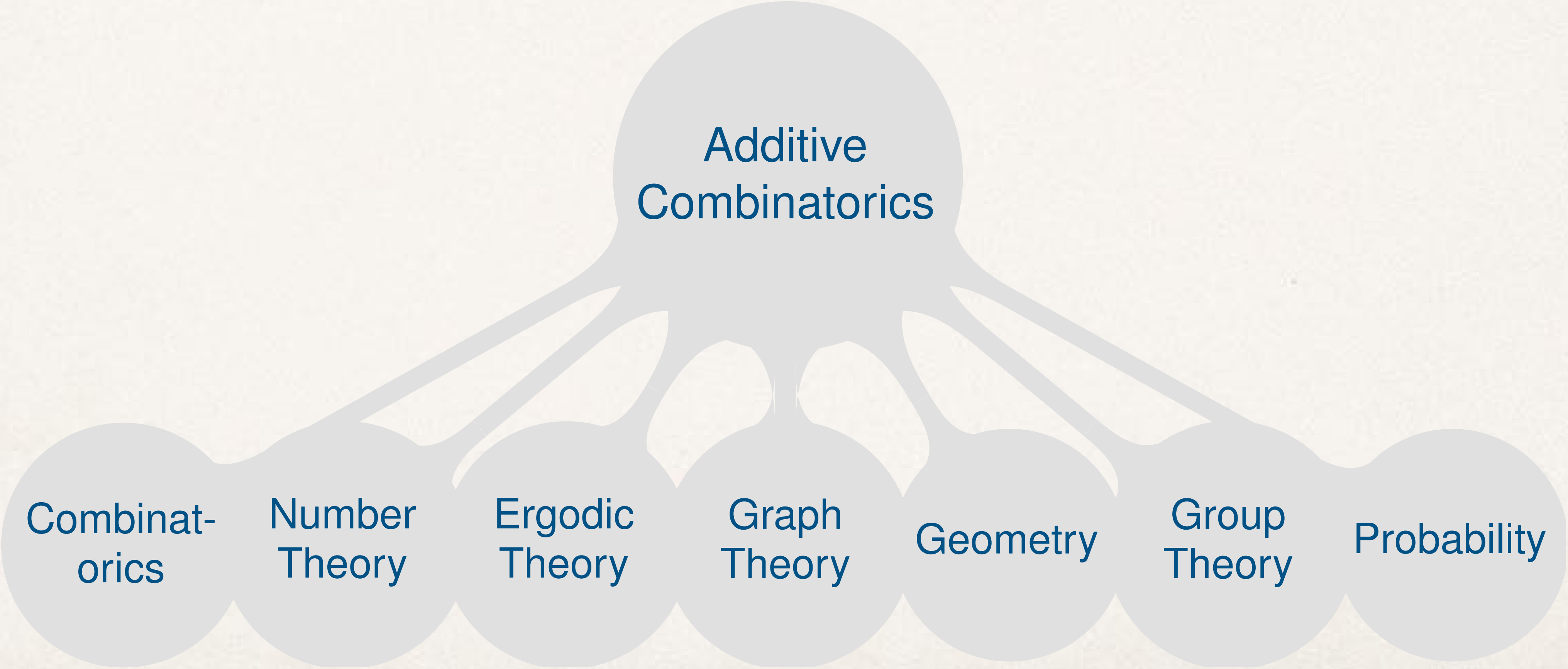
An *ϵ-regular partition* is where the edges between different parts behave "almost randomly"

The key tool in the study of large graphs

Every subset of the integers with positive *upper asymptotic density* contains a 3-term arithmetic progression.

Additive combinatorics

The study of the additive structure of sets, with numerous applications across mathematics



This topic concerns the *sumset* $A + B = \{a + b : a \in A, b \in B\}$

for a given abelian group $(G, +)$

and the *iterated sumset*: the n-fold sum $nA = A + \cdots + A$

Plünnecke–Ruzsa inequality: an upper bound on $mB - nB$

Khovanskii's theorem: $|nA|$ grows like a polynomial for sufficiently large n

Kneser's theorem and the *Cauchy–Davenport theorem: lower bounds for* $|A + B|$

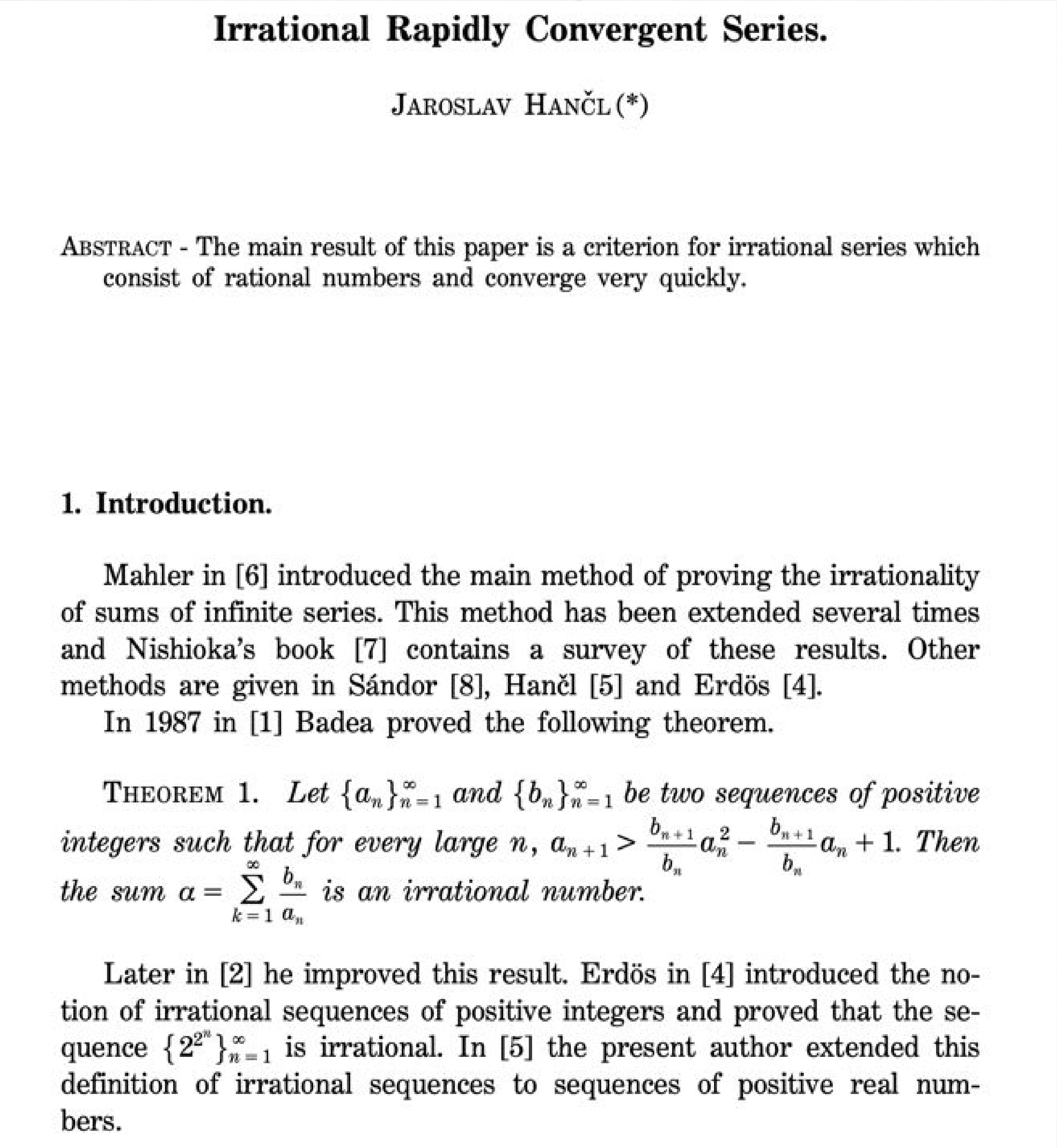
Balog–Szemerédi–Gowers: a deep result bearing on Szemerédi's theorem

Combinatorial structures

- dozens of varieties of block designs, hypergraphs, graphs and the relationships among them
- E.g. *Fisher's inequality* for balanced incomplete block designs
- probabilistic and generating function methods
- advanced techniques using Isabelle's *locales*

PhD work of Chelsea Edmonds

Some Papers We Formalised



Irrational Rapidly Convergent Series.

