Ronald L. Graham · Jaroslav Nešetřil
Steve Butler *Editors*

The Mathematics of Paul Erdős I

Second Edition

Springer

The Mathematics of Paul Erdős I

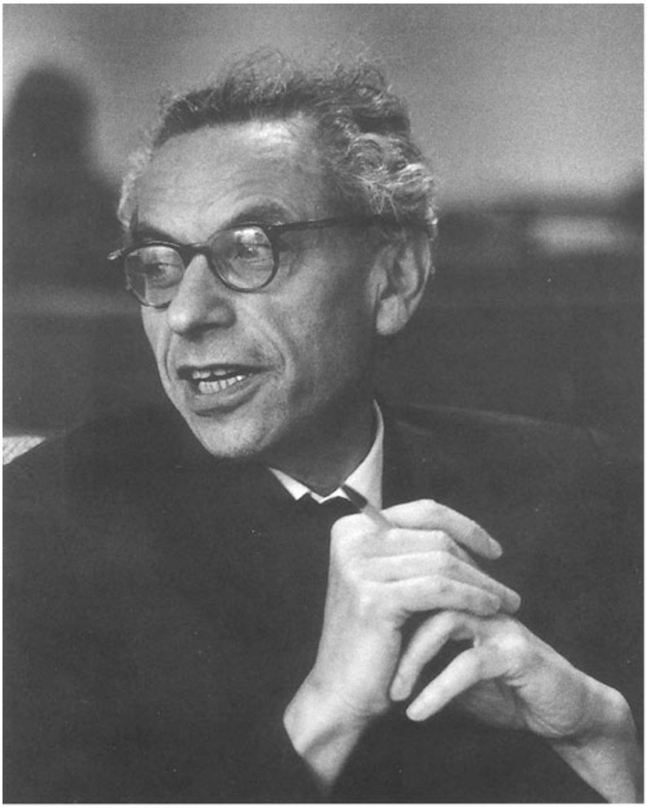


Photo by Adrian Bondy

Paul Erdős p g o m
l.d. a.d. l.d c.d n.d

Ronald L. Graham • Jaroslav Nešetřil
Steve Butler
Editors

The Mathematics of Paul Erdős I

Second Edition

Springer

Editors

Ronald L. Graham
Department of Mathematics
Computer Science and Engineering
University of California, San Diego
La Jolla, CA, USA

Jaroslav Nešetřil
Department of Applied Mathematics
and Computer Science Institute
Charles University
Prague, Czech Republic

Steve Butler
Department of Mathematics
Iowa State University
Ames, IA, USA

ISBN 978-1-4614-7257-5 ISBN 978-1-4614-7258-2 (eBook)
DOI 10.1007/978-1-4614-7258-2
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013940428

Mathematics Subject Classification: 00-02, 01A70, 01A75, 00B15, 00B30, 11-02, 51-02, 52-02, 60-02, 60C99, 68R05, 68R10

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Cover design: K. Marx

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Second Edition

In 2013 the world mathematical community is celebrating the 100th anniversary of Paul Erdős' birth. His personality is remembered by many of his friends, former disciples, and over 500 coauthors, and his mathematics is as alive and well as if he was still among us. In 1995/1996 we were preparing the two volumes of The Mathematics of Paul Erdős not only as a tribute to the achievements of one of the great mathematicians of the twentieth century but also to display the full scope of his œuvre, the scientific activity which transcends individual disciplines and covers a large part of mathematics as we know it today. We did not want to produce just a "festschrift".

In 1995/1996 this was a reasonable thing to do since most people were aware of the (non-decreasing) Erdős activity only in their own particular area of research. For example, we combinatorialists somehow have a tendency to forget that the main activity of Erdős was number theory.

In the busy preparation of the volumes we did not realize that at the end, when published, our volumes could be regarded as a tribute, as one of many obituaries and personal recollections which flooded the scientific (and even mass) media. It had to be so; the old master left.

Why then do we think that the second edition should be published? Well, we believe that the quality of individual contributions in these volumes is unique, interesting and already partly historical (and irreplaceable—particularly in Part I of the first volume). Thus it should be updated and made available especially in this anniversary year. This we feel as our duty not only to our colleagues and authors but also to students and younger scientists who did not have a chance to meet the wandering scholar personally. We decided to prepare a second edition, asked our authors for updates and in a few instances we solicited new contributions in exciting new areas. The result is then a thoroughly edited volume which differs from the first edition in many places.

On this occasion we would like to thank all our authors for their time and work in preparing their articles and, in many cases, modifying and updating them. We are fortunate that we could add three new contributions: one by

Joel Spencer (in the way of personal introduction), one by Larry Guth in Part IV of the first volume devoted to geometry, and one by Alexander Razborov in Part I of the second volume devoted to extremal and Ramsey problems. We also wish to acknowledge the essential contributions of Steve Butler who assisted us during the preparation of this edition. In fact Steve's contributions were so decisive that we decided to add him as co-editor to these volumes. We also thank Kaitlin Leach (Springer) for her efficiency and support. With her presence at the SIAM Discrete Math. conference in Halifax, the whole project became more realistic.

However, we believe that these volumes deserve a little more contemplative introduction in several respects. The nearly 20 years since the first edition was prepared gives us a chance to see the mathematics of Paul Erdős in perspective. It is easy to say that his mathematics is alive; that may sound cliché. But this is in fact an understatement for it seems that Erdős' mathematics is flourishing. How much it changed since 1995 when the first edition was being prepared. How much it changed in the wealth of results, new directions and open problems. Many new important results have been obtained since then. To name just a few: the distinct distances problem, various bounds for Ramsey numbers, various extremal problems, the empty convex 6-gon problem, packing and covering problems, sum-product phenomena, geometric incidence problems, etc. Many of these are covered by articles of this volumes and many of these results relate directly or indirectly to problems, results and conjectures of Erdős. Perhaps it is not as active a business any more to solve a particular Erdős problem. After all, the remaining unsolved problems from his legacy tend to be the harder ones. However, many papers quote his work and in a broader sense can be traced to him.

There may be more than meets the eye here. More and more we see that the Erdős problems are attacked and sometimes solved by means of tools that are not purely combinatorial or elementary, and which originate in the other areas of mathematics. And not only that, these connections and applications merge to new theories which are investigated on their own and some of which belong to very active areas of contemporary mathematics. As if the hard problems inspire the development of new tools which then became a coherent group of results that may be called theories. This phenomenon is known to most professionals and was nicely described by Tim Gowers as two cultures. [W. T. Gowers, *The two cultures of mathematics*, in Mathematics: Frontiers and Perspectives (Amer. Math. Soc., Providence, RI, 2000), 65–78.] On one side, problem solvers, on the other side, theory builders. Erdős' mathematics seems to be on one side. But perhaps this is misleading. As an example, see the article in the first volume *Unexpected applications of polynomials in combinatorics* by Larry Guth and the article in the second volume *Flag algebras: an interim report* by Alexander Razborov for a wealth of theory and structural richness. Perhaps, on the top level of selecting problems and with persistent activity in solving them, the difference between the two sides becomes less clear. (Good) mathematics presents a whole.

Time will tell. Perhaps one day we shall see Paul Erdős not as a theory builder but as a man whose problems inspired a wealth of theories.

People outside of mathematics might think of our field as a collection of old tricks. The second edition of mathematics of Paul Erdős is a good opportunity to see how wrong this popular perception of mathematics is.

La Jolla, USA R.L. Graham
Prague, Czech Republic J. Nešetřil

Just a few lines to
remember Szeged restaurant
in Budapest on July 23/96
Paul Erdős

Joachim Reuse

Laci Lovász
Walter Deuber
Miklós Laczkovich
Noga Alon
Moshe Rosenfeld
Miki Simonovits
Vince Grolmusz
NURIT Alon (NOGA'S BOSS)

Rob Tijdeman
Bernie Tijdeman
Vera T. Sós
Vesztergombi Kati
Baranyai Zs
Jiří Matoušek
Mária Grebová

IN MEMORIAM

Paul Erdős

26.3.1913–20.9.1996

The week before these volumes were scheduled to go to press, we learned that Paul Erdős died on September 20, 1996. He was 83. Paul died while attending a conference in Warsaw, on his way to another meeting. In this respect, this is the way he wanted to "leave". In fact, the list of his last month's activities alone inspires envy in much younger people.

Paul was present when the completion of this project was celebrated by an elegant dinner in Budapest for some of the authors, editors and Springer representatives attending the European Mathematical Congress. He was especially pleased to see the first copies of these volumes and was perhaps surprised (as were the editors) by the actual size and impact of the collection (On the opposite page is the collection of signatures from those present at the dinner, taken from the inside cover of the mock-up for these volumes). We hope that these volumes will provide a source of inspiration as well as a last tribute to one of the great mathematicians of our time. And because of the unique lifestyle of Paul Erdős, a style which did not distinguish between life and mathematics, this is perhaps a unique document of our times as well.

R.L. Graham
J. Nešetřil

Preface to the First Edition

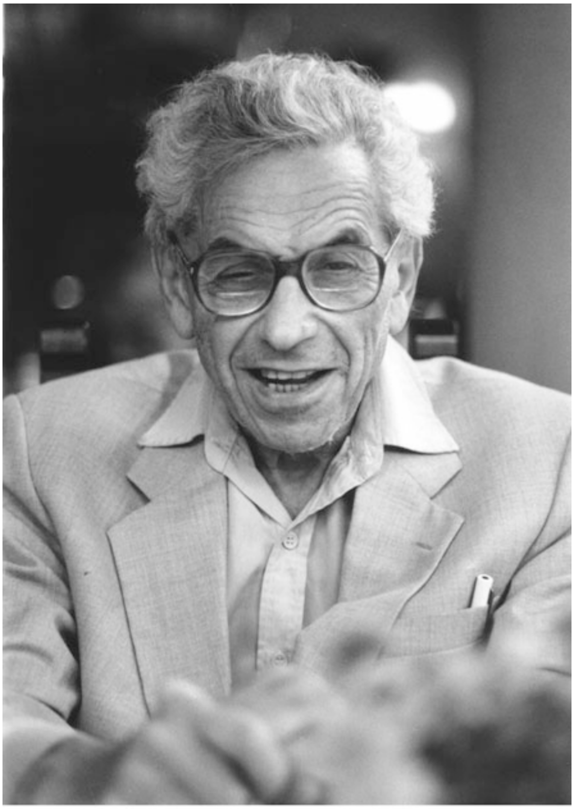
In 1992, when Paul Erdős was awarded a Doctor Honoris Causa by Charles University in Prague, a small conference was held, bringing together a distinguished group of researchers with interests spanning a variety of fields related to Erdős' own work. At that gathering, the idea occurred to several of us that it might be quite appropriate at this point in Erdős' career to solicit a collection of articles illustrating various aspects of Erdős' mathematical life and work. The response to our solicitation was immediate and overwhelming, and these volumes are the result.

Regarding the organization, we found it convenient to arrange the papers into six chapters, each mirroring Erdős' holistic approach to mathematics. Our goal was not merely a (random) collection of papers but rather a thoroughly edited volume composed in large part by articles explicitly solicited to illustrate interesting aspects of Erdős and his life and work. Each chapter includes an introduction which often presents a sample of related Erdős' problems "in his own words". All these (sometimes lengthy) introductions were written jointly by editors.

We wish to thank the nearly 70 contributors for their outstanding efforts (and their patience). In particular, we are grateful to Béla Bollobás for his extensive documentation of Paul Erdős' early years and mathematical high points; our other authors are acknowledged in their respective chapters. We also want to thank A. Bondy, G. Hahn, I. Ouhel, K. Marx, J. Načeradský and Ché Graham for their help and for the use of their works. At various stages of the project, the book was supported by AT&T Bell Laboratories, GAČR 2167 and GAUK 351. We also are indebted to Dr. Joachim Heinze and Springer Verlag for their encouragement and support. Finally, we would like to record our extreme debt to Susan Pope (at AT&T Bell Laboratories) who somehow (miraculously) managed to convert more than 50 manuscripts of all types into the attractive form they now have.

Here then is a unique portrait of a man who has devoted his whole being to "proving and conjecturing" and to the pursuit of mathematical knowledge and understanding. We hope that this will form a lasting tribute to one of the great mathematicians of our time.

Murray Hill, USA R.L. Graham
Praha, Czech Republic J. Nešetřil



Paul Erdős. Photo by George Csicsery.

Paul Erdős on the Queen Elizabeth.
Photo by Ronald Taft.

Paul Erdős in the mountains.

Paul Erdős in 1958.

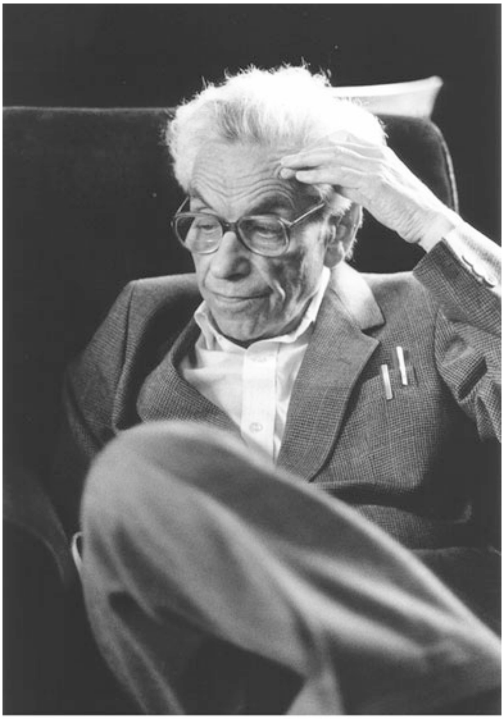
Paul Erdős with Béla Bollobás. Photo by George Csicsery.

Paul Erdős with Vera Sós.

Paul Erdős with Richard Rado.

Paul Erdős with epsilons.

Paul Erdős with Mel Nathanson.



Paul Erdős. Photo by George Csicsery.

Paul Erdős in 1941.

Paul Erdős in the 1950s.

Paul Erdős with Leo Moser.

Paul Erdős receiving a Dr.h.c. degree from Charles University. Photo by Gena Hahn.

Contents

Part III Randomness and Applications

Part IV Geometry

Paul Erdős: Life and Work

Béla Bollobás

B. Bollobás (✉)
Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB, England

Trinity College, Cambridge, CB2 1TQ, England, UK

Department of Mathematical Sciences, University of Memphis,
Memphis, TN 38152, USA
e-mail: B.Bollobas@dpmms.cam.ac.uk

Dipping into the mathematical papers of Paul Erdős is like wandering into *Aladdin's Cave*. The beauty, the variety and the sheer wealth of all that one finds is quite overwhelming. There are fundamental papers on number theory, probability theory, real analysis, approximation theory, geometry, set theory and, especially, combinatorics. These great contributions to mathematics span over six decades; Erdős and his collaborators have left an indelible mark on the mathematics of the twentieth-century. The areas of probabilistic number theory, partition calculus for infinite cardinals, extremal combinatorics, and the theory of random graphs have all practically been created by Erdős, and no-one has done more to develop and promote the use of probabilistic methods throughout mathematics.

Erdős is the mathematician *par excellence*: he thrives on mathematics, living in a state of continuous excitement; he raises, answers and communicates questions, picking up the problems of others and making incisive contributions to them with lightning speed.

Considering what a mild-mannered man he is, it is surprising that everything about Erdős and his mathematics is extreme. He has written over 1,400 papers, more than any mathematician since Euler, and has more than 400 coauthors. If the *Guinness Book of Records* had categories related to mathematical activities, Paul Erdős would hold many of the records by a margin one could not even attempt to estimate, like the thousands of problems posed, the millions of miles travelled, the tens of thousands of mathematical discussions held, the thousands of different beds slept in, the thousands of lectures delivered at different universities, the hundreds of mathematicians helped, and so on.

Today we live in the age of big mathematical theories, bringing together many sophisticated branches of mathematics. These powerful theories can be very successful in solving down-to-earth problems, as in the case of Andrew Wiles's wonderful proof of *Fermat's Last Theorem*. But no matter how important and valuable these big theories are, they cannot constitute all of mathematics. There are a remarkable number of basic mathematical questions that we would love to answer (nay, we *should* answer!) which seem

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_1,

to withstand all our assaults. There is a danger that we turn our backs on such questions, persuading ourselves that they are not interesting when in fact we mean only that we cannot tackle them with our favourite theories. Of course, such an attitude would not be in the proper spirit of science; surely, we should say that we do want to answer these questions, by *whatever means.* And if there are no theories to help us, no bulldozers to move the earth, then we must rely on our bare hands and ingenuity. It is not that we do not want to use big theories to crack our problems, but that the big theories around are unable to say anything deep about our questions. And, with luck, our hands-on approach will tie up with available theories or, better still, will lead to new, more sensitive theories.

Ernst Straus, who as a young man was Einstein's assistant, reported that the reason why Einstein had chosen physics over mathematics was that mathematics was so full of beautiful and attractive questions that one might easily waste one's life working on the "wrong" questions. Einstein was confident that in physics he could identify the "central" questions, and he felt that it was the duty of a scientist to pursue these questions and not let himself be seduced by any problem-no matter how difficult or attractive it might be.

The philosophy of Erdős has been completely different. Throughout his long career, he has been happy to pursue the beautiful problems he encountered, and has raised many others. But this is not an ad hoc process: Erdős has an amazing instinct for discerning beautiful problems that, while appearing innocuous, in fact go right to the heart of the matter. These problems are not chosen indiscriminately; they frequently lead to the discovery of unexpected and exciting phenomena. Like Ramanujan, Erdős uses particular instances of problems to explore an area. Rather than taking whole countries in one sweeping move, he prefers first to occupy some nearby castles, from which he can weigh up the unknown territory before making his next move.

For over 60 years now, Erdős has been the world's most celebrated problem solver and problem poser. Unrivalled, king, nonpareil, ... He has been called an occidental Ramanujan, a modern-day Euler, the Mozart of mathematics. These glowing epithets accurately capture the different facets of Paul Erdős—each is correct in its own way. He has a unique talent to pose penetrating questions. It is easy to ask questions that lead nowhere, questions that are either impossibly hard or too easy. It is a completely different matter to raise, as Erdős does, innocent-looking problems whose solutions shed light on the shape of the mathematical landscape.

An important feature of the problems posed by Erdős is that they carry differing monetary rewards. Needless to say, this is done in jest, but the prizes do indicate Erdős's assessment of the difficulty of the problems. How different this is from the annoying habit of some mathematicians, who casually mention a problem as if they hadn't even thought about it, when in fact they are telling you the central problem they have been working on for a long time!

Two features of his mathematical *œuvre* stand out: his mastery of *elementary* methods and his advocacy of *random* methods. Starting with his very first papers, Erdős championed *elementary* methods in diverse branches of mathematics. He showed, again and again, that elementary methods often succeed against overwhelming odds. In many brilliant proofs he showed that, rather than bringing somewhat foreign machinery to bear on some problems, and thereby trying to fit a square peg into a round hole, one can progress considerably further by facing the complications, going deep into the problem, and tailoring our approach to the intrinsic difficulties of the problem. This philosophy can pay unexpected dividends, as shown by Charles Read's solution of the Invariant Subspace Problem, Miklós Laczkovich's solution of Tarski's problem of "Squaring the Circle", and Tim Gowers' recent solutions of Banach's last unsolved problems, including the Hyperplane Problem.

As to *probabilistic* methods, by now it is widely acknowledged that these can be remarkably effective in tackling main-line questions in diverse areas of mathematics that have nothing to do with probability. It is worth remembering, though, that when Erdős started it all, the idea was very startling indeed. That today we take it in our stride is a sign of the tremendous success of the random method, which is very much his method, still frequently called the *Erdős method.*

Paul Erdős was born on 26th March 1913, in Budapest. His parents were teachers of mathematics and physics; his father translated a book on aircraft design from English into Hungarian. The young Paul did not go to elementary school, but was brought up by his devoted mother, Anna, and, for 3 years, between the ages of 3 and 6, he had a German *Fräulein.* His exceptional talent for mathematics was evident by the time he was 3: his agility at mental arithmetic impressed all comers, and he was not yet 4 when he discovered negative numbers for himself. With the outbreak of the First World War, his father was drafted into the Austro-Hungarian army, and served on the Eastern Front. He was taken prisoner by the Russians, and sent to Siberia to a prisoner of war camp, from which he returned only after about 6 years.

After the unconditional surrender of Hungary at the end of the War, the elected government resigned, as it could not accept the terms of the Allies. These terms left Hungary only the rump of her territory, and in March 1919 the communists took over the country, with the explicit aim of repelling the Allies. The communists formed a *Dictatorship of the Proletariat*, usually referred to as the *Commune*, after its French equivalent in 1871, and set about defending the territory and forcibly reforming the social order.

The Commune could not resist the invasion by the Allies and the Hungarian "white" officers under *Admiral Horlhy*, and it fell after a struggle of 3 months. Unfortunately for the Erdős family, Anna Erdős had a minor post under the Commune, and when Horthy came to power, she lost her job, never to teach again. Later she worked as a technical editor.

He studied elementary school privately with his mother. After that, in 1922, the young Erdős went to Tavaszmező gymnasium, the first year as a private pupil, the second and third years as a normal student, and the fourth year again as a private pupil. After the fourth year he attended St. Stephen's School (Szent István Gimnázium) where his father was a high school teacher. At this time Erdős also received significant instruction from his parents as well. As it happened, my father entered the school just as Erdős left it, so they share many classmates, although they met only many years later.

By the early 1920s the *Mathematical Journal for Secondary Schools* (Középiskolai Matematikai Lapok) was a successful journal, catering for pupils with talent for mathematics. The journal had been founded in 1895 by a visionary young man, Dániel Arany, who hoped to raise the level of mathematics in the whole of Hungary by enticing students to mathematics through beautiful problems. The backbone of the journal was the year-long competition. Every month a number of problems were set for each age group; the readers were invited to submit their solutions, which were marked, and the best published under the names of the authors.

The young Erdős became an ardent reader of this journal, and his love of mathematics was greatly fanned by the intriguing problems in it. In some sense, Erdős's earliest publications date to this time, with the appearance of his solutions in the journal. On one occasion Paul Erdős and Paul Turán were the only ones who managed to solve a particular problem, and their solution was published under their joint names. This was Erdős's first "joint paper" with Turán, whom he had not even met at the time, and who later became one of his closest friends and most important collaborators.

Mathematicians, and especially young mathematicians, learn much from each other. Erdős was very lucky in this respect, for when at the age of 17 he entered the Pázmány Péter Tudományegyetem (the science university of Budapest) he found there an excellent group of about a dozen youngsters devoted to mathematics. Not surprisingly, Erdős became the focal point of this group, but the long mathematical discussions stimulated him greatly.

This little group included Paul Turán, the outstanding number theorist; Tibor Gallai, the excellent combinatorialist; Dezső Lázár, who was later tragically killed by the Nazis; George Szekeres and Esther Klein, who later married and subsequently emigrated to Australia; László Alpár, who became an important member of the Hungarian Mathematical Institute; Márta Svéd, another member of the group who went to live in Australia; and several others. Not only did they discuss mathematics at the university, but also in the afternoons and evenings, when they used to meet at various public places, especially by the *Statue of Anonymous*, commemorating the first chronicler of Hungarian history.

Two of Erdős's professors stand out: Lipót Fejér, the great analyst, and Dénes König, who introduced Erdős to graph theory. The lectures of König led to the first results of Erdős in graph theory: in answer to a question posed in the lectures, in 1931 he extended Menger's theorem to infinite graphs. Erdős never published his proof, but it was reproduced in König's classic, published in 1936.

As an undergraduate, Erdős worked mostly on number theory, obtaining several substantial results. He was not even 20, when in Berlin the great Issai Schur lectured on Erdős's new proof of Bertrand's postulate. He wrote his doctoral dissertation as a second year undergraduate, and it was not long before he got into correspondence with several mathematicians in England, including Louis Mordell, the great number theorist in Manchester, and Richard Rado and Harold Davenport in Cambridge. All three became Erdős's close friends.

When in 1934 Erdős finished university, he accepted Mordell's invitation to Manchester. He left Hungary for England in the autumn of 1934, not knowing that he would never again live in Hungary permanently. On 1st October 1934 he was met at the railway station in Cambridge by Davenport and Rado, who took him to Trinity College, and they immediately embarked on the first of their many long mathematical discussions. Next day Erdős met Hardy and Littlewood, the giants of English mathematics, before hurrying on to Mordell.

Mordell put together an amazing group of mathematicians in Manchester, and Erdős was delighted to join them. First he took up the *Bishop Harvey Goodwin Fellowship*, and was later awarded a *Royal Society Fellowship*. He was free to do research under Mordell's guidance, and he was soon producing papers with astonishing rapidity. In 1937 Davenport left Cambridge to join Mordell and Erdős, and their life-long friendship was soon cemented. I have a special reason to be grateful for the Erdős-Davenport friendship: many years later, I was directed to my present home, Trinity College, Cambridge, only because Davenport was a Fellow here, and he was a good friend of Erdős.

In 1938 Erdős was offered a fellowship at the Institute for Advanced Study in Princeton, so he soon thereafter sailed for the U.S., where he was to spend the next decade. The war years were rather hard on Erdős, as it was not easy to hear from his parents in Budapest, and when he received news, it was never good. His father died in August 1942, his mother later had to move to the Ghetto in Budapest, and his grandmother died in 1944. Many of his relatives were murdered by the Nazis.

In spite of being cut off from his home, Erdős continued to pour forth wonderful mathematics at a prodigious rate. Having arrived in America, he spent a year and a half at Princeton, before starting on his travels. He visited Philadelphia, Purdue, Notre Dame, Stanford, Syracuse, Johns Hopkins, to mention but a few places, and the pattern was set: like a Wandering Scholar of the Middle Ages, Erdős never stopped again. In addition to the many important papers he wrote by himself, he collaborated more and more with mathematicians from diverse areas, writing outstanding joint papers with Mark Kac, Wintner, Kai Lai Chung, Ivan Niven, Arye Dvoretzky, Shizuo Kakutani, Arthur A. Stone, Leon Alaoglu, Irving Kaplansky, Alfred Tarski, Gabor Szegő, William Feller, Fritz Herzog, George Piranian, and others. Through correspondence, he continued his collaboration with Paul Turán, Harold Davenport, Chao Ko and Tibor Gallai (Grünwald).

In 1954, he left the U.S. to attend the International Congress of Mathematicians in Amsterdam. He had also asked for a reentry permit at that time but his request was denied. So he left without a reentry permit since in his own words, "Neither sam nor joe can restrict my right to travel." Left without a country, Israel came to his aid, offering him employment at the Hebrew University in Jerusalem, and a passport. He arrived in Israel on 30th November 1954, and from then on he has been to Israel practically every year. Before leaving Israel for Europe in July 1955, he applied for a return visa to Israel. When the officials asked him whether he wanted to become an Israeli citizen, he politely refused, saying that he did not believe in citizenship.

After the upheaval following his trip to Amsterdam, he first returned to the U.S. in 1959; the relationship between Erdős and the U.S. Immigration Department was finally normalized in 1963, and since then he has had no problems with them.

In the *Treaty of Yalta*, Hungary was placed within the Soviet sphere of influence; the communists, aided by the Russians, took over the government, and turned Hungary into a *People's Republic.* For ordinary Hungarians, leaving Hungary even for short trips to the West became very difficult. Nevertheless, in 1955 Erdős managed to return to Hungary for a short time, when his good friend, George Alexits, pulled strings and convinced the officials that, if Erdős were to enter the country, he should be allowed to leave.

Later Erdős could return to Hungary at frequent intervals, in order to spend more and more time with his mother, as well as to collaborate with a large number of Hungarian mathematicians, especially Turán and Rényi, In those dark days, Erdős was *the* main link between many Hungarian mathematicians and the West.

As a young pupil, I first heard him lecture during one of his visits: not only did he talk about fascinating problems but he also cut a flamboyant figure, with his suntan, Western suit and casual mention of countries I was sure I could never visit. I got to know him during his next visit: in 1958, having won the National Competition, I was summoned to the elegant hotel he stayed in with his mother. They could not have been kinder: Erdős told to me a host of intriguing questions, and did not talk down to me, while his mother (whom, as most of their friends, I learned to call *Annus Néni* or Aunt Anna) treated me to cakes, ice cream and drinks. Three years later they got to know my parents, and from then on they were frequent visitors to our house, especially for Sunday lunches. My father, who was a physician, looked after both Erdős and Annus Néni.

Seeing them together, there was no doubt that they were very happy in each other's company: these were blissful days for both of them. Erdős thoroughly enjoyed being with his mother, and she was delighted to have her son back for a while. They looked after each other lovingly; each worried whether the other ate well and slept enough or, perhaps, was a little tired.

Annus Néni was fiercely proud of her wonderful son, loved to see the many signs that her son was a great mathematician, and revelled in her role as the *Queen Mother of Mathematics*, surrounded by all the admirers and well-wishers. She was never far from Erdős's mathematics either: she kept Erdős's hundreds of reprints in perfect order, sending people copies on demand.

Annus Néni was not young, having been born in 1880, but her health was good and she was very sharp. To compensate for the many years when they had been kept apart, Annus Néni started to travel with her son in her 80s; their first trip together being to Israel in November 1964. From then on they travelled much together: to England in 1965, many times to other European countries and the U.S., and towards the end of 1968 to Australia and Hawaii. When, tinged with envy, we told her that it must be wonderful to see the world, she replied "*You know that I don't travel because I like it but to be with my son.*" It was a tragedy for Erdős when, in 1971, Annus Néni died during a trip to Calgary. Her death devastated him and for years afterwards he was not quite himself. He still hasn't recovered from the blow, and it is most unlikely that he ever will.

Erdős's brushes with officialdom were not quite over: the communists also managed to upset him. In 1973 there was an international meeting in Hungary, to celebrate his 60th birthday. Erdős's friends from Israel were denied a visa to enter Hungary; this outraged him so much that for 3 years he did not return to Hungary.

With the collapse of communism and with the end of the Cold War, Erdős has entered a golden age of travel: not only can he go freely wherever he wants to, but he is even welcomed by officials everywhere.

Having started as a mathematical *prodigy*, by now Erdős is the *doyen* of mathematicians, with more friends in mathematics than the number of people most of us meet in a lifetime. As he likes to put it in his inimitable way, he has progressed from *prodigy to dotigy.* As a Member of the Hungarian Academy of Sciences, Erdős has a permanent position in Budapest. During summer months, he frequently stays in the Guest House of the Academy, two doors away from my mother, visiting Vera Sós, András Hajnal, Miklós Simonovits, András Sárközy, Miklós Laczkovich, and inspiring many others. Another permanent position awaits him in Memphis, where he stays and works with Ralph Faudree, and his other friends, Dick Schelp and Cecil Rousseau. In Israel he visits all the universities, including the Technion in Haifa, Tel Aviv, Jerusalem and the Weizman Institute. But for years now, Erdős has had many other permanent ports of call, including Kalamazoo, where Yousef Alavi looks after him; New Jersey and the New York area, where he stays with Ron Graham and Fan Chung and talks to many others as well, including János Pach, Joel Spencer, Mel Nathanson, Peter Winkler, Endre Szemerédi, Joseph Beck and Herb Wilf; Calgary, mostly because of Eric Milner, Richard Guy and Norbert Sauer; Atlanta, with Dick Duke, Vojtěch Rödl, Ron Gould and Dwight Duffus. And the list could go on and on, with Athens, Baton Rouge, Berlin, Bielefeld, Boca Raton, Bonn, Boston, Cambridge, Chicago, Los Angeles, Lyon, Minneapolis, Paris, Poznań, Prague, Urbana, Warsaw, Waterloo, and many others.

Honours have been heaped upon Erdős, although he could not care less. Every fifth year there is an International Conference in Cambridge on his birthday, and in 1991 Cambridge also awarded him a prestigious Honorary Doctorate, as did the Charles University of Prague a year later, and many other universities since. On the occasion of his 80th birthday, he was honoured at a spate of conferences, not only in Cambridge, but also in Kalamazoo, Boca Raton, Prague and Keszthély.

Nowadays Erdős lectures in more places than ever, interspersing his mathematical problems with stories about mathematicians and his remarks about life. He dislikes cold but, above all, hates old age and stupidity, and so he appreciates the languages in which these evils sound similar. Thus, *old* and *cold* and *alt* and *kalt* go hand in hand in English and German, and in no other language he knows. But Hindi is better still because the two greatest evils sound almost the same: *buddha* is old and *budu* is stupid.

Erdős is fond of paraphrasing poems, especially Hungarian poems, to illustrate various points. The great Hungarian poet at the beginning of this century, Endre Ady, Wrote: *Legyen átkozott aki a helyembe áll! (Let him be cursed who takes my place!)* As a mathematician builds the work of others, so that his immortality depends on those who continue his work, Erdős professes the opposite: *Let him be blessed who takes my place!*

But Erdős does not wait for posterity to find people to continue his work: his extraordinary number of collaborators ensures that many people carryon his work all around the world. The collaborators who particularly stand out are Paul Turán, Harold Davenport, Richard Rado, Mark Kac, Alfréd Rényi, András Hajnal, András Sarközy, Vera Sós and Ron Graham: they have all done much major work with Erdős. In a moment we shall see a brief account of some of this work. Needless to say, our review of Erdős's mathematics will be woefully brief and inadequate, and will also reflect the taste of the reviewer.

Erdős wrote his first paper as a first-year undergraduate, on Bertrand's postulate that, *for every* $n \geq 1$*, there is a prime* p *satisfying* $n < p \leq 2n$. Bertrand's postulate was first proved by Chebyshev, but the original proof was rather involved, and in 1919 Ramanujan gave a considerably simpler proof of it. In his fundamental book, *Vorlesungen über Zahlentheorie*, published in Leipzig in 1927, Landau gave a rather simple proof of the assertion that for some $q > 1$ and every $n \geq 1$, there is a prime between n and qn. However, Landau's q could not be taken to be 2. In his first paper, Erdős sharpened Landau's argument, and by studying the prime factors of the binomial coefficient $\binom{2a}{a}$, gave a simple and elementary proof of Bertrand's postulate.

Erdős was quick to develop further the ideas in his first paper. In 1932, Breusch made use of L-functions to generalize Bertrand's postulate to the arithmetic progressions $3n+1$, $3n+2$, $4n+1$ and $4n+3$: for every $m \geq 7$ there are primes of the form $3n+1$, $3n+2$, $4n+1$ and $4n+3$ between m and $2m$. By constructing expressions containing, as factors, all terms of the arithmetic progression at hand, and rather few other factors, Erdős managed to give an elementary proof of Breusch's theorem, together with various extensions of it to other arithmetic progressions. These results constituted the Ph.D. thesis Erdős wrote as a second-year undergraduate, and published in Sárospatak in 1934.

Schur, who had been Breusch's supervisor in Berlin, was quick to recognize the genius of the author of the beautiful elementary proof of Breusch's theorem. When, a little later, Erdős proved a conjecture of Schur on abundant numbers, and solved another problem of Schur, Erdős became "der Zauberer von Budapest" ("the magician of Budapest")—no small praise from the great German for a young man of 20.

Abundant numbers figured prominently among the early problems tackled by Erdős. In his lectures on number theory, Schur conjectured that the abundant numbers have positive density: $\lim_{x\to\infty} A(x)/x$ exists, where $A(x)$ is the number of abundant numbers not exceeding x. (A natural number n is *abundant* if $\sigma(n)$, the sum of its positive divisors, is at least $2n$.) The beautiful elementary proof Erdős gave of this conjecture led him straight to other problems concerning the distribution of the values of real-valued additive arithmetical functions $f(n)$, that is functions $f : \mathbb{N} \to \mathbb{R}$ satisfying $f(ab) = f(a) + f(b)$ whenever $(a,b) = 1$.

These problems were first investigated by Hardy and Ramanujan in 1917, but were more or less forgotten for over a decade. As eventually proved by Erdős and Wintner in 1939, a real-valued additive arithmetical function $f(n)$ behaves rather well if the following three series are convergent:

$$\sum_{|f(p)|>1} 1/p, \qquad \sum_{|f(p)|\le 1} f(p)/p \quad \text{and} \quad \sum_{|f(p)|\le 1} f(p)^2/p,$$

with the summations over primes p. To be precise, the three series above are convergent if and only if $\lim_{x\to\infty} A(x)/x$ exists for every real c, where $A_c(x)$ stands for the number of natural numbers $n \le x$ with $f(n) < c$.

In 1934, Turán gave a marvelous proof of an extension of the Hardy-Ramanujan theorem on the "typical number of divisors" of a natural number. Writing $\nu(n)$ for the number of *distinct* prime factors of n (so that $\nu(12) = 2$), Turán proved that

$$\sum_{n=1}^{N} \{\nu(n) - \log\log n\}^2 = N \log\log N + o(N \log\log N).$$

It is a little disappointing that Hardy, one of the greatest mathematicians alive, failed to recognize the immense significance of this new proof. Erdős, on the other hand, not only saw the significance of the paper, but was quick to make use of the probabilistic approach and so became instrumental in the birth of a very fruitful new branch of mathematics, *probabilistic number theory*. In a ground-breaking joint paper he wrote with Kac in 1939, Erdős proved that if a bounded real-valued arithmetical function $f(n)$ satisfies $\sum_p f(p)^2/p = \infty$ then, for every $x \in \mathbb{R}$,

$$\lim_{m\to\infty} A_x(m)/m = \Phi(x),$$

where $A_x(m)$ is the number of positive integers $n \le m$ satisfying

$$f(n) < \sum_{p\le m} f(p)/p + x\Big(\sum_{p\le m} f(p)^2/p\Big)^{1/2},$$

and, as usual

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

is the *standard normal distribution*. In other words, the arithmetical function $f(n)$ satisfies the Gaussian law of error! It took the mathematical community quite a while to appreciate the significance and potential of results of this type.

Note that for $\nu(n)$, the number of prime factors of n, the Erdős-Kac theorem says that if $x \in \mathbb{R}$ is fixed then

$$\lim_{m\to\infty} \frac{1}{m} \big|\{n \le m \text{ and } \nu(n) \le \log\log m + x(\log\log m)^{1/2}\}\big| = \Phi(x).$$

Starting with his very first papers, Erdős championed "*elementary*" methods in number theory. That the number theorists in the 1930s appreciated elementary methods was due, to some extent, to Shnirelman's great success in studying integer sequences, with a view of attacking, perhaps, the Goldbach conjecture. To study integer sequences, Shnirelman introduced a density, now bearing his name: an integer sequence $0 \le a_1, < a_2 < \dots$ is said to have *Shnirelman density* α if

$$\inf_{x\ge 1} \frac{1}{x} \sum_{a_n \le x} 1 = \alpha.$$

Thus if $a_1 > 1$ then the Shnirelman density of the sequence $(a_n)_{n=1}^{\infty}$ is 0.

Khintchine discovered the rather surprising fact that if $(a_n)_{n=1}^{\infty}$ is an integer sequence of Shnirelman density α with $0 < \alpha < 1$, and $(b_n)_{n=1}^{\infty}$ is the sequence of squares $0^2, 1^2,\ 2^2, \dots$, then the "sum-sequence" $(a_n + b_m)$ has Shnirelman density *strictly greater* than α. The original proof of this result, although elementary, was rather involved.

When Landau lectured on Khintchine's theorem in 1935 in Cambridge, he presented a somewhat simplified proof he had found with Buchstab. Nevertheless, talking to Landau after his lecture, Erdős expressed his view that the proof should be considerably simpler and, to Landau's astonishment, as soon as the next day he came up with a "proper" proof that was both elementary and short. In addition, the new proof also made it clear what the result had to do with squares: all one needs is that every positive integer is the sum of at most four squares. If (b_n) is such that every positive integer is the sum of at most k terms b_n, then the sum-sequence $(a_n + b_m)$ has Shnirelman density at least $\alpha + \alpha(1-\alpha)/2k$. It says much about Landau, that he immediately included this beautiful theorem of Erdős into the Cambridge "Tract" he was writing at the time (*Neue Ergebnisse der additiven Zahlentheorie*, published in 1937).

The difference between consecutive primes has attracted much attention. Writing p_n for the nth prime, the *twin prime conjecture* states that $p_{n+1} - p_n$ is infinitely often equal to 2, that is $\liminf_{n\to\infty}(p_{n+1} - p_n) = 2$. At the moment we seem to be very far from a proof of this conjecture; in fact, there seems to be no hope to prove that $\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$. The *Prime Number Theorem*, asserting that $\pi(x) \sim x/\log x$, where $\pi(x)$ is the number of primes $p \le x$, implies that $c = \liminf_{n\to\infty}(p_{n+1} - p_n)/\log p_n \le 1$, but Erdős was the first to prove, in 1940, that $c < 1$. Later Rankin showed that $c \le 59/60$, and then Selberg that $c \le 15/16$. Subsequent improvements were obtained by Bombieri and Davenport and by Huxley; the present record, $c \le 0.248$, is held by Maier.

Independently, Erdős and Ricci showed that the set of limit points of the sequence $(p_{n+1} - p_n)/\log p_n$ has positive Lebesgue density and yet, *no number is known to be a limit point.*

Concerning *large* gaps between consecutive primes, Backlund proved in 1929 that $\limsup_{n\to\infty}(p_{n+1} - p_n)/\log p_n \ge 2$. In quick succession, this was improved by Brauer and Zeitz (1930), by Westzynthius (1931), and then by Ricci (1934), to

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n \log\log\log p_n} > 0.$$

By making use of the method of Brauer and Zeitz, Erdős proved in 1934 that

$$\limsup_{n\to\infty} \frac{(p_{n+1}-p_n)(\log\log\log p_n)^2}{\log p_n \log\log p_n} > 0.$$

In 1938 this result was improved by Rankin, who smuggled a factor $\log\log\log\log p_n$ into the denominator: there is a $c > 0$ such that

$$p_{n+1} - p_n > c\frac{\log p_n \log\log p_n \log\log\log\log p_n}{(\log\log\log p_n)^2} \tag{1}$$

for infinitely many values of n. It seems to be extremely difficult to improve this result, to the extent that Erdős is offering (according to him, perhaps a little rashly) \$10,000 for a proof that (1) holds for every c. The original value of c given by Rankin was improved by Maier and Pomerance in 1990.

Although in the 1930s elementary methods were spectacularly successful in additive number theory and in the study of additive arithmetical functions, they did not seem to be suitable for the study of the distribution of primes.

It was not only a desire for diverse proofs that urged mathematicians to search for elementary proofs of results proved by deep analytical methods: many mathematicians, including Hardy, felt that if the *Prime Number Theorem* (PNT) could be proved by elementary methods then the *Riemann Hypothesis* itself might yield to a similar attack. This belief was reinforced by the result of Norbert Wiener in 1930 that the prime number theorem is *equivalent* to the fact that the zeta function $\zeta(s) = \zeta(\sigma + it)$ has no zero on the line $\sigma = 1$.

Next to the PNT, Dirichlet's classical theorem on primes in an arithmetical progression, proved in 1837, was a test case for the power of elementary methods. In 1948 Atle Selberg found an ingenious elementary proof of Dirichlet's theorem; indeed, Selberg proved that if k and l are relatively prime numbers then

$$\liminf_{x\to\infty} \frac{1}{\log x} \sum_{p\le x;\, p\equiv l \pmod{k}} p^{-1}\log p > 0.$$

Shortly after this, Selberg proved the following *fundamental formula*:

$$\sum_{p\le x} (\log p^2) + \sum_{pq\le x} \log p \log q = 2x\log x + O(x), \tag{2}$$

where p and q run over primes. This formula is an easy consequence of the PNT, but what caused the excitement was that Selberg gave a completely elementary proof. Thus the fundamental formula could be a starting point for elementary proofs of various theorems in number theory which previously seemed inaccessible by elementary methods.

Using Selberg's fundamental formula, Erdős quickly proved that $p_{n+1}/p_n \to 1$ as $n \to \infty$, where p_n is, as before, the nth prime. Even more, Erdős proved (in an entirely elementary way) that if $c > 1$ then

$$\liminf_{n\to\infty} \frac{\log x}{x}(\pi(cx) - \pi(x)) > 0. \tag{3}$$

Erdős communicated this proof of (3) to Selberg, who, 2 days later, using (2), (3) and the ideas in the proof of (3), deduced the PNT $\lim_{n\to\infty} \frac{\pi(x)\log x}{x} = 1$. Thus an elementary proof of the PNT was found!

A little later Selberg found it possible to argue directly from (2), without making any use of (3); this is the way he wrote up his paper in the autumn of 1948. In a separate paper, Erdős stated (2), referred to Selberg's final proof of the PNT (not published at the time), gave his own proof of (3), Selberg's deduction of the PNT from (2) and (3), and a joint simplified deduction of the PNT from (2). In the *Mathematical Reviews* the great Cambridge mathematician A.E. Ingham found it convenient to review these two papers together. As he wrote, "All previous proofs have been by '*transcendental*' arguments involving some appeal to the theory of functions of a complex variable. Successive proofs have moderated the demands on this theory, or invoked alternative analytical theories (e.g., Fourier transforms), but there remained a nucleus of complex variable theory, namely the proposition that

Riemann zeta-function $\zeta(s) = \zeta(\sigma + it)$ has no zeros on the line $\sigma = 1$; and this could hardly be avoided, except by a radically new approach, since the PNT is in a clearly definable sense 'equivalent' to this property of $\zeta(s)$. It has long been recognized that an 'elementary' proof of the PNT, not depending on analytical ideas remote from the problem itself, would (if indeed possible) constitute a discovery of the first importance for the logical structure of the theory of the distribution of primes. An elementary (though not easy) proof is given, in various forms, in these two papers".

"In principle, [the papers] open up the possibility of a new approach, in which the old logical arrangement is reversed and analytical properties of $\zeta(s)$ are deduced from arithmetical properties of the sequence of primes. How far the practical effects of this revolution of ideas penetrate into the structure of the subject, and how much of the theory will ultimately have to be rewritten, it is too early to say."

For the startling elementary proof of the Prime Number Theorem Selberg was awarded a *Fields Medal*, and Erdős a *Cole Prize*, given every fourth year to the author of the best paper in algebra and number theory published in an American journal.

Let us say a few words about the contributions of Erdős to asymptotic formulae. One of the glorious achievements of the Hardy-Ramanujan partnership was the striking formula for $p(n)$, the number of different partitions of n (ignoring the order of the summands). By using powerful analytic methods that eventually led to the celebrated circle method of Hardy and Littlewood, in 1918 Hardy and Ramanujan gave an extremely good approximation for $p(n)$; later Rademacher improved the approximation a little and turned it into an *analytic expression* for $p(n)$. A weak form of the Hardy-Ramanujan result states that

$$p(n) \sim \frac{1}{4\sqrt{3}n} e^{\pi\sqrt{2n/3}},$$

and Hardy and Ramanujan also gave an elementary proof of

$$\log p(n) \sim \pi\sqrt{2n/3}.$$

Two decades later, Erdős set about proving that elementary methods can go considerably further, and in 1942 he proved that

$$p(n) \sim \frac{a}{n} e^{\pi\sqrt{2n/3}}$$

for some positive constant a.

Taking his cue from the Hardy-Ramanujan result mentioned a little earlier, that most integers n have about $\log\log n$ prime factors, Erdős also proved, with Lehner, that "almost all" partitions of a positive integer n contain about

$$A(n) = \frac{1}{\pi}\sqrt{3n/2}\log n$$

summands. Furthermore, there is also a beautiful distribution about $A(n)$: for $x \in \mathbb{R}$, the probability that a random partition of n has at most $A(n) + x\sqrt{n}$ summands tends to

$$e^{-\frac{\sqrt{6}}{\pi}e^{-\pi\sqrt{x/6}}}.$$

Some years later, in 1946, Erdős returned to another variant of this problem. Given $n \in N$, what is the *most likely* number of summands in a random partition of n? Writing $k_0(n)$ for this number, it is not clear that $k_0(n)$ is well-defined although, as was shown later by Szekeres, this is the case. However, what does seem to be clear is that $k_0(n)$ is about $A(n)$. Erdős proved that, in fact,

$$k_0(n) = A(n) + \frac{\sqrt{6}}{\pi}\left(\log\frac{\sqrt{6}}{\pi}\right)\sqrt{n} + o(\sqrt{n}).$$

Another circle of problems that has occupied Erdős for over 60 years originated with a question raised by Sidon when Erdős and Turán went to see him. Given a sequence $\mathcal{S}$ of natural numbers and $k \in \mathbb{N}$, write $r_k(n)$ for the number of representations of n in the form

$$n = a_1 + a_2 + \cdots + a_k,$$

with $a_i \in \mathcal{S}$ and $1 \le a_1 < a_2 < \ldots < a_k$. Call $\mathcal{S}$ an *asymptotic basis of order* k if $r_k(n) \ge 1$ whenever n is sufficiently large. In 1932 Sidon asked Erdős the following question. Is there an asymptotic basis of order 2 such that $r_2(n) = o(n^\epsilon)$ for every $\epsilon > 0$? The young Erdős confidently reassured Sidon that he would come up with such a sequence. Erdős was right, but it took him over 20 years: he proved in 1954 in Acta (Szeged) that for some constant c there is a sequence $\mathcal{S}$ such that

$$1 \le r_2(n) < c \log n$$

if n is large enough.

What can one say about $r_k(n)$ rather than $r_2(n)$? In 1990, Erdős and Tetali proved that for every $k \ge 2$ there are positive constants c_1, c_2, and a sequence $\mathcal{S}$ such that $c_1 \log n \le r_k(n) \le c_2 \log n$ if n is large enough. In fact, Erdős and Tetali gave two proofs of this theorem; the easier of the two gets the result as a fairly simple consequence of Janson's powerful and ingenious correlation inequality. The related conjecture of Erdős and Turán, made in 1941, that if $r_2(n) \ge 1$ for all sufficiently large n then $\limsup_{n\to\infty} r_2(n) = \infty$, is still far from being solved, although it seems possible that much more is true, namely if $r_2(n) \ge 1$ whenever n is large enough then $\limsup_{n\to\infty} r_2(n)/\log n > 0$.

In 1956, Erdős and Fuchs proved a remarkable theorem somewhat related to Sidon's problem but originating in a result of Hardy and Landau. Let us write $r(n)$ for the number of lattice points in $\mathbb{Z}^2$ in the circle of radius $\sqrt{n}$, so that $r(n)$ is the number of *integer* solutions of the inequality $x^2 + y^2 \le n$.

Gauss was the first to prove that $r(n)$ stays rather close to its expectation, namely $r(n) - \pi n = O(n^{1/2})$. In 1906, Sierpiński returned to the study of $r(n)$, and showed that, in fact, $r(n) - \pi(n) = O(n^{1/3})$. The question whether this bound is essentially best possible or could be improved: intrigued many of the best number theorists in the first few decades of this century: including Hardy, Littlewood, Landau and Walfisz. In 1925 Hardy and Landau gave an exact expression for $r(n) - \pi(n)$ is terms of Bessel functions. They showed also that $r(n)$ does not stay too close to its expectation $\pi(n)$, namely that

$$\limsup_{n\to\infty} \frac{|r(n) - \pi(n)|}{(n \log n)^{1/4}} > 0.$$

Erdős and Fuchs, proved that this result has *nothing to do* with the sequence of squares $0^2, 1^2, \ldots$ but it holds in great generality. Indeed, let $0 \leq a_1 \leq a_2 \leq \cdots$ be any sequence of integers, and for $n \in \mathbb{N}$ let $r^*(n)$ be the number of solutions of the inequality $a_i + a_j \leq n$. Then, as proved by Erdős and Fuchs for *every positive real* α we have

$$\limsup_{n\to\infty} \frac{|r^*(n) - \alpha n|}{(n \log n)^{1/4}} > 0.$$

Erdős contributed much to the theory of diophantine approximation. Recall that a sequence $(\phi_n) \subset [0, 1]$ is said to be *uniformly distributed* if for all $0 \leq \alpha < \beta \leq 1$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\kappa \leq n, \alpha \leq \phi_\kappa \leq \beta} 1 = \beta - \alpha. \tag{4}$$

Weyl proved in 1916 that $(\phi_n) \subset [0, 1]$ is uniformly distributed if, and only if,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} e^{2\pi i \kappa \phi_j} = 0$$

for every non-zero integer k, Needless to say, this necessary and sufficient condition gives no information about the speed in (4). To get some information about the speed of convergence, one needs a “finite” form of Weyl’s criterion. A finite form was given by van der Corput and Koksma in 1936, but a stronger conjecture of Koksma in his 1936 book on Diophantine approximation remained unproved until 1948, when Erdős and Turán proved the following remarkable theorem.

Let $\phi_1, \ldots, \phi_n \in [0, 1]$, and set $s_k = \sum j = 1^n e^{2\pi i \kappa \phi_j}$. Suppose that $|s_k| \leq \psi(k)$ for $k = 1, \ldots, m$. Then for all $0 \leq \alpha < \beta \leq 1$ we have

$$\left|(\beta - \alpha)n - \sum_{\alpha \leq \phi_j \leq \beta} 1\right| \leq C\left\{\frac{n}{m} + \sum_{k=1}^{n} \psi(k)/k\right\}$$

for some absolute constant C.

This result has had numerous applications, starting with the following beautiful theorem from the original Erdős-Turán paper. For $n \geq 2$, let

$$f(z) = z^n + a_{n_1} z^{n-1} + \cdots + a_1 z + a_0 = \prod_{j=1}^{n} (z - z_j)$$

be such that $z_j| \geq 1$ for every j. For $0 < \theta < 1$ set $M_\theta = \max_{|z|=\theta} |f(z)|$ and define $g(n,\theta), 2 \leq g(n,\theta) \leq n$, by

$$M_\theta / \sqrt{|a_0|} = e^{n/g(n,\theta)}.$$

Then for all $0 \leq \alpha < \beta \leq 2\pi$ we have

$$\left| \frac{\beta - \alpha}{2\pi} n - \sum_{\alpha \leq \arg\ z_j \leq \beta} 1 \right| < C \log(4/\theta) \frac{n}{\log g(n,\theta)},$$

where C is an absolute constant.

Note that if $M_\theta/\sqrt{|a_0|}$ is "not too large", say at most $e^{\sqrt{n}}$, then the error term above is $O(n/\log n)$.

Other applications were found by Egerváry and Turán, Környei, and others. When, in 1988, Laczkovich cracked Tarski's 50-year old problem on squaring the circle, he made substantial use of this theorem of Erdős and Turán from 1948.

There are very few people who have contributed more to the fundamental theorems in probability theory than Paul Erdős; here we shall state only a small fraction of the major results of Erdős in probability theory. The *law of the iterated logarithm* was proved around 1930 by Khintchine and Kolmogorov, with further contributions from Lévy. To state this fundamental result, let $X_1, X_2, \ldots$ be independent Bernoulli random variables, with $\mathbb{P}(X_n = -1) = \mathbb{P}(X_n = 1) = \frac{1}{2}$ for every n, and set $S_n = \sum_{i=1}^n X_i$. The law of the iterated logarithm states that $\limsup_{n\to\infty} S_n/\sqrt{2n \log\log n} = 1$ almost surely. Putting it another way, for $t \in [0,1]$, let $t = 0.\epsilon_1(t)\epsilon_2(t)\ldots$ be its dyadic expansion, or equivalently, set $\epsilon_n(t) = 0$ or 1 according as the integer part of $2^n t$ is even or odd. (Thus the variables $X_n(t) = 2\epsilon_n(t) - 1$ are as above.) Set $f_n(t) = \sum_{k=1}^n \epsilon_k(t) - \frac{n}{2}$. Then the law of iterated logarithm states that

$$\limsup_{n\to\infty} \frac{f_n(t)}{(\frac{n}{2} \log\log n)^{1/2}} = 1$$

for almost every $t \in (0,1]$.

Let $\phi(n)$ be a monotone increasing non-negative function defined for all sufficiently large integers. Following Lévy, this function $\phi(n)$ is said to belong to the *upper class* if, for almost all t,

$$f_n(t) \leq \phi(n)$$

provided n is sufficiently large, and it belongs to the *lower class* if, for almost all t, $f_n(t) > \phi(n)$ for infinitely many values of n. Then the law of the iterated logarithm states that $\phi(n) = (1+\epsilon)(\frac{n}{2}\log\log n)^{1/2}$ belongs to the upper class if $\epsilon > 0$, and to the lower class if $\epsilon < 0$.

In 1942, Erdős considerably sharpened this assertion when he proved that a function

$$\left(\tfrac{n}{2\log\log n}\right)^{1/2}\left\{\log\log n + \tfrac{3}{4}\log_3 n + \tfrac{1}{2}\log_4 n + \dots \right.$$
$$\left. \tfrac{1}{2}\log_{k-1} n + \left(\tfrac{1}{2}+\epsilon\right)\log_k n\right\}$$

belongs to the upper class if $\epsilon > 0$, and to the lower class if $\epsilon > 0$. (We write $\log_k$ for the k times iterated logarithm.) Not surprisingly, Erdős gave an elementary proof, and made no use of the results of Khintchine and Kolmogorov. Furthermore, as he indicated, the result could easily be extended to the case of Brownian motion. Some years later, Erdős returned to this topic in a joint paper with K.L. Chung.

In addition to the papers that were instrumental in creating *probabilistic number theory*, Erdős wrote some important papers with Mark Kac proving several basic results of *probability theory.* Let $X_1, X_2, \dots, X_n$ be independent random variables, each with mean 0 and expectation 1. As before, set $S_k = \sum_{l=1}^{k} X_l, k = 1, \dots, n$. In 1946, Erdős and Kac determined the limiting distributions of $\max_{1\le k\le n} S_k$ and $\max_{1\le k\le n}|S_k|$, which turned out to be independent of the distribution of the X_i.

Although this result was important, the *method of proof* was even more so: Erdős and Kac proved that if the theorem can be established for one particular sequence of independent random variables satisfying the conditions of the theorem, then the conclusion of the theorem holds for *all* sequences of independent random variables satisfying the conditions of the theorem. Erdős and Kac called this the *invariance principle.* Since then, this principle has been widely applied in probability theory.

Erdős and Kac promptly proceeded to apply their powerful invariance principle to extending a beautiful result of Paul Lévy, proved in 1939. To state this result, let $X_1, X_2, \dots$ be independent random variables, each with mean 0 and variance 1, such that the central limit theorem holds for the sequence. As before, let $S_k = X_1 + \cdots + X_k$, and let N_n be the number of S_k, $1 \le k \le n$, which are positive. Erdős and Kac proved in 1947 that, in this case,

$$\lim_{n\to\infty} \mathbb{P}(N_n/n < x) = \frac{2}{\pi}\arcsin x^{1/2}$$

for all $x, 0 \le x \le 1$. Thus N_n/n tends in distribution to the *arc sin distribution.*

What Paul Lévy had proved in 1939 is that this *arcsin law* holds in the binomial case $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = 1/2$.

In 1953 Erdős returned to this theme. In a joint paper with Hunt he proved that if $X_1, X_2, \dots$ are independent zero-mean random variables with the same continuous distribution which is symmetric about 0 then, almost surely,

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{\sin S_k}{k} = 0.$$

In his joint papers with Dvoretzky, Kac and Kakutani, Erdős contributed much to the theory of random walks and Brownian motion. For example, in 1940, Paul Lévy proved that almost all paths of a Brownian motion in the plane have double points. This was extended by Dvoretzky, Erdős and Kakutani in 1950: they proved that for $n \leq 3$ almost all paths of a Brownian motion in $\mathbb{R}^n$ have double points, but for $n \geq 4$ almost all paths of a Brownian motion in $\mathbb{R}^n$ are free of double points. In 1954, in a paper dedicated to Albert Einstein on his 75th birthday, Dvoretzky, Erdős and Kakutani returned to this topic, and proved that almost all paths of a Brownian motion in the plane have k-multiple points for every k, $k = 2, 3 \ldots$; in fact, for almost all paths the set of k-multiple points is dense in the plane.

Let us say a few words about classical measure theory. A subset of a metric space is of *first category* if it is a countable union of nowhere dense sets. There are a good many striking similarities between the class of nullsets and the class of sets of first category *on the line*. Indeed, both are σ-ideals (i.e. σ-rings closed under taking subsets), both include all countable sets and contain some sets of cardinality c, both classes have power 2^c, both classes are invariant under translation, neither class contains an interval, in fact, the complement of any set of either class is a set dense in $\mathbb{R}$, the complement of any set of either class contains a member of the class with cardinality c, and so on.

Of course, neither class includes the other; also, it is easily seen that $\mathbb{R} = A \cup B$, with A of first category and B a nullset. Nevertheless, the existence of numerous common properties suggests that the two σ-ideals are *similar* in the sense that there is a one-to-one mapping $f : \mathbb{R} \to \mathbb{R}$ such that $f(E)$ is a nullset if and only if E is of first category. In 1934 Sierpiński proved that this is indeed the case, provided we assume the continuum hypothesis. Sierpiński went on to ask whether the stronger assertion is also true that, assuming the continuum hypothesis, there is a function *simultaneously* mapping the two classes into each other. In 1943 Erdős answered this question in the affirmative.

Assuming the continuum hypothesis, there is a one-to-one map $f : \mathbb{R} \to \mathbb{R}$ such that $f(E)$ is a nullset if and only if E is of first category, and $f(E)$ is of first category if and only if E is a nullset. In fact, f can be chosen to satisfy $f = f^{-1}$.

Of the many results of Erdős in approximation theory, let us mention some beautiful theorems concerning *Lagrange interpolation.* Let $X = \{x_{i,n}\}, n = 1, 2, \ldots, i = 1, 2, \ldots, n,$ be a triangular matrix with

$$-1 \leq x_{1,n} < x_{2,n} < \cdots < x_{n,n} \leq 1 \tag{5}$$

for every n. The values $x_{i,n}$ are the *nodes* of interpolation. As usual, for $1 \leq k \leq n$, define the *fundamental polynomials* $l_{k,n}(X)$ as

$$l_{k,n}(X) = \prod_{i \neq k}(x - x_{i,n}) / \sum_{j=1}^{n} \prod_{i \neq j}(x_{k,n} - x_{i,n}).$$

so that $l_{k,n}(X)$ is the unique polynomial of degree $n-1$ with zeros at $x_{i,n}, i \neq k$, with $l_{k,n}(X_{k,n}) = 1$.

The *Lebesgue functions* and the *Lebesgue constants* of the interpolation are

$$\lambda_n(x) = \sum_{k=1}^{n} |l_{k,n}(x)| \text{ and } \lambda_n = \max_{-1 \leq x \leq 1} \lambda_n(x).$$

In fact, one frequently considers a generalization of the Lebesgue constants as well: for $-1 \leq a < b \leq 1$ set

$$\lambda_n(a, b) = \max_{a \leq x \leq b} \lambda_n(x),$$

so that $\lambda_n = \lambda_n(-1, 1)$.

Faber showed before the First World War that if X is any set of nodes satisfying (5) then $\lambda_n \geq \frac{1}{6} \log n$ for every n.

Much research was done on improving this inequality. After a series of papers with Turán, in 1942 Erdős proved the asymptotically best possible inequality that

$$\lambda_n > \frac{2}{\pi} \log n + O(1)$$

for every matrix X.

In 1931 Faber's inequality was extended by Bernstein, who proved that there is an absolute constant $c > 0$ such that

$$\lambda_n(a, b) \geq c \log n,$$

provided $-1 \leq a < b \leq 1$, and n is *sufficiently large, depending on* (a, b). In other words, the L_∞-norm of the restriction of $\lambda_n(X)$ to the interval (a, b) grows at least as fast as $c \log n$.

In a beautiful paper, written jointly with Szabados, Erdős proved in 1978 the *much stronger result* that a similar assertion holds for the *normalized* L_1*-norms.*

There is an absolute constant $c > 0$ *such that if* X *is an arbitrary system of nodes satisfying (5),* $-1 \leq a < b \leq 1$*, and* n *is sufficiently large, then*

$$\int_a^b \lambda_n(x)\, dx \geq c(b - a) \log n.$$

In the special case $a = -1$, $b = 1$, the result had been announced by Erdős in 1961, but the proof in the Erdős-Szabados paper in 1978 was along different lines and simpler.

Let us turn to a substantial extension of some classical results of Faber and Bernstein. Given a system X of nodes satisfying (5), and a function F on $[-1, 1]$, let

$$L_n(F, X, x) = \sum_{k=1}^{n} F(x_{k,n}) l_{k,n}(x)$$

be the nth *Lagrange interpolation polynomial of* F. Thus $L_n(F, X, x)$ is the unique polynomial of degree at most $n-1$ whose value at x_{kn} is $F(x_{kn}), 1 \leq k \leq n$. Extending a result of Faber from 1914, Bernstein proved in 1931 that, for every triangular matrix X satisfying (5), there is a continuous function F and a point x_0, $-1 \leq x_0 \leq 1$, such that

$$\limsup_{n\to\infty} |L_n(F, X, x_0)| = \infty. \tag{6}$$

In 1936, Géza Grünwald and Marcienkiewicz proved that if X is the "good" Chebyshev matrix then for some continuous function F relation (6) holds for almost every x_0, and 1978 Privalov proved the same assertion for the class of Jacobi matrices.

After these results concerning special classes of matrices, in 1980 Erdős and Vértesi proved the striking result that a similar assertion holds for *every matrix X satisfying* (5): *there is always a continuous function F such that* (6) *holds for almost every x_0*. The proof is intricate and ingenious.

To conclude our brief list of results on approximation theory, let us return to an early major result of Erdős. It has been known since Newton that interpolation polynomials can be used to approximate definite integrals of functions. Indeed, as proved by Stieltjes, if the nth row of X consists of the roots of the nth Legendre polynomial then

$$\lim_{n\to\infty} \int_{-1}^{1} l_n(F, X, x)\, dx = \int_{-1}^{1} F(x)\, dx$$

for every Riemann integrable function F.

Later this result was extended to other matrices X formed by the zeros of polynomials that were orthogonal in $[-1, 1]$ with respect to a weight function of the form $(l - x)^\alpha (1 + x)^\beta$ for some α and β. However, this was not known for any general class of weight functions; furthermore, the result of Stieltjes could not even be sharpened to

$$\lim_{n\to\infty} \int_{-1}^{1} |L_n(F, X, x) - F(x)|\, dx = 1.$$

In 1937, Erdős and Turán solved both problems. Let $p(x) \geq M > 0$ be Riemann integrable over $[-1, 1]$, and let $\omega_0(x), \omega_1(x), \ldots$ be orthogonal polynomials in $[-1, 1]$ with respect to $p(x)$, with $\omega_n(x)$ being a monic polynomial of degree n. Let $A_n,\ B_n$ be constants with $B_n \leq 0$ such that

$$R_n(x) = \omega_n(x) + A_n\omega_{n-1}(x) + B_n\omega_{n-2}(x)$$

has n different roots in $[-1,1]$, and let X be the set of nodes formed by the roots of the polynomials $R_1, R_2, \ldots$. Erdős and Turán proved that in this case every Riemann integrable function $F(x)$ on $[-1,1]$ satisfies

$$\lim_{n\to\infty}\int_{-1}^{1}|L_n(F,X,x)-F(x)|\,dx=0.$$

Much of Erdős's work in real analysis concerns so-called *Tauberian theorems.* The origin of these results is a theorem of Tauber stating that if $\sum a_n x^n \to s$ as $x \to 1-$, and $na_n \to 0$ as $n \to \infty$, then $\sum a_n$ is convergent (to sum s). Hence if $na_n \to 0$ and $\sum a_n$ is Cesàro summable then $\sum a_n$ is convergent. Soon after the turn of the century, Landau, Hardy and Littlewood founded a flourishing branch of analysis by making extensive use of deep results resembling this theorem of Tauber. These *Tauberian theorems* claim that if a series is summable with a certain method of summation *and* satisfies certain additional conditions then it is also summable with a weaker method of summation. For example, Hardy and Littlewood proved in 1911 that *if* $\sum a_n$ *is Borel summable* (i.e. $\lim_{x\to\infty} e^{-x}\sum s_n x^n/n!$ exists, where $s_n = a_1 + \ldots + a_n$) *and* $\sqrt{n}a_n \to 0$ *then* $\sum a_n$ *is convergent.* The second part of the elementary proof of the PNT was, essentially, such a Tauberian theorem.

Shortly after the elementary proofs of the Prime Number Theorem were found, Erdős proved that the PNT can be deduced from Selberg's fundamental formula *alone*, without any reference to other properties of the sequence of primes. What Erdős needed was the following Tauberian theorem: if $a_n \geq 0$ and $\sum_{k=1}^{n} a_k(S_{n-k}+k) = n^2 + O(n)$ then $s_n = n + O(1)$. Here, as before, $s_n = a_1 + \cdots + a_n$.

Hardy and Littlewood also considered *lacunary* series and proved, among others, that under certain lacunarity conditions Abel-summability implies summability. In 1943, Meyer-König proved a similar lacunarity theorem for Euler summability: *if* $\sum a_n$ *is Euler summable (i.e.* $\lim_{n\to\infty} 2^{-n}\sum_{k=0}^{n}\binom{n}{k}s_k$ *exists) and* $a_n = 0$ *except if* $n = n_i$*, where* $n_1 < n_2 < \cdots$ *satisfies* $n_{i+1}/n_i \geq c > 1$*, then* $\sum a_n$ *is convergent.* Meyer-König went on to conjecture the much stronger assertion that instead of $n_{i+1}/n_i \geq c > 1$ it suffices to demand that $n_{i+1} - n_i > A\sqrt{n_i}$ for some $A > 0$. In 1952 Erdős came very close to proving this conjecture: he showed that the assertion is true if $A > 0$ is *sufficiently large.*

Another Tauberian theorem of Erdős, proved with Feller and Pollard in 1949, is important in the theory of Markov chains. Let $p_0, p_1, \ldots$ be non-negative, with $\sum p_k = 1$ and $\mu = \sum kp_k$, and suppose that $P(z) = \sum_0^\infty p_k z^k$ is not a power series in z^t for any integer $t > 1$. Then $|P(z)| < 1$ for $|z| < 1$; in particular, $(1 - P(z))^{-l}$ is analytic in $|z| < 1$, say

$$\frac{1}{1-P(z)} = \sum_{k=0}^{\infty} u_k z^k.$$

The Erdős-Feller-Pollard Theorem states that $\lim_{k\to\infty} u_k = 1/\mu$ if $\mu < \infty$ and $u_k \to \infty$ if $\mu = \infty$. The theorem has important consequences in probability theory, and in 1951 de Bruijn and Erdős also used it to study recursion formulae.

In a beautiful paper written with Niven in 1948, Erdős extended a result relating the zeros of a complex polynomial to the zeros of its derivative. Among other results, Erdős and Niven proved that *if $r_1, r_2, \ldots, r_n$ are the zeros of a complex polynomial, and $R_1, R_2, \ldots, R_{n-1}$ are the zeros of its derivative then*

$$\frac{1}{n}\sum_{j=1}^{n} |z - r_j| \geq \frac{1}{n-1}\sum_{j=1}^{n-1} |z - R_j|$$

for every $z \in \mathbb{C}$, with equality if, and only if, all the, zeros r_j are on a half-line emanating from z.

In a difficult paper written with Szegő in 1942, Erdős tackled a problem concerning real polynomials. Extending Markov's classical theorem that if a polynomial f of degree n satisfies $|f(x)| \leq 1$ for $-1 \leq x \leq 1$, then $|f'(x)| \leq n^2$ for $-1 \leq x \leq 1$, Schur proved in 1919 that *if f is a polynomial of degree n with $|f(x)| \leq 1$ for $-1 \leq x \leq 1$, then $|f'(x_0) \leq \frac{1}{2}n^2$, provided $-1 \leq x_0 \leq 1$ and $f''(x_0) = 0$.* Writing m_n for the smallest constant that would do in the inequality above instead of $\frac{1}{2}$, Erdős and Szegő proved that for $n > 3$ the extremum $m_n n^2$ is attained for $x_0 = 1$ (or -1) and the so-called *Zolotarev polynomials.* This enabled Erdős and Szegő to determine $\lim_{n\to\infty} m_n$ as well (which turned out to be $0.3124\ldots$).

Whatever branch of mathematics Erdős works in, in spirit and attitude he is a *combinatorialist*: his strength is the hands-on approach, making use of ingenious elementary methods. Therefore it is not surprising that Erdős helped to shape twentieth-century combinatorics as no-one else: with his results, problems, and influence on people, much of combinatorics in this century owes its existence to Erdős.

One of the fundamental results in combinatorics is a theorem (to be precise, a pair of theorems) proved by F.P. Ramsey in 1930. Erdős was the first to realize the tremendous importance of this "super pigeon-hole principle", and did much to turn *Ramsey's finite theorem* into *Ramsey theory*, a rich branch of combinatorics, as witnessed by the excellent monograph of *Graham, Rothschild and Spencer.* In a seminal paper written in 1935, Erdős and his co-author, George Szekeres, tackled the following beautiful problem of Esther Klein: can we find, for a given n, a number $N(n)$ such that from any set of N points in the plane it is possible to select n points forming a convex polygon? Erdős and Szekeres showed that the existence of $N(n)$ is an easy consequence of Ramsey's theorem for finite sets. In fact, they discovered Ramsey's theorem for themselves, and were told only later that they had been beaten to it by Ramsey. It is remarkable that Ramsey, working in Cambridge, and Erdős and Szekeres, working in Budapest, arrived at the same result independently and in totally different ways, but within a few years of each other. As it happened, the proof given by Erdős and Szekeres is much simpler than the original, and it also gives much better upper bounds for the various *Ramsey numbers.* In particular, they proved that if $k, l \geq 2$ then

$$R(k,l) \leq \binom{k+l-2}{k-l},$$

where the Ramsey number $R(k,l)$ is the smallest value of n for which every graph of order n contains either a complete graph of order k or l independent vertices.