JAROSLAV HANČL (*)

ABSTRACT - The main result of this paper is a criterion for irrational series which consist of rational numbers and converge very quickly.

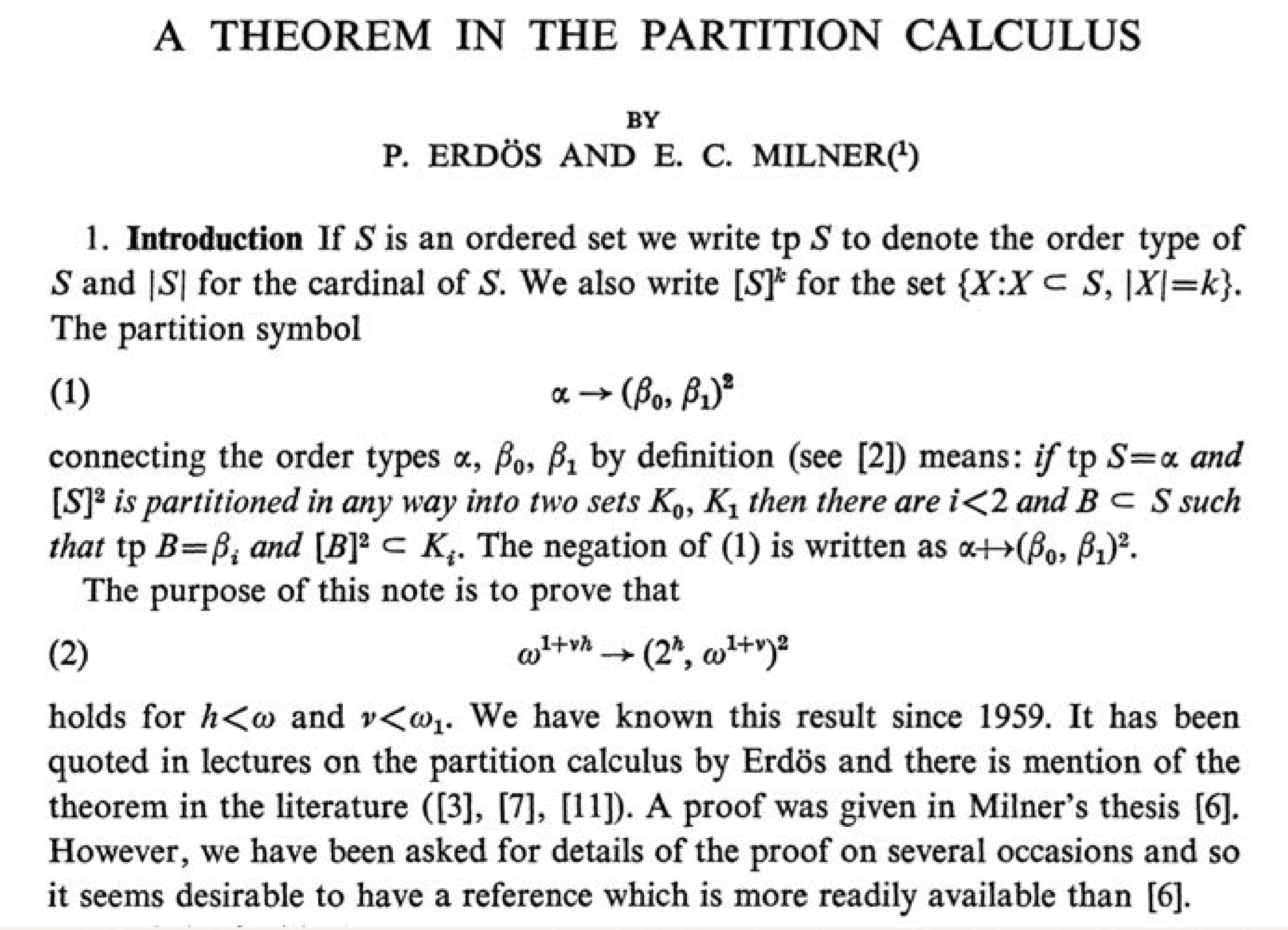
1. Introduction.

Mahler in [6] introduced the main method of proving the irrationality of sums of infinite series. This method has been extended several times and Nishioka's book [7] contains a survey of these results. Other methods are given in Sándor [8], Hančl [5] and Erdös [4].

In 1987 in [1] Badea proved the following theorem.

THEOREM 1. *Let* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *be two sequences of positive integers such that for every large* n, $a_{n+1} > \frac{b_{n+1}}{b_n}a_n^2 - \frac{b_{n+1}}{b_n}a_n + 1$. *Then the sum* $\alpha = \sum_{k=1}^{\infty} \frac{b_n}{a_n}$ *is an irrational number.*

Later in [2] he improved this result. Erdös in [4] introduced the notion of irrational sequences of positive integers and proved that the sequence $\{2^{2^n}\}_{n=1}^{\infty}$ is irrational. In [5] the present author extended this definition of irrational sequences to sequences of positive real numbers.



A THEOREM IN THE PARTITION CALCULUS

BY

P. ERDÖS AND E. C. MILNER[1]

1. **Introduction** If S is an ordered set we write tp S to denote the order type of S and $|S|$ for the cardinal of S. We also write $[S]^k$ for the set $\{X : X \subset S, |X| = k\}$. The partition symbol

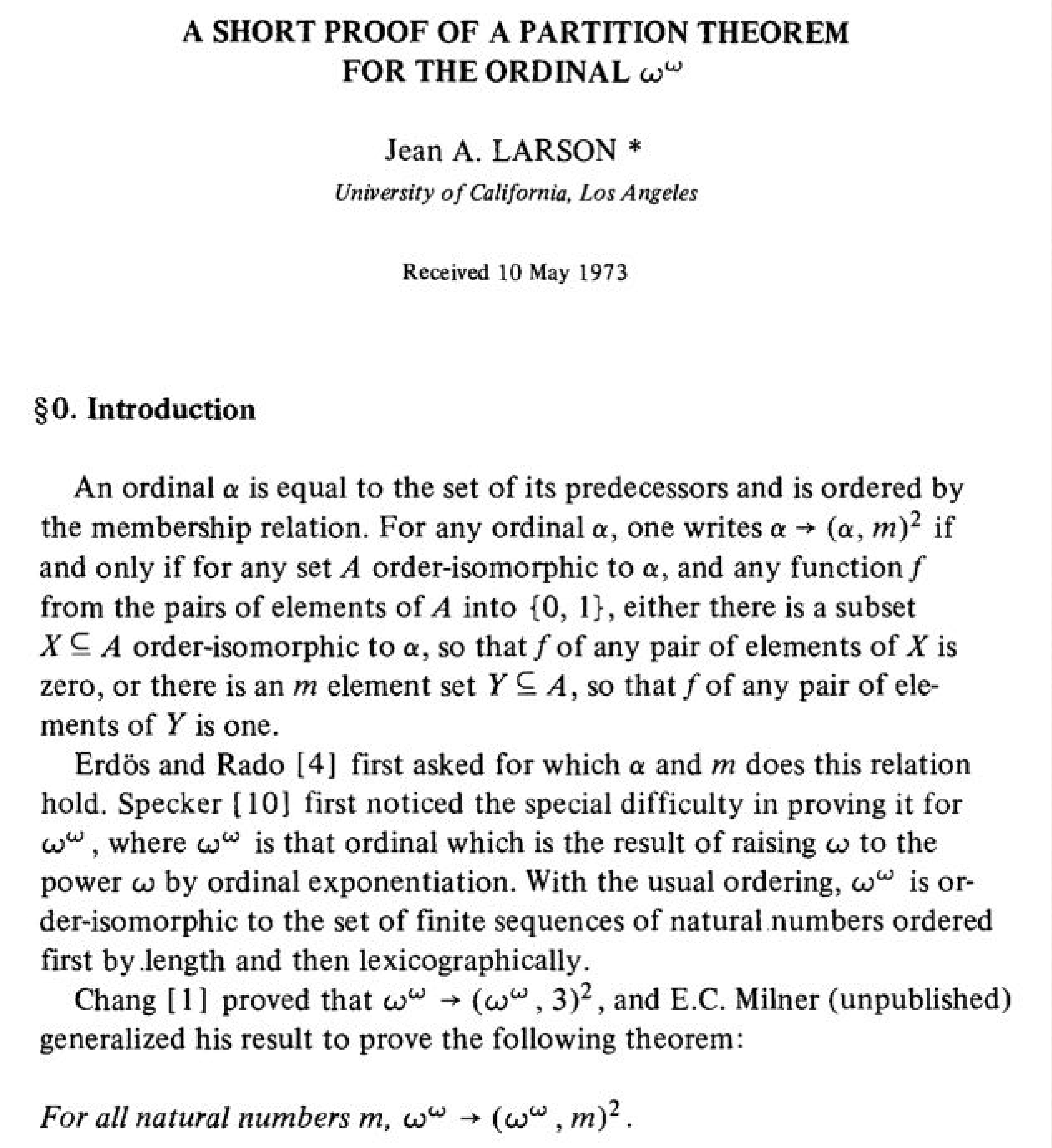
$$\alpha \rightarrow (\beta_0, \beta_1)^2 \tag{1}$$