In view of the simplicity of the proof of the Erdős-Szekeres bound, it is amazing that over 50 years had to pass before the bound above was improved appreciably. In 1986 Rödl showed that there is a positive constant $c > 0$ such that

$$R(k,l) \leq \binom{k+l-2}{k-1} / \log^c(k+l),$$

and, simultaneously and independently, Thomason replaced the power of the logarithm by a power of $k+l$. To be precise, Thomason proved that

$$R(k,l) \leq k^{-1/2+A\sqrt{\log k}} \binom{k+l-2}{k-l}$$

for some absolute constant $A > 0$ and all k, l with $k \geq l \geq 2$.

Concerning the lower bounds for $R(k,l)$, especially $R(k,k)$, the situation seems to be even more peculiar. It is not even obvious that $R(k,k)$ is not bounded from above by a polynomial of k. Indeed, it was again Erdős, who gave, in 1947, the following lower bound: *if* $\binom{n}{k}2^{-\binom{k}{2}+1} < 1$ *then* $R(k,k) > n$. Erdős's proof is remarkable for its simplicity and its influence on combinatorics. Although there are very few mathematicians who do not know this proof, we present it here, since it is delightful and brief. Consider the set of all $2^{\binom{n}{2}}$ graphs on $\{1, 2, \ldots, n\}$. What is the average number of complete subgraphs of order k? Since each of the $\binom{n}{k}$ possible complete subgraphs of order k is contained in $2^{\binom{n}{2}-\binom{k}{2}}$ of our graphs, the average is $\binom{n}{k}2^{-\binom{k}{2}} < 1/2$. Similarly, the average number of complete subgraphs of order k in the *complements* of our graphs is also $\binom{n}{k}2^{-\binom{k}{2}} < 1/2$. Consequently, there is some graph G on $\{1, 2, \ldots, n\}$ such that neither G nor its complement $\bar{G}$ contains a complete graph of order k. Hence $R(k,k) > n$, as claimed.

It took over three decades to improve this wonderfully simple lower bound: in 1977 Spencer showed that an immediate consequence of the Erdős-Lovász Local Lemma is that

$$R(k,k) \geq k2^{k/2}\left(\frac{\sqrt{2}}{e} + o(1)\right),$$

which is only about a factor 2 improvement. Needless to say, the combinatorialists are eagerly awaiting a breakthrough that more or less eliminates the gap between the upper and lower bounds for $R(k,k)$, but judging by the speed of improvements on the original bounds of Erdős, we are in for a long wait.

Erdős did not fail to notice that the other theorem of Ramsey from 1930, concerning infinite sets, also had a tremendous potential. In the 1950s and 1960s, mostly with his two great collaborators, Rado and Hajnal, Erdős revolutionized combinatorial set theory.

Ramsey's theorem concerning infinite sets, in its simplest form, states that if G is an infinite graph then *either* G *or* its complement $\bar{G}$ contains an infinite complete graph. While this is very elegant, in order to express more complicated results succinctly, it is convenient to rely on the Erdős-Rado *arrow notation.* Given cardinals r, a and b_γ, $\gamma \in \Gamma$, where Γ is an indexing set, the partition relation

$$a \to (b_\gamma)^r_{\gamma\in\Gamma}$$

is said to hold if, given any partition $\bigcup_{\gamma\in\Gamma} I_\gamma$, of the set $A(r)$ of all subsets of cardinality r of a set A with $|A| = a$, there is a $\gamma \in \Gamma$ and a subset B_γ of A with $|B_\gamma| = b_\gamma$ such that $B^r_\gamma \subset I_\gamma$. The same notation is used to express the analogous assertion when some or all the symbols r, a and b_γ denote *order types* rather than cardinalities. If Γ is a small set then one tends to write out all the b_γs.

Thus, in this notation, the infinite Ramsey theorem is that

$$\aleph_0 \to (\aleph_0, \aleph_0)^r$$

for every integer r, with $r = 2$ being the case of graphs.

In 1933, Sierpiński proved that there is a graph of cardinality $2^{\aleph_0}$ which has *neither* an uncountable complete graph *nor* an uncountable independent set:

$$2^{\aleph_0} \not\to (\aleph_1, \aleph_1)^2,$$

so the "natural" extension of Ramsey's theorem is false. Sierpiński's result says that one can partition the pairs of real numbers in such a way that every uncountable subset of $\mathbb{R}$ contains a pair from both classes. This partition somewhat resembles a Bernstein subset of $\mathbb{R}$.

If we are happy with one of the classes being merely *countably infinite* then Ramsey's theorem extends to all cardinals. This was proved in 1941 by Dushnik and Miller for regular cardinals, and extended by Erdős to singular cardinals. Thus,

$$\kappa \to (\kappa, \aleph_0)^2.$$

In the language of graphs, this means that if a graph on κ vertices does not contain a complete subgraph on κ vertices then it contains an infinite independent set.

The Erdős-Rado collaboration on partition problems started in 1949. One of their first results is an attractive assertion concerning $\mathbb{Q}$, the set of rationals. *If G is a graph on $\mathbb{Q}$ then either G or its complement $\bar{G}$ contains a complete graph whose vertex set is dense in an interval.* Years later this was considerably extended by Galvin and Laver.

After a good many somewhat ad hoc results, in 1956 Erdős and Rado gave the first systematic treatment of "arrow relations" for cardinals; in their fundamental paper, "*A partition calculus in set theory*", they set out to establish a 'calculus' of partitions. Among many other results, they proved that *if $\lambda \geq 2$ and $\rho \geq \aleph_0$ are cardinals then*

$$(\lambda^\rho)^+ \to ((\lambda^\rho)^+, (\rho^+)_\rho)^2,$$

but

$$\lambda^\rho \nrightarrow ((\lambda \cdot \rho)^+, \rho^+)^2.$$

In the special case $\lambda = 2$ and $\rho = \aleph_0$, the last relation is precisely Sierpiński's theorem.

In proving their positive results, Erdős and Rado used so called "*tree arguments*", arguments resembling the usual proof of Ramsey's infinite theorem, but relying on sequences of *transfinite length.* Another important ingredient is a *stepping-up lemma*, enabling one to deduce arrow relations about larger cardinals from similar relations about smaller ones. Thus the trivial relation $\aleph_1 \to (\aleph_1)^1_{\aleph_0}$ implies that

$$(2^{\aleph_0})^+ \to (\aleph_1)^2_{\aleph_0}.$$

In 1965, in a monumental paper "*Partition relations for cardinal numbers*", running to over 100 pages, Erdős, Hajnal and Rado presented an almost complete theory of the partition relation above for cardinals, assuming the *generalized continuum hypothesis.* For years after its publication, its authors lovingly referred to their paper as GTP, the *Giant Triple Paper.*

In fact, Erdős had taken an interest in extensions of Ramsey's theorem for infinite sets well before the Dushnik-Miller result appeared. In 1934, in a letter to Rado, he asked whether if we split the countable subsets of a set A of cardinality a into two classes then there is an infinite subset B of A, all of whose countable subsets are in the same class. In the arrow notation, does $a \to (\aleph_0, \aleph_0)^{\aleph_0}$ hold for some cardinal a? Almost by return mail, Rado sent Erdős his counterexample (which is, by now, well known), constructed with the aid of the axiom of choice. Later this question led to the study of partitions restricted in some way, including the study of Borel and analytic partitions, and to many beautiful results of Galvin, Mathias, Prikry, Silver, and others.

The first results concerning partition relations for *ordinals* were also obtained in 1954. In November 1954, on his way to Israel, Erdős passed through Zürich. He told his good friend Specker that he was offering \$20 for a proof or disproof of the conjecture of his with Rado that $\omega^2 \to (\omega^2, n)^2$. Within a few days, Specker sent Erdős a proof which is, by now, well known. Erdős had high hopes of building on Specker's proof to deduce that $\omega^n \to (\omega^n, 3)^2$ for every integer $n \geq 3$, but could prove only $\omega^{2n} \to (\omega^{n+1}, 4)^2$. A little later Specker produced an example showing that

$$\omega^n \not\to (\omega^n, 3)^2$$

for every integer $n > 3$.

Neither Specker's proof, nor his (counter)example worked for

$$\omega^\omega \to (\omega^\omega, 3)^2.$$

Erdős rated this problem so highly that eventually, in the late 1960s, he offered $250 for a proof or counterexample. The prize was won by Chang in 1969 with a very complicated *proof*, which was later simplified by Milner and Jean Larson.

The remaining problems are far from being easy, and Erdős is now offering $1,000 for a complete characterization of the values of α and n for which

$$\omega^{\omega^\alpha} \to (\omega^{\omega^\alpha}, n)^2$$

holds.

The theory of partition relations for ordinals took off after Cohen introduced *forcing methods* and Jensen created his theory of the constructible universe. Not surprisingly, in many questions "independence reared its ugly head", as Erdős likes to say. In addition to Erdős, Hajnal and Rado, a host of excellent people working on combinatorial set theory contributed to the growth of the field, including Baumgartner, Galvin, Larson, Laver, Máté, Milner, Prikry and Shelah. An account of most results up to the early 1980s can be found in the excellent monograph "*Combinatorial Set Theory: Partition Relations for Cardinals*" by Erdős, Hajnal, Máté and Rado, published in 1984.

In 1940 Turán proved a beautiful result concerning graphs, vaguely related to Ramsey's theorem. *For $3 \leq r \leq n$, every graph of order n that has more edges than an $(r-1)$-partite graph of order n contains a complete graph of order r.* It was once again Erdős who, with Turán, Gallai and others, showed that Turán's theorem is just the starting point of a large and lively branch of combinatorics, *extremal graph theory.* In order to formulate the quintessential problem of extremal graph theory, let us recall some notation. As usual, we write $|G|$ for the *order* (i.e. number of vertices) and $e(G)$ for the *size* (i.e. number of edges) of a graph G. Given graphs G and H, the expression $H \subset G$ means that H is a *subgraph* of G. Let F be a fixed graph, usually called the *forbidden graph.* Set

$$ex(n; F) = \max\{e(G) : |G| = n \text{ and } F \not\subset G\}.$$

and

$$EX(n;F) = \{G : |G| = n, e(G) = ex(n;F), \text{ and } F \not\subset G\}.$$

We call $ex(n;F)$ the *extremal function*, and $EX(n;F)$ the set of *extremal graphs* for the *forbidden graph* F. Then the basic problem of extremal graph theory is to determine, or at least estimate, $ex(n;F)$ for a given graph F and, at best, to determine $EX(n;F)$. From here it is but a short step to the problem of excluding several forbidden graphs, i.e. to the functions $ex(n;F_l,\ldots,F_k)$ and $EX(n,F_1,\ldots,F_k)$ for a finite family $F_1,\ldots,F_k$ of forbidden graphs.

Writing K_r for the complete graph of order r, and $T_k(n)$ for the unique k-partite graph of order n and maximal size (so that $T_k(n)$ is the *k-partite Turán graph* of order n), Turán proved, in fact, that $EX(n;K_r) = \{T_{r-1}(n)\}$, i.e. $T_{r-l}(n)$ is the *unique* extremal graph, and so $ex(n;K^\Gamma) = t_{r-1}(n)$, where $t_{r-1}(n) = e(T_{r_1}(n))$ is the size of $T_{r-1}(n)$.

As it happens, Erdős came very close to founding extremal graph theory *before* Turán proved his theorem: in 1938, in connection with sequences of integers no one of which divided the product of two others, proved that for a *quadrilateral* C^4 we have $ex(n;C^4) = O(n^{3/2})$. However, at the time Erdős failed to see the significance of problems of this type: one of the very few occasions when Erdős was "blind".

Before we mention some of the important results of Erdős in extremal graph theory, let us remark that in 1970 (!) Erdős proved the following beautiful extension of Turán's theorem (so the *rest of the world* had been blind). Let G be a graph without a K_r, with degree sequence $(d_i)_{\mathbf{1}}^n$. Then there is an $(r-1)$-partite graph G^* (which, a fortiori, contains no K_r either) with degree sequence $(d_i^*)_1^n$, such that $d_i \leq d_i^*$ for every i. In this theorem, the achievement is in the audacity of *stating* the result: once it is stated, the proof follows easily.

Erdős conjectured another extension of Turán's theorem which was proved in 1981 by Bollobás and Thomason. The conjecture was sharpened by Bondy. Let $|G| = n$ and $e(G) > t_{r-1}(n)$. Then *every vertex x of maximal degree d in G is such that the neighbours span a subgraph with more than* $t_{r-2}(d)$ edges. In this instance it is also true that once the full assertion has been made, the proof is just about trivial; in fact, it is simply a minor variant of the proof of the previous theorem of Erdős.

It is fitting that the *fundamental theorem of extremal graph theory* is a result of Erdős, and his collaborator, Stone. Note that, by Turán's theorem, the maximal size of a K_r-free graph of order n is about $\frac{r-2}{r-1}\binom{n}{2}$, in fact, trivially,

$$\frac{r-2}{r-1}\binom{n}{2} \leq t_r(n) \leq \frac{r-2}{r-1}\frac{n^2}{2}.$$

Writing, as usual, $K_r(t)$ for the *complete r-partite graph with t vertices in each class*, Erdős and Stone proved in 1946 that if $r \geq 2, t \geq 1$ and $\epsilon > 0$ are fixed and n is sufficiently large then *every graph of order n and size at least* $\left(\frac{r-2}{r-1}+\epsilon\right)\binom{n}{2}$ *contains a* $K_r(t)$. In other words, *even ϵn^2 more edges* than can be found in a Turán graph *guarantee not only a K_r but a "thick" K_r*, one in which every vertex has been replaced by a group of t vertices.

Prophetically, Erdős and Stone entitled their paper "*On the structure of linear graphs*"; this is indeed the significance of the paper: it not only gives us much information about the *size* of extremal graphs, but it is also the starting point for the study of the *structure* of extremal graphs. If F is a non-empty r-chromatic graph, i.e. $\chi(F) = r \geq 2$, then, precisely by the definition of the chromatic number, F is not a subgraph of $T_{r-l}(n)$, so $ex(n;F) \geq t_{r-l}(n) \geq \frac{r-2}{r-1}\binom{n}{2}$. On the other hand, $F \subset K_r(t)$ if t is large enough (say, $t \geq |F|$), so if $\epsilon > 0$ and n is large enough then

$$ex(n,F) < \left(\frac{r-2}{r-1}+\epsilon\right)\binom{n}{2}.$$

In particular, if $\chi(F) = r \geq 2$ then

$$\lim_{n\to\infty} ex(n,F)/\binom{n}{2} = \frac{r-2}{r-1},$$

that is the asymptotic density of the extremal graphs with forbidden subgraph $F(r-2)/(r-1)$. Needless to say, the same argument can be applied to the problem of forbidding any finite family of graphs: given graphs $F_1, F_2, \ldots, F_k$, with min $\chi(F_i) = r \geq 2$, we have

$$\lim_{n\to\infty} ex(n,F_1,\ldots,F_k)/\binom{n}{2} = \frac{r-2}{r-1}.$$

Starting in 1966, in a series of important papers Erdős and Simonovits went considerably further than noticing this instant consequence of the Erdős-Stone theorem. Among other results, Erdős and Simonovits proved that if $G \in EX(n;F)$, with $\chi(F) = r \geq 2$, then G can be obtained from $T_{r-1}(n)$ by the addition and deletion of $o(n^2)$ edges. Later this was refined to several results concerning the *structure* of extremal graphs. Here is an example, showing how very close to a Turán graph an extremal graph has to be. *Let $F_1, \ldots, F_k$ be fixed graphs, with $r = \min\chi(F_i)$, and suppose that F_1 has an r-colouring in which one of the colour classes contains t vertices. Let $G_n \in EX(n;F_1,\ldots,F_k)$. Then, as $n \to \infty$,*

(i) The minimal degree of G_n is $((r-2)/(r-1)+o(1))n$,

(ii) The vertices of G_n can be partitioned into $r-1$ classes such that each vertex is joined to at most as many vertices in its own class as in any other class,

(iii) For every $\epsilon > 0$ there are at most $c_\epsilon = c(\epsilon;F_1,\ldots,F_k)$ vertices joined to at least ϵn vertices in their own class,

(iv) There are $0(n^{2-1/t})$ edges joining vertices in the same class,

(v) Each class has $n/(r-1)+O(n^{1-1/2t})$ vertices.

Returning to the Erdős-Stone theorem itself, let us remark that Erdős and Stone also gave a bound for the speed of growth of t for which $K_r(t)$ is guaranteed to be a subgraph of every graph with n vertices and at least $((r-2)/(r-1)+\epsilon)\binom{n}{2}$ edges.

Let us write $t(n,r,\epsilon)$ for the maximal value of t that will do. Erdős and Stone proved that $t(n,r,\epsilon) \geq (\log_{r-1}(n))^{1-\delta}$ for fixed $r \geq 2$, $\epsilon > 0$ and $\delta > 0$, and large enough n, where $\log_k(n)$ is the k times iterated logarithm of n. They also thought it plausible though unproved that $\log_{r-1}(n)$ would be about the "best" value.

This assertion was conjectured in several subsequent papers by Erdős, so it was rather surprising when, in 1973, Erdős and Bollobás proved that, for fixed $r \geq 2$ and $0 < \epsilon < I/(r-1)$ the correct *order* of $t(n,r,\epsilon)$ is, in fact, $\log n$.

A little later, in 1976, Erdős, Bollobás and Simonovits sharpened this result, and the dependence of the implicit constant on r and ϵ was finally settled by Chvátal and Szemerédi, who proved that *there are positive absolute constants c_1 and c_2 such that*

$$c_1 \frac{\log n}{\log(1/\epsilon)} \leq t(n,r,\epsilon) \leq c_2 \frac{\log n}{\log(1/\epsilon)}$$

whenever $r \geq 2$ and $0 < \epsilon < 1/(r-1)$.

For a bipartite graph F, the general Erdős-Stone theorem is not sensitive enough to provide non-trivial information about $ex(n;F)$, since all it tells us is that $ex(n;F) = o(n^2)$. It was, once again, Erdős, who proved several of the fundamental results about $ex(n;F)$ when F is bipartite. In particular, he proved with Gallai in 1959 that for a path P_l of length l we have

$$ex(n;p_l) \leq \frac{l-1}{2}n.$$

By taking vertex-disjoint unions of complete graphs of order l, we see that this inequality is, in fact, an equality whenever $l|n$. The determination of $ex(n;P_l)$ was completed by Faudree and Schelp in 1973.

Another ground-breaking result of Erdős concerns supersaturated graphs, i.e. graphs with *slightly* more edges than the extremal graph. An unpublished result of Rademacher from 1941 claims that a graph of order n with *more than* $\lfloor n^2/4 \rfloor = t_2(n)$ edges contains not only one triangle but at least $\lfloor n/2 \rfloor$ triangles. In 1962 Erdős extended this result considerably; he showed that for some constant $c > 0$ every graph with n vertices and $\lfloor n^2/4 \rfloor + k$ edges has at least $k\lfloor n/2 \rfloor$ triangles, provided $0 \leq k \leq cn$. Later, this led to a spate of related results by Erdős himself, Moon and Moser, Lovász and Simonovits, Bollobás and others.

Erdős, the problem-poser par excellence, could not fail to notice how much potential there is in combining Ramsey-type problems with Turán-type problems.

The extremal graph for K_r, namely the Turán graph $T_{r-1}(n)$, is *stable* in the sense that if a K_r-free graph G on n vertices has almost as many edges as $T_{r-1}(n)$, then G is rather similar to $T_{r-1}(n)$; in particular, it has a large independent set. Putting it another way, if G is K_r-free and does not have a large independent set then $e(G)$ is much smaller than $t_{r-l}(n)$.

This observation led Erdős and Sós to the prototype of *Ramsey-Turán problems.* Given a graph H and a natural number l, let $f(n; H, l)$ be the smallest integer m for which every graph of order n and size more than m *either* contains H as a subgraph, *or* has at least l independent vertices.

Erdős and Sós were especially interested in the case $H = K_r$ and $l = o(n)$, and so in the function

$$l(r) = \lim_{\epsilon\to 0} \lim_{n\to\infty} f(n; K_r, \lfloor \epsilon n \rfloor)/\binom{n}{2}.$$

It is easily seen that $l(3) = 0$, and in 1969 Erdős and Sós proved that $l(r) = (r-3)/(r-1)$ whenever $r \geq 3$ is odd.

The stumbling block in determining $l(r)$ for even values of r was the case $r = 4$. Szemerédi proved in 1972 that $f(4) \leq 1/4$, but it seemed likely that $f(4)$ is, in fact, 0. Thus it was somewhat of a surprise when in 1976 Erdős and Bollobás constructed a graph on a k-dimensional sphere that shows $f(4) = 1/4$. In fact, this graph is rather useful in a number of other questions as well; it would be desirable to construct an infinite family of graphs in this vein.

Erdős, Hajnal, Sós and Szemerédi completed the determination of $l(r)$ in 1983 when they showed that $l(r) = (3r-10)/(3r-4)$ whenever $r \geq 4$ is even. Note that the condition that our graph does not have more than $o(n)$ independent vertices, *does* force the graph to have considerably fewer edges: Turán's theorem tells us that without the condition on the independence number the limit would be $(r-2)/(r-1)$.

Erdős was still a young undergraduate, when he became interested in *extremal problems* concerning *set systems.* It all started with his fascination with *Sperner's theorem* on the maximal number of subsets of a finite set with no subset contained in another. Sperner proved in 1928 that if the ground set has n elements then the maximum is attained by the system of all $\lfloor n/2 \rfloor$-subsets. Erdős was quick to appreciate the beauty and importance of this result, and throughout his career frequently returned to problems in this vein.

In 1939, Littlewood and Offord gave estimates of the number of real roots of a random polynomial of degree n for various probability spaces of polynomials. In the course of their work, they proved that for some constant $c > 0$, if $z_l, z_2, \ldots, z_n$ are complex numbers with $|z_i| \geq 1$ for each i, then of the 2^n sums of the form $\pm z_1 \pm z_2 \pm \ldots \pm z_n$ no more than

$$cr2^n(\log n)n^{-1/2} \tag{7}$$

fall into a circle of radius r.

On seeing the result, Erdős noticed immediately the connection with Sperner's theorem, especially in the real case. In fact, Sperner's theorem implies the following best possible assertion. *If $x_1, \ldots, x_n$ are real numbers of modulus at least 1, then no more than $\binom{n}{n/2}$ of the sums $\pm x_1 \pm x_2 \pm \ldots \pm x_n$ fall in an open interval of length* 2. From here it was but a short step to show that the maximal number of sums that can fall in an open interval of length $2r$ is precisely the sum of the r largest binomial coefficients $\binom{n}{k}$.

Concerning the complex case, Erdős improved the Littlewood-Offord bound (7), to an *essentially* best possible bound, by removing the factor $\log n$. More importantly, Erdős conjectured that the Sperner-type bound holds not only for real numbers, as he noticed, by for vectors of norm at least 1 in a Hilbert space. This beautiful conjecture was proved 20 years later by Kleitman and, independently, by Katona. In 1970, Kleitman gave a strikingly elegant proof of the even stronger assertion that if $x_1, x_2, \ldots, x_n$ are vectors of norm at least 1 in a normal space, then there are at most $\binom{n}{n/2}$ sums of the form $\pm x_1 \pm x_2 \pm \ldots \pm x_n$ such that any two of them are at a distance less than 2.

With Offord, in 1956 Erdős tackled the original Littlewood-Offord problem concerning random polynomials. They concentrated on the class of 2^n polynomials of the form $f_n(x) = \pm x^n \pm x^{n-1} \pm \ldots \pm 1$. Refining the result of Littlewood and Offord, they proved that, with the exception of $o((\log\log n)^{-1/3})2^n$ polynomials, the equations $f_n(x) = 0$ have

$$\frac{2}{\pi}\log n + o\big((\log n)^{2/3}\log\log n\big)$$

real roots.

Let us turn to some results concerning *hypergraphs*, the objects most frequently studied in the extremal theory of set systems. For a positive integer r, an *r-uniform hypergraph*, also called an *r-graph* or *r-uniform set system*, is a pair $(X, \mathcal{A})$, where X is a set and $\mathcal{A}$ is a subset of $X^{(r)}$, the set of all r-subsets of X. *The vertex set* of this hypergraph is X, and $\mathcal{A}$ is the set of *(hyper) edges*. The vertex set is frequently taken to be $[n] = \{1, \ldots, n\}$, and our hypergraph is often referred to as a "collection of r-subsets of $[n]$". For $r = 2$ an r-graph is just a graph. Although r-graphs seem to be innocuous generalizations of graphs, they are much more mysterious than graphs.

The most influential paper of Erdős on hypergraphs, "*Intersection theorems for systems of finite sets*", written jointly with Chao Ko and Richard Rado, has a rather curious history. The research that the paper reports on was done in 1938 in England. However, at the time there was rather little interest in pure combinatorics, and the authors went their different ways: Erdős went to Princeton, Chao Ko returned to China, and Rado stayed in England. As a result of this, the paper was published only in 1961.

In its simplest form, the celebrated Erdős-Ko-Rado theorem states the following. Let $\mathcal{A} \subset [n]^{(r)}$, that is let *$\mathcal{A}$ be a collection of r-subsets of the set $[n] = \{1, 2, \ldots, n\}$. If $n \geq 2r$ and $\mathcal{A}$ is intersecting, that is if $A \cap B \neq \emptyset$*

whenever $A, B \in \mathcal{A}$*, then* $|A| \leq \binom{n-1}{r-1}$. Taking $\mathcal{A} = \{A \in [n]^{(r)} : 1 \in A\}$, we see that the bound is best possible. This result has been the starting point of much research in combinatorics. By now there are a good many proofs of it, including a particularly ingenious and elegant proof found by Katona in 1972.

The more general Erdős-Ko-Rado theorem states that if $1 \leq t \leq r$*,* $\mathcal{A} \subset [n]^{(r)}$ *and* $\mathcal{A}$ *is* t*-intersecting, that is if* $|A \cap B| \geq t$ *whenever* $A, B \in \mathcal{A}$*, then* $|\mathcal{A}| \leq \binom{n-t}{r-t}$*, provided* n *is large enough, depending on* r *and* t*.*

In the original paper it was proved that $n \geq t + (r-t)\binom{r}{t}^3$ will do. Once again, the bound on $|\mathcal{A}|$ is best possible, as shown by a collection of r-subsets containing a fixed t-set. But the bound on n given by Erdős, Ko and Rado is far from being best possible. For $i = 0, 1, \ldots, r-t$, let

$$\mathcal{A}_i = \{A \in [n]^{(r)} : \ |A \cap [t+2i]| \geq t+i\}.$$

It is clear that each $\mathcal{A}_i$ is a t-intersecting system, and it so happens that $|\mathcal{A}_1| > |\mathcal{A}_0|$ if $n < (t+1)(r-t+1)$. Thus the best we can hope for is that the Erdős-Ko-Rado bound $\binom{n-t}{r-t}$ holds whenever $n \geq (t+1)(r-t+1)$.

It took many years to prove that this is indeed the case. In 1976 Frankl came very close to proving it: he showed it for all t except the first few values, namely for all $t \geq 15$. Finally, by ingenious arguments involving vector spaces, Richard Wilson gave a complete (and self-contained) proof of it in 1984.

The Erdős-Ko-Rado theorem inspired so much research that in 1983 Deza and Frankl considered it appropriate to write a paper entitled "*The Erdős-Ko-Rado theorem – 22 years later*".

The first Erdős-Rado paper that appeared in print, in 1950, contained their *canonical Ramsey theorem for* r*-graphs*, to be precise, for $\mathbb{N}^{(r)}$, the collection of r-subsets of $\mathbb{N}$. This is yet another Erdős paper which had much influence on the development of Ramsey theory, especially through the work of Graham, Leeb, Rothschild, Spencer, Nešetřil, Rödl, Deuber, Voigt and Prömel. To formulate this result, let $X \subset \mathbb{N}$ or, for that matter, let X be any ordered set, and let r be an integer. A partition of $X^{(r)}$ into some classes (finitely or infinitely many) is said to be *canonical* if there is a set $I \subset [r]$ such that two r-sets $A = (a_1, \ldots, a_r), B = (b_1, \ldots, b_r) \in X^{(r)}$ belong to the same class if, and only if, $a_i = b_i$ for every $i \in I$. Here we assumed that $a_1 < \ldots < a_r$ and $b_1 < \ldots < b_r$. Thus in a canonical partition, A and B belong to the same class if, and only, for each $i \in I$, the ith element of A is identical with the ith element of B.

Note that if all we care about is whether two r-sets belong to the same class or not, then for every ordered set X with more than r elements, $X^{(r)}$ has precisely 2^r distinct canonical partitions, one for each subset I of $[r]$. If X is infinite then there is only one canonical partition with *finitely* many classes: this is the canonical partition belonging to $I = \emptyset$, in which all r-sets belong to the same class.

The Erdős-Rado canonical Ramsey theorem claims that *if we partition* $\mathbb{N}^{(r)}$ *into any number of classes then there is always an infinite sequence of integers* $x_1 < x_2 < \ldots$ *on which the partition is canonical.* If $\mathbb{N}^{(r)}$ is partitioned into only finitely many classes then, as it was just remarked, on $X = \{x_1, x_2, \ldots\}$ the canonical distribution belongs to $I = \emptyset$, that is all r-sets of X belong to the same class. Thus Ramsey's theorem for infinite sets is an instant consequence of the Erdős-Rado result.

The canonical Ramsey theorem has attracted much attention: it has been extended to other settings many times over, notably by Erdős, Rado, Galvin, Taylor, Deuber, Graham, Voigt, Nešetřil and Rödl.

In 1952 Erdős and Rado gave a rather good upper bound for the Ramsey number $R^{(r)}(n, \ldots, n)_k = R^{(r)}(n; k)$ concerning (r)-graphs: $R^{(r)}(n; k)$ is the minimal integer m such that if $[m]^{(r)}$ is partitioned into k classes then there is always a subset $N \in [m]^{(n)}$ all of whose r-sets are in the same class. Putting it slightly differently, $R^{(r)}(n; k)$ is the minimal integer m such that every k-colouring of the edges of a complete r-graph of order m contains a monochromatic complete r-graph of order n. Writing $\exp_s k$ for the s times iterated exponential so that $\exp_1 k = k, \exp_2 k = k^k$, and $\exp_3 k = k^{k^k}$, Erdős and Rado proved that

$$R^{(r)}(n; k)^{1/n} < \exp_{r-1} k.$$

Although this seems a rather generous bound, in their GTP, Erdős, Hajnal and Rado proved that for $r \geq 3$ we have

$$R^{(r)}(n; k)^{1/n} > \exp_{r-2} k.$$

Erdős believes that the right order is given by the upper bound.

An important question concerning hypergraphs is to what extent the Erdős-Stone theorem can be carried over to them. The density $d(G)$ of an r-graph $G = (X, \mathcal{A})$ of order n is

$$d(G) = |\mathcal{A}| / \binom{n}{r},$$

so that $0 \leq d(G) \leq 1$ for every hypergraph. Call $0 \leq \alpha \leq 1$ a *jump-value* for r-graphs if there is a $\beta = \beta_r(\alpha) > \alpha$ such that for every $\alpha' > \alpha$ and positive integer m there is an integer n such that every r-graph of order at least n and density at least α' contains a subgraph of order at least m and density at least β. An immediate consequence of the Erdős-Stone theorem is that *every* α in the range $0 \leq \alpha < 1$ is a jump-value.

In 1965 Erdős proved that, for every $r \geq 1$, 0 is a jump-value for r-graphs and, in fact, $\beta_r(0) = r!/r^r$ will do. This is because if $\alpha' > 0$, $m \geq 1$ and n is sufficiently large then every r-graph of order at least n and density at least α' contains a $K_r^{(r)}(m)$, a complete r-partite r-graph with m vertices in each class. Clearly,

$$d(K_r^{(r)}(m)) = m^r / \binom{rm}{r} \sim \frac{r!}{r^r}.$$

This seems to indicate that every α, $0 \leq \alpha < 1$, is a jump-value for r-graphs for every $r \geq 3$ as well. Nevertheless, for years no progress was made with the problem so that, eventually, Erdős was tempted to offer \$1,000 for a proof or disproof of this assertion. In 1984, Frankl and Rödl won the coveted prize when they showed that $1 - l^{-(r-1)}$ is *not* a jump-value for r-graphs if $r \geq 3$ and $l > 2r$. In spite of this beautiful result, we are *very far* from a complete characterization of jump-values.

The important topic of Δ-systems was also initiated by Erdős. A family of sets $\{A_\gamma\}_{\gamma \in \Gamma}$ is called a Δ-*system* if any two sets have precisely the same intersection, that is if the intersection of any two of them is $\bigcap_{\gamma \in \Gamma} A_\gamma$. Given cardinals n and p, let $f(n;p)$ be the maximal cardinal m for which every collection of m sets, each of size (at most) n, contains a Δ-system of size p. In 1960, Erdős and Rado determined $f(n;p)$ for infinite cardinals but found that surprising difficulties arise when n and p are finite. Even the case $p = 3$ seems very difficult, so that they could not resolve their conjecture that

$$f(n;3) \leq c^n \tag{8}$$

for some constant c.

As Erdős and Rado pointed out, it is rather trivial that $f(n;3) > 2^n$. Indeed, let $\mathcal{A}$ be the collection of n-subsets of a $2n$-set $\{x_1, \ldots, x_n, y_1, \ldots, y_n\}$ containing precisely one of x_i and y_i for each i. Then $|\mathcal{A}| = 2^n$ and $\mathcal{A}$ does not contain a Δ-system on three sets.

Abbott and Hanson have improved this bound to $f(n;3) > 10^{n/2}$, but due to the very slow progress with the upper bound, for years now Erdős has offered \$1,000 for a proof or disproof of (8). Recently, Kostochka has made some progress with the problem when he proved that $f(n;3) < n! \left(\frac{c \log\log n}{\log\log\log n} \right)^{-n}$.

Let us turn to a flourishing area of mathematics that was practically created by Erdős. This is the theory of *random graphs*, started by Erdős and then, a little later, founded by Erdős and Rényi,

Throughout his career, Erdős had a keen eye for problems likely to yield to either combinatorial or probabilistic attacks. Thus it is not surprising that he had such a tremendous success in combining *combinatorics and probability*.

At first, Erdős used random methods to tackle problems in mainstream graph theory. We have already mentioned the delightful probabilistic argument Erdős used in 1947 to give a lower bound for the Ramsey number $R(k,k)$. A little later, in 1959, a more difficult result was proved by Erdős by random methods: *for every $k \geq 3$ and $g \geq 3$ there is a graph of chromatic number k and girth g.* Earlier results in this vein had been proved by Tutte, Zykov, Kelly and Mycielski, but before this beautiful result of Erdős, it had not even been known that such graphs exist for any $k \geq 6$.

Later ingenious *constructions* were given by Lovász, Nešetřil and Rödl, but these constructions lead to considerably larger graphs than obtained by Erdős.

In a companion paper, published in 1961, Erdős turned to lower bounds for the Ramsey numbers $R(3,l)$, and proved by similar probabilistic arguments that $R(3,l) > c_0 l^2/(\log l)^2$ for some positive constant c_0. In 1968, Graver and Yackel gave a good upper bound for $R(3,l)$, which was improved, in 1972, by Yackel. As expected, further improvements were harder to come by. In 1980, Ajtai, Komlós and Szemerédi proved that $R(3,l) < c_1 l^2/\log l$; the difficult proof was simplified a little later by Shearer. Very recently, J.B. Kim improved greatly the lower bound due to Erdős, and so now we know that the order of $R(3,l)$ is $l^2/\log l$.

Almost simultaneously with his beautiful applications of random graphs to extremal problems, Erdős, with Rényi, embarked on a *systematic* study of random graphs. The first Erdős-Rényi paper on random graphs, in 1959, is about the connectedness of $G_{n,M}$, the random graph with vertex set $[n] = \{1,2,\ldots,n\}$, with M randomly chosen edges. Extending an unpublished result of Erdős and Whitney, they proved, among others, that if $c \in \mathbb{R}$ and $M = M(n) = \lfloor \frac{1}{2} n(\log n + c) \rfloor$ then

$$\lim_{n\to\infty} \mathbb{P}(G_{n,M} \text{ is connected}) = e^{-e^{-c}}.$$

This implies, in particular, that if $M = M(n) = \lfloor \frac{1}{2} n(\log n + \omega(n)) \rfloor$ then

$$\lim_{n\to\infty} \mathbb{P}(G_{n,M} \text{ is connected}) = \begin{cases} 0 & \text{if} \quad \omega(n) \to -\omega, \\ e^{-e^{-c}} & \text{if} \quad \omega(n) \to c \in \mathbb{R}, \\ 1 & \text{if} \quad \omega(n) \to \infty. \end{cases}$$

The result is easy to remember if one notes that the "obstruction" to the connectedness of a random graph is the existence of isolated vertices: if $|\omega(n)|$ is not too large, say at most $\log\log n$, then $G_{n,M}$ is *very likely* to be connected if it has no isolated vertices (and if it does have isolated vertices then, a fortiori, it is disconnected).

By now, quite rightly, this is viewed as a rather simple result, but when it was proved, it was very surprising. To appreciate it, note that a graph of order n with as few as $n-1$ edges *need not be disconnected*, and a graph of order n with as many as $(n-1)(n-2)/2$ edges *need not be connected.*

A little later, in 1960, in a monumental paper, entitled "*On the evolution of random graphs*", Erdős and Rényi laid the foundation of the *theory of random graphs.* As earlier, they studied the random graphs $G_{n,M}$ with n labelled vertices and M random edges for large values of n, as M increased from 0 to $n\binom{}{2}$. They introduced basic concepts like "threshold function", "sharp threshold function", "typical graph", "almost every graph", and so on. An important message of the paper was that most monotone properties of graphs appear rather suddenly. A property Q_n of graphs of order n is said to be *monotone increasing* if Q_n is closed under the addition of edges. Thus being connected, containing a triangle or having diameter at most five are all monotone increasing properties. Erdős and Rényi showed that for many a fundamental structural monotone increasing property Q_n there is a threshold function, that is a function $M^*(n)$ such that

$$\lim_{n\to\infty} \mathbb{P}(G_{n,M} \text{ has } Q_n) = \begin{cases} 0 & \text{if } M(n)/M^*(n) \to 0, \\ 1 & \text{if } M(n)/M^*(n) \to \infty. \end{cases}$$

Later it was noticed by Bollobás and Thomason that in this weak sense every monotone increasing property of set systems has a threshold function; recently a considerably deeper result has been proved by Friedgut and Kalai, which takes into account the automorphism group of the property, and so is much more relevant to properties of graphs.

The more technical part of the "*Evolution*" paper concerns cycles, trees, the number of components and, most importantly, the emergence of the giant component. Erdős and Rényi showed that if $M(n) = \lfloor cn \rfloor$ for some constant $c > 0$ then, with probability tending to 1, the largest component of $G_{n,M}$ is of order $\log n$ if $c < \frac{1}{2}$, it jumps to order $n^{2/3}$ if $c = \frac{1}{2}$ and it jumps again, this time right up to order n if $c > \frac{1}{2}$. Quite understandably, Erdős and Rényi considered this "double jump" to be one of the most striking features of random graphs.

By now, all this is well known, but in 1960 this was a striking discovery indeed. In fact, for over two decades not much was added to our knowledge of this *phase transition* or, as called by many a combinatorialist, the *emergence of the giant component.* The investigations were reopened in 1984 by the author of these lines with the main aim of deciding what happens around $M = \lfloor n/2 \rfloor$; in particular, what *scaling*, what *magnification* we should use to see the giant component growing continuously. It was shown, among others, that if $M = n/2 + s$ and $s = o(n)$ but slightly larger than $n^{2/3}$ then, with probability tending to 1, there is a unique largest component, with about $4s$ vertices, and the second largest component has no more than $(\log n)n^2/s^2$ vertices. Thus, in a rather large range, on average every new edge adds four new vertices to the giant component!

With this renewed attack on the phase transition the floodgates opened, and quite a few more precise studies of the behaviour of the components near the point of phase transition were published, notably by Stepanov (1988), Flajolet, Knuth and Pittel (1989), Łuczak (1990, 1991) and others. To cap it all, in 1993 Knuth, Pittel, Janson and Łuczak published a truly prodigious (over 120 pages) study, "*The birth of the giant component*", giving very detailed information about the random graph $G_{n,M}$ near to its phase transition.

Erdős and Rényi also stated several problems concerning random graphs, thereby influencing the development of the subject. In 1966, they themselves solved the problem of 1-*factors*: if n is even and $M = M(n) = \lfloor \frac{n}{2}(\log n + c) \rfloor$ then the probability that $G_{n,M}$ has a 1-factor tends to $e^{-e^{-c}}$ as $n \to \infty$; the "obstruction" is, once again, the existence of isolated vertices.

The *Hamilton cycle* problem was a much harder nut to crack. As a Hamiltonian graph is connected (and has minimal degree 2), it is rather trivial that if, with probability tending to 1, $G_{n,M}$ has a Hamilton cycle and if $M = M(n)$ is written as

$$M = M(n) = \frac{n}{2}(\log n + \log\log n + \omega(n)),$$

then we must have $\omega(n) \to \infty$. On the other hand, it is far from obvious that a "typical" $G_{n,M}$ is Hamiltonian, even if $M = \lfloor cn \log n \rfloor$ for some large constant c. This beautiful assertion was proved in 1976 by Pósa, making use of his celebrated lemma. Several more years passed, before Komlós and Szemerédi proved in 1983 that $\omega(n) \to \infty$ also suffices to ensure that a "typical" $G_{n,M}$ is Hamiltonian. A little later Bollobás proved a sharper, hitting time type result that had been conjectured by Erdős and Spencer, connecting Hamiltonicity with having minimal degree at least 2.

The *chromatic number* problem from the 1960 "*Evolution*" paper of Erdős and Rényi was the last to fall. In 1988 the author of this note proved that *picking one of the* $2^{\binom{n}{2}}$ *graphs on* $[n]$ *at random, with probability tending to* 1*, the chromatic number of the random graph is asymptotic to* $\frac{\log 2}{2} n / \log n$. Earlier results had been obtained by Grimmett and McDiarmid, Bollobás and Erdős, Matula, Shamir and Spencer, and others, and subsequent refinements were proved by Frieze, Łuczak, McDiarmid and others.

The tremendous success of the theory of random graphs in shedding light on a variety of combinatorial, structural problems concerning graphs foreshadows the use of random methods in other branches of mathematics. Graphs carry only a minimal structure so they are bound to yield to detailed statistical analysis. However, as we acquire more expertise in applying results of probability theory, we should be able to subject more complicated structures to statistical analysis. In keeping with this philosophy, having founded, with Rényi, the theory of random graphs, Erdős turned to "the theory of random groups" with another great collaborator, Paul Turán. In a series of seven substantial papers, published between 1965 and 1972, Erdős and Turán laid the foundations of *statistical group theory.*

For simplicity, let us consider the symmetric group S_n, and let π_n be a random element of S_n, with each of the $n!$ possibilities equally likely. Thus π_n *is a random permutation of* $[n] = \{1, 2, \ldots, n\}$, and every function of π_n is a *random variable.* One of the simplest of these random variables is the *order* $O(\pi_n)$ of a permutation π_n.

Concerning $g(n) = \max_{\pi \in S_n} O(\pi)$, the maximal order of a permutation, it was already shown by Landau in 1909 that

$$\lim_{n\to\infty} \frac{\log g(n)}{\sqrt{n \log n}} = 1.$$

Thus $O(\pi_n)$ is *always* small compared to the order of the group S_n, although it *can* be rather large!

In contrast, for a single cycle of length n has order n, although such cycles constitute a non-negligible fraction, namely a fraction $1/n$, of all possible permutations. What is then the order of most elements of S_n?

As the starting point of their investigations, Erdős and Turán proved that for a "*typical*" permutation π_n, the order $O(\pi)$ is much smaller than the maximum $g(n) = \exp\{(n \log n)^{1/2}(1 + o(l))\}$, and much larger than n. In fact, if $\omega(n) \to \infty$ (arbitrarily slowly, as always) then

$$\lim_{n\to\infty} \mathbb{P}(|\log O(\pi_n) - \frac{1}{2}\log^2 n| \geq \omega(n)\log^{3/2} n) = 0.$$

Thus the "typical" order is about $\frac{1}{2}\log^2 n$.

Erdős and Turán went on to prove that, asymptotically, $O(\pi_n)$ has a *log-normal* distribution: as $n \to \infty$,

$$\sqrt{3}(\log O(\pi_n) - \frac{1}{2}\log^2 n)/\log^{3/2} n$$

tends, in distribution, to the standard normal distribution, i.e. if $x \in \mathbb{R}$ then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{3}(\log O(\pi_n) - \frac{1}{2}\log^2 n)}{\log^{3/2} n}\right) = \Phi(x).$$

Having established this central limit theorem, which by now is known as the *Erdős-Turán law*, they went on to study the number $W(n)$ of *different* values of $O(\pi_n)$. (Thus $W(n)$ is the number of non-isomorphic cyclic subgroups of S_n.) Erdős and Turán proved that

$$W(n) = \exp\left\{\pi\sqrt{\frac{2n}{3\log n}} + O\left(\frac{\sqrt{n}\log\log n}{\log n}\right)\right\},$$

and, with the exception of $o\big(W(n)\big)$ values, all are of the form

$$\exp\left\{(1 + o(1))\frac{\sqrt{6}\log 2}{\pi}\sqrt{n\log n}\right\}.$$

In S_n there are $p(n)$ conjugacy classes, where $p(n)$ is the partition function mentioned earlier, and studied in detail by Hardy and Ramanujan. As the order of a permutation $\pi \in S_n$ depends only on its conjugacy class K, it is natural to ask what the distribution of $O(K)$ is if the $p(n)$ conjugacy classes are considered equiprobable. Here we have written $O(K)$ for the order of any permutation in K. Erdős and Turán proved that, with probability tending to 1,

$$O(K) = \exp((A_0 + o(1))\sqrt{n}),$$

where

$$A_0 = \frac{2\sqrt{6}}{\pi}\sum_{j\neq 0}\frac{(-1)^{j+1}}{3j^2 + j} \approx 1.81.$$

All these results are proved by hard analysis, using Tauberian theorems and contour integration, somewhat resembling the Hardy-Ramanujan analysis; there is no reference to soft analysis or general theorems in probability theory or group theory that would get round the hard work. Thus it is not surprising that, over the years, many of the results of Erdős and Turán have been given shorter, more probabilistic proofs, that lead to sharper results. In particular, the Erdős-Turán law was studied by Best in 1970, Bovey in 1980, Nicolas in 1985 and Arratia and Tavaré in 1992. To date, the last word on the topic is due to Barbour and Tavaré, who used the *Ewens sampling formula*, derived by Ewens in 1972 in the context of population genetics, to give a beautiful proof of the Erdős-Turán law with a sharp error estimate. It is fascinating that, in order to get a small error term, Barbour and Tavaré had to adjust slightly the approximating normal distribution:

$$\sup_x \bigl|\mathbb{P}\bigl[\{\tfrac{1}{3}\log^3 n\}^{-1/2}(\log O(\pi_n) - \tfrac{1}{2}\log^2 n + \log n \log\log n) \le x\bigr] - \Phi(x)\bigr| \\ = O\bigl((\log n)^{-1/2}\bigr).$$

Numerous other problems of statistical group theory have been studied, including the problem of *random generation.* Dixon proved in 1969 that almost all pairs of elements of S_n generate S_n or the alternating group A_n, and recently Kantor and Lubotzky proved analogues of this result for finite classical groups. Because of problems arising in computational Galois theory, one is also interested in a considerably stronger condition than mere generation. The elements $x_1, \ldots, x_m$ of a group G are said to generate G *invariably* if G is generated by $y_1, \ldots, y_m$ whenever y_i is conjugate to x_i. for $i = 1, 2, \ldots, m$. Dixon showed in 1992 that for some constant $c > 0$, with probability tending to 1, $c(\log n)^{1/2}$ randomly chosen permutations generate S_n invariably. In 1993 Łuczak and Pyber, confirming a conjecture of McKay, proved that *for every $\epsilon > 0$ there is a constant $C = C(\epsilon)$ such that C random elements generate S_n with probability at least $1 - \epsilon$.*

Łuczak and Pyber also proved a conjecture of Cameron; they showed that *the fraction of elements of S_n that belong to transitive subgroups other than S_n or A_n tends to 0 as $n \to \infty$.*

Needless to say, in spite of these powerful results, many important questions remain unanswered, indicating that statistical group theory is still in its infancy.

When writing about the contributions of Erdős to mathematics, it would be unforgivable not to emphasize the enormous influence he exerts through his uncountably many problems. At the International Congress of Mathematicians in Paris in 1900, David Hilbert emphasized with great eloquence the importance of problems for mathematics. “The clearness and ease of comprehension insisted on for a mathematical theory I should still more demand for a mathematical problem, if it is to be perfect. For what is clear and easily comprehended attracts; the complicated repels us.”

For lack of space, we shall confine ourselves to one more of the problems of Erdős that have been solved, and to three particularly beautiful unsolved questions.

There is no doubt that the most difficult Erdős problem solved to date is the problem on arithmetic progressions. In 1927 van der Waerden proved the following conjecture of Baudet: *if the natural numbers are partitioned into two classes then at least one of the classes contains arbitrarily long arithmetic progressions.* Over the years, this beautiful Ramsey-type result has been the starting point of much research. Quite early on, in 1936, Erdős and Turán suspected that *partitioning* the integers is an overkill: it suffices to take a "large" set of integers. Thus they formulated the following conjecture: *if A is a set of natural numbers with positive upper density, that is, if*

$$\limsup_{n\to\infty} \frac{|A\cap[n]|}{n} > 0,$$

then A contains arbitrarily long arithmetic progressions.

Roth was the first to put a dent in this Erdős-Turán conjecture when, in 1952, he proved that A must contain arithmetic progressions of length 3. Length 4 was much harder: Szemerédi proved it only in 1969. Having warmed up on length 4, in 1974 Szemerédi proved the full conjecture; the long and difficult proof is a real *tour de force* of combinatorics. The story did not end there: in 1977 Fürstenberg gave another proof of Szemerédi's theorem, using tools of ergodic theory; the methods of this proof and the new problems it naturally led to revolutionized ergodic theory.

Let us turn then to the three unsolved Erdős problems we promised. The first asks for a substantial extension of Szemerédi's theorem. Let $a_1 < a_2 < \dots$ be a sequence of natural numbers such that $\sum 1/a_n = \infty$. Is it true then that the sequence contains arbitrarily long arithmetic progressions? It is not even known that Roth's theorem holds in this case, i.e., that the sequence contains an arithmetic progression with three terms. If this is not enough to indicate that this problem is rather hard, it is worth noting that Erdős offers \$5,000 for a solution. A rather special case of the conjecture would be that the *primes* contain arbitrarily long arithmetic progressions.

The last two are also rather old conjectures, but each carries "only" a \$500 price-tag. Let $f(n)$ be the minimal number of distinct distances determined by n distinct points in the plane. Erdős conjectured in 1946 that

$$f(n) > \frac{cn}{\sqrt{\log n}}$$

for some absolute constant $c > O$. The lattice points show that, if true, this is best possible. Chung, Szemerédi and Trotter have proved that $f(n)$ is at least about $n^{4/5}$.

The third problem is from the 1961 paper of Erdős, Ko and Rado; it is, in fact, the last unsolved problem of that paper. (However, Ahlswede and Khachatrian have just announced a proof of the conjecture.)

Let $\mathcal{A}$ be a 2-intersecting family of $2n$-subsets of $[4n] = \{1, 2, \ldots, 4n\}$. Thus if $A, B \in \mathcal{A}$ then $A, B \subset [4n]$, $|A| = |B| = 2n$, and $|A \cap B| \geq 2$. Then the Erdős-Ko-Rado conjecture states that

$$|\mathcal{A}| \leq \frac{1}{2}\binom{2n}{2n} - \frac{1}{2}\binom{2n}{n}^2.$$

It is easily seen that, if true, this inequality is best possible. Indeed, let $\mathcal{A}$ be the collection of $2n$-subsets of $[4n]$, containing at least $n+1$ of the first $2n$ natural numbers. Then $\mathcal{A}$ is clearly 2-intersecting, and for every $2n$-subset A on $[4n]$, the system contains precisely one of A and its complement $\bar{A}$, unless A (and so $\bar{A}$ as well) contains precisely n of the first $2n$ natural numbers.

It is widely known that vast amounts of thought and ingenuity are required in order to earn \$500 on an Erdős problem; even so, this problem may be far harder than its price-tag suggests.

Although this brief review does not come close to doing justice to the mathematics of Paul Erdős, it does indicate that he has enriched the mathematics of this century as very few others have. He has clearly earned a *mathematical Oscar for lifetime achievement*, several times over. May he continue to prove and conjecture for many years to come.

Added in Proof

Sadly, this was not to be. On 20 September 1996, while attending a mini-semester at the Banach Center in Warsaw, Professor Paul Erdős was killed by a massive heart attack. Although in the last year he started to show signs of aging, his death was premature and entirely unexpected.

We combinatorialists have just become orphans.

B.B.

Erdős Magic

Joel Spencer

J. Spencer (✉)
Courant Institute, New York University, New York, NY 10012, USA
e-mail: spencer@cims.nyu.edu

A version of this article was first published by the Institute of Mathematics and its Applications (www.ima.org.uk) in 'Mathematics Today', February 2013

> If I have seen further, it is by standing on the shoulders of Hungarians—Peter Winkler

Paul Erdős was a giant of twentieth century mathematics. Born in 1913 he was a child prodigy whose talents were well cultivated in his native Budapest. Dubbed *Der Zauberer von Budapest* he moved onto the world stage in his early 20s. And there he stayed, extraordinarily active right until his death in 1996.

In 1999 when his longtime friend and collaborator Vera Sós organized a memorial conference we participants were stunned. Komjáth on infinite graphs, Pomerance on special arithmetic functions, Hajnal on partition relations, Graham on Ramsey theory, Simonovits on Extremal Graph theory, Lubinsky on interpolation, Alon and this author on probabilistic methods—none of us individually had realized the *breadth* of Erdős's contributions.

Paul's place in the mathematical pantheon will be a matter of strong debate for in that rarefied atmosphere he had a unique style. The late Ernst Straus described this style in a commemoration of Erdős's 70th birthday.

> In our century, in which mathematics is so strongly dominated by "theory constructors" he has remained the prince of problem solvers and the absolute monarch of problem posers. One of my friends - a great mathematician in his own right - complained to me that "Erdős only gives us corollaries of the great metatheorems which remain unformulated in the back of his mind." I think there is much truth to that observation but I don't agree that it would have been either feasible or desirable for Erdős to stop producing corollaries and concentrate on the formulation of his metatheorems. In many ways Paul Erdős is the Euler of our times. Just as the "special" problems that Euler solved pointed the way to analytic and algebraic number theory, topology, combinatorics, function spaces, etc.; so the methods and results of Erdős's work already let us see the outline of great new disciplines, such as combinatorial and probabilistic number theory,

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_2,

> combinatorial geometry, probabilistic and transfinite combinatorics and graph theory, as well as many more yet to arise from his ideas.

Straus, who worked as an assistant to Albert Einstein, noted that Einstein chose physics over mathematics because he feared that one would waste one's powers in pursuing the many beautiful and attractive questions of mathematics without finding the central questions. Straus goes on,

> Erdős has consistently and successfully violated every one of Einstein's prescriptions. He has succumbed to the seduction of every beautiful problem he has encountered—and a great many have succumbed to him. This just proves to me that in the search for truth there is room for Don Juans like Erdős and Sir Galahad's like Einstein.

Tim Gowers, in his beautiful and influential essay "The Two Cultures of Mathematics" discusses this distinction between theory constructors and problem solvers in detail. He gives Erdős as the archetypal problem solver and gives a spirited defense of the problem solver mode of doing mathematics. He worries that such a defense is necessary:

> . . . mathematicians in the theory-building areas often regard what they are doing as the central core (Atiyah uses this exact phrase) of mathematics, with subjects such as combinatorics thought of as peripheral and not particularly relevant to the main aims of mathematics.

Today, however, discrete math is a respected area of mathematics. Gowers himself was awarded the Fields Medal in 1998, Terence Tao was awarded the Fields Medal in 2006, László Lovász was awarded the Kyoto Prize in 2010, all for work with a strong combinatorial component. In 2012 the Abel Prize was awarded to Endre Szemerédi. Szemerédi's work is almost entirely combinatorial and follows closely in the footsteps of Paul Erdős. This respect certainly did not exist a 100, even 50 years ago. Discrete Math was often dismissed as "puzzle math" and the phrase "the slums of topology" (sometimes attributed to J. H. C. Whitehead) was widely and disparagingly used to describe it.

The rise of the Discrete has, I feel, two main causes. The first was The Computer, how wonderful that this physical object has led to such intriguing mathematical questions. The second, with due respect to the many others, was the constant attention of Paul Erdős with his famous admonition "Prove and Conjecture!" Ramsey Theory, Extremal Graph Theory, Random Graphs, how many turrets in our mathematical castle were built one brick at a time with Paul's theorems and, equally important, his frequent and always penetrating conjectures.

My own research specialty, The Probabilistic Method, owes its existence to Paul Erdős. I have proposed the term *Erdős Magic* for the area as a tribute to Erdős. It began in 1947 with a 3 page paper in the Bulletin of the American Mathematical Society. Paul proved the existence of a graph having certain Ramsey property without actually constructing it. In modern language he showed that an appropriately defined random graph would have the property with positive probability and hence there must exist a graph with the property. For the next 20 years Paul was a “voice in the wilderness”, his colleagues admired his amazing results but adoption of the methodology was slow. But Paul persevered—he was always driven by his personal sense of mathematical aesthetics in which he had supreme confidence—and today the method is widely used in both Discrete Math and in Theoretical Computer Science.

My own introduction to Erdős was not atypical. I had managed to solve one of his ten dollar problems. Erdős would give such small monetary awards for solutions of his conjectures. To receive such an award was heaven incarnate. We met and I nervously explained my argument. Instantly he saw how my methods could be used on a very different problem. This became our first joint paper and I became one of his disciples. This occurred in the late 1960s, a tumultuous time when “do your own thing” was the admonition that resonated so powerfully. But while others spoke of it, this was Paul's modus operandi. He had no job; he worked constantly. He had no home; the world was his home. Possessions were a nuisance, money a bore. He lived on a web of trust, traveling ceaselessly from Center to Center, spreading his mathematical pollen.

What drew so many of us into his circle? What explains the joy we have in speaking of this gentle man? Why do we love to tell Erdős stories? I've thought a great deal about this and I think it comes down to a matter of belief, or faith. We know the beauties of mathematics and we hold a belief in its transcendent quality. In Kronecker's immortal words: Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk. Mathematical truth is immutable, it lies outside physical reality. When we show, for example, that two n-th powers never add to an n-th power for $n \geq 3$ we have discovered a Truth. This is our belief, this is our core motivating force. Yet our attempts to describe this belief to our nonmathematical friends is akin to describing the Almighty to an atheist. Paul embodied this belief in mathematical truth. His enormous talents and energies were given entirely to the Temple of Mathematics. He harbored no doubts about the importance, the absoluteness, of his quest. To see his faith was to be given faith. The religious world has a name for such people—they are called saints. We knew him as Uncle Paul.

I do hope that one cornerstone of Paul's, if you will, theology will long survive. I refer to The Book. The Book consists of all the theorems of mathematics. For each theorem there is in The Book just one proof. It is the most aesthetic proof, the most insightful proof, what Paul called The Book Proof. When one of Paul's myriad conjectures was resolved in an “ugly” way Paul would be very happy in congratulating the prover but would add, “Now, let's look for The Book Proof.” This platonic ideal spoke strongly to those of us in his circle. The mathematics was there, we had only to discover it.

The intensity and the selflessness of the search for truth were described by the writer Jorge Luis Borges in his story *La biblioteca de Babel.* The narrator is a worker in a library that contains on its infinite shelves all wisdom. He wanders its infinite corridors in search of what Paul Erdős might have called The Book. He cries out,

> To me, it does not seem unlikely that on some shelf of the universe there lies a total book. I pray the unknown gods that some man—even if only one man, and though it have been thousands of years ago!—may have examined and read it. If honor and wisdom and happiness are not for me, let them be for others. May heaven exist though my place be in hell. Let me be outraged and annihilated but may Thy enormous Library be justified, for one instant, in one being.

In the summer of 1985 I drove Paul to Yellow Pig Camp—a mathematics camp for talented high school students. It was a beautiful day—the students loved Uncle Paul and Paul enjoyed nothing more than the company of eager young minds. In my introduction to his lecture I discussed The Book but I made the mistake of describing it as being "held by God." Paul began his lecture with a gentle correction that I shall never forget. "You don't have to believe in God," he said, "but you should believe in The Book."

I. Early Days

Introduction

We do not have much to add here. We only want to express our gratitude to Arthur Stone, Cedric Smith, William Tutte and Irving Kaplansky for their personal recollections of those days when many contemporary theories were being created. In those early days, Erdős' interests were almost entirely devoted to number theory and, as seen from his own contribution, as well as from the other contributions to this chapter, number theory is an old love which does not fade. But so is an early combinatorial puzzle solved by four young Cambridge undergraduates (three of which are among authors of this chapter). C. Smith related this puzzle to P. Erdős while W. Tutte, in his characteristic style, connects his very recent research to his own early beginnings.

But perhaps it is very fitting to include two quotations by two of Erdős' closest and earliest collaborators and colleagues from their own recollections of the early days in Budapest in the 1930s.

I have known Paul Erdős long enough to be a bit personal. We met daily at the university, made excursions practically every Sunday with a group of other fellow students, steady members of which were (among others) Gy. Szekeres and T. Gallai. The main subject of conversations was mathematics; since Gallai and Erdős attended the graph theoretical lectures of D. König, graph theory was discussed often. The first graph-theoretical result of Erdős, an extension of Menger's theorem to infinite graphs, arose as early as 1931; it was an answer to a question of König a few weeks after König raised it in his course. This was published in König's classical book in 1936 (probably nowhere else)...

Erdős' main interest in the thirties was number theory but gradually more and more combinatorial moments occurred even in these works. As a clear sign of this I quote from a paper of his entitled "On sequences of integers no one of which divides the product of two others and on some related problems" which appeared in 1938: "The argument was really based upon the following theorem for graphs. Let $2k$ points be given. We split them into two classes each containing k of them. The points of two classes are connected by segments such that the segments form no closed quadrilateral. Then the number of segments is less than $3k^{3/2}$." No doubt he was not far from the discovery of extremal graph theory before 1938.

P. Turán (Art of Counting, p. xvii)

We had a very close circle of young mathematicians, foremost among them Erdős, Turán and Gallai; friendships were forged which became the most lasting that I have ever known and which outlived the upheavals of the thirties, a vicious world war and our scattering to the four corners of the world. I myself was an "outsider," studying chemical engineering at the Technical University, but often joined the mathematicians at weekend excursions in the charming hill country around Budapest and (in summer) at open air meetings on the benches of the city park.

Paul, then still a young student but already with a few victories in his bag, was always full of problems and his sayings were already a legend. He used to address us in the same fashion as we would sign our names under an article and this habit became universal among us; even today I often call old members of the circle by a distortion of their initials.

"Szekeres Gy., open up your wise mind." This was Paul's customary invitation—or was it an order?—to listen to a proof or a problem of his. Our discussions centered around mathematics, personal gossip, and politics. It was the beginning of a desperate era in Europe. Most

> of us in the circle belonged to that singular ethnic group of European society which drew its cultural heritage from Heinrich Heine and Gustav Mahler, Karl Marx and Cantor, Einstein and Freud, later to become the principal target of Hitler's fury.
>
> G. Szekeres (Art of Computing, p. xix, xx)

And we are happy to include here a standard Erdős article of those days. Well slightly nonstandard this time: My favorite problems from all fields. . .

Some of My Favorite Problems and Results

Paul Erdős

P. Erdős (Deceased)
Mathematical Institute, Hungarian Academy of Sciences, Budapest Pf. 127,
1364, Hungary

1. Introduction

Problems have always been an essential part of my mathematical life. A well chosen problem can isolate an essential difficulty in a particular area, serving as a benchmark against which progress in this area can be measured. An innocent looking problem often gives no hint as to its true nature. It might be like a "marshmallow," serving as a tasty tidbit supplying a few moments of fleeting enjoyment. Or it might be like an "acorn," requiring deep and subtle new insights from which a mighty oak can develop.

As an illustration of how hard it can be to judge the difficulty of a problem, I'd like to tell the following anecdote concerning the great mathematician David Hilbert. Hilbert lectured in the early 1920s on problems in mathematics and said something like this—probably all of us will see the proof of the Riemann Hypothesis, some of us (but probably not I) will see a proof of Fermat's last theorem but none of us will see the proof that $2^{\sqrt{2}}$ is transcendental. In the audience was Carl Ludwig Siegel, whose deep research contributed decisively to the proof by Kusmin a few years later of the transcendence of $2^{\sqrt{2}}$. In fact, shortly thereafter Gelfond and a few weeks later Schneider independently proved the α^{β} is transcendental if α and β are algebraic, β is irrational and $\alpha \neq 0, 1$.

In this note I would like to describe a variety of my problems which I would classify as my favorites. Of course, I can't guarantee that they are all "acorns," but because many have thwarted the efforts of the best mathematicians for many decades (and have often acquired a cash reward for their solutions), it may indicate that new ideas will be needed, which can in turn, lead to more general results, and naturally, to further new problems. In this way, the cycle of life in mathematics continues forever.

2. Number Theory

My first serious problem was formulated in 1931 and it is still wide open. Denote by $f(n)$ the largest number of integers $1 \leq a_1 < a_2 < \cdots < a_k \leq n$ all of whose subset sums $\sum_{i=1}^{k} \varepsilon_i a_i$ are distinct, where $\varepsilon_i = 0$ or 1. The powers

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_3,

of 2 have this property and I conjectured that

$$f(n) < \frac{\log n}{\log 2} + c \tag{1}$$

for some absolute constant c. I offer \$500 for a proof or disproof. The inequality

$$f(n) < \frac{\log n}{\log 2} + \frac{\log\log n}{\log 2} + c_1$$

is almost immediate, since there are 2^k sums of the form $\sum_i \varepsilon_i a_i$ and they must be all distinct and all are less than kn. In 1954, Leo Moser and I proved by using the second moment method that

$$f(n) < \frac{\log n}{\log 2} + \frac{\log\log n}{2\log 2} + c_2$$

which is the current best upper bound.

Conway and Guy found 24 integers all $\leq 2^{22}$ for which all subset sums are distinct, which implies $f(2^n) \geq n+2$ for $n \geq 22$. Perhaps

$$f(2^n) \leq n+2 \quad \text{for all } n?$$

Perhaps the following variant of the problem is more suitable for computation. Let $1 \leq b_1 < b_2 < \cdots < b_m$ be a sequence of integers for which all the subset sums $\sum_{i=1}^{m} \varepsilon_i b_i$, $\varepsilon_i = 0$ or 1, are distinct. Is it true that

$$\min b_m > 2^{m-c} \tag{2}$$

for some absolute constant c? Inequality (2) is of course equivalent to (1). The determination of the exact value of $\min b_m$ is perhaps hopeless but for small m the value of $\min b_m$ can no doubt be determined by computation, and I think this would be of some interest.

Perhaps my favorite problem of all concerns covering congruences. It was really surprising that it had not been asked before. A system of congruences

$$a_i \pmod{n_i}, \qquad n_1 < n_2 < \ldots < n_k \tag{3}$$

is called a *covering system* if every integer satisfies at least one of the congruences in (3). The simplest covering system is $0 \pmod 2$, $0 \pmod 3$, $1 \pmod 4$, $5 \pmod 6$ and $7 \pmod{12}$. The main problem is: Is it true that for every c one can find a covering system all of whose moduli are larger than c? I offer \$1,000 for a proof or disproof.

Choi [2] found a covering system with $n_1 = 20$, and a Japanese mathematician whose name I do not remember found such a system with $n_1 = 24$. If the answer to this question is positive, denote by $f(n)$ the smallest integer k for which there is a covering system

$$a_i \pmod{n_i}, \qquad 1 \leq i \leq k, \quad n_1 = t, \quad k = f(t).$$

It would be of some mild interest to determine $f(t)$ for the few values of $t = n_1$ for which we know that covering systems exist.

Many further unsolved problems can be asked about covering systems. Selfridge and I asked: Is there a covering system all of whose moduli are odd? Schinzel asked: Is there a covering system where $n_i \nmid n_j$ for $i \neq j$? Schinzel used such covering systems for the study of irreducibility of polynomials. Related to this is a question of Herzog and Schőnheim: Does there exist a finite group which can be partitioned into cosets of different sizes?

More generally, let $n_1 < n_2 < \cdots$ be a sequence of integers. Is there a reasonable condition which would imply that there is a covering system whose moduli are among the n_i? Quite likely there is no such condition. Let us now drop the condition that the set of moduli is finite, but to avoid triviality we insist that in the congruence $a_i \pmod{n_i}$, only the integers $\geq n_i$ are considered. When if ever can we find such a system?

Perhaps it is of some interest to tell how I came upon the problem of covering congruences. In 1934 Romanoff [20] proved that the lower density of integers of the form $2^k + p$, with p prime, is positive. This was surprising since the number of sums $2^k + p \leq x$ is cx. Romanoff in a letter in 1934 asked me if there were infinitely many odd numbers not of the form $2^k + p$. Using covering congruences I proved in [8] there is an infinite arithmetic progression of odd numbers no term of which is of the form $2^k + p$. Independently, Van der Corput also proved that there are infinitely many odd numbers not of the form $2^k + p$. In [3] Crocker proved there are infinitely many odd numbers not of the form $2^{k_1} + 2^{k_2} + p$, but his proof only gives the number of integers $\leq x$ which are not of this form is $> c \log\log x$. This surely can be improved but I am not at all sure if the upper density of the integers not of the form $2^{k_1} + 2^{k_2} + p$ is positive. One could ask the following (probably unattackable) problem: Is it true that there is an r so that every integer is the sum of a prime and at most r powers of 2? Gallagher [14] proved (improving a result of Linnik) that for every ε there is an r_ε so that the lower density of the integers which are the sum of a prime and r_ε powers of 2 is at least $1 - \varepsilon$. No doubt, lower density always could be replaced by density, but a proof that the density of the integers of the form $2^k + p$ exists seems untouchable.

I think that every arithmetic progression contains infinitely many integers of the form $2^{k_1} + 2^{k_2} + p$. Thus, covering congruences cannot be used to improve the result of Crocker.

On diophantine equations, my most important paper is my result with Selfridge: The product of consecutive integers is never a power. Perhaps in fact for $k > 3$ the product $\prod_{i=1}^{k}(n+i)$, $n > k$, always has a prime factor $p > k$ for which $p^2 \nmid \prod_{i=1}^{k}(n+i)$. Unfortunately, this seems hopeless at present. Perhaps the product $\prod_{j=1}^{m}(n+jk)$, $m > 3$, (i.e., the product of more than

3 terms of an arithmetic progression) is never a power. Ramachandra, Shorey and Tijdeman have important results in this direction (see the paper "Some methods of Erdős applied to finite arithmetic progressions" by Shorey and Tijdeman in this volume).

The following Tauberian theorem is connected to the elementary proof of the prime number theorem by Selberg and myself. Let $a_k > 0$, $s_t = \sum_{i \le t} a_i$ and suppose

$$\sum_{k=1}^{n} a_k(k + s_{n-k}) = n^2 + O(n). \tag{4}$$

Then $s_n = n + O(1)$.

This result is clearly related to Selberg's fundamental inequality

$$\sum_{p<x} (\log p)^2 + \sum_{pq<x} \log p \log q = 2x \log x + O(x).$$

My original proof of (4) was very complicated. It was simplified by Siegel and later by Shapiro. I am certain further results of this type can be obtained.

Let me now move on to some problems which originally arose in connection with van der Waerden's classic theorem on arithmetic progressions. Nearly 80 years ago, Schur conjectured that if we color the integers with ℓ colors, there is always formed a monochromatic arithmetic progression of k terms. In 1927, van der Waerden found an ingenious proof, and, in fact, he proved that there is a least integer $W(k, \ell)$ such that for every partition of the integers $1 \le t \le W(k, \ell)$ into ℓ classes, there is always an arithmetic progression of k terms belonging to a single class. Let $W(k)$ denote $W(k, 2)$. The proof given by van der Waerden gave a very poor estimate for $W(k)$, e.g., it was not even primitive recursive. It was a great achievement a few years ago when Shelah gave a primitive recursive bound for $W(k)$. Probably, his bound is still much too large and perhaps $W(k) < 2^{2^k}$. The first nontrivial lower bound $W(k) > 2^{k/2}$ was given by Rado and myself. The current best bound is $W(p+1) \ge p \cdot 2^p$, p prime, due to Berlekamp. I believe $\lim_{n\to\infty} \frac{f(n)}{2^n} = \infty$ has been proved by Beck (but this is not yet published). The first task would be to prove that $W(k) > (2+\varepsilon)^k$, and perhaps $W(k)^{1/k} \to \infty$. Graham offers \$1,000 to prove $W(k) \le 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ (where there are k 2's in the tower) for all k.

Over 60 years ago, Turán and I thought that this was a Turán type of problem and not a Ramsey type (this terminology did not exist at the time!). In fact, let $r_k(n)$ be the smallest integer for which every sequence of integers $1 \le a_1 < a_2 < \cdots < a_t \le n$ with $t \ge r_k(n)$ contains an arithmetic progression of k terms. It is easy to see that $\lim_{n\to\infty} \frac{r_k(n)}{n}$ exists for any k. We conjectured that $\frac{r_k(n)}{n} \to 0$ as $n \to \infty$. This of course would imply van der Waerden's theorem. We did not at first realize the difficulty of our conjecture and hoped that one might be able to get better estimates for $W(k)$ this way. At first we thought $r_k(n) < n^{1-\varepsilon}$ but 50 years ago, R. Salem and D. C. Spencer proved

$$r_3(n) > n^{1-c \log \log n}$$

and in 1946, Behrend proved

$$r_3(n) > n \exp(-c\sqrt{\log n})$$

which is the current record. Forty years ago Roth proved

$$r_3(n) > cn/\log\log n.$$

The current record is due to Heath-Brown and Szemerédi:

$$r_3(n) \leq n/(\log n)^{\alpha}, \quad \alpha \approx 1/4.$$

I offer \$500 for a proof that $r_3(n) < n/(\log n)^c$ for every c, and \$1,000 for any asymptotic formula for $r_k(n)$. This is probably unattackable at present.

In the early 1970s I offered \$1,000 for a proof that $r_k(n)/n \to 0$ as $n \to \infty$ and this was accomplished by Szemerédi in 1974. His proof is a masterpiece of combinatorial reasoning, and his regularity lemma, introduced in an early form in his proof, has subsequently found many applications in combinatorics and graph theory. Unfortunately, his proof used van der Waerden's theorem and so could not be used for the estimation of $W(k)$. Furstenberg also proved $r_k(n)/n \to 0$ as $n \to \infty$ by using ergodic theory, and his methods already have many applications to various problems in combinatorial number theory. In fact, there is a growing number of results in combinatorial number theory which only can be proved by methods of ergodic theory (so far). Unfortunately, Furstenberg's methods do not help in estimating $W(k)$.

An old conjecture in number theory states that for every k there are k primes in an arithmetic progression. This problem seems unattackable; the largest currently known arithmetic progression of primes has 22 terms. Many years ago I made the following conjecture which, if true, would settle the problem: Let $a_l < a_2 < \cdots$ be a sequence of integers satisfying

$$\sum_{k=1}^{\infty} \frac{1}{a_k} = \infty. \tag{5}$$

Then the a_k's contain arbitrarily long arithmetic progressions. I offer \$5,000 for a proof (or disproof) of this. Neither Szemerédi nor Furstenberg's methods are able to settle this but perhaps the next century will see its resolution.

Perhaps for every $\varepsilon > 0$ there is an $n_0(\varepsilon, k)$ so that if $n_0(\varepsilon, k) < a_1 < a_2 < \cdots$ is a sequence of integers which does not contain a k-term arithmetic progression then $\sum_i \frac{1}{a_i} < \varepsilon$.

Here is a related problem which is completely hopeless at present. Is it true that for every k there are k *consecutive* primes in arithmetic progression? In the present state of science, the problem is unattackable even for $k = 3$.

It is well known that there are infinitely many triples of primes forming an arithmetic progression but it is not yet known if this holds for quadruples.

Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers which does not contain a k-term arithmetic progression. Is it true that

$$\sum_i \frac{1}{a_i} \leq \left(\frac{1}{2} + o(1)\right) \ln W(k)? \tag{6}$$

If true, (6) is trivially best possible, and would be a considerable strengthening of (5). The inequality

$$\sum_i \frac{1}{a_i} < c \ln W(k)$$

would already be a sensational result.

I should point out that it is very difficult to determine exact values of $W(k)$. So far, we only know

$$W(3) = 9, \quad W(4) = 35, \quad W(5) = 178.$$

$W(6)$ is certainly unknown and perhaps beyond the range of computers. For more complete references on these topics, the reader can consult the books of Graham, Rothschild, and Spencer [15], Nešetřil and Rödl [19], Erdős and Graham [12], and Guy [16].

I now move on to Sidon sequences and related problems. Let $A = \{a_1 < a_2 < \cdots\}$ be a finite or infinite sequence of integers, and let

$$A(x) = \sum_{a_i \leq x} 1.$$

Denote by $f(n)$ the number of solutions of $n = a_i + a_j$. I first met Sidon in 1932. He posed two problems. The first was this: Does there exist an infinite sequence A for which for all $n > n_0$, $f(n) > 0$, but so that for every $\varepsilon > 0$, $f(n)/n^{\varepsilon} \to 0$ as $n \to \infty$. In other words, A should be a basis of order 2 but the number of solutions of $n = a_i + a_j$ should be small. I liked the problem very much and optimistically assured Sidon that I would construct such a sequence in a few days. In fact, it took me 20 years to prove there is a sequence A for which

$$c_1 \log n < f(n) < c_2 \log n. \tag{7}$$

This is of course much stronger than what Sidon asked for, but I never have found a *constructive* proof of (7), or for that matter, even for Sidon's original question. My proof shows that if you select a random sequence A in which each n is put in A with probability $c\left(\frac{\log n}{n}\right)^{1/2}$ then almost all such A satisfy (7).

Turán and I conjectured that if $f(n) > 0$ for $n > n_0$ then $\limsup_{n\to\infty} f(n) = \infty$. I offer \$500 for a proof or disproof of this. I also offer \$100 for a constructive proof of Sidon's original question. I conjecture there is no sequence A for which

$$f(n)/\log n \to c, \qquad 0 < c < \infty. \tag{8}$$

In other words, (7) is nearly best possible. I offer \$500 for a proof or disproof of (8). Sárközy and I proved that

$$\frac{|f(n) - c\log n|}{(\log n)^{1/2}} \to 0 \tag{9}$$

is not possible for any c, which is much weaker than (8). Perhaps an even stronger result than (8) holds. Put

$$c_1 = \liminf \frac{f(n)}{\log n}, \qquad c_2 = \limsup \frac{f(n)}{\log n}.$$

Then there is an absolute constant c so that for all A, $c_2 - c_1 > c$. This conjecture may be a bit too optimistic but I could never find a counterexample.

Now let us discuss Sidon's second question. Find a sequence A for which $A(x)$ is as large as possible and for which $f(n) = 0$ or 1, i.e., the sums $a_i + a_j$, $i < j$, are all distinct. Such sequences are now called Sidon sequences. Sidon was led to these problems by his study of lacunary trigonometric series. The greedy algorithm easily gives an infinite Sidon sequence A which for every x satisfies

$$A(x) > cx^{1/3}. \tag{10}$$

It took about 50 years until Ajtai, Komlós and Szemerédi (finally) improved (10). They showed it is possible to have

$$A(x) > c(x \log x)^{1/3}. \tag{11}$$

Probably there is a Sidon sequence A for which $A(x) > x^{1/2-\varepsilon}$ but this at present is far beyond reach. Rényi and I proved that for every $\varepsilon > 0$ there is a sequence A with $A(x) > n^{1/2-\varepsilon}$ for which $f(n) < c(\varepsilon)$. I proved $\limsup \frac{A(n)}{n^{1/2}} \geq \frac{1}{2}$ which was strengthened by Krückerberg to $\limsup \frac{A(n)}{n^{1/2}} \geq \frac{1}{\sqrt{2}}$ but the truth is surely 1. Does there exist such a sequence with $A(x) > x^{1/2}/(\log x)^c$? A sharpening of our old conjecture with Turán would state: If $a_n < Cn^2$ for all n then $\limsup f(n) = \infty$. In fact, for what functions $t(n) \to \infty$ does $a_n < n^2 t(n)$ imply $\limsup_{n\to\infty} f(n) = \infty$.

Here is an old conjecture of mine: Let $a_1 < a_2 < \cdots$ be an infinite sequence for which all the triple sums $a_i + a_j + a_k$ are distinct: Is it then true that $\limsup a_n/n^3 = \infty$. I offer \$500 for a proof or disproof of this. Turán and I observed that Sidon sequences behave quite differently. We proved that if $a_1 < a_2 < \cdots < a_k \leq n$ is a Sidon sequence, and we denote by $S(n) = \max k$, then

$$S(n) < n^{1/2} + cn^{1/4}. \tag{12}$$

Chowla and I noticed that the construction of Singer for perfect difference sets immediately gives

$$S(n) > n^{1/2} - n^{1/2-\varepsilon}. \tag{13}$$

I am fairly sure that

$$S(n) = n^{1/2} + O(n^{\varepsilon}). \tag{14}$$

for every $\varepsilon > 0$, and it might even be true that

$$S(n) = n^{1/2} + O(1). \tag{15}$$

I offer \$1,000 for clearing up these two problems. Incidentally, Singer proved that there are $p+1$ residues $a_1, a_2, \dots, a_{p+1} \pmod{p^2+p+1}$ (where p is a prime power) so that every nonzero residue $t \pmod{p^2+p+1}$ can be expressed uniquely in the form $t \equiv a_i - a_j \pmod{p^2+p+1}$, $1 \leq i, j \leq p+1$. This beautiful result easily gives (13).

Given an infinite sequence $S = a_1 < \cdots < a_n < \cdots$ of positive integers, denote by $R_n(S) = R_n$ the number of representations of the integer n in the form $n = a_i + a_j$. My result with Fuchs states that for any positive constant c, the relation

$$\sum_{n=0}^{N} R_n = cN + o(N^{1/2} \log^{-1/2} N)$$

cannot hold (irrespective of the nature of S). I hope that the Erdős-Fuchs result will survive us by centuries. We proved our result when I visited Cornell in the summer of 1954 and the paper appeared in 1956. Fuchs also considered our result important. Montgomery and Vaughan now have a somewhat sharper result.

Many years ago, I conjectured that every finite Sidon sequence $a_1 < \cdots < a_t$ can be completed into a perfect difference set of Singer. In other words, there is a $p = q^{\alpha}$ and a Singer set

$$a_1 < a_2 < \cdots < a_t < a_{t+1} < \cdots < a_{p+1} < p^2 + p + 1$$

which is a perfect difference set for $p^2 + p + 1$. I now feel this conjecture is perhaps too optimistic and I would be very happy for a proof of the following weaker conjecture: Let $a_1 < a_2 < \cdots < a_t$ be a Sidon sequence. Then for every $\varepsilon > 0$ there is a Sidon sequence $a_1 < \cdots < a_t < a_{t+1} < \cdots < a_n$ for which $a_n < (1+\varepsilon)n^2$. If true, this conjecture would imply that there is a Sidon sequence for which

$$\liminf_{n\to\infty} \frac{a_n}{n^2} = 1. \tag{16}$$

In view of (12), this would be best possible.

I proved that for every Sidon sequence, $\liminf_{n\to\infty} a_n/n^2 = \infty$, and in fact, more precisely,

$$\limsup_{n\to\infty} \frac{a_n}{n^2 \log n} > 0. \tag{17}$$

The reader will notice that (11) and (17) are very far apart. I conjecture that (11) and probably (17) can be improved a great deal. I expect there is a Sidon sequence so that for every $\varepsilon > 0$ and $n > n_0$,

$$a_n < n^{2+\varepsilon}. \tag{18}$$

I offer $1,000 for a proof or disproof of (18).

Sárközy and I recently conjectured that if $a_1 < a_2 < \cdots$ and $b_1 < b_2 < \cdots$ are two infinite sequences for which $a_n/b_n \to 1$, and if $g(n)$ denotes the number of solutions of $n = a_i + b_j$, then if $g(n) > 0$ for all n then

$$\limsup_{n\to\infty} g(n) = \infty. \tag{19}$$

If true, (19) would greatly strengthen our old conjecture with Turán: $f(n) > 0$ for all $n \Rightarrow \limsup_{n\to\infty} f(n) = \infty$. It is not even clear to me that this implies $f(n) \geq 2$ for infinitely many n. Our conjecture is perhaps too optimistic. Observe that the condition $\frac{a_n}{b_n} \to 1$ cannot be replaced by $c_1 < a_n/b_n < c_2$. To see this, let the a_k's be the integers of the form $\sum_i \varepsilon_i 2^{2i}$, $\varepsilon_i = 0$ or 1, and let the b_k's be the integers of the form $\sum_i \varepsilon_i 2^{2i+1}$. Clearly, every n can be uniquely expressed in the form $a_i + b_j$. Perhaps if $\varepsilon > 0$ is sufficiently small then

$$1 - \varepsilon < a_n/b_n < 1 + \varepsilon$$

and $f(n) > 0$ will imply $\limsup f(n) = \infty$. (However, this conjecture also seems too optimistic.)

Let me conclude this topic with one more conjecture. Let $a_1 < a_2 < \cdots$, be an infinite sequence of integers, and let $h(n)$ denote the number of integers not exceeding n of the form $a_i + a_j$, Suppose that for every $\varepsilon > 0$, $h(n)/n^{1-\varepsilon} \to \infty$ as $n \to \infty$. Is it then true that $\limsup f(n) = \infty$? Perhaps even this conjecture is too much to ask for and one instead should first look for a counterexample. For more complete references and many more results in this area, the reader should consult the excellent book of Halberstam and Roth [17]. (I understand a new and greatly enlarged edition is in preparation.)

I conjectured long ago that if $f(n) = \pm 1$ then for all c there is a d such that

$$\max_n \left| \sum_{k=1}^{n} f(kd) \right| > c.$$

In fact, perhaps

$$\max_n \left| \sum_{\substack{k=1 \\ \ell < n/d}}^{\ell} f(kd) \right| > c \log n.$$

If true, this would be best possible. I certainly offer \$500 to settle this annoying problem.

Next I would like to make some remarks about probabilistic number theory (cf. the excellent books of Elliott [5, 6]). First, though, I have to discuss the preprobabilistic era (i.e., when I did not know probability). Denote by $\sigma(n)$ the sum of the divisors of n. If $\sigma(n) = 2n$ then n is *perfect*, if $\sigma(n) \geq 2n$ then n is *abundant*, and otherwise n is *deficient*. Bessel-Hagen stated in his book that it did not seem to be known whether or not the density of the set of abundant numbers existed. Behrend, Chowla and Davenport (using Fourier analysis) proved that the density did exist. I independently gave a different proof, using elementary methods.

A number n is called primitive abundant if n is abundant but every proper divisor of n is deficient. Denote by $A(x)$ the number of primitive abundant numbers less than x. I proved

$$\frac{x}{\exp c_1(\log x \log\log x)^{1/2}} < A(x) < \frac{x}{\exp c_2(\log x \log\log x)^{1/2}}.$$

Ivic simplified the proof and obtained better constants but

$$A(x) = \frac{x}{\exp(1+o(1))c(\log x \log\log x)^{1/2}}$$

is still open. Michael Avidon, who is a student of Pomerance, found a better bound. The sum of the reciprocals of the primitive abundant numbers is therefore convergent and consequently the density of the abundant numbers exists. Later I proved that the distribution function of $\sigma(n)$ is purely singular.

Denote by $\nu(n)$ the number of prime factors of n. In 1917, Hardy and Ramanujan proved that for almost all n,

$$\bigl|\nu(n) - \log\log n\bigr| < f(n)(\log\log n)^{1/2}$$

for any $f(n) \to \infty$. Turán found a very simple proof of this and in 1935, I proved that the density of integers n for which $\nu(n) > \log\log n$ is 1/2. Also I proved that the density of integers for which $\sigma(n) > \sigma(n+1)$ is 1/2. I used Brun's method and the Central Limit Theorem (which I did not know at the time) in the binomial case. This was easy without using probability theory.

Now suppose $f(n)$ is an additive function. I proved that the distribution function of $f(n)$ exists provided the following three series converge:

$$\sum_{|f(p)|\geq 1} \frac{1}{p}, \quad \sum_{|f(p)|\leq 1} \frac{f(p)}{p}, \quad \sum_{|f(p)|\leq 1} \frac{f^2(p)}{p}.$$

This is analogous to the "three series" theorem of Kolmogoroff, which of course I did not know. I conjectured that the convergence of the three series was both necessary and sufficient for the existence of the distribution function but this I could not prove because of gaps in my knowledge. In March of 1939, Kac gave a talk at the Institute for Advanced Study in Princeton.

He stated the following conjecture: Let $f(n)$ be an additive function for which $|f(p)| < 1$, $f(p^\alpha) = f(p)$ (this is assumed only for the sake of simplicity), and

$$\sum_{p<x} \frac{f(p)}{p} = A(x), \quad \sum_{p<x} \frac{f^2(p)}{p} = B(x) \to \infty.$$

Then the density of integers n for which

$$f(n) < A(n) + c\sqrt{2B(n)}$$

is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1} \exp(-x^2)dx.$$

He further stated that he could prove it if $f(n) = \sum_{p|n} f(p)$ is replaced by $f_k(n) = \sum_{\substack{p|n \\ p<p_k}} f(p)$. I realized that if in fact Kac could prove this, then by using Brun's method, I could prove his conjecture. After the lecture we got together and soon saw that by combining our knowledge (i.e., the Central Limit Theorem and Brun's method) we could indeed prove the conjecture. Thus, we would say with a little impudence that probabilistic number theory was born. Using our theorem with Kac, Wintner and I proved that the convergence of the three series mentioned really is both necessary and sufficient for the existence of the distribution function. Of course, many nice problems remain open in this field. I refer the reader to the book of Elliott.

Davenport and I proved in 1935 that for any integer sequence $1 < a_1 < a_2 < \cdots$, the sequence of multiples always has a logarithmic density which equals its lower density. Suppose $a_1 < a_2 < \cdots$ is an integer sequence with positive upper logarithmic density. Davenport and I proved that there is an infinite subsequence where a_{i_r} divides $a_{i_{r+1}}$. I also proved that if $\sum_i \frac{1}{a_i \log a_i}$ is large then the sequence cannot be primitive, i.e., $a_i \nmid a_j$ cannot hold.

Now the following problem is annoying: Let $1 < \alpha_1 < \alpha_2 < \cdots$ be a sequence of real numbers and assume that for all $i \neq j, k$ we have

$$|k\alpha_i - \alpha_j| \geq 1. \tag{20}$$

Note that if the α_i's are integers then (20) implies that no α_i divides any α_j, $i \neq j$. Does (20) imply

$$\sum_i \frac{1}{\alpha_i \log \alpha_i} = \infty \quad \text{or} \quad \frac{1}{\log x} \sum_{\alpha_i < x} \frac{1}{\alpha_i} \to \infty \quad \text{as} \quad x \to \infty?$$

I offer \$500 for settling this annoying diophantine problem.

Let $\varepsilon_n \to 0$ arbitrarily slowly. Then the density of integers which have a divisor in $(n, n^{1+\varepsilon_n})$ tends to 0 as $n \to \infty$. This is a best possible strengthening of a result of Besicovitch. I conjectured infinitely long ago that the density of integers which have two divisors d_1 and d_2 with $d_1 < d_2 < 2d_1$

exists and is 1. This, and much more was finally proved by Maier and Tenenbaum (for which they collected \$250). For more results in this subject, see the books of Halberstam-Roth [17] and Hall-Tenenbaum [18].

Let me close this section with a few problems about (my old friends) the primes. In fact, my very first paper was on a new proof of Chebyshev's theorem: "Chebyshev said it and I'll say it again; there is always a prime between n and $2n$". Kalmár and I independently found in 1939 a very simple proof that $\prod_{p \leq n} p \leq 4^n$, and I found a simple proof that $\sum \frac{1}{p} = \infty$, but perhaps these are trifles.

Set $d_n = p_{n+1} - p_n$ (where, as usual, p_n denotes the nth prime). I proved in 1934 that for infinitely many n we have

$$d(n) > \frac{c \log n \log\log n}{(\log\log\log n)^2}$$

and in 1938, Rankin added a factor of $\log\log\log\log n$ to the numerator. I used to offer \$10,000 for a proof that

$$d_n > \frac{c \log n \log\log n \log\log\log\log n}{(\log\log\log n)^2}$$

holds for every c and infinitely many n. However, I would now like to offer only \$5,000 for this conjecture, and instead offer \$10,000 for a proof that

$$d_n > (\log n)^{1+\varepsilon} \tag{21}$$

for some $\varepsilon > 0$. Of course, this would also imply the previous conjecture, so I suppose that (21) would actually cost me \$15,000!

Ricci and I proved that the set of limit points of $d_n / \log_n$ has positive measure. No doubt they are everywhere dense. Maier has the sharpest partial results and he has collected \$250 more than once for some of these. Maier and I had dinner at Pomerance's house one night and afterwards, Maier drove me to my hotel. I told Maier to stop at the library on the way so I could find my paper which would show that I owed him \$250. Sure enough, we found the reference and I handed Maier the well-deserved \$250. The next day Pomerance said jokingly that this was certainly an expensive taxi ride.

In 1937, Kalmár and I proved that for every $\varepsilon > 0$, there is an elementary proof for

$$(1-\varepsilon)\frac{n}{\ln n} < \pi(n) < (1+\varepsilon)\frac{n}{\ln n}, \qquad n > n_0(\varepsilon). \tag{22}$$

Our proof did not give an elementary proof of the prime number theorem since it was based on the following fact. Let $\delta = \delta(\varepsilon)$ be small but fixed. Then one can find a k so that for every $k < t < k^2$,

$$\sum_{n=1}^{t} \mu(n) < \delta t \tag{23}$$

where μ is the usual Möbius function. That such a k exists follows from the prime number theorem but (23) can be shown by a finite computation. This was a curious logical situation which was perhaps folklore but surely deserved publication.

This is what happened. In the Spring of 1939 I met Rosser at a meeting at Duke University and I told him of our result. Rosser in fact already had a manuscript where he proved the analogous result for arithmetic progressions as well. Thus, Kalmár and I decided not to publish and agreed that Rosser could mention our result in his paper. However, Rosser's paper never appeared. (It seems the referee, who was not a number theorist, had difficulties in reading it, and Rosser eventually lost interest). Diamond and I still felt that these results deserved publication and we tried to reconstruct our old proof with Kalmár. In any case, we found a proof which was perhaps not identical to the original and which appeared around 1980 in Enseignement Math. Incidentally, the Tauberian theorem of Ingham seems to imply that the prime number theorem should follow from the general asymptotic formula for $n!$ but no one has ever found an elementary proof of Ingham's Tauberian theorem.

3. Polynomials

Let me now turn to polynomials. Turán and I obtained many important results on the distribution of the roots of polynomials based on the asymptotic properties of the polynomials. I also obtained many inequalities about polynomials with Gallai (who in those days was called Grünwald), Offord and many others. As an example, I state an elementary result with Gallai (which somebody as a joke called it a generalization of a theorem of Archimedes): Let the polynomial $f(x)$ have real roots $f(-1) = f(1) = 0$ and no other roots in $[-1, 1]$, and suppose $\sup_{-1 \leq x \leq 1} f(x) = 1$. Then

$$\int_{-1}^{1} f(x)dx \leq 4/3$$

with equality only for $f(x) = 1 - x^2$.

I have a long paper on polynomials with Herzog and Piranian in which we state many problems and results. Here I only want to mention one: For every $\varepsilon_n > 0$ there is a polynomial $f_n(z) = \prod_{i=1}^{n}(z - z_i), |z_i| = 1$, for which the measure of the set for which $|f_n(z)| < 1$ is less than ε_n. It would be of some interest to determine the dependence of ε_n on n, e.g., perhaps $\varepsilon_n > 1/(\log n)^c$.

Offord and I proved that among all polynomials

$$\sum_{k=1}^{n} \varepsilon_k z^k, \quad \varepsilon_k = \pm 1,$$

for all but $o(2^n/(\log n \log\log n)^{1/2})$ polynomials the number of real roots is

$$\frac{2}{\pi}\log n + O\big((\log n)^{2/3}\log\log n\big)$$

(this sharpened earlier results of Littlewood and Offord).

Clarkson and I (and independently, Laurent Schwartz) proved that if $\sum_k 1/n_k < \infty$ and $f(x)$ is a continuous function on $(-1,1)$ which can be approximated by polynomials $g_{n,k}(x) = \sum_{i=1}^k a_i x^{n_i}$ then $f(x)$ is analytic in the unit circle. A very well known theorem of Müntz and Szász asserts that if $\sum_k 1/n_k = \infty$ then every continuous function can be approximated by polynomials $g_n(x) = \sum_i a_i x^{n_i}$, and that $\sum\limits_k 1/n_k = \infty$ is necessary for this, as well. Our result makes this result clearer.

To end this section I would like to mention an old result on polynomials which was later followed up by a number of mathematicians. Let $f_n(x)$ be a polynomial of degree n with only real roots, none in $(-1,1)$, and with $|f(x)| \le 1$ for $-1 \le x \le 1$. Then

$$\sup_{-1\le x\le 1} |f'(x)| \le \frac{en}{2}.$$

Here, $\frac{en}{2}$ is best possible. If we take $-1+c < x < 1-c$ then we get $|f'(x)| < \frac{4}{c^2}\sqrt{n}$ and we only have to assume that $f(x)$ has no roots in the interior of the unit circle.

4. Combinatorics

One of my very favorite problems here is the following old conjecture of Faber, Lovász and myself: Let $G_1, \dots, G_n$ be n edge-disjoint complete graphs on n vertices. We conjectured more than 20 years ago that the chromatic number of $\bigcup_{i=1}^n G_i$ is n. I offer \$500 for a proof or disproof. Not long ago Kahn proved that the chromatic number of $\bigcup_{i=1}^n G_i$ is at most $(1+o(1))n$. I immediately gave him a consolation prize of \$100. It might be of interest to determine the maximum of the chromatic number of $\bigcup_{i=1}^n G_i$ if we require that $G_i \bigcap G_j$, $i \ne j$, is triangle-free, or should have size at most 1, but it is not clear we get a nice answer in these cases.

A family of sets A_i, $i = 1, 2, \dots$, is called a *strong* $\triangle$*-system* if all the intersections $A_i \bigcap A_j$, $i \ne j$, are identical. The family is called a *weak* $\triangle$*-system* if we only assume that all the *sizes* $|A_i \bigcap A_j|$, $i \ne j$, are the same. Rado and I [9, 10] investigated the following question: Denote by $f(n,k)$ the smallest integer for which every family of sets A_i, $1 \le i \le f(n,k)$, with $|A_i| = n$ for all i contains k sets which form a strong $\triangle$-system. In particular, we proved

$$2^n < f(n,3) < 2^n n!$$

Rado and I conjectured that

$$f(n,3) < c_3^n \tag{24}$$

for some constant c_3. No doubt, it is true that

$$f(n,k) < c_k^n.$$

I offer \$1,000 for a proof or disproof of (24). Recently, Kostochka proved (see his article in this volume)

$$f(n,3) < n!\left(\frac{c\log\log n}{\log\log\log n}\right)^{-n}.$$

I gave Kostochka a consolation prize of \$100. More recently, Axenovich, Fonder-Flass and Kostochka improved this to

$$f(n,3) < (n!)^{1/2+\varepsilon}$$

for every $\varepsilon > 0$ provided $n > n_0(\varepsilon)$.

Let $f(n) \to \infty$ arbitrarily slowly. Is it true that there is a graph G of infinite chromatic number such that for every n, every subgraph of G of n vertices can be made bipartite by the omission of at most $f(n)$ edges? I offer \$250 for a proof or disproof. It would be of interest to prove or disprove the existence of a graph G of infinite chromatic number for which $f(n) = o(n^{\varepsilon})$ or $f(n) = o\big((\log n)^c\big)$.

Many years ago I asked: Is there a sequence A of density 0 for which there is a constant $c(A)$ so that for $n > n_0(A)$, every $G(n, c(A)n)$ contains a cycle whose length is in A? (Here, $G(n,e)$ denotes a graph with n vertices and e edges). This question seems very interesting to me and I certainly offer \$100 for an answer. I am almost certain that if A is the sequence of powers of 2 then no such constant exists. What if A is the sequence of squares? I have no guess. Let $f(n)$ be the smallest integer for which every $G(n, f(n))$ contains a cycle of length 2^k for some k. I think that $f(n)/n \to \infty$ but that $f(n) < n(\log n)^c$ for some $c > 0$.

Next, I would like to discuss some problems connected to Ramsey's theorem. Denote by $r_k^{(\ell)}(p_1,\dots,p_\ell) = n$ the smallest integer so that if you color the k-tuples of $|S| = n$ by ℓ colors, there will always be for some i a subset of S of size p_i all of whose k-tuples have color i. (Often, we will omit writing the superscript (ℓ).) The existence $r_k^{(\ell)}(p_1,\dots,p_\ell)$ was first proved by Ramsey. It will be convenient to use the arrow notation introduced by Rado: $n \to (p_1,\dots,p_\ell)_k^{(\ell)}$ means that $r_k^{(\ell)}(p_1,\dots,p_\ell) \le n$. Of course, $n \not\to (p_1,\dots,p_\ell)^{k^{(\ell)}}$ means that $r_k^{(\ell)}(p_1,\dots,p_\ell)_k > n$. Also, we will use the square bracket notation introduced by Hajnal, Rado and myself: $n \to [p_1,\dots,p_\ell]_k^{(\ell)}$ means that if we color the k-tuples of a set S by ℓ colors, there always is for at least one i, a subset S_i of S of size p_i, no k-tuple of which is colored with color i. If all the p_i are equal to the same p, we will simply write $n \to (p)_k^{(\ell)}$ and $n \to [p]_k^{(\ell)}$.

As has been written elsewhere (see [11]), Ramsey's theorem was rediscovered in 1933 by Szekeres. He and I proved

$$cn2^{n/2} < r_2(n,n) < \binom{2n-2}{n-1} \tag{25}$$

or in the arrow notation

$$\binom{2n-2}{n-1} \to (n)^2_2 \quad \text{and} \quad cn2^{n/2} \nrightarrow (n)^2_2.$$

In other words, if one 2-colors the edges of a complete graph on $\binom{2n-2}{n-1}$ vertices, there is always a monochromatic complete subgraph $K(n)$ on n vertices.

Denote by $f(n)$ the smallest integer for which $f(n) \to (n)^2_2$ holds, so that $f(n) = r^2_2(n,n)$. I offer \$100 for a proof that $\lim_{n\to\infty} f(n)^{1/n}$ exists, and \$250 for the value c of this limit. It follows from (25) that $\sqrt{2} \le c \le 4$. Perhaps $c = 2$? Very little progress has been made in resolving these questions. Spencer has improved the constant in (25), and Thomason showed $f(n) < \binom{2n-2}{n-1}/n^{1/2-\varepsilon}$. My proof of the lower bound of (25) is nonconstructive. I offer \$100 for a constructive proof that $f(n) > (l+\varepsilon)^n$. Frankl and Wilson have a constructive proof that $f(n) > n^{c\log n}$. It is now known that

$$\frac{c_1n^2}{\log n} < r_2(3,n) < \frac{c_2n^2}{\log n}.$$

The upper bound is due to Ajtai, Komlós and Szemerédi. The recent lower bound was proved by J. H. Kim using very clever probability arguments. It would be nice to have an asymptotic formula for $r_2(3,n)$. I used to think that the probability method would give

$$r_2(4,n) > n^{3-\varepsilon}$$

and, in fact, more generally,

$$r_2(k,n) > \frac{n^{k-1}}{(\log n)^2}$$

for fixed k as $n \to \infty$. It now seems that I am wrong and new ideas will be required. The current record for a lower bound of $r_2(4,n)$ is $cn^{5/2}$ due to Spencer. The proof of Ajtai, Komlós and Szemerédi gives

$$r_2(\ell,n) < cn^{\ell-1}/\log n$$

for fixed ℓ.

For more general Ramsey numbers, much less is known. Hajnal, Rado and I proved

$$2^{cn^2} < r_3(n,n) < 2^{2^n}. \tag{26}$$

We believe the upper bound is closer to the truth, although Hajnal and I have a result which seems to favor the lower bound. We proved that if we color the triples of a set of n elements by two colors, there is always a set of size $s = \lfloor (\log n)^{1/2} \rfloor$ on which the distribution is unbalanced, i.e., one of the colors contains at least $(\frac{1}{2} + \varepsilon)\binom{s}{3}$ triples. This is in strong contrast to the case $k = 2$, where it is possible to 2-color the pairs of an n-set so that in every set of size $f(n)\log n$, where $f(n) \to \infty$, both colors get asymptotically the same number of pairs. We would begin to doubt seriously that the upper bound in (26) is correct if we could prove that in any 2-coloring of the triples of an n-set, some set of size $s = \lfloor (\log n)^{\varepsilon} \rfloor$ for which at least $(1-\eta)\binom{s}{3}$ triples have the same color. However, at the moment we can prove nothing like this. Hajnal proved

$$r_3(n,n,n,n) > 2^{c2^n}$$

which very strongly favors the upper bound in (26).

I now turn to an old conjecture of Graham and myself which lies at the interface of Ramsey theory and number theory. Is it true that no matter how one k-colors the integers ≥ 2 one can always find a solution to

$$1 = \sum_{a \in A} \frac{1}{a}, \quad \text{for a finite monochromatic subset } A? \tag{27}$$

We could never prove this even for $k = 2$. If the answer is in the affirmative, then determine or estimate the smallest integer $f(k)$ for which any k-coloring of $\{2, 3, \ldots, f(k)\}$ has the desired property. One can conjecture that if the integers are k-colored then for one of the color classes, *every* positive rational can be represented as a finite sum of the $\sum_{a \in A} \frac{1}{a}$. In fact, let $1 < a_1 < \cdots \leq a_k \leq n$ be a sequence of integers satisfying $\sum_{i=1}^{k} \frac{1}{a_i} > f(n)$. Is it true that if

$$\frac{f(n)}{(\log\log n)^2} \to \infty$$

then there is always a subsequence of the a_i's the sum of whose reciprocals sum to 1. The strongest conjecture we could not disprove states there is an absolute constant c so that if

$$\sum_{a_i < n} \frac{1}{a_i} > (c + \varepsilon)(\log\log n)^2$$

then (27) has a solution among the a_i's, but if $c + \varepsilon$ is replaced by $c - \varepsilon$ then this no longer holds. Perhaps all this is a bit too optimistic, but we do believe that if

$$\sum_{a_i < n} \frac{1}{a_i} > c \log n$$

then (27) has a solution in the a_i's, which if true, would show that our problem is a "Turán"-type problem.

I have worked quite a lot on problems in extremal graph theory. Since there are several excellent sources for such problems (the book of Bollobás

[1] and the survey paper of Simonovits [21]). I will restrict myself to just a few nice problems here.

Let H be a graph. The Turán number $T_n(H)$ of H is the smallest integer e_n for which every $G(n, e_n)$ contains H as a subgraph. Simonovits and I conjectured long ago that if H is bipartite and every induced subgraph of H has a vertex of degree $< r$, then

$$T_n(H) < cn^{2-1/(r-1)}. \tag{28}$$

This conjecture is open even for $r = 3$. We further conjecture that if H has an induced subgraph, every vertex of which has degree $\geq r$, then

$$T_n(H) > n^{2+\varepsilon-1/(r-1)}$$

I offer \$500 for a proof or disproof of each of our conjectures.

Denote by $f_n(H)$ the number of graphs on n vertices which do not contain H as a subgraph. Clearly $f_n(H) > 2^{T_n}(H)$. I conjectured

$$f_n(H) < 2^{(1+o(1))T_n(H)}. \tag{29}$$

This is open even for $H = C_4$. It is well known that $T_n(C_4) = (\frac{1}{2} + o(1))n^{3/2}$. On the other hand Kleitman and Winston only proved

$$f_n(C_4) < 2^{cn^{3/2}}.$$

Simonovits and I also conjectured that every $G(n, T_n(H))$ contains at least two copies of H. This is also open for C_4.

5. Geometry

I want to conclude this paper with some problems in geometry. Since there are several excellent sources for such problems (the book by Croft, Falconer and Guy [4] and the survey article of Erdős-Purdy [13]), I will limit my remarks.

Let $x_1, \ldots, x_n$ be n distinct points in the plane. Denote by $f(n)$ the minimum possible number of distinct distances $d(x_i, x_j)$. Thus, any set of n points in the plane determines at least $f(n)$ distinct distances. I conjectured [7] in 1946 that

$$f(n) > cn/\sqrt{\log n}. \tag{30}$$

The lattice points show that (30) if true is best possible. I offer \$500 for a proof or disproof of (30).

Denote by $g(n)$ the maximum number of times the same distance can occur, i.e., $g(n)$ is the maximum number of pairs for which $d(x_i, x_j) = 1$, where the maximum is taken over all configurations $x_1, \ldots, x_n$. I conjectured in my 1946 paper that

$$g(n) < n^{1+c/\log\log n} \tag{31}$$

for some $c > 0$. The lattice points again show that (31) if true is best possible. I offer \$500 for a proof or disproof of (31). The best results so far are $f(n) > n^{3/4}$ and $g(n) < n^{5/4} + \varepsilon$ for any $\varepsilon > 0$ and $n > n_0(\varepsilon)$.

Let $x_1, \ldots, x_n$ be n distinct points in the plane. Let $h(n)$ be the largest integer so that for every x_i there are $\geq h(n)$ points equidistant from x_i. Is it true that for $n > n_0(\varepsilon)$ we have $h(n) = o(n^{\varepsilon})$? I offer \$500 for a proof but only \$100 for a counterexample.

Szemerédi conjectured that if $x_1, \ldots, x_n$ are n points in the plane with no three on a line then they determine at least $\lfloor n/2 \rfloor$ distinct distances (but he could only prove $n/3$).

Let $x_1, \ldots, x_n$ be a convex polygon in the plane. Consider the $\binom{n}{2}$ distances $d(x_i, x_j)$ and assume that the distance u_i occurs s_i times. Clearly

$$\sum_i s_i = \binom{n}{2}.$$

I conjectured and Fishburn proved that

$$\sum_i s_i^2 < cn^3.$$

I also conjectured that $\sum_i s_i^2$ is maximal for the regular n-gon for $n > n_0$. If convexity is not assumed then I conjectured that

$$\sum_i s_i^2 < n^{3+\varepsilon} \tag{32}$$

for any $\varepsilon > 0$ and $n > n_0(\varepsilon)$. I offer \$500 for a proof or disproof of (32).

An old conjecture of mine states that if $x_1, \ldots, x_n$ are n points with no five points on a line then the number of lines containing four of the points is $o(n^2)$. I offer \$100 for a proof or disproof. An example of Grünbaum shows that the number of these lines can be $> cn^{3/2}$ for some $c > 0$, and perhaps $n^{3/2}$ is the correct upper bound. Sylvester observed that one can give n points in the plane so that the number of points passing through exactly three of our points is as large as $\frac{n^2}{6} - cn$ for some constant $c > 0$.

Purdy and I considered the following related problem. If we no longer insist that no five of the x_i can be on a line then the lattice points in the plane show that we can get cn^2 distinct lines each containing four (or more) of our points, and in fact, $c_1 n^2$ containing exactly four of them. Denote by $f(n)$ the maximum number of distinct lines which pass through at least four of our points. Determine or estimate $f(n)$. Perhaps if there are cn^2 distinct lines each containing more than three points, then there is an $h(n) \to \infty$ so that there is a line containing $h(n)$ distinct points. We can not even prove $h(n) \geq 5$ but we suspect $h(n) \to \infty$, and perhaps $h(n) > \varepsilon n^{1/2}$ for some $\varepsilon > 0$. It is easy to see that $h(n) < cn^{1/2}$ for some $c > 0$.

Finally, let me state the Erdős-Klein-Szekeres problem. This has quite a nice history which can be found in [11]. Let $f(n)$ be the smallest integer r so that any configuration of r points in the plane with no three on a line must contain the vertices of a convex n-gon. It is known that $f(4) = 5$ and $f(5) = 9$. Szekeres conjectured that $f(n) = 2^{n-2} + 1$. It is known that

$$2^{n-2} + 1 \leq f(n) \leq \binom{2n-4}{n-2} + 1$$

and these bounds have remained unchanged for many years. I would certainly pay $500 for a proof of Szekeres' conjecture.

References

1. B. Bollobás, *Extremal Grophy Theory*, Academic Press, London, 1978.
2. S. L. G. Choi, *Covering the set of integers by congruence classes of distinct moduli*, Mathematics of Computation 25 (1971), 885–895.
3. R. Crocker, *On the sum of a prime and two powers of two*, Pacific J. Math. 36 (1971), 103–107.
4. H. T. Croft, K. J. Falconer and R. K. Guy, *Unsolved Problems in Geometry*, Springer-Verlag, New York, 1991.
5. P. D. T. A. Elliot, *Probabilistic Number Theory I, Mean-Value Theorems*, Springer-Verlag, New York, 1979.
6. P. D. T. A. Elliot, *Probabilistic Number Theory II, Central Limit Theorems*, Springer-Verlag, New York, 1980.
7. P. Erdős, *On sets of distances of n points*, Amer. Math. Monthly 53 (1946), 248–250.
8. P. Erdős, *On integers of the form $2^k + p$ and some related problems*, Summa Brasiliensis Math. II (1950), 113–123.
9. P. Erdős and R. Rado, *Intersection theorems for systems of sets*, J. London Math. Soc. 35 (1960), 85–90.
10. P. Erdős and R. Rado, *Intersection theorems for systems of sets II*, J. London Math. Soc. 44 (1969), 467–479.
11. P. Erdős, *The Art of Counting*, J. Spencer, ed., MIT Press, Cambridge, MA, 1973.
12. P. Erdős and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, Monograph 28, l'Enseignement Math., 1980.
13. P. Erdős and G. Purdy, *Combinatorial geometry, in Handbook of Combinatorics*, R. L. Graham, M. Grötschel and L. Lovász, eds., North Holland, Amsterdam, 1994.
14. P. X. Gallagher, *Primes and powers of 2*, Invent. Math. 29 (1975), 125–142.
15. R. L. Graham, B. Rothschild and J. Spencer, *Ramsey Theory*, 2nd ed., Wiley, New York, 1990.
16. R. K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag, New York, 1981.
17. H. Halberstam and K. F. Roth, *Sequences*, Springer-Verlag, New York, 1983.
18. R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge University Press, Cambridge, 1988.
19. J. Nešetřil and V. Rödl, *Mathematics of Ramsey Theory*, Alg. and Comb. 5, Springer, New York, 1990.
20. N. P. Romanoff, *Über einige Sätze der additiven Zahlentheorie*, Math. Annalen 109 (1934), 668–678.
21. M. Simonovits, *Extremal Graph Theory, in Selected Topics in Graph Theory,* L. Beineke and R. J. Wilson, eds., vol. 2, Academic Press, New York, 1983.

Integers Uniquely Represented by Certain Ternary Forms

Irving Kaplansky

I. Kaplansky (Deceased)
Mathematical Sciences Research Institute, Berkeley, CA 94720, USA

1. Dedication

This paper has no connection with the two papers jointly authored by Paul Erdős and myself; nor does it overlap any of the many conversations we had. But I feel it is appropriate to dedicate the paper to him. It has the flavor of the mathematics we both particularly enjoyed: very explicit problems challenging us to answer “yes” or “no”.

Let me take the occasion to express my admiration to Paul for his style, his enthusiasm, and his incredible ability to do something effective on just about any problem posed to him. I shall mention just one sample: the so-called Erdős-Kaplansky lemma which appears on page 67 of Jacobson’s *Structure of Rings.* This came about when I mentioned to him the then unsolved problem of determining the dimension of the full dual of an infinite dimensional vector space over a division ring D. The key point needed is to exhibit as many linearly independent sequences as the cardinal number of D. When D is commutative, Vandermonde does the trick, but a different idea is needed to cope with noncommutativity. He listened politely, but of course we all (or almost all) do that. But within twenty-four hours he did much more: he returned with a novel workable idea. Needless to say, Paul is equally at home in the finite or infinite, and this was a dandy example of a delicate Zornification. (Incidentally, although it is very late in the day, I take this publication as an opportunity to publicly ask Jake to delete my name in a future printing. The person who merely asked the question should not have his or her name attached to a theorem.)

To conclude this dedication I mention a theorem which I learned from him (along with a slick proof): any countably infinite set has continuum many infinite subsets such that any two have finite intersection. There have been two occasions where this was just what I needed: on page 41 of *Linear Algebra and Geometry—A Second Course*, and in the paper *Representations of separable algebras* in volume 19 of Duke. If I had never met Paul, two problems might still be awaiting solution.

Happy birthday, Paul!

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_4,

2. Introduction

In a recent paper [1] Arno settled two questions that go back to Gauss: he showed that the classical list of 54 imaginary quadratic fields with class number 4 is complete, and deduced the completeness of the following well known list of 33 numbers not divisible by 4 which are uniquely expressible as a sum of three squares:

$$\begin{array}{ccccccccccc} 1, & 2, & 3, & 5, & 6, & 10, & 11, & 13, & 14, & 19, & 21, \\ 22, & 30, & 35, & 37, & 42, & 43, & 46, & 58, & 67, & 70, & 78, \\ 91, & 93, & 115, & 133, & 142, & 163, & 190, & 235, & 253, & 403, & 427. \end{array} \tag{1}$$

Remark 1. *(a) Through some mysterious slip the number 19 disappeared from Arno's list.*

(b) Until Sect. 7 uniqueness is always to be taken in the simple minded sense, ignoring order and signs.

(c) To get all integers which are uniquely expressible as a sum of three squares, take all products of the numbers in (1) by powers of 4.

It is timely to seek similar results for ternary forms other than $x^2 + y^2 + z^2$. Now there is substantial literature concerning formulas for the number of representations of integers by ternary forms. The relation between this literature and what I do here will be discussed in Sect. 7. In the body of the paper, in an elementary and self-contained way, I deduce from Arno's theorem results on three ternary forms. The final section explores three other forms.

The three selected forms are $x^2+y^2+2z^2$, $x^2+2y^2+2z^2$, and $x^2+2y^2+4z^2$. I chose these because in Dickson's book [4, pp. 96–97] they are discussed right after sums of three squares and used later in the book in treating Waring's problem for cubes. So this is my homage to Dickson, who introduced me to integral quadratic forms in a course during the summer of 1938, possibly the last time he presented his famous elementary course on number theory.

In stating Theorems 1–3 three lists extracted from (1) play a role. The first consists of the even numbers in (1):

$$2, 6, 10, 14, 22, 30, 42, 46, 58, 70, 78, 142, 190. \tag{2}$$

For the second list divide the entries of (2) by 2:

$$1, 3, 5, 7, 11, 15, 21, 23, 29, 35, 39, 71, 95. \tag{3}$$

The last consists of the entries in (1) which are congruent to 1 (mod 4):

$$1, 5, 13, 21, 37, 93, 133, 253. \tag{4}$$

Theorem 1. *The numbers uniquely represented by $x^2 + y^2 + 2z^2$ consist of (3) and 4^k6 $(k = 0, 1, 2 \ldots)$.*

Theorem 2. *The numbers uniquely represented by $x^2 + 2y^2 + 4z^2$ consist of (2), (4), and 4^k3 $(k = 0, 1, 2 \ldots)$.*

Theorem 3. *The numbers uniquely represented by* $x^2+2y^2+4z^2$ *consist of* (3) *and four even numbers:* 2*,* 10*,* 26*,* 74*.*

Note that in Theorem 3 the list is finite.

Although the information is not needed in this paper, it is probably helpful for the reader to have available the integers not represented by the four forms under discussion.

$$x^2+y^2+z^2 \quad \text{and} \quad x^2+2y^2+2z^2: \quad 4^k(8n+7),$$

$$x^2+y^2+2z^2 \quad \text{and} \quad x^2+2y^2+4z^2: \quad 4^k(16n+14).$$

3. Four One to One Correspondences

In this section we establish various equalities between numbers of representations. In each case a map will be defined, followed by a display of its inverse. In the first instance the routine verification that the product both ways is the identity will be presented; after that the details will be left to the reader.

Let A be a given odd integer. We define three numbers N, P, Q as follows: N is the number of representations of $2A$ by $x^2+y^2+z^2$, P the number of representations of A by $x^2+y^2+2z^2$, and Q the number of representations of A by $x^2+y^2+4z^2$.

Lemma 1. $N = P = Q$.

Proof of $N = P$*.* Given a representation $A = u^2+v^2+2w^2$ we pass to $2A = (u+v)^2+(u-v)^2+(2w)^2$. For the reverse map, suppose that $2A = r^2+s^2+t^2$ is given. Note that two of r, s, t must be odd and the third even. Say t is even. We pass to $A = [(r+s)/2]^2+[(r-s)/2]^2+2(t/2)^2$. After $u, v, ww \to u+v, |u-v|, 2w$, to perform the second map we note that u and v have opposite parities, so that $u+v$ and $|u-v|$ are odd. Thus the second map sends $u+v$, $|u-v|$, $2w$ back into u, v, w. For the product the other way, $r, s, t \to (r+s)/2, |r-s|/2, t/2$ and then back to r, s, t. □

Proof of $P = Q$*.* Given a representation $A = u^2+2v^2+4w^2$, we pass to $A = u^2+2v^2+(2w)^2$. For the reverse, given $A = r^2+s^2+2t^2$ we note that r and s have opposite parities. Assume that s is even and pass to $A = r^2+4(s/2)^2+2t^2$. □

Lemma 2. *Let* B *be any integer. Then the number of representations of* B *by* $x^2+y^2+2z^2$ *equals the number of representations of* $2B$ *by* $x^2+2y^2+2z^2$*.*

Proof. Given $B = u^2+v^2+2w^2$ we pass to $2B = 2u^2+2v^2+(2w)^2$. For the reverse, given $2B = r^2+2s^2+2t^2$ we note that r is even and pass to $B = 2(r/2)^2+s^2+t^2$. □

Lemma 3. *Assume* $C \equiv 1 \pmod 4$. *Then the number of representations of* C *by* $x^2+y^2+z^2$ *equals the number of representations of* C *by* $x^2+2y^2+2z^2$.

Proof. From $C = u^2 + 2v^2 + 2w^2$ we pass to $C = u^2 + (v+w)^2 + (v-w)^2$. For the reverse, write $C = r^2 + s^2 + t^2$ and note that $C \equiv 1 \pmod 4$ implies that two of r, s, t are even and the third odd. Say r and s are even. We pass to $C = 2[(r+s)/2]^2 + 2[(r-S)/2]^2 + t^2$. □

4. Partial Results Obtainable Without Reference to Class Numbers

The next lemma will be used in proving Lemma 5.

Lemma 4. *In a representation by* $x^2+y^2+2z^2$ *of a number divisible by* 8, x, y, *and* z *must be even.*

Proof. We have

$$8D = x^2 + y^2 + 2z^2. \tag{5}$$

Necessarily x and y have the same parity. Suppose that they are odd. Then, since the square of any odd number is congruent to $1 \pmod 8$, we have $x^2 + y^2 \equiv 2 \pmod 8$. The element $2z^2$ is congruent to 2 or $0 \pmod 8$, according as z is odd or even. In either case (5) is contradicted. Therefore x and y are even and it follows that z is even. □

Lemmas 5 and 6 are of course portions of Theorems 1 and 2. They are presented at this point to emphasize that their proofs do not require class number information. All that is needed is the following sharpening of the three square theorem, first published by Legendre in 1798: if a number is not divisible by 4 and not congruent to $7 \pmod 8$ then it has a representation as $x^2+y^2+z^2$ with the greatest common divisor of x, y, and z equal to 1. This quickly implies that the entries in (1) are square free. See [2, p. 304] for more details.

In short: Lemmas 5 and 6 could have been proved 200 years ago.

Lemma 5. *The even numbers uniquely represented by* $x^2+y^2+2z^2$ *are the multiples of* 6 *by powers of* 4.

Proof. Let $2F$ be uniquely represented by $x^2+y^2+2z^2$. Note in the first place that F is a sum of three squares. This follows from the form of the numbers represented by $x^2+y^2+z^2$ and $x^2+y^2+2z^2$, as exhibited at the end of Sect. 2. But we can see this at once directly: if $2F = u^2+v^2+2w^2$, u and v must have the same parity and we deduce $F = [(u+v)/2]^2 + [(u-v)/2]^2 + w^2$. A similar remark is needed in Lemma 6 and will be left to the reader.

Write $F = r^2+s^2+t^2$. Then $2F = (r+s)^2+(r-s)^2+2t^2$. We can permute r, s, and t and so r and s are eligible to play the role of t. We get more than one representation of $2F$ by $x^2 + y^2 + 2z^2$ unless $r = s = t = G$, say. Thus $F = 3G^2$. Furthermore, this argument is applicable to any expression of F as a sum of three squares and so $F = G^2 + G^2 + G^2$ is the only representation of F a sum of three squares. By the remarks preceding the statement of Lemma 5, we see that F is a multiple of 3 by a power of 4, and so $2F$ has the form $4^k 6$. We still have to verify that these numbers $4^k 6$ are uniquely represented by $x^2 + y^2 + 2z^2$. The number 6 is checked by inspection, and Lemma 4 looks after the others. □

Lemma 6. *Assume that H is congruent to $3 \pmod 4$ and that H is uniquely represented by $x^2 + 2y^2 + 2z^2$, then $H = 3$.*

Proof. The argument is similar. We observe that H is a sum of three squares. When we write $H = r^2 + s^2 + t^2$ all three are odd. With any choice for r we can write

$$H = r^2 + 2[(s + t)/2]^2 + 2[(s - t)/2]^2.$$

Therefore $r = s = t = J$, say. Moreover $H = J^2 + J^2 + J^2$ is the only representation of H as a sum of three squares. It follows that H must be 3.□

This section concludes with another preliminary lemma.

Lemma 7. *Suppose that $2K$ is uniquely represented by $x^2+2y^2+4z^2$. Then K is uniquely represented by $x^2+2y^2+2z^2$, and in this unique representation $y = z$.*

Proof. Suppose that

$$K = u^2 + 2v^2 + 2w^2 = r^2 + 2s^2 + 2t^2.$$

Then

$$\begin{aligned} 2K &= 2u^2 + (2v)^2 + 4w^2 = 2u^2 + 4v^2 + (2w)^2 \\ &= 2r^2 + (2s)^2 + 4t^2 = 2r^2 + 4s^2 + (2t)^2. \end{aligned}$$

The assumed uniqueness shows that $u = r$ and that $v = w = s = t$. □

5. Proofs of Theorems 1–3

Proof of Theorem 1. Lemma 1 has looked after the odd numbers uniquely representable by $x^2 + y^2 + 2z^2$, showing that they are given by (3). Lemma 5 looks after the even numbers. □

Proof of Theorem 2. Lemma 2 takes care of even numbers, Lemma 4 accounts for the odd ones congruent to $1 \pmod 4$, and Lemma 6 asserts that for those congruent to $3 \pmod 4$ the only possibility is 3. □

Proof of Theorem 3. Lemma 1 looks after odd numbers. For even numbers, Lemma 7 is ready. It calls for us to examine all the numbers in Theorem 2 and check whether in their unique representation by $x^2+2y^2+2z^2$ we have $y=z$. The survivors get doubled. The work is facilitated by initially discarding all numbers in Theorem 2 which are not represented by the binary form x^2+4y^2. It turns out that the survivors are 1, 5, 13, and 37; their doubles are 2, 10, 26, and 74. □

6. The Literature

As remarked in Sect. 2, there is a substantial literature on the number of representations of integers by ternary forms. Here the representations are being counted in the inflated sense, meaning that both the order and the signs of the summands are taken into account. During the 1920s there was a flurry of work, including the papers [3, 5], and [9] by Bell, Jones, and Uspensky.

(It gives me pleasure to cite a research paper by Eric Temple Bell; I had the privilege of meeting him during a 1950 visit to Caltech. Several generations of readers, including myself, have appreciated his expository and historical writing. Let me mention also his science fiction, written under the pseudonym John Taine. I specially recommend *Green Fire*; *The Purple Sapphire* and *Quayle's Invention* are also excellent. *The Iron Star* is still in print. Constance Reid has informed me that her current project is a biography of Bell.)

A fairly definitive result appears as Theorem 86 on page 194 of [6]. It applies to all positive definite ternary forms. As is to be expected, what is being counted is the total number of representations by all the forms in a genus. Separating out a single form in a genus is an extra enterprise; the papers [7] and [8] make a contribution to this.

Let me cite a sample from [5], using the notation of that paper. Jones writes $N(n)$ for the number of representations of n by $x^2+y^2+z^2$ and $A(n)$ for the number of representations of n by $x^2+y^2+2z^2$. Then $A(n)=N(2n)/3$ when n is odd; this is quite parallel to the $N=P$ portion of Lemma 1, and leads quickly to Theorem 1 for odd numbers. The next statement is $A(2n)=N(n)$. This has no counterpart in my paper; it is a typical example of the fact that counting *all* the representations usually leads to neater formulas. But if the target is unique representation in the simple minded sense then for even numbers we must argue as in Lemma 5.

It is only for a handful of forms that we are ready to derive results on unique simple minded representations; for the others we have to wait for more information on class numbers.

7. Concluding Remarks

(a) *The form* $x^2+2y^2+3z^2$. This form is treated in [4, p. 101] by a reduction to $x^2+2y^2+2z^2$. Nevertheless, this is one of the forms that is waiting for more information on class numbers. But I was curious and Noam Elkies generously took the time to program and run a computation up to 16,383($=2^{14}-1$). For odd numbers the resulting list was (6).

$$1, 5, 7, 13, 17, 23, 47, 55. \tag{6}$$

For even numbers the result is recorded in Theorem 5 below. This was sufficiently striking that I decided to return to the drawing board to see whether I could prove it *now*. It turned out that there was enough information available to derive the needed preliminary theorem on a case with two representations.

Theorem 4. *The even numbers with exactly two representations by* $x^2+y^2+2z^2$ *are*

$$2, 4, 12, 22, 38, 44, 86, 134, 326 \tag{7}$$

and multiples of the entries in (7) *by powers of* 4.

From this Theorem 5 was deducible.

Theorem 5. *The even numbers uniquely represented by* $x^2+2y^2+3z^2$ *are the odd powers of* 2.

Theorem 5 is an addition to the growing list of theorems suggested by a computer and then proved.

The proofs of Theorems 4 and 5 follow the pattern of those of Theorems 1 –3, with appropriate minor changes. I am omitting them, offering them as exercises for the reader.

(b) *The forms* $x^2+2y^2+3z^2$ *and* $x^2+2y^2+3z^2$.

One way of measuring progress is by the size of the discriminant. The forms $x^2+y^2+z^2$ and $x^2+y^2+2z^2$ are the only ones with discriminants 1 and 2, respectively. There are two forms of discriminant 3: $x^2+y^2+3z^2$ and $x^2+2y^2+2z^2$. However, the latter is so closely related to $x^2+2y^2+3z^2$ that it merits no attention at this time. So I regard $x^2+y^2+3z^2$ as the next challenge. It, too, has to wait for class number information but right now it is another tempting target for computation.

It is convenient to bring in $x^2+3y^2+3z^2$ as well, because then we can ignore multiples of 3, which simply bounce us back and forth between these two forms. Elkies found (8) and (9) for the numbers prime to 3 upto 16,383 uniquely represented by $x^2+y^2+3z^2$ and $x^2+3y^2+3z^2$, respectively. (Note that the entries in (9) have to be congruent to 1 (mod 3) to be eligible for any kind of representation.) To get the full answer for $x^2+y^2+3z^2$ adjoin multiples of (8) by even powers of 3 and multiples of (9) by odd powers of 3. For the other form the parities are reversed.

$$\begin{array}{ccccccccccccc} 1, & 2, & 7, & 10, & 11, & 14, & 19, & 22, & 23, & 26, & 31, & 34, & 38, \\ 46, & 47, & 55, & 59, & 70, & 71, & 86, & 94, & 115, & 119, & 154, & 166. & \end{array} \tag{8}$$

$$1, 10, 13, 22, 34, 37, 46, 58, 82, 85, 130, 142, 190, 253. \tag{9}$$

The lists (6), (8) and (9) could be compared with those obtainable by assuming the completeness of the existing lists of imaginary quadratic fields for class numbers bigger than 4. I have not undertaken this comparison as yet.

(c) *Numbers of the form* $8n + 3$ *with two representations as a sum of three squares.* These numbers need not wait for more class number information: they can be treated right now. The task is entirely straightforward. There is, however, one thing to note. It will not be the case here that the numbers are square free. This introduces a complication; the technique needed to cope with it is available in pages 307–308 of [2].

(d) *The form* $xy + xz + yz$. Here is a different chapter in the subject: the representation of an integer n by $xy + xz + yz$. This is a ternary quadratic form, but it is indefinite. If we allow x, y, and z to take both positive and negative values, then any n would have an infinite number of representations. The situation is remedied by a restriction to positive values. More precisely, we take n positive and allow x, y, z to be 0 as well as positive (note that in fact only one of the three can be 0). Then there is in this context again a connection between the number of representations and class numbers. Indeed the number of representations of n (in the inflated style where the order counts, but of course this time signs do not enter) is three times the number of equivalence classes of positive definite binary quadratic forms of discriminant n. To be honest, for this to work perfectly requires a little creative accounting: representations that use a zero are given weight 1/2, and the forms $a(x^2 + y^2)$ and $a(x^2 + xy + y^2)$ get weights 1/2 and 1/3, respectively.

The idea goes back to Liouville. It was taken up again in the 1920s by Mordell and Bell. The note [10] by R. F. Whitehead presents a model concise proof in half a page. Starting with Whitehead's note, a curious reader can trace the earlier references.

Suppose we examine the number of representations in the simple minded style. It takes only a glance to see that the integers with exactly one representation are 1 and 2. Those with exactly two representations are 4, 5, 10, 13, 22, 37, and 58. This could have been proved around 1970 as soon as Baker and Stark (separately) found the classical list of imaginary quadratic fields with class number 2 to be complete. Of course more can be done now that we are up to class number 4, but I leave that story for another day.

In closing I heartily thank Noam Elkies. In addition to the computations noted above, he ran others that confirmed the correctness (upto 16,383) of Theorems 1–4.

References

1. Steven Arno, The imaginary quadratic fields of class number 4, *Acta Arithmetica* **60** (1992), 321–334.
2. Paul T. Bateman and Emil Grosswald, Positive integers expressible as a sum of three squares in essentially one way, *J. of Number Theory* **19** (1984) , 301–308.
3. E. T. Bell, The numbers of representations of integers in certain forms $ax^2 + by^2 + cz^2$, *Amer. Math. Monthly* **31** (1924), 126–131.
4. Leonard Eugene Dickson, *Modern Elementary Theory of Numbers*, Univ. of Chicago Press, 1939.
5. Burton W. Jones, The number of representations by certain positive ternary quadratic forms, *Amer. Math. Monthly* **36** (1929), 73–77.
6. Burton W. Jones, *The Arithmetic Theory of Quadratic Forms*, Carus Monograph 10, Math. Assoc. of America, 1950.
7. G. A. Lomadze, Formulas for the number of representations of numbers by certain regular and semi-regular ternary quadratic forms belonging to two-class genera , Acta Arith. **34** (1977), 131–162. (Russian).
8. Gordon Pall, Representation by quadratic forms, *Can. J. of Math.* **1** (1949), 344–364.
9. J. V. Uspensky, On the number of representations of integers by certain ternary quadratic forms, *Amer. J. of Math.* **51** (1929), 51–60.
10. R. F. Whitehead, On the number of solutions in positive integers of the equation $yz + zx + xy = n$, *Proc. Lon. Math. Soc.* **21** (1922), xx.

(Added July 26, 1993). The existence in a countable set of uncountably many infinite sets with finite intersection was problem B-4 in the 1991 Putnam examination; see the Amer. Math. Monthly 98 (1991), p. 322. Of the 199 top contestants, 52 got it. We have some very bright undergraduate students!

(Added February 15, 1995). Constance Reid's biography is out: *The Search for E. T. Bell*, Math. Assoc. of America, 1993.

Did Erdős Save Western Civilization?

Cedric A. B. Smith

C.A.B. Smith (Deceased)
The Galton Laboratory, Department of Genetics and Biometry,
University College London, Wolfson House, 4 Stephenson Way,
London NW1 2HE, England

1. Introduction

If you stand on the famous Chain bridge in Budapest, you will see below you the broad sweep of the Danube. But this broad river arose from the confluence of many small streams. Indeed, there is a point near St. Moritz, where if a rain drop happens to fall a few centimeters to the north, it will make its way into the Rhine, and so to the North Sea. If it falls a little to the west, it will join the Adda and the Po, and end up in the Adriatic, whereas to the east it would run into the Inn, the Danube, and the Black Sea. An apparently negligible movement at the start can make a difference of hundreds of kilometers later on.

We can compare this to the flow of events in life. Each event has its own sequence of consequences. The consequences of different events will flow together, and evolve to and fro in an unpredictable manner like the sinuous bends of a river. But there is one important difference. A river is the union of its component streams. If one stream dries up, it will make only a trivial difference to the final flow of the river. But an event is the intersection of previous events, in the sense that if only one apparently small and unimportant component fails, it may make a very great difference to the final consequences.

2. The Erdős Conjecture

Here we look at a typical conjecture of Erdős's, one which was apparently a conjecture of no special importance, and trace the flow of events resulting from it and its combination with other flows, to see how, very plausibly, it became of considerable importance to us all.

Around 60 years ago Erdős was a young man in Cambridge. He suggested the conjectures

(a) Any dissection of a square into a finite number of smaller squares must contain at least two squares of equal size,
(b) At least two such equal squares must touch. (We take the word "dissection" to have its obvious meaning. To be strictly accurate, we

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_5,

> would say that the union of the component squares is the original square, and any two component squares intersect, if at all, in part or all of an edge. But, in what follows, we will try to present the discussion in a simple and obvious way, leaving matters of complete rigor to be supplied, if desired, by the reader.)

This conjecture was noticed by W. R. Dean, then a Cambridge lecturer (later a London professor). He sometimes visited Christ's Hospital (a famous boys' school). On one such visit he mentioned the conjecture, conceivably thinking that it might be decided by some of the more capable mathematically inclined pupils. At least one of the pupils, Arthur H. Stone, took note of the conjecture. Soon afterwards, he gained a scholarship to Trinity College Cambridge.

3. The Cambridge Students

Among the other schoolboys sitting the scholarship examination was one, Cedric A. B. Smith, feeling very miserable. His great ambition was to study mathematics at Cambridge. But he had just failed one vital examination. And now he was faced with an impossibly difficult exam paper. The situation seemed desperate, and he was near to bursting into tears. But at that moment someone, rejoicing in a warm sunny summer's day, walked past, cheerfully whistling the Mexican waltz, "Over the Waves." Life once more seemed worth living—perhaps the exam paper was not totally impossible. And if, unlike Arthur Stone, Cedric Smith did not get a scholarship, at least he was admitted to Trinity, and his ambition was fulfilled. (Incidentally, both he and Arthur were told that their applied mathematics was much better than their pure, which explains how it comes about that Stone is now a leading topologist.)

So, not long afterwards, Cedric Smith walked into his very first university lecture, on geometry, with great excitement. What new revelations were in store? He already knew much geometry—about angles, areas, lines, rectangles, triangles, circles, conics, poles and polars, tetrahedra, dodecahedra, and all that nonsense. But the lecturer mentioned none of that. He added points together, and multiplied points by noncommutative numbers. At the end, Smith said to the young man next to him, "That was very confusing." The reply was, "Not at all. I thought it was a very good lecture. When is the next lecture?"—"At 11."—"No it isn't, it's at 10."

So the two walked into a 10 o'clock lecture. They did not know that there was a typing error in the timetable, and that it was an advanced lecture. They did not know that, because the lecturer's Russian accent was so thick that for 30 min they did not understand one word.

After the lecture, Smith found that his new friend was R. Leonard Brooks, the future discoverer of Brooks's Theorem in Graph Theory (though that was not obvious at the time.)

Later on, when Mr. Besicovitch produced some weird mispronunciation of an English word, and the class roared with laughter, he sternly defended himself. "Feefty meelion peeple speek yore kind of Eenglesh. Fife handred meelion peeple speek my kind of Eengleesh."

Came the Christmas vacation. Smith went out shopping, and met someone he thought he had seen in lectures. They both stopped. The other young man said:

"What are you doing here?"— "I live here."

There followed a long silence, then Smith asked:

"What are you doing here?"—"I live here."

Another long silence, then the other said:

"I must do the shopping."

and walked off. But when they got back to Cambridge, it turned out that the other was Arthur H. Stone, and that, like Brooks, he had a room in New Court. Smith introduced Stone to Brooks, while Brooks introduced us to a student of chemistry, William T. Tutte, also of New Court, saying that Tutte was good at chess.

While we knew that we were real mathematicians, we were still broad minded enough to talk to someone who was only a chemist. Tutte put a problem to us: find a semipotential function, i.e., a function $S(x)$ such that

$$S(S(x)) = \exp x. \tag{1}$$

We couldn't. So Tutte said: let a be such that $S(a) = a$. Then

$$\exp a = S(S(a)) = S(a) = a. \tag{2}$$

For example, we might have $a = 0.318 + 1.337i$. Differentiate (1) repeatedly, and substitute a for x in each relation. We find the successive derivatives of the function $S(x)$ at a, and hence its Taylor series.

This looked plausible, though we weren't then sure if it converged. Here is a sketch of a justification. Write $\ln_2 x$ for $\ln\ln x$, $\ln_3 x$ for $\ln\ln\ln x$, and so on. If x is near a, the iterated sequence $\ln_n x$ converges geometrically to a, with asymptotic ratio $\ln' a = 1/2$. Hence as n increases, the limit

$$f(x) = \lim[a^n(\ln_n x - a)] \tag{3}$$

exists, and satisfies $f(\exp x) = a \cdot f(x)$, i.e.,

$$\exp x = f^{-1}(a \cdot f(x)). \tag{4}$$

So $S(x)$, defined as $f^{-1}(\sqrt{a}f(x))$ satisfies (1), and its Taylor series can be found exactly as Tutte did.

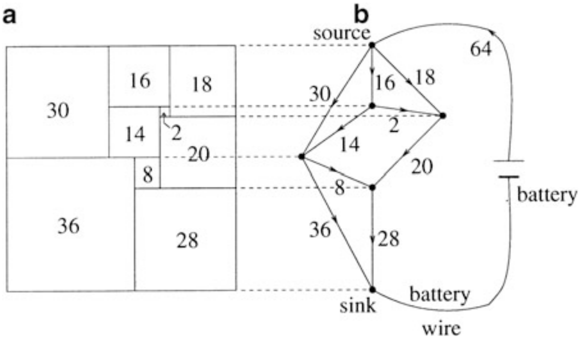


Fig. 1 (**a**) Moroń's squared rectangle, with horizontal side 64 and vertical side 66. (**b**) An electrical network representing Moroń's rectangle, with total current 64 and potential drop 66

4. Squaring a Square

Stone then introduced us to the question of whether a square can be divided into unequal squares (though he didn't then know that it came from Erdős.) We spent some 3 years working on the problem.

The first advance was to realize it was easy to find rectangles divided into unequal squares, as in Fig. 1a. We drew a rough figure, something like Fig. 1a. Then the conditions that the interior squares fitted together gave a set of homogeneous linear equations. We solved these (using suitable short cuts). The result was always a rectangle of uniquely defined shape, although clearly it could be magnified in size by any constant factor. Unfortunately, the two sides were never equal, and never had any very simple ratio. (We later found that the rectangle shown in Fig. 1a had been already discovered by Moroń [1925].)

The second stage was to replace the rectangle by an electrical network of wires, all of unit resistance, as shown in Fig. 1b.

Each horizontal line in the rectangle becomes a node in the network. Each square becomes a wire (edge) joining the two nodes corresponding to the horizontal lines between which the square lies. Each wire carries a current equal to the side of the corresponding square. Since the wires have unit resistance, this equals the potential drop along the wire. The linear relations stating that the squares fit together become Kirchhoff's (1847) laws: the first law, that the total current entering any node must equal the total current leaving it, except for the "source" and "sink" nodes at the top and bottom respectively, and the second law that the total change in potential round a circuit must be zero.

It follows that the total current is equal to the horizontal side of the corresponding rectangle, and the total potential drop between source and sink is equal to the vertical side.

We can, if we wish, “complete” the network by adding a “battery wire” joining the sink and source, containing a battery providing the electromotive force necessary to drive the currents through the network, as in Fig. 1b. Obviously, we can rotate the rectangle through a right angle, so that “horizontal” and “vertical” are interchanged. If we do this we get a new network. The relation of this to the former one is simply that the two completed networks are topological duals.

We can again calculate the currents (= sides of component squares) by noting that Kirchhoff’s laws provide linear equations, and the solutions are given by determinants. But a more interesting way is to use spanning trees, following Kirchhoff [1847].

For the moment, consider a spanning tree in the network, one not including the battery wire, as in Fig. 2. Imagine that only the wires in this tree are now conductive: we could imagine the other wires cut. Let a unit current flow from source to sink, as in Fig. 2. It will follow a uniquely defined path. Repeat for every such spanning tree, and add the currents. We get the currents shown in Fig. 1b.

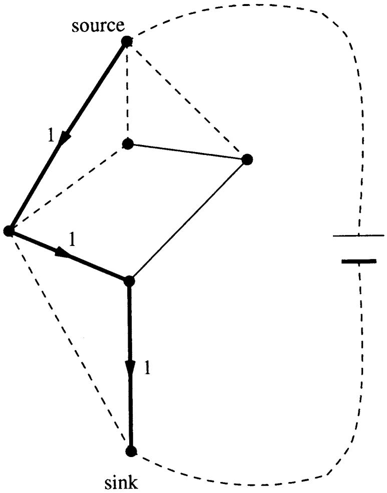


Fig. 2 A unit current flowing through a spanning tree

To show that in general this procedure gives the network currents, we show that Kirchhoff’s two laws hold. Clearly Kirchhoff’s first law holds for the current in each spanning tree, as in Fig. 2, and hence in the total.