connecting the order types α, β_0, β_1 by definition (see [2]) means: *if* tp $S = \alpha$ *and* $[S]^2$ *is partitioned in any way into two sets* K_0, K_1 *then there are* $i < 2$ *and* $B \subset S$ *such that* tp $B = \beta_i$ *and* $[B]^2 \subset K_i$. The negation of (1) is written as $\alpha \nrightarrow (\beta_0, \beta_1)^2$.

The purpose of this note is to prove that

$$\omega^{1+\nu h} \rightarrow (2^h, \omega^{1+\nu})^2 \tag{2}$$

holds for $h < \omega$ and $\nu < \omega_1$. We have known this result since 1959. It has been quoted in lectures on the partition calculus by Erdös and there is mention of the theorem in the literature ([3], [7], [11]). A proof was given in Milner's thesis [6]. However, we have been asked for details of the proof on several occasions and so it seems desirable to have a reference which is more readily available than [6].



A SHORT PROOF OF A PARTITION THEOREM FOR THE ORDINAL ω^ω

Jean A. LARSON *

University of California, Los Angeles

Received 10 May 1973

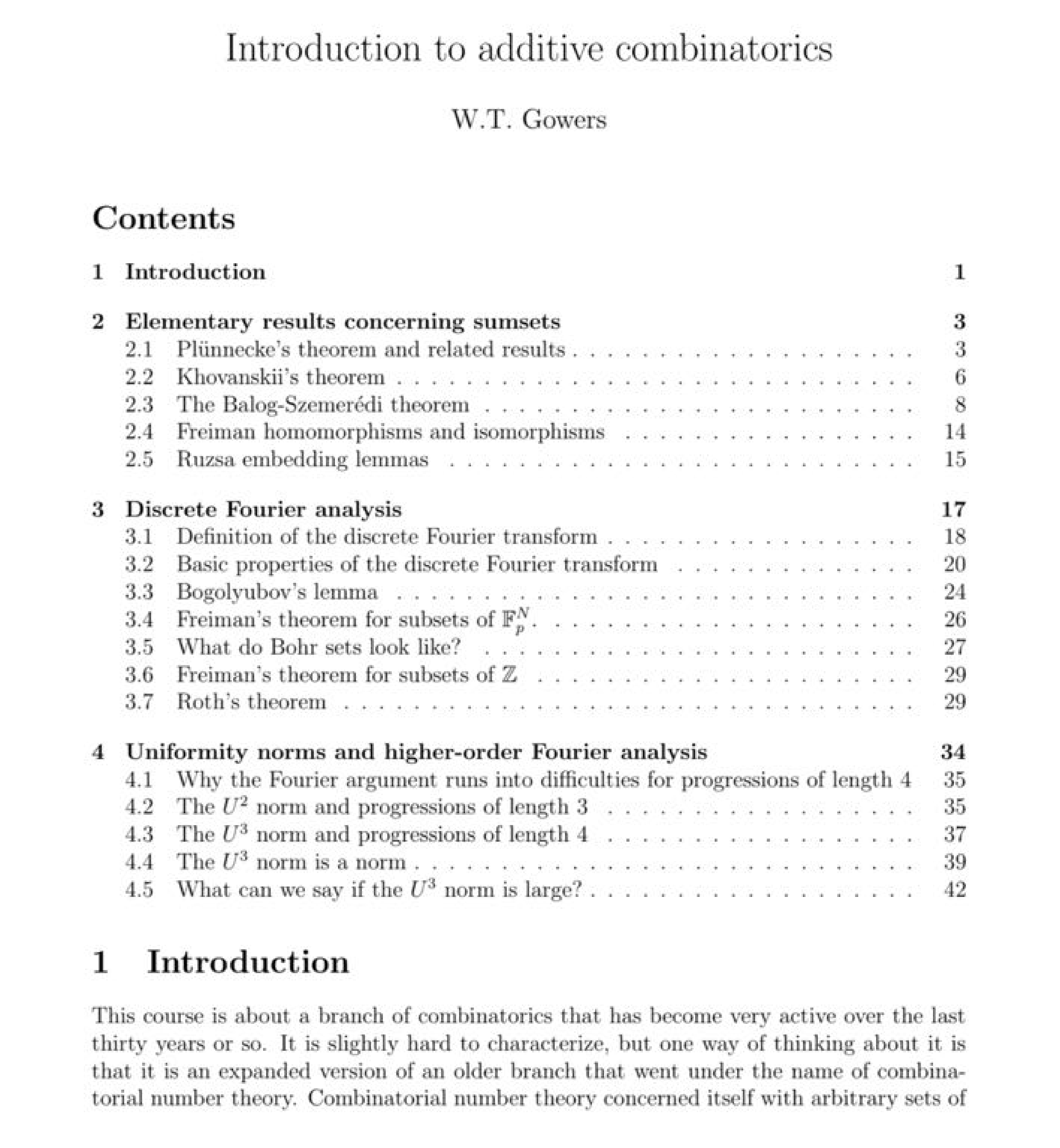
§0. Introduction

An ordinal α is equal to the set of its predecessors and is ordered by the membership relation. For any ordinal α, one writes $\alpha \to (\alpha, m)^2$ if and only if for any set A order-isomorphic to α, and any function f from the pairs of elements of A into $\{0, 1\}$, either there is a subset $X \subseteq A$ order-isomorphic to α, so that f of any pair of elements of X is zero, or there is an m element set $Y \subseteq A$, so that f of any pair of elements of Y is one.

Erdös and Rado [4] first asked for which α and m does this relation hold. Specker [10] first noticed the special difficulty in proving it for ω^ω, where ω^ω is that ordinal which is the result of raising ω to the power ω by ordinal exponentiation. With the usual ordering, ω^ω is order-isomorphic to the set of finite sequences of natural numbers ordered first by length and then lexicographically.

Chang [1] proved that $\omega^\omega \to (\omega^\omega, 3)^2$, and E.C. Milner (unpublished) generalized his result to prove the following theorem:

For all natural numbers m, $\omega^\omega \to (\omega^\omega, m)^2$.