To verify Kirchhoff’s second law, we introduce a new construction. The trees already considered are those spanning trees in the completed network which do not include the battery edge. Now take a spanning tree which does include the battery edge, and again suppose that all edges not in this tree are cut. Then the nodes will divide into two classes, those connected through the tree to the positive pole of the battery, and those connected to the negative pole. To nodes in the first set give potential 1, and to those in the second set potential 0. Repeat for all spanning trees containing the battery edge, and add the potentials. We get a set of potentials which necessarily obey Kirchhoff's second law. It remains to show that the current in any wire, as derived by the first construction, is equal to the difference in potential between its end nodes, as derived by the second construction. But that is a consequence of the theorem, that if e_1 and e_2 are two edges in a graph, then a spanning tree including e_1 but not e_2 provides in the obvious way one including e_2 but not e_1. Take e_1 and e_2 to be the battery wire and the wire in question.

From these results it follows that the total current flowing through the network (= horizontal side of the corresponding rectangle) is equal to the number of spanning trees in the network which do not include the battery wire. We called this the "complexity" of the (uncompleted) network. The total potential drop (= the vertical side of the rectangle) equals the number of spanning trees in the completed network which do include the battery wire. But spanning trees in graphs are the same as the bases of the corresponding matroids. In fact we were dealing with regular patroids [Smith 1972, 1974].

A surprise occurred when Brooks cut up a squared rectangle into its component squares to form a jigsaw, and challenged his mother to put them together again to form a rectangle. She did so, but it was not the rectangle he had started with. The phenomenon was explained by Tutte. As an example, consider the electrical networks of Fig. 3a, b, corresponding to two rectangles made up of the same squares differently arranged. (Both have total current = horizontal side 1,025 and total potential drop = vertical side 592). (These are not the rectangles tried out on Brooks's mother, but two new rectangles constructed by Tutte.)

Tutte pointed out that the nodes P, Q, in Fig. 3a are symmetrically related to A, B, and C, so they have equal potentials equal to the average of the potentials of A, B, and C. So one can pick Q up and put it down on top of P without changing currents, thus deriving Fig. 3b.

Tutte now went on to change Fig. 3a–c, by reflecting the interior of the triangle in a horizontal mirror, leaving the external connections unaltered, obtaining in Fig. 3c, d, another pair of rectangles composed of the same squares.

Acknowledgment. Figure 3 was taken (with suitable modifications) from Figs. 31 and 32 of J. D. Skinner, *Squared Squares: Who's Who and What's What*. With kind permission from the author.

Now a new surprise, all four networks had the same total currents, and total potential drops, so that all four corresponding rectangles had the same horizontal and vertical sides, though not all the component squares were shared. Tutte explained that as follows. The uncompleted networks 3a and 3c are topologically isomorphic, hence contain the same number of spanning trees, and hence the same total currents. If we add battery wires, then any

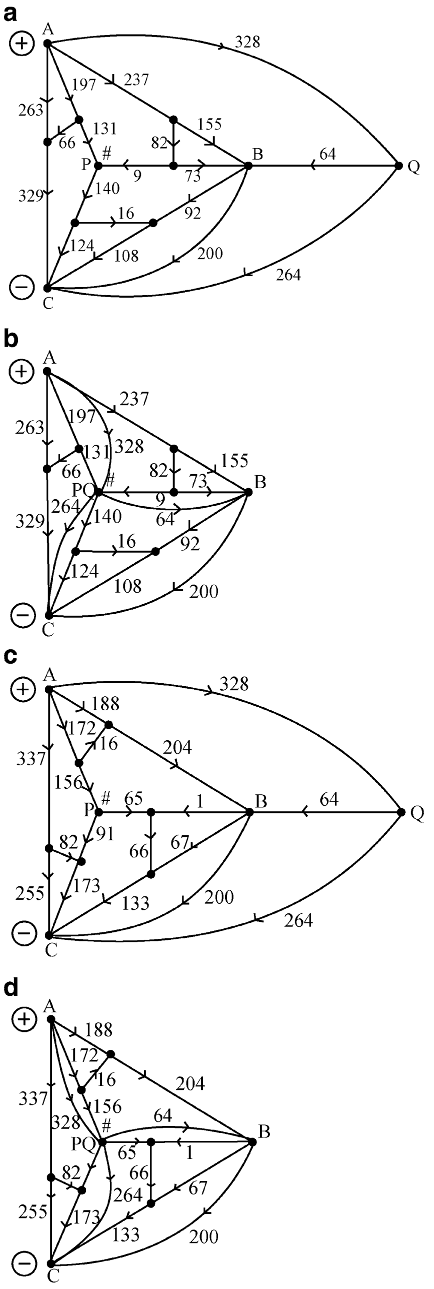


Fig. 3 Four electrical networks, each representing a squared rectangle with horizontal side 1,025 (= total current) and vertical side 592 (= total potential drop). (**a**) and (**b**) correspond to rectangles composed of the same squares differently arranged, and so also do (**c**) and (**d**)

spanning tree including the battery wire and the wire BC in 3a gives rise to one including the battery wire and the wire $A'B'$ in 3c. So the numbers of spanning trees including the battery wire are the same in 3a and 3c, so the total potential drops are equal.

Brooks, on the one hand, and Stone and Smith on the other, simultaneously and independently carried these ideas further, and each succeeded in producing two rectangles of the same size and shape containing all different squares, with the exception that two corner squares were the same. By placing these two rectangles so that the equal corner squares were superimposed, they each succeeded in producing a square divided into unequal squares, by the construction shown in Fig. 4.

Erdős's conjecture was incorrect!! (Stone et al. 1940). (But Sprague 1939, beat us by a short head.)

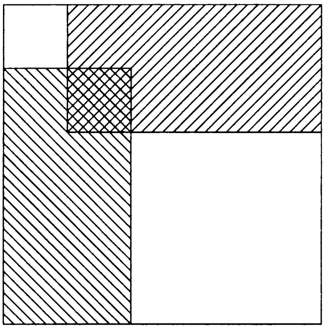


Fig. 4 A method of obtaining a square dissected into unequal squares from two squared rectangles having only one square in common, and that at a corner in each rectangle

A number of other workers have since been interested in the problem, and the state of investigation up to date is described in Skinner [1993]. However, from the point of view of the present discussion, the important fact is that by this time that, notwithstanding that we knew that Tutte was a chemist, we were extremely impressed by his mathematical ability.

5. Wartime Developments

The year 1939 was rather like the present (1993), in that there was severe unemployment. The students of Trinity College, Cambridge, organized a camp in the Lake District to give some unemployed both a holiday, with excursions, and also an opportunity to do some work making a lakeside path. One of the participants was a Tutor (later professor) Patrick Duff. But one day it rained hard, Patrick Duff's tent collapsed, and he was taken into the big house to dry. Cedric Smith followed to talk to him. Patrick Duff said:

"Do you know Bill Tutte?"–

"Yes." –

"We're very worried. He's no good" –

"He got a first class degree." –

"His supervisor is disappointed" –

"Well, he's very good at maths." –

"Prove it."

So Smith sent a letter to the College, detailing Tutte's achievements. But Trinity sounded rather unimpressed.

Quite soon war broke out. Stone had just arrived in Princeton with a Visiting Fellowship. He wrote back to the remaining three, saying that he had met a Hungarian named Erdős, who pronounced English in a Hungarian manner, so that "pineapple upside down cake" became "pinnayopp-play oopshiday dovn tsockay," But that did not prevent them collaborating mathematically.

Quite soon Tutte joined a group at Bletchley Park. What they did there was a profound secret. But the group contained some of Britain's best mathematicians, and was strongly suspected to dealing with codes and ciphers. Why was Tutte asked to go to Bletchley, and not to a chemical establishment? Presumably through Smith's letter, declaring him to be an excellent mathematician, even though previously the letter had a cool reception.

For 3 years Smith was a porter at Addenbrooke's Hospital, Cambridge. One day near the end of the war he was cycling past the hospital when he saw Prof. Duff, and waved. Prof. Duff shouted "Stop! STOP!! STOP!!!"

"What's the matter!!??" –

"Do you know Bill Tutte's address?" –

"Yes, Why??"

"We've been told to elect him to a Fellowship at Trinity. But they didn't tell us why, or what his address is. So we can't send a telegram to congratulate him."

The past may now be only a memory. But during the war the Germans had occupied most of western Europe, so that Britain had to get all vital supplies across the Atlantic, constantly menaced by German submarines. The situation was not at all good. It would have helped to be able to decode German messages. The Poles had secretly examined a German coding machine as it was transported across the Polish Corridor, so the coding device was known. It was simple, but fiendishly clever, and was found very difficult to decode by even the top mathematicians at Bletchley. Very plausible rumor says that Tutte supplied the vital clue, possibly avoiding military defeat. The consequences of a Nazi victory would have been horrific—civilization was saved. And it all began with a conjecture by Erdős!

6. Postwar Developments

To round off this story, we may briefly say what eventually became of the various dramatis personae. W. R. Dean became Professor at University College London. Leonard Brooks became an Income Tax Inspector—but continued to develop mathematical ideas. Arthur Stone was another Fellow of Trinity, and after a time at Manchester, he and his wife Dorothy both became Professors of Mathematics at the University of Rochester, NY. Bill Tutte, now truly a mathematician, via Toronto became Professor at Waterloo, then Fellow of the Royal Society of Canada, Distinguished Professor, Fellow of the Royal Society of London, and Founding President of the Institute of Combinatorics and its Applications. Not bad for only a chemist!

As for Cedric Smith, he heard that the Galton Laboratory needed a statistician. (The Galton lab is the human genetics section at University College London.) This seemed an interesting possibility for a first academic post. He applied, and was interviewed by Prof. Lionel Penrose, and by Prof. J. B. S. Haldane, who greeted him with "you know why you've come here. Prof. Penrose thinks there might be a job for you. But I don't think so." Nevertheless it did turn out to be a reasonable first job, for Smith is still there after 47 years.

This gave him a good chance to get to know the famous Erdős in person. Thus, after he had got married, his mother-in-law complained of the phone ringing and a deep voice (guess who) saying, "How are you? Vair is your slave? Is he preaching?" (in conventional English = Where is your husband? Is he lecturing?). Said she, indignantly, "I haven't got a slave."

The University of London ran a series of seminars. Erdős attended every One. That is to say, he sat there for the first 15 min, with a look on his face as if to say "I know all that." Then he walked out. There came the time when Erdős himself was the speaker. Could he walk out on himself? It seemed improbable. We waited to see. After 15 min he turned, faced the audience, said "Rado can explain all this better than I can.", and sat down.

Later there came an invitation from the Bolyai Mathematical Society to attend a combinatorial conference in Hungary, and so to see Erdős in his own town. Smith and Tutte went to the Hungarian Legation in London to get visas. "What can we do for you gentlemen?" asked the official.—"We want to go to a mathematical conference."—"We know all about you, gentlemen. We can't give you visas."—"Why not?"—"Permission has not yet come through from Budapest."—"Well," said Smith, "if you don't give me a visa now, I'm not going." Anxious phone calls followed, then the official said, "You are very fortunate, gentlemen. Permission has just come through from Budapest."

Smith took the visa home, then looked at it. As he remembers, it said something like "Egy bemenet, semmi kimenet." He phoned his father-in-law and asked, "Doesn't that mean 'one entry, no exit?' "—"Yes. What about it?"–"I would like to come back."–"Well, if you're worried, I'll phone the Hungarian legation." They replied, "We are not mathematicians. We are only lawyers. We say that if you go in you must come out. What's worrying you?"

So next day Smith found himself, feeling hungry, in a train crossing Belgium. He went to the dining car. Opposite him sat someone reading a German chemical treatise. Smith speaks little German, and stayed silent.

However, when Smith held up a banknote to pay the bill the German leaned over and said "C'est trop." Smith tried to say, in schoolboy French, that he only wanted change. The Frenchman asked, "Quelle est votre nationalité?", to which Smith replied "Anglais." There was a long silence, which did not seem too polite, so Smith asked the Frenchman, "Quelle est votre nationalité?"—"Oh, I'm American. 'What are you going to Frankfurt for?"—"I'm not going to Frankfurt."—"Then where are going?"—"To Budapest."—"Budapest?", said the American, "Budapest? that's where I was born. What are you going to Budapest for?"—"A mathematical conference."—"A mathematical conference", said the Hungarian, "I know some mathematicians. Do you know a mathematician called Erdős?"—"Everyone knows Erdős."—"His father was my mathematics schoolteacher."—"He will be at the conference."—"Give him my regards. Do you know a mathematician called Stone?"—"There are lots of mathematicians called Stone. But the only one I know went to America and married a mathematician."—"Was her name Margaret?"—"No, her name is Dorothy."—"Then that must be the Stone I know."—"He will be at the conference."— "Give him my regards."

At Wien Smith found an old second class carriage, more decrepit than anything he ever remembered seeing, labeled "Orient express, Budapest, Bucuresti", and drawn by an ancient steam engine.

When it got to Hegyeshalom they put an electric locomotive onto the train, and it rushed through the darkness until it cam to a hill of lights, and a broad river. Itt a nagy szép híres főváros Budapest. Ilyen boldogság!! But what to do next?

Here he was in the home town of not only Erdős but also of very many other famous mathematicians. Dr. Surányi, the Society Secretary met him and took him to his hotel. A group of mathematicians were there, possibly including Erdős. At the next table sat someone wearing a St. Johns College Cambridge tie. Smith nervously asked him "Don't I know your tie?"—"No, you don't. You think it's a St. Johns tie. It isn't. It's my old school tie. The stripes are narrower." Smith looked up. There was Gabriel Dirac, come for the conference, wearing a genuine St. Johns tie. The ties were compared. There was no visible difference.

The conference went wonderfully well, with Hungarian hospitality and organization. But that is a story for another occasion.

Erdős is not only one of a line of most distinguished mathematicians coming from Hungary. He must also be one of the most respected, and one treated with the greatest affection by all mathematicians. It is both a great pleasure and a great honor to greet him on his 80th birthday, "much love to Uncle Paul",

"Sok szeretettel a Pali Bácsikának!!"

References

1. Brooks, R. L., Smith, C. A. B., Stone, A. H., Tutte, W. T. (1940): The Dissection of Rectangles into Squares. Duke Math. J. **7**, 312–340.
2. Kirchhoff, G. (1847): Ueber die Auflősung der Gleichungen, auf welche man bei der Untersüchung der linearen Vertheilung galvanischer Strőme geführt wird. Annalen der Physik und Chemie **72**, 497–.
3. Moroń, Z. (1925): 0 rozkladach prostokatów na kwadraty: Przegląd Matematyczno-Fizyczny **3** 152–153.
4. Skinner, J. D. (II) (1993): Squared Squares, Who's Who and What's What. Lincoln, Nebraska, J. D. Skinner.
5. Smith, C. A. B. (1972): Electric Currents in Regular Matroids. In Welsh, D. J. A, and Woodall, D. R., eds., Combinatorics, 262–284, Southend-on-sea, Institute of Mathematics and Applications.
6. Smith, C. A. B. (1974): Patroids. Journal of Combinatorial Theory **16**, 64–76.
7. Sprague, R. P. (1939): Beispiel einer Zerlegung der Quadrats in lauter verschiedene Quadrate. Matematische Zeitschrift **45**, 607–608.

Encounters with Paul Erdős

Arthur H. Stone

A.H. Stone (Deceased)
Department of Mathematics, Northeastern University, Boston, MA 02115, USA

1. Encounters with Paul Erdős

My first encounter with Paul Erdős was curiously indirect. As a pre-undergraduate at Cambridge (England) in 1934, I learned from one of the Trinity College tutors that a mathematician named Erdős, passing through Cambridge, had mentioned an intriguing conjecture (attributed to Lusin, I believe), implying that a square could not be dissected into a finite number of unequal smaller square pieces. I passed this problem on to three fellow students, and we eventually found methods that produced counterexamples [1]. Of recent years the advent of high-speed computing has given rise to a considerable industry listing large numbers of dissections of squares into unequal squares ([2] and [6] for example), an industry that could continue indefinitely as there are infinitely many different dissections of this kind.

I first met Erdős a few years later (around 1940[1]) at Princeton, where I was a graduate student. Nearly all the mathematicians there were friendly and approachable, but Erdős excelled them all in that he was always willing to listen, with attention and encouragement, to other people's mathematics, even to lowly graduate students. He listened to my thesis results as they emerged, as he had listened to my wife's (already completed) thesis. This, we found, was and is typical of him.

When I next met Erdős we were at Purdue (1942?[2]), during World War II. His arrival there, for a stay of some months, began a little inauspiciously. Before setting out for the mathematics department, he had engaged a room and left his luggage there. At the mathematics department he found he needed something from his room, and Douglas Olds, of the Purdue faculty, volunteered to walk there with him. As Professor Olds told it, they walked and walked and walked, until he asked Erdős "Is it much further?" "Fascism Stalinism!" was the reply "I thought *you* were taking *me* there".

As this indicates, Paul was (and is) strongly interested in politics, with rather liberal views. He disapproved both of Communism as practiced in

[1] P. Erdős remarks: autumn 1939

[2] P. Erdős remarks: 1943

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_6,

Joedom and of some of the American reactions to it. “Joe” and “Sam” often rated disapproving remarks. He also disapproved, in general, of Providence, which he would refer to as the “S. F.” (Supreme Fascist). At that time he espoused the cause of China—not then “Red China”—and used to raise quite a bit of money for this cause by volunteering to drink small quantities of “poison” (i.e., alcoholic beverages) for so many dollars for China. (Nowadays the U.S. custom for raising money for good causes is for volunteers to walk so many miles at so much per mile, paid for by other volunteers. This at least encourages a healthier lifestyle.)

Naturally we continued to discuss mathematics, and wrote a couple of joint papers [3, 4]. Paul’s method for writing joint papers was, of course, for him to convey just the essence of his share of the argument; it was up to the co-author to write the actual paper. (I believe Alexandre Dumas père used a somewhat similar system.) Unlike many mathematicians, he then showed little interest in music; but there was one record (a “78” of course) I had that he was very fond of—Myra Hess playing her piano arrangement “Jesu Joy of Man’s Desiring” from Bach’s cantata. It is a beautiful piece, and he would often ask to hear it.

My later encounters with Paul Erdős have been somewhat hit-or-miss; for example, last Spring my wife and I left our home near Boston to attend a conference in Columbus at which he was expected to speak, but instead he gave a talk near Boston precisely during our absence. As someone said, the Heisenberg uncertainty principle applies to him; you cannot determine simultaneously his position and his velocity. We have sometimes successfully attended the same conference—for example a combinatorial one in Calgary, Canada (1969), resulting in [5], and one on “Real Analysis” in Smolenice (1991). And we have corresponded, the usual Erdős letter beginning “Dear Stone, Let $p_1,\ p_2,\ldots,p_n$ be distinct primes” (or points or whatnot), and listing some interesting but (to me) hopelessly difficult conjectures.

In addition to the foregoing traits, here are two more that even a casual acquaintance with Paul would reveal. First, he delighted in small children and infants. (They were the “superbosses” , in his terminology, outranking older females who were merely “bosses”.) Second, it may be observed that I have been a bit vague about some dates. But he would remember them exactly. He remembers events of interest to his friends, too; as someone said, he remembers the incidence matrix.

2. Some (Very) Elementary Number Theory

Theorem 1. *The two Diophantine equations*

$$x^2 = \lambda m^y \pm 1$$

have at most two nontrivial solutions between them.

Here $\lambda \geq 1$ and $m \geq 2$ are given integers, and for "nontriviality" we require the unknown integers x and y to be greater than 1.

It is convenient to refer to the equations as $(*,+)$ and $(*,-)$ respectively, and to abbreviate "nontrivial solution" by "solution". Thus the assertion is that $(*,\pm)$ has at most 2 solutions.

Proof. Write $(*,\pm)$ as the pair of Pell equations $x^2 = \mu z^2 \pm 1$ where, if y is odd, $\mu = \lambda m$ and $z = m^{(y-1)/2}$, and if y is even, $\mu = \lambda m^2$ and $z = m^{(y-2)/2}$. In either case we have $m \mid \mu$ (so $\mu > 1$) and $z \geq 1$. We can assume that μ is not a square, else there are no (nontrivial) solutions.

If the equation $x^2 = \mu z^2 - 1$ has solutions, let (x_1, z_1) be the smallest positive one. If not, let (x_1, z_1) be the smallest positive solution of $x^2 = \mu z^2 + 1$ (which certainly exists). In either case, note that $(x_1, m) = 1$. It is well known (see e.g. [10]) that all (nontrivial) solutions of $x^2 = \mu z^2 \pm 1$ are of the form (x_r, z_r) $(r = 1, 2, \ldots)$ where

$$x_r + \mu^{1/2} z_r = (x_1 + \mu^{1/2} z_1)^r, \quad x_r - \mu^{1/2} z_r = (x_1 - \mu^{1/2} z_1)^r.$$

Thus z_r can be expressed as the finite sum

$$z_r = \binom{r}{1} x_1^{r-1} z_1 + \binom{r}{3} x_1^{r-3} z_1^3 \mu + \binom{r}{5} x_1^{r-5} z_1^5 \mu^2 + \cdots \tag{1}$$

Lemma 1. *If rz_1 is a multiple of m^t for some nonnegative integer t, and if $(m,3) = 1$, then (for $s = 1, 2, \ldots, \left[\frac{r-1}{2}\right]$) $\binom{r}{2s+1} z_1 m^s$ is a multiple of m^{t+1}.*

Let p be a prime factor of m, say with multiplicity $\alpha \geq 1$, and first suppose $p \geq 5$. In the expression

$$\binom{r}{2s+1} z_1 m^s = \frac{r(r-1)\cdots(r-2s)}{1 \cdot 2 \cdots (2s+1)} z_1 m^s,$$

count the number of occurrences of p as a factor. In the numerator, the factors $rz_1 m^s$ provide $\alpha(t+s)$ occurrences of p. In the denominator, the number is

$$\left[\frac{2s+1}{p}\right] + \left[\frac{2s+1}{p^2}\right] + \cdots < (2s+1)(p^{-1} + p^{-2} + \cdots \text{to } \infty) = \frac{2s+1}{p-1}$$

which is less than $\frac{1}{4}(2s+1) < s$. So $\binom{r}{2s+1} z_1 m^s$ is a multiple of p^β where, $\beta = \alpha(t+s) - (s-1) \geq \alpha(t+1)$.

In the remaining case $p = 2$, the same argument shows that the number of occurrences of 2 as factor in the denominator is at most $2s$. In the numerator, besides $\alpha(t+s)$ occurrences from $rz_1 m^s$, s of the factors $r-1, r-2, \ldots, r-2s$ are even and (assuming for the moment that $s \geq 2$) at least one of them is a multiple of 4. So $\binom{r}{2s+1} z_1 m^s$ (where $s \geq 2$) is a multiple of 2^γ where $\gamma = \alpha(t+s) + (s+1) - 2s = \beta \geq \alpha(t+1)$ as before.

But if $s = 1$, $\binom{r}{2s+1} z_1 m^s = \frac{r(r-1)(r-2)}{1 \cdot 2 \cdot 3} z_1 m$, a multiple of 2^δ where (because $(r-1)(r-2)$ is even) $\delta \geq \alpha(t+1)$; and Lemma 1 follows.

Remark 1. *The lemma would fail if $p = 3$ were allowed; for instance (with $r = 6$, $z_1 = 5$, $m = 3$, $s = 1$) $6 \cdot 5$ is a multiple of 3 but $\binom{6}{3} \cdot 5 \cdot 3$ is not a multiple of* 9.

Lemma 2. *If $x \geq 2$ and $z \geq 1$ are integers such that $x^2 = \mu z^2 \pm 1$, and if z is a power of m and $(m, 3) = 1$ and $m \mid \mu$, then (given μ) x and z are uniquely determined.*

We know $x = x_r$ and $z = z_r$ for some $r \geq 1$. First suppose $x_1 > 1$; we show that $r = 1$. Observe that r is necessarily odd, for otherwise (1) shows that $x_1 \mid z_r$; but $(x_1, m) = 1$, contradicting that z is a power of m. So we can assume, for a contradiction, that $r \geq 3$. There are now at least 2 terms on the right of (1). Let m^t be the highest power of m dividing the first of these, namely $\binom{r}{1} x_1^{r-1} z_1$. By Lemma 1 the other terms are multiples of m^{t+1}. Thus z_r is greater than m^t but not divisible by m^{t+1}, again contradicting that z is a power of m.

Finally, if $x_1 = 1$, we must have $\mu = 2$ and the trivial "solution" (x_1, z_1) has been excluded here because $x \geq 2$. The same argument as before shows that $r < 3$, so necessarily $r = 2$.

To deal with the case in which $3 \mid m$, we first show that Lemma 1 continues to hold provided $9 \mid m$ and $s > 1$. We state this as

Lemma 3. *If rz_1 is a multiple of m^t for some nonnegative integer t, and if $9 \mid m$, then (for $s = 2, 3, \ldots, \left[\frac{r-1}{2}\right]$), $\binom{r}{2s+1} z_1 m^s$ is a multiple of m^{t+1}.*

In view of Lemma 1 it suffices to check the occurrences of the prime 3 in $\frac{r(r-1)\ldots(r-2s)}{1\cdot 2\cdot \ldots (2s+1)} z_1 m^s$ where $s \geq 2$. In the numerator there are $\alpha(t+s)$ from $rz_1 m^s$, as before, together with at least one from $(r-1)\ldots(r-2s)$. In the denominator, as before, there are fewer than $(2s+1)/2$, hence at most s. Altogether $\binom{r}{2s+1} z_1 m^s$ is a multiple of 3^θ where $\theta = \alpha(t+s) + 1 - s \geq \alpha(t+1)$, as before, and Lemma 3 follows.

A modified form of Lemma 2 now holds:

Lemma 4. *If $x \geq 2$ and $z \geq 2$ are integers such that $x^2 = \mu z^2 \pm 1$, and if z is a power of m, and $3 \mid m$ and $3m \mid \mu$, then (given μ) x and z are uniquely determined.*

As in Lemma 2 we have that $x = x_r$ and $z = z_r$ for some positive integer r. We show that r must be just one of 1 and 3.

Define t as in Lemma 2, and first suppose $r \geq 5$. There are then at least 3 terms on the right of (1). By Lemma 3, all but the first 2 of these are multiples of m^{t+1}. The sum of the first two can be written

$$r x_1^{r-3} z_1 u \quad \text{where} \quad u = \frac{1}{3}\mu z_1^2 \left(\binom{r-1}{2} + 3 \right) \pm 1.$$

Now we observe that $(u, m) = 1$. In fact, if p is a prime factor of m other than 3 we clearly have $p \mid (u \mp 1)$, so p is not a factor of u. And if $p = 3$ this still holds because $9 \mid \mu$. Since $(x_1, m) = 1$ also, we see that the sum of the first two terms in (1) is a multiple of m^t but not of m^{t+1}, contradicting that z is a power of m.

As in Lemma 2, r cannot be even here, so r can only be 1 or 3. We show these alternatives are mutually exclusive.

Suppose not. Then both $z_3 = 3x_1^2 z_1 + z_1^3\mu$, and z_1, are powers of m; so their quotient, $3x_1^2 + z_1^2\mu$, is also a power of m, and in particular a multiple of m. Hence $m \mid 3x_1^2$; but $(m, x_1) = 1$, so $m \mid 3$, forcing $m = 3$. We also have $x_1^2 = \mu z_1^2 + 1$ (the alternative -1 being excluded here because it is not a quadratic residue $\mod 3$), so $3x_1^2 + z_1^2\mu = 4\mu z_1^2 + 3$ is a power of $m = 3$, and so also is $(4\mu z_1^2 + 3)/3 = 12(\mu/9)z_1^2 + 1$. This is impossible, as it is greater than 1 and congruent to $1 \mod 3$.

The theorem now follows readily from Lemma 2 when m is not a multiple of 3, and from Lemma 4 when $3 \mid m$ by the following modification of the initial transformation to Pell equations. Put

$$\mu = \lambda m^3 \quad \text{if } y \text{ is odd}, \quad \lambda m^2 \quad \text{if } y \text{ is even}$$

and

$$z = m^{(y-3)/2} \text{ or } m^{(y-2)/2} \text{ respectively.}$$

Then the equations $(*, \pm)$ reduce to $x^2 = \mu m^2 + 1$ (the $-$ sign being impossible now because $3 \mid \mu$). In each case, Lemmas 2 and 4 show that, if (nontrivial) solutions for x and z exist, they are unique, given μ, As there are two possible values for μ in each case, there are at most two (nontrivial) solutions altogether for $(*, \pm)$. □

Remarks

(i) For certain values of λ and m there actually are two nontrivial solutions to $(*, \pm)$; for instance, $5^2 = 3 \cdot 2^3 + 1$, $7^2 = 3 \cdot 2^4 + 1$; and $11^2 = 15 \cdot 2^3 + 1$, $31^2 = 15 \cdot 2^6 + 1$. There can also be additional trivial "solutions", for instance $2^2 = 3 \cdot 2^0 + 1$ and $4^2 = 15 \cdot 2^0 + 1$.

(ii) From the way in which the two possible values of μ arise, if there are two solutions to $(*, \pm)$, one will have an odd value of y and the other an even one.

(iii) If λ is a square, the equations $(*, \pm)$ can have at most *one* solution between them, for y must be odd and there is only one value of μ to consider.

(iv) The equation $(*, -)$ can have solutions only if -1 is a quadratic residue $\mod \lambda m$. But I do not know whether $(*, -)$ can then have more than one (nontrivial) solution, nor whether the two equations $(*, \pm)$ can have one solution each.

(v) The equations $x^2 = 10^y \pm 1$ have no solutions other than the trivial $3^2 = 10^1 - 1$. For, as in (iii), we have only one value of μ to consider;

namely $\mu = 10$; and then $x_1 = 3$, $z_1 = 1$. (Thus the numbers $100\ldots001$, $11\ldots111$ in the usual scale of ten, and of more than one digit, are never squares.) Of course, much more than this is known; see [9].

(vi) *Background* A problem in the American Mathematical Monthly (E 2511, January 1975, p. 73), due to M. Olitsky, reduces to solving the Diophantine equations $x^2 + 1 = 5^y$, $x^2 + 1 = 2 \cdot 5^y$. The solution (ibid., April 1976, p. 291) mentioned references [7–9] below. And the equations suggested the present generalization to $x^2 = \lambda m^y \pm 1$.

References

1. R. L. Brooks, C. A. B. Smith, A. H. Stone and W. T. Tutte, The dissection of rectangles into squares, Duke Math. J. **7** (1940) 312–340.
2. C. J. Bouwkamp and A. J. W. Duijvestijn, Catalogue of Simple Perfect Squared Squares of orders 21 through 25, Eindhoven University of Technology 1992.
3. P. Erdős and A. H. Stone, Some remarks on almost periodic transformations, Bull. Amer. Math. Soc. **51** (1945) 126–130.
4. P. Erdős and A. H. Stone, On the structure of linear graphs, Bull. Amer. Math. Soc. **52** (1946) 1087–1091.
5. P. Erdős and A. H. Stone, On the sum of two Borel sets, Proc. Amer. Math. Soc. **25** (1970) 304–306.
6. Jasper Dale Skinner **II**, Squared Squares: Who's Who and What's What, Lincoln, Nebraska, 1993.
7. C. Engelman, On close-packed double-error-correcting codes on p symbols, 1. R. E. Transactions on Information Theory, Correspondence, January 1961, 51–52.
8. V. A. Lebesgue, Sur l'impossibilité en nombres entiers de l'équation $x^m = y^2 + 1$, Nouv. Ann. Math. **9** (1850) , 178–181.
9. L. J. Mordell, Diophantine Equations, Academic Press 1969, esp. p. 301.
10. I. Niven and H. S. Zuckerman, Introduction to the Theory of Numbers, Wiley, New York 1960.

On Cubic Graphs of Girth at Least Five

William T. Tutte

W.T. Tutte (Deceased)
Department of Combinatorics and Optimization, University of Waterloo,
Waterloo, ON N2L 3G1, Canada

It is an honor to be asked to contribute a paper to so historically important a collection. Yet it can be embarrassing too. In my case I ask distractedly "What can I write about? The researches I have completed have been published already, or at least have been submitted to Journals. The work I am engaged upon is incomplete, may be anticipated, perhaps even fallacious. And what else can there be?"

But perhaps I can say something about what I am doing now, mathematically, though my discourse may be more speculative than demonstrative. Something about how it came to interest me and what I hope to get from it.

It seems that someone once conjectured that every cubic graph without a Tait coloring contains a subdivided Petersen graph (or is a Petersen graph itself). Some time last summer (1994), this conjecture having come up several times in conversation, I reacted against it. "So what?" I asked, "Probably every sufficiently nonplanar cubic graph has that property!"

On later reflection I realized that in this hasty dismissal I had another conjecture, and even one that a graph-theorist might reasonably hope to prove.

Over the next few months that conjecture, having persisted in my thoughts, changed its form somewhat and acquired some qualifications. I recalled that the Petersen graph was famous as the smallest cubic graph of girth 5. I also knew of cubic graphs of girth less than 5 that obviously did not contain a Petersen graph. So I began to think of the new conjecture as being about cubic graphs of girth 5 or more.

After these months of incubation the conjecture reformulated itself as the problem of classifying the "critical" cubic graphs. A critical cubic graph was to be defined as one of girth ≥ 5 that contained (in subdivision) no other cubic graph of girth ≥ 5. Obviously any finite cubic graph of so high a girth would either be critical itself or would contain a subdivided critical cubic graph. One hoped there would not be too many critical cubic graphs. Perhaps only the Petersen graph and the graph of the regular dodecahedron?

There was an elementary theorem about critical cubic graphs that seemed relevant to the problem of classification. Suppose we have a cubic graph G of girth at least 5 and delete an edge A, suppressing its now-divalent ends to get a small cubic graph H. Then H also has girth at least 5 unless A "impinges" on a pentagon P of G. "Impinges" means that A is not an edge of P but does

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_7,

have one end in P. Since G contains a subdivision of H we infer that if G is critical each of its edges must impinge upon a pentagon of G.

From this theorem there developed a theory of "semicubic" subgraphs of a critical cubic graph G. (A graph is semicubic if each of its vertices has valency either 2 or 3). Let us define a "free" edge of a semicubic graph J of girth 5 as one that impinges on no pentagon of J. Let us say that such a graph J is "open" if it has a free edge and "closed" otherwise. Given such a J with a free edge A it was found possible to make a list of all the semicubic graphs of girth 5 that could be obtained from J by uniting it with a pentagon that was impinged upon by A. Then one had a theorem saying that if a critical graph G contains J then it must contain also one of those more complicated derivatives of J. Each of these derivatives, if open, had to be treated in the same way, and so on. The end result was a list of closed semicubic graphs such that each critical graph G must contain as a subgraph some member of the list.

Getting this theorem was a long and tedious process. Getting from it to the desired classification was a long and arduous task. "O Murphy" I groaned, "Wherefore doth thy Law throw me always the hard ones?"

In the course of the investigation I encountered three more critical cubic graphs that I thought interesting. One consisted of a nonagon and three outside vertices each joined to the nonagon by three edges. Another could be derived from a cube by the following construction. Two opposite sides of one face are bisected and the points of bisection are joined by a new segment. The opposite face of the cube is treated likewise, but so that the two new segments are perpendicular. Then the two new segments are bisected and their mid-points are joined by a new edge.

Another critical cubic graph can be constructed from a rectangle with two long sides and two short ones. Inside it, as a frame we draw two pentagons having one edge in common. Their vertices opposite the common edge are joined to the mid-points of the short sides of the rectangle. Their other vertices are joined to the points of trisection of the long sides. All these joins are such as to preserve planarity. Finally each pair of opposite corners of the rectangle are joined by a new edge, to give a critical cubic graph of crossing number 1.

After adding these three graphs to the Petersen graph and the dodecahedral graph I now had a list of five "interesting" cubic graphs. I called them the Z-graphs, and noted that they were all cyclically 5-connected.

There was a gaggle of other critical cubic graphs that I thought less interesting. Each of these had an isthmus or a 2-bond, or a 3-bond that separated circuits. Perhaps one could, in a way, eliminate these by showing that a cubic graph K of girth at least 5 and of sufficiently high cyclic connectivity must contain a subdivision of a Z-graph? I an now satisfied that there is such a theorem. It requires only that K be cyclically 4-connected. More hard work, and with many false starts!

Work on the original conjecture cannot be regarded as complete until we have an extensive and simply defined class of cubic graphs that must contain (in subdivision) a Petersen graph and not merely "some Z-graph"). Perhaps there is a hint of such a theorem in the fact that the Petersen graph is, in a way, more nonplanar than the other Z-graphs. The dodecahedral graph is planar. The one derived from a rectangle has crossing number 1. The one from a cube can be made planar by deleting one edge, and the nonagonal one by identifying two vertices. But the Petersen graph has crossing number 2 and it cannot be made planar by either of those two tricks. If sufficiently high cyclic connectivity in K ensures the appearance (in subdivision) of a cyclically 5-connected Z-graph, then perhaps sufficiently complex nonplanarity will ensure the appearance of a Petersen graph, that most nonplanar of the Z-graphs?

Meanwhile I am happy to present some of the facts in a book dedicated to the Master.

II. Number Theory

Introduction

It is difficult to estimate the relative impact of Erdős' research in different areas of mathematics. But it is a fact that Erdős started with number theory (e.g., out of his first 60 papers only 2 are not related to number theory) and that among his publications, the number theory papers have highest frequency. His achievements are well known and are amply mirrored by contributions to this chapter (which is the largest of all the chapters of these volumes).

The papers by Ahlswede and Khachatrian, Konyagin and Pomerance, Nathanson, Nicolas, Schinzel, Shorey and Tijdeman, Sárközy and Sós, and Tenenbaum survey and relate to various parts of Erdős' research, and they complement in various respects his own recollections given in an earlier chapter.

Some of these papers are research articles, such as the papers by Ahlswede and Cai, Sárközy, Tenenbaum and Bergelson et al. (which includes Erdős himself as a coauthor).

Although we believe this is a representative sample of Erdős' activities in this area, many problems and particular research directions are not covered. The reader should bear in mind that Erdős himself considered the probabilistic methods in number theory together with his work on prime numbers as his main contributions to number theory. Probabilistic methods are covered by the next section as well.

In 1995/1996, when the content of these volumes was already crystallizing, we asked Paul Erdős to isolate a few problems, both recent and old, for each of the eight main parts of this book. To this part on infinity he contributed the following collection of problems and comments.

Erdős in his own words

Here is a purely computational problem (this problem cannot be attacked by other means at present). Call a prime p *good* if every even number $2r \leq p-3$ can be written in the form $q_1 - q_2$ where $q_1 \leq p$, $q_2 \leq p$ are primes. Are there infinitely many good primes? The first bad prime is 97 I think. Selfridge and Blecksmith have tables of the good primes up to 10^{37} at least, and they are surprisingly numerous.

I proved long ago that every $m < n!$ is the distinct sum of $n-1$ or fewer divisors of $n!$. Let $h(m)$ be the smallest integer, if it exists, for which every integer less than m is the distinct sum of $h(m)$ or fewer divisors of m. Srinivasan called the numbers for which $h(m)$ exists practical. It is well known and easy to see that almost all numbers m are not practical. I conjectured that there is a constant $c \geq 1$ for which for infinitely many m we have $h(m) < (\log\log m)^c$. M. Vose proved that $h(n!) < cn^{1/2}$. Perhaps $h(n!) < c(\log n)^{C_2}$. I would be very glad to see a proof of $h(n!) < n^\epsilon$.

A practical number n is called a *champion* if for every $m > n$, we have $h(m) > h(n)$. For instance, 6 and 24 are champions, as $h(6) = 2$, the next practical number is 24, $h(24) = 3$, and for every $m > 24$, we have $h(m) > 3$. It would be of some interest to prove some results about champions. A table of the champions $< 10^6$ would be of some interest. I conjecture that $n!$ is not a champion for $n > n_0$.

The study of champions of various kinds was started by Ramanujan (Highly composite numbers, *Collected Papers of Ramanujan*). See further my paper with Alaoglu on highly composite and similar numbers and many papers of J. L. Nicolas and my joint papers with Nicolas.

The following related problem is perhaps of some mild interest, in particular, for those who are interested in numerical computations. Denote by $g_r(n)$ the smallest integer which is not the distinct sum of r or fewer divisors of n. A number n is an r-champion if for every $t < n$ we have $g_r(n) > g_r(t)$. For $r = 1$ the least common multiple M_m of the integers $\leq m$ is a champion for any m, and these are all the 1-champions. Perhaps the M_m are r-champions too, but there are other r-champions; e.g., 18 is a 2-champion.

Let $f_k(n)$ be the largest integer for which you can give $f_k(n)$ integers $a_i \leq n$, for which you can not find $k+1$ of those that are relatively prime. I conjectured that you get $f_k(n)$ by taking the multiple $\leq n$ of the first k primes. This was proved for small k by Ahlswede, and Khachatrian disproved it for $k \geq 212$. Perhaps if $n \geq (1+\epsilon)p_k^2$, where p_k is kth prime, the conjecture remains true.

Let $n_1 < n_2 < \ldots$ be an arbitrary sequence of integers. Besicovitch proved more than 60 years ago that the set of the multiples of the n_i does not have to have a density. In those prehistoric days this was a great surprise. Davenport and I proved that the set of multiples of the $\{n_i\}$ have a logarithmic density and the logarithmic density equals the lower density of the set of multiples of the $\{n_i\}$. Now the following question is perhaps of interest: Exclude one or several residues mod n_i (where only the integers $\geq n_i$ are excluded).

Is it true that the logarithmic density of the integers which are not excluded always exists? This question seems difficult even if we only exclude one residue mod n_i for every n_i.

For a more detailed explanation of these problems see the excellent books by Halberstam and Roth, *Sequences*, Springer-Verlag, and by Hall and Tenenbaum, *Divisors*, Cambridge University Press.

Tenenbaum and I recently asked the following question: let $n_1 < n_2 < \dots$ be an infinite sequence of positive integers. Is it then true that there always is a positive integer k for which almost all integers have a divisor of the form $n_i + k$? In other words, the set of multiples of the $n_i + k$ $(1 \le i < \infty)$ has density 1. Very recently Ruzsa found a very ingenious counterexample.

Tenenbaum thought that perhaps for every $\epsilon > 0$ there is a k for which the density of the multiples of the $n_i + k$ has density $> 1 - \epsilon$.

In a paper (Proc. London Math. Soc. (1970) dedicated to the memory of Littlewood) Sárközy and I state the following problem: Let $1 \le a_i < a_2 < \dots < a_{n+2} \le 3n$ be $n+2$ integers. Prove that there always are three of them $a_i < a_j < a_k$ for which $a_j + a_k \equiv 0 \pmod{a_i}$. The integers $2n \le t \le 3n$ show that $n+1$ integers do not suffice.

Perhaps a proof or disproof will be easy. As far as I know, the problem has been rather forgotten.

Many more problems are contained in the book P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, the second edition of which should appear soon.

*****

So much for Paul Erdős. The progress on his problems since 1995/1996 has been considerable and many of his results and problems became the subject of intensive study. For example research on sum sets (reported in this book in the Nathanson article) led to the celebrated Freiman–Ruzsa theorem, which in turn has been instrumental in the Green–Tao proof of arithmetic progressions in primes (see the Ramsey theory chapter in Vol II). For sum sets, see the comprehensive book:

I. Z. Ruzsa, Sumsets and Structure. In: Combinatorial Number Theory and Additive Group Theory (A. Geroldiner, I. Z. Ruzsa, eds.) Advanced Courses in Mathematics CRM, Birkhäuser, 2009, 87–210.

In another, yet broadly related, direction, the celebrated sum–product theorem of Erdős and Szemerédi became the basis for many similar results dealing with the sum–product phenomenon in both the finite and infinite setting. This in turn has found many applications.

P. Erdős, E. Szemerédi, On sums and products of integers, Studies in Pure Mathematics, Birkhauser, Basel, 1983, 213–218.

T. Tao, The sum-product phenomenon in arbitrary rings, Contributions to Discrete Math. 4,2 (2009), 59–82.

J. Solymosi, Incidence and Spectra of Graphs. In: Combinatorial Number Theory and Additive Group Theory (A. Geroldiner, I. Z. Ruzsa, eds.) Advanced Courses in Mathematics CRM, Birkhäuser, 2009, 299–314.

It is only fitting to remark that these questions are close to those considered by L. Guth article in Part IV devoted to geometry. The progress in combinatorial number theory in the last decade was spectacular. Perhaps P. Erdős didn't expect that. Here is an evidence of this: He mentions in the introduction of his article in this volume (Some of my favourite problems and results) an anecdote about Hilbert's wrong estimation on what is a difficult problem. Well the same happened to him: In the above article he states that

"An old conjecture in number theory states that for every k there are k primes in an arithmetic progression. This problem seems unattackable".

Well as we know the opposite is exactly what happened. But the Erdős related conjecture is still open and (quoting him again) "neither Szemerédi nor Furstenberg" (nor Gowers nor Green and Tao) "are able to settle this."

For the general development of combinatorial number theory, see recently published book:

T. Tao, V. Vu, Additive Combinatorics, Cambridge University Press, 2006.

Note also that since 2000 there has been published the electronic journal of combinatorial number theory, *Integers*, which has P. Erdős as part of its logo (www.integers-ejcnt.org).

Cross-Disjoint Pairs of Clouds in the Interval Lattice

Rudolf Ahlswede and Ning Cai

R. Ahlswede (Deceased)
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131,
33501 Bielefeld, Germany

N. Cai (✉)
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131,
33501 Bielefeld, Germany

The State Key Laboratory of ISN, Xidian University, Xi'an,
Shaanxi 710071, China
e-mail: caining@mail.xidian.edu.cn

Summary Let $\mathcal{I}_n$ be the lattice of intervals in the Boolean lattice $\mathcal{L}_n$. For $\mathcal{A},\mathcal{B}\subset\mathcal{I}_n$ the pair of clouds $(\mathcal{A},\mathcal{B})$ is cross-disjoint, if $I\cap J=\emptyset$ for $I\in\mathcal{A},\ J\in\mathcal{B}$. We prove that for such pairs $|\mathcal{A}||\mathcal{B}|\leq 3^{2n-2}$ and that this bound is best possible.

Optimal pairs are up to obvious isomorphisms unique. The proof is based on a new bound on cross intersecting families in $\mathcal{L}_n$ with a weight distribution. It implies also an Intersection Theorem for multisets of Erdős P, Schőnheim J (1969) On the set of non pairwise coprime division of a number. In: Proc. of the Colloquium on Comb. Math. Dalaton Füred, pp 369–376.

1. The Results

Consider the set $[n]=\{1,2,\ldots,n\}$, the set of all its subsets $\mathcal{L}_n$, and the lattice of intervals $\mathcal{I}_n=\{I=[A,B]:A,B\in\mathcal{L}_n\}$, where $[A,B]=\{C\in\mathcal{L}_n: A\subset C\subset B\}$, if $A\subset B$, and $[A,B]=I_\emptyset$ (the empty interval), if $A\not\subset B$. The lattice operations $\wedge$ and $\vee$ are defined by

$$[A,B]\wedge[A',B']=[A,B]\cap[A',B'], \tag{1}$$

$$[A,B]\vee[A',B']=[A\cap A',B\cup B']. \tag{2}$$

Here the empty interval $I_\emptyset$, is represented by $[[n],\emptyset]$. The pair $(\mathcal{A},\mathcal{B})$ with $\mathcal{A},\mathcal{B}\subset\mathcal{I}_n\setminus\{I_\emptyset\}$ is cross-disjoint, if

$$I\wedge J=I_\emptyset \text{ for } I\in\mathcal{A},\ J\in\mathcal{B}. \tag{3}$$

Let us denote the set of those pairs by $\mathcal{D}_n$.

Theorem 1. *For $n=1,2,\ldots$*

$$\max\{|\mathcal{A}||\mathcal{B}|:(\mathcal{A},\mathcal{B})\in\mathcal{D}_n\}=3^{2n-2}.$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_8,

Equality is assumed for

$$\mathcal{A}^* = \{I \in \mathcal{I}_n : I = [A,B], 1 \notin B\}, \quad \mathcal{B}^* = \{I \in \mathcal{I}_n : I = [A,B], 1 \in A\}.$$

All optimal pairs are obtained by replacing 1 *in the definition of* $\mathcal{A}^*$ *and* $\mathcal{B}^*$ *by any element* m *of* $[n]$*, and by exchanging the roles of these two sets.*

We shall relate *cross-disjoint* pairs of clouds from $\mathcal{I}_n$ to *cross-intersecting* pairs of clouds from $\mathcal{L}_n$ with a suitable weight.

Recall from [1] that $(\mathcal{U}, \mathcal{V})$ with $\mathcal{U}, \mathcal{V} \subset \mathcal{L}_n$ is cross-intersecting, if

$$U \cap V \neq \emptyset \text{ for } U \in \mathcal{U} \text{ and } V \in \mathcal{V}. \tag{4}$$

We denote the set of these pairs by $\mathcal{P}_n$. Furthermore we introduce the weight $w : \mathcal{L}_n \to \mathbb{N}$ by

$$w(A) = 2^{n-|A|} \text{ for } A \in \mathcal{L}_n. \tag{5}$$

Theorem 2. *For* $(\mathcal{U}, \mathcal{V}) \in \mathcal{P}_n$

$$W(\mathcal{U})W(\mathcal{V}) \triangleq \sum_{U \in \mathcal{U}} w(U) \cdot \sum_{V \in \mathcal{V}} w(V) \leq 3^{2(n-1)}$$

and the bound is best possible. Moreover, for any optimal pair $(\mathcal{U}, \mathcal{V})$ *there exists a* $t \in [n]$ *such that* $\mathcal{U} = \mathcal{V} = \{A \in \mathcal{L}_n : t \in A\}$.

2. Another Description for $\mathcal{I}_n \setminus \{I_\emptyset\}$

We associate $[A,B] \in \mathcal{I}_n \setminus \{I_\emptyset\}$ with a ternary sequence $\Psi([A,B]) = (x_1, x_2, \dots, x_n)$, where

$$x_t = \begin{cases} 0 & \text{if } t \notin B \\ 1 & \text{if } t \in A \\ 2 & \text{if } t \in B \setminus A. \end{cases}$$

$\Psi : \mathcal{I}_n \setminus \{I_\emptyset\} \to \{0,1,2\}^n$ is bijective.

If $\Psi([A,B]) = x^n$ and $\Psi([A',B']) = y^n$, then

$$[A,B] \wedge [A',B'] = I_\emptyset \Leftrightarrow \exists t \in [n] : \{x_t, y_t\} = \{0,1\}. \tag{6}$$

For $(\mathcal{A}, \mathcal{B}) \in \mathcal{D}_n$ the associated pair $(\mathcal{X}, \mathcal{Y}) = (\Psi(\mathcal{A}), \Psi(\mathcal{B}))$ has the property:

$$\text{For } x^n \in \mathcal{X}, y^n \in \mathcal{Y} \quad \{x_t, y_t\} = \{0,1\} \text{ for some } t \in [n]. \tag{7}$$

We can view $(\mathcal{X}, \mathcal{Y})$ as families of cross-disjoint subcubes of the n-dimensional unit cube or as families of cross-disjoint cylinder sets in $\{0,1\}^n$ in the sense of measure or probability theory. In this interpretation 2 stands for the set $\{0,1\}$.

Henceforth we consider pairs $(\mathcal{X},\mathcal{Y})$; $\mathcal{X},\mathcal{Y} \subset \{0,1,2\}^n$; satisfying (7). We call them cross-disjoint. The set of these pairs is denoted by $\mathcal{D}_n^*$. Our first goal in proving Theorem 1 is to show that for $(\mathcal{X},\mathcal{Y}) \in \mathcal{D}_n^*$

$$|\mathcal{X}||\mathcal{Y}| \leq 3^{2(n-1)}. \tag{8}$$

3. Down-Up-Shifts

The proof of Theorem 1 goes in several steps. At first we show here that any $(\mathcal{X},\mathcal{Y}) \in \mathcal{D}_n^*$ can be transformed into another pair in $\mathcal{D}_n^*$ with the same cardinalities and with invariance under down-up-shifts. They are defined as follows.

For any $\mathcal{Z} \subset \{0,1,2,\}^n$ and any $t \in [n]$ set

$$\begin{aligned} d_t(\mathcal{Z}) = \big\{ &(z_1,\dots,z_{t-1},i,z_{t+1},\dots,z_n) : i = 0,\dots,j-1; j \geq 1 \quad \text{and} \\ &|\{z : (z_1,\dots,z_{t-1},z,z_{t+1},\dots,z_n) \in \mathcal{Z}\}| = j \big\}. \end{aligned} \tag{9}$$

This is the down-shift of $\mathcal{Z}$ in the t-th component. Similarly, $u_t(\mathcal{Z})$, the up-shift of $\mathcal{Z}$ in the t-th component is obtained by exchanging 0 and 1 in the t-th component of the sequences in $d_t(\mathcal{Z})$. We formulate an immediate consequence of our definitions.

Lemma 1. *For any $(\mathcal{X},\mathcal{Y}) \in \mathcal{D}_n^*$ and $t \in [n]$, then also $(d_t(\mathcal{X}), u_t(\mathcal{Y})) \in \mathcal{D}_n^*$.*

We say that $(\mathcal{X},\mathcal{Y})$ with $\mathcal{X},\mathcal{Y} \subset \{0,1,2\}^n$ is down-up-extremal, if

$$(d_t(\mathcal{X}), u_t(\mathcal{Y})) = (\mathcal{X},\mathcal{Y}) \text{ for all } t \in [n]. \tag{10}$$

4. Relation to Cross-Intersection in $\mathcal{L}_n$

Next we introduce the mappings $\sigma_i : \{0,1,2\}^n \to \mathcal{L}_n$ by

$$\sigma_i(z^n) = \{t : z_t = i, 1 \leq t \leq n\} \text{ for } i = 0,1. \tag{11}$$

We also put for $\mathcal{Z} \subset \{0,1,2\}^n$

$$\sigma_i(\mathcal{Z}) = \{\sigma_i(z^n) : z^n \in \mathcal{Z}\}. \tag{12}$$

These mappings make it possible to convert cross-disjoint pairs of clouds from the interval lattice $\mathcal{I}_n$ into cross-intersecting pairs of clouds from the Boolean lattice $\mathcal{L}_n$.

More precisely we have the following result.

Lemma 2. *Suppose that $(\mathcal{X},\mathcal{Y})$ with $\mathcal{X},\mathcal{Y} \subset \{0,1,2\}^n$ is down-up-extremal. Then $(\mathcal{X},\mathcal{Y})$ is cross-disjoint exactly if $(\sigma_0(\mathcal{X}),\sigma_1(\mathcal{Y}))$ is cross-intersecting, that is, $X \cap Y \neq \emptyset$ for $X \in \sigma_0(\mathcal{X})$ and $Y \in \sigma_1(\mathcal{Y})$.*

Proof. Suppose that $(\mathcal{X}, \mathcal{Y})$ is cross-disjoint, but that $\big(\sigma_0(\mathcal{X}), \sigma_1(\mathcal{Y})\big)$ is not cross-intersecting. Then there exist $x^n \in \mathcal{X}, y^n \in \mathcal{Y}$, and a non-empty set $E \subset [n]$ such that $(x_t, y_t) = (1, 0)$ for $t \in E$ and $\{x_t, y_t\} \neq \{1, 0\}$ for $t \notin E$. However, since $(\mathcal{X}, \mathcal{Y})$ is down-up-extremal, the sequence x'^n obtained from x^n by replacing for $t \in E$ with $x_t = 1$ by $x'_t = 0$ must be in $\mathcal{X}$ and this sequence is not disjoint with y^n. This contradiction proves that $(\sigma_0(\mathcal{X}), \sigma_1(\mathcal{Y}))$ is cross-intersecting. The reverse implication is obvious. □

5. Theorem 1 from Theorem 2

Notice that for $A \subset [n]$

$$|\sigma_0^{-1}(A)| = |\sigma_1^{-1}(A)| = 2^{n-|A|}. \tag{13}$$

Therefore also for any $A, B \subset [n]$ and $\mathcal{X}, \mathcal{Y} \subset \{0,1,2\}^n$

$$|\sigma_0^{-1}(A) \cap \mathcal{X}| \leq 2^{n-|A|}, |\sigma_1^{-1}(B) \cap \mathcal{Y}| \leq 2^{n-|B|}. \tag{14}$$

Now in upper bounding $|\mathcal{X}||\mathcal{Y}|$ for $(\mathcal{X}, \mathcal{Y}) \in \mathcal{D}_n^*$ we can assume by Lemma 1 that $(\mathcal{X}, \mathcal{Y})$ is down-up-extremal and by Lemma 2 that $(\sigma_0(\mathcal{X}), \sigma_1(\mathcal{Y})) \in \mathcal{P}_n$. Hence Theorem 2 implies that for $\mathcal{U} = \sigma_0(\mathcal{X})$ and $\mathcal{V} = \sigma_1(\mathcal{Y})$

$$W(\mathcal{U})W(\mathcal{V}) \leq 3^{2(n-1)}$$

and thus by (14) and (13)

$$\begin{aligned} |\mathcal{X}||\mathcal{Y}| &= \sum_{U \in \mathcal{U}} |\sigma_0^{-1}(U) \cap \mathcal{X}| \cdot \sum_{V \in \mathcal{V}} |\sigma_1^{-1}(V) \cap \mathcal{Y}| \\ &\leq \sum_{U \in \mathcal{U}} 2^{n-|U|} \cdot \sum_{V \in \mathcal{V}} 2^{n-|V|} = W(\mathcal{U}) \cdot W(\mathcal{V}) \leq 3^{2n-1}. \end{aligned} \tag{15}$$

The characterization of the optimal pairs follows from the one in Theorem 2. We use right away the sequence terminology. If $(\mathcal{X}, \mathcal{Y}) \in \mathcal{D}_n^*$ is optimal, then applications of operations (d_t, u_t) and σ_0, σ_1 lead to an optimal $(\mathcal{U}, \mathcal{V}) \in \mathcal{P}_n$ by (15).

By the uniqueness part of Theorem 2 for some $t \in [n]$, without loss of generality say $t = n$, we have $\mathcal{U} = \mathcal{V} = \{A \in \mathcal{L}_n : t \in A\}$. Furthermore $(\sigma_0^{-1}(\mathcal{U}), \sigma_1^{-1}(\mathcal{V})) = (\{0,1,2\}^{n-1} \times \{0\}, \{0,1,2\}^{n-1} \times \{1\})$.

It remains to be seen that (d_n^{-1}, u_n^{-1}) leads to no non-isomorphic pairs. We have

$$d_n^{-1}(\{0,1,2\}^{n-1} \times \{0\}) = \mathcal{X}(0) \times \{0\} \,\dot{\cup}\, \mathcal{X}(1) \times \{1\} \,\dot{\cup}\, \mathcal{X}(2) \times \{2\},$$

$$u_n^{-1}(\{0,1,2\}^{n-1} \times \{1\}) = \mathcal{Y}(0) \times \{0\} \,\dot{\cup}\, \mathcal{Y}(1) \times \{1\} \,\dot{\cup}\, \mathcal{Y}(2) \times \{2\},$$

where by the optimality the $\mathcal{X}(i)$'s and also the $\mathcal{Y}(i)$'s partition $\{0,1,2\}^{n-1}$. Therefore for some i and some j, $(2,2,\dots,2) \in \mathcal{X}(i) \cap \mathcal{Y}(j)$. But now by (7) necessarily $\{i,j\} = \{0,1\}$ and $\mathcal{X}(i') = \emptyset$ for $i \neq i', \mathcal{Y}(i') = \emptyset$ for $j \neq j'$. We have arrived at the desired form.

6. Auxiliary Results for Proving Theorem 2

Obviously in deriving an upper bound on $W(\mathcal{U})W(\mathcal{V})$ for $(\mathcal{U},\mathcal{V}) \in \mathcal{P}_n$ we can always assume that $\mathcal{U}$ and $\mathcal{V}$ are upsets.

Moreover we can replace $(\mathcal{U},\mathcal{V})$ by the pair of images $(S_{ij}(\mathcal{U}), S_{ij}(\mathcal{V}))$ under the familiar left-shifting S_{ij}:

For any $\varepsilon \subset \mathcal{L}_n$, and $i < j$,

$$S_{ij}(E) = \begin{cases} E\Delta\{i,j\} & \text{if } i \notin E, j \in E \text{ and } E\Delta\{i,j\} \notin \varepsilon, \\ E & \text{otherwise.} \end{cases} \tag{16}$$

for $E \in \varepsilon$ and

$$S_{ij}(\varepsilon) = \{S_{ij}(E) : E \in \varepsilon\}.$$

Just verify that $(\mathcal{U},\mathcal{V}) \in \mathcal{P}_n$ implies $(S_{ij}(\mathcal{U}), S_{ij}(\mathcal{V})) \in \mathcal{P}_n$, that $|S_{ij}(\mathcal{U})| = |\mathcal{U}|, |S_{ij}(\mathcal{V})| = |\mathcal{V}|$, and that

$$W(S_{ij}(\mathcal{U})) = W(\mathcal{U}), W(S_{ij}(\mathcal{V})) = W(\mathcal{V}). \tag{17}$$

Clearly, finitely many applications of left-shifting operators results in a pair, which is invariant under further such operations. We call such a pair left-shifted.

Let now $(\mathcal{U},\mathcal{V}) \in \mathcal{P}_n$ be a pair of left-shifted upsets.

For the analysis of such pairs we introduce the following sets and families of sets.

For $A \subset [n]$ its projection on $[n-1]$ is

$${}_pA = A \cap [n-1] \tag{18}$$

and for $\mathcal{A} \subset \mathcal{L}_n$ we define

$${}_p\mathcal{A} = \{{}_pA : A \in \mathcal{A}\}. \tag{19}$$

Furthermore we partition $\mathcal{A}$ into

$$\mathcal{A}_0 = \{A \in \mathcal{A} : n \notin A\}, \qquad \mathcal{A}_1 = \{A \in \mathcal{A} : n \in A\}. \tag{20}$$

Thus also $\mathcal{U}_0$, $\mathcal{U}_1$, $\mathcal{V}_0$, $\mathcal{V}_1$, ${}_p\mathcal{U}_i$ and ${}_p\mathcal{V}_i$ $(i = 0, 1)$ are well-defined.

Since $\mathcal{U}$ and $\mathcal{V}$ are upsets

$${}_p\mathcal{U}_0 \subset {}_p\mathcal{U}_1 \quad \text{and} \quad {}_p\mathcal{V}_0 \subset {}_p\mathcal{V}_1. \tag{21}$$

Lemma 3. *If $(\mathcal{U},\mathcal{V}) \in \mathcal{P}_n$ cannot be enlarged without violating the cross-intersection property and $\mathcal{U},\mathcal{V}$ are left-shifted upsets, then for all $U \in \mathcal{U}_1$ with ${}_pU \in_p \mathcal{U}_1 \setminus {}_p\mathcal{U}_0$ there exists a $\mathcal{V} \in \mathcal{V}_1$ with ${}_p\mathcal{V} \in {}_p\mathcal{V}_1 \setminus {}_p\mathcal{V}_0$ such that*

(i) ${}_pU \cap {}_pV = \emptyset, {}_pU \cup {}_pV = [n-1]$ and

(ii) For all V' with ${}_pV' \in {}_p\mathcal{V}_1 \setminus_p \mathcal{V}_0$ and $V' \neq V$ necessarily ${}_pU \cap {}_pV' \neq \emptyset$.

Exchanging the roles of $\mathcal{U}$ and $\mathcal{V}$ gives analogous statements. Furthermore

(iii) ${}_p\mathcal{U}_0 = {}_p\mathcal{U}_1 \Leftrightarrow {}_p\mathcal{V}_0 = {}_p\mathcal{V}_1$.

Proof. (i) For every $U \in \mathcal{U}_1$ with ${}_pU \in {}_p\mathcal{U}_1 \setminus {}_p\mathcal{U}_0$ there must exist a $V \in \mathcal{V}_1$ with ${}_pV \in_p \mathcal{V}_1 \setminus_p \mathcal{V}_0$ with ${}_pU \cap_p V = \emptyset$, because otherwise U is intersecting on $[n-1]$ with all $V^* \in \mathcal{V}_1$, and by (21) with all $V^* \in \mathcal{V}$, and thus one can enlarge $\mathcal{U}$ by $U \setminus \{n\}$ in contradiction to our assumptions.

Furthermore for this $V {}_pU \cup {}_pV = [n-1]$, because otherwise for some $i \in [n-1] \setminus {}_pU \cup {}_pV U^* = S_{in}(U) \in \mathcal{U}$ by assumption and $U^* \cap V = \emptyset$ in contradiction to the fact that $(\mathcal{U}, \mathcal{V}) \in \mathcal{P}_n$.

(ii) The forgoing argument shows that ${}_pV' \not\subset {}_pV$. and that necessarily ${}_pU \cap {}_pV' \neq \emptyset$, because ${}_pU \cup {}_pV = [n-1]$.

(iii) This follows from (i), (ii) and the analogous statements obtained by exchanging the roles of U and V. □

7. Proof of Theorem 2

We proceed by induction on n. The case $n = 1$ is verified by inspection. For $n \geq 2$ we can consider a $(\mathcal{U}, \mathcal{V})$ satisfying the assumptions of Lemma 3.

Case: ${}_p\mathcal{U}_0 = {}_p\mathcal{U}_1$.

By (iii) in Lemma 3 we have also ${}_p\mathcal{V}_0 = {}_p\mathcal{V}_1$. However $({}_p\mathcal{U}_0, {}_p\mathcal{V}_0) \in \mathcal{P}_{n-1}$ and by the induction hypothesis

$$W({}_p\mathcal{U}_0)W({}_p\mathcal{V}_0) \leq 3^{2(n-2)}. \tag{22}$$

Now just calculate that in the present case

$$W(\mathcal{U})W(\mathcal{V}) = [W({}_p\mathcal{U}_0) \cdot 2 + W({}_p\mathcal{U}_1)] \cdot [W({}_p\mathcal{V}_0) \cdot 2 + W({}_p\mathcal{V}_1)]$$
$$= 3W({}_p\mathcal{U}_0) \cdot 3W({}_p\mathcal{V}_0) \leq 3^{2(n-1)}.$$

Case: ${}_p\mathcal{U}_0 \neq_p \mathcal{U}_1$.

By Lemma 3 ${}_p\mathcal{V}_0 \neq {}_P\mathcal{V}_1$ and there are subsets $U \in \mathcal{U}_1$, $\mathcal{V} \in \mathcal{V}_1$ satisfying (i). For all $V' \in \mathcal{V}_0$ necessarily ${}_pU \cap_p V' \neq \emptyset$ and again by Lemma 3 also for all $V' \in \mathcal{V}_1 \setminus \{V\}$, $V' \neq V$, ${}_pU \cap_p V' \neq \emptyset$. This means that $(\mathcal{U} \cup \{U \setminus \{n\}\}, \mathcal{V} \setminus \{V\}) \in \mathcal{P}_n$ and symmetrically $(\mathcal{U} \setminus \{U\}, \mathcal{V} \cup \{V \setminus \{n\}\}) \in \mathcal{P}_n$.

Moreover we see that

$$W(\mathcal{U} \cup \{U \setminus \{n\}\}) = W(\mathcal{U}) + 2^{n-|U|+1}, \tag{23}$$

$$W(\mathcal{V} \setminus \{V\}) = W(\mathcal{V}) - 2^{n-|V|} = W(\mathcal{V}) - 2^{|U|-1} \quad \text{(by (i))}, \tag{24}$$

$$W(\mathcal{U} \setminus \{U\}) = W(\mathcal{U}) - 2^{n-|U|}, \tag{25}$$

and

$$W(\mathcal{V} \cup \{V - \{n\}\}) = W(\mathcal{V}) + 2^{n-|V|+1} = W(\mathcal{V}) + 2^{|U|} \quad \text{(by (i))}. \tag{26}$$

By the optimality of $(\mathcal{U}, \mathcal{V})$ we conclude with (23) and (24) that

$$\begin{aligned} W(\mathcal{U})W(\mathcal{V}) &\geq W(\mathcal{U} \cup \{U \setminus \{n\}\})W(\mathcal{V} \setminus \{V\}) \\ &= (W(\mathcal{U}) + 2^{n-|U|+1})(W(\mathcal{V}) - 2^{|U|-1}) \\ &= W(\mathcal{U})W(\mathcal{V}) - 2^{|U|-1}W(\mathcal{U}) + 2^{n-|U|+1}W(\mathcal{V}) - 2^n \qquad (27) \end{aligned}$$

and with (25) and (26) that

$$\begin{aligned} W(\mathcal{U})W(\mathcal{V}) &\geq W(\mathcal{U} \setminus \{U\})W(V \cup \{V \setminus \{n\}\}) \\ &= (W(\mathcal{U}) - 2^{n-|U|})(W(\mathcal{V}) + 2^{|U|}) \\ &= W(\mathcal{U})W(\mathcal{V}) + 2^{|U|}W(\mathcal{U}) - 2^{n-|U|}W(\mathcal{V}) - 2^n. \qquad (28) \end{aligned}$$

Now (27) and (28) yield

$$-2^{|U|-1}W(\mathcal{U}) + 2^{n-|U|+1}W(\mathcal{V}) \leq 2^n \qquad (29)$$

and

$$2^{|U|}W(\mathcal{U}) - 2^{n-|U|}W(\mathcal{V}) \leq 2^n. \qquad (30)$$

The double of the left hand side in (29) plus the left hand side of (30) equals $3 \cdot 2^{n-|U|}W(\mathcal{V})$ and satisfies

$$3 \cdot 2^{n-|U|}W(\mathcal{V}) \leq 3 \cdot 2^n.$$

This is equivalent to

$$W(\mathcal{V}) \leq 2^{|U|}. \qquad (31)$$

Similarly, by doubling the left hand side of (30) and adding to it the left hand side of (29) leads to the inequality

$$W(\mathcal{U}) \leq 2^{n-|U|+1}. \qquad (32)$$

The two inequalities imply

$$W(\mathcal{U})W(\mathcal{V}) \leq 2^{n+1} < 3^{2(n-1)} \text{ for } n \geq 2. \qquad (33)$$

We calculate that for $\mathcal{U} = \mathcal{V} = \{A \subset [n] : 1 \in A\}$

$$W(\mathcal{U})W(\mathcal{V}) = 3^{2(n-1)}.$$

Finally we prove uniqueness. We have learnt already that for optimal left-shifted pairs $(\mathcal{U}, \mathcal{V})$ necessarily ${}_p\mathcal{U}_0 = {}_p\mathcal{U}_1, {}_p\mathcal{V}_0 = {}_p\mathcal{V}_1$ and that by the induction hypothesis $W({}_p\mathcal{U}_0)W({}_p\mathcal{V}_0) = 3^{2(n-2)}$. Thus $\mathcal{U} = \mathcal{V} = \{A \subset [n] : 1 \in A\}$. In general, every optimal pair $(\mathcal{U}^*, \mathcal{V}^*)$ can be left-shifted to $(\mathcal{U}, \mathcal{V})$. Since the left-shifting operators don't change cardinalities of subsets, there must be a singleton $\{t\}$ in both, $\mathcal{U}^*$ and $\mathcal{V}^*$. Consequently we have $\mathcal{U}^* = \mathcal{V}^* = \{A \subset [n] : t \in A\}$.

8. A Common Generalization of Theorem 2 and a Theorem of Erdős-Schőnheim [9]

In deriving their Intersection Theorem for multisets Erdős and Schőnheim established first an Intersection Theorem with weights for $\mathcal{L}_n$. Those weights $w(A), A \in \mathcal{L}_n$, are *increasing* in $|A|$, whereas our weights $w(A) = 2^{n-|A|}$ used in Theorem 2 are *decreasing* in $|A|$. The latter does not allow to just choose the "heavier" one of A and $A^c = [n] \setminus A$ in order to construct an optimal configuration. This difference makes things more difficult in our case. Nevertheless we can give a unified approach.

Let $\mathcal{W} = \{W_i : 1 \leq i \leq n\}$ be positive reals which give rise to the weight w on $\mathcal{L}_n$:

$$w(A) = \prod_{t \in A} W_t \text{ for } A \subset [n] \tag{34}$$

and

$$W(\mathcal{A}) = \sum_{A \in \mathcal{A}} w(A) \text{ for } \mathcal{A} \subset \mathcal{L}_n. \tag{35}$$

Define

$$\alpha(n, w) = \max\{W(\mathcal{A}) : \mathcal{A} \subset \mathcal{L}_n \text{ is intersecting}\} \tag{36}$$

(i.e. $A \cap B \neq \emptyset$ for $A, B \in \mathcal{A}$).

We recall first a result of [9].

Theorem 3.

$$\alpha(n, w) \leq \frac{1}{2} \max_{A \subset [n]} \big(w(A), w(A^c)\big) \tag{37}$$

and the bound is best possible when $W_i \geq 1$ for $i \in [n]$.

Proof. Clearly an intersecting $\mathcal{A}$ can have at most one of the sets A and A^c as member. □

One can construct an optimal intersecting family $\mathcal{A}(n, w)$ in the case

$$w_i \leq 1 \text{ for } i \in [n] \tag{38}$$

as follows:

(a) If $w(A) > w(A^c)$, then $A \in \mathcal{A}(n, w)$.
(b) If $w(A) = w(A^c)$ and $|A| > |A^c|$, then $A \in \mathcal{A}(n, w)$.
(c) If $w(A) = w(A^c)$ and $|A| = |A^c|$, then take anyone of A, A^c into $\mathcal{A}(n, w)$ and keep the other out of $\mathcal{A}(n, w)$. Clearly, $W(\mathcal{A}(n, w)) = \frac{1}{2} \sum_{A \subset [n]} \max\{w(A), w(A^c)\}$.