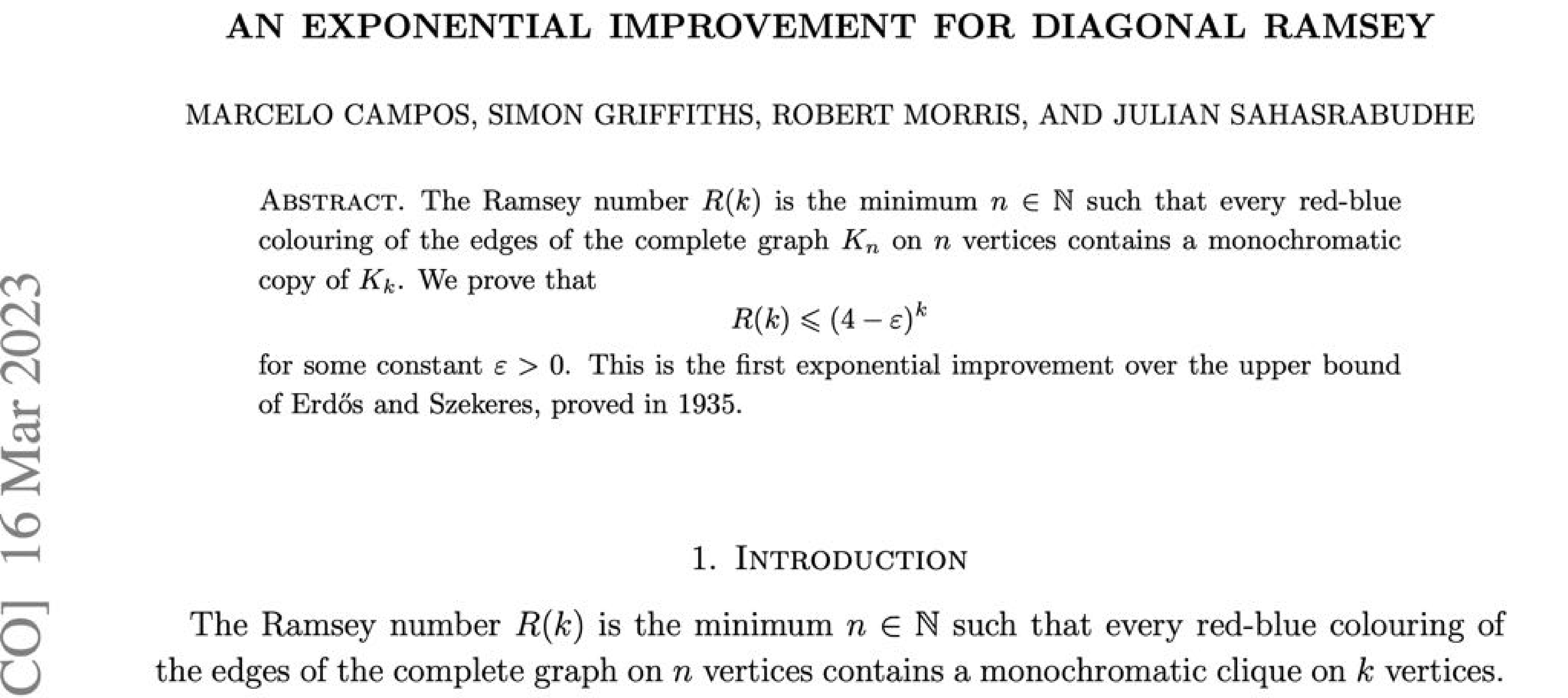
Introduction to additive combinatorics

W.T. Gowers

Contents

1 Introduction

This course is about a branch of combinatorics that has become very active over the last thirty years or so. It is slightly hard to characterize, but one way of thinking about it is that it is an expanded version of an older branch that went under the name of combinatorial number theory. Combinatorial number theory concerned itself with arbitrary sets of



AN EXPONENTIAL IMPROVEMENT FOR DIAGONAL RAMSEY

MARCELO CAMPOS, SIMON GRIFFITHS, ROBERT MORRIS, AND JULIAN SAHASRABUDHE

Abstract. The Ramsey number $R(k)$ is the minimum $n \in \mathbb{N}$ such that every red-blue colouring of the edges of the complete graph K_n on n vertices contains a monochromatic copy of K_k. We prove that

$$R(k) \leqslant (4 - \varepsilon)^k$$

for some constant $\varepsilon > 0$. This is the first exponential improvement over the upper bound of Erdős and Szekeres, proved in 1935.

1. Introduction

The Ramsey number $R(k)$ is the minimum $n \in \mathbb{N}$ such that every red-blue colouring of the edges of the complete graph on n vertices contains a monochromatic clique on k vertices.

What Did ALEXANDRIA Achieve?

no borders between mathematical topics

…and no topics off-limits

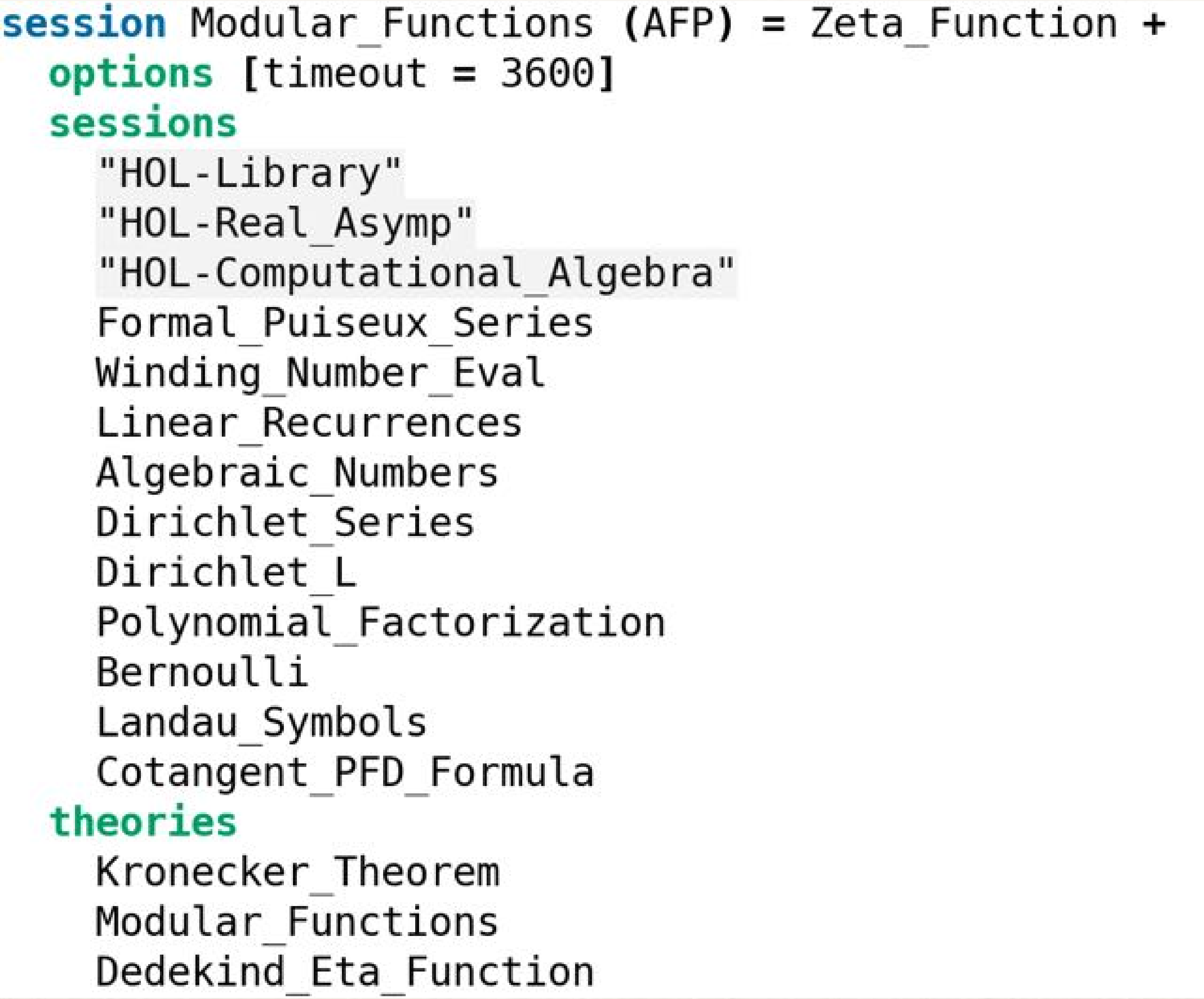
good automation

good performance

legible proofs

sophisticated, modern mathematics

No borders between topics



```
session Modular_Functions (AFP) = Zeta_Function +
  options [timeout = 3600]
  sessions
    "HOL-Library"
    "HOL-Real_Asymp"
    "HOL-Computational_Algebra"
    Formal_Puiseux_Series
    Winding_Number_Eval
    Linear_Recurrences
    Algebraic_Numbers
    Dirichlet_Series
    Dirichlet_L
    Polynomial_Factorization
    Bernoulli
    Landau_Symbols
    Cotangent_PFD_Formula
  theories
    Kronecker_Theorem
    Modular_Functions
    Dedekind_Eta_Function
```

- We combined *probability* with *combinatorics*
- … *transfinite recursion* with *holomorphic functions*
- we are perfectly fine without *dependent types*
- with locales we can handle **multiple inheritance** (“diamonds”)

Performance matters too!

- *14 seconds* for Szemerédi’s regularity lemma
- 15s for Erdős–Straus theorem on irrational series
- 50s for ordinal partitions
- 1:11 for Balog–Szemerédi–Gowers
- 1:04 for Grothendieck schemes
- 1:03 for Roth’s theorem on arithmetic progressions

Run on a 2019 iMac, 3.6 GHz 8-Core Intel Core i9

What we get from Lean fans

You never prove anything hard

… only the work of two Fields medalists (Roth, Gowers), an Abel prize winner (Szemerédi) and Paul Erdős.

You need dependent types

We can get $T(i) = T(j)$ from $i = j$ without worrying about **definitional equality**

Our proofs are nicer

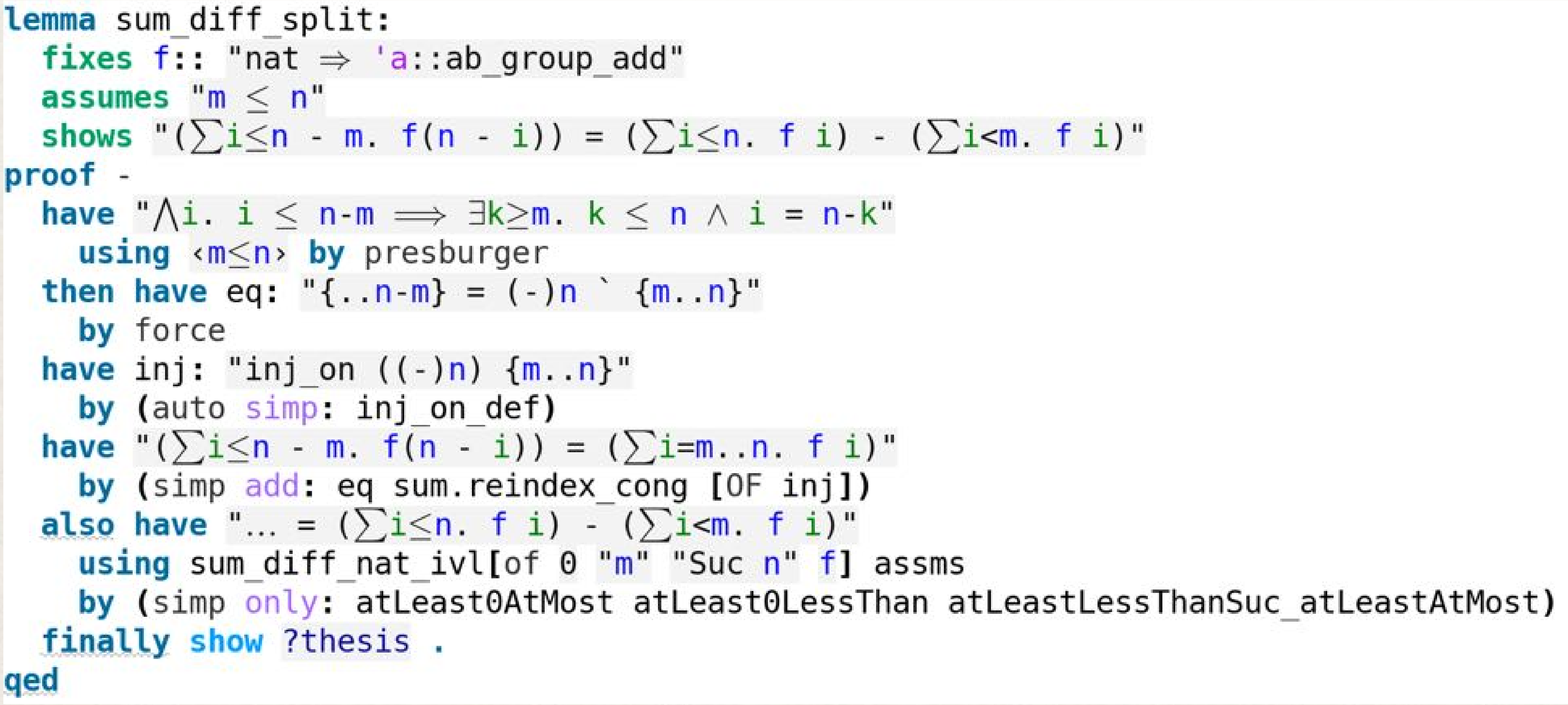
Oh?

On the Legibility of Proofs

Is a proof a proof just because Lean agrees it's one? In some ways, it's as good as the people who convert the proof into inputs for Lean.

– Andrew Granville

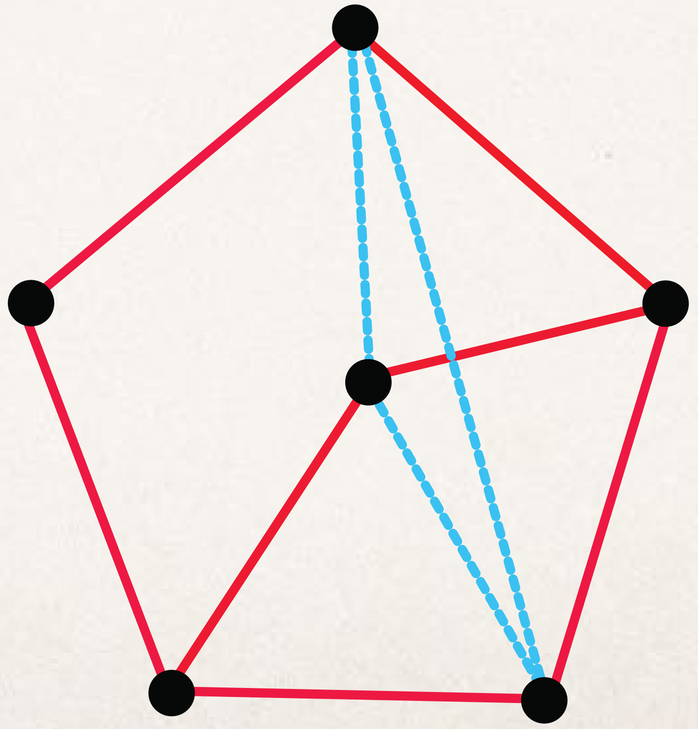
A small summation identity



```
lemma sum_diff_split:
  fixes f:: "nat ⇒ 'a::ab_group_add"
  assumes "m ≤ n"
  shows "(∑i≤n - m. f(n - i)) = (∑i≤n. f i) - (∑i<m. f i)"
proof -
  have "⋀i. i ≤ n-m ⟹ ∃k≥m. k ≤ n ∧ i = n-k"
    using ‹m≤n› by presburger
  then have eq: "{..n-m} = (-)n ` {m..n}"
    by force
  have inj: "inj_on ((-)n) {m..n}"
    by (auto simp: inj_on_def)
  have "(∑i≤n - m. f(n - i)) = (∑i=m..n. f i)"
    by (simp add: eq sum.reindex_cong [OF inj])
  also have "… = (∑i≤n. f i) - (∑i<m. f i)"
    using sum_diff_nat_ivl[of 0 "m" "Suc n" f] assms
    by (simp only: atLeast0AtMost atLeast0LessThan atLeastLessThanSuc_atLeastAtMost)
  finally show ?thesis .
qed
```