By (38) and (a)–(c) $\mathcal{A}(n, w)$ is an upset and also intersecting, because for $A, B \in \mathcal{A}(n, w) A \cap B = \emptyset$ implies $A^c \supset B$ and thus $A^c \in \mathcal{A}(n, w)$ in contradiction to (a)–(c).

However, without condition (38) the $\mathcal{A}(n, w)$ described above need not be an upset or intersecting.

For example when $w_i < 1$ for all $i \in [n]$, then the biggest weight is assigned to the empty set, which cannot occur in an intersecting family. Therefore (37) is not tight.

Fortunately an analysis of the proof of our Theorem 2 leads us to the right generalization.

First of all by relabelling we can always assume that

$$w_1 \geq w_2 \geq \cdots \geq w_n. \tag{39}$$

Let now m be the largest index with $w_m \geq 1$, if it exists, and otherwise set $m = 0$. Set $\mathcal{W}' = \{w_i : i \in [m]\}, w'(B) = \prod\limits_{i \in B} w_i$ for $B \subset [m]$.

Next define

$$\mathcal{A}^*(n, w) = \begin{cases} \{A \subset [n] : A \cap [m] \in \mathcal{A}(m, w')\} & \text{if } m \geq 1, \\ \{A \subset [n] : 1 \in A\} & \text{if } m = 0. \end{cases} \tag{40}$$

Clearly,

$$\mathcal{A}^*(n, w) = \mathcal{A}(n : w), \text{ if (38) holds.} \tag{41}$$

Theorem 4.

$$\alpha(n, w) = W(A^*(n, w)). \tag{42}$$

Proof. We use induction on $n - m$.

The case $n - m = 0$ or $n = m$ is the case covered by Theorem 3.

Case: $n - m > 0$

Suppose that $\mathcal{A}$ is an optimal intersecting family, that is,

$$W(\mathcal{A}) = \alpha(n, w). \tag{43}$$

Since (39) holds, the left-pushing operator S_{ij} can be applied, because it does not decrease the total weight. We can therefore assume that $\mathcal{A}$ is invariant under such operations. Also we can assume that $\mathcal{A}$ is an upset, because adding an $A' \subset [n]$ to $\mathcal{A}$ with $A' \supset A$ for some $A \in \mathcal{A}$ does not affect the intersection property and could only increase the total weight.

We use again the projection p on $[n-1]$ and our earlier definitions ${}_pA, {}_p\mathcal{A}, \mathcal{A}_i$, and ${}_p\mathcal{A}_i (i = 0, 1)$. Since $\mathcal{A}$ is an upset

$${}_p\mathcal{A}_0 \subset {}_p\mathcal{A}_1. \tag{44}$$

Case: ${}_p\mathcal{A}_0 = {}_p\mathcal{A}_1$.

Since ${}_p\mathcal{A}_0$ is intersecting, by the induction hypothesis in this case

$$\begin{aligned} W(\mathcal{A}) &= W({}_p\mathcal{A}_0) + w_n W({}_p\mathcal{A}_1) = (1 + w_n) W({}_p\mathcal{A}_1) \\ &\leq (1 + w_n) W(\mathcal{A}^*(n-1, w'')) = W(\mathcal{A}^*(n, w)), \end{aligned}$$

where $w'' = (w_i)_{i=1}^{n-1}$ and the last identity follows with definition (40).

Case: ${}_p\mathcal{A} \neq {}_p\mathcal{A}_1$.

Here there is an $A \in \mathcal{A}_1$ with $A \setminus \{n\} \notin \mathcal{A}_0$ and there must be a $B \in \mathcal{A}_1$ with

$$B \cap A = \{n\}, \tag{45}$$

because otherwise one can enlarge $\mathcal{A}$ by $A \setminus \{n\}$. Now the same ideas as used in the proof of Lemma 3 apply and give

$${}_pA \cup {}_pB = [n-1] \tag{46}$$

and consequently that the B with these properties is unique, because otherwise there is an $i \in {}_pA \cup {}_pB$ and, since $S_{in}(A) \in \mathcal{A}$, by (45)

$$S_{in}(A) \cap B = ({}_pA \cup \{i\}) \cap (B_p \cup \{n\}) = \emptyset.$$

This is a contradiction.

Since A and B can be exchanged, we can assume that.

$$w(A) \geq w(B) \tag{47}$$

and consequently that

$$w(A \setminus \{n\}) = \frac{w(A)}{w_n} > w(A) \geq w(B), \tag{48}$$

because $w_n < 1$, if $n - m > 0$.

Finally, since B is the unique member of $\mathcal{A}$ satisfying (45) $(\mathcal{A}\setminus\{B\})\cup\{A\setminus\{n\}\}$ is intersecting and by (48) has bigger weight than $\mathcal{A}$. This contradicts the optimality of $\mathcal{A}$. The case ${}_p\mathcal{A}_0 \neq {}_p\mathcal{A}_1$ cannot arise. □

9. Maximal Families of Disjoint Intervals

One might wonder what can be said about families $\mathcal{A} \subset \mathcal{I}_n$ with

$$A \wedge B = I_\emptyset \text{ for } A, B \in \mathcal{A}. \tag{49}$$

The family $\mathcal{A}$ corresponds to a set $\mathcal{A}^* \subset \{0,1,2\}^n$ by the mapping Ψ of Sect. 2. $\mathcal{A}^*$ has the property:

$$\text{for all } x^n, y^n \in \{0,1,2\}^n \text{ for some } t \in [n]\{x_t, y_t\} = \{0,1\}. \tag{50}$$

One readily verifies that $|\mathcal{A}| \leq 2^n$, and equality occurs for $\mathcal{A}^* = \{0,1\}^n$. In fact the problem is equivalent to Shannon's zero error capacity problem for the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

As Shannon noticed in [11], it equals $\log_2 2 = 1$.

References

1. R. Ahlswede and Z. Zhang, “On cloud-antichains and related configurations”, Discrete Mathematics 85, 225–245, 1990.
2. D.E. Daykin, D.J. Kleitman, and D.B. West, “The number of meets between two subsets of a lattice”, J. Comb. Theory, Ser. A, 26, 135–156, 1979.
3. R. Ahlswede and L.H. Khachatrian, “Sharp bounds for cloud-antichains of length two”, SFB 343, Preprint. 92–012.
4. R. Ahlswede and L.H. Khachatrian, “Towards equality characterization in correlation inequalities”, SFB 343, Preprint 93–027, to appear in European J. of Combinatorics.
5. R. Ahlswede and L.H. Khachatrian, “Optimal pairs of incomparable clouds in multisets”, Graphs and Combinatorics 12, 97–137, 1996.
6. P. Seymour, “On incomparable families of sets”, Mathematica 20, 208–209, 1973.
7. D.J. Kleitman, “Families of non-disjoint subsets”, J. Comb. Theory 1, 153–155, 1966.
8. R. Ahlswede and D.E. Daykin, “An inequality for the weights of two families of sets, their unions and intersections”, Z. Wahrscheinlichkeitstheorie u. verw. Geb. 43, 183–185, 1978.
9. P. Erdős and J. Schonheim, “On the set of non pairwise coprime division of a number”, Proc. of the Colloquium on Comb. Math. Dalaton Füred, 369–376, 1969.
10. P. Erdős, M. Herzog and J. Schőnheim, “An extremal problem on the set of noncoprime divisors of a number”, Israel J. Math., vol. 8, 408–412, 1970.
11. C.E. Shannon, “The zero-error capacity of a noisy channel”, IRE Trans. Inform. Theory, vol. IT-2, no. 3, 8–19, 1956.

Classical Results on Primitive and Recent Results on Cross-Primitive Sequences

Rudolf Ahlswede and Levan H. Khachatrian

R. Ahlswede (Deceased) · L.H. Khachatrian (Deceased)
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501
Bielefeld, Germany

Summary. When the kind invitation of Ron Graham and Jaroslav Nešetřil, to write in honour of Paul Erdős about aspects of his work, reached us, our first reaction was to follow it with great pleasure. Our second reaction was not as clear: Which one among the many subjects in mathematics, to which he has made fundamental contributions, should we choose?

Finally we just followed the most natural idea to write about an area which just had started to fascinate us: Density Theory for Integer Sequences.

More specifically we add here to the classical theory of primitive sequences and their sets of multiples results for cross-primitive sequences, a concept, which we introduce. We consider both, density properties for finite and infinite sequences. In the course of these investigations we naturally come across the main theorems in the classical theory and the predominance of results due to Paul Erdős becomes apparent. Several times he had exactly proved the theorems we wanted to prove! Many of them belong to his earliest contribution to mathematics in his early twenties.

Quite luckily our random approach led us to the perhaps most formidable period in Erdős' work. It reminds us about a statement, which K. Reidemeister [18, ch. 8] made about Carl Friedrich Gauss: "…Aber das Epochale ist doch die geniale Entdeckung des Jünglings: die Zahlentheorie."

1. Classical Results

At first we set up our notation. $\mathbb{N}$ denotes the set of positive integers and $\mathbb{P} = \{p_1, p_2, \ldots\} = \{2, 3, 5, \ldots\}$ denotes the set of all primes. For the number $u, v \in \mathbb{N}$ we write $u \mid v$, if u divides v. Further (u, v) stands for the largest common divisor and $\langle u, v \rangle$ denotes the smallest common multiple of u and v.

In case $(u, v) = 1$, u and v are said to be relatively prime (or coprimes). The greatest prime factor of u is written as $p^+(u)$. For $i \leq j$, $[i, j]$ equals $\{i, i+1, \ldots, j\}$ and $(i, j]$ equals $\{i+1, \ldots, j\}$. Any set $A \subset \mathbb{N}$ can also be viewed as an increasing sequence $(a_i)_{i=1}^{\infty}$ where $A = \{a_i : i \in \mathbb{N}\}$, and vice versa. We reserve the letter A for such sets or sequences. It is convenient to use the abbreviations $A(x) = A \cap [1, x]$ and $|B|$ for the cardinality of any set B. We also use $\phi(x, y) = |\{n \in [1, x] : p^+(n) > y\}|$.

The lower and upper asymptotic density of A are

$$\underline{\mathbf{d}}\ A = \liminf_{x \to \infty} \frac{|A(x)|}{x} \qquad \text{and} \qquad \overline{\mathbf{d}}\ A = \limsup_{x \to \infty} \frac{|A(x)|}{x}. \tag{1}$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_9,

If $\underline{\mathbf{d}}\ A = \overline{\mathbf{d}}\ A$, then A possesses the asymptotic density $\mathbf{d}\ A = \underline{\mathbf{d}}\ A = \overline{\mathbf{d}}\ A$. Related quantities are

$$\underline{\delta}\ A = \liminf_{x \to \infty} \frac{1}{\log\ x} \sum_{a_i \leq x} \frac{1}{a_i} \qquad \text{and} \qquad \overline{\delta}\ A = \limsup_{x \to \infty} \frac{1}{\log\ x} \sum_{a_i \leq x} \frac{1}{a_i}, \tag{2}$$

the logarithmic lower and upper density of A. If $\underline{\delta}\ A = \overline{\delta}\ A$, then A possesses logarithmic density $\delta\ A = \underline{\delta}\ A = \overline{\delta}\ A$.

In the first half of the century there was noticeable interest in the study of density properties of the *set of multiples*

$$M(A) = \{m \in \mathbb{N} : \text{ for some } a \in A \quad a \mid m\} \tag{3}$$

of infinite sequences A of positive integers. This naturally relates to the study of primitive sequences.

A sequence $A = (a_i)_{i=1}^{\infty}$ is primitive, if

$$a_i \nmid a_j \text{ for } i \neq j. \tag{4}$$

One readily verifies that every A contains a unique subsequence $P(A)$ which is primitive and satisfies

$$M\big(P(A)\big) = M(A). \tag{5}$$

Actually,

$$P(A) = \{a \in A : \nexists\ b \in A, b \neq a \text{ and } b \mid a\} \tag{6}$$

One question of Chowla (see [2]) opened the subject: Does $\mathbf{d}\ M(A)$ exist for every $A \subset \mathbb{N}$?

This can readily be shown to be the case for all finite A, however, this was open for a longer time and finally settled in the negative by Besicovitch [2] in the infinite case.

Theorem 1 (Besicovitch [2]). *For every $\varepsilon > 0$ there is an $A \subset \mathbb{N}$ with*

$$\overline{\mathbf{d}}\ M(A) \geq \frac{1}{2} \quad \textit{and} \quad \underline{\mathbf{d}}\ M(A) \leq \varepsilon.$$

Actually, the A's are constructed as unions of suitable intervals. The primitive sequence $P(A)$ generating the $M(A)$ of Theorem 1 gives the next famous result.

Theorem 2 (Besicovitch [2]). *For every $\varepsilon > 0$ there is primitive sequence A' with*

$$\overline{\mathbf{d}}\ A' \geq \frac{1}{2} - \varepsilon \quad \textit{and} \quad \underline{\mathbf{d}}\ A' \leq \varepsilon.$$

This shows that a question of Davenport and Erdős (see [7, 13]), whether every primitive sequence has asymptotic density 0, has a negative answer.

We derive next an upper bound on $\overline{\mathbf{d}}\ A$ because it is instructive and beautiful. For any primitive $A = \{a_1, \dots, a_\alpha\} \subset [1, 2n]$ let d_i denote the greatest odd divisor of a_i. Then necessarily $d_1, \dots, d_\alpha$ are all distinct and hence

$$|A| = \alpha \leq n. \tag{7}$$

Theorem 3 (**Behrend** [3]). *For every primitive A, $\overline{\mathbf{d}}\ A \leq \frac{1}{2}$.*

Example 1 (**Everybody**). *$\{n+1, \dots, 2n\}$ is primitive and has density $\frac{1}{2}$.*

This simple fact is very relevant in the analysis of infinite primitive sequences.

Now Paul Erdős enters the scene.

Theorem 4 (**Erdős** [5]). *For a primitive $A \not\supset \{1\}$*

$$\sum_{i=1}^{\infty} \frac{1}{a_i \log a_i} < \infty.$$

It is an open problem of Erdős whether $\sum\limits_{i=1}^{\infty} \frac{1}{a_i \log a_i} \leq \sum\limits_{i=1}^{\infty} \frac{1}{p_i \log p_i}$.

By Abel summation it can be shown that for any set $B \subset \mathbb{N}$

$$0 \leq \underline{\mathbf{d}}\ B \leq \underline{\delta}\ B \leq \overline{\delta}\ B \leq \overline{\mathbf{d}}\ B \leq 1. \tag{8}$$

Since $\frac{1}{\log n} \sum_{N < a_i \leq n} \frac{1}{a_i} \leq \sum\limits_{N < a_i \leq n} \frac{1}{a_i \log a_i}$, by Theorem 4 $\delta\ A = 0$ for primitive A. Also by (8) $\underline{\mathbf{d}}\ A = 0$. We state this result.

Theorem 5 (**Erdős** [5]). *For every primitive sequence A, $\underline{\mathbf{d}}\ A = \delta A = 0$ or (equivalently)*

$$\frac{1}{\log n} \sum_{a_i \leq n} \frac{1}{a_i} = o(1) \ \textit{as}\ n \to \infty. \tag{9}$$

Logarithmic density has turned out to be an appropriate measure! Also, what can be said about the speed in (9)?

Theorem 6 (**Behrend** [3]). *There is a constant γ such that for every primitive sequence A*

$$\frac{1}{\log n} \sum_{a_i \leq n} \frac{1}{a_i} \leq \gamma \frac{1}{(\log \log n)^{\frac{1}{2}}} \ \textit{for}\ n \geq 3. \tag{10}$$

In the proof the general case is reduced to A's consisting entirely of square-free integers and their analysis is based on Sperner's Lemma [1]! This gave a strong impetus also to combinatorial extremal theory starting with [12] and continuing with [24], ..., [30] and many, many others.

Theorem 6 is best possible in the sense that γ cannot be replaced by $o(n)$.

Theorem 7 (**S. Pillai** [8]). *There exists a positive constant c, such that to every $x \leq 3$ corresponds a primitive set A_x with*

$$\frac{1}{\log x} \sum_{a_i \leq x} \frac{1}{a_i} > \frac{c}{(\log \log x)^{\frac{1}{2}}}.$$

Subsequently Erdős, Sárközy, and Szemerédi [20] showed that c can be chosen as $(2\pi)^{-\frac{1}{2}} - \varepsilon$ for any $\varepsilon > 0$ and that this is best possible.

The last three theorems concern in essence only finite primitive sequences. Related to infinite primitive sequences in the true sense is the following.

Theorem 8 (Erdős, Sárközy, Szemerédi [21]). *For every infinite primitive sequence A*

$$\sum_{a_i \leq x} \frac{1}{a_i} = o\left(\frac{\log x}{(\log \log x)^{\frac{1}{2}}} \right)$$

and this bound is best possible.

We draw attention also to a survey paper of Erdős, Sárközy, and Szemerédi [22] and to a related paper of Pomerance and Sárközy [23].

Concerning $\overline{\mathbf{d}}\ A$ there is the following improvement of Theorem 3.

Theorem 9 (Erdős [14]). *Let A be an infinite primitive sequence, then for every $a \in A$ of the form $a = 2^u(2v+1) \leq n; u, v \geq 0$,*

$$|A(n)| \leq n - \left\lfloor \frac{1}{2} n \right\rfloor - \left\lfloor \frac{1}{2} \left(\frac{n}{3^u(2v+1)} - 1 \right) \right\rfloor.$$

Hence, $\overline{\mathbf{d}}\ A < \frac{1}{2}$. After Besicovitch's negative answer to Chowla's question, it is natural to address the next question: Under which conditions on A does $\mathbf{d}\ M(A)$ or $\delta\ M(A)$ exist? Davenport-Erdős and Erdős answered all these questions: We derive here the simplest and most transparent result, Theorem 10 below, in order to explain the important role of a quantity, which we consider to be a density concept for sets of multiples and denote as μ.

Since A is fixed, we write M for $M(A)$. Further we denote by $M_m = M_m(A)$ the set of multiples of the first m elements of A, namely $a_1, a_2, \dots, a_m$. M_m can be represented as the union of a finite number of congruence classes, and therefore possesses asymptotic density. If we denote by $M^{(i)}(n)$ the natural numbers, not exceeding n, which are divisible by a_i but not divisible by anyone of $a_1, \dots, a_{i-1}$, then we have

$$M_m(n) = \dot{\bigcup_{i=1}^{m}} M^{(i)}(n). \tag{11}$$

By inclusion-exclusion for every $i = 1, 2, 3, \dots$,

$$|M^{(i)}(n)| = \left\lfloor \frac{n}{a_i} \right\rfloor - \sum_{j<i} \left\lfloor \frac{n}{\langle a_i, a_j \rangle} \right\rfloor + \sum_{k<j<i} \left\lfloor \frac{n}{\langle a_k, a_i, a_j \rangle} \right\rfloor - \cdots$$

and

$$\mathbf{d}\ M^{(i)} = \lim_{n\to\infty} \frac{|M^{(i)}(n)|}{n} = \frac{1}{a_i} - \sum_{j<i} \frac{1}{\langle a_j, a_i\rangle} + \cdots .$$

Therefore by (11)

$$\mathbf{d}\ M_m = \sum_{i=1}^{m} \mathbf{d}\ M^{(i)}. \tag{12}$$

Since $0 < \sum\limits_{i=1}^{m} \mathbf{d}\ M^{(i)} < 1$ and $\mathbf{d}\ M^{(i)} \geq 0$, $\lim\limits_{m\to\infty} \mathbf{d}\ \mathbf{M_m} = \sum_{\mathrm{i}=1}^{\infty} \mathbf{d}\ \mathbf{M^{(i)}}$ exists. We define now the "density" μ by

$$\mu\ A = \lim_{m\to\infty} \mathbf{d}\ M_m(A), \qquad A \subset \mathbb{N}. \tag{13}$$

Since $M_m(A) \subset M(A)$, we see immediately that

$$\mu\ A \leq \underline{\mathbf{d}}\ M(A). \tag{14}$$

Suppose now that $\sum\limits_{i=1}^{\infty} a_i^{-1} < \infty$. Then $\overline{\mathbf{d}}\ M(A) \leq \mathbf{d}\ M_m(A) + \sum\limits_{i=m+1}^{\infty} \frac{1}{a_i}$ and thus $\overline{\mathbf{d}}\ M(A) \leq \mu\ A \leq \underline{\mathbf{d}}\ M(A)$.

Theorem 10 (**[17]**). *If* $\sum\limits_{i=1}^{\infty} a_i^{-1} < \infty$, *then* $\mathbf{d}M(A)$ *exists and equals* μA.

Here are the highlights.

Theorem 11 (**Davenport-Erdős [7], also [13]**). *For any* $A \subset \mathbb{N}$, $M(A)$ *has logarithmic density and*

$$\delta M(A) = \underline{\mathbf{d}}\ M(A) = \mu A.$$

Theorem 12 (**Erdős [11]**). *A necessary and sufficient condition for* $\mathbf{d}\ M(A)$ *to exist is*

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \sum_{n^{1-\varepsilon} < a_i \leq n} |M^{(i)}(n)| = 0. \tag{15}$$

Even though condition (15) looks complicated, it yields a useful sufficient condition.

Theorem 13 (**Erdős [11]**). *If* $A \subset \mathbb{N}$ *satisfies for some constant* c, $|A(n)| \leq \frac{cn}{\log\ n}$ *for* $n \geq 2$, *then* $\mathbf{d}\ M(A)$ *exists.*

The case $A \subset \mathbb{P}$ is included here. The result is best possible in the following sense.

Theorem 14 (**Erdős [12]**). *For any monotonically increasing function* $\Psi : \mathbb{N} \to \mathbb{R}_+$ *with* $\lim\limits_{n\to\infty} \Psi(n) = \infty$ *there exists an* $A \subset \mathbb{N}$ *such that*

$$|A(n)| \leq const \, \frac{n \, \Psi(n)}{\log n} \quad \textit{for large } n,$$

but $\mathbf{d}\ M(A)$ *does not exist.*

We present now two further results with many applications.

The first of them was probably motivated by Example 1. It shows how the set of multiples of certain intervals behaves in density. This is the key idea in Besicovitch's construction [2]. Erdős improved the length of the intervals.

Theorem 15 (Erdős [5]). *The intervals* $(T^{1-\varepsilon}, T] \subset \mathbb{N}$ *satisfy*

$$\lim_{\substack{\varepsilon \to 0 \\ T \to \infty}} \mathbf{d}\ M((T^{1-\varepsilon}, T]) = 0.$$

The second result is the famous Behrend Lemma in a dual formulation, that is, for $X \subset \mathbb{N}$ we use $M(X)$ instead of $\mathbb{N} \setminus M(X)$.

Lemma 1 (Behrend [10]). *Let* $A, B \subset \mathbb{N}$ *be finite, then*

$$\mathbf{d}\ M(A) \cdot \mathbf{d}\ M(B) \leq \mathbf{d}\ \big(M(A) \cap M(B)\big).$$

Moreover, equality holds exactly if the primitive sets $P(A)$ and $P(B)$ satisfy $(a,b) = 1$ for all $a \in P(A), b \in P(B)$.

Finally there are also several papers concerning the growth of $\phi(x,y)$ [15]. We use later only the following result.

Theorem 16 (Chowla and Vijayaraghavan [15]).

$$\lim_{x \to \infty} \frac{\phi(x, x^\theta)}{x} = \log \frac{1}{\theta}, \ \textit{for } \frac{1}{2} \leq \theta < 1.$$

Remark 1. *We apologize for not including in our sketch several results of basic nature such as Rogers inequality [17] and others. Our selection is guided by our present research interest. The reader may consult the books by Halberstam and Roth [17] and Hall and Tenenbaum [19].*

2. New Results

We introduce a seemingly basic and new concept.

The pair of sets (or sequences) (A, B) with $A, B \subset \mathbb{N}$ is called *cross-primitive*, if

$$a \nmid b \text{ and } b \nmid a \text{ for all } a \in A, b \in B. \tag{16}$$

It is convenient to denote the set of all cross-primitive pairs (A, B) with $A, B \subset \mathbb{N}(x)$ (resp. $\mathbb{N}$) by $\text{Cross}(x)$ (resp. $\text{Cross}(\infty)$). We are again interested in density properties. We begin with the finite case and define

$$c(x) = \max_{(A,B) \in \text{Cross}(x)} \frac{|A| \cdot |B|}{x^2}.$$

Theorem 17. *For all* $x \in \mathbb{N}$, $c(x) < \frac{1}{4}$ *and*

$$\lim_{x\to\infty} c(x) = \frac{1}{4}.$$

Remark 2. *As analogue for a primitive sequence see the simple Example 1 and (7). We believe that our construction is optimal for large* x*. Erdős thinks that the deviation of* $\max\limits_{(A,B)\in \mathrm{Cross}(x)} |A||B|$ *from* $\frac{x^2}{4}$ *is of the order* x^α *for some* $\alpha > 1$*.*

The infinite case shows more complex behaviour and that's the case where also several classical results on primitive sequences are used.

Theorem 18.

$$\max_{(A,B)\in \mathrm{Cross}(\infty)} \underline{\mathbf{d}}\ A \cdot \underline{\mathbf{d}}\ B = \frac{1}{16}.$$

The maximum is assumed for a pair with densities.

One auxiliary result for proving this Theorem deserves special attention. It is an infinite form of Behrend's Lemma 1, but by no means an easy extension. On the other hand, it involves the essence of the Davenport-Erdős Theorem 11 and expresses it in an elegant way.

Lemma 2. *For arbitrary* $A, B \subset \mathbb{N}$

$$\underline{\mathbf{d}}\ M(A) \cdot \underline{\mathbf{d}}\ M(B) \le \underline{\mathbf{d}}\ (M(A) \cap M(B)).$$

We use another auxiliary result, which should be known to the experts, but we could not find stated in the literature.

Lemma 3. *For any* $0 < \lambda \le 1$*, and any* $q_1 \in \mathbb{P}$, $\frac{1}{q_1} \le \lambda$*, there exists a set of primes* $Q = \{q_1 < q_2 < \cdots\} \subset \mathbb{P}$ *with*

$$\mathbf{d}\ M(Q) = \lambda.$$

Finally, we enter the world of pathologies discovered by Besicovitch.

Theorem 19.

$$\sup_{(A,B)\in \mathrm{Cross}(\infty)} \overline{\mathbf{d}}\ A \cdot \overline{\mathbf{d}}\ B = 1.$$

The supremum cannot be assumed: For $A = \{a_1, \ldots\}$, $\overline{\mathbf{d}}\ B \le 1 - \frac{1}{a_1}$*.*

Remark 3. *The construction in the proof of this Theorem can be used to show that in Lemma 2* $\underline{\mathbf{d}}$ *cannot be replaced by* $\overline{\mathbf{d}}$*.*

3. Proof of Theorem 17

Since for $(A,B) \in \text{Cross}(x)$, A and B must be disjoint and $1 \notin A \cup B$ necessarily $|A| + |B| < x$ and $\frac{|A|\cdot|B|}{x^2} < \frac{1}{4}$. Therefore also $c(x) < \frac{1}{4}$ for all $x \in \mathbb{N}$.

To complete the proof, we have to construct a sequence $(A_x, B_x)_{x=1}^{\infty}$ with $(A_x, B_x) \in$ Cross(x) and

$$\lim_{x\to\infty} \frac{|A_x|\cdot|B_x|}{x^2} = \frac{1}{4}. \tag{17}$$

We define for a $\theta, \frac{1}{2} \leq \theta < 1$, which we adjust later,

$$A_x = \{a \in \mathbb{N} : x^{1-\theta} \leq a \leq x \text{ and } p^+(a) \leq x^\theta\},$$

$$B_x = \{b \in \mathbb{N} : b \leq x \text{ and } p^+(b) > x^\theta\},$$

and observe that for a θ in the specified interval $(A_x, B_x) \in$ Cross(x). Hence

$$|A_x| \geq x - x^{1-\theta} - |B_x|. \tag{18}$$

Now Theorem 16 says that $\lim\limits_{x\to\infty} \frac{|B_x|}{x} = \log\frac{1}{\theta}$ and since $x^{1-\theta} = o(x)$,

$$|B_x| \sim x\log\frac{1}{\theta},\ |A_x| \gtrsim x(1-\log\frac{1}{\theta}). \tag{19}$$

We choose now θ such that $\log\frac{1}{\theta} = 1-\log\frac{1}{\theta} = \frac{1}{2}$, that is, $\theta = e^{-\frac{1}{2}} \sim 0.6065 > \frac{1}{2}$. Clearly, (19) implies now (17).

Remark 4. *A good estimate of $|B_x|$ is possible, because B_x can be partitioned according to the biggest prime in the decomposition of its members. These biggest primes are essentially known in magnitude by the Prime Number Theorem. Our first proof followed this line. Then we learnt about [15].*

4. Proof of Lemma 2

Behrend's Lemma implies that for every $n \in \mathbb{N}$

$$\mathbf{d}\ M(A(n)) \cdot \mathbf{d}\ M(B(n)) \leq \mathbf{d}\ \big(M(A(n)) \cap M(B(n))\big). \tag{20}$$

Since $A(n)$, $B(n)$ and thus also $A(n) \cap B(n)$ are monotonically increasing in n, we have

$$M(A(n)) \cap M(B(n)) \subset M(A) \cap M(B)$$

and therefore

$$\underline{\mathbf{d}}\ \big(M(A(n)) \cap M(B(n))\big) \leq \underline{\mathbf{d}}\ \big(M(A) \cap M(B)\big). \tag{21}$$

Since $\mathbf{d}\ \big(M(A(n)) \cap M(B(n))\big) = \underline{\mathbf{d}}\ (M(A(n)) \cap M(B(n)))$, (20) and (21) imply

$$\mathrm{d}\ M(A(n)) \cdot \mathbf{d}\ M(B(n)) \le \underline{\mathbf{d}}\ (M(A) \cap M(B)). \tag{22}$$

Now by Theorem 11

$$\lim_{n\to\infty} \mathbf{d}\ M(A(n)) = \underline{\mathbf{d}}\ M(A), \lim_{n\to\infty} \mathbf{d}\ M(B(n)) = \underline{\mathbf{d}}\ M(A)$$

and therefore

$$\underline{\mathbf{d}}\ M(A)\underline{\mathbf{d}}\ M(B) \le \underline{\mathbf{d}}\ (M(A) \cap M(B)).$$

5. Proof of Lemma 3

For any $Q = \{q_1 < q_2 < \ldots\} \subset \mathbb{P}$ by the Prime Number Theorem (or weaker versions)

$$|Q \cap [1, n]| < \text{const} \cdot \frac{n}{\log\, n}.$$

Therefore by Theorem 13 $M(Q)$ possesses asymptotic density and by Theorem 11

$$\underline{\mathbf{d}}\ M(Q) = \mathrm{d}\ M(Q) = \sum_{i=1}^{\infty} q^{(i)},$$

where

$$q^{(i)} = \frac{1}{q_i} - \sum_{j<i} \frac{1}{q_j q_i} + \sum_{k<j<i} \frac{1}{q_k q_j q_i} - \ldots, \tag{23}$$

because $Q \subset \mathbb{P}$. Therefore

$$\sum_{i=1}^{\infty} q^{(i)} = 1 - \prod_{i=1}^{\infty} \left(1 - \frac{1}{q_i}\right)$$

and now the statement follows from $\sum_{i=1}^{\infty} \frac{1}{p_i} = \infty$, because $-\log(1 - \frac{1}{q_i}) > \frac{1}{q_i}$ and for any null sequence $\{a_i\}_{i=1}^{\infty}$ of positive numbers with $\sum_{i=1}^{\infty} a_i = \infty$ any real number $r > 0$ equals $\sum_{j=1}^{\infty} a_{ij}$ for a suitable subsequence $\{a_{ij}\}_{j=1}^{\infty}$.

6. Proof of Theorem 18

We show first that for $(A, B) \in \text{Cross}(\infty)$

$$\underline{\mathbf{d}}\ A \cdot \underline{\mathbf{d}}\ B \le \frac{1}{16}. \tag{24}$$

We associate with (A, B) the sets

$$A^* = M(A) \setminus (M(A) \cap M(B)),$$
$$B^* = M(B) \setminus (M(A) \cap M(B)),$$

and observe that also $(A^*, B^*) \in \text{Cross}(\infty)$. Moreover, we notice that

$$A \subset A^* \text{ and } B \subset B^* \tag{25}$$

and that

$$M(A) \cap M(B) = M(C), \tag{26}$$

where

$$C = \{\langle a, b\rangle : a \in A, b \in B\}. \tag{27}$$

By our definitions and properties (25) and (26) we have also

$$A \cap [1, x] \subset (M(A) \cap [1, x]) \setminus (M(C) \cap [1, x]) \tag{28}$$

and therefore

$$|A \cap [1, x]| \leq |M(A) \cap [1, x]| - |M(C) \cap [1, x]|. \tag{29}$$

Let now $(x_i)_{i=1}^{\infty}$ be an increasing sequence of positive integers with

$$\lim_{i\to\infty} \frac{|M(A) \cap [1, x_i]|}{x_i} = \underline{\mathbf{d}}\ M(A). \tag{30}$$

Then by (29) and (30)

$$\underline{\mathbf{d}}\ A \leq \liminf_{i\to\infty} \frac{|A \cap [1, x_i]|}{x_i} \leq \underline{\mathbf{d}}\ M(A) - \liminf_{i\to\infty} \frac{|M(C) \cap [1, x_i]|}{x_i} \leq \underline{\mathbf{d}}\ M(A) - \underline{\mathbf{d}}\ M(C). \tag{31}$$

Now we lower bound $\mathbf{d}\ M(C)$ by Lemma 2:

$$\underline{\mathbf{d}}\ M(C) = \underline{\mathbf{d}}\ (M(A) \cap M(B)) \geq \underline{\mathbf{d}}\ M(A) \cdot \underline{\mathbf{d}}\ M(B)$$

and conclude that

$$\underline{\mathbf{d}}\ A \leq \underline{\mathbf{d}}\ M(A) - \underline{\mathbf{d}}\ M(A) \cdot \underline{\mathbf{d}}\ M(B) = \underline{\mathbf{d}}\ M(A)(1 - \underline{\mathbf{d}}\ M(B)). \tag{32}$$

Symmetrically

$$\underline{\mathbf{d}}\ B \leq \underline{\mathbf{d}}\ M(B)(1 - \underline{\mathbf{d}}\ M(A)) \tag{33}$$

and thus finally

$$\underline{\mathbf{d}}\ A \cdot \underline{\mathbf{d}}\ B \leq \underline{\mathbf{d}}\ M(A)(1 - \underline{\mathbf{d}}\ M(A)) \cdot \underline{\mathbf{d}}\ M(B)(1 - \underline{\mathbf{d}}\ M(B)) \leq \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}.$$

We construct now $(A, B) \in \text{Cross}(\infty)$ with

$$\mathbf{d}\ A \cdot \mathbf{d}\ B = \frac{1}{16}. \tag{34}$$

By Lemma 3 there is a $Q = \{q_1 \subset q_2 \dots\} \subset \mathbb{P}$ with

$$\mathbf{d}\ M(Q) = \frac{1}{2} \text{ and } q_1 > 2. \tag{35}$$

Set

$$A = \{a \in \mathbb{N} : 2 \mid a \text{ and } q_i \nmid \text{ for all } q_i \in Q\},$$
$$B = \{b \in \mathbb{N} : 2 \nmid b \text{ and } b \in M(Q)\}.$$

Equivalently

$$A = M(\{2\}) \setminus (M(\{2\}) \cap M(Q)) \text{ and } B = M(Q) \setminus (M(\{2\}) \cap M(Q)).$$

Also, it is clear that $(A, B) \in \text{Cross}(\infty)$ and that $M(\{2\}) \cap M(Q) = M(C)$, where

$$C = \{2q_i : q_i \in Q\}.$$

Obviously $|C \cap [1, n]| \leq \text{ const}\frac{n}{\log n}$ and again by Theorem 13 $M(C)$ has asymptotic density and is given by the formula $\mathbf{d}\ M(C) = \sum\limits_{i=1}^{\infty} q_*^{(i)}$, where

$$q_*^{(i)} = \frac{1}{2q_i} - \sum_{j<i} \frac{1}{2q_j q_i} + \sum_{k<j<i} \frac{1}{2q_k q_j q_i} \dots = \frac{q^{(i)}}{2}.$$

Hence $\mathbf{d}\ M(C) = \frac{1}{2} \sum\limits_{i=1}^{\infty} q^{(i)} = \frac{\mathbf{d}\ M(Q)}{2} = \frac{1}{4}$. Therefore

$$\mathbf{d}\ A = \mathbf{d}\ M(\{2\}) - \mathbf{d}\ (C) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}, \mathbf{d}\ B = \mathbf{d}\ M(Q) - \mathbf{d}\ (C) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4},$$

and (34) holds.

7. Proof of Theorem 19

Let us fix $\delta > 0, \delta_i > 0$ for $i \in \mathbb{N}$, $\sum\limits_{i=1}^{\infty} \delta_i = \delta$ and $0 < \theta < 1$.

By Theorem 15 for a $\delta_i > 0$ there are positive numbers $T(\delta_i)$ and $\lambda(\delta_i)$ such that

$$\mathbf{d}\ M([T^{1-\lambda_i}, T]) < \delta_i \quad \text{for} \quad T > T(\delta_i), \lambda_i < \lambda(\delta_i). \tag{36}$$

We fix arbitrary $\lambda_i \in (0, \lambda(\delta_i))$ for $i \in \mathbb{N}$ and $\lambda^* \in (0, 1)$. Now (36) and the definition of density tell us that for $T_j > T(\delta_j)$ $(j = 1, 2, \dots, i)$ and $S > S(\lambda^*, T_1, T_2, \dots, T_i, \lambda_1, \dots, \lambda_i, \delta_1, \dots, \delta_i)$ (suitable) simultaneously

$$\frac{|M([T_j^{1-\lambda_j}, T_j]) \cap [S^{1-\lambda^*}, S]|}{S - S^{1-\lambda^*}} < \delta_j \text{ for } j \leq i. \tag{37}$$

Now let R_1 be an integer with the properties

$$R_1 > T(\delta_1) \text{ and } \frac{1}{R_1^{\lambda_1}} < \theta. \tag{38}$$

We fix the interval $[R_1^{1-\lambda_1}, R_1]$. Let L_1 be an integer with the properties

$$L_1^{1-\lambda_1} > R_1 \text{ and } L_1^{1-\lambda_1} > S(\lambda_1, R_1, \lambda_1, \delta_1). \tag{39}$$

We choose the interval $[L_1^{1-\lambda_1}, L_1]$. Furthermore, we choose R_2, L_2 such that $L_1 < R_2^{1-\lambda_2}$, $R_2 < L_2^{1-\lambda_2}$, $R_2^{1-\lambda_2} > \max\{T(\delta_2), S(\lambda_2, L_1, \lambda_1, \delta_1)\}$, $\frac{1}{R_2^{\lambda_2}} < \theta$ and $L_2^{1-\lambda_2} > S(\lambda_2, R_1, R_2, \lambda_1, \lambda_2, \delta_1, \delta_2)$. We fix now intervals $[R_2^{1-\lambda_2}, R_2]$ and $[L_2^{1-\lambda_2}, L_2]$. Continuing this procedure, for every $i \in \mathbb{N}$ we choose R_i, L_i such that $L_{i-1} < R_i^{1-\lambda_i} < L_i^{(1-\lambda_i)^2}$, $R_i^{1-\lambda_i} > \max\{T(\delta_i), S(\lambda_i, L_1, \dots, L_{i-1}, \lambda_1, \dots, \lambda_{i-1}, \delta_1, \dots, \delta_{i-1})\}$, $\frac{1}{R_i^{\lambda_i}} < \theta$, and $L_i^{1-\lambda_i} > S(\lambda_i, R_1, \dots, R_i, \lambda_1, \dots, \lambda_i, \delta, \dots, \delta_i)$. We fix intervals $[R_i^{1-\lambda_i}, R_i]$ and $[L_i^{1-\lambda_i}, L_i]$. By our construction one has for every $i \in \mathbb{N}$

$$\frac{|M([R_j^{1-\lambda_j}, R_j]) \cap [L_i^{1-\lambda_i}, L_i]|}{L - L^{1-\lambda_i}} < \delta_j \text{ for all } j \le i \tag{40}$$

and analogously

$$\frac{|M([L_j^{1-\lambda_j}, L_j]) \cap [R_i^{1-\lambda_i}, R_i]|}{R_i - R_i^{1-\lambda_i}} < \delta_j \text{ for all } j \le i-1. \tag{41}$$

Now we introduce (disjoint) sets

$$A^* = \bigcup_{i=1}^{\infty} [R_i^{1-\lambda_i}, R_i], B^* = \bigcup_{i=1}^{\infty} [L_i^{1-\lambda_i}, L_i] \tag{42}$$

and consider the sets

$$A = A^* \setminus M(B^*), B = B^* \setminus M(A^*). \tag{43}$$

It is clear from this definition that $(A, B) \in \text{Cross}(\infty)$. The upper densities $\overline{\mathbf{d}}A$ and $\overline{\mathbf{d}}\,B$ are lower bounded now with the help of (40) and (41). For every $i \in \mathbb{N}$ the number of integers in A, which do not exceed R_i is at least

$$\begin{aligned}
&|[R_i^{1-\lambda_i}, R_i] \setminus (M(B^*) \cap [R_i^{1-\lambda_i}, R_i])| \\
&\ge (R_i - R_i^{1-\lambda_i}) - \sum_{j=1}^{i-1} |M([L_j^{1-\lambda_j}, L_j]) \cap [R_i^{1-\lambda_i}, R_i]| \\
&> (R_i - R_i^{1-\lambda_i}) - (R_i - R_i^{1-\lambda_i}) \cdot \sum_{j=1}^{i-1} \delta_j > (R_i - R_i^{1-\lambda_i})(1-\delta).
\end{aligned}$$

Therefore, for every $i \ge 1$, $\frac{|A \cap [1, R_i]|}{R_i} > \frac{R_i - R_i^{1-\lambda_i}}{R_i}(1-\delta) = \left(1 - \frac{1}{R_i^{\lambda_i}}\right)$ $(1-\delta) > (1-\theta)(1-\delta)$, because $\frac{1}{R_i^{\lambda_i}} < \theta$ for all $i \in \mathbb{N}$. Hence, $\overline{\mathbf{d}}\,A \ge (1-\theta)(1-\delta)$. Similarly $\overline{\mathbf{d}}\,B \ge (1-\theta)(1-\delta)$ and therefore

$$\overline{\mathbf{d}}\,A \cdot \overline{\mathbf{d}}\,B \ge (1-\theta)^2(1-\delta)^2.$$

The result follows, because θ and δ can be made arbitrarily small.

8. Concluding Remarks

The notion of cross-primitive pairs can be generalized to that of cross-primitive k-tuples of sets $(A_1, \dots, A_k)$. The understanding here is that any pair (A_i, A_j) $(i \neq j)$ is cross-primitive. Cross(x) then becomes Cross$_k(x)$. We guess that

1. $\lim\limits_{x \to \infty} \max\limits_{(A_1,\dots,A_k) \in \mathrm{Cross}_k(x)} \prod_{i=1}^{k} \frac{|A_i|}{x} = (\frac{1}{k})^k$
2. $\max\limits_{(A_1,\dots,A_k) \in \mathrm{Cross}_k(\infty)} \prod_{i=1}^{k} \underline{\mathbf{d}}\ A_i = (\frac{1}{k})^k (\frac{k-1}{k})^{k(k-1)}$
3. $\sup\limits_{(A_1,\dots,A_k) \in \mathrm{Cross}_k(\infty)} \prod_{i=1}^{k} \overline{\mathbf{d}}\ A_i = 1.$

References

1. E. Sperner, “Ein Satz über Untermengen einer endlichen Menge”, Math. Z. 27, 544–548, 1928.
2. A.S. Besicovitch, “On the density of certain sequences of integers”, Math. Annal. 110, 336–341, 1934.
3. F. Behrend, “On sequences of numbers not divisible one by another”, J. London Math. Soc. 10, 42–44, 1935.
4. P. Erdős, “On primitive abundant numbers”, J. London Math. Soc. 10, 49–58, 1935.
5. P. Erdős, “Note on sequences of integers no one of which is divisible by any other”, J. Lond. Math. Soc. 10, 136–128, 1935.
6. P. Erdős, “Generalization of a theorem of Besicovitch”, J. London Math. Soc. 11, 92–98, 1936.
7. H. Davenport and P. Erdős, “On sequences of positive integers”, Acta Arithm. 2,147–151,1937.
8. S. Pillai, “On numbers which are not multiples of any other in the set”, Proc. Indian Acad. Sci. A 10,392–394, 1939.
9. P. Erdős, “Integers with exactly k prime factors”, Ann. Math. II 49, 53–66, 1948.
10. F. Behrend, “Generalization of an inequality of Hulbrom and Rohrbach”, Bull. Ann. Math. Soc. 54, 681–684, 1948.
11. P. Erdős, “On the density of some sequences of integers”, Bull. Ann. Math. Soc. 54, 685–692, 1948.
12. N.G. De Bruijn, C. van E. Tengbergen, and D. Kruyswijk, “On the set of divisors of a number”, Nieuw Arch. f. Wisk. Ser II, 23, 191–193, 1949–51.
13. H. Davenport and P. Erdős, “On sequences of positive integers”, J. Indian Math. Soc. 15, 19–24, 1951.
14. P. Erdős, Aufgabe 395 in Elem. Math. Basel 16, 21, 1961.
15. S.D. Chowla and T. Vijayaraghavan, On the largest prime divisors of numbers, J. of the Indian Math. Soc. 11,31–37, 1947.
16. L.E. Dickson, “Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors”, American J. of Math. 35, 413–426, 1913.
17. H. Halberstam and K.F. Roth, “Sequences”, Oxford University Press, 1966, Springer Verlag, New York, Heidelberg, Berlin 1983.

18. K. Reidemeister, “Raum und Zahl”, Springer-Verlag, Berlin, Göttingen, Heidelberg 1957.
19. R.R. Hall and G. Tenenbaum, “Divisors”, Cambridge Tracts in Mathematics 90, Cambridge University Press, Cambridge, New York, 1988.
20. P. Erdős, O. Sárközy, and E. Szemerédi, “On an extremal problem concerning primitive sequences”, J. London Math. Soc. 42,484–488, 1967.
21. P. Erdős, O. Sárközy, and E. Szemerédi, “On a theorem of Behrend”, J. Australian Math. Soc. 7, 9–16, 1967.
22. P. Erdős, O. Sárközy, and E. Szemerédi, “On divisibility properties of sequences of integers”, Coll. Math. Soc. J. Bolyai 2, 35–49, 1970.
23. C. Pomerance and A. Sárközy, “On homogeneous multiplicative hybrid problems in number theory”, Acta Arith. 49, 291–302, 1988.
24. K. Yamamoto, “Logarithmic order of free distributive lattices”, J. Math. Soc. Japan 6, 343–353, 1954.
25. L.D. Meshalkin, “A generalization of Sperner’s theorem on the number of subsets of a finite set”, Theor. Probability Appl. 8, 203–204, 1963.
26. D. Lubell, “A short proof of Sperner’s theorem”, J. Combinatorial Theory 1, 299, 1966.
27. B. Bollobás, “On generalized graphs”, Acta Math. Acad. Sci. Hungar. 16, 447–452, 1965.
28. R. Ahlswede and Z. Zhang, “An identity in combinatorial extremal theory”, Adv. in Math., Vol. 80, No.2, 137–151, 1990.
29. R. Ahlswede and N. Cai, “A generalization of the AZ-identity”, Combinatorica 13 (3), 241–247, 1993.
30. R. Ahlswede and Z. Zhang, “On cloud-antichains and related configurations”, Discrete Mathematics 85, 225–245, 1990.
31. R. Ahlswede and L.H. Khachatrian, “On extremal sets without coprimes”, Acta Arithmetica, LXVI 1, 89–99, 1994.
32. R. Ahlswede and L.H. Khachatrian, “Towards characterising equality in correlation inequalities”, Preprint 93–027, SFB 343 “Diskrete Strukturen in der Mathematik”, Universität Bielefeld, to appear in European J. of Combinatorics.
33. R. Ahlswede and L.H. Khachatrian, “Optimal pairs of incomparable clouds in Multisets”, Preprint 93, SFB 343 “Diskrete Strukturen in der Mathematik”, Universität Bielefeld, to appear in Graphs and Combinatorics.
34. R. Ahlswede and L.H. Khachatrian, “The maximal length of cloud-antichains”, Discrete Mathematics, Vol. 131,9–15, 1994.
35. R. Ahlswede and L.H. Khachatrian, “Sharp bounds for cloud-antichains of length two”, Preprint 92–012, SFB 343 “Diskrete Strukturen in der Mathematik”, Universität Bielefeld.
36. H. Heilbronn, “On an inequality in the elementary theory of numbers”, Cambr. Phil. Soc. 33, 207–209, 1937.
37. H. Rohrbach, “Beweis einer zahlentheoretischen Ungleichung”, J. Reine u. Angew. Math. 177, 193–196, 1937.

Dense Difference Sets and Their Combinatorial Structure

Vitaly Bergelson*, Paul Erdős, Neil Hindman*, and Tomasz Łuczak

V. Bergelson (✉)
Department of Mathematics, Ohio State University, Columbus, OH 43210, USA
e-mail: vitaly@math.ohio-state.edu

P. Erdős
Mathematical Institute of the Hungarian Academy of Sciences, P.O. Box 127,
Realtanoda u. 13-15, H-1364 Budapest, Hungary

N. Hindman
Department of Mathematics, Howard University, Washington, DC 20059, USA
e-mail: nhindman@aol.com

T. Łuczak
Department of Discrete Mathematics, Faculty of Mathematics and CS,
Adam Mickiewicz University, Poznan, Poland
e-mail: tomasz@amu.edu.pl

Summary. We show that if a set B of positive integers has positive upper density, then its difference set $D(B)$ has extremely rich combinatorial structure, both additively and multiplicatively. If on the other hand only the density of $D(B)$ rather than B is assumed to be positive, one is not guaranteed any multiplicative structure at all and is guaranteed only a modest amount of additive structure.

1. Introduction

Given a subset B of the set $\mathbb{N}$ of positive integers, denote by $D(B)$ its "difference set". That is $D(B) = \{x - y : x, y \in B \text{ and } x > y\}$. We are concerned here with difference sets which are "large" in one of two senses. That is, we ask either that $\bar{d}(B) > 0$ or that $\bar{d}(D(B)) > 0$ where

$$\bar{d}(A) = \limsup_{n\to\infty} \big|A \cap \{1, 2, \ldots, n\}\big|/n.$$

We show in Sect. 2 that if $\bar{d}(B) > 0$, then $D(B)$ has an incredibly rich algebraic structure. We show for example that given any function $f : \mathbb{N} \longrightarrow \mathbb{N}$, there must exist a sequence $\langle x_n \rangle_{n=1}^{\infty}$ so that $\{\sum_{n\in F} a_n \cdot x_n : F$ is a finite nonempty subset of $\mathbb{N}$ and for each $n \in F, 1 \leq a_n \leq f(n)\} \cup \{\prod_{n\in F} x_n^{a_n} : F$ is a finite nonempty subset of $\mathbb{N}$ and for each $n \in F, 1 \leq a_n \leq f(n)\} \subseteq D(B)$.

* The first and third authors acknowledge support received from the National Science Foundation (USA) via grants DMS-9103056 and DMS-9025025 respectively. They also thank the US-Israel Binational Science Foundation for travel support.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_10,

With no sort of largeness assumptions at all (beyond the requirement that B should have at least three members) one must always be able to get some a and b with $\{a, b, a+b\} \subseteq D(B)$. (Given $x < y < z$ in B, let $a = y - x$ and $b = z - y$.) Infiniteness by itself doesn't help much. Indeed, it is easy to see that if $B = \{2^n : n \in \mathbb{N}\}$, then for no a, b, and c is $\{a, b, c, a+b, a+c, b+c, a+b+c\} \subseteq D(B)$. On the other hand, we show in Sect. 3 that if $\bar{d}(D(B)) > 0$, one can always find a, b, and c with $\{a, b, c, a+b, a+c, b+c, a+b+c\} \subseteq D(B)$.

We have not been able to determine whether $D(B)$ (where $\bar{d}(D(B)) > 0$) must contain some 4 elements with all of their sums. However, we do show in Sect. 3 that one can find sets B with $\bar{d}(D(B))$ arbitrarily close to $1/2$ such that $D(B)$ contains no five elements and all of their sums. We also show that we can find sets B with $\bar{d}(D(B))$ arbitrarily close to 1 such that $D(B)$ does not contain any $\{a, b, a \cdot b\}$.

2. The Difference Set of a Set of Positive Density

We show here that if $\bar{d}(B) > 0$, then $D(B)$ has a rich additive and multiplicative structure. Many of the results in this section are from the dissertation of the first author [2]. We begin by stating a well known result about sets of positive upper density, whose proof we leave as an exercise.

Lemma 1. *Let $A \subseteq \mathbb{N}$ such that $\bar{d}(A) > 0$ and let $k \in \mathbb{N}$ such that $1/k < \bar{d}(A)$. Then given any $t_1, t_2, \ldots, t_k$ in $\mathbb{N}$ there exist some $i < j$ in $\{1, 2, \ldots, k\}$ with $\bar{d}((A - t_i) \cap (A - t_j)) > 0$.*

Note by way of contrast that it is easy to get two disjoint sets both with upper density equal to 1. It is an immediate consequence of Lemma 1 that if $\bar{d}(B) > 0$, then $D(B)$ is an IP*-set. That is, given any sequence $\langle x_n \rangle_{n=1}^{\infty}$ in $\mathbb{N}$ there is some finite nonempty subset F of $\mathbb{N}$ such that $\sum_{n \in F} x_n \in D(B)$. (To see this: for each i, let $a_i = \sum_{n=1}^{i} x_n$ and pick $i < j$ such that $\bar{d}\big((B - a_i) \cap (B - a_j)\big) > 0$. Then $\sum_{n=i+1}^{j} x_n \in D(B)$.) Therefore, by Bergelson and Hindman [4, Theorem 2.6] there is some sequence $\langle x_n \rangle_{n=1}^{\infty}$ with $\{\sum_{n \in F} x_n : F$ is a finite nonempty subset of $\mathbb{N}\} \cup \{\prod_{n \in F} x_n : F$ is a finite nonempty subset of $\mathbb{N}\} \subseteq D(B)$. We show in Theorem 5 below that a stronger conclusion holds, (without invoking any results from [4]).

We shall utilize in our proofs two results from ergodic theory. The first of these is Furstenberg's famous correspondence principle which was first used in his proof of Szemerédi's Theorem [6, 7]. Recall that a *measure preserving system* is a quadruple $(X, \mathcal{B}, \mu, T)$ where X is a nonempty set, $\mathcal{B}$ is a σ-algebra of subsets of X, μ is a nonnegative σ-additive measure defined on $\mathcal{B}$ with $\mu(X) = 1$ (so that $(X, \mathcal{B}, \mu)$ is a probability measure space), and T is an invertible measure preserving transformation of X. (That is, T is continuous, and whenever $B \in \mathcal{B}$, $T^{-1}B \in \mathcal{B}$ and $\mu(T^{-1}B) = \mu(B)$.)

Theorem 1 (Bergelson [1, Proposition 2.1] and Furstenberg [6, Theorem 1.1]). *Let $B \subseteq \mathbb{N}$ with $\bar{d}(B) > 0$. There exist a measure preserving system $(X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ such that $\mu(A) = \bar{d}(B)$ and for all $n \in \mathbb{N}$, $\bar{d}(B \cap (B - n)) \geq \mu(A \cap T^n A)$.*

Given measure preserving systems $(X_1, \mathcal{B}_1, \mu_1, T_1)$ and $(X_2, \mathcal{B}_2, \mu_2, T_2)$ we follow standard practice and denote by $(X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_2, \mu_1 \times \mu_2, T_1 \times T_2)$ the system where $X_1 \times X_2$ is the cartesian product, $\mathcal{B}_1 \times \mathcal{B}_2$ is the σ-algebra generated by sets of the form $A_1 \times A_2$ for $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$, $\mu_1 \times \mu_2$ is the measure on $\mathcal{B}_1 \times \mathcal{B}_2$ determined by $(\mu_1 \times \mu_2)(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$ and $T_1 \times T_2$ is the measure preserving transformation defined by $(T_1 \times T_2)((x_1, x_2)) = (T_1(x_1), T_2(x_2))$.

Theorem 2. *Let $T_1, T_2, \ldots, T_k$ be invertible commuting transformations of a probability measure space $(X, \mathcal{B}, \mu)$. Assume that $p_1(n), p_2(n), \ldots, p_k(n)$ are polynomials with integer coefficients such that $p_i(0) = 0$ for $i \in \{1, 2, \ldots, k\}$. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then there exists $n \in \mathbb{N}$ such that $\mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \ldots T_k^{p_k(n)} A) > 0$.*

Proof. This is exactly [3, Theorem 4.2] except that the conclusion there has $n \in \mathbb{Z} \setminus \{0\}$. To derive this version we utilize the product space $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$. For $i \in \{1, 2, \ldots, k\}$, let $S_i = T_i \times \iota$, where ι is the identity. For $i \in \{k+1, k+2, \ldots, 2k\}$, let $S_i = \iota \times T_{i-k}$ and let $p_i(n) = p_{i-k}(-n)$. Then $S_1, S_2, \ldots, S_{2k}$ are invertible commuting transformations of $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu)$ and $(\mu \times \mu)(A \times A) > 0$ so pick (using [3, Theorem 4.2]) $n \in \mathbb{Z} \setminus \{0\}$ such that $(\mu \times \mu)((A \times A) \cap S_1^{p_1(n)} S_2^{p_2(n)} \ldots S_{2k}^{p_2k(n)}(A \times A)) > 0$. If $n > 0$ we see from the first coordinate that $\mu(A \cap T_1^{p_1(n)} T_2^{p_2(n)} \ldots T_k^{p_k(n)} A) > 0$. If $n < 0$ we see from the second coordinate that $\mu(A \cap T_1^{p_1(-n)} T_2^{p_2(-n)} \ldots T_k^{p_k(-n)} A) > 0$. □

We shall see in Theorem 5 that whenever $\bar{d}(B) > 0$, $D(B)$ contains sums and products from a sequence where terms are allowed to repeat a restricted number of times. We present first a special case so we may introduce the proof in a relatively uncomplicated setting.

Theorem 3. *Let $B \subseteq \mathbb{N}$ with $\bar{d}(B) > 0$. Then there is some sequence $\langle x_n \rangle_{n=1}^{\infty}$ such that $\{\sum_{n \in F} a_n x_n : F$ is a finite nonempty subset of $\mathbb{N}$ and for each $n \in F$, $a_n \in \{1, 2\}\} \cup \{\prod_{n \in F} x_n^{a_n} : F$ is a finite nonempty subset of $\mathbb{N}$ and for each $n \in F$, $a_n \in \{1, 2\}\} \subseteq D(B)$.*

Proof. Pick by Theorem 1 a measure preserving system $(X, \mathcal{B}, \mu, T)$ and some $A \in \mathcal{B}$ such that $\mu(A) = \bar{d}(B)$ and for each $n \in \mathbb{N}$, $\bar{d}(B \cap (B - n)) \geq \mu(A \cap T^n A)$. Observe that $\{n \in \mathbb{N} : \mu(A \cap T^n A) > 0\} \subseteq D(B)$. For $m \in \mathbb{N}$ and a sequence $\langle x_n \rangle_{n=1}^{m}$ in $\mathbb{N}$ let $E(\langle x_n \rangle_{n=1}^{m}) = \{\sum_{n \in F} a_n x_n : F$ is a nonempty subset of $\{1, 2, \ldots, m\}$ and for each $n \in F, a_n \in \{1, 2\}\}$ and let $C(\langle x_n \rangle_{n=1}^{m}) = \{\prod_{n \in F} x_n^{a_n} : F$ is a nonempty subset of $\{1, 2, \ldots, m\}$ and for each $n \in F, a_n \in \{1,2\}\}$. We construct a sequence $\langle x_n \rangle_{n=1}^{\infty}$ by induction so that for each m, $E(\langle x_n \rangle_{n=1}^{m} \cup C(\langle x_n \rangle_{n=1}^{m}) \subseteq \{n \in \mathbb{N} : \mu(A \cap T^n A) > 0\}$ which will suffice by our observation.

To ground the induction consider the measure space $(X \times X \times X, \mathcal{B} \times \mathcal{B} \times \mathcal{B}, \mu \times \mu \times \mu)$, let $S_1 = (T \times \iota \times \iota), S_2 = (\iota \times T \times \iota), S_3 = (\iota \times \iota \times T)$, $p_1(n) = n$, $p_2(n) = 2n$, and $p_3(n) = n^2$. (Recall that ι is the identity.) Pick by Theorem 2 some $x_1 \in \mathbb{N}$ such that $(\mu \times \mu \times \mu)((A \times A \times A) \cap S_1^{p_1(x_1)} S_2^{p_2(x_1)} S_3^{p_3(x_1)}(A \times A \times A)) > 0$. From the first coordinate we see that $\mu(A \cap T_1^{x_1} A) > 0$, from the second coordinate we see that $\mu(A \cap T_1^{2x_1} A) > 0$, and from the third coordinate we see that $\mu(A \cap T_1^{x_1^2} A) > 0$. Since $E(\langle x_n \rangle_{n=1}^{1}) = \{x_1, 2x_1\}$ and $C(\langle x_n \rangle_{n=1}^{1}) = \{x_1, x_1^2\}$, the grounding is complete.

Now let $m \in \mathbb{N}$ be given and assume we have chosen $\langle x_n \rangle_{n=1}^{m-1}$ with $E((x_n)_{n=1}^{m-1}) \cup C(\langle x_n \rangle_{n=1}^{m-1}) \subseteq \{n \in \mathbb{N} : \mu(A \cap T^n A) > 0\}$. Let $b = 3^{m-1}$ and enumerate (with repetitions if need be) $\{0\} \cup E(\langle x_n \rangle_{n=1}^{m-1})$ as $\langle y_j \rangle_{j=1}^{b}$ and enumerate $\{1\} \cup C(\langle x_n \rangle_{n=1}^{m-1})$ as $\langle z_j \rangle_{j=1}^{b}$. Now consider the measure space $(\bigtimes_{j=1}^{4b} X, \bigtimes_{j=1}^{4b} \mathcal{B}, \bigtimes_{j=1}^{4b} \mu)$. Let $H = \bigtimes_{j=1}^{b}((A \cap T^{y_j} A) \times (A \cap T^{y_j} A) \times A \times A)$, let $\bar{\mu} = \bigtimes_{j=1}^{4b} \mu$, and note that $\bar{\mu}(H) > 0$. (Our induction hypothesis tells us that each $\mu(A \cap T^{y_j} A) > 0$.) Let $S_1 = \bigtimes_{j=1}^{b}(T \times \iota \times \iota \times \iota)$, $S_2 = \bigtimes_{j=1}^{b}(\iota \times T \times \iota \times \iota)$, $S_3 = \bigtimes_{j=1}^{b}(\iota \times \iota \times T^{z_j} \times \iota)$, and $S_4 = \bigtimes_{j=1}^{b}(\iota \times \iota \times \iota \times T^{z_j})$. Let $p_1(n) = n$, $p_2(n) = 2n$, $p_3(n) = n$, and $p_4(n) = n^2$. Pick by Theorem 1, some $x_m \in \mathbb{N}$ such that $\bar{\mu}(H \cap S_1^{p_1(x_m)} S_2^{p_2(x_m)} S_3^{p_3(x_m)} S_4^{p_4(x_m)} H) > 0$.

To see that $E(\langle x_n \rangle_{n=1}^{m}) \subseteq \{n \in \mathbb{N} : \mu(A \cap T^n A) > 0\}$, let $\emptyset \neq F \subseteq \{1, 2, \ldots, m\}$ and for each $n \in F$, let $a_n \in \{1,2\}$. If $m \notin F$, then $\sum_{n \in F} a_n x_n \in E(\langle x_n \rangle)_{n=1}^{m-1})$, so we assume $m \in F$. Pick $j \in \{1, 2, \ldots, b\}$ such that $\sum_{n \in F} a_n x_n = y_j + a_m x_m$. If $a_m = 1$, we see by looking at coordinate $4j - 3$ that $\mu(A \cap T^{y_j} A \cap T^{x_m}(A \cap T^{y_j} A)) > 0$; in particular $\mu(A \cap T^{y_j + x_m} A) > 0$. If $a_m = 2$, we see by looking at coordinate $4j - 2$ that $\mu(A \cap T^{y_j} A \cap T^{2x_m}(A \cap T^{y_j} A)) > 0$; in particular $\mu(A \cap T^{y_j + 2x_m} A) > 0$.

To see that $C(\langle x_n \rangle_{n=1}^{m}) \subseteq \{n \in \mathbb{N} : \mu(A \cap T^n A) > 0\}$, let $\emptyset \neq F \subseteq \{1, 2, \ldots, m\}$ and for each $n \in F$, let $a_n \in \{1,2\}$. If $m \notin F$, then $\prod_{n \in F} x_n^{a_n} \in C(\langle x_n \rangle_{n=1}^{m-1})$, so we assume $m \in F$. Pick $j \in \{1, 2, \ldots, b\}$ such that $\prod_{n \in F} x_n^{a_n} = z_j \cdot x_z^{a_m}$. If $a_m = 1$, we see by looking at coordinate $4b - 1$ that $\mu(A \cap (T^{z_j})^{x_m} A) > 0$ so that $\mu(A \cap T^{z_j x_m} A) > 0$. If $a_m = 2$, we see by looking at coordinate $4b$ that $\mu(A \cap (T^{z_j})^{x_m^2} A) > 0$ so that $\mu(A \cap T^{z_j x_m^2 A}) > 0$. □

We observe in fact that if one has sets $B_1, B_2, \ldots, B_n$ with each $\bar{d}(B_i) > 0$, then the conclusion of Theorem 3 applies to $\bigcap_{i=1}^{n} D(B_i)$. To see this one simply starts with the product system $(\bigtimes_{i=1}^{n} X_i, \bigtimes_{i=1}^{n} \mathcal{B}_i, \bigtimes_{i=1}^{n} \mu_i, \bigtimes_{i=1}^{n} T_i)$ where $(X_i, \mathcal{B}_i, \mu_i, T_i)$ is the system given by Theorem 1 for B_i.

Recall that a set $B \subseteq \mathbb{N}$ is an IP* set if and only if whenever $\langle x_n \rangle_{n=1}^{\infty}$ is a sequence in $\mathbb{N}$, one has $FS(\langle x_n \rangle_{n=1}^{\infty}) \cap B \neq \emptyset$. We pause now to observe that neither of the conclusions of Theorem 3 follow from the fact that $D(B)$ is an IP* set.

Theorem 4. *There is an* IP^* *set* A *such that for no sequence* $\langle x_n\rangle_{n=1}^{\infty}$ *is* $\{\sum_{n\in F} a_n x_n$: F *is a finite nonempty subset of* $\mathbb{N}$ *and for each* $n\in\mathbb{N}$, $a_n\in\{1,2\}\}\subseteq A$ *and for no sequence* $\langle y_n\rangle_{n=1}^{\infty}$ *is* $\{\prod_{n\in F} y_n^{a_n}$: F *is a finite nonempty subset of* $\mathbb{N}$ *and for each* $n\in\mathbb{N}, a_n\in\{1,2\}\}\subseteq A$.

Proof. Let $B=\mathbb{N}\setminus\{x^2 : x\in\mathbb{N}\}$. Since one clearly cannot get any sequence $\langle x_n\rangle_{n=1}^{\infty}$ with $\{\sum_{n\in F} x_n : F$ is a finite nonempty subset of $\mathbb{N}\}\subseteq\{x^2 : x\in\mathbb{N}\}$, one has that B is an IP^* set. And no sequence $\langle y_n\rangle_{n=1}^{\infty}$ has any $y_n^2\in B$.

Now by Deuber et al. [5, Theorem 3.14], there is a partition $\mathbb{N}=C_1\cup C_2$ such that for no sequence $\langle x_n\rangle_{n=1}^{\infty}$ is $\{\sum_{n\in F} x_n : F$ is a finite nonempty subset of $\mathbb{N}\}\subseteq C_1$ and for no sequence $\langle y_n\rangle_{n=1}^{\infty}$ is $\{\sum_{n\in F_1} y_n+\sum_{n\in F_2} 2y_n : F_1$, and F_2 are finite nonempty subsets of $\mathbb{N}$ and $\max F_1<\min F_2\}\subseteq C_2$. Then C_2 is an IP^* set. Let $A=B\cap C_2$. Since the intersection of two IP^* sets is again an IP^* set (see [4]), we have that A is as required. □

The next theorem is our major result of this section. Considerably stronger statements are in fact available with the same proof. However, we are trying to keep the results easily comprehensible.

Theorem 5. *Let* $B\subseteq\mathbb{N}$ *with* $\bar{d}(B)>0$ *and let* $f:\mathbb{N}\to\mathbb{N}$. *Then there is some sequence* $\langle x_n\rangle_{n=1}^{\infty}$ *such that* $\{\sum_{n\in F} a_n x_n$: F *is a finite nonempty subset of* $\mathbb{N}$ *and for each* $n\in F, a_n\in\{1,2,\ldots,f(n)\}\}\cup\{\prod_{n\in F} x_n^{a_n}$: F *is a finite nonempty subset of* $\mathbb{N}$ *and for each* $n\in F$, $a_n\in\{1,2,\ldots,f(n))\}\subseteq D(B)$.

Proof. We describe how to modify the proof of Theorem 3. First define $E(\langle x_n\rangle_{n=1}^{m})$ and $C(\langle x_n\rangle_{n=1}^{m})$ analogously. At the grounding level one takes the measure space $(\times_{i=1}^{2f(1)-1}X,\times_{i=1}^{2f(1)-1}\mathcal{B},\times_{i=1}^{2f(1)-1}\mu)$. One lets $P_i(n)=i\cdot n$ for $i\in\{1,2,\ldots,f(1)\}$ and lets $p_i(n)=n^{i-f(1)-1}$ for $i\in\{f(1)+1,f(1)+2,\ldots,2f(1)-1\}$.

At the induction stage, one lets $b=\prod_{i=1}^{m-1}(f(i)+1)$ and enumerates $E(\langle x_n\rangle_{n=1}^{m-1})\cup\{0\}$ as $\langle y_i\rangle_{j=1}^{b}$ and enumerates $\{1\}\cup C(\langle x_n\rangle_{n=1}^{m-1})$ as $\langle z_j\rangle_{j=1}^{b}$. Then one uses the measure space $(\times_{j=1}^{b\cdot 2\cdot f(m)}X,\times_{j=1}^{b\cdot 2\cdot f(m)}\mathcal{B},\times_{j=1}^{b\cdot 2\cdot f(m)}\mu)$, and lets $H=\times_{j=1}^{b}(\times_{i=1}^{f(m)}(A\cap T^{y_j}A)\times\times_{i=1}^{f(m)}A)$. Using the obvious definitions of $S_1,S_2,\ldots,S_{2\cdot f(m)}$ and $p_1,p_2,\ldots,p_{2\cdot f(m)}$ one completes the proof. □

3. Additive Structure in Dense Difference Sets

For the remainder of the paper we look at difference sets $D(B)$ where we no longer require that $\bar{d}(B)>0$, but only that $\bar{d}(D(B))>0$.

Because difference sets are defined additively one would not necessarily expect them to have any multiplicative structure. On the other hand, Theorem 5 might make one suspect that they would have some multiplicative structure. We begin this section by showing that they need not.

Theorem 6. *Let $\epsilon > 0$. There is a set B such that $\bar{d}(D(B)) > 1 - \epsilon$ and there do not exist a and b in $\mathbb{N}$ with $\{a, b, a \cdot b\} \subseteq D(B)$.*

Proof. Pick $\alpha \in \mathbb{N}$ such that $1/2^\alpha < \epsilon$. Define a sequence $\langle f(r)\rangle_{r=0}^\infty$ by $f(0) = 2+\alpha$ and $f(r+1) = 2(f(r)+\alpha)+1$. Let $\langle x_n\rangle_{n=1}^\infty$ enumerate $\bigcup_{r=0}^\infty \{2^{f(r)}, 2^{f(r)}+1, \ldots, 2^{f(r)+\alpha}-1\}$ in increasing order and note that for all n in $\mathbb{N}, x_n < 2^{f(n)}$. Let $B = \{2^{f(n)} : n \in \mathbb{N}\} \cup \{2^{f(n)} + x_n : n \in \mathbb{N}\}$. Then $D(B) = \{x_n : n \in \mathbb{N}\} \cup \{2^{f(n)} + x_n - 2^{f(m)} : m, n \in \mathbb{N}$ and $m < n\} \cup \{2^{f(n)} - 2^{f(m)} : m, n \in \mathbb{N}$ and $m < n\} \cup \{2^{f(n)} + x_n - 2^{f(m)} - x_m : m, n \in \mathbb{N}$ and $m < n\} \cup \{2^{f(n)} - 2^{f(m)} - x_m : m, n \in \mathbb{N}$ and $m < n\}$. Now given any $r \in \mathbb{N}$ we have $|\{x_n : n \in \mathbb{N}\} \cap \{1, 2, \ldots, 2^{f(r)+\alpha}\}| > 2^{f(r)+\alpha} - 2^{f(r)}$ so $\bar{d}(D(B)) \geq 1 - 1/2^\alpha > 1 - \epsilon$.