Aside: Ramsey's theorem

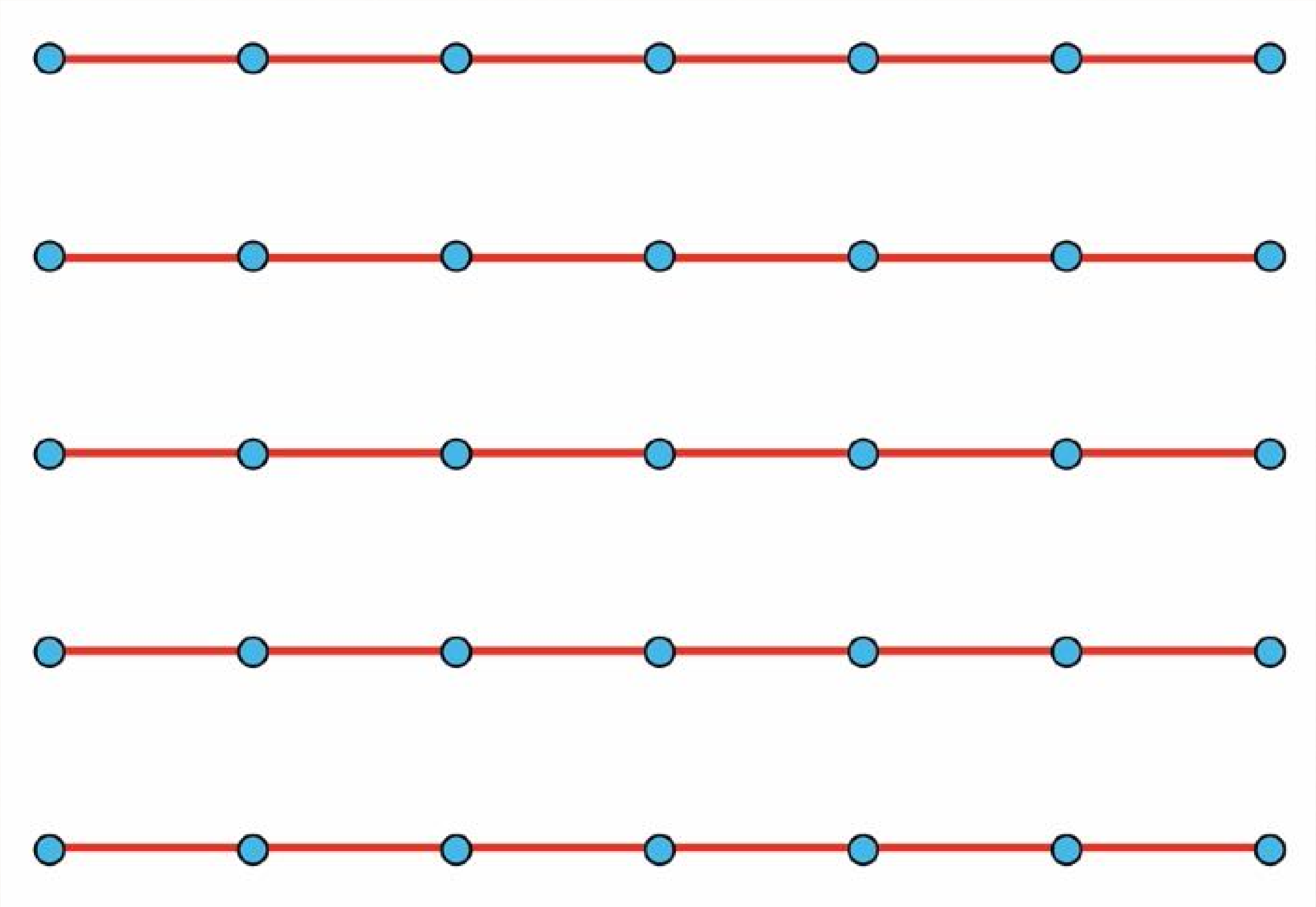
For all m and n there exists a number $R(m, n)$ such that every graph with at least $R(m, n)$ vertices contains a *clique* of size m or an *anti-clique* of size n

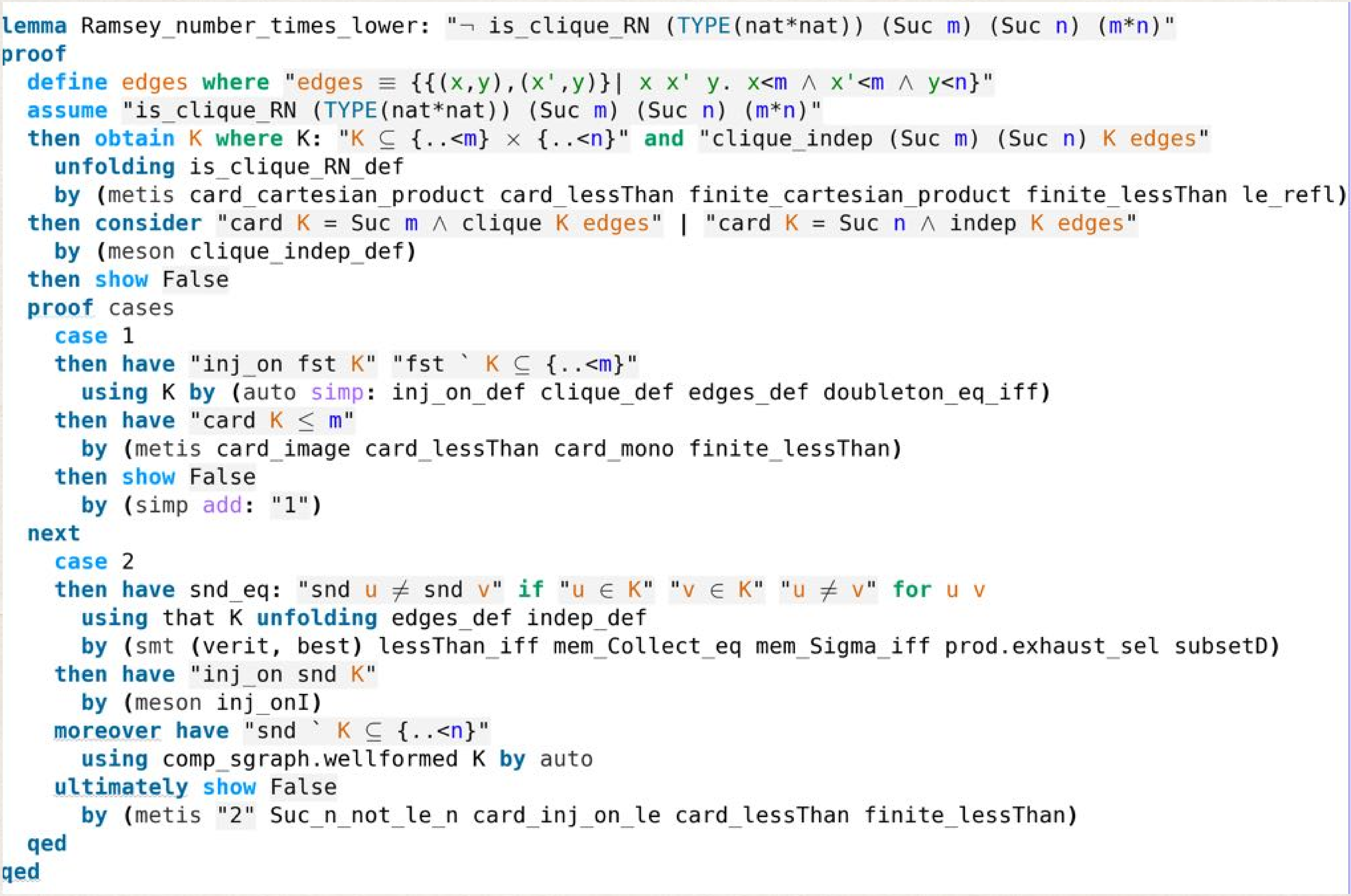


Proving $R(m+1, n+1) > mn$

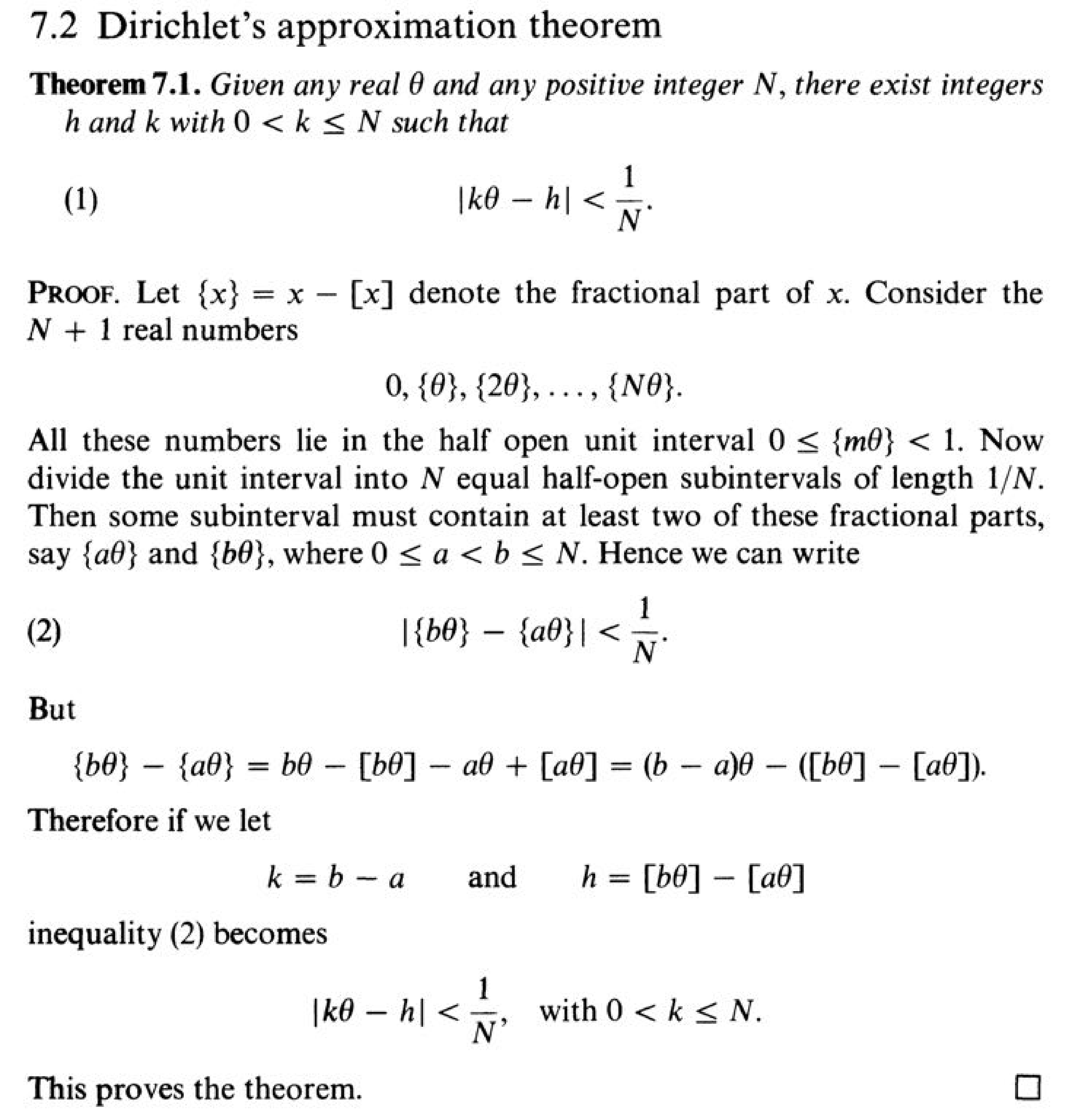
- Construct a graph with $m \times n$ vertices, containing
  - No clique of size $m+1$, and
  - No independent set (anticlique) of size $n+1$
- The vertices are pairs (x, y)
- The edges join every (x, y) with (x', y)

Our $m \times n$ graph, with its edges





```
lemma Ramsey_number_times_lower: "¬ is_clique_RN (TYPE(nat*nat)) (Suc m) (Suc n) (m*n)"
proof
  define edges where "edges ≡ {{(x,y),(x',y)}| x x' y. x<m ∧ x'<m ∧ y<n}"
  assume "is_clique_RN (TYPE(nat*nat)) (Suc m) (Suc n) (m*n)"
  then obtain K where K: "K ⊆ {..<m} × {..<n}" and "clique_indep (Suc m) (Suc n) K edges"
    unfolding is_clique_RN_def
    by (metis card_cartesian_product card_lessThan finite_cartesian_product finite_lessThan le_refl)
  then consider "card K = Suc m ∧ clique K edges" | "card K = Suc n ∧ indep K edges"
    by (meson clique_indep_def)
  then show False
  proof cases
    case 1
    then have "inj_on fst K" "fst ` K ⊆ {..<m}"
      using K by (auto simp: inj_on_def clique_def edges_def doubleton_eq_iff)
    then have "card K ≤ m"
      by (metis card_image card_lessThan card_mono finite_lessThan)
    then show False
      by (simp add: "1")
  next
    case 2
    then have snd_eq: "snd u ≠ snd v" if "u ∈ K" "v ∈ K" "u ≠ v" for u v
      using that K unfolding edges_def indep_def
      by (smt (verit, best) lessThan_iff mem_Collect_eq mem_Sigma_iff prod.exhaust_sel subsetD)
    then have "inj_on snd K"
      by (meson inj_onI)
    moreover have "snd ` K ⊆ {..<n}"
      using comp_sgraph.wellformed K by auto
    ultimately show False
      by (metis "2" Suc_n_not_le_n card_inj_on_le card_lessThan finite_lessThan)
  qed
qed
```



7.2 Dirichlet's approximation theorem

Theorem 7.1. *Given any real θ and any positive integer N, there exist integers h and k with $0 < k \le N$ such that*

$$|k\theta - h| < \frac{1}{N}. \tag{1}$$

PROOF. Let $\{x\} = x - [x]$ denote the fractional part of x. Consider the $N + 1$ real numbers

$$0, \{\theta\}, \{2\theta\}, \ldots, \{N\theta\}.$$

All these numbers lie in the half open unit interval $0 \le \{m\theta\} < 1$. Now divide the unit interval into N equal half-open subintervals of length $1/N$. Then some subinterval must contain at least two of these fractional parts, say $\{a\theta\}$ and $\{b\theta\}$, where $0 \le a < b \le N$. Hence we can write

$$|\{b\theta\} - \{a\theta\}| < \frac{1}{N}. \tag{2}$$

But

$$\{b\theta\} - \{a\theta\} = b\theta - [b\theta] - a\theta + [a\theta] = (b - a)\theta - ([b\theta] - [a\theta]).$$