If $a = x_n$, then for some $r \in \mathbb{N} \cup \{0\}$, we have $2^{f(r)} \leq a < 2^{f(r)+\alpha}$. If $a \in D(B) \setminus \{x_n : n \in \mathbb{N}\}$ then there exist m and n in $\mathbb{N}$ with $m < n$ such that $2^{f(n)} - 2^{f(m)} - x_m \leq a \leq 2^{f(n)} + x_n - 2^{f(n)}$. Since $2^{f(n)} - 2^{f(m)} - x_m > 2^{f(n)-1}$ we conclude that for any $a \in D(B)$ there is some $n \in \mathbb{N} \cup \{0\}$ with $2^{f(n)-1} < a < 2^{f(n)+\alpha}$. Now suppose we have $a \leq b$ in $D(B)$ such that $a \cdot b \in D(B)$. Pick $m \leq n \leq r$ in $\mathbb{N} \cup \{0\}$ such that $2^{f(m)-1} < a < 2^{f(m)+\alpha}$, $2^{f(n)-1} < b < 2^{f(n)+\alpha}$, and $2^{f(r)-1} < a \cdot b < 2^{f(r)+\alpha}$. If $n < r$ we have $2^{f(n)-1} < a \cdot b < 2^{f(m)+f(n)+2\alpha}$ so $f(r) \leq f(m) + f(n) + 2\alpha \leq 2f(n) + 2\alpha < f(r)$, a contradiction. Thus $n = r$ so that $2^{f(m)+f(n)-2} < a \cdot b < 2^{f(r)+\alpha} = 2^{f(n)+\alpha}$. Then $f(m) < \alpha + 2 = f(0) \leq f(m)$, a contradiction. □

Theorem 7. *Let $B \subseteq \mathbb{N}$ and assume $\bar{d}(D(B)) > 0$. There exist a, b, c in $\mathbb{N}$ such that $\{a, b, c, a+b, a+c, b+c, a+b+c\} \subseteq D(B)$.*

Proof. If $\bar{d}(B) > 0$ we are done by Theorem 5 so we assume $\bar{d}(B) = 0$. Enumerate B in order as $\langle x_n\rangle_{n=1}^\infty$. The result of this theorem is almost free. That is given any $r > s > n > k$, if we let $a = x_r - x_s$, $b = x_s - x_n$, and $c = x_n - x_k$, then $a + b = x_r - x_n$, $b + c = x_s - x_k$, and $a + b + c = x_r - x_k$. The only problem then is to find $r > s > n > k$ such that $x_r - x_s + x_n - x_k \in D(B)$.

Let $\alpha = \bar{d}(D(B))$ and pick $l \in \mathbb{N}$ such that $1/l < \alpha$. For each t we have $\bar{d}(B - t) = \bar{d}(B) = 0$. Let $E = D(B) \bigcup_{k=1}^l (B - x_k)$. Then $E = \{x_r - x_s : r, s \in \mathbb{N}$ and $r > s > l\}$ and $\bar{d}(E) = \alpha$. Pick by Lemma 1 some $k < n \leq l$ such that $\bar{d}((E - x_k) \cap (E - x_n)) > 0$. In particular $(E - x_k) \cap (E - x_n) \neq \emptyset$ so pick $r > s > l$ and $t > m > l$ such that $x_r - x_s - x_k = x_t - x_m - x_n$. Then $r > s > l \geq n > k$ and $x_r - x_s + x_n - x_k = x_t - x_m$ as required. □

We now set out to show that we can produce sets B with $\bar{d}(D(B))$ arbitrarily close to $1/2$ such that $D(B)$ does not contain $FS(\langle a_n\rangle_{n=1}^5)$ for any a_1, a_2, a_3, a_4, a_5. (Here $FS(\langle a_n\rangle_{n=1}^m) = \{\sum_{n \in F} a_n : \emptyset \neq F \subseteq \{1, 2, \ldots, m\}\}$.) We first introduce the sets B (whose dependence on α is suppressed).

Definition 1. *Fix $\alpha \in \mathbb{N}$ with $\alpha > 4$. Let $\langle x_n\rangle_{n=1}^\infty$ enumerate in increasing order* $(\mathbb{N}2 + 1) \cap (\bigcup_{t=0}^\infty \{2^{\alpha t+2}, 2^{\alpha t+2} + 1, \ldots, 2^{\alpha t+\alpha-2} - 1\})$. *Let $B = \{2^{\alpha n} : n \in \mathbb{N}\} \cup \{2^{\alpha n} + x_n : n \in \mathbb{N}\}$.*

One sees immediately that one can get a_1, a_2, a_3, and a_4 with $FS(\langle a_n\rangle_{n=1}^4) \subseteq D(B)$. Indeed let $s < m$ be given, pick l and r such that $2^{\alpha r+2} < x_l < x_l + 2^{\alpha m} - 2^{\alpha s} < 2^{\alpha r+\alpha-2}$, let $x_k = x_l + 2^{\alpha m} - 2^{\alpha s}$, and pick v and t such that $2^{\alpha t+2} < x_v < x_v + 2^{\alpha k} - 2^{\alpha s} < 2^{\alpha t+\alpha-2}$. Then let $a_1 = 2^{\alpha m} - 2^{\alpha s} = x_k - x_l$, $a_2 = 2^{\alpha l} - 2^{\alpha m}$, $a_3 = 2^{\alpha k} - 2^{\alpha l}$, and $a_4 = x_v$. Then $FS(\langle a_n\rangle_{n=1}^4) \subseteq D(B)$. In fact, one can show that any sequence of length 4 with its sums contained in $D(B)$ must fit this description. The computations are longer and more painful than those on which we are embarking, so we omit them.

Definition 2. *Let α and $\langle x_n\rangle_{n=1}^\infty$ be as in Definition 1. Then $A_1 = \{x_n : n \in \mathbb{N}\}$, $A_2 = \{2^{\alpha n} + x_n - 2^{\alpha m} : n, m \in \mathbb{N}$ and $m < n\}$, $A_3 = \{2^{\alpha n} - 2^{\alpha m} - x_m : n, m \in \mathbb{N}$ and $m < n\}$, $A_4 = \{2^{\alpha n} - 2^{\alpha m} : n, m \in \mathbb{N}$ and $m < n\}$, and $A_5 = \{2^{\alpha n} + x_n - 2^{\alpha m} - x_m : n, m \in \mathbb{N}$ and $m < n\}$.*

Observe that $D(B) = \bigcup_{i=1}^5 A_i$.

We next prove two lemmas to aid in our computations.

Lemma 2. *Let $n_1, n_2, m_1, m_2 \in \mathbb{N}$ and let γ_1, γ_2, δ_1, $\delta_2 \in \{0, 1\}$ with $n_2 \geq n_1$ and $m_2 \geq m_1$. If $2^{\alpha n_2} + 2^{\alpha n_1} + \gamma_2 x_{n_2} + \gamma_1 x_{n_1} = 2^{\alpha m_2} + 2^{\alpha m_1} + \delta_2 x_{m_2} + \delta_1 x_{m_1}$, then*

(1) $(n_2, n_1, \gamma_2, \gamma_1) = (m_2, m_1, \delta_2, \delta_1)$ or
(2) $n_2 = n_1$ and $(n_2, n_1, \gamma_2, \gamma_1) = (m_2, m_1, \delta_1, \delta_2)$.

Proof. We assume without loss of generality that $n_2 \geq m_2$. If we had $n_2 > m_2$ we would have $2^{\alpha m_2} + 2^{\alpha m_1} + \delta_2 x_{m_2} + \delta_1 x_{m_1} < 4 \cdot 2^{\alpha m_2} = 2^{\alpha m_2+2} < 2^{\alpha n_2} < 2^{\alpha n_2} + 2^{\alpha n_1} + \gamma_2 x_{n_2} + \gamma_1 x_{n_1}$, a contradiction. Thus $n_2 = m_2$. Assume first that $\gamma_2 \neq \delta_2$ and assume without loss of generality that $\gamma_2 = 1$ and $\delta_2 = 0$, Then $x_{n_2} = 2^{\alpha m_1} - 2^{\alpha n_1} + \delta_1 x_{m_1} - \gamma_1 x_{n_1}$. We claim $m_1 = n_1$. If we had $m_1 < n_1$ we would have $x_{n_2} \leq 2^{\alpha m_1} - 2^{\alpha n_1} + \delta_1 x_{m_1} < 2 \cdot 2^{\alpha m_1} - 2^{\alpha n_1} < 0$. Suppose now $m_1 > n_1$. Then $x_{n_2} < 2^{\alpha m_1} + \delta_1 x_{m_1} < 2^{\alpha m_1+1}$ and $x_{n_2} \geq 2^{\alpha m_1} - 2^{\alpha n_1} - \gamma_1 x_{n_1} > 2^{\alpha m_1} - 2 \cdot 2^{\alpha n_1} > 2^{\alpha m_1 - 1}$. But for some r we have $2^{\alpha r+2} < x_{n_2} < 2^{\alpha r+\alpha-2}$, a contradiction. Thus $m_1 = n_1$ so $x_{n_2} = (\delta_1 - \gamma_1) \cdot x_{n_1}$ and hence $\delta_1 = 1, \gamma_1 = 0$ and $n_1 = n_2$ so that conclusion (2) holds.

Now assume $\gamma_2 = \delta_2$. Then we have $2^{\alpha m_1} + \gamma_1 n_1 = 2^{\alpha m_1} + \delta_1 m_1$. As in the first paragraph we see $n_1 = m_1$ so $\gamma_1 n_1 = \delta_1 n_1$ so $\gamma_1 = \delta_1$. □

Lemma 3. *Let $n_1, n_2, n_3, m_1, m_2, m_3 \in \mathbb{N}$ and let γ_1, γ_2, γ_3, δ_1, δ_2, $\delta_3 \in \{0, 1\}$ with $n_3 \geq n_2 \geq n_1$ and $m_3 \geq m_2 \geq m_1$. Assume $2^{\alpha n_3} + 2^{\alpha n_2} + 2^{\alpha n_1} + \gamma_3 x_{n_3} + \gamma_2 x_{n_2} + \gamma_1 x_{n_1} = 2^{\alpha m_3} + 2^{\alpha m_2} + 2^{\alpha} m_1 + \delta_3 x_{m_3} + \delta_2 x_{m_2} + \delta_1 x_{m_1}$. Then some one of the following conclusions holds. In any event we have $\gamma_1 + \gamma_2 + \gamma_3 = \delta_1 + \delta_2 + \delta_3$ and $\max\{n_1, n_2, n_3\} = \max\{m_1, m_2, m_3\}$.*

(1) $(n_3, n_2, n_1, \gamma_3, \gamma_2, \gamma_1) = (m_3, m_2, m_1, \delta_3, \delta_2, \delta_1)$
(2) $n_2 = n_1$ and $(n_3, n_2, n_1, \gamma_3, \gamma_2, \gamma_1) = (m_3, m_2, m_1, \delta_3, \delta_2, \delta_1)$
(3) $n_3 = n_2$ and $(n_3, n_2, n_1, \gamma_3, \gamma_2, \gamma_1) = (m_3, m_2, m_1, \delta_2, \delta_3, \delta_1)$
(4) $n_3 = n_2 = n_1$ and $(n_3, n_2, n_1, \gamma_3, \gamma_2, \gamma_1) = (m_3, m_2, m_1, \delta_1, \delta_2, \delta_3)$
(5) $(n_3, n_2, \gamma_3, \gamma_2, \gamma_1) = (m_3, m_2, \delta_3, \delta_2, \delta_1)$ and $\gamma_3 \neq \gamma_2$ and $n_1 \neq m_1$.

Proof. We assume without loss of generality that $n_3 \geq m_3$, If we had $n_3 > m_3$ we would have $2^{\alpha m_3} + 2^{\alpha m_2} + 2^{\alpha m_1} + \delta_3 x_{m_3} + \delta_2 x_{m_2} + \delta_1 x_{m_1} < 6 \cdot 2^{\alpha m_3} < 2^{\alpha n_3} <^{\alpha n_3} + 2^{\alpha n_2} + 2^{\alpha n_1} + \gamma_3 x_{n_3} + \gamma_2 x_{n_2} + \gamma_1 x_{n_1}$, a contradiction. Thus we must have $n_3 = m_3$. If also $\gamma_3 = \delta_3$ we have $2^{\alpha n_2} + 2^{\alpha n_1} + \gamma_2 x_{n_2} + \gamma_1 x_{n_1} = 2^{\alpha m_2} + 2^{\alpha m_1} \delta_2 x_{m_2} + \delta_1 x_{m_1}$ so Lemma 2 applies and yields conclusion (1) or conclusion (2).

Thus we assume $\gamma_3 \neq \delta_3$ and assume without loss of generality that $\gamma_3 = 1$ and $\delta_3 = 0$. Then $x_{n_3} = 2^{\alpha m_2} - 2^{\alpha n_2} + 2^{\alpha m_1} - 2^{\alpha n_1} + \delta_2 x m_2 - \gamma_2 x_{n_2} + \delta_1 x m_1 - \gamma_1 x_{n_1}$. We observe that if we had $m_2 < n_2$ we would have $x_{n_3} < 4 \cdot 2^{\alpha m_2} - 2^{\alpha n_2} < 0$.

Consequently $m_2 \geq n_2$. We claim in fact $m_2 = n_2$ so suppose instead that $m_2 > n_2$. Then $x_{n_3} < 4 \cdot 2^{\alpha m_2} = 2^{\alpha m_2 + 2}$ and $x_{n_3} > 2^{\alpha m_2} - 4 \cdot 2^{\alpha n_2} > 2^{\alpha m_2 - 1}$. But there is some $r \in \mathbb{N}$ such that $2^{\alpha r + 2} < x_{n_3} < 2^{\alpha r + \alpha - 2}$, a contradiction. Thus $m_2 = n_2$ as claimed. Consequently we have $x_{n_3} = 2^{\alpha m_1} - 2^{\alpha n_1} + (\delta_2 - \gamma_2) x_{n_2} + \delta_1 x_{m_1} - \gamma_1 x_{n_1}$.

Case 1. $\delta_2 = \gamma_2$. Then we have $x_{n_3} = 2^{\alpha m_1} - 2^{\alpha n_1} + \delta_2 x_{m_1} - \gamma_1 x_{n_1}$. Reasoning as above we conclude $m_1 = n_1$. Then $x_{n_3} = (\delta_1 - \gamma_1) \cdot x_{n_1}$ so $\delta_1 = 1$, $\gamma_1 = 0$, and $n_3 = n_1$. Then conclusion (4) holds.

Case 2. $\delta_2 \neq \gamma_2$. We claim that we must have $\delta_2 = 1$ and $\gamma_2 = 0$. To see this suppose instead $\delta_2 = 0$ and $\gamma_2 = 1$. Then $x_{n_3} = 2^{\alpha m_1} - 2^{\alpha n_1} - x_{n_2} + \delta_1 x_{m_1} - \gamma_1 x_{n_1}$. One cannot have $n_1 > m_1$ for then one would have $x_{n_3} < 0$. If we had $n_1 = m_1$ we would have $x_{n_3} = -x_{n_2} + (\delta_1 - \gamma_1) x_{n_1}$. Since $x_{n_3} > 0$ one would have to have $\delta_1 = 1$ and $\gamma_1 = 0$. But then one would have $x_{n_3} + x_{n_2} = x_{n_1}$ forcing x_{n_1} to be even. Thus one must have $n_1 < m_1$.

Now we claim that $x_{n_2} > 2^{\alpha m_1 - 1}$. Suppose instead that $x_{n_2} < 2^{\alpha m_1 - 1}$. Now $x_{n_3} < 2 \cdot 2^{\alpha m_1}$ and for some $r 2^{\alpha r + 2} < x_{n_3} < 2^{\alpha r + \alpha - 2}$ so $x_{n_3} < 2^{\alpha m_1 - 2}$. That is $2^{\alpha m_1} - 2^{\alpha n_1} - x_{n_2} + \delta_1 x_{m_1} - \gamma_1 x_{n_1} < 2^{\alpha m_1 - 2}$ so $2^{\alpha m_1} + \delta_1 x_{m_1} < 2^{\alpha m_1 - 2} + 2^{\alpha n_1} + x_{n_2} + \gamma_1 x_{n_1} < 2^{\alpha m_1 - 2} + 2^{\alpha m_1 - 2} + 2^{\alpha m_1 - 1} = 2^{\alpha m_1}$, a contradiction. Thus we have $x_{n_2} > 2^{\alpha m_1 - 1}$.

But now for some s we have $2^{\alpha s + 2} < x_{n_2} < 2^{\alpha s + \alpha - 2}$ so $x_{n_2} > 2^{\alpha m_1 + 2}$. But then we have $x_{n_3} = 2^{\alpha m_1} - 2^{\alpha n_1} - x_{n_2} + \delta_1 x_{m_1} - \gamma_1 x_{n_1} < 2^{\alpha m_1} + \delta_1 x_{m_1} - 2^{\alpha m_1 + 2} < 0$, a contradiction. Thus we have established that $\delta_2 = 1$ and $\gamma_2 = 0$.

Then we have that $x_{n_3} = 2^{\alpha m_1} - 2^{\alpha n_1} + x_{m_2} + \delta_1 x_{m_1} - \gamma_1 x_{n_1}$. Since $x_{n_3}, x_{m_2}, x_{m_1}$, and x_{n_1} are all odd we conclude $\delta_1 = \gamma_1$. If also $m_1 = n_1$ we conclude that $x_{n_3} = x_{m_2}$ so $n_3 = m_2 = n_2$ and conclusion (3) holds. Thus we assume $m_1 \neq n_1$. In this case conclusion (5) holds. □

We now begin an embarrassingly long sequence of computational lemmas.

Lemma 4. *If $a, b \in A_1 \cup A_2$ then $a + b \notin D(B)$.*

Proof. Suppose $a, b \in A_1 \cup A_2$ and $a+b \in D(B)$. Then $a+b$ is even so $a+b \in A_4 \cup A_5$. Pick $s < r$ and $\delta \in \{0,1\}$ such that $a+b = 2^{\alpha r} - 2^{\alpha s} + \delta(x_r - x_s)$. We consider three cases.

Case 1. $a, b \in A_1$. Pick $n, m \in \mathbb{N}$ such that $a = x_n$ and $b = x_m$. Then $x_n + x_m + 2^{\alpha s} + \delta x_s = 2^{\alpha r} + \delta x_r$ so adding $2^{\alpha n} + 2^{\alpha m}$ to both sides we get by Lemma 3 that $1 + 1 + \delta = \delta$, a contradiction.

Case 2. $a, b \in A_2$. Pick $m < n$ and $l < k$ such that $a = 2^{\alpha n} + x_n - 2^{\alpha m}$ and $b = 2^{\alpha k} + x_k - 2^{\alpha l}$. Then $2^{\alpha n} + x_n - 2^{\alpha m} + 2^{\alpha k} + x_k - 2^{\alpha l} = 2^{\alpha r} - 2^{\alpha s} + \delta(x_r - x_s)$ so $2^{\alpha n} + 2^{\alpha k} + 2^{\alpha s} + x_n + x_k + \delta x_s = 2^{\alpha r} + 2^{\alpha m} + 2^{\alpha l} + \delta x_r$ so by Lemma 3, $1 + 1 + \delta = \delta$, a contradiction.

Case 3. Not case 1 or case 2. Without loss of generality $a \in A_1$ and $b \in A_2$. Pick n such that $a = x_n$ and pick $l < k$ such that $b = 2^{\alpha k} + x_k - 2^{\alpha l}$. Then $x_n + 2^{\alpha k} + x_k - 2^{\alpha l} = 2^{\alpha r} - 2^{\alpha s} + \delta(x_r - x_s)$ so we again get a contradiction using Lemma 3. □

Lemma 5. *If $a, b \in A_3$, then $a + b \notin D(B)$.*

Proof. Pick $n > m$ and $k > l$ such that $a = 2^{\alpha n} - 2^{\alpha m} - x_m$ and $b = 2^{\alpha k} - 2^{\alpha l} - x_l$. Suppose $a+b \in D(B)$, in which case since it is even, $a+b \in A_4 \cup A_5$. Pick $\delta \in \{0,1\}$ and $s < r$ such that $a + b = 2^{\alpha r} - 2^{\alpha s} + \delta(x_r - x_s)$. Then $2^{\alpha n} + 2^{\alpha k} + 2^{\alpha s} + \delta x_s = 2^{\alpha r} + 2^{\alpha m} + 2^{\alpha l} + x_l + x_m + \delta x_r$ so that by Lemma 3, $\delta = 1 + 1 + \delta$, a contradiction. □

Lemma 6. *Let $m < n$ and $l < k$ be given and let $a = 2^{\alpha n} - 2^{\alpha m} - x_m$ and $b = 2^{\alpha k} - 2^{\alpha l}$. If $a + b \in D(B)$, then $l = n$.*

Proof. Since $a + b$ is odd we have $a + b \in A_1$ or $a + b \in A_2$ or $a + b \in A_3$. We show first that the first two possibilities cannot hold. Indeed if we had $a + b \in A_1$, then for some r, $2^{\alpha n} - 2^{\alpha m} - x_m + 2^{\alpha k} - 2^{\alpha l} = x_r$ so that $2^{\alpha n} + 2^{\alpha k} + 2^{\alpha r} = 2^{\alpha m} + 2^{\alpha l} + 2^{\alpha r} + x_m + x_r$ so that by Lemma 3, $1 + 1 = 0$. A similar contradiction is obtained from the assumption that $a + b \in A_2$. Thus we may pick $s < r$ such that $a + b = 2^{\alpha r} - 2^{\alpha s} - x_s$. Then $2^{\alpha n} + 2^{\alpha k} + 2^{\alpha s} + x_s = 2^{\alpha r} + 2^{\alpha m} + 2^{\alpha l} + x_m$. By Lemma 3 we have that $\max\{n, k, s\} = \max\{r, m, l\}$. Since $l < k \leq \max\{n, k, s\}$ we have $l \neq \max\{r, m, l\}$. Similarly $m \neq \max\{r, m, l\}$ and $s \neq \max\{n, k, s\}$. Thus $\max\{r, m, l\} = r$. Assume first $k \leq n$. Then $n = \max\{n, k, s\}$ so $n = r$ so $2^{\alpha k} + 2^{\alpha s} + x_s = 2^{\alpha m} + 2^{\alpha l} + x_m$. By Lemma 2 we have $\max\{k, s\} = \max\{m, l\}$. Since $l < k$ we have $l \neq \max\{m, l\}$ so $l < m$ so conclusion (2) of Lemma 2 cannot hold.

If we had $k \leq s$ we would have $(m, l) = (s, k)$, while $l < k$. Thus $s < k$ so $(k, s, 0, 1) = (m, l, 1, 0)$, a contradiction. Thus we have $n < k$ so that $k = \max\{n, k, s\}$ and hence $k = r$. Then $2^{\alpha n} + 2^{\alpha s} + x_s = 2^{\alpha m} + 2^{\alpha l} + x_m$.

By Lemma 2 $\max\{n, s\} = \max\{m, l\}$. Since $m < n$ we have $m \neq \max\{m, l\}$ so $(l, m) = (n, s)$ or $(l, m) = (s, n)$. The latter is impossible since $m < n$ so in particular $n = l$. □

Lemma 7. *Let $l < k$ and $m < n$ in $\mathbb{N}$ be given with $k \geq n$ and let μ, $\tau \in \{0,1\}$. Let $a = 2^{\alpha k} - 2^{\alpha l} + \tau(x_k - x_l)$ and let $b = 2^{\alpha n} - 2^{\alpha m} + \mu(x_n - x_m)$ and assume that $a + b \in D(B)$. Then some one of the following holds:*

(1) $n = l$ and $\mu = \tau$;
(2) $n = l$ and $\mu = 0$ and $\tau = 1$ and there is some $v < m$ such that $x_k - x_l = 2^{\alpha v} - 2^{\alpha m} - 2^{\alpha v}$;
(3) $n \leq l$ and $\mu = 1$ and $\tau = 0$ and there is some $v > m$ such that $x_k - x_l = 2^{\alpha m} + x_v - x_m$; if $n < l$, then $v = n$; or
(4) $n \leq l$ and $\mu = r = 0$ and $x_k - x_l = 2^{\alpha n} - 2^{\alpha m}$.

Proof. Since $a + b$ is even we must have $a + b \in A_4 \cup A_5$. So pick $r > s$ in $\mathbb{N}$ and $\nu \in \{0,1\}$ such that $a + b = 2^{\alpha r} - 2^{\alpha s} + \nu(x_r - x_s)$. Then $2^{\alpha k} + 2^{\alpha n} + 2^{\alpha s} + \tau x_k + \mu x_n + \nu x_s = 2^{\alpha r} + 2^{\alpha l} + 2^{\alpha m} + \nu x_r + \tau x_l + \mu x_m$. By Lemma 3, $\max\{k,n,s\} = \max\{r,l,m\}$. Since $l < k, m < n$, and $s < r$ we have $\max\{r,l,m\} = r$ and $s \neq \max\{k,n,s\}$. Since $k \geq n, k = \max\{k,n,s\}$.

Case 1. $n \geq s$. If we had $m \geq l$ we would then have $k \geq n \geq s$ and $r > m \geq l$ so that by Lemma 3 we would have $(k,n) = (r,m)$ while $m < n$. Thus $l > m$. We then have $k \geq n \geq s$ and $r > l > m$ so by Lemma 3 some one of the following holds:

(a) $(k,n,s,\tau,\mu,\nu) = (r,l,m,\nu,\tau,\mu)$,
(b) $n = s$ and $(k,n,s,\tau,\mu,\nu) = (r,l,m,\nu,\mu,\tau)$,
(c) $k = n$ and $(k,n,s,\tau,\mu,\nu) = (r,l,m,\tau,\nu,\mu)$,
(d) $k = n = s$ and $(k,n,s,\tau,\mu,\nu) = (r,l,m,\mu,\tau,\nu)$, or
(e) $(k,n,\tau,\mu,\nu) = (r,l,\tau,\nu,\mu)$, and $\tau \neq \mu$ and $s \neq m$.

If $\mu = \tau$ we have that conclusion (1) of the current lemma holds. So assume $\mu \neq \tau$. This eliminates (a) and (d) above. The fact that $m < l$ eliminates (b) above. The fact that $l < k$ eliminates (c) above. Thus we have (e) must hold. Observe also that $\tau \neq \nu$. (If so one would have $2^{\alpha n} + 2^{\alpha s} + \mu x_n + \nu x_s = 2^{\alpha l} + 2^{\alpha m} + \tau x_l + \mu x_m$ so that by Lemma 2 one would have $m = s$, which is forbidden by (e).)
There are thus two possibilities. First one could have $\mu = \nu = 0$ and $\tau = 1$. In this case $2^{\alpha s} + x_k = 2^{\alpha m} + x_l$ so $2^{\alpha m} - 2^{\alpha s} = x_k - x_l > 0$ so $s < m$ and conclusion (2) of the current lemma holds. Second one could have $\mu = \nu = 1$ and $\tau = 0$. In this case $2^{\alpha s} + x_l + x_s = 2^{\alpha m} + x_k + x_m$ so that $x_k - x_l = 2^{\alpha s} - 2^{\alpha m} + x_s - x_m$ and conclusion (3) of the current lemma holds.

Case 2. $n < s$. Since $s < r = k$ we have then $k > s > n$. By Lemma 3 we then have that $(k,s) = (r,l)$ or $(k,s) = (r,m)$. Since $m < n$, the latter alternative is impossible and hence $m < l$. Also $l < k = r$ so we have $r > l > m$. Since $n \neq m$ we have only one possibility from Lemma 3, namely that $(k,s,\tau,\nu,\mu) = (r,l,\tau,\nu,\mu)$ and $\tau \neq \nu$. Since $k = r$ and $s = l$ we then have $2^{\alpha n} + \tau x_k + \mu x_n + \nu x_l = 2^{\alpha m} + \nu x_k + r x_l + \mu x_m$. Suppose $\tau = 1$. Then we have $\nu = 0$ so $x_k - x_l = 2^{\alpha m} - 2^{\alpha n} + \mu(x_m - x_n) < 0$, which is impossible.

Thus $\tau = 0$ and $\nu = 1$ and hence $x_k - x_l = 2^{\alpha n} - 2^{\alpha m} + \mu(x_n - x_m)$. If $\mu = 1$ this gives conclusion (3) of the current lemma while if $\mu = 0$ it gives conclusion (4). □

Lemma 8. *Assume $a \geq b \geq c$ and $\{a, b, c\} \subseteq A_4 \cup A_5$ and $\{a+b, a+c, b+c, a+b+c\} \subseteq D(B)$. Then there exist $k > l > m > s$ in $\mathbb{N}$ such that $a = 2^{\alpha k} - 2^{\alpha l}, b = 2^{\alpha l} - 2^{\alpha m}$, and $c = 2^{\alpha m} - 2^{\alpha s} = x_k - x_l$.*

Proof. Since a, b, and c are in $A_4 \cup A_5$ we have $k > l, n > m$, and $r > s$ in $\mathbb{N}$ and τ, μ, ν in $\{0, 1\}$ such that $a = 2^{\alpha k} - 2^{\alpha l} + \tau(x_k - x_l), b = 2^{\alpha n} - 2^{\alpha m} + \mu \cdot (x_n - x_m)$ and $c = 2^{\alpha r} - 2^{\alpha s} + \nu \cdot (x_r - x_s)$. Since $a \geq b \geq c$ we have $k \geq n \geq r$. Applying Lemma 7 to $a + b$ we have one of:

(1) $n = l$ and $\mu = \tau$;
(2) $n = l$ and $\mu = 0$ and $\tau = 1$ and there is some $v < m$ such that $x_k - x_l = 2^{\alpha m} - 2^{\alpha v}$;
(3) $n \leq l$ and $\mu = 1$ and $\tau = 0$ and there is some $v > m$ such that $x_k - x_l = 2^{\alpha v} - 2^{\alpha m} + x_v - x_m$; if $n < l$, then $v = n$; or
(4) $n < l$ and $\mu = \tau = 0$ and $x_k - x_l = 2^{\alpha n} - 2^{\alpha m}$.

Applying Lemma 7 to $b + c$ we have one of:

(1)′ $r = m$ and $\nu = \mu$;
(2)′ $r = m$ and $\nu = 0$ and $\mu = 1$ and there is some $t < s$ such that $x_n - x_m = 2^{\alpha s} - 2^{\alpha t}$.
(3)′ $r \leq m$ and $\nu = 1$ and $\mu = 0$ and there is some $t > s$ such that $x_n - x_m = 2^{\alpha t} - 2^{\alpha s} + x_t - x_s$; if $r < m$, then $t = r$; or
(4)′ $r < m$ and $\nu = \mu = 0$ and $x_n - x_m = 2^{\alpha r} - 2^{\alpha s}$.

Now from (1)′, (2)′, (3)′ and (4)′ we see that in any event $r \leq m$ and from (1), (2), (3) and (4) we see that $n \leq l$. Thus $r \leq m < n \leq l$. Thus applying Lemma 7 to $a + c$ we have one of:

(3)* $r < l$ and $\nu = 1$ and $\tau = 0$ and $x_k - x_l = 2^{\alpha r} - 2^{\alpha s} + x_r - x_s$; or
(4)* $r < l$ and $\nu = \tau = 0$ and $x_k - x_l = 2^{\alpha r} - 2^{\alpha s}$.

We show first that (1) must hold. From (3)* or (4)* we conclude $\tau = 0$ so (2) cannot hold.

Now suppose that (3) or (4) holds and pick $v > m$ and $\gamma \in \{0, 1\}$ such that $x_k - x_l = 2^{\alpha v} - 2^{\alpha m} + \gamma \cdot (x_v - x_m)$. Since (3)* or (4)* holds pick $\lambda \in \{0, 1\}$ such that $x_k - x_l = 2^{\alpha r} - 2^{\alpha s} + \lambda \cdot (x_r - x_s)$. Then $2^{\alpha v} + 2^{\alpha s} + \gamma x_v + \lambda x_s = 2^{\alpha r} + 2^{\alpha m} + \lambda x_r + \gamma x_s$. Since $s < r$ and $m < v$ we conclude from Lemma 2 that $(v, s) = (r, m)$. But we have already observed that $r \leq m$ so $r \leq m = s < r$, a contradiction.

We have thus established that (1) holds. In particular we know $\mu = \tau$ from (1) and $\tau = 0$ from (3)* or (4)* so $\mu = \tau = 0$. We now show that (1)′ holds. Since $\mu = 0$ we know (2)′ cannot hold.

Since (1) holds we know that $a = 2^{\alpha k} - 2^{\alpha l}$ and $b = 2^{\alpha l} - 2^{\alpha m}$ so that $a+b+c = 2^{\alpha k} - 2^{\alpha m} + 2^{\alpha r} - 2^{\alpha s} + \nu \cdot (x_r - x_s)$. Also $a+b+c \in A_4 \cup A_5$ so pick $w > u$ in $\mathbb{N}$ and $\rho \in \{0,1\}$ such that $a+b+c = 2^{\alpha w} - 2^{\alpha u} + \rho \cdot (x_w - x_u)$. Then $2^{\alpha k} + 2^{\alpha r} + 2^{\alpha u} + \nu x_r + \rho x_u = 2^{\alpha w} + 2^{\alpha m} + 2^{\alpha s} + \rho x_w + \nu x_s$. Now $\max\{k,r,u\} = \max\{w,m,s\}$ and $m < k$ and $s < r$ so $w = \max\{w,m,s\}$. Also $m \geq r > s$ so we have $w > m > s$. Since $r \leq m < k$ and $u < w$ we have $k = \max\{k,r,u\}$. Thus $k = w$. We suppose $(3)'$ or $(4)'$ holds and consider two cases.

Case 1. $m = r$. Then $(4)'$ cannot hold so $(3)'$ holds and hence $\nu = 1$. We also conclude that $r \geq u$. (For if $r < u$ then by Lemma 3 we have $(k,u) = (w,m)$ so $m = u > r = m$.) Now since $w > m > s$ the only possibilities in Lemma 3 are for conclusion (1) or (5) to hold. If conclusion (1) held we would have $(k,r,u,0,1,\rho) = (w,m,s,\rho,0,1)$ which is impossible. Thus $(k,r,0,1,\rho) = (w,m,0,\rho,1)$ so $\rho = 1$. Thus we have $2^{\alpha u} + x_m + x_u = 2^{\alpha s} + x_k + x_s$ so $x_k - x_m = 2^{\alpha u} - 2^{\alpha s} + x_u - x_s$ and hence $u > s$. Also by $(3)'$ pick $t > s$ such that $x_n - x_m = 2^{\alpha t} - 2^{\alpha s} + x_t - x_s$. Since $\nu = 1$, $(3)^*$ holds so we have $x_k - x_l = 2^{\alpha r} - 2^{\alpha s} + x_r - x_s$. Since $l = n$ we then have $x_k - x_m = 2^{\alpha t} + 2^{\alpha r} - 2 \cdot 2^{\alpha s} + x_t + x_r - 2 \cdot x_s$. Thus $2^{\alpha u} - 2^{\alpha s} + x_u - x_s = 2^{\alpha t} + 2^{\alpha r} - 2 \cdot 2^{\alpha s} + x_t + x_r - 2 \cdot x_s$ so that $2^{\alpha u} + 2^{\alpha s} + x_u + x_s = 2^{\alpha t} + 2^{\alpha r} + x_t + x_r$. Thus by Lemma 2 we have $(u,s) = (t,r)$ or $(u,s) = (r,t)$.

But $r > s$ and $t > s$, a contradiction.

Case 2. $m > r$. Then from $(3)'$ or $(4)'$ we have that $x_n - x_m = 2^{\alpha r} - 2^{\alpha s} + \nu \cdot (x_r - x_s)$. Now $w > m > s$ and $(w,m) \neq (k,r)$ so by Lemma 3 we must have $k > u > r$. Since $s < r$ we must then have conclusion (5) of Lemma 3 must hold and consequently $\rho \neq 0$, i.e, $p = 1$. Thus $2^{\alpha r} + \nu x_r + x_m = 2^{\alpha s} + \nu x_s + x_k$ so that $x_k - x_m = 2^{\alpha r} - 2^{\alpha s} + \nu \cdot (x_r - x_s)$ so $x_k - x_m = x_n - x_m$ and hence $k = n$. Since $n = l < k$, this is a contradiction.

Thus we have established that $(1)'$ holds. Thus $\mu = \tau = \nu$ so $(3)^*$ does not hold so $(4)^*$ holds. The conjunction of (1), $(1)'$, and $(4)^*$ is precisely the conclusion of this lemma. □

Lemma 9. *Let* a_1, a_2, a_3, *and* a_4 *in* $\mathbb{N}$ *be given such that* $FS(\langle a_n \rangle_{n=1}^4) \subseteq D(B)$. *Then there is some* $i \in \{1,2,3,4\}$ *such that* $a_i \in A_1 \cup A_2$ *and* $\{a_j : j \in \{1,2,3,4\}$ *and* $j \neq i\} \subseteq A_4 \cup A_5$.

Proof. Suppose first that $\{a_1, a_2, a_3, a_4\} \subseteq A_4 \cup A_5$ and assume without loss of generality that $a_1 \geq a_2 \geq a_3 \geq a_4$. Applying Lemma 8 to a_1, a_2, and a_3 we pick $k > l > m > s$ in $\mathbb{N}$ such that $a_1 = 2^{\alpha k} - 2^{\alpha l}$, $a_2 = 2^{\alpha l} - 2^{\alpha m}$, and $a_3 = 2^{\alpha m} - 2^{\alpha s}$. Applying Lemma 8 to a_1, a_3, and a_4 we conclude that $m = l$, a contradiction.

Now by Lemma 4 at most one i has $a_i \in A_1 \cup A_2$ and by Lemma 5 at most one i has $a_i \in A_3$ so to complete the proof it suffices to show that no $a_i \in A_3$. Suppose we have some $a_i \in A_3$ and assume without loss of generality that $a_1 \in A_3$.

Case 1. Some j has $a_j \in A_1 \cup A_2$. Without loss of generality $a_2 \in A_1 \cup A_2$. We may further assume without loss of generality that $a_3 \geq a_4$. Since $FS(< a_1 + a_2, a_3, a_4 >) \subseteq A_4 \cup A_5$ we have by Lemma 8 some $k > l \geq m > s$ such that $a_3 = 2^{\alpha k} - 2^{\alpha l}$ and $a_4 = 2^{\alpha m} - 2^{\alpha s}$. (If $a_1 + a_2$ is between a_3 and a_4 we have $l > m$. Otherwise equality holds.) Pick $u > v$ in $\mathbb{N}$ such that $a_1 = 2^{\alpha u} - 2^{\alpha v} - x_v$. Since $a_1 + a_3 \in D(B)$ we have by Lemma 6 that $l = u$. Since $a_1 + a_4 \in D(B)$ we have by Lemma 6 that $s = u$. But $s < l$, a contradiction.

Case 2. $\{a_2, a_3, a_4\} \subseteq A_4 \cup A_5$. Without loss of generality $a_2 \geq a_3 \geq a_4$. Then by Lemma 8 we have some $k \geq l > m > s$ such that $a_2 = 2^{\alpha k} - 2^{\alpha l}, a_3 = 2^{\alpha l} - 2^{\alpha m}$, and $a_4 = 2^{\alpha m} - 2^{\alpha s}$. Applying Lemma 6 to (a_1, a_2) and (a_1, a_4) we again get $l = u = s$, a contradiction. □

We temporarily abandon our assumption that α has a fixed value in order to state the next theorem.

Theorem 8. *Let $\epsilon > 0$ be given. There is a set $B \subseteq \mathbb{N}$ with $\bar{d}(D(B)) > 1/2 - \epsilon$ such that no a_1, a_2, a_3, a_4, and a_5 have $FS(\langle a_n \rangle_{n=1}^5) \subseteq D(X)$.*

Proof. Pick $\alpha \in \mathbb{N}$ such that $1/2^{\alpha-5} < \epsilon$. Define B as in Definition 1. Observe that $A_1 \subseteq D(B)$ and $\bar{d}(A_1) \geq 1/2 - 1/2^{\alpha-5}$ since $|A_1 \cap \{1, 2, \ldots, 2^{\alpha t + \alpha - 2}\}| \geq \frac{1}{2}((2^{\alpha t + \alpha - 2} - 2^{\alpha t + 2}))$.

Suppose now one has a_1, a_2, a_3, a_4, and a_5 with $FS(\langle a_n \rangle_{n=1}^5) \subseteq D(B)$. Applying Lemma 9 first to a_1, a_2, a_3, and a_4 one has without loss of generality that $a_1 \in A_1 \cup A_2$ and $\{a_2, a_3, a_4\} \subseteq A_4 \cup A_5$. Applying Lemma 9 to a_2, a_3, a_4, and a_5 one sees that $a_5 \in A_1 \cup A_2$. Then applying Lemma 4 to a_1 and a_5 one obtains a contradiction. □

We close with two questions which are raised by Theorems 7 and 8.

Question 1. *If $B \subseteq \mathbb{N}$ and $\bar{d}(D(B)) > 0$, must there exist a_1, a_2, a_3, and a_4 in $\mathbb{N}$ with $FS(\langle a_n \rangle_{n=1}^4) \subseteq D(B)$?*

Since always $\bar{d}(B \cap (B - t)) \geq 2 \cdot \bar{d}(B) - 1$, one easily sees that if $\bar{d}(D(X)) > 1 - 1/2^{m-1}$, there will exist $a_1, a_2, \ldots, a_m$ with $FS(\langle a_i \rangle_{i=1}^m) \subseteq D(X)$. (See [8, Theorem 4.5].) To utilize this to obtain $FS(\langle a_n \rangle_{n=1}^5)$ one needs $\bar{d}(D(X)) > 1 - 1/16$.

Question 2. *If $\bar{d}(D(X)) = 1/2$ or even if $\bar{d}(D(X)) > 1/2$ must there exist a_1, a_2, a_3, a_4, and a_5 in $\mathbb{N}$ with $FS(\langle a_n \rangle_{n=1}^5) \subseteq D(X)$?*

Reference

1. V. Bergelson, *A density statement generalizing Schur's Theorem*, J. Comb. Theory (Series A) **43** (1986), 336–343.
2. V. Bergelson, *Applications of ergodic theory to combinatorics*, Dissertation (in Hebrew), Hebrew University of Jerusalem: 1984.
3. V. Bergelson, *Sets of recurrence of* $\mathbb{Z}^m$ *actions and properties of sets of differences in* $\mathbb{Z}^m$, J. London Math. Soc. (2) **31** (1985), 295–304.
4. V. Bergelson and N. Hindman, *On* IP* *sets and central sets*, Combinatorica **14** (1994), 269–277.
5. W. Deuber, N. Hindman, I. Leader, and H. Lefmann, *Infinite partition regular matrices*, Combinatorica, **15** (1995), 333–355.
6. H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J . Anal. Math. **31** (1977), 204–256.
7. H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton Univ. Press, Princeton, 1981.
8. N. Hindman, *On density, translates, and pairwise sums of integers*, J. Comb. Theory (Series A) **33** (1982), 147–157.

Integer Sets Containing No Solution to $x + y = 3z$

Fan R. K. Chung and John L. Goldwasser

F.R.K. Chung (✉)
Department of Mathematics, University of California, San Diego, La Jolla,
CA 92037, USA
e-mail: fan@ucsd.edu

J.L. Goldwasser
Department of Mathematics, West Virginia University, Morgantown,
WV 26506, USA
e-mail: jgoldwas@math.wvu.edu

Summary. We prove that a maximum subset of $\{1, 2, \ldots, n\}$ containing no solutions to $x + y = 3z$ has $\lceil \frac{n}{2} \rceil$ elements if $n \neq 4$, thus settling a conjecture of Erdős. For $n \geq 23$ the set of all odd integers less than or equal to n is the unique maximum such subset.

1. Introduction

Many classical problems in computational number theory focus upon subsets S of positive integers with the property that for all x, y, z in S, we have $x + y \neq kz$, for a fixed positive integer k. The history can be dated from 1916, when Schur [5], in work related to Fermat's Last Theorem, proved that the set of positive integers cannot be partitioned into finitely many sum-free sets, i.e., sets having no solution to $x + y = z$. This result is a Ramsey-type theorem which predates Ramsey's Theorem. In 1927, van der Waerden [9] considered subsets having no solution to $x + y = 2z$, or in other words, subsets containing no three-term arithmetic progression. His celebrated theorem states that the positive integers cannot be partitioned into finitely many subsets each of which contains no k-term arithmetic progressions. In 1952, Roth proved that a set of positive upper density contains three-term progressions [4]. This was improved by Szemerédi to four-term progressions [7] and later to the general k-term progressions [8].

A problem which appears in several undergraduate combinatorics texts is to show that a maximum subset of $\{1, \ldots, n\}$ containing no solutions to $x + y = z$ (x, y, z not necessarily distinct) has size $\lceil \frac{n}{2} \rceil$. In Sect. 2 of this paper we find all such subsets and prove the following theorem:

Theorem 1. *A maximum subset of $\{1, \ldots, n\}$ containing no solution to $x + y = z$ (x, y, z not necessarily distinct) has size $\lceil \frac{n}{2} \rceil$. If $n \geq 3$ is odd there are precisely two maximum subsets: the odd integers less than or equal to n and $\{x \in Z \mid \frac{n+1}{2} \leq x \leq n\}$. If $n \geq 4$ is even there are at least three maximum*

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_11,

subsets: $\{x \in Z \mid \frac{n+1}{2} \leq x \leq n\}$ *and the two maximum subsets for the odd number* $n-1$*. For even* $n \geq 10$ *these three are the only ones. For smaller even numbers,* $\{1,4\}, \{2,5,6\}, \{1,4,6\}$*, and* $\{2,3,7,8\}$ *are the only additional ones.*

Let $f^*(n,2)$ denote the maximum size subset of $\{1,\ldots,n\}$ containing no three-term arithmetic progressions (such subsets contain no solutions to $x+y=2z$, but now, for the problem to make sense, x,y,z are distinct). Roth first showed [4] that

$$f^*(n,2) = O\left(\frac{n}{\log\log n}\right).$$

The current best bounds are, for appropriate absolute constant c_i,

$$ne^{-c_1\sqrt{\log n}} < f^*(n,2) < \frac{c_2 n}{(\log n)^{c_3}}$$

where the lower bound was proved by Salem and Spencer [6] (see also Behrend [1]), and the upper bound was proved by Heath-Brown and Szemerédi [3].

Erdős conjectured that a maximum subset of $\{1,\ldots,n\}$ having no solutions to $x+y=3z$ (x,y,z not necessarily distinct) has size no more than a small constant more than $\lceil \frac{n}{2} \rceil$. In Sect. 3 we verify this conjecture by proving the following theorem:

Theorem 2. *Let* T_n *be a subset of* $\{1,\ldots,n\}$ *of maximum size such that* $x+y=3z$ *has no solutions with* $x,y,z \in T_n$ *(*x,y,z *not necessarily distinct). If* $n \neq 4$ *then* $|T_n| = \lceil \frac{n}{2} \rceil$*.*

In Sect. 4 we show that for sufficiently large n there is a unique maximum such subset:

Theorem 3. *If* $n \geq 23$ *and* T_n *is a subset of maximum size of* $\{1,\ldots,n\}$ *having no solutions to* $x+y=3z$ *then* T_n *is the set of all odd integers less than or equal to* n*.*

We use the standard notation $\lceil\ \rceil$ and $\lfloor\ \rfloor$ for least integer not less than and greatest integer not greater than, respectively. For a and b nonnegative integers we let $[a,b]$ denote the set of all integers x such that $a \leq x \leq b$.

2. Maximum Sum-Free Sets of $\{1,\ldots,n\}$

Proof of Theorem 1. First we show the maximum size is always $\lceil \frac{n}{2} \rceil$. Let U_n be a maximum sum-free subset of $\{1,\ldots,n\}$ and let p be the largest integer in U_n. Then at most one integer in each of the pairs $(i, p-i)$, $i = 1,2,\ldots,\lceil \frac{p-2}{2} \rceil$ is in U_n, so $|U_n| \leq \lceil \frac{p}{2} \rceil \leq \lceil \frac{n}{2} \rceil$. Clearly, there are subsets which attain this bound, so $|U_n| = \lceil \frac{n}{2} \rceil$.

To characterize the maximum subsets we consider two cases depending on the parity of n.

Case 1 (n odd). Let $n \geq 5$ be the smallest odd integer such that there exists a maximum sum-free subset U_n of $\{1, \ldots, n\}$ which is not the odd integers less than or equal to n or $\left[\frac{n+1}{2}, n\right]$. Clearly $n \in U_n$ (or else $|U_n| \leq \lceil \frac{n-1}{2} \rceil < \lceil \frac{n}{2} \rceil$), so if $n-1 \notin U_n$ then, by the minimality of n, either U_n is the set of all odd integers less than or equal to n (which is impossible by assumption) or $U_n = [\frac{n-1}{2}, n-2] \cup \{n\}$ which is impossible since $\frac{n-1}{2} + \frac{n+1}{2} = n$. So we can assume $n-1$ and n are in U_n.

Let G be the graph with vertex set $V = \{v_i \mid i \in [2, \frac{n-3}{2}] \cup [\frac{n+1}{2}, n-2]\}$ of size $n-4$ where $\{v_i, v_j\}$ is an edge of G if and only if $i + j = n$ or $i + j = n-1$. Then G is the path $v_{n-2}, v_2, v_{n-3}, v_3, \ldots, v_{\frac{n-3}{2}}, v_{\frac{n+1}{2}}$ (with an odd number of vertices). Since $n-1$ and n are in U_n, 1 and $\frac{n-1}{2}$ are not, so the other $\frac{n-3}{2}$ integers in U_n must be the indices of an independent set of vertices in G (i.e., no two of them adjacent). The only sufficiently large independent set in G is the maximum independent set which has indices $[\frac{n+1}{2}, n-2]$, so $U_n = [\frac{n+1}{2}, n]$.

Case 2 (n even). If $n \geq 4$ is even and $n \notin U_n$, then certainly U_n must be one of the two maximum subsets for the odd integer $n-1$. It is easy to check that the statement in the theorem about when n is 4, 6, or 8 is correct. Let $n \geq 10$ be the smallest even integer such that there exists a maximum sum-free subset U_n of $\{1, \ldots, n\}$ which contains n but is not $[\frac{n}{2}+1, n]$. If $n-1 \notin U_n$, we let $U_{n-2} = U_n \cap [1, n-2]$, so that $|U_{n-2}| = \frac{n-2}{2}$. U_{n-2} cannot be the odd integers less than or equal to $n-3$ because $3 + (n-3) = n$. And U_{n-2} cannot be $[\frac{n-2}{2}, n-3]$ or $[\frac{n}{2}, n-2]$ because $\frac{n}{2} \notin U_n$. So U_{n-2} cannot be any of the three kinds of maximum subsets for even n described in the theorem. So by the minimality of n we would have to have $n = 10$ and $U_{n-2} = \{2, 3, 7, 8\}$. This cannot be because $2 + 8 = 10$. Hence, as with the odd case, $n-1$ and n are both in U_n.

Now let H be the graph with vertex set $V = \{v_i \mid i \in [2, \frac{n-2}{2}] \cup [\frac{n+2}{2}, n-2]\}$ of size $n-4$ where $\{v_i, v_j\}$ is an edge of H if and only if $i + j = n$ or $i + j = n-1$ or $\{i, j\} = \{n-2, \frac{n-2}{2}\}$. Then H is the cycle $v_{n-2}, v_2, v_{n-3}, v_3, \ldots, v_{\frac{n+2}{2}}, v_{\frac{n-2}{2}}$. Since $n-1$ and n are in $U_{n,1}$ and $\frac{n}{2}$ are not, so the other $\frac{n-4}{2}$ integers in U_n must be the indices of an independent set of vertices in H. There are two possibilities: $[2, \frac{n-2}{2}]$ and $[\frac{n+2}{2}, n-2]$. If $n \geq 10$ the first of these cannot occur because it contains 2 and 4 (If $n = 6$ the first possibility gives us $\{2, 5, 6\}$ while if $n = 8$ it gives $\{2, 3, 7, 8\}$.). So we have $U_n = [\frac{n}{2} + 1, n]$ which completes the proof. □

3. The Size of a Set with No Solution to $x + y = 3z$

We observe that if T_n is a set containing no solutions to $x + y = 3z$ and if $w \in T_n$, then $\frac{1}{2}w$, $\frac{2}{3}w$, $\frac{3}{2}w$, and $2w$ cannot be in T_n (because x, y, z need not be distinct).

Proof of Theorem 2. The set of all odd integers less than or equal to n has no solutions to $x + y = 3z$ so $|T_n| \geq \lceil \frac{n}{2} \rceil$. It is easy to check that $|T_n| = \lceil \frac{n}{2} \rceil$ for $n = 1, 2, 3, 5$ (there are three ways to choose T_5: $\{1, 3, 4\}$, $\{1, 3, 5\}$, or $\{1, 4, 5\}$. The first of these shows that $|T_4| = 3$.) So it remains to show $|T_n| \leq \lceil \frac{n}{2} \rceil$ for $n \geq 6$. Let n be the smallest integer greater than or equal to 6 such that $|T_n| > \lceil \frac{n}{2} \rceil$. We can assume n is even and $n \in T_n$ (otherwise $n-1$ is a smaller counter-example).

Case 1. T_n has no integer x such that $\frac{n}{3} < x \leq \frac{2n}{3}$. By the minimality of n at most $\lceil \lfloor \frac{n}{3} \rfloor / 2 \rceil$ of the integers in $[1, \lfloor \frac{n}{3} \rfloor]$ are in T_n provided $\lfloor \frac{n}{3} \rfloor \neq 4$. So

$$|T_n| \leq \left\lceil \frac{\lfloor \frac{n}{3} \rfloor}{2} \right\rceil + \left| \left[\lfloor \frac{2n}{3} \rfloor + 1, n \right] \right|$$

$$= \begin{cases} \frac{n}{2} & \text{if } n = 0 \text{ or } 2 \pmod 6 \\ \frac{n}{2} + 1 & \text{if } n = 4 \pmod 6). \end{cases}$$

So we are done if $n \neq 4 \pmod 6$ and $\lfloor \frac{n}{3} \rfloor \neq 4$. If $\lfloor \frac{n}{3} \rfloor = 4$ then $n = 12$ or $n = 14$ and the respective candidates for a set of size greater than $\frac{n}{2}$ are $\{1, 3, 4, 9, 10, 11, 12\}$ and $\{1, 3, 4, 10, 11, 12, 13, 14\}$. However, neither is acceptable because $1 + 11 = 3 \cdot 4$.

It remains only to consider $n = 6k + 4$ $(k = 1, 2, 3, \ldots)$, in which case the candidate for a counter-example is to choose $k + 1$ of the integers in $[1, 2k + 1]$ and all the integers in $[4k + 3, 6k + 4]$. If $2k + 1 \notin T_n$ then more than half of the first $2k$ integers are in T_n, so $2k = 4$, $n = 16$, and the candidate is $\{1, 3, 4, 11, 12, 13, 14, 15, 16\}$ which fails again because it contains 1, 4, and 11. So $2k + 1 \in T_n$ and $6k + 3$ is a forbidden sum. Since $[4k + 3, 6k + 2] \subseteq T_n$ it follows that $T_n \cap [1, 2k] = \emptyset$. This is impossible since $k + 1$ of the first $2k + 1$ integers are in T_n.

Case 2. T_n has an integer x such that $\frac{n}{3} < x \leq \frac{2n}{3}$.

In fact $x \neq \frac{2n}{3}$ since $n \in T_n$. Assume x is the largest integer in T_n such that $\frac{n}{3} < x < \frac{2n}{3}$. Then the integers in $W = [3x - n, n]$ can be arranged in pairs as follows:

$$(3x - n + j, n - j) \quad j = 0, 1, 2, \ldots n - \left\lceil \frac{3x}{2} \right\rceil$$

Since the sum of the integers in each pair is $3x$, at most one integer from each pair can be in T_n.

If x is even then $|W|$ is odd and one of the pairs is $(\frac{3x}{2}, \frac{3x}{2})$. In this case T_n contains at most $\frac{1}{2}(|W| - 1)$ integers from W and, by the minimality of n, at most $\frac{n-|W|+1}{2}$ integers from $[1, 3x - n - 1]$. So $|T_n| \leq \frac{n}{2}$.

If x is odd, then $|W|$ is even and at most $\frac{1}{2}|W|$ integers from W can be in T_n. So at most $\frac{1}{2}(n - |W|)$ integers from $[1, 3x - n - 1]$ can be in T_n provided $3x - n - 1 \neq 4$. So we are done except for the possibility that $3x - n = 5$. In this case one of the pairs of integers in W is $(\frac{3x-1}{2}, \frac{3x+1}{2}) = (\frac{n+4}{2}, \frac{n+6}{2})$. Since x is the largest integer in T_n which is less than $\frac{2n}{3}$ and since $x = \frac{n+5}{3} < \frac{n+4}{2} < \frac{n+6}{2}$, we must have $\frac{n+6}{2} > \frac{2n}{3}$ (so that one integer in this pair can be in T_n). Solving this gives $n < 18$. Since $n \geq 6$, the only possibilities are $n = 10$, $x = 5$ and $n = 16$, $x = 7$. The first of these is impossible because $n \in T_n$ but T_n cannot contain both 5 and 10. For the second possibility, since $7 \in T_n$ certainly $14 \notin T_n$ so the only candidate is $\{1, 3, 4, 7, 11, 12, 13, 15, 16\}$. But T_n cannot contain 1, 4, and 11 so the proof is complete. □

4. Maximum Sets with No Solutions to $x + y = 3z$

Choosing lots of smaller integers to go into a set T_n which has no solutions to $x + y = 3z$ clearly eliminates some of the larger integers from inclusion. If the smaller included integers follow a simple pattern it may be possible to get a simple description of the eliminated larger integers.

Lemma 1. *Let w be an odd integer greater than or equal to 3. If T is a set which contains no solutions to $x + y = 3z$ and if T contains all odd positive integers less than or equal to w, then T contains no even integer less than $3w$.*

Proof. The result is easy to verify for $w = 3$. If $w \geq 5$ and v is any even number less than $3w$, then (precisely) one of the integers $v + 1, v + 3, v + 5$ is equal to $3t$ where t is an odd number less than or equal to w. □

In proving Theorem 3 we will make frequent use of the maximum size subsets of $\{1, \ldots, n\}$ with no solutions to $x + y = 3z$ for $n \leq 22$. We have calculated them all and display them for even n between 6 and 22 inclusive. To get all such maximum subsets for $n = 2p - 1$ for $p = 3, 4, \ldots, 11$, just choose the ones for $n = 2p$ which do not include the integer $2p$.

Proof of Theorem 3. Suppose the theorem is false and let $n \geq 23$ be the smallest counter-example with T_n a subset of $\{1, \ldots, n\}$ of size $\lceil \frac{n}{2} \rceil$ which contains an even integer. It is easy to see that 23 cannot be added to any of the maximum subsets of $\{1, \ldots, 22\}$ listed in the table without producing a solution to $x + y = 3z$, so $n \geq 24$. By the minimality of the counter-example we can assume n is even and $n \in T_n$. We divide the proof into cases (and subcases) along the lines of the proof of Theorem 2.

Case 1. T_n has no integer x such that $\frac{n}{3} < x \leq \frac{2n}{3}$. Let y be the largest integer in T_n such that $y \leq \frac{n}{3}$. Then $3y$ is a forbidden sum and at most one member of each pair $(i, 3y-i)$ $i = 1, 2, \ldots, y$ can be in T_n. Since T_n has no integers strictly between y and $2y$

$$\frac{n}{2} = |T_n| \leq y + |[3y, n]| = n - 2y + 1.$$

So, $y \leq \lfloor \frac{1}{4}(n+2) \rfloor$ and, since $T_n \cap [y+1, \lfloor \frac{2n}{3} \rfloor] = \emptyset$,

$$\begin{aligned}\frac{n}{2} &\leq |T_n \cap [1, y]| + \left\lceil \frac{n}{3} \right\rceil \\ &\leq \left\lceil \frac{y}{2} \right\rceil + \left\lceil \frac{n}{3} \right\rceil \qquad (if \quad y \neq 4) \\ &\leq \left\lfloor \frac{n+6}{8} \right\rfloor + \left\lceil \frac{n}{3} \right\rceil .\end{aligned}$$

The only even solutions greater than 22 for this inequality are $n = 26, 28, 34$ with corresponding values $y = 7, 7, 9$ respectively. If $n = 34$ and $y = 9$ then T_n must contain five of the first nine integers, which (see Table 1) must be $\{1, 3, 5, 7, 9\}$, and everything in $[23, 34]$. Since $24 + 3 = 3 \cdot 9$ this cannot happen. If $y = 7$ and $n = 26$ or $n = 28$, then T_n contains everything in $[18, 26]$ or $[19, 28]$ respectively. But $7 \in T_n$, so 21 is a forbidden sum, and, since 19 and 20 are in T_n (in both cases), neither 1 nor 2 can be in T_n. Since four of the first seven positive integers must be in T_n this is a contradiction (see Table 1). If $y = 4$, then $|T_n \cap [1, y]|$ could be as much as 3 in the above inequalities, but $\frac{n}{2} \leq 3 + \lceil \frac{n}{3} \rceil$ has no solutions for $n \geq 24$. So Case 1 cannot occur.

Case 2. T_n has an integer y such that

$$\frac{n}{3} < y \leq \frac{2n}{3} \tag{1}$$

Let x be the largest integer in T_n satisfying (1). Then $3x$ is a forbidden sum and the integers in $W = [3x - n, n]$ can be arranged in pairs $\left(\lfloor \frac{3x}{2} \rfloor - i, \lceil \frac{3x}{2} \rceil + i\right)$ $i = 0, 1, \ldots, n - \lceil \frac{3x}{2} \rceil$ such that the sum of the integers in each pair is $3x$. If x is even then $\frac{3x}{2}$ is paired with itself. All other pairs (for x even or odd) have distinct integers.

If x is even then, because of the above pairing,

$$|T_n \cap W| \leq \frac{|W| - 1}{2} \tag{2}$$

But by Theorem 2,

$$T_n \cap [1, 3x - n - 1] \leq \left\lceil \frac{3x - n - 1}{2} \right\rceil = \frac{n - |W| + 1}{2} \tag{3}$$

Table 1 Maximum subsets of $1, \ldots, n$ with no solutions to $x + y = 3z$ for $6 \leq n \leq 22$ (except the set of all odd integers less than or equal to n)

n	T_n	
6	1 3 4	1 5 6
	1 4 5	2 5 6
8	1 3 4 7	1 5 6 8
	1 5 6 7	2 5 6 8
	2 5 6 7	1 6 7 8
		2 6 7 8
10	1 3 4 7 10	1 7 8 9 10
	1 3 7 9 10	2 7 8 9 10
	1 4 7 9 10	3 7 8 9 10
12	1 3 4 7 10 12	1 3 9 10 11 12
14	1 3 9 10 11 12 13	1 3 10 11 12 13 14
	1 3 9 10 11 12 14	3 4 10 11 12 13 14
	1 3 4 10 12 13 14	
16	1 3 4 7 10 12 13 16	1 3 4 12 13 14 15 16
	3 4 7 11 12 13 15 16	1 3 11 12 13 14 15 16
	1 3 4 7 12 13 15 16	3 4 11 12 13 14 15 16
	1 3 7 11 12 13 15 16	
18	1 3 4 12 13 14 15 16 17	1 3 4 13 14 15 16 17 18
20	1 3 4 13 14 15 16 17 18 19	1 3 4 14 15 16 17 18 19 20
	1 3 4 13 14 15 16 17 18 20	
22	1 3 4 7 10 12 13 16 19 21 22	1 3 5 15 16 17 18 19 20 21 22
	1 3 4 15 16 17 18 19 20 21 22	1 4 5 15 16 17 18 19 20 21 22

Since $\frac{n-|W|+1}{2} + \frac{|W|-1}{2} = \frac{n}{2}$, equality must hold in (2) and (3). So T_n contains precisely one integer out of each pair of distinct integers above. And $T_n \cap [1, 3x - n - 1]$ must be a maximum subset of $[1, 3x - n - 1]$ containing no solutions to $x + y = 3z$.

If x is odd then $|W|$ is even, so T_n must contain precisely one integer out of each pair of W and $T_n \cap [1, 3x-n-1] = \frac{3x-n-1}{2}$, unless $3x-n-1 = 4$. If $3x - n - 1 = 4$ then it would be possible to have $T_n \cap [1, 4] = \{1, 3, 4\}$ and T_n contain precisely one integer from all but one of the pairs and no integer from that one pair.

Subcase 2a. $\lceil \frac{3x+1}{2} \rceil \leq \frac{2}{3}n$.

If x is even then both integers of the pair $(\frac{3x}{2} - 1, \frac{3x}{2} + 1)$ are less than or equal to $\frac{2}{3}n$. Since x is the largest integer in T_n which is less than or equal to $\frac{2}{3}n$, neither integer in this pair is in T_n which is a contradiction.

If x is odd, neither integer in the pair $(\lfloor \frac{3x}{2} \rfloor, \lceil \frac{3x}{2} \rceil)$ can be in T_n, which again is a contradiction unless $3x - n - 1 = 4$. But since $n \geq 24$

$$\frac{3x+3}{2} = \frac{n+8}{2} \leq \frac{2}{3}n$$

so neither integer in the pair $(\lfloor \frac{3x}{2} \rfloor - 1, \lceil \frac{3x}{2} \rceil + 1)$ can be in T_n either, so Subcase 2a cannot occur.

Subcase 2b. Assume the two following inequalities:

$$\left\lceil \frac{3x+1}{2} \right\rceil > \frac{2}{3}n \tag{4}$$

and

$$3x - n > 23 \tag{5}$$

Since $T_n \cap [1, 3x-n-1]$ is a maximum subset of $[1, 3x-n-1]$ containing no solutions to $x+y=3z$ and since $23 \leq 3x-n-1 < n$, by the minimality of n as a counter-example, $T_n \cap [1, 3x-n-1]$ must be the set of all odd integers less than or equal to $3x-n-1$. By Lemma 1, T_n contains no even integer less than or equal to 68, so $n \geq 70$.

If x is even, then inequality (4) becomes

$$x > \frac{4n-6}{9}. \tag{6}$$

And $3x-n-1$ is odd so, by Lemma 1, T_n contains no even integers less than or equal to $3(3x-n-1)-1$, which by (6) is greater than $n-10$. If x is odd then (4) becomes

$$x > \frac{4n-3}{9} \tag{7}$$

and T_n contains no even integers less than or equal to $3(3x-n-2)-1$ which by (7) is also greater than $n-10$. So T_n contains at most five even integers, and hence there are at most five odd integers less than n which are not in T_n. Let $m_i = 2\lceil \frac{n}{6} \rceil + 2i - 1$ and $p_i = 3m_i - n$, $i = 1, 2, \ldots, 11$. It is easy to check that $\{m_i\}$ and $\{p_i\}$ are each sets of 11 distinct odd integers less than or equal to n and clearly not both m_i and p_i can be in T_n for any $i = 1, \ldots, 11$. Hence there are at least six odd integers less than n which are not in T_n, which shows Subcase 2b cannot occur.

Subcase 2c. Inequality (4) holds but (5) does not.

Hence

$$\frac{2}{3}n < \left\lceil \frac{3x+1}{2} \right\rceil \leq \frac{n}{2} + 12 \tag{8}$$

from which it follows that $n \leq 70$. So only a finite number of possibilities remain to be checked, and this could be done one by one (by hand or computer). We prefer to avoid this by considering the following possibilities.

2c(i). Assume (8) holds and also assume

$$\left\lceil \frac{n}{3} \right\rceil + 3 \leq x \leq \frac{n}{2} \tag{9}$$

If x is even, inequality (8) simplifies to

$$\frac{4n-6}{9} < x \leq \frac{n+22}{3} \tag{10}$$

And if x is even then, as we showed before, $|T_n \cap [1, 3x-n-1]| = \frac{3x-n}{2}$. By inequality (9), $3x - n - 1 \geq 9 > 5$, so $3x - n - 1 \in T_n$. Since $T_n \cap \left[x+1, \lfloor \frac{2n}{3} \rfloor\right] = \emptyset$, we must have $[3x - \lfloor \frac{2n}{3} \rfloor + 1, 2x - 1] \subseteq T_n$, since these integers are each paired with an excluded integer in the pairing of $[3x - n, n]$ we discussed before ($3x - \lfloor \frac{2n}{3} \rfloor$ might not be in T_n because it might be equal to $\frac{3x}{2}$). But $3(3x - n - 1)$ is a forbidden sum so $T_n \cap [7x - 3n - 2, 6x - 3n + \lfloor \frac{2n}{3} \rfloor - 4] = \emptyset$. Let $a = 7x - 3n - 2$, $b = 6x - 3n + \lfloor \frac{2n}{3} \rfloor - 4$, and $c = 3x - n - 1$. With x satisfying (9) and (10) it is easy to check that $a \geq 0$ and $c \geq 9$. Since $T_n \cap [a, b] = \emptyset$ and $c \in T_n$ we certainly have a contradiction if $a \leq c \leq b$. We also have a contradiction if $a \leq b < c$ and $b - a + 1 \geq c - b + 2$ because then

$$\begin{aligned} \frac{c+1}{2} = |T_n \cap [1, c]| &\leq |T_n \cap [1, a-1]| + (c - b) \\ &\leq \frac{a}{2} + c - b \\ &\leq \frac{c-1}{2}. \end{aligned}$$

Hence we have a contradiction if the conditions $a \leq c$ and $2b \geq a + c + 1$ are both satisfied, i.e., if

$$n - \left\lfloor \frac{2n}{3} \right\rfloor + 3 \leq x \leq \frac{2n+1}{4}.$$

But that is precisely our assumption in (9).
If x is odd then inequality (8) simplifies to

$$\frac{4n-3}{9} < \frac{n+23}{3} \tag{11}$$

and the argument is similar. In this case it turns out that $[3x - \lfloor \frac{2n}{3} \rfloor, 2x - 1] \subseteq T_n$ and that $9x - 3n - 3$ or $9x - 3n - 6$ is a forbidden sum, but we still get a contradiction with inequality (9).

2c(ii). Assume (8) holds and $x > \frac{n}{2}$.

If x is even there are eight ordered pairs of values for $n \geq 24$ and x which satisfy (8) when $x > \frac{n}{2}$. We list them as triples $(n, x, 3x - n - 1)$:

$(24, 14, 17)$	$(30, 16, 17)$
$(26, 14, 15)$	$(32, 18, 21)$
$(26, 16, 21)$	$(34, 18, 19)$
$(28, 16, 19)$	$(38, 20, 21)$

Since $|T_n \cap [1, 3x - n - 1]| = \frac{3x-n}{2}$ we see (by the table) that if $3x-n-1$ is equal to 21 or 15 then T_n must contain all the odd integers less than or equal to 15 and (by Lemma 1) no even integers at all. That eliminates four triples. Since $3x-n-1 \in T_n$ and x is the largest integer in T_n less than $\frac{2n}{3}$, we cannot have $x < 3x-n-1 \leq \frac{2n}{3}$. That eliminates three more triples, leaving only $(24, 14, 17)$. If $3x-n-1 = 17$, then $15 \in T_n$ (by the table). But $x = 14 < 15 \leq \frac{2}{3} \cdot 24$, so we get a contradiction here as well.

If x is odd there are ten such triples $(n, x, 3x - n - 1)$:

$$\begin{array}{ll} (24,13,14) & (30,17,20) \\ (24,15,20) & (32,17,18) \\ (26,15,18) & (34,19,22) \\ (28,15,16) & (36,19,20) \\ (28,17,22) & (40,21,22) \end{array}$$

We will show just the argument for $(32, 17, 18)$ here. Since $\lfloor \frac{2n}{3} \rfloor = 21$, we must have $T_n \cap [18, 21] = \emptyset$. Since $|T_n \cap [1, 18]| = 9$, by the table (or by Theorem 2) 17 must be in T_n. Also, $|T_n \cap [22, 32]|$ must be 7. Since 51 is a forbidden sum, at most one integer in each of the pairs $(22, 29)$, $(23, 28)$, $(24, 27)$, $(25, 26)$ is in T_n. By the table $15 \in T_n$, so $30 \in T_n$, which means $|T_n \cap [22, 32]| \leq 6$ a contradiction. The other nine triples can be disposed of with similar (and mostly simpler) arguments.

2c(iii). Assume (8) holds and $x < \lceil \frac{n}{3} \rceil + 3$.

The only possibility here is $n = 28$ and $x = 12$. Then $3x - n = 8$, so $|T_n \cap [1, 7]| = 4$ and T_n contains precisely one member of each pair $(18 - i, 18 + i)$ $i = 1, 2, \ldots, 10$. Since $\lfloor \frac{2n}{3} \rfloor = 18$, $T_n \cap [13, 18] = \emptyset$, so $[19, 23] \subseteq T_n$. But 21 is a forbidden sum, so neither 1 nor 2 can be in T_n. This is a contradiction since (by the table) if $|T_n \cap [1, 7]| = 4$, either 1 or 2 must be in T_n. □

5. Related Problems and Remarks

For k a positive integer not equal to 2, let $f(n, k)$ be the maximum size of a subset S of $\{1, \ldots, n\}$ such that there are no solutions to $x + y = kz$ with x, y, z (not necessarily distinct) integers in S (the problem does not make sense for $k = 2$). In this paper we determined $f(n, 1)$ and $f(n, 3)$ for all n and found all maximum such subsets. The determination of $f(n, k)$ when $k \geq 4$ has a very different flavor, and we have some results in this direction [2].

Let $g(t)$ be the maximum "size" (appropriately defined) of a subset S of the closed interval $[0, 1]$ having no solutions to $x + y = tx$ where t is a fixed positive number. Finding $g(t)$ is a continuous analog of finding $f(n, k)$. It turns out there is a strong connection between these problems if $k \geq 4$, but not for $k = 3$. For $k = 1$ we remark that the maximum set $\left[\lfloor \frac{n+1}{2} \rfloor, n\right]$ of Theorem 1 does have a continuous analog, while the set of all odd integers less than or equal to n does not.

In [2] we show that if $k \geq 3$ then the positive integers can be partitioned into finitely many subsets each of which has no solutions to $x + y = kz$.

For k any positive integer, let $f^*(n, k)$ be the maximum size of a subset of $\{1, \ldots, n\}$ having no solutions to $x + y = kz$ where x, y, z must be distinct. Clearly, $f(n, k) \leq f^*(n, k)$. It is easy to show that $f^*(n, 1) = \lceil \frac{n+1}{2} \rceil$. There are many values of n for which $f(n, 3)$ is smaller than $f^*(n, 3)$. For example, the set $[1, 4] \cup [12, 18]$ shows that $f^*(n, 3) \geq 11$. However, we join Paul Erdős in conjecturing that $f^*(n, 3) = f(n, 3) = \lceil \frac{n}{2} \rceil$ for sufficiently large n. We also suspect there is a unique maximum set for sufficiently large n. One could do some computer work to obtain some information as to the likelihood of this conjecture being correct. To get a proof one could follow the lines of our proofs in this paper. Unfortunately, the minor inconvenience of $f(4, 3)$ being greater than 2 would become the major headache of $f^*(n, 3)$ being greater than $\lceil \frac{n}{2} \rceil$ for many small values of n.

Of course the problem of narrowing the bounds for $f^*(n, 2)$ is one of the most intriguing problems in combinatorics.

Reference

1. F. A. Behrend, On sets of integers which contain no three in arithmetic progression, *Proc. Nat. Acad. Sci.* **23** (1946), 331–332.
2. F. R. K. Chung and J. L. Goldwasser, Maximum subsets of $(0, 1]$ with no solutions to $x + y = kz$, *Elec. J. of Comb.* **3** (1996), R1, 23pp.
3. D. R. Heath-Brown, Integer sets containing no arithmetic progressions, *The Journal of the London Math. Soc.* **35** (1987), 385–394.
4. K. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.
5. I. Schur, Uber die Kongruenz $x^m + y^m = z^m$ (mod p), *J. ber. Deutch. Math. Verein.* **25** (1916), 114–116.
6. R. Salem and D. C. Spencer, On sets of integers which contain no three terms in arithmetical progressions, *Proc. Nat. Acad. Sci.* USA **28** (1942), 561–563.
7. E. Szemerédi , On sets of integers containing no four elements in arithmetic progression, *Acta. Arith.* **20** (1969), 89–104.
8. E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta. Arith.* **27** (1975), 199–245.
9. B. L. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk* **15** (1927), 212–216.

On Primes Recognizable in Deterministic Polynomial Time

Sergei Konyagin* and Carl Pomerance**

S. Konyagin
Steklov Institute of Mathematics, 119991, Russia
e-mail: konyagin23@gmail.com

C. Pomerance (✉)
Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA
e-mail: carl.pomerance@dartmouth.edu

For Paul Erdős on his eightieth birthday

Summary. We discuss some simple deterministic algorithms that establish primality for a robust set of primes in polynomial time. The first 6 sections comprise the intact original article published in the first edition of this volume in 1997. The last 2 sections discuss developments in this fast-moving field to early 2013, and refer to the prior sections in the past tense. The bibliography for the original article and the new update have been combined.

1. Introduction

In this paper we present several algorithms that can find proofs of primality in deterministic polynomial time for some primes. In particular we show this for any prime p for which the complete prime factorization of $p-1$ is given. We can also show this when a completely factored divisor of $p-1$ is given that exceeds $p^{1/4+\varepsilon}$. And we can show this if $p-1$ has a factor F exceeding p^{ε} with the property that every prime factor of F is at most $(\log p)^{2/\varepsilon}$. Finally, we present a deterministic polynomial time algorithm that will prove prime more than $x^{1-\varepsilon}$ primes up to x. The key tool we use is the idea of a smooth number, that is, a number with only small prime factors. We show an inequality for their distribution that perhaps has independent interest.

It is known that if one assumes the Riemann hypothesis for Dirichlet L-functions, then the prime recognition problem is in the complexity class P. Thus, from Miller and Bach we know that for every odd composite number n there is some integer a in the range $1 < a < \min\{n, 2(\log n)^2\}$ such that n is

* Supported in part by the Cultural Initiative Fund and the Russian Academy of Natural Sciences

** Supported in part by an NSF grant

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_12,

not a strong probable prime to the base a, and so n is proved composite. If an odd number n is a strong probable prime to every base a in the above range, then n is prime. Thus assuming the above extended Riemann hypothesis, every prime p can be deterministically supplied with a proof of its primality in $O((\log p)^3)$ arithmetic steps with integers at most p.

The results in this paper do not rely on the truth of any unproved hypotheses.

It has been known since Lucas that it is easy to find a proof of primality for a prime p if the complete factorization of $p-1$ is known. Indeed one merely has to present a primitive root for p and prove it is one using the prime factorization of $p-1$. Though we know no fast deterministic algorithm for finding a primitive root for a prime p, the probabilistic method of just choosing random integers until a primitive root is found works very well in practice. The expected number of tries is $O(\log\log p)$. In fact, one can show that the expected number of random choices to find a set of numbers which generate $(\mathbb{Z}/p\mathbb{Z})^*$ as a group is $O(1)$. (Note also that it is a simple matter to deterministically fashion a primitive root out of a set of generators with knowledge of the complete prime factorization of $p-1$.) Our algorithm requires $O((\log p)^{10/7})$ tries, and does not guarantee that it will find a primitive root or a set of generators, but it does prove primality and it is deterministic.

It is also known (see Brillhart, Lehmer, Selfridge [8]) that if one has a fully factored divisor F of $p-1$, where $F > p^{1/3}$, then one can quickly decide if p is prime or composite. Again, this involves choosing numbers at random. We show how the prime or composite nature of p can be decided deterministically and in polynomial time. In addition, we only require $F > p^{1/4+\varepsilon}$. In another algorithm we only need $F > p^{\varepsilon}$, but for the method to be fast, F must be smooth.

While many of the algorithms in this paper are only of theoretical interest, it is likely that at least some of the ideas have practical value. In particular, an algorithm we present below which allows one to decide whether n is prime or composite, when it is known that all prime factors of n are $1 \mod F$ with $F \geq n^{3/10}$, should be a practical addition to the Brillhart, Lehmer, Selfridge "$n-1$ test".

It is to be expected that some of the ideas presented here would be of use in the "$n+1$ test" and the combined "n^2-1 test". These elementary tests are often used in conjunction with the Jacobi sums test (see [6] and references there), and it is possible that a few ideas presented here will be of use in that context as well. However, as stated above, our primary emphasis in this paper is theoretical and not practical.

Let $\psi(x,y)$ denote the number of integers $n \leq x$ free of prime factors exceeding y. In [15], Erdős and van Lint show that in some sense $\psi(x,y)$ can be approximated by the binomial coefficient $\binom{\pi(y)+[u]}{[u]}$ where $u = \log x/\log y$ and $\pi(y)$ denotes the number of primes not exceeding y. In fact an elementary combinatorial argument shows that $\psi(x,y) \geq \binom{\pi(y)+[u]}{[u]}$, so one side of the approximation is easy. In this paper we obtain the lower bound $x/(\log x)^u = x^{1-\log\log x/\log y}$, which is valid whenever $x \geq 4$ and $2 \leq y \leq x$. This lower bound for $\psi(x,y)$ is attractive for its simplicity and near universality. However, one should note that it is a good approximation to $\psi(x,y)$ only in the range $(\log x)^{1+\varepsilon} \leq y \leq \exp((\log x)^{\varepsilon})$. The inequality in the special case $y = (\log x)^2$ was previously established by Lenstra [19], and for a similar purpose.

Our main idea in this paper is to build up a large subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$ using a small set of generators. Specifically, if p is the least prime factor of n and a is an integer with $1 < a < p$, let $\mathcal{G}_n(a)$ denote the subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$ generated by $j \mod n$ for $j = 2, 3, \ldots, a$. From the above estimate for $\psi(x,y)$, we have

$$\#\mathcal{G}_n([(\log n)^c]) \geq \psi(n, (\log n)^c) > n^{1-1/c},$$

whenever $n \geq 4$ and $2 \leq (\log n)^c < p$, with p the least prime factor of n. Thus we can create an "exponentially large" subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$ with a "polynomially sized" set of generators. The idea of using smooth number estimates to show that one has built up a large subgroup of $(\mathbb{Z}/p\mathbb{Z})^*$ for p prime was first used in 1926 by Vinogradov [28] to estimate the least positive residue mod p that is not a k-th power.

It was previously shown by Pintz, Steiger and Szemerédi [22] that there are infinitely many primes p that can be proved prime in deterministic polynomial time. They require for their primes p that $p-1$ has a divisor which is a power of 3 and exceeds $p^{1/3}$. Thus they could only show there are more than $x^{2/3-\varepsilon}$ such primes up to x. As mentioned above, we replace the "2/3" with 1.

Our result in Sect. 3 on deciding if n is prime or composite in deterministic polynomial time, when the complete prime factorization of $n-1$ is given, was anticipated by Fellows and Koblitz [16], though their algorithm is not as fast as ours.

We mention a few other results that are somewhat relevant. Adleman and Huang [1] have given a probabilistic algorithm for primality proving that has expected polynomial time. Much earlier, Solovay and Strassen [27] had given a probabilistic algorithm for compositeness proving that has expected polynomial time. In [23], the second author showed that for every prime p there is a proof that p is prime that can be verified in $O(\log p)$ arithmetic steps with integers at most p. Previously, Pratt [25] had shown via Lucas's test the existence of a primality proof that requires $O((\log p)^2)$ arithmetic steps. These two papers show only the existence of these proofs; they do not show how to find them quickly.

We wish to thank W. R. Alford, Ronald Burthe, Andrew Granville, Hendrik Lenstra, and Jeff Shallit for some helpful remarks.

2. A Lower Bound for the Distribution of Smooth Numbers

We say an integer n is y-smooth if no prime factor of n exceeds y. Let $\psi(x,y)$ denote the number of integers n in $[1,x]$ that are y-smooth. In this section we are going to prove the following theorem.

Theorem 1. *If $x \ge 4$ and $2 \le y \le x$, then $\psi(x,y) > x^{1-\log\log x/\log y}$.*

We begin with a few lemmas. Let $\pi(x)$ denote the number of primes p with $p \le x$.

Lemma 1. *For $x \ge 37$ we have $\pi(x) - \pi(x^{1/2}) > (7/9)x/\log x$.*

This lemma follows from Rosser and Schoenfeld [26, Theorem 1] and a simple calculation. The next lemma is well known; we give the proof for completeness.

Lemma 2. *Let p_k denote the kth prime. For $x \ge 1$ we have*

$$\psi(x,p_k) > \frac{(\log x)^k}{k!} \prod_{j=1}^{k} \frac{1}{\log p_j}.$$

Proof. An integer $n \le x$ which is p_k-smooth has its prime factorization in the form $p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ where $a_1, a_2, \dots, a_k$ are non-negative integers and $\sum a_j \log p_j \le \log x$. Thus $\psi(x,p_k)$ is the number of lattice points $(a_1, a_2, \dots, a_k) \in \mathbb{Z}^k$ with each $a_j \ge 0$ and $\sum a_j \log p_j \le \log x$. Putting each such lattice point at the "lower left" corner of a unit cube with edges parallel to the axes, a region is described which is strictly larger than the simplex

$$\left\{ (y_1, \dots, y_k) \in \mathbb{R}^k : \text{ each } y_j \ge 0, \sum_{j=1}^{k} y_j \log p_j \le \log x \right\}.$$

Thus $\psi(x,p_k)$ exceeds the k-dimensional volume of this simplex, which is

$$(\log x)^k (k!)^{-1} \prod_{j=1}^{k} (\log p_j)^{-1}. \qquad \square$$

Proof of Theorem 1. We verify the theorem directly for pairs x, y with $2 \le y < 37$ and $x < 120$. Assume now that $2 \le y < 37$ and $x \ge 120$. Since the theorem is trivial when $\log y < \log\log x$, we may assume $y \ge 3$. It is not hard to show that

$$x^{1-\log\log x/\log 37} < \frac{(\log x)^2}{\log 3 \log 4} \tag{1}$$

for $x \geq 120$. But the left side of (1) is greater than $x^{1-\log\log x/\log y}$ and the right side of (1) is less than $\psi(x,3)$ by Lemma 2. Since $\psi(x,3) \leq \psi(x,y)$, the theorem holds in this range.

Now assume $37 \leq y \leq x$. Let $u = \log x/\log y$ and let $\{u\} = u - [u]$ denote the fractional part of u. If m is a positive integer with $m \leq y^{\{u\}}$ and n is a product of $[u]$ not necessarily distinct primes in the interval $(y^{1/2}, y]$, then $N = mn$ is y-smooth and $N \leq y^{\{u\}}y^{[u]} = y^u = x$. Moreover, since m has at most one prime factor in $(y^{1/2}, y]$, it follows that the number of representations of N as a product mn in this way is at most $[u]+1$. In fact, if $\{u\} \leq 1/2$, then N has at most one representation as mn. We conclude that

$$\psi(x,y) \geq \begin{cases} [y^{\{u\}}](\pi(y)-\pi(y^{1/2}))^{[u]}/([u]+1)!, & \text{if } \{u\} > 1/2 \\ [y^{\{u\}}](\pi(y)-\pi(y^{1/2}))^{[u]}/[u]!, & \text{if } \{u\} \leq 1/2. \end{cases} \tag{2}$$

Note that using $y \geq 37$, we have

$$\left[y^{\{u\}}\right] > \begin{cases} \frac{6}{7}y^{\{u\}}, & \text{if } \{u\} > 1/2 \\ \frac{1}{2}y^{\{u\}}, & \text{if } \{u\} \leq 1/2. \end{cases} \tag{3}$$

Also, from Lemma 1, we have

$$(\pi(y)-\pi(y^{1/2}))^{[u]} > \left(\frac{7}{9}\cdot\frac{y}{\log y}\right)^{[u]} = \left(\frac{7u}{9}\right)^{[u]}\frac{y^{[u]}}{(\log x)^{[u]}}.$$

Thus

$$\begin{aligned} y^{\{u\}}(\pi(y)-\pi(y^{1/2}))^{[u]} &> \left(\frac{7u}{9}\right)^{[u]}(\log x)^{\{u\}}\frac{y^u}{(\log x)^u} \\ &= \left(\frac{7u}{9}\right)^{[u]}(\log x)^{\{u\}}x^{1-\log\log x/\log y}. \end{aligned}$$

Using this inequality with (2) and (3) we have that the theorem will hold if we show

$$\left(\frac{7u}{9}\right)^{[u]}(\log x)^{\{u\}} \geq \begin{cases} \frac{7}{6}([u]+1)!, & \text{if} \{u\} > 1/2 \\ 2[u]! & \text{if} \{u\} \leq 1/2. \end{cases} \tag{4}$$

We now show

$$\left(\frac{7k}{9}\right)^{k} > 2(k+1)! \text{ for every integer } k \geq 6. \tag{5}$$

This holds by inspection for $k = 6, 7, 8, 9$. For any non-negative integer k, the arithmetic-geometric mean inequality implies that

$$\left(\frac{k+2}{2}\right)^{k+1} \geq (k+1)!.$$

Using this and the easily verified inequality

$$\left(\frac{7k}{9}\right)^{k} > 2\left(\frac{k+2}{2}\right)^{k+1} \text{ for } k \geq 10,$$

we have (5). Note that (5) implies (4) when $u \geq 6$.

Suppose that $3 \leq u < 6$ and $\{u\} \leq 1/2$. We verify for $k = 3, 4, 5$ that

$$\left(\frac{7k}{9}\right)^{k} > 2k!,$$

so that (4) holds for these values of u.

Suppose now that $2.5 < u < 6$ and $\{u\} > 1/2$. We verify for $k = 2, 3, 4, 5$ that

$$\left(\frac{7(k+1/2)}{9}\right)^{k} \left(\log(37^{k+1/2})\right)^{1/2} > \frac{7}{6}(k+1)!,$$

so that (4) holds for these values of u. (We use that $x = y^u \geq 37^u$.)

Now suppose $2.2 \leq u \leq 2.5$. We have

$$\left(\frac{7u}{9}\right)^{[u]} (\log x)^{\{u\}} \geq \left(\frac{7(2.2)}{9}\right)^{2} (\log(37^{2.2}))^{0.2} > 4 = 2[u]!,$$

so the theorem holds here too.

In the range $1 \leq u < 2.2$ we use another estimate for $\psi(x, y)$. The number of integers up to x divisible by a prime p is $[x/p]$. Thus

$$\psi(x, y) \geq [x] - \sum_{y<p\leq x} \left[\frac{x}{p}\right] > x - 1 - x \sum_{y<p\leq x} \frac{1}{p} \tag{6}$$

where p runs over primes.

First assume that $1.6 \leq u < 2.2$. Then $x \geq 37^{1.6} > 286$. It follows from Theorem 5 in Rosser and Schoenfeld [26] that

$$\sum_{y<p\leq x} \frac{1}{p} < \log\log x - \log\log y + \frac{1}{2(\log x)^2} + \frac{1}{2(\log y)^2}$$

$$\leq \log 2.2 + \frac{1}{2(\log(37^{1.6}))^2} + \frac{1}{2(\log 37)^2} < 0.85.$$

Thus from (6) we have

$$\psi(x, y) > 0.15x - 1 > 0.14x.$$

But

$$x^{1-\log\,\log\,x/\log\,y} = \frac{x}{(\log\,x)^u} \leq \frac{x}{(\log(37^{1.6}))^{1.6}} < 0.07x,$$

so the theorem holds in this range.

Finally assume $1 \leq u < 1.6$. Then from Theorem 5 and its Corollary in [26] we have that

$$\sum_{y<p\leq x} \frac{1}{p} < \log\,\log\,x - \log\,\log\,y + \frac{1}{(\log\,x)^2} + \frac{1}{2(\log\,y)^2}$$

$$\leq \log\,1.6 + \frac{1}{(\log\,37)^2} + \frac{1}{2(\log\,37)^2} < 0.59,$$

so that from (6) we have

$$\psi(x, y) > 0.41x - 1 > 0.38x.$$

But

$$x^{1-\log\,\log\,x/\log\,y} \leq \frac{x}{\log\,37} < 0.28x,$$

so we have the theorem here as well. This concludes the proof of Theorem 1. □

3. When $n - 1$ Is Fully Factored

In this section we present and analyze two deterministic algorithms that will decide if a positive integer n is prime or composite when the complete prime factorization of $n-1$ is known. The first algorithm uses the Brillhart, Lehmer, Selfridge "$n-1$ test" (see [8]). The second algorithm is somewhat faster and uses a new result presented below.

We begin with a factorization algorithm that is very fast, but unfortunately is usually unsuccessful in factoring composite numbers.

The base B factorization method. We are input integers n, B with $n > B \geq 2$. This algorithm attempts to find a nontrivial factorization of n.

Step 1 Write n in the base B : $n = c_d B^d + c_{d-1} B^{d-1} + \ldots + c_0$, where $c_0, \ldots, c_d$ are integers in the interval $[0, B-1]$ and $c_d > 0$.

Step 2 Compute $c = \gcd(c_0, \ldots, c_d)$. If $c > 1$, return c as a proper factor of n and stop.

Step 3 Factor $f(x) = c_d x^d + \ldots + c_0$ into irreducible polynomials in $\mathbb{Z}[x]$ with the algorithm of [18].

Step 4 If $f(x)$ is irreducible in $\mathbb{Z}[x]$, the algorithm has failed, so stop. If $f(x) = g_1(x)g_2(x)\ldots g_k(x)$ where each $g_i(x)$ is irreducible in $\mathbb{Z}[x]$, then return $g_1(B)g_2(B)\ldots g_k(B)$ as a nontrivial factorization of n and stop.

That each $g_i(B)$ is a proper factor of n in Step 4 follows from [7]. Thus the algorithm is correct. From the analysis in [18], it follows that the running time of the algorithm is $(\log n)^{O(1)}$.

We shall only be applying the base B factorization method in the cases $d = 2, 3$ and in these cases it should be considered "overkill" to use the algorithm of [18] to factor $f(x)$ in Step 3. In particular, if $d = 2$, then $c_2x^2 + c_1x + c_0$ factors if and only if $c_1^2 - 4c_0c_2$ is a square, in which case it is trivial to write down the factorization. Further, it is easy to detect squares and take square roots of squares with a binary search. Thus the time for Step 3 in the case $d = 2$ is $O(\log n)$ arithmetic steps with integers at most n. (Newton's method is even better than a binary search; its complexity is $O(\log\log n)$ arithmetic steps with integers at most n.)

When $d = 3$ we can again use a binary search in Step 3. In particular, $f(x)$ factors if and only if it has a rational root, and if one rational root is found, we can reduce the problem to the quadratic case. It is more convenient to replace $f(x)$ with $c_3^2 f(x) = g(c_3x)$, since $g(x)$ factors if and only if it has an integer root. However, every integer root of g divides $g(0)$, so if $g(0) \neq 0$, then every integer root is in the interval $[-|g(0)|, |g(0)|]$ and may be located with essentially a binary search. Thus again Step 3 can be accomplished in $O(\log n)$ arithmetic steps with integers at most n. (Note that Newton's method could be applied here as well.)

We now describe an algorithm based on the Brillhart, Lehmer, Selfridge $n - 1$ test.

Algorithm 1. *We are input an integer $n > 4$ and the complete prime factorization of $n - 1$. This deterministic algorithm decides if n is prime or composite.*

Let $F(1) = 1$. For $a = 2, 3, \ldots, [(\log n)^{3/2}]$ do the following:

Step 1 *If a is composite, let $F(a) = F(a-1)$ and go to Step 7. If $a^{F(a-1)} \equiv 1 \mod n$, let $F(a) = F(a-1)$ and go to Step 7. Verify that $a^{n-1} \equiv 1 \mod n$. If not, declare n composite and stop.*

Step 2 *Using the prime factorization of $n - 1$, find the least positive divisor $E(a)$ of $n - 1$ with $a^{E(a)} \equiv 1 \mod n$.*

Step 3 *Verify that $(a^{E(a)/q} - 1, n) = 1$ for each prime factor q of $E(a)$. If not, declare n composite and stop.*

Step 4 *Let $F(a) = \mathrm{lcm}\{F(a-1), E(a)\}$. Compute $F(a)$.*

Step 5 *If $F(a) \geq n^{1/2}$, declare n prime and stop.*

Step 6 *If $n^{1/3} \leq F(a) < n^{1/2}$, attempt to factor n by the base $F(a)$ factorization method. If n is factored nontrivially, declare n composite and stop. If n is not factored, declare n prime and stop.*

Step 7 *If $a < [(\log n)^{3/2}]$, get the next a. Otherwise declare n composite and stop.*

Proof of correctness. Since $(\log n)^{3/2} < n$ for every integer $n > 1$, Step 1 is correct by Fermat's little theorem. It is clear that Step 3 is correct from the definition of $E(a)$.

Suppose r is a prime factor of n and we have reached Step 4 of the algorithm for a particular a. Consider the subgroup $\mathcal{G}_r(a)$ of $(\mathbb{Z}/r\mathbb{Z})^*$ defined in the Introduction. We shall show that $\#\mathcal{G}_r(a) = F(a)$. For each prime j with $j \leq a$ we have $j^{F(a)} \equiv 1 \mod n$, so that $j^{F(a)} \equiv 1 \mod r$. Thus $\#\mathcal{G}_r(a) | F(a)$. Further, if j is a prime with $j \leq a$ and $F(j) > F(j-1)$, then from Step 3 the order of j in $(\mathbb{Z}/r\mathbb{Z})^*$ is $E(j)$. Since $F(a)$ is the least common multiple of those numbers $E(j)$ with j prime, $j \leq a$, and $F(j) > F(j-1)$, we have $F(a) | \#\mathcal{G}_r(a)$. Thus $\#\mathcal{G}_r(a) = F(a)$, as asserted.

We conclude that if we have reached Step 4 of the algorithm for a particular a, then for each prime factor r of n we have $r \equiv 1 \mod F(a)$. Thus the correctness of Steps 5 and 6 follows from [8].

Suppose now that $a = [(\log n)^{3/2}]$ and we have reached Step 7. Thus $F(a) < n^{1/3}$. Suppose n is prime. For every a-smooth integer m in the range $1 \leq m \leq n$ we have $m \mod n \in \mathcal{G}_n(a)$. Thus

$$F(a) = \#\mathcal{G}_n(a) \geq \psi(n, a) = \psi(n, (\log n)^{3/2}) > n^{1/3},$$

where the last inequality follows from Theorem 1 and the fact that $(\log n)^{3/2} > 2$ for $n > 4$. This is a contradiction and so Step 7 is correct.

We conclude that Algorithm 1 is correct. □

Analysis of runtime. We measure the runtime by the number of arithmetic steps with integers no larger than n. By an arithmetic step we mean addition, subtraction, multiplication, division with remainder, greatest common divisor, and finding an inverse for a member of $(\mathbb{Z}/n\mathbb{Z})^*$. Using naive arithmetic, an arithmetic step can be accomplished in $O((\log n)^2)$ bit operations. Using the FFT, an arithmetic step can be accomplished in $O_\varepsilon((\log n)^{1+\varepsilon})$ bit operations for each $\varepsilon > 0$.

One can use the sieve of Eratosthenes to prepare a list of all of the primes up to $(\log n)^{3/2}$ in time $O((\log n)^{3/2} \log \log n)$. For each prime number a, Step 1 can be accomplished in $O(\log n)$ arithmetic steps. Since the number of such primes is $O((\log n)^{3/2}/\log \log n)$, an upper bound for the time spent in Step 1 is $O((\log n)^{5/2}/\log \log n)$.

To do Step 2 we use a variation of the algorithm of [9]. First consider the case where $n-1$ is squarefree; say $n - 1 = q_1 \dots q_k$ with $q_1, \dots, q_k$ distinct primes. Then to find $E(a)$ it suffices to find the set of q_i which divide $E(a)$. But $q_i | E(a)$ if and only if $a^{(n-1)/q_i} \not\equiv 1 \mod n$. The algorithm of [9] computes all of the residues $a^{\prod_{j\neq i} q_j} \mod n = a^{(n-1)/q_i} \mod n$ for $i = 1, \dots, k$. It breaks the computation into steps where at a particular step we are taking a residue $x \mod n$ and computing $x^{q_j} \mod n$ for some j. Each q_j is used $O(\log(k+1))$ times, so that the total number of arithmetic operations with integers at most n is

$$O\left(\sum_{j=1}^{k} \log\, q_j\, \log(k+1)\right) = O(\log\, n\, \log(k+1)).$$

Now consider the general case where we no longer assume that $n-1$ is squarefree. Say $n-1 = q_1^{\alpha_1} \dots q_k^{\alpha_k}$ with the q_i's distinct primes and the α_i's positive integers. We have $q_i^{\beta_i} || E(a)$ if and only if β_i is the least non-negative integer with $a^{(n-1)q_i^{\beta_i-\alpha_i}} \equiv 1 \mod n$. To compute the β_i's we combine the ideas from the squarefree case with a binary search. In the first step, we let $m_1 = q_1^{[\alpha_1/2]} \dots q_k^{[\alpha_k/2]}$ and let $a_1 = a^{m_1} \mod n$. We use the algorithm of [9] with a_1 and the numbers $q_i^{\alpha_i - [\alpha_i/2]}$ to decide if $\beta_i \le [\alpha_i/2]$ or $[\alpha_i/2] < \beta_i \le \alpha_i$ for each $i = 1, \dots, k$. In the first case we replace $q_i^{[\alpha_i/2]}$ in m_1 with $q_i^{[\alpha_i/4]}$. In the second case we replace $q_i^{[\alpha_i/2]}$ in m_1 with $q^{\alpha_i - [\alpha_i/4]}$. Thus we have a number m_2, we form $a_2 = a^{m_2} \mod n$ and again we use the algorithm of [9], this time with the q_i's raised to exponents about $\alpha_i/4$. Continuing in this fashion we compute the β_i's and thus $E(a)$. In the l-th step of this algorithm we are using the algorithm of [9] with numbers whose product is about $n^{2^{-l}}$. Thus the number of arithmetic operations for the l-th step is $O(2^{-l} \log n \log(k+1))$. Moreover, we can compute a_l from a_{l-1} in $O(2^{-l} \log\, n)$ arithmetic operations. Thus summing over l, the number of steps for Step 2 for a particular value of a is $O(\log\, n \log(k+1))$. The number of values of a for which we perform Step 2 is $O(\log\, n)$. (To see this, note that we only perform Step 2 when $F(a) > F(a-1)$ and that $F(a)$ is the product of the integers $F(j)/F(j-1)$ for $2 \le j \le a$.) Since $k = O(\log\, n)$, we have that the total time spent in Step 2 is $O((\log\, n)^2 \log \log\, n)$ arithmetic steps with integers at most n.

For Step 3, note that to compute the greatest common divisor it is sufficient to work with $a^{E(a)/q} \mod n$ rather than $a^{E(a)/q}$. If $E(a) = q_1^{\beta_1} \dots q_{k'}^{\beta_{k'}}$ where $q_1, \dots, q_{k'}$, are distinct primes and $\beta_1, \dots, \beta_{k'}$, are positive integers, let $a_1 = \prod a^{q_i^{\beta_i - 1}} \mod n$. We use the algorithm of [9] to compute $a_1{}^{\prod_{j\neq i} q_j} \mod n = a^{E(a)/q_i} \mod n$ for each i. The number of steps is $O(\log\, E(a) \log(k'+1)) = O(\log\, n \log(k+1))$, where $k \ge k'$ is the number of distinct prime factors of $n-1$. Thus as with Step 2, the total time spent in Step 3 is $O((\log\, n)^2 \log\log\, n)$ arithmetic steps with integers at most n.

Steps 4, 5, and 7 are each $O(1)$ arithmetic steps for each a and, as remarked above, Step 6 is $O(\log\, n)$ arithmetic steps for each a. Note that we visit Steps 4, 5, and 6 for $O(\log n)$ values of a and we visit Step 7 for $O((\log n)^{3/2})$ values of a. Thus the total time for all of these steps is $O((\log n)^2)$ arithmetic steps with integers at most n.

We conclude that in the worst case, Algorithm 1 runs in $O((\log n)^{5/2}/\log\log n)$ arithmetic steps with integers at most n. We have proved the following theorem.

Theorem 2. *Given an integer $n > 4$ and the complete prime factorization of $n-1$, Algorithm 1 correctly decides if n is prime or composite. Further, Algorithm 1 uses at most $O((\log n)^{5/2}/\log\log n)$ arithmetic steps with integers at most n.*

We remark that in some cases when Algorithm 1 declares n composite, a nontrivial factorization of n may also be found. In particular, this is true in Steps 3 and 6. However most composite inputs will be proved composite in Step 1 with $a = 2$, in which case no nontrivial factorization of n is produced.

The next algorithm may be considered an extension of the Brillhart, Lehmer, Selfridge $n-1$ test. We shall use it as a subroutine in an improved version of Algorithm 1 we present below.

Algorithm 2. *This deterministic algorithm finds the complete prime factorization of n when input integers n, F such that $F \geq n^{3/10} > 1$ and each prime factor of n is $1 \mod F$.*

Step 1 *If $n \leq 243$, factor n by trial division and stop.*

Step 2 *If $F \geq n^{1/3}$, use the method of [8] and stop. (That is, if $F \geq n^{1/2}$, declare n prime; if $n^{1/3} \leq F < n^{1/2}$, use the base F factorization method to factor n. Note that if the base F factorization succeeds in factoring n, then it produces the prime factorization of n, while if it fails, then n is prime.)*

Step 3 *We have $n^{3/10} \leq F < n^{1/3}$. Attempt to factor n by the base F factorization method. If this succeeds in splitting n, report it as the complete prime factorization of n and stop. Let c_1, c_2, c_3 be the base F "digits" of n, so that $n = c_3F^3 + c_2F^2 + c_1F + 1$.*

Step 4 *Let $c_4 = c_3F + c_2$ so that $n = c_4F^2 + c_1F + 1$. If either $c_4x^2 + c_1x + 1$ or $(c_4 - 1)x^2 + (c_1 + F)x + 1$ are reducible in $\mathbb{Z}[x]$, this may lead to a factorization of n as in the base F factorization method. If so, report this factorization as the prime factorization and stop.*

Step 5 *Develop the continued fraction for c_1/F and let $u/v, u'/v'$ be consecutive convergents with $v < F^2/\sqrt{n} \leq v'$. Let $u_0 = \pm u'$, $v_0 = \pm v'$ be such that $uv_0 + u_0v = 1$.*

Step 6 *For each integer d with $|d - c_4v/F| < 2n^{3/2}/F^5$ (≤ 2) do the following. Find all integral roots s of the polynomial*

$$f_d(x) = y^3 - c_1y^2 + c_4y + Fzy + z$$

where $y = dv_0 + vx, z = -du_0 + ux$. For any integral root s found with $(dv_0+vs)F+1$ a nontrivial factor of n, report this number and its cofactor in n as the prime factorization of n and stop. If n is not split in this step, then declare n prime and stop.

Proof of correctness. We first show that if n is factored in Step 3, then this step produces the complete prime factorization of n. First, it is clear that n has at most three prime factors. Thus if $f(x) = c_3x^3 + c_2x^2 + c_1x + 1$ factors into three linear factors in $\mathbb{Z}[x]$, then these give, upon substituting F for x, the prime factorization of n. Suppose conversely that n has three prime factors. Thus there are positive integers a_1, a_2, a_3 with $n = (a_1F + 1)(a_2F+1)(a_3F+1)$. Since $n > 243$ we have $F^4 > 3n$, so that $a_1a_2a_3 < F/3$. Thus $a_1a_2 + a_1a_3 + a_2a_3 \le 3a_1a_2a_3 < F$ and $a_1 + a_2 + a_3 < F$. We conclude that $c_3 = a_1a_2a_3,\ c_2 = a_1a_2 + a_1a_3 + a_2a_3,\ c_1 = a_1 + a_2 + a_3$ and that $f(x) = (a_1x+1)(a_2x+1)(a_3x+1)$. That is, the base F factorization method will find the complete prime factorization of n.

Suppose now that n has exactly two prime factors so that there are positive integers a_1, a_2 with $n = (a_1F + 1)(a_2F + 1)$. Assume $a_1 \le a_2$. If we obtain any nontrivial splitting of n in any step of the algorithm, evidently this gives the complete prime factorization of n. We now show that if n has not been factored in Steps 3 and 4 of the algorithm and if n is composite, then it will be factored in Step 6.

Since $n = c_4F^2 + c_1F + 1 = a_1a_2F^2 + (a_1 + a_2)F + 1$, there is some integer $t \ge 0$ with

$$a_1a_2 = c_4 - t,\ a_1 + a_2 = c_1 + tF. \tag{7}$$

From the failure of Step 4 to find a_1, a_2, we have $t \ge 2$. Thus

$$a_2 \ge \frac{a_1 + a_2}{2} \ge \frac{c_1 + 2F}{2} \ge F \tag{8}$$

and

$$a_1 < \frac{n}{a_2F^2} \le \frac{n}{F^3}. \tag{9}$$

We have from (7) that

$$t \le \frac{a_1 + a_2}{F} \le \frac{a_1a_2 + 1}{F} < \frac{c_4}{F} < \frac{n}{F^3}. \tag{10}$$

Also (7) gives us the equation

$$a_1c_1 + a_1tF = a_1^2 + c_4 - t. \tag{11}$$

From the elementary theory of continued fractions we have

$$\left|\frac{c_1}{F} - \frac{u}{v}\right| \le \frac{1}{vv'} \le \frac{\sqrt{n}}{vF^2}. \tag{12}$$

Using (11) we have

$$\begin{aligned} a_1u + a_1tv - \frac{c_4v}{F} &= a_1v\left(\frac{u}{v} - \frac{c_1}{F}\right) + (a_1c_1 + a_1tF)\frac{v}{F} - \frac{c_4v}{F} \\ &= a_1v\left(\frac{u}{v} - \frac{c_1}{F}\right) + (a_1^2 + c_4 - t)\frac{v}{F} - \frac{c_4v}{F} \\ &= a_1v\left(\frac{u}{v} - \frac{c_1}{F}\right) + (a_1^2 - t)\frac{v}{F}. \end{aligned}$$

Thus from (9), (10), (12) and the fact that $v < F^2/\sqrt{n}$ we have that

$$\left|a_1u + a_1tv - \frac{c_4v}{F}\right| < a_1v\frac{\sqrt{n}}{vF^2} + \left(\frac{n}{F^3}\right)^2\frac{v}{F} < \frac{n}{F^3}\frac{\sqrt{n}}{F^2} + \frac{n^2}{F^7}\frac{F^2}{\sqrt{n}} = \frac{2n^{3/2}}{F^5}. \tag{13}$$

Let $d = a_1u + a_1tv$. Note that the general solution to $yu + zv = d$ is given by

$$y = dv_0 + vs,\ \ z = du_0 - us,$$

where s runs over the integers. Let s be the unique integer with

$$a_1 = dv_0 + vs,\ \ a_1t = du_0 - us.$$

From (11) we have that s satisfies

$$(dv_0 + vs)^2 + c_4 - \frac{du_0 - us}{dv_0 + vs} = (dv_0 + vs)c_1 + (du_0 - us)F.$$

Thus from (13) we see that s is an integral root for one of the polynomials $f_d(x)$ presented in Step 6. This concludes the proof of correctness of Algorithm 2. □

Since the computation of the convergents u/v and u'/v' in Step 5 of the algorithm can be made part of the extended Euclidean algorithm for c_1 and F, it is clear that the runtime of Algorithm 2 is dominated by the calculations of the possible integer roots of the four cubic polynomials in Step 6. Thus Algorithm 2 runs in $O(\log n)$ arithmetic operations with integers at most n if a binary search is used to find the roots as discussed in connection with the base B factorization method above.

We now use Algorithm 2 in the framework of Algorithm 1.

Algorithm 3. *We are input an integer $n > 5$ and the complete prime factorization of $n - 1$. This deterministic algorithm decides if n is prime or composite.*

Let $F(1) = 1$. For $a = 2, 3, \ldots, [(\log n)^{10/7}]$ do the following:

Steps 1–4 *These are exactly the same as in Algorithm 1 except that when $F(a) = F(a-1)$ we go to Step 6.*

Step 5 *If $F(a) \geq n^{3/10}$, find the complete prime factorization of n with Algorithm 2 and stop.*
Step 6 *If $a < [(\log n)^{10/7}]$, get the next a. Otherwise declare n composite and stop.*

We have already proved in connection with Algorithm 1 that if we do Steps 1–4 for $j = 2, 3, \ldots, a$, and we have not stopped, then every prime factor of n is $1 \mod F(a)$. Thus Algorithm 2 is appropriate to use in Step 5. Suppose $a = [(\log n)^{10/7}]$ and we are in Step 6. If n is prime and $\mathcal{G}_n(a)$ is as before, then, as with the proof of correctness of Algorithm 1, we have

$$F(a) = \#\mathcal{G}_n(a) \geq \psi(n, a) = \psi(n, (\log n)^{10/7}) > n^{3/10}$$

by Theorem 1. Thus Step 6 is correct. We conclude that Algorithm 3 is correct.

We have the following theorem.

Theorem 3. *Given an integer $n > 5$ and the complete prime factorization of $n-1$, Algorithm 3 correctly decides if n is prime or composite. Moreover, it uses at most $O((\log n)^{17/7}/\log\log n)$ arithmetic operations with integers at most n.*

4. When $n - 1$ Is Partially Factored

In this section we describe a deterministic polynomial time algorithm that decides if n is prime or composite when n and a fully-factored divisor F of $n-1$ are input with $F > n^{1/4+\varepsilon}$.

Algorithm 4. *We are input an integer n and a number ε with $n > e^3$, $0 < \varepsilon \leq 3/4$ and $(\log n)^{5/(4\varepsilon)} < n$. We are also input integers F, R with $n - 1 = FR$ and $F > n^{1/4+\varepsilon}$, and we are input the complete prime factorization of F. This deterministic algorithm decides if n is prime or composite.*

Let $F(1) = 1$. For $a = 2, 3, \ldots, [(\log n)^{5/(4\varepsilon)}]$ do the following:

Step 1 *If a is composite, let $F(a) = F(a-1)$ and go to Step 7. If $a^{RF(a-1)} \equiv 1 \mod n$, let $F(a) = F(a-1)$ and go to Step 7. Verify that $a^{n-1} \equiv 1 \mod n$. If not, declare n composite and stop.*
Step 2 *Using the prime factorization of F, compute $E(a)$, the order of $a^R \mod n$ in $(\mathbb{Z}/n\mathbb{Z})^*$. Thus $E(a)$ is the least positive divisor of F with $a^{RE(a)} \equiv 1 \mod n$.*
Step 3 *For each prime factor q of $E(a)$, verify that $(a^{RE(a)/q} - 1, n) = 1$. If not, declare n composite and stop.*
Step 4 *Let $F(a) = \mathrm{lcm}\{F(a-1), E(a)\}$. Compute $F(a)$.*
Step 5 *If $F(a) \geq n^{3/10}$, get the complete prime factorization of n by Algorithm 2. In particular, if n is prime, declare it so and stop; if n is composite declare it so and stop.*

Step 6 *If $F(a) > n^{1/4+\varepsilon/5}$, attempt to factor n by the base $F(a)$ factorization method. If this succeeds in splitting n, then declare n composite and stop. Let c_1, c_2, c_3 be the base $F(a)$ "digits" of n so that $n = c_3F(a)^3 + c_2F(a)^2 + c_1F(a) + 1$. Let $c_4 = c_3F(a) + c_2$. If either $c_4x^2 + c_1x + 1$ or $(c_4 - 1)x^2 + (c_1 + F(a))x + 1$ are reducible in $\mathbb{Z}[x]$, this may lead to nontrivial factorization of n by substituting $F(a)$ for x. If so, declare n composite and stop.*

Step 7 *If $a < [(\log n)^{5/(4\varepsilon)}]$ get the next a. If $a = [(\log n)^{5/(4\varepsilon)}]$ and $F(a) \le n^{1/4+\varepsilon/5}$ declare n composite and stop. If $a = [(\log n)^{5/(4\varepsilon)}]$ and $F(a) > n^{1/4+\varepsilon/5}$, declare n prime and stop.*

Proof of correctness. Since $(\log n)^{5/(4\varepsilon)} < n$, Step 1 is correct. Step 3 is clearly correct. We recall the definition of $\mathcal{G}_r(a)$ from the Introduction. If we have passed Step 3 of the algorithm for $2, 3, \ldots, a$, then $F(a)|\#\mathcal{G}_r(a)$ and $\#\mathcal{G}_r(a)|RF(a)$ for each prime factor r of n. In particular $r \equiv 1 \mod F(a)$. Thus it is appropriate to use Algorithm 2 in Step 5.

It is clear that Step 6 is correct since it only declares n composite when it succeeds in splitting n. Suppose we are in Step 7 and $a = [(\log n)^{5/(4\varepsilon)}]$. If n is prime, then as in the analysis of Algorithm 1, we have

$$RF(a) \ge \#\mathcal{G}_n(a) \ge \psi(n, a) = \psi(n, (\log n)^{5/(4\varepsilon)}) > n^{1-4\varepsilon/5}$$

by Theorem 1. Since $R < n^{3/4-\varepsilon}$, we thus have $F(a) > n^{1/4+\varepsilon/5}$. We conclude that if $F(a) \le n^{1/4+\varepsilon/5}$, then Step 7 is correct in declaring n composite. Suppose finally we are in Step 7, $a = [(\log n)^{5/(4\varepsilon)}]$, $F(a) > n^{1/4+\varepsilon/5}$ and n is composite. Since Step 6 was not able to split n, we have as with the analysis of Algorithm 2 that n is the product of two primes. (It is here where we use the hypothesis $n > e^3$ since this assures that $F(a)^4 > n^{1+4\varepsilon/5} > n((\log n)^{5/(4\varepsilon)})^{4\varepsilon/5} = n \log n > 3n$.) Say $n = r_1r_2$ where $r_1 \le r_2$ are primes and $r_1 = b_1F(a) + 1, r_2 = b_2F(a) + 1$, where b_1, b_2 are positive integers. Again from the failure of Step 6 to factor n and the argument for Algorithm 2 (see (11) and (9)) we have

$$b_2 \ge \frac{b_1 + b_2}{2} \ge F(a), \quad b_1 < \frac{n}{b_2F(a)^2} \le \frac{n}{F(a)^3}.$$

Thus

$$\begin{aligned}
\#\mathcal{G}_{r_2}(a) &\le (RF(a), r_2 - 1) \le (n - 1, r_2 - 1) \\
&= (b_1b_2F(a)^2 + (b_1 + b_2)F(a), b_2F(a)) \\
&= (b_1b_2F(a) + b_1 + b_2, b_2)F(a) = (b_1, b_2)F(a) \\
&\le b_1F(a) = \frac{b_1}{b_2}b_2F(a) < \frac{n}{F(a)^4}r_2 < n^{-4\varepsilon/5}r_2.
\end{aligned}$$

On the other hand, as before, we have

$$\#\mathcal{G}_{r_2}(a) \geq \psi(r_2, a) = \psi(r_2, (\log\, n)^{5/(4\varepsilon)}) \geq \psi(r_2, (\log\, r_2)^{5/(4\varepsilon)}) \geq r_2^{1-4\varepsilon/5}$$

by Theorem 1. These last two displays are incompatible. Thus Step 7 is correct in declaring n prime when $a = [(\log\, n)^{5/(4\varepsilon)}]$ and $F(a) > n^{1/4+\varepsilon/5}$. This concludes the proof of correctness for Algorithm 4. □

The runtime analysis for Algorithm 4 is argued analogously to that of Algorithm 1. We have the following theorem.

Theorem 4. *On input of an integer $n > e^3$, a number ε in the range $0 < \varepsilon \leq 3/4$ with $(\log\, n)^{5/(4\varepsilon)} < n$, integers F, R with $n - 1 = FR$ and $F > n^{1/4+\varepsilon}$, and the complete prime factorization of F, Algorithm 4 correctly decides if n is prime or composite. The runtime is $O((\log\, n)^{1+5/(4\varepsilon)}/\log\log n)$ arithmetic operations with integers at most n.*

5. Primes Recognizable in Deterministic Polynomial Time

The algorithms in the preceding two sections to determine whether n is prime or not all hypothesized substantial information about the factorization of $n-1$ (or knowledge about the prime factors of n). In this section we present an algorithm that can be applied to any number n. For most numbers, this algorithm will not be very efficient—in fact it will take exponential time. However there are also many primes for which the algorithm will work in polynomial time—more than $x^{1-\varepsilon}$ of them up to x. Before we give this result, we first state the algorithm.

Algorithm 5. *Suppose we are input a positive integer $n > 5 \times 10^{14}$. This deterministic algorithm decides if n is prime or composite.*

Step 1 *Continue using trial division on $n - 1$ until a fully factored divisor F of $n - 1$ is found with $F > n^{1/3}$.*
Step 2 *Use Algorithm 4 with inputs n, $\varepsilon = 1/12$, F, $R = (n - 1)/F$.*

It is clear that Algorithm 5 is correct. It is also clear that for some numbers it is a terrible algorithm. For example, if n is even, one might well spend exponential time discovering that n is composite. Nevertheless, Algorithm 5 is able to prove prime quite a few numbers in polynomial time.

Theorem 5. *For each $\varepsilon > 0$ there are numbers k and x_0 such that if $x \geq x_0$, then the number of primes $p \leq x$ which Algorithm 5 proves prime in at most $(\log\, p)^k$ arithmetic steps with integers at most p exceeds $x^{1-\varepsilon}$.*

The proof of this theorem depends strongly on the distribution of primes p for which $p-1$ has a large smooth divisor. We establish such a result now.

Theorem 6. *There are effectively computable positive constants c_1, x_1 with the following property. Suppose $x \geq x_1$, $\log x \leq y \leq x^{1/20}$ and $N(x,y)$ is the number of primes $p \leq x$ such that $p-1$ has a y-smooth divisor exceeding $x^{1/3}$. Then $N(x,y) \geq x/(c_1 \log x)^{1+u/3}$, where $u = \log x/\log y$.*

Proof. Let x_2 be the number $x_{\varepsilon,\delta}$ in Theorem 1 of [4], where $\varepsilon = 1/11$ and $\delta = 1/60$. If $x \geq x_2$, let $d_1, d_2, \ldots, d_k$ be the possible "exceptional moduli" corresponding to x in this theorem, so that they all exceed $\log x$ and $k = k(x) = O(1)$. Let q_i denote the greatest prime factor of d_i for $i = 1, \ldots, k$.

Let

$$\mathcal{P} = \{q \text{ prime } : y/2 < q \leq y\} \setminus \{q_1, \ldots, q_k\}.$$

Thus if an integer d is composed solely of primes from $\mathcal{P}$, then no d_i divides d. From Mertens's Theorem we have that if x is sufficiently large, then

$$\frac{1}{2\log y} < \sum_{q\in\mathcal{P}} \frac{1}{q} < \sum_{q\in\mathcal{P}} \frac{1}{q-1} < \frac{1}{\log y}. \tag{14}$$

Let $v = \lceil (\log(x^{1/3}))/\log(y/2) \rceil$ and let $\mathcal{D}(\mathcal{P}, v)$ denote the set of integers d composed of v not necessarily distinct primes from $\mathcal{P}$. If $d \in \mathcal{D}(\mathcal{P}, v)$, then clearly $d > x^{1/3}$ and d is y-smooth. Further, if x is sufficiently large, then

$$d \leq y^{1+(\log(x^{1/3}))/\log(y/2)} = y(x^{1/3})^{\log y/\log(y/2)} \leq x^{2/5}.$$

Thus if x is sufficiently large and $d \in \mathcal{D}(\mathcal{P}, v)$, we have from Theorem 1 in [4] that

$$\pi(x,d,1) := \sum_{p\leq x, d|p-1} 1 \geq \frac{9}{10\varphi(d)} \cdot \frac{x}{\log x}, \tag{15}$$

where p runs over primes and φ denotes Euler's function.

For n a positive integer, let $(n, \mathcal{P})$ denote the largest divisor of n composed of only primes from $\mathcal{P}$. For p a prime, let $d(p,v)$ denote the number of divisors of $p-1$ which come from $\mathcal{D}(\mathcal{P}, v)$. Note that $d(p,v) = 1$ if and only if there is some $d \in \mathcal{D}(\mathcal{P}, v)$ with $d|p-1$ and $((p-1)/d, \mathcal{P}) = 1$. Thus

$$\begin{aligned} N(x,y) &\ge \sum_{\substack{p\le x\\ d(p,v)>0}} 1 \ge \sum_{\substack{p\le x\\ d(p,v)=1}} 1 = \sum_{d\in\mathcal{D}(\mathcal{P},v)} \sum_{\substack{p\le x, d\mid p-1\\ ((p-1)/d,\mathcal{P})=1}} 1 \\ &= \sum_{d\in\mathcal{D}(\mathcal{P},V)} \sum_{\substack{p\le x\\ d\mid p-1}} 1 - \sum_{d\in\mathcal{D}(\mathcal{P},v)} \sum_{\substack{p\le x, d\mid p-1\\ ((p-1)/d,\mathcal{P})>1}} 1 \\ &\ge \sum_{d\in\mathcal{D}(\mathcal{P},V)} \sum_{\substack{p\le x\\ d\mid p-1}} 1 - \sum_{d\in\mathcal{D}(\mathcal{P},v)} \sum_{q\in\mathcal{P}} \sum_{\substack{p\le x\\ dq\mid p-1}} 1 \\ &= \sum_{d\in\mathcal{D}(\mathcal{P},v)} \pi(x,d,1) - \sum_{d\in\mathcal{D}(\mathcal{P},v)} \sum_{q\in\mathcal{P}} \pi(x,dq,1). \end{aligned} \tag{16}$$

From (15) we have

$$\sum_{d\in\mathcal{D}(\mathcal{P},v)} \pi(x,d,1) \ge \frac{9}{10}\cdot\frac{x}{\log x} \sum_{d\in\mathcal{D}(\mathcal{P},v)} \frac{1}{\varphi(d)} \tag{17}$$

if x is sufficiently large. From the Brun-Titchmarsh inequality we have

$$\begin{aligned} \sum_{d\in\mathcal{D}(\mathcal{P},v)} \sum_{q\in\mathcal{P}} \pi(x,dq,1) &\ll \sum_{d\in\mathcal{D}(\mathcal{P},v)} \sum_{q\in\mathcal{P}} \frac{x}{\varphi(dq)\log(x/(dq))} \\ &\ll \frac{x}{\log x}\left(\sum_{d\in\mathcal{D}(\mathcal{P},v)} \frac{1}{\varphi(d)}\right) \sum_{q\in\mathcal{P}} \frac{1}{q-1} \\ &< \frac{x}{\log x\,\log y} \sum_{d\in\mathcal{D}(\mathcal{P},v)} \frac{1}{\varphi(d)}, \end{aligned}$$

where we use (14) for the last inequality. Putting this estimate and (17) into (16) we have for all sufficiently large x that

$$N(x,y) \ge \frac{4}{5}\cdot\frac{x}{\log x} \sum_{d\in\mathcal{D}(\mathcal{P},v)} \frac{1}{\varphi(d)}. \tag{18}$$

We now estimate this last sum. We have from (14) that

$$\sum_{d\in\mathcal{D}(\mathcal{P},v)} \frac{1}{\varphi(d)} > \sum_{d\in\mathcal{D}(\mathcal{P},v)} \frac{1}{d} \geq \frac{1}{v!}\left(\sum_{q\in\mathcal{P}} \frac{1}{q}\right)^{v}$$
$$= \exp(-v \log v - v \log\log y + O(v))$$
$$= \exp(-\frac{1}{3}u \log u - \frac{1}{3}u \log\log y + O(u))$$
$$= \exp(-\frac{1}{3}u \log\log x + O(u)).$$

Putting this in (18) we have the theorem. □

Proof of Theorem 5. First note that from Theorem 6, if $k \geq 1$ is arbitrary, then the number of primes $p \leq x$ for which $p-1$ has a $(\log x)^k$-smooth divisor F with $F > x^{1/3}$ is at least $x^{1-1/(3k)+o(1)}$ as $x \to \infty$. For such primes p, Step 1 of Algorithm 5 takes at most $O((\log x)^k)$ arithmetic steps with integers at most p. The number of arithmetic steps with integers at most p to complete Step 2 of Algorithm 5 is $O((\log p)^{16}/\log\log p)$.

It suffices to establish the theorem for values of ε satisfying $0 < \varepsilon < 1/48$. In this case let $K > K' > 1/(3\varepsilon)$ be arbitrary. If $p > x^{1/3}$ and c is any constant, then $(\log p)^K > c(\log x)^{K'}$ for all large x. Thus if x is large, p is a prime with $p \leq x$ and $p-1$ has a $(\log x)^{K'}$-smooth divisor $F > x^{1/3}$, then Algorithm 5 takes at most $(\log p)^K$ steps with integers at most p to prove p prime. By the above, the number of such primes is at least $x^{1-1/(3K')+o(1)}$ for $x \to \infty$. As $1 - 1/(3K') > 1 - \varepsilon$, the theorem follows. □

6. More Primes Recognizable in Deterministic Polynomial Time

In this section we describe a deterministic algorithm that recognizes many more primes in polynomial time than our previous methods. Covered is any prime n with a divisor F of $n-1$ exceeding n^{ε} and such that all of the prime factors of F are at most $(\log n)^k$. The running time is about $O((\log n)^{2/\varepsilon} + (\log n)^k)$. We also show that for most such primes, the $2/\varepsilon$ can be reduced to $1/\varepsilon$. A corollary of this algorithm is an improved version of Theorem 5. There if we were willing to spend time $(\log n)^k$ on trying to prove n prime, we would succeed for about $x^{1-(3k)^{-1}}$ primes up to x. With the methods of this section we will succeed for about $x^{1-k^{-2}}$ primes up to x.

If n is prime and the order of b in $(\mathbb{Z}/n\mathbb{Z})^*$ is E, then for any a with $a^E \equiv 1 \mod n$, there is some exponent $j \in \{1, 2, \ldots, E\}$ with $a \equiv b^j \mod n$. Thus if it is shown that no such j exists, then it is proved that n is composite. How difficult is it to do this test? If we have already prepared the complete set $\{b^j \mod n : j = 1, 2, \ldots, E\}$, then testing if there is some j with $a \equiv b^j \mod n$

can be accomplished by a binary search in $O(\log E)$ steps. Thus we have the initial step of preparing the set of powers of b, which takes E steps, and then each subsequent test takes $O(\log E)$ steps.

In the case when $E = q^\beta$ with q prime and $\beta \geq 1$, we can do a precomputation taking q steps, with each subsequent test taking $O(\beta^2 \log q)$ steps. Here is how. Suppose the order of $b \mod n$ in $(\mathbb{Z}/n\mathbb{Z})^*$ is q^β. Assuming that we have already computed $b^{q^{\beta-1}} \mod n$ (if not, this takes an additional $O(\beta \log q)$ steps for the precomputation), we can compute the set

$$\mathcal{B} = \{b^{jq^{\beta-1}} \mod n : j = 1, \ldots, q\}$$

in q steps. Now suppose we are presented with some integer a with the order of $a \mod n$ in $(\mathbb{Z}/n\mathbb{Z})^*$ equal to q^α, with $0 \leq \alpha \leq \beta$, and we wish to see if there is some integer j with $a \equiv b^j \mod n$. If $\alpha = 0$ or 1, then if j exists it is a multiple of $q^{\beta-1}$, and so we test for membership of $a \mod n$ in $\mathcal{B}$ by a binary search. As an induction hypothesis suppose $1 < \alpha \leq \beta$ and we have already described how to find j for any a' for which the order of $a' \mod n$ in $(\mathbb{Z}/n\mathbb{Z})^*$ properly divides q^α. Note that the order of $a^q \mod n$ is $q^{\alpha-1}$, so we may use our inductively described algorithm to search for some integer j_0 with $a^q = b^{j_0} \mod n$. Suppose we have found j_0. Then it must be that $q^{\beta-(\alpha-1)}$ divides j_0 and, in particular, $q|j_0$. Then $ab^{-j_0q^{-1}} \mod n$ has order dividing q. But we have already described how in this case we may search for an integer j_1 with $ab^{-j_0q^{-1}} \equiv b^{j_1} \mod n$. If these searches are successful, we may take $j = j_1 + j_0q^{-1}$. Totaling up the time spent, we have used α binary searches in the set $\mathcal{B}$, we have done $\alpha - 1$ modular multiplications, and we have done $\alpha - 1$ modular powerings with exponents at most q^β (in fact, at most $q^{\beta-1}$). The latter computation dominates, taking $O(\alpha\beta \log q) = O(\beta^2 \log q)$ arithmetic steps with integers the size of n.

The computation of $\mathcal{B}$, which is the precomputation step of this method, we call "Set up (b, q^β)". The subsequent search for an exponent j we call "Test $(b, q^\beta; a)$". With these subroutines we are now ready to describe the main algorithm of this section.

Algorithm 6. *We are given positive integers n, F, R and a positive number ε such that $n > 4$, $2 \leq (\log n)^{2/\varepsilon} < n$, $n - 1 = FR$, and $F > n^\varepsilon$. This algorithm decides if n is prime or composite.*

Let $F(1) = 1$. For $a = 2, 3, \ldots, \lceil(\log n)^{2/\varepsilon}\rceil$ do the following.

Step 1. *Check if n is even or if n is a nontrivial power. If so, declare n composite and stop.*

Step 2. *Verify that $a^{n-1} \equiv 1 \mod n$ holds. If not, declare n composite and stop.*

Step 3. *Compute $E(a)$, the order of $a^R \mod n$ in $(\mathbb{Z}/n\mathbb{Z})^*$. Let $F(a) = \operatorname{lcm}\{E(a), F(a-1)\}$. For each prime $q|F(a)/F(a-1)$, verify that $(a^{RE(a)/q} - 1, n) = 1$ holds, declaring n composite and stopping if not.*

Step 4. *For each prime q and positive integers α, β with $q^\alpha \parallel (E(a), F(a-1))$ and $q^\beta \parallel F(a-1)$ do Test $(b_q, q^\beta; a^{RE(a)q^{-\alpha}})$. If this test proves n composite, then declare this and stop.*

Step 5. *For each prime q and positive integer β with $q | F(a)/F(a-1)$ and $q^\beta \parallel F(a)$, let $b_q = a^{RE(a)q^{-\beta}} \mod n$ and do Set up (b_q, q^β).*

Step 6. *If $F(a)^{\log a} > n^{\log\log n}$, declare n prime and stop. If $a < \lceil (\log n)^{2/\varepsilon} \rceil$, get the next a. If $a = \lceil (\log n)^{2/\varepsilon} \rceil$, declare n composite and stop.*

Proof of correctness. Suppose we have made it to Step 6 and $a = \lceil (\log n)^{2/\varepsilon} \rceil$. Suppose n is prime. From Theorem 1 we have $\#\mathcal{G}_n(a) > n^{1-\varepsilon/2}$. Every $b \in \mathcal{G}_n(a)$ satisfies $b^{(n-1)F(a)/F} \equiv 1 \mod n$, so $\#\mathcal{G}_n(a) \leq (n-1)F(a)/F < n^{1-\varepsilon}F(a)$. Thus $F(a) > n^{\varepsilon/2}$, so that $F(a)^{\log a} > n^{\log\log n}$. Hence it is correct to declare n composite when this inequality fails.

Suppose n is composite and suppose we have made it to Step 6 for a particular a. From Step 1 we know that n is divisible by at least two distinct odd primes. From Step 3 we know that each prime factor of n is $1 \mod F(a)$. Let $\mathcal{F} = \{b \mod n : b^{n-1} \equiv 1 \mod n\}$. By the Chinese remainder theorem, this subgroup of $(\mathbb{Z}/n\mathbb{Z})^*$ is isomorphic to the direct product of the groups $\mathcal{F}_p = \{b \mod p^\alpha : b^{n-1} \equiv 1 \mod p^\alpha\}$ where p^α runs over the prime powers with $p^\alpha \parallel n$. Since p is an odd prime, $(\mathbb{Z}/p^\alpha\mathbb{Z})^*$ is cyclic so that $\#\mathcal{F}_p = (n-1, \varphi(p^\alpha)) = (n-1, p-1)$. Thus for each prime power q^β with $q^\beta \parallel F(a)$ we have $q^\beta | \#\mathcal{F}_p$. Since n has at least two distinct prime factors p, the number of $b \mod n \in \mathcal{F}$ with $b^{q^\beta} \equiv 1 \mod n$ is at least $q^{2\beta}$. From Step 2, $\mathcal{G}_n(a)$ is a subgroup of $\mathcal{F}$. And from Step 4 we have that $\mathcal{G}_n(a)$ has exactly q^β members $b \mod n$ with $b^{q^\beta} \equiv 1 \mod n$. Thus the index of $\mathcal{G}_n(a)$ in $\mathcal{F}$ is at least the product of the prime powers q^β with $q^\beta \parallel F(a)$, which is $F(a)$. We have from Theorem 1 that

$$n^{1-\log\log n/\log a} < \#\mathcal{G}_n(a) \leq \frac{1}{F(a)} \prod_{p|n} (n-1, p-1) < \frac{n}{F(a)},$$

so that $F(a) < n^{\log\log n/\log a}$. Hence it is correct to declare n prime when $F(a)^{\log a} > n^{\log\log n}$. This concludes the proof of correctness. □

Analysis of runtime. For each integer $k \leq \log n/\log 2$ we can check if n is a k-th power by computing $[n^{1/k}]$ with a binary search and seeing if $[n^{1/k}]^k = n$. When $k \geq 3$, the binary search may begin with $[n^{1/(k-1)}]$. Thus the number of arithmetic operations to see if n is a k-th power is $O(\frac{\log k}{k} \log n)$, so the total number of arithmetic operations for Step 1 is $O(\log n(\log\log n)^2)$.

The time for Step 2 is at most $O((\log n)^{1+2/\varepsilon})$.

For each a, the time for Step 3 is $O(\log n \log\log n)$. Thus the total time for Step 3 is $O((\log n)^{1+2/\varepsilon} \log\log n)$.

As we have seen, each implementation of Test $(b_q, q^\beta; a)$ in Step 4 takes time $O(\beta^2 \log q)$. Thus the total time for Step 4 is $O((\log n)^{2+2/\varepsilon})$.

Each time we do Set up (b_q, q^β) in Step 5, it takes time $O(q + \beta \log q)$. Thus if Ω is the total number of prime factors of F, counted with multiplicity, and if each prime factor q of F satisfies $q \le B$: then the total time for Step 5 is $O(B\Omega + \Omega^2 \log n) = O(B \log n + (\log n)^3)$.

Thus the total number of arithmetic steps with integers at most n is $O((\log n)^{2+2/\varepsilon} + B \log n)$.

We have the following theorem.

Theorem 7. *On input of positive integers n, F, R and a positive number ε with $n > 4$, $2 \le (\log n)^{2/\varepsilon} < n$, $n-1 = FR$ and $F > n^\varepsilon$, Algorithm 6 correctly decides if n is prime or composite. The running time is $O((\log n)^{2+2/\varepsilon} + B \log n)$ arithmetic steps with integers at most n, where B is the largest prime factor of F.*

We remark that the term $(\log n)^{2+2/\varepsilon}$ in the running time may be replaced with $\varepsilon(\log n)^{2+2/\varepsilon}/\log\log n$ if we perform Steps 2 to 4 only for prime values of a.

The time bound in Theorem 7 is only an upper bound. With the aid of the next result we will be able to show that most primes for which Algorithm 6 is applicable are proved prime in a considerably shorter time.

Theorem 8. *For $2 \le y \le x$ let $R(x, y)$ denote the number of primes p in the range $y < p \le x$ such that $(\mathbb{Z}/p\mathbb{Z})^*$ is not generated by the set $\{a \bmod p : 1 \le a \le y\}$; that is, such that $\mathcal{G}_p([y]) \ne (\mathbb{Z}/p\mathbb{Z})^*$. Then $R(x, y) < 8x^{2 \log\log(x^2)/\log y}$.*

Proof. Let $\mathcal{Z}$ denote the set of y-smooth numbers up to x^2. Suppose p is a prime counted by $R(x, y)$. Then $\#\mathcal{G}_p([y])$ divides $(p-1)/q$ for some prime q. Thus the set $\mathcal{Z}$ occupies at most $(p-1)/q$ residue classes mod p, so there are at least $p/2$ residue classes mod p free of elements of $\mathcal{Z}$. Thus by the large sieve (see Theorem 3, p. 159 in [12]) we have

$$\#\mathcal{Z} \le \frac{4x^2}{\frac{1}{2} \cdot R(x, y)}.$$

Thus from Theorem 1 we have

$$R(x, y) \le \frac{8x^2}{\#\mathcal{Z}} = \frac{8x^2}{\psi(x^2, y)} < 8x^{2 \log\log(x^2)/\log y}, \tag{19}$$

which completes the proof of the theorem. □

Because of the use of Theorem 1 in the proof, Theorem 8 is only fairly good when y is a power of $\log x$. If we let $v = 2 \log x/\log y$ and use a stronger estimate for $\psi(x^2, y)$ (see [10]) we may obtain the following result which is valid in the same range as is Theorem 8:

$$R(x,y) \le \exp(v(\log v + \log\log v - 1 + (\log\log v - 1)/\log v) \\ + O(v(\log\log v/\log v)^2)).$$

This estimate is implicit in the dissertation of Pappalardi [21, Sect. 3.3]. The estimate has an inexplicit constant, though an actual numerical value could be provided in principle. We remark that Vinogradov, Linnik and Fridlender have discussed problems related to the estimation of $R(x,y)$.

Theorem 9. *Let $x, \varepsilon > 0$ be arbitrary and let $E(x,\varepsilon)$ denote the number of primes p with $(\log p)^{2/\varepsilon} < p \le x$ for which there is some integer $F > p^\varepsilon$ with $F|p-1$ and such that if Algorithm 6 is run on $n = p$, F, ε, then $F(a)^{\log a} \le p^{\log\log p}$ for $a = \lceil(\log p)^{1/\varepsilon}\rceil$. Then $E(x,\varepsilon) < 9x^{3\varepsilon} + e^{32}$.*

Proof. Suppose Algorithm 6 is run on $n = p$, F, ε where $F|p-1$ and $F > p^\varepsilon$. If $F(a) = F$ with $a = \lceil(\log p)^{1/\varepsilon}\rceil$, then $F(a)^{\log a} > p^{\log\log p}$. Thus we may assume that if p is counted by $E(x,\varepsilon)$, then $F(a) < F$ with $a = \lceil(\log p)^{1/\varepsilon}\rceil$. Thus $(\mathbb{Z}/p\mathbb{Z})^*$ is not generated by $\{j \mod p : 1 \le j \le a\}$. We conclude from Theorem 8 that

$$\begin{aligned} E(x,\varepsilon) - E(x^{1/2},\varepsilon) &\le R(x, \lceil(\log x^{1/2})^{1/\varepsilon}\rceil) \\ &\le R(x, (\log x^{1/2})^{1/\varepsilon}) \\ &< 8x^{2\varepsilon \log\log(x^2)/\log\log(x^{1/2})}. \end{aligned}$$

Note that if $x \ge e^{32}$, then $\log\log(x^2)/\log\log(x^{1/2}) \le 3/2$. Thus if k is that positive integer with $x^{2^{-k}} < e^{32} \le x^{2^{-(k-1)}}$ then

$$\begin{aligned} E(x,\varepsilon) &= \sum_{i=0}^{k-1}(E(x^{2^{-i}},\varepsilon) - E(x^{2^{-i-1}},\varepsilon)) + E(x^{2^{-k}},\varepsilon) \\ &< \sum_{i=0}^{k-1} 8(x^{2^{-i}})^{3\varepsilon} + E(e^{32},\varepsilon) \\ &< 9x^{3\varepsilon} + e^{32}, \end{aligned}$$

which concludes the proof of the theorem. □

We remark that Theorem 8 can also be used in the context of Algorithm 4 to give a small improvement in that algorithm for most primes.

We now give an algorithm similar to Algorithm 5.

Algorithm 7. *Suppose we are input a positive number ε and an integer $n > 4$ with $2 \le (\log n)^{2/\varepsilon} < n$. This algorithm attempts to decide if n is prime or composite.*

Step 1. *Using trial division, find the largest divisor F of $n-1$ composed of primes up to $(\log n)^{1+1/\varepsilon}$. If $F \le n^{\varepsilon}$, then stop for the algorithm has failed.*

Step 2. *Use Algorithm 6 on n, F, ε, terminating with failure if the parameter a exceeds $\lceil (\log n)^{1/\varepsilon} \rceil$.*

From the proof of Theorem 6, for each number ε with $0 < \varepsilon < 1$, there is a number $x_2(\varepsilon)$, such that if $x \ge x_2(\varepsilon)$, the number N of primes $p \le x$ for which $p-1$ has a $(\log p)^{1+1/\varepsilon}$-smooth divisor exceeding p^{ε} satisfies $N > 2x^{1-\varepsilon^2}$. For such primes p we make it past Step 1 of Algorithm 7. From Theorem 9 we have that if $0 < \varepsilon \le 1/4$, then at least half of these primes are proved prime in Step 2 of Algorithm 7, though $x_2(\varepsilon)$ may have to be adjusted. We thus have the following result.

Theorem 10. *Let ε be any number with $0 < \varepsilon \le 1/4$. There is a number $x_2(\varepsilon)$ such that if $x \ge x_2(\varepsilon)$ then the number of primes $p \le x$ which Algorithm 7 proves prime exceeds $x^{1-\varepsilon^2}$.*

We remark that the running time of Algorithm 7 is $O((\log n)^{2+1/\varepsilon})$ arithmetic steps with integers at most n. Thus Theorem 10 improves on Theorem 5 since there if one wants to prove prime $x^{1-\varepsilon^2}$ primes up to x, the bound for the running time is about $(\log n)^{1/(3\varepsilon^2)}$.

7. Update on Primality Testing: Background

The previous sections comprised the original article in the first edition of this volume published in 1997. This section and the next one summarize the state of primality testing in 2013, 16 years later.

The subject of primality testing concerns the creation and analysis of efficient algorithms for deciding whether a given integer $n > 1$ is prime or composite. This subject is closely related to, but distinct from, factoring. Some algorithms, such as trial division, can accomplish both tasks, but the most efficient methods are tailored to one or the other.

From a practical point of view, the story of primality testing is a simple one. In real-world applications one does not require mathematical certitude, a tiny possibility of error being acceptable, so various random algorithms that have been known for decades and are easy to implement may be used. An example, commonly known as the Miller–Rabin test, runs in $O((\log n)^{2+\varepsilon})$ bit operations (using fast arithmetic subroutines) and almost certainly returns a correct verdict on the primality of a given input n, with the bonus that a composite verdict is mathematically correct. Even the simple base-2 Fermat congruence $2^{n-1} \equiv 1 \pmod{n}$ when applied to a large *random* input n almost certainly steers one right (a number n for which the congruence holds is almost certainly prime, a number n for which it does not hold is definitely composite). Indeed, as shown by Erdős [13], composite numbers n satisfying $2^{n-1} \equiv 1 \pmod{n}$ are much scarcer than primes. For more details on these and similar tests see [11] and the references there.