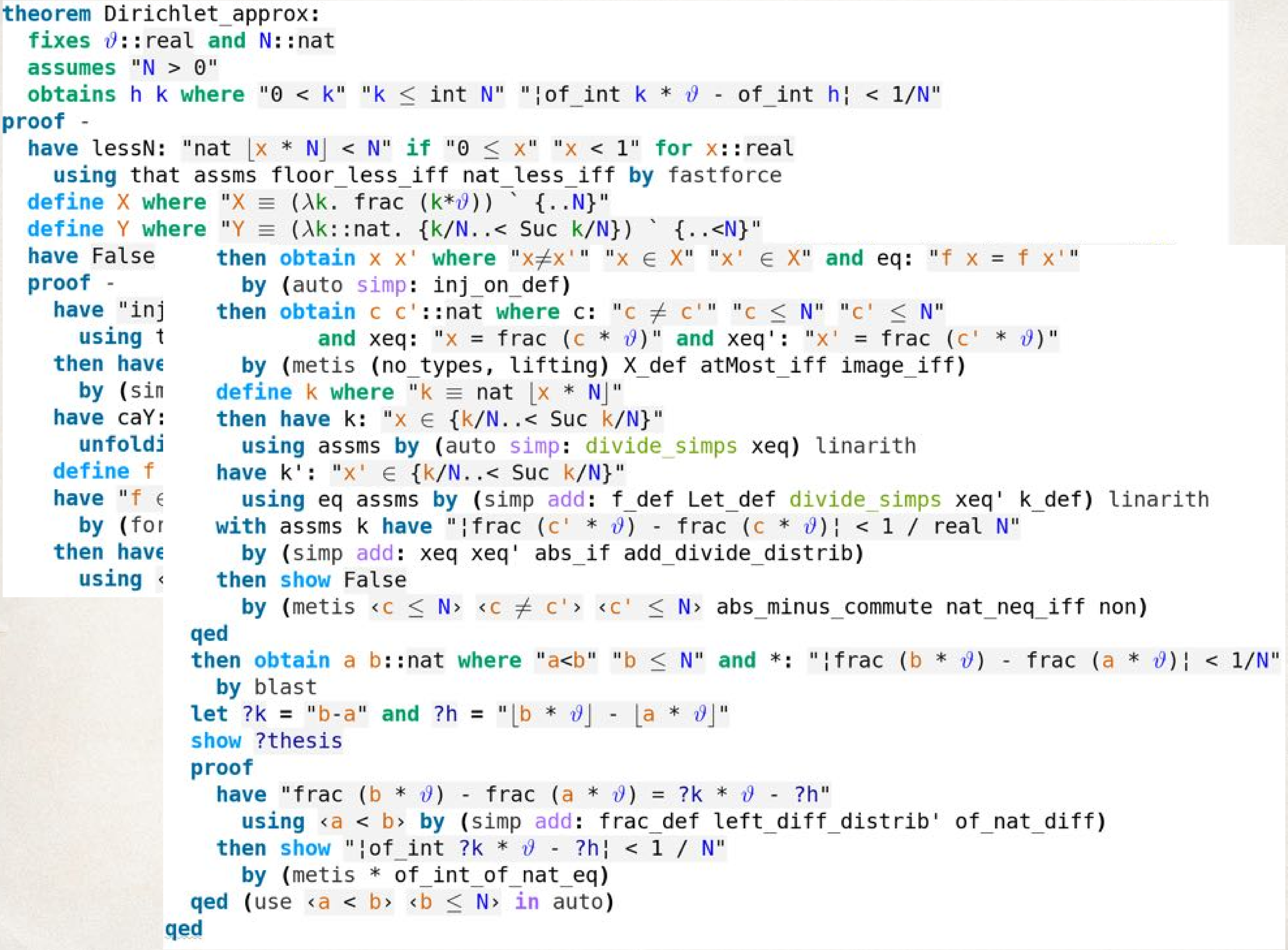
Therefore if we let

$$k = b - a \qquad \text{and} \qquad h = [b\theta] - [a\theta]$$

inequality (2) becomes

$$|k\theta - h| < \frac{1}{N}, \quad \text{with } 0 < k \le N.$$

This proves the theorem. □



```
theorem Dirichlet_approx:
  fixes ϑ::real and N::nat
  assumes "N > 0"
  obtains h k where "0 < k" "k ≤ int N" "¦of_int k * ϑ - of_int h¦ < 1/N"
proof -
  have lessN: "nat ⌊x * N⌋ < N" if "0 ≤ x" "x < 1" for x::real
    using that assms floor_less_iff nat_less_iff by fastforce
  define X where "X ≡ (λk. frac (k*ϑ)) ` {..N}"
  define Y where "Y ≡ (λk::nat. {k/N..< Suc k/N}) ` {..<N}"
  have False
  proof -
    have "inj
      using t
    then have
      by (sim
    have caY:
      unfoldi
    define f
    have "f ∈
      by (for
    then have
      using ‹
```

```
    then obtain x x' where "x≠x'" "x ∈ X" "x' ∈ X" and eq: "f x = f x'"
      by (auto simp: inj_on_def)
    then obtain c c'::nat where c: "c ≠ c'" "c ≤ N" "c' ≤ N"
            and xeq: "x = frac (c * ϑ)" and xeq': "x' = frac (c' * ϑ)"
      by (metis (no_types, lifting) X_def atMost_iff image_iff)
    define k where "k ≡ nat ⌊x * N⌋"
    then have k: "x ∈ {k/N..< Suc k/N}"
      using assms by (auto simp: divide_simps xeq) linarith
    have k': "x' ∈ {k/N..< Suc k/N}"
      using eq assms by (simp add: f_def Let_def divide_simps xeq' k_def) linarith
    with assms k have "¦frac (c' * ϑ) - frac (c * ϑ)¦ < 1 / real N"
      by (simp add: xeq xeq' abs_if add_divide_distrib)
    then show False
      by (metis ‹c ≤ N› ‹c ≠ c'› ‹c' ≤ N› abs_minus_commute nat_neq_iff non)
  qed
  then obtain a b::nat where "a<b" "b ≤ N" and *: "¦frac (b * ϑ) - frac (a * ϑ)¦ < 1/N"
    by blast
  let ?k = "b-a" and ?h = "⌊b * ϑ⌋ - ⌊a * ϑ⌋"
  show ?thesis
  proof
    have "frac (b * ϑ) - frac (a * ϑ) = ?k * ϑ - ?h"
      using ‹a < b› by (simp add: frac_def left_diff_distrib' of_nat_diff)
    then show "¦of_int ?k * ϑ - ?h¦ < 1 / N"
      by (metis * of_int_of_nat_eq)
  qed (use ‹a < b› ‹b ≤ N› in auto)
qed
```

Why should proofs be legible?

- Legible proofs can yield *insights*
- No need to *trust* the proof if you can actually read it
- Proofs can be maintained, refactored, reused

Lessons and Conclusions

It is in principle impossible to set up a system of formulas that would be equivalent to intuitionistic mathematics, for the possibilities of thought cannot be reduced to a finite number of rules set up in advance.

– Arend Heyting (1930)

Thus we are led to conclude that, although everything mathematical is formalisable, it is nevertheless impossible to formalise all of mathematics in a *single* formal system, a fact that intuitionism has asserted all along.

–Kurt Gödel (1935)

- But simple type theory (higher-order logic) worked fine for practically everything

(Whitehead and Russell were basically right)

- We found nothing that we couldn’t handle, and never had to redo a development

- Although we never had to fight the formalism, newcomers do struggle with *the system*

What areas of mathematics can we formalise?

Everything we tried: combinatorics, number theory, algebra, complex analysis, quantum computation, …

What sorts of proofs can we formalise?

Err… *Correct* proofs that don't have *large gaps*

[and where AC is admissible]

Some Obstacles

- The immensity and variety of mathematics
  - Organising libraries (including variant entries)
  - finding things in these libraries
- The difficulty of proving the obvious (recall de Bruijn’s observation)

Many thanks to my postdocs

Anthony Bordg

Angeliki Koutsoukou-Argyraki

Wenda Li

Yiannos Stathopoulos

… and to my many colleagues and students

Many thanks to the European Research Council for funding Advanced Grant ALEXANDRIA (Project GA 742178)