But a problem as fundamental as deciding primality cries out for a thorough mathematical analysis. Here too, where no possibility of error is to be tolerated, there is both a theoretical and practical side. The practical primality tester has specific numbers n in mind that are to be tested, and wishes to implement an algorithm that will give a completely correct answer. It is possible for such an algorithm to use randomness, where coins are flipped (figuratively), but there is no doubt in the output, the only issue being the running time of the algorithm. A simple but illustrative example is that of finding a quadratic nonresidue for a given prime p. This is an integer k such that the congruence $x^2 \equiv k \pmod{p}$ has no integral solutions. We know that for an odd prime p exactly $(p-1)/2$ choices for k in $\{1, \ldots, p-1\}$ are quadratic nonresidues. Moreover, via either Euler's criterion or the law of quadratic reciprocity for Jacobi symbols, it is possible to decide quickly (and deterministically) if a candidate k works or not. So a random and quick method to find a quadratic nonresidue k is to choose randomly from $\{1, \ldots, p-1\}$ until one is found. This simple algorithm runs in expected polynomial time. Remarkably, without assuming an unproved hypothesis (the Extended Riemann Hypothesis), we know no deterministic method for finding a quadratic nonresidue that runs in polynomial time.

Long before our article was published, we had the Adleman–Huang test [1], a random primality test with running time expected to be polynomial (and, as opposed to the Miller–Rabin test, there is no doubt in the output). Based on the arithmetic of Jacobian varieties of hyperelliptic curves of genus 2 (and also on elliptic curves), it is a very difficult result, requiring an entire volume for its analysis. Other tests, based on elliptic curves (practical improvements of the Goldwasser–Kilian test) are not theoretically complete, but stand as excellent practical primality proving algorithms for those who do not wish to have any possibility of error. Again, see [11] for more on this.

And this brings us to the last holdout of the theorist: a deterministic primality test that runs in polynomial time. Such a test has long been known (in fact, a version of the Miller–Rabin test), but it relies on the Extended Riemann Hypothesis in a similar way as the quadratic nonresidue problem mentioned above. Withot any unproved hypothesis, we had the APR test [2] with complexity $O((\log n)^{c \log\log\log n})$, tantalizingly close to being polynomial. We also had an interesting result of Pintz, Steiger, and Szemerédi that presented a set of primes, with counting function to x of about $x^{2/3}$, which could be recognized in deterministic polynomial time. These primes p were characterized by $p-1$ being divisible by a very large power of 3.

In our paper we showed that any prime p can be proved prime by a deterministic algorithm in polynomial time, provided we have a fully-factored divisor $F > p^{\varepsilon}$ of $p-1$. Furthermore, a simple procedure identifies more than

$x^{1-\varepsilon}$ primes p up to x with such a fully-factored divisor in $p-1$, and so we have many primes that are recognizable in polynomial time. The case when $F > p^{1/2+\varepsilon}$ was done earlier by Fürer [17] (we only learned of this paper recently), and the case when $p-1$ itself is fully factored was rediscovered by Fellows and Koblitz [16].

Our paper had one practical component for those interested in implementing a primality test. The so-called "$n-1$ test" of Brillhart, Lehmer, and Selfridge requires a fully-factored divisor of $p-1$ larger than $p^{1/3}$. Our paper was able to reduce the exponent 1/3 in this practical test to 3/10. (For positive exponents smaller than 3/10 our algorithm still has polynomial complexity, but it is not so practical.) More recently, an analog of this improvement was accomplished for the "$n+1$ test", see [11], though it is no longer deterministic.

8. Update on Primality Testing: Derandomization and the AKS Algorithm

By far the most important development since our article was the AKS algoritm [3], named for its inventors, Agrawal, Kayal, and Saxena. Their algorithm is deterministic and it distinguishes between primes and composites in polynomial time. Further it does not depend on any unproved hypotheses for its analysis.

Like the algorithms in our paper and in many other approaches to primality testing, the AKS algorithm either recognizes n as composite by a series of simple tests, or if n passes all of these tests, a group is built up that is so large that n is inescapably prime. (For more on this line of thought see [24].)

Two analyses of the AKS algorithm are presented in [3], a more elementary analysis using effective tools and running time $O((\log n)^{10.5+\varepsilon})$, and an analysis using ineffective tools and running time $O((\log n)^{7.5+\varepsilon})$. Both of these estimates are upper bounds for the true running time, conjectured to be $O((\log n)^{6+\varepsilon})$. A version of the AKS algorithm with this running time and effective tools is presented in the preprint [20] and is described in [11].

Unfortunately, the AKS algorithm has not proved to be numerically competitive with previous primality tests. Even a random version with expected running time $O((\log n)^{4+\varepsilon})$ (see [5]) is not competitive.

In [3] a conjecture is made that suggests a version of the AKS algorithm has running time $O((\log n)^{3+\varepsilon})$. Using a heuristic of Erdős [14] on Carmichael numbers, Lenstra and Pomerance (unpublished) have given a plausibility argument that this AKS conecture is false.

Since we knew already a random polynomial-time algorithm for primality testing, the AKS test might be thought of as a derandomization, even though it bears little resemblance to the Adleman–Huang test. Similarly, the Fürer and Fellows–Koblitz algorithms for proving the primality of a prime p where a large part of $p-1$ is factored are derandomizations of an algorithm of Lucas, as improved by Proth, Pocklington, and Lehmer early in the twentieth century. Our paper as well contains a derandomization (and extension) of the Brillhart, Lehmer, Selfridge $n-1$-test, as mentioned. In [29], Źrałek applied some of the methods of our paper to derandomize a factorization algorithm, namely the $p-1$ method of Pollard. Here, one is expected to find quickly those prime factors p of n which have the additional property that all of the prime factors of $p-1$ are small. (For this reason, some implementers of the RSA cryptosystem use prime factors p, q where both $p-1, q-1$ have large prime factors, so-called safe primes.) The Pollard $p-1$ method uses randomness and Źrałek derandomizes it. In a later paper he again uses similar ideas, this time to factor polynomials over some finite fields $\mathbf{F}_p$.

References

1. L. M. Adleman and M.-D. A. Huang, *Primality testing and two-dimensional abelian varieties over finite fields*, Lecture Notes in Math. **1512**, Springer-Verlag, Berlin, 1992, 142 pp.
2. L. M. Adleman, C. Pomerance, and R. S. Rumely, *On distinguishing prime numbers from composite numbers*, Ann. of Math. **117** (1983), 173–206.
3. M. Agrawal, N. Kayal, and N. Saxena, PRIMES *is in* P, Ann. of Math. **160** (2004), 781–793.
4. W. R. Alford, A. Granville, and C. Pomerance, There are infinitely many Carmichael numbers, Ann. of Math. **140** (1994), 703–722.
5. D. Bernstein, *Proving primality in essentially quartic random time*, Math. Comp. **76** (2007), 389–403.
6. W. Bosma and M.-P. van der Hulst, Primality proving with cyclotomy, Ph.D. thesis, Amsterdam (1990).
7. J. Brillhart, M. Filaseta, and A. Odlyzko, On an irreducibility theorem of A. Cohn, Can. J. Math. **33** (1981), 1055–1099.
8. J. Brillhart, D. H. Lehmer and J. L. Selfridge, New primality criteria and factorizations of $2^m \pm 1$, Math. Comp. **29** (1975), 620–647.
9. J. P. Buhler, R. E. Crandall and M. A. Penk, Primes of the form $n! \pm 1$ and $2 \cdot 3 \cdot 5 \ldots p \pm 1$, Math. Comp. **38** (1982), 639–643.
10. E. R. Canfield, P. Erdős and C. Pomerance, On a problem of Oppenheim concerning "factorisatio numerorum", J. Number Theory **17** (1983), 1–28.
11. R. Crandall and C. Pomerance, *Prime numbers: a computational perspective*, 2nd ed., Springer, New York, 2005.
12. H. Davenport, *Multiplicative number theory*, 2nd ed., Springer-Verlag, New York, 1980.
13. P. Erdős, *On almost primes*, Amer. Math. Monthly **57** (1950), 404–407.
14. P. Erdős, *On pseudoprimes and Carmichael numbers*, Publ. Math. Debrecen **4** (1956), 201–206.
15. P. Erdős and J. H. van Lint, On the number of positive integers $\leq x$ and free of prime factors $> y$, Simon Stevin **40** (1966/67), 73–76.
16. M. R. Fellows and N. Koblitz, Self-witnessing polynomial-time complexity and prime factorization, Designs, Codes and Cryptography **2** (1992), 231–235.

17. M. Fürer, *Deterministic and Las Vegas primality testing algorithms*, in Proceedings of ICALP 85 (July 1985). Nafplion, Greece. W. Brauer, ed., Lecture Notes in Computer Science **194**, Springer-Verlag, Berlin, 1985, pp. 199–209.
18. A. K. Lenstra, H. W. Lenstra, Jr. and L. Lovász, Factoring polynomials with rational coefficients, Math. Ann. **261** (1982), 515–534.
19. H. W. Lenstra, Jr., Miller's primality test, Information Processing Letters **8** (1979), 86–88.
20. H. W. Lenstra, jr and Carl Pomerance, *Primality testing with Gaussian periods*, www.math.dartmouth.edu/~carlp/aks041411.pdf.
21. F. Pappalardi, On Artin's conjecture for primitive roots, Ph.D. thesis, McGill University (1993).
22. J. Pintz, W. L. Steiger and E. Szemerédi, Infinite sets of primes with fast primality tests and quick generation of large primes, Math. Comp. **53** (1989), 399–406.
23. C. Pomerance, Very short primality proofs, Math. Comp. **48** (1987), 315–322.
24. C. Pomerance, *Primality testing: variations on a theme of Lucas*, Congressus Numerantium **201** (2010), 301–312.
25. V. R. Pratt, Every prime has a succinct certificate, SIAM J. Comput. **4** (1975), 214–220.
26. J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, Illinois J. Math. **6** (1962), 64–94.
27. R. Solovay and V. Strassen, A fast Monte Carlo test for primality, SIAM J. Comput. **6** (1977) 84–85; *erratum* 7 (1978), 118.
28. I. M. Vinogradov, On the bound of the least non-residue of nth powers, Bull. Acad. Sci. USSR **20** (1926), 47–58 (= Trans. Amer. Math. Soc. **29** (1927), 218–226).
29. B. Źrałek, *A deterministic version of Pollard's* $p-1$ *algorithm*, Math. Comp. **79** (2010), 513–533.

Ballot Numbers, Alternating Products, and the Erdős-Heilbronn Conjecture

Melvyn B. Nathanson

M.B. Nathanson (✉)
Department of Mathematics, Lehman College (CUNY), Bronx, NY 10468, USA
e-mail: melvyn.nathanson@lehman.cuny.edu

1. Introduction

Let A be a subset of an abelian group. Let hA denote the set of all sums of h elements of A with repetitions allowed, and let $h^\wedge A$ denote the set of all sums of h distinct elements of A, that is, all sums of the form $a_1 + \cdots + a_h$, where $a_1, \ldots, a_h \in A$ and $a_i \neq a_j$ for $i \neq j$.

Let A be a set of k congruence classes modulo a prime p. The Cauchy-Davenport theorem states that

$$|2A| \geq \min(p, 2k-1),$$

and, by induction,

$$|hA| \geq \min(p, hk-h+1)$$

for every $h \geq 2$. Erdős and Heilbronn conjectured 30 years ago that

$$|2^\wedge A| \geq \min(p, 2k-3).$$

They did not include this conjecture in their paper on addition of residue classes [10], but Erdős has frequently mentioned this problem in lectures and papers (for example, Erdős-Graham [9, p. 95]). Dias da Silva and Hamidoune recently prove this conjecture. They used linear algebra and the representation theory of the symmetric group to show that

$$|h^\wedge A| \geq \min(p, hk-h^2+1)$$

for every $h \geq 2$.

The purpose of this paper is to give a complete and elementary exposition of this proof. Instead of representation theory, we will use the combinatorics of the h-dimensional ballot numbers.

2. Multi-dimensional Ballot Numbers

The standard basis for $\mathbf{R}^h$ is the set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$, where

$$\mathbf{e}_1 = (1, 0, 0, 0, \ldots, 0)$$

$$\mathbf{e}_2 = (0, 1, 0, 0, \ldots, 0)$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_13,

$$\vdots$$

$$\mathbf{e}_h = (0, 0, 0, \ldots, 0, 1).$$

The lattice $\mathbf{Z}^h$ is the subgroup of $\mathbf{R}^h$ generated by the set $\{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$, so $\mathbf{Z}^h$ is the set of vectors in $\mathbf{R}^h$ with integral coordinates. Let

$$\mathbf{a} = (a_0, a_1, \ldots, a_{h-1}) \in \mathbf{Z}^h$$

and

$$\mathbf{b} = (b_0, b_1, \ldots, b_{h-1}) \in \mathbf{Z}^h.$$

A *path* in $\mathbf{Z}^h$ is a finite sequence of lattice points

$$\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$$

such that

$$\mathbf{v}_j - \mathbf{v}_{j-1} \in \{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$$

for $j = 1, \ldots, m$. Let $\mathbf{v}_{j-1}$, $\mathbf{v}_j$ be successive points on a path. We call this a *step* in the direction $\mathbf{e}_i$ if

$$\mathbf{v}_j = \mathbf{v}_{j-1} + \mathbf{e}_i.$$

The vector $\mathbf{a}$ is called *nonnegative* if $a_i \geq 0$ for $i = 0, 1, \ldots, h-1$. We write

$$\mathbf{a} \leq \mathbf{b}$$

if $\mathbf{b} - \mathbf{a}$ is a nonnegative vector.

Let $P(\mathbf{a}, \mathbf{b})$ denote the number of paths from $\mathbf{a}$ to $\mathbf{b}$. The path function $P(\mathbf{a}, \mathbf{b})$ is translation invariant in the sense that

$$P(\mathbf{a} + \mathbf{c}, \mathbf{b} + \mathbf{c}) = P(\mathbf{a}, \mathbf{b})$$

for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{Z}^h$. In particular,

$$P(\mathbf{a}, \mathbf{b}) = P(\mathbf{0}, \mathbf{b} - \mathbf{a}).$$

The path function satisfies the boundary conditions

$$P(\mathbf{a}, \mathbf{a}) = 1,$$

and

$$P(\mathbf{a}, \mathbf{b}) > 0 \text{ if and only if } \mathbf{a} \leq \mathbf{b},$$

If $\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$ is a path, then

$$\mathbf{v}_{m-1} = \mathbf{b} - \mathbf{e}_i$$

for some $i = 1, \ldots, h$, and there is a unique path from $\mathbf{b} - \mathbf{e}_i$, to $\mathbf{b}$. It follows that the path counting function $P(\mathbf{a}, \mathbf{b})$ also satisfies the difference equation

$$P(\mathbf{a}, \mathbf{b}) = \sum_{i=0}^{h-1} P(\mathbf{a}, \mathbf{b} - \mathbf{e}_i).$$

Let $\mathbf{a} \leq \mathbf{b}$. For $i = 0, 1, \ldots, k-1$, every path from $\mathbf{a}$ to $\mathbf{b}$ contains exactly $b_i - a_i$ steps in the direction $\mathbf{e}_{i+1}$. Let

$$m = \sum_{i=0}^{h-1} (b_i - a_i).$$

Every path from $\mathbf{a}$ to $\mathbf{b}$ has exactly m steps, and the number of different paths is the multinomial coefficient

$$P(\mathbf{a}, \mathbf{b}) = \frac{(\sum_{i=0}^{h-1} (b_i - a_i))!}{\prod_{i=0}^{h-1} (b_i - a_i)!} = \frac{m!}{\prod_{i=0}^{h-1} (b_i - a_i)!}. \tag{1}$$

Let $h \geq 2$. There are h candidates in an election. The candidates will be labelled by the integers $0, 1, \ldots, h-1$. Suppose that m_0 votes have already been cast, and that candidate i has received a_i votes. Then

$$m_0 = a_0 + a_1 + \cdots + a_{h-1}.$$

We shall call

$$\mathbf{v}_0 = \mathbf{a} = (a_0, a_1, \ldots, a_{h-1})$$

the initial *ballot vector*. There are m remaining voters, each of whom has one vote, and these votes will be cast sequentially. Let $v_{i,k}$ denote the number of votes that candidate i has received after k additional votes have been cast. We represent the distribution of votes at step k by the ballot vector

$$\mathbf{v}_k = (v_{0,k}, v_{1,k}, \ldots, v_{h-1,k}).$$

Then

$$v_{0,k} + v_{1,k} + \cdots + v_{h-1,k} = k + m_0$$

for $k = 0, 1, \ldots, m$. Let

$$\mathbf{v}_m = \mathbf{b} = (b_0, b_1, \ldots, b_{h-1})$$

be the final ballot vector. It follows immediately from the definition of the ballot vectors that

$$\mathbf{v}_k - \mathbf{v}_{k-1} \in \{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$$

for $k = 1, \ldots, m$, and so

$$\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$$

is a path in $\mathbf{Z}^h$ from $\mathbf{a}$ to $\mathbf{b}$. Therefore, the number of distinct sequences of m votes that can lead from the initial ballot vector $\mathbf{a}$ to the final ballot vector $\mathbf{b}$ is the multinomial coefficient

$$\frac{\left(\sum_{i=0}^{h-1} (b_i - a_i)\right)!}{\prod_{i=0}^{h-1} (b_i - a_i)!} = \frac{m!}{\prod_{i=0}^{h-1} (b_i - a_i)!}.$$

Let $\mathbf{v} = (v_1, \ldots, v_h)$ and $\mathbf{w} = (w_1, \ldots, w_h)$ be vectors in $\mathbf{R}^h$. The vector $\mathbf{v}$ will be called *increasing* if

$$v_1 \leq v_2 \leq \cdots \leq v_h,$$

and *strictly increasing* if

$$v_1 < v_2 < \ldots < v_h.$$

Now suppose that the initial ballot vector is

$$\mathbf{a} = (0, 0, 0, \ldots, 0)$$

and that the final ballot vector $\mathbf{b} = (b_0, b_1, \ldots, b_{h-1})$ is nonnegative and increasing. Let

$$m = b_0 + b_1 + \cdots + b_{h-1}.$$

Let $B(b_0, b_1, \ldots, b_{h-1})$ denote the number of ways that m votes can be cast so that all of the k-th ballot vectors are nonnegative and increasing. This is the classical *h-dimensional ballot number*. Observe that

$$B(0, 0, \ldots, 0) = 1$$

and

$$B(b_0, b_1, \ldots, b_{h-1}) > 0$$

if and only if $(b_0, b_1, \ldots, b_{h-1})$ is a nonnegative, increasing vector. These boundary conditions and the difference equation

$$B(b_0, b_1, \ldots, b_{h-1}) = \sum_{i=0}^{h-1} B(b_0, \ldots, b_{i-1}, b_i - 1, b_{i+1}, \ldots, b_{h-1})$$

completely determine the function $B(b_0, b_1, \ldots, b_{h-1})$.

There is an equivalent combinatorial problem. Suppose that the initial ballot vector is

$$\mathbf{a}^* = (0, 1, 2, \ldots, h-1),$$

and that the final ballot vector

$$\mathbf{b} = (b_0, b_1, \ldots, b_{h-1})$$

is nonnegative and strictly increasing. Let

$$m = \sum_{i=0}^{h-1} (b_i - i) = \sum_{i=0}^{h-1} b_i - \binom{h}{2}.$$

Let $\hat{B}(b_0, b_1, \ldots, b_{h-1})$ denote the number of ways that m votes can be cast so that all of the ballot vectors $\mathbf{v}_k$ are nonnegative and strictly increasing. We shall call this the *strict h-dimensional ballot number*.

A path $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m$ in $\mathbf{Z}^h$ will be called *strictly increasing* if every lattice point $\mathbf{v}_k$ on the path is strictly increasing. Then $\hat{B}(b_0, b_1, \ldots, b_{h-1})$ is the number of strictly increasing paths from $\mathbf{a}^*$ to $\mathbf{b} = (b_0, \ldots, b_{h-1})$.

The strict h-dimensional ballot numbers satisfy the boundary conditions

$$\hat{B}(0, 1, \ldots, h-1) = 1$$

and

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) > 0$$

if and only if $(b_0, b_1, \ldots, b_{h-1})$ is a nonnegative, strictly increasing vector. These boundary conditions and the difference equation

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) = \sum_{i=0}^{h-1} \hat{B}(b_0, \ldots, b_{i-1}, b_i - 1, b_{i+1}, \ldots, b_{h-1})$$

completely determine $\hat{B}(b_0, b_1, \ldots, b_{h-1})$.

There is a simple relationship between the h-dimensional ballot numbers $B(b_0, b_1, \ldots, b_{h-1})$ and $\hat{B}(b_0, b_1, \ldots, b_{h-1})$. The lattice point

$$\mathbf{v} = (v_0, v_1, \ldots, v_{h-1})$$

is nonnegative and strictly increasing if and only if the lattice point

$$\mathbf{v}' = \mathbf{v} - (0, 1, 2, \ldots, h-1) = \mathbf{v} - \mathbf{a}^*$$

is nonnegative and increasing. It follows that

$$\mathbf{a}^* = \mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m = \mathbf{b}$$

is a path of strictly increasing vectors from $\mathbf{a}^*$ to $\mathbf{b}$ if and only if

$$\mathbf{0}, \mathbf{v}_1 - \mathbf{a}^*, \mathbf{v}_2 - \mathbf{a}^*, \ldots, \mathbf{b} - \mathbf{a}^*$$

is a path of increasing vectors from $\mathbf{0}$ to $\mathbf{b} - \mathbf{a}^*$. Thus,

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) = B(b_0, b_1 - 1, b_2 - 2, \ldots, b_{h-1} - (h-1)).$$

For $1 \leq i < j \leq h$, let $H_{i,j}$ be the hyperplane in $\mathbf{R}^h$ consisting of all vectors $(x_1, \ldots, x_h)$ such that $x_i = x_j$. There are $\binom{h}{2}$ such hyperplanes. A path

$$\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m = \mathbf{b}$$

will be called *intersecting* if there exists at least one vector $\mathbf{v}_k$ on the path such that $\mathbf{v}_k \in H_{i,j}$ for some hyperplane $H_{i,j}$.

The symmetric group S_h acts on $\mathbf{R}^h$ as follows: For $\sigma \in S_h$ and $\mathbf{v} = (v_0, v_1, \ldots, v_{h-1}) \in \mathbf{R}^h$, let

$$\sigma\mathbf{v} = (v_{\sigma(0)}, v_{\sigma(1)}, \ldots, v_{\sigma(h-1)}).$$

A path is intersecting if and only if there is a transposition $\tau = (i, j) \in S_h$ such that $\tau\mathbf{v}_k = \mathbf{v}_k$ for some lattice point $\mathbf{v}_k$ on the path.

Let $I(\mathbf{a},\mathbf{b})$ denote the number of intersecting paths from $\mathbf{a}$ to $\mathbf{b}$. Let $J(\mathbf{a},\mathbf{b})$ denote the number of paths from $\mathbf{a}$ to $\mathbf{b}$ that do not intersect any of the hyperplanes $H_{i,j}$. Then

$$P(\mathbf{a},\mathbf{b}) = I(\mathbf{a},\mathbf{b}) + J(\mathbf{a},\mathbf{b}).$$

Lemma 1. *Let* $\mathbf{a}$ *be a lattice point in* $\mathbf{Z}^h$*, and let* $\mathbf{b} = (b_0, \ldots, b_{h-1})$ *be a strictly increasing lattice point in* $\mathbf{Z}^h$*. A path from* $\mathbf{a}$ *to* $\mathbf{b}$ *is strictly increasing if and only if it intersects none of the hyperplanes* $H_{i,j}$*, and so*

$$\hat{B}(b_0, \ldots, b_{h-1}) = J(\mathbf{a}^*, \mathbf{b}).$$

Proof. Let $\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$ be a path, and-let

$$\mathbf{v}_k = (v_{0,k}, v_{1,k}, \ldots, v_{h-1,k})$$

for $k = 0, 1, \ldots, m$. If the path is strictly increasing, then every vector on the path is strictly increasing, and so the path does not intersect any of the hyperplanes $H_{i,j}$. Conversely, if the path is not strictly increasing, then there exists a greatest integer k such that the lattice point $\mathbf{v}_{k-1}$ is not strictly increasing. Then $1 \leq k \leq m$, and

$$v_{j,k-1} \leq v_{j-1,k-1}$$

for some $j = 1, \ldots, h-1$. Since the vector $\mathbf{v}_k$ is strictly increasing, we have

$$v_{j-1,k} \leq v_{j,k} - 1.$$

Since $\mathbf{v}_{k-1}$ and $\mathbf{v}_k$ are successive vectors in a path, we have

$$v_{j-1,k-1} \leq v_{j-1,k}$$

and

$$v_{j,k} - 1 \leq v_{j,k-1}.$$

Combining these inequalities, we obtain

$$v_{j,k-1} \leq v_{j-1,k-1} \leq v_{j-1,k} \leq v_{j,k} - 1 \leq v_{j,k-1}.$$

This implies that

$$v_{j,k-1} = v_{j-1,k-1}$$

and so the vector v_{k-1} lies on the hyperplane $H_{j-1,j}$. Therefore, if $\mathbf{b}$ is a strictly increasing vector, then a path from $\mathbf{a}$ to $\mathbf{b}$ is strictly increasing if and only if it is non-intersecting, and so $J(\mathbf{a},\mathbf{b})$ is equal to the number of strictly increasing paths from $\mathbf{a}$ to $\mathbf{b}$. It follows that $J(\mathbf{a}^*,\mathbf{b})$ is equal to the strict ballot number $\hat{B}(b_0, \ldots, b_{h-1})$. This completes the proof. □

Lemma 2. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be strictly increasing vectors. Then*

$$P(\sigma\mathbf{a},\mathbf{b}) = I(\sigma\mathbf{a},\mathbf{b})$$

for every $\sigma \in S_h, \sigma_h \neq \mathrm{id}$.

Proof. If $\mathbf{a}$ is strictly increasing and if $\sigma \in S_h$, $\sigma \neq \mathrm{id}$, then $\sigma\mathbf{a}$ is not strictly increasing, and so every path from $\sigma\mathbf{a}$ to $\mathbf{b}$ must intersect at least one of the hyperplanes $H_{i,j}$. This completes the proof. □

Lemma 3. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be strictly increasing lattice points. Then*

$$\sum_{\sigma \in S_h} \varepsilon(\sigma) I(\sigma\mathbf{a}, \mathbf{b}) = 0.$$

Proof. Since $\mathbf{a}$ is strictly increasing, it follows that there are $h!$ distinct lattice points of the form $\sigma\mathbf{a}$, where $\sigma \in S_h$, and none of these lattice points lies on a hyperplane $H_{i,j}$. Let Ω be the set of all intersecting paths that start at any one of the $h!$ lattice points $\sigma\mathbf{a}$ and end at $\mathbf{b}$. We shall construct an involution from the set Ω to itself.

Let $\sigma \in S_h$, and let

$$\sigma\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$$

be a path that intersects at least one of the hyperplanes. Let k be the least integer such that $\mathbf{v}_k \in H_{i,j}$ for some $i < j$. Then $k \geq 1$ since $\mathbf{a}$ is strictly increasing, and the hyperplane $H_{i,j}$ is uniquely determined since $\mathbf{v}_k$ lies on a path. Consider the transposition $\tau = (i, j) \in S_h$. Then

$$\tau\mathbf{v}_k = \mathbf{v}_k \in H_{i,j},$$

$$\tau\sigma\mathbf{a} \neq \sigma\mathbf{a},$$

and

$$\tau\sigma\mathbf{a} = \tau\mathbf{v}_0, \tau\mathbf{v}_1, \ldots, \tau\mathbf{v}_k = \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_m = \mathbf{b}$$

is an intersecting path in Ω from $\tau\sigma\mathbf{a}$ to $\mathbf{b}$. Moreover, k is the least integer such that a lattice point on this new path lies in one of the hyperplanes, and $H_{i,j}$ is still the unique hyperplane containing $\mathbf{v}_k$. Since τ^2 is the identity permutation for every transposition τ, it follows that if we apply the same mapping to this path from $\tau\sigma\mathbf{a}$ to $\mathbf{b}$, we recover the original path from $\sigma\mathbf{a}$ to $\mathbf{b}$. Thus, this mapping is an involution on the set Ω of intersecting paths from the $h!$ lattice points $\sigma\mathbf{a}$ to $\mathbf{b}$. Moreover, if σ is an even (resp. odd) permutation, then an intersecting path from $\sigma\mathbf{a}$ is sent to an intersecting path from $\tau\sigma\mathbf{a}$, where τ is a transposition and so $\tau\sigma$ is an odd (resp. even) permutation. It follows that the number of intersecting paths that start at even permutations of $\mathbf{a}$ is equal to the number of intersecting paths that start at odd permutations of $\mathbf{a}$. This means that

$$\sum_{\substack{\sigma \in S_h \\ \epsilon(\sigma)=1}} I(\sigma\mathbf{a}, \mathbf{b}) = \sum_{\substack{\sigma \in S_h \\ \epsilon(\sigma)=-1}} I(\sigma\mathbf{a}, \mathbf{b}).$$

This statement is equivalent to the Lemma. □

Let $[x]_k$ denote the polynomial $x(x-1)\cdots(x-k+1)$. If $b_i, \sigma(i)$ are nonnegative integers, then

$$\begin{aligned}[b_i]_{\sigma(i)} &= b_i(b_i-1)(b_i-2)\cdots(b_i-\sigma(i)+1)\\ &= \begin{cases} b_i!/(b_i-\sigma(i))! & \text{if } \sigma(i) \leq b_i \\ 0 & \text{if } \sigma(i) > b_i. \end{cases}\end{aligned}$$

Theorem 1. *Let $h \geq 2$, and let $b_0, b_1, \ldots, b_{h-1}$ be integers such that*

$$0 \leq b_0 < b_1 < \cdots < b_{h-1}.$$

Then

$$\hat{B}(b_0+b_1+\cdots+b_{h-1}) = \frac{(b_0+b_1+\cdots+b_{h-1}-\binom{h}{2})!}{b_0!b_1!\cdots b_{h-1}!} \prod_{0\leq i<j\leq h-1}(b_j-b_i).$$

Proof. Let $\mathbf{a}^* = (0,1,2,\ldots,h-1)$, and let $\mathbf{b} = (b_0,b_1,\ldots,b_{h-1}) \in \mathbf{Z}^h$. Applying the previous lemmas, we obtain

$$\begin{aligned}
&\hat{B}(b_0,b_1,\ldots,b_{h-1}) =\\
&\quad = J(\mathbf{a}^*,\mathbf{b})\\
&\quad = P(\mathbf{a}^*,\mathbf{b}) - I(\mathbf{a}^*,\mathbf{b})\\
&\quad = P(\mathbf{a}^*,\mathbf{b}) + \sum_{\substack{\sigma\in S_h\\ \sigma\neq \mathrm{id}}} \varepsilon(\sigma) I(\sigma\mathbf{a}^*,\mathbf{b})\\
&\quad = P(\mathbf{a}^*,\mathbf{b}) + \sum_{\substack{\sigma\in S_h\\ \sigma\neq \mathrm{id}}} \varepsilon(\sigma) P(\sigma\mathbf{a}^*,\mathbf{b})\\
&\quad = \sum_{\sigma\in S_h} \varepsilon(\sigma) P(\sigma\mathbf{a}^*,\mathbf{b})\\
&\quad = \sum_{\substack{\sigma\in S_h\\ \sigma a^*\leq b}} \varepsilon(\sigma) \frac{(b_0+\cdots+b_{h-1}-\binom{h}{2})!}{\prod_{i=0}^{h-1}(b_i-\sigma(i))!}\\
&\quad = \frac{\left(b_0+\cdots+b_{h-1}-\binom{h}{2}\right)!}{b_0!b_1!\cdots b_{h-1}!} \sum_{\substack{\sigma\in S_h\\ \sigma a^*\neq b}} \varepsilon(\sigma)[b_0]_{\sigma(0)}[b_1]_{\sigma(1)}\cdots[b_{h-1}]_{\sigma(h-1)}\\
&\quad = \frac{(b_0+\cdots+b_{h-1}-\binom{h}{2})!}{b_0!b_1!\cdots b_{h-1}!} \sum_{\sigma\in S_h} \varepsilon(\sigma)[b_0]_{\sigma(0)}[b_1]_{\sigma(1)}\cdots[b_{h-1}]_{\sigma(h-1)}
\end{aligned}$$

$$= \frac{\left(b_0+\cdots+b_{h-1}-\binom{h}{2}\right)!}{b_0!b_1!\cdots(b_{h-1}!} \begin{vmatrix} 1 & [b_0]_1 & [b_0]_2 & \cdots & [b_0]_{h-1} \\ 1 & [b_1]_1 & [b_1]_2 & \cdots & [b_1]_{h-1} \\ \vdots & & & & \\ 1 & [b_{h-1}]_1 & [b_{h-1}]_2 & \cdots & [b_{h-1}]_{h-1} \end{vmatrix}$$

$$= \frac{\left(b_0+\cdots+b_{h-1}-\binom{h}{2}\right)!}{b_0!b_1!\cdots b_{h-1}!} \prod_{0\le i<j\le h-1} (b_j-b_i).$$

This completes the proof. □

We state the following corollary with the notation that is used later in the proof of the Erdős-Heilbronn conjecture.

Corollary 1. *Let $h \geq 2$, let p be a prime number, and let $i_0, i_1, \ldots, i_{h-1}$ be integers such that*

$$0 \le i_0 < i_1 < \cdots < i_{h-1} < p$$

and

$$i_0 + i_1 + \cdots i_{h-1} < \binom{h}{2} + p.$$

Then

$$\hat{B}(i_0, i_1, \ldots, i_{h-1}) \not\equiv 0 \pmod{p}.$$

Proof. This follows immediately from the Theorem. □

3. A Review of Linear Algebra

Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear operator. Let $I : V \to V$ be the identity operator. For every nonnegative integer i, we define $T^i : V \to V$ by

$$T^0(\mathbf{v}) = I(\mathbf{v}) = \mathbf{v},$$

$$T^i(\mathbf{v}) = T(T^{i-1}(\mathbf{v}))$$

for all $\mathbf{v} \in V$. To every polynomial

$$p(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 \in F[x]$$

we associate the linear operator $p(T) : V \to V$ defined by

$$p(T) = c_n T^n + c_{n-1}T^{n-1} + \cdots + c_1 T + c_0 I.$$

The set of all polynomials $p(x)$ such that $p(T) = 0$ forms a nonzero, proper ideal J in the polynomial ring $F[x]$. Since every ideal in $F[x]$ is principal, there exists a unique monic polynomial $p_T(x) = p_{T,V}(x) \in J$ such that $p_T(x)$

divides every other polynomial in J. This polynomial is called the *minimal polynomial* of T over the vector space V.

A subspace W of V is called *invariant* with respect to T if $T(W) \subseteq W$, that is, if $T(\mathbf{w}) \in W$ for all $\mathbf{w} \in W$. Then T restricted to the subspace W is a linear operator on W with minimal polynomial $p_{T,W}(x)$. Since $p_{T,V}(T)(\mathbf{w}) = \mathbf{0}$ for all $\mathbf{w} \in W$, it follows that $P_{T,W}(x)$ divides $p_{T,V}(x)$, and so

$$\deg(p_{T,W}) \leq \deg(p_{T,V}), \tag{2}$$

where $\deg(p)$ denotes the degree of the polynomial p.

For $\mathbf{v} \in V$, the *cyclic subspace* with respect to T generated by $\mathbf{v}$ is the smallest subspace of V that contains $\mathbf{v}$ and is invariant under the operator T. We denote this subspace by $C_T(\mathbf{v})$, Let $\mathbf{v}_i = T^i(\mathbf{v})$ for $i = 0, 1, 2, \ldots$. Then $C_T(\mathbf{v})$ is the subspace generated by the vectors

$$\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), T^3(\mathbf{v}), \ldots\} = \{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots\}$$

and

$$\dim(C_T(\mathbf{v})) = l,$$

where l is the smallest integer such the vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_l$ are linearly dependent. This means that there exist scalars $c_0, c_1, \ldots, c_{l-1}$ in the field F such that

$$\mathbf{v}_l + c_{l-1}\mathbf{v}_{l-1} + \cdots + c_1\mathbf{v}_1 + c_0\mathbf{v}_0 = \mathbf{0}.$$

Let

$$p(x) = x^l + c_{l-1}x^{l-1} + \cdots + c_1 x + c_0.$$

Then

$$\begin{aligned} p(T)(\mathbf{v}_0) &= T^l(\mathbf{v}_0) + c_{l-1}T^{l-1}(\mathbf{v}_0) + \cdots + c_1 T(\mathbf{v}_0) + c_0 I(\mathbf{v}_0) \\ &= \mathbf{v}_l + c_{l-1}\mathbf{v}_{l-1} + \cdots + c_1\mathbf{v}_1 + c_0\mathbf{v}_0 \\ &= \mathbf{0}, \end{aligned}$$

and so

$$p(T)(\mathbf{v}_i) = p(T)(T^i(\mathbf{v}_0)) = T^i(p(T)(\mathbf{v}_0)) = T^i(\mathbf{0}) = \mathbf{0}$$

for $i = 0, 1, 2, \ldots$. Therefore, $p(T) = 0$ on the cyclic subspace $C_T(\mathbf{v}) = C$, and so $p(x)$ is divisible by the minimal polynomial $P_{T,C}(x)$, and

$$m = \deg(p_{T,C}) \leq \deg(p) = l.$$

On the other hand, since

$$p_{T,C}(T)(\mathbf{v}) = \mathbf{0},$$

it follows that the vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent, and so

$$l \leq m.$$

This implies that $l = m$ and so, by inequality (2),

$$\dim(C_T(\mathbf{v})) = \deg(p_{T,C}) \leq \deg(p_{T,V})$$

for all $\mathbf{v} \in V$.

If $T(\mathbf{f}) = a\mathbf{f}$ for some $a \in F$ and some nonzero vector $\mathbf{f} \in V$, then a is called an *eigenvalue* of T and $\mathbf{f}$ is called an *eigenvector* of T with eigenvalue a. The *spectrum* of T, denoted $\sigma(T)$, is the set of all eigenvalues of T. If V has a basis consisting entirely of eigenvectors of T, then T is called a *diagonal operator*.

The following inequality plays a central role in the proof of the Erdős-Heilbronn conjecture.

Lemma 4. *Let T be a diagonal linear operator on a finite-dimensional vector space V, and let $\sigma(T)$ be the spectrum of T. Then*

$$\dim(C_T(\mathbf{v})) \leq |\sigma(T)| \tag{3}$$

for every $\mathbf{v} \in V$.

Proof. Let $a \in \sigma(T)$, and let $\mathbf{f}$ be an eigenvector with eigenvalue a. Let W be the one-dimensional subspace generated by $\mathbf{f}$. Then W is invariant with respect to T, and $p_{T,W}(x) = x - a$. It follows that $x - a$ divides $p_{T,V}(x)$, and so

$$\prod_{a \in \sigma(T)} (x - a)$$

divides $p_{T,V}(x)$. Let dim $(V) = k$. If T is a diagonal linear operator, then V has a basis $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ of eigenvectors, and

$$\prod_{a \in \sigma(T)} (T - a)(\mathbf{f}_i) = \mathbf{0}$$

for $i = 0, 1, \ldots, k-1$. It follows that $\prod_{a \in \sigma(T)}(T - a)(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$, and so

$$p_{T,V}(x) = \prod_{a \in \sigma(T)} (x - a).$$

In particular, the degree of $p_{T,V}(x)$ is equal to the number of distinct eigenvalues of T. It follows that, if T is a diagonal operator on a finite-dimensional vector space V, then

$$\dim(C_T(\mathbf{v})) \leq \deg(p_{T,V}) = |\sigma(T)|$$

for every $\mathbf{v} \in V$. This completes the proof. □

Lemma 5. *Let $T : V \to V$ be a linear operator on the vector space V, and let $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ be eigenvectors of T with distinct eigenvalues. Let*

$$\mathbf{v}_0 = \mathbf{f}_0 + \mathbf{f}_1 + \cdots + \mathbf{f}_{k-1},$$

and let $C_T(\mathbf{v}_0)$ be the cyclic subspace generated by $\mathbf{v}_0$. Then

$$\dim(C_T(\mathbf{v}_0)) = k$$

and

$$\{\mathbf{v}_0, T(\mathbf{v}_0), T^2(\mathbf{v}_0), \ldots, T^{k-1}(\mathbf{v}_0)\}$$

is a basis for $C_T(\mathbf{v}_0)$. If $\dim(V) = k$, then $C_T(\mathbf{v}_0) = V$.

Proof. We first show that the vectors $\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}$ are linearly independent. If they are linearly dependent, then there is a minimal subset of the vectors $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ that is linearly dependent, say, $\mathbf{f}_0, \ldots, \mathbf{f}_{l-1}$. Moreover, $l \geq 2$ since $\mathbf{f}_i \neq \mathbf{0}$ for $i = 0, 1, \ldots, k-1$. There exist nonzero scalars $c_0, c_1, \ldots, c_{l-1}$ such that $\sum_{i=0}^{l-1} c_i \mathbf{f}_i = \mathbf{0}$. Let $a_i \in \sigma((T))$ be the eigenvalue corresponding to the eigenvector $\mathbf{f}_i$. Then

$$T\left(\sum_{i=0}^{l-1} c_i(\mathbf{f}_i)\right) = \sum_{i=0}^{l-1} c_i T(\mathbf{f}_i) = \sum_{i=0}^{l-1} c_i a_i \mathbf{f}_i = \mathbf{0}.$$

Since

$$\sum_{i=0}^{l-1} c_i a_{l-1} \mathbf{f}_i = a_{l-1} \sum_{i=0}^{l-1} c_i \mathbf{f}_i = \mathbf{0},$$

it follows that

$$\sum_{i=0}^{l-1} c_i(a_i - a_{l-1})\mathbf{f}_i = \sum_{i=0}^{l-2} c_i(a_i - a_{l-1})\mathbf{f}_i = \mathbf{0},$$

which contradicts the minimality of l, since $c_i(a_i - a_{l-1}) \neq 0$ for $i < l - 1$. Thus, the vectors $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ are linearly independent, and span a k-dimensional subspace W of V. Moreover, W is an invariant subspace, since it has a basis of eigenvectors of T. Since

$$\mathbf{v}_0 = \mathbf{f}_1 + \cdots + \mathbf{f}_k \in W,$$

it follows that

$$C_T(\mathbf{v}_0) \subseteq W$$

and so

$$\dim(C_T(\mathbf{v}_0)) \leq \dim(W) = k.$$

The vector $T^i(\mathbf{v}_0) \in C_T(\mathbf{v}_0)$ for every nonnegative integer i. Since

$$T^i(\mathbf{v}_0) = a_0^i \mathbf{f}_0 + a_1^i \mathbf{f}_1 + \cdots + a_{k-1}^i \mathbf{f}_{k-1},$$

the matrix of the set of vectors $\{\mathbf{v}_0, T(\mathbf{v}_0), T^2(\mathbf{v}_0), \ldots, T^{k-1}(\mathbf{v}_0)\}$ with respect to the basis $\{\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}\}$ is

$$\begin{pmatrix} 1 & a_0 & a_0^2 & \cdots & a_0^{k-1} \\ 1 & a_1 & a_1^2 & \cdots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{k-1} & a_{k-1}^2 & \cdots & a_{k-1}^{k-1} \end{pmatrix},$$

and its determinant is the Vandermonde determinant

$$\prod_{0 \leq i < j \leq h} (a_j - a_i) \neq 0.$$

It follows that $\{\mathbf{v}_0, T(\mathbf{v}_0), T^2(\mathbf{v}_0), \ldots, T^{k-1}(\mathbf{v}_0)\}$ is a set of linearly independent vectors, and so

$$\dim(C_T(\mathbf{v}_0)) \geq k = \dim(W).$$

Therefore, $\dim(C_T(\mathbf{v}_0)) = k$. If $\dim(\mathbf{V}) = k$, then $C_T(\mathbf{v}_0) = V$. This completes the proof. □

4. Alternating Products

Let $\wedge^h V$ denote the h-th *alternating product* of the vector space V. Then $\wedge^h V$ is a vector space whose elements are linear combinations of expressions of the form

$$\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1},$$

where $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{h-1} \in V$. These *wedge products* have the property that

$$\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} = \mathbf{0}$$

if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$, and

$$\mathbf{v}_{\sigma(0)} \wedge \mathbf{v}_{\sigma(1)} \wedge \cdots \wedge \mathbf{v}_{\sigma(h-1)} = \varepsilon(\sigma)\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1}$$

for all $\sigma \in S_h$.

If $\{\mathbf{e}_0, \ldots, \mathbf{e}_{k-1}\}$ is a basis for V, then a basis for $\wedge^h V$ is the set of all wedge products of the form

$$\mathbf{e}_{i_0} \wedge \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_{h-1}},$$

where

$$0 \leq i_0 < i_1 < \ldots < i_{h-1} \leq k-1,$$

and

$$\dim(\wedge^h V) = \binom{k}{h}.$$

Every linear operator $T : V \to V$ induces a linear operator

$$DT : \bigwedge^h V \to \bigwedge^h V$$

that acts on wedge products according to the rule

$$DT(\mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}) =$$
$$\sum_{j=0}^{h-1} \mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}} \wedge T(\mathbf{v}_{i_j}) \wedge \mathbf{v}_{i_{j+1}} \cdots \wedge \mathbf{v}_{i_{h-1}}. \tag{4}$$

The operator DT is called the *derivative* of T.

Lemma 6. *Let T be a diagonal linear operator on V, and let $\sigma(T)$ be the spectrum of T. For $h \geq 2$, let $DT : \bigwedge^h V \to \bigwedge^h V$ be the derivative of T. If T has distinct eigenvalues, that is, if $|\sigma(T)| = \dim(V)$, then*

$$\sigma(DT) = h^\wedge \sigma(T)$$

and

$$|h^\wedge \sigma(T)| \geq \dim(C_{DT}(\mathbf{w})) \tag{5}$$

for every $\mathbf{w} \in \wedge^h V$.

Proof. Let $\sigma(T) = \{a_0, a_1, \ldots, a_{k-1}\}$, and let $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ be a basis of eigenvectors of V such that $T(\mathbf{f}_i) = a_i \mathbf{f}_i$ for $i = 0, 1, \ldots, k-1$. Then (4) implies that

$$DT(\mathbf{f}_{i_0} \wedge \cdots \wedge \mathbf{f}_{i_{h-1}}) = (a_{i_0} + \cdots + a_{i_{h_1}})(\mathbf{f}_{i_0} \wedge \cdots \wedge \mathbf{f}_{i_{h-1}}).$$

It follows that DT is a diagonal linear operator on $\wedge^h V$, and its spectrum $\sigma(DT)$ consists of all sums of h distinct eigenvalues of T, that is,

$$\sigma(DT) = h^\wedge \sigma(T).$$

Applying inequality (3) to the vector space $\wedge^h V$ and the operator DT, we obtain

$$|h^\wedge \sigma(T)| = |\sigma(DT)| \geq \dim(C_{DT}(W))$$

for every $\mathbf{w} \in \wedge^h V$. This completes the proof. □

Theorem 2. *Let T be a linear operator on the finite-dimensional vector space V. Let $h \geq 2$, and let $DT : \wedge^h V \to \wedge^h V$ be the derivative of T. For $\mathbf{v}_0 \in V$, define*

$$\mathbf{v}_i = T^i(\mathbf{v}_0) \in V$$

for $i \geq 1$, and let

$$\mathbf{w} = \mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} \in \wedge^h V.$$

Then for every $r \geq 0$

$$\begin{aligned}(DT)^r(\mathbf{w}) &= (DT)^r(\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1}) \\ &= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1})\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}},\end{aligned}$$

where the sum is over all integer lattice points $(i_0, i_1, \ldots, i_{h-1}) \in \mathbf{Z}^h$ *such that*

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq r + h - 1$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r,$$

and where $\hat{B}(i_0, i_1, \ldots, i_{h-1})$ *is the strict* h*-dimensional ballot number corresponding to the lattice point* $(i_0, i_1, \ldots, i_{h-1})$.

Proof. The proof will be by induction on r. Let $r = 0$. Since

$$\hat{B}(0, 1, 2, \ldots, h-1) = \mathbf{1},$$

we have

$$\begin{aligned}(DT)^0(\mathbf{w}) &= \mathbf{w} \\ &= \mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} \\ &= \hat{B}(0, 1, 2, \ldots, h-1)\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1}.\end{aligned}$$

Suppose the result holds for some integer $r \geq 0$. Then

$$\begin{aligned}&(DT)^{r+1}(\mathbf{w}) = DT((DT)^r(\mathbf{w})) \\ &= DT\left(\sum \hat{B}(i_0, i_1, \ldots, i_{h-1})\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}\right) \\ &= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1})DT(\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}) \\ &= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1})\sum_{j=0}^{h-1}(\mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{j-1}} \wedge T(\mathbf{v}_{i_j}) \wedge \mathbf{v}_{i_{j+1}} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}) \\ &= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1})\sum_{j=0}^{h-1}(\mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{j-1}} \wedge \mathbf{v}_{i_{j+1}} \wedge \mathbf{v}_{i_{j+1}} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}) \\ &= \sum C(i_0, i_1, \ldots, i_{h-1})\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}},\end{aligned}$$

where the last sum is over all integer lattice points $(i_0, i_1, \ldots, i_{h-1}) \in \mathbf{Z}^n$ such that

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq r + h$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r + 1,$$

and the integer $C(i_0, i_1, \ldots, i_{h-1})$ satisfies the difference equation

$$C(i_0, i_1, \ldots, i_{h-1}) = \sum_{j=0}^{h-1} \hat{B}(i_0, \ldots, i_{j-1}, i_j - 1, i_{j+1}, \ldots, i_{h-1})$$

This difference equation determines the strict h-dimensional ballot numbers, and so

$$C(i_0, i_1, \ldots, i_{h-1}) = \hat{B}(i_0, i_1, \ldots, i_{h-1}).$$

Therefore, the result holds in the case $r + 1$. This completes the induction.□

5. Erdős-Heilbronn, Concluded

Theorem 3. *Let p be a prime number, and let $A \subseteq \mathbf{Z}/p\mathbf{Z}$, where $|A| = k$. Let $2 \leq h \leq k$. Then*

$$|h^\wedge A| \geq \min(p, hk - h^2 + 1).$$

Proof. Let $A = \{a_0, a_1, \ldots, a_{k-1}\}$. Let V be a vector space of dimension k over the field $\mathbf{Z}/p\mathbf{Z}$, and let $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ be a basis for V. We define the diagonal linear operator $T : V \to V$ by

$$T(\mathbf{f}_i) = a_i \mathbf{f}_i$$

for $i = 0, 1, \ldots, k - 1$. The spectrum of T is

$$\sigma(T) = A.$$

Let

$$\mathbf{v}_0 = \mathbf{f}_0 + \mathbf{f}_1 + \cdots + \mathbf{f}_{k-1},$$

and define

$$\mathbf{v}_{i+1} = T(\mathbf{v}_i) = T^i(\mathbf{v}_0)$$

for $i \geq 0$. By Lemma 5, the cyclic subspace generated by $\mathbf{v}_0$ is V, and the set of vectors $\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$ is a basis for V. The alternating product $\bigwedge^h V$ is a vector space with a basis consisting of the $\binom{k}{h}$ wedge products of the form

$$\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}},$$

where

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq k - 1.$$

Let

$$\mathbf{w} = \mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} \in \bigwedge^h V.$$

By inequality (5),

$$|h^\wedge A| = |\sigma(DT)| \geq \dim C_{DT}(\mathbf{w}).$$

Therefore, it suffices to prove that

$$\dim C_{DT}(\mathbf{w}) \geq \min(p, hk - h^2 + 1).$$

This is equivalent to proving that the vectors

$$\mathbf{w}, (DT)(\mathbf{w}), (DT)^2(\mathbf{w}), \ldots, (DT)^n(\mathbf{w})$$

are linearly independent in the alternating product $\bigwedge^h V$, where

$$n = \min(p, hk - h^2 + 1) - 1 = \min(p - 1, hk - h^2).$$

Let $0 \leq r \leq n$. By Theorem 2, the vector $(DT)^r(\mathbf{w})$ is a linear combination of the vectors

$$\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$$

such that

$$0 \leq i_0 < i_1 < \ldots < i_{h-1} \leq r + h - 1$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r. \tag{6}$$

Let I be the interval of integers $[0, k-1]$. Since

$$h^\wedge I = \left[\binom{h}{2}, hk - \binom{h+1}{2}\right] = \binom{h}{2} + [0, hk - h^2],$$

it follows that there is at least one basis vector $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h_1}}$ in the expansion of $(DT)^r(\mathbf{w})$ such that

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq k - 1 < k \leq p$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r \leq \binom{h}{2} + n < \binom{h}{2} + p.$$

By Theorem 2, the coefficient of this basis vector is the strict h-dimensional ballot number $\hat{B}(i_0, i_1, \ldots, i_{h-1})$ and

$$\hat{B}(i_0, i_1, \ldots, i_{h-1}) \not\equiv 0 \pmod{p}$$

by Corollary 1.

Suppose that the vector $\mathbf{v}_{i_0}\wedge\mathbf{v}_{i_1}\wedge\cdots\wedge\mathbf{v}_{ih_1}$ satisfies $i_l \geq k$ for some l. Since every vector $\mathbf{v}_l \in V$ with $l \geq k$ is a linear combination of $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$, it follows that $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{ih_1}$ is a linear combination of basis vectors of the form $\mathbf{v}_{j_0} \wedge \mathbf{v}_{j_1} \wedge \cdots \wedge \mathbf{v}_{j_{h-1}}$, where

$$0 \leq j_0 < j_1 < \cdots < j_{h-1} \leq k-1$$

and

$$j_0 + j_1 + \cdots + j_{h-1} < \binom{h}{2} + r.$$

It follows that $(DT)^r(\mathbf{w})$ is a linear combination of basis vectors $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$ such that either

$$i_0 + i_1 + \cdots + i_{h-1} < \binom{h}{2} + r$$

or

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r,$$

and in the latter case the basis vector appears with a coefficient that is nonzero modulo p. This implies that the vectors $\mathbf{w}, (DT)(\mathbf{w}), \ldots, (DT)^n(\mathbf{w})$ are linearly independent in the cyclic subspace $C_{DT}(\mathbf{w})$, and the proof of the Erdős-Heilbronn conjecture is complete. □

6. Remarks

This proof only requires that A be a subset of a field, and does not require that the field be $\mathbf{Z}/p\mathbf{Z}$. Let F be an arbitrary field. Let p be the characteristic of F if the characteristic is positive, and let $p = \infty$ if the characteristic is zero. Then we have, in fact, proved that if $A \subseteq F$ and $|A| = k \leq p$, then $|h^\wedge A| \geq \min(p, hk - h^2 + 1)$ for all $h \geq 1$.

The Cauchy-Davenport theorem was proved by Cauchy [3] in 1813. Davenport [5] rediscovered the result in 1935. I. Chowla [4] immediately extended the Cauchy-Davenport theorem to composite moduli. Other generalizations have been obtained by Pillai [14], Shatrovskii [20], Pollard [15, 16], Brakemaier [2], and Hamidoune [12]. Davenport [6] discovered in 1947 that Cauchy had proved the Cauchy-Davenport theorem first.

The Erdős-Heilbronn conjecture originated in the 1960s. Partial results on the Erdős-Heilbronn conjecture were obtained by Rickert [18], Mansfield [13], Rödseth [19], Pyber [17], and Freiman, Low, and Pitman [11]. Dias da Silva and Hamidoune [8] proved the conjecture by using results from representation theory and linear algebra. This algebraic technique had previously been applied to additive number theory by Dias da Silva and Hamidoune [7] and Spigler [21].

The derivation of the formula for the strict ballot number $\hat{B}(b_0, \ldots, b_{n-1})$ follows a paper of Zeilberger [22].

Using combinatorial ideas from the proof of Dias da Silva and Hamidoune, Alon, Nathanson, and Ruzsa [1] found a different proof of the Erdős-Heilbronn conjecture that uses only the simplest properties of polynomials.

References

1. N. Alon, M. B. Nathanson, and I. Z. Ruzsa. Adding distinct congruence classes modulo a prime. *Amer. Math. Monthly*, 102, 1995.
2. W. Brakemeier. Eine Anzahlformel von Zahlen modulo *n*. *Monat. Math.*, 85:277–282, 1978.
3. A. L. Cauchy. Recherches sur les nombres. J. *École polytech.*, 9:99–116, 1813.
4. I. Chowla. A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring's problem. *Proc. Indian Acad. Sci., Section A*, 1:242–243, 1935.
5. H. Davenport. On the addition of residue classes. *J. London Math. Soc.*, 10:30–32, 1935.
6. H. Davenport. A historical note. *J. London Math. Soc.,* 22:100–101, 1947.
7. J. A. Dias da Silva and Y. O. Hamidoune. A note on the minimal polynomial of the Kronecker sum of two linear operators. *Linear Algebra and its Applications*, 141:283–287, 1990.
8. J. A. Dias da Silva and Y. O. Hamidoune. Cyclic spaces for Grassmann derivatives and additive theory. *Bull. London Math. Soc.*, 26:140–146, 1994.
9. P. Erdős and R. L. Graham. *Old and New Problems and Results in Combinatorial Number Theory.* L'Enseignement Mathématique, Geneva, 1980.
10. P. Erdős and H. Heilbronn. On the addition of residue classes mod *p*. *Acta Arith.*, 9:149–159, 1964.
11. G. A. Freiman, L. Low, and J. Pitman. The proof of Paul Erdős' conjecture of the addition of different residue classes modulo prime number. In *Structure Theory of Set Addition, 7–11 June 1993, CIRM Marseille, pages* 99–108, 1993.
12. Y. O. Hamidoune. A generalization of an addition theorem of Shatrowsky. *Europ. J. Combin.*, 13:249–255, 1992.
13. R. Mansfield. How many slopes in a polygon? *Israel J. Math.*, 39:265–272, 1981.
14. S. S. Pillai. Generalization of a theorem of Davenport on the addition of residue classes. *Proc. Indian Acad. Sci. (Series A)*, 6:179–180,1938.
15. J. M. Pollard. A generalization of a theorem of Cauchy and Davenport. *J. London Math. Soc.*, 8:460–462, 1974.
16. J. M. Pollard. Additive properties of residue classes. *J. London Math. Soc.*, 11:147–152,1975.
17. L. Pyber. On the Erdős-Heilbronn conjecture. Personal communication.
18. U.-W. Rickert. *Über eine Vermutung in der additiven Zahlentheorie.* PhD thesis, Tech. Univ. Braunschweig, 1976.
19. O. Rödseth, Sums of distinct residues mod *p*. *Acta Arith.*, 65:181–184, 1993.
20. L. Shatrovskii. A new generalization of Davenport's-Pillai's theorem on the addition of residue classes. *Doklady Akad. Nauk CCCR*, 45:315–317, 1944.
21. R. Spigler. An application of group theory to matrices and to ordinary differential equations. *Linear Algebra and its Applications*, 44:143–151, 1982.
22. D. Zeilberger. André's reflection proof generalized to the many-candidate ballot problem. *Discrete Math.*, 44:325–326, 1983.

On Landau's Function $g(n)$

Jean-Louis Nicolas

J.-L. Nicolas (✉)
Institut Camille Jordan, Mathématiques, Université de Lyon, CNRS,
Université Lyon 1, 43 Bd. du 11 Novembre 1918, F-69622 Villeurbanne
Cedex, France
e-mail: jlnicola@in2p3.fr; http://math.univ-lyon1.fr/~nicolas/

1. Introduction

Let S_n be the symmetric group of n letters. Landau considered the function $g(n)$ defined as the maximal order of an element of S_n; Landau observed that (cf. [9])

$$g(n) = \max \operatorname{lcm}(m_1, \ldots, m_k) \tag{1}$$

where the maximum is taken on all the partitions $n = m_1 + m_2 + \cdots + m_k$ of n and proved that, when n tends to infinity

$$\log g(n) \sim \sqrt{n \log n}. \tag{2}$$

More precise asymptotic estimates have been given in [11, 22, 25]. In [25] and [11] one also can find asymptotic estimates for the number of prime factors of $g(n)$. In [8] and [3], the largest prime factor $P^+(g(n))$ of $g(n)$ is investigated. In [10] and [12], effective upper and lower bounds of $g(n)$ are given. In [17], it is proved that $\lim_{n\to\infty} g(n+1)/g(n) = 1$. An algorithm able to calculate $g(n)$ up to 10^{15} is given in [2] (see also [26]). The sequence of distinct values of $g(n)$ is entry A002809 of [24]. A nice survey paper was written by W. Miller in 1987 (cf. [13]).

My very first mathematical paper [15] was about Landau's function, and the main result was that $g(n)$, which is obviously non decreasing, is constant on arbitrarily long intervals (cf. also [16]). I first met A. Schinzel in Paris in May 1967. He told me that he was interested in my results, but that P. Erdős would be more interested than himself. Then I wrote my first letter to Paul with a copy of my work. I received an answer dated of June 12 1967 saying "I sometimes thought about $g(n)$ but my results were very much less complete than yours". Afterwards, I met my advisor, the late Professor Pisot, who, in view of this letter, told me that my work was good for a thesis.

The main idea of my work about $g(n)$ was to use the tools introduced by S. Ramanujan to study highly composite numbers (cf. [19, 20]). P. Erdős was very well aware of this paper of Ramanujan (cf. [1, 4–6]) as well as of the symmetric group and the order of its elements, (cf. [7]) and I think that he enjoyed the connection between these two areas of mathematics. Anyway, since these first letters, we had many occasions to discuss Landau's function.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_14,

Let us define $n_1 = 1$, $n_2 = 2$, $n_3 = 3$, $n_4 = 4$, $n_5 = 5$, $n_6 = 7$, etc. $\dots, n_k$ (see a table of $g(n)$ in [16, p. 187]), such that

$$g(n_k) > g(n_k - 1). \tag{3}$$

The above mentioned result can be read:

$$\overline{\lim}(n_{k+1} - n_k) = +\infty. \tag{4}$$

Here, I shall prove the following result:

Theorem 1.

$$\underline{\lim}(n_{k+1} - n_k) < +\infty. \tag{5}$$

Let us set $p_1 = 2, p_2 = 3, p_3 = 5, \dots$, $p_k =$ the k-th prime. It is easy to deduce Theorem 1 from the twin prime conjecture (i.e. $\underline{\lim}(p_{k+1} - p_k) = 2$) or even from the weaker conjecture $\underline{\lim}(p_{k+1} - p_k) < +\infty$. (cf. Sect. 4 below.) But I shall prove Theorem 1 independently of these deep conjectures. Moreover I shall explain below why it is reasonable to conjecture that the mean value of $n_{k+1} - n_k$ is 2; in other terms one may conjecture that

$$n_k \sim 2k \tag{6}$$

and that $n_{k+1} - n_k = 2$ has infinitely many solutions. Due to a parity phenomenon, $n_{k+1} - n_k$ seems to be much more often even than odd; nevertheless, I conjecture that:

$$\underline{\lim}(n_{k+1} - n_k) = 1. \tag{7}$$

The steps of the proof of Theorem 1 are first to construct the set G of values of $g(n)$ corresponding to the so called superior highly composite numbers introduced by S. Ramanujan, and then, when $g(n) \in G$, to build the table of $g(n+d)$ when d is small. This will be done in Sects. 4 and 5. Such values of $g(n+d)$ will be linked with the number of distinct differences of the form $P-Q$ where P and Q are primes satisfying $x - x^\alpha \leq Q \leq x < P \leq x + x^\alpha$, where x goes to infinity and $0 < \alpha < 1$. Our guess is that these differences $P - Q$ represent almost all even numbers between 0 and $2x^\alpha$, but we shall only prove in Sect. 3 that the number of these differences is of the order of magnitude of x^α, under certain strong hypothesis on x and α, and for that a result due to Selberg about the primes between x and $x + x^\alpha$ will be needed (cf. Sect. 2).

To support conjecture (6), I think that what has been done here with $g(n) \in G$ can also be done for many more values of $g(n)$, but, unfortunately, even assuming strong hypotheses, I do not see for the moment how to manage it.

I thank very much E. Fouvry who gave me the proof of Proposition 2.

1.1 Notation

p will denote a generic prime, p_k the k-th prime; P, Q, P_i, Q_j will also denote primes. As usual $\pi(x) = \sum_{p \le x} 1$ is the number of primes up to x.

$|S|$ will denote the number of elements of the set S. The sequence n_k is defined by (3).

2. About the Distribution of Primes

Proposition 1. *Let us define $\pi(x) = \sum_{p\le x} 1$, and let α be such that $\frac{1}{6} < \alpha < 1$, and $\varepsilon > 0$. When ξ goes to infinity, and $\xi' = \xi + \xi/\log \xi$, then for all x in the interval $[\xi, \xi']$ but a subset of measure $O((\xi' - \xi)/\log^3 \xi)$ we have:*

$$\left| \pi(x + x^\alpha) - \pi(x) - \frac{x^\alpha}{\log x} \right| \le \varepsilon \frac{x^\alpha}{\log x} \tag{8}$$

$$\left| \pi(x) - \pi(x - x^\alpha) - \frac{x^\alpha}{\log x} \right| \le \varepsilon \frac{x^\alpha}{\log x} \tag{9}$$

$$\left| \frac{x}{\log x} - \frac{Q^k - Q^{k-1}}{\log Q} \right| \ge \frac{\sqrt{x}}{\log^4 x} \text{ for all primes } Q, \text{ and } k \ge 2. \tag{10}$$

Proof. This proposition is an easy extension of a result of Selberg (cf. [21]) who proved that (8) holds for most x in (ξ, ξ'). In [18], I gave a first extension of Selberg's result by proving that (8) and (9) hold simultaneously for all x in (ξ, ξ') but for a subset of measure $O((\xi' - \xi)/\log^3 \xi)$. So, it suffices to prove that the measure of the set of values of x in (ξ, ξ') for which (10) does not hold is $O((\xi' - \xi)/\log^3 \xi)$.

We first count the number of primes Q such that for one k we have:

$$\frac{\xi}{\log \xi} \le \frac{Q^k - Q^{k-1}}{\log Q} \le \frac{\xi'}{\log \xi'}. \tag{11}$$

If Q satisfies (11), then $k \le \frac{\log \xi'}{\log 2}$ for ξ' large enough. Further, for k fixed, (11) implies that $Q \le (\xi')^{1/k}$, and the total number of solutions of (11) is

$$\le \sum_{k=2}^{\log \xi'/\log 2} (\xi')^{1/k} = O(\sqrt{\xi'}) = O(\sqrt{\xi}).$$

With a more careful estimation, this upper bound could be improved, but this crude result is enough for our purpose. Now, for all values of $y = \frac{Q^k - Q^{k-1}}{\log Q}$ satisfying (11), we cross out the interval $\left(y - \frac{\sqrt{\xi'}}{\log^4 \xi'}, y + \frac{\sqrt{\xi'}}{\log^4 \xi'}\right)$. We also cross out this interval whenever $y = \frac{\xi}{\log \xi}$ and $y = \frac{\xi'}{\log \xi'}$. The total sum of the lengths of the crossed out intervals is $O\left(\frac{\xi}{\log^4 \xi}\right)$, which is smaller than the length of the interval $\left(\frac{\xi}{\log \xi}, \frac{\xi'}{\log \xi'}\right)$ and if $\frac{x}{\log x}$ does not fall into one of these forbidden intervals, (10) will certainly hold. Since the derivative of the function $\varphi(x) = x/\log x$ is $\varphi'(x) = \frac{1}{\log x} - \frac{1}{\log^2 x}$ and satisfies $\varphi'(x) \sim \frac{1}{\log \xi}$ for all $x \in (\xi, \xi')$, the measure of the set of values of $x \in (\xi, \xi')$ such that $\varphi(x)$ falls into one of the above forbidden intervals is, by the mean value theorem $O\left(\frac{\xi}{\log^3 \xi}\right)$, and the proof of Proposition 1 is completed. □

3. About the Differences Between Primes

Proposition 2. *Suppose that there exists $\alpha, 0 < \alpha < 1$, and x large enough such that the inequalities*

$$\pi(x + x^\alpha) - \pi(x) \geq (1 - \varepsilon)x^\alpha/\log x \tag{12}$$

$$\pi(x) - \pi(x - x^\alpha) \geq (1 - \varepsilon)x^\alpha/\log x \tag{13}$$

hold. Then the set

$$E = E(x, \alpha) = \{P - Q; P, Q \textit{ primes},\ x - x^\alpha < Q \leq x < P \leq x + x^\alpha\}$$

satisfies:

$$|E| \geq C_2 x^\alpha$$

where $C_2 = C_1\alpha^4(1 - \varepsilon)^4$ and C_1 is an absolute constant ($C_1 = 0.00164$ works).

Proof. The proof is a classical application of the sieve method that Paul Erdős enjoys very much. Let us set, for $d \leq 2x^\alpha$,

$$r(d) = |\{(P, Q); x - x^\alpha < Q \leq x < P \leq x + x^\alpha, P - Q = d\}|.$$

Clearly we have

$$|E| = \sum_{\substack{0<d\leq 2x^\alpha \\ r(d)\neq 0}} 1 \tag{14}$$

and

$$\sum_{0<d\leq 2x^\alpha} r(d) = (\pi(x + x^\alpha) - \pi(x))(\pi(x) - \pi(x - x^\alpha)) \geq (1 - \varepsilon)^2 x^{2\alpha}/\log^2 x. \tag{15}$$

Now to get an upper bound for $r(d)$, we sift the set

$$A = \{n; x - x^\alpha < n \leq x\}$$

with the primes $p \leq z$. If p divides d, we cross out the n's satisfying $n \equiv 0 \pmod p$, and if p does not divide d, the n's satisfying

$$n \equiv 0 \pmod p \quad \text{or} \quad n \equiv -d \pmod p$$

so that we set for $p \le z$:

$$w(p) = \begin{cases} 1 & \text{if } p \text{ divides } d \\ 2 & \text{if } p \text{ does not divide } d. \end{cases}$$

By applying the large sieve (cf. [14, Corollary 1]), we have

$$r(d) \le \frac{|A|}{L(z)}$$

with

$$L(z) = \sum_{n \le z} \left(1 + \frac{3}{2} n |A|^{-1} z\right)^{-1} \mu(n)^2 \left(\prod_{p|n} \frac{w(p)}{p - w(p)}\right)$$

(μ is the Möbius function), and with the choice $z = (\frac{2}{3}|A|)^{1/2}$, it is proved in [23] that

$$\frac{|A|}{L(z)} \le 16 \prod_{p \ge 3} \left(1 - \frac{1}{(p-1)^2}\right) \frac{|A|}{\log^2(|A|)} \prod_{\substack{p|d \\ p>2}} \frac{p-1}{p-2}.$$

The value of the above infinite product is $0.6602\ldots < 2/3$. We set $f(d) = \prod_{\substack{p|d \\ p>2}} \frac{p-1}{p-2}$, and we observe that $|A| \ge x^\alpha - 1$, so that for x large enough

$$r(d) \le \frac{32}{3\alpha^2} \frac{|A|}{\log^2 x} f(d). \tag{16}$$

Now, for the next step, we shall need an upper bound for $\sum_{n \le x} f^2(n)$. By using the convolution method and defining

$$h(n) = \sum_{a|n} \mu(a) f^2(n/a)$$

one gets $h(2) = h(2^2) = h(2^3) = \ldots = 0$ and, for $p \ge 3$, $h(p) = \frac{2p-3}{(p-2)^2}$, $h(p^2) = h(p^3) = \ldots = 0$, so that

$$\begin{aligned} \sum\nolimits_{n \le x} f^2(n) &= \sum\nolimits_{n \le x} \sum\nolimits_{a|n} h(a) = \sum\nolimits_{a \le x} h(a) \left\lfloor \frac{x}{a} \right\rfloor \\ &\le x \sum\nolimits_{a=1}^{\infty} \frac{h(a)}{a} = x \prod\nolimits_{p \ge 3} \left(1 + \frac{2p-3}{p(p-2)^2}\right) \qquad (17) \\ &= 2.63985\ldots x \le \frac{8}{3} x. \end{aligned}$$

From (15) and (16), one can deduce

$$\frac{(1-\varepsilon)^2 x^{2\alpha}}{\log^2 x} \le \sum_{\substack{0<d\le 2x^\alpha \\ r(d)\ne 0}} r(d) \le \frac{32}{3\alpha^2} \frac{|A|}{\log^2 x} \sum_{\substack{0<d\le 2x^\alpha \\ r(d)\ne 0}} f(d)$$

which implies

$$\sum_{\substack{0<d\le 2x^\alpha\\ r(d)\neq 0}} f(d) \ge \frac{3\alpha^2 x^{2\alpha}(1-\varepsilon)^2}{32|A|}.$$

By Cauchy-Schwarz's inequality, one has

$$\Bigg(\sum_{\substack{0<d\le 2x^\alpha\\ r(d)\neq 0}} 1\Bigg)\Bigg(\sum_{\substack{0<d\le 2x^\alpha\\ r(d)\neq 0}} f^2(d)\Bigg) \ge \frac{9\alpha^4 x^{4\alpha}(1-\varepsilon)^4}{1{,}024|A|^2}$$

and, by (14) and (17)

$$|E| \ge \frac{9\alpha^4 x^{4\alpha}(1-\varepsilon)^4}{1{,}024|A|^2} \Bigg/ \frac{8}{3}(2x^\alpha) = \frac{27}{16{,}384}\frac{x^{3\alpha}(1-\varepsilon)^4}{|A|^2}.$$

Since $|A| \le x^\alpha + 1$, and x has been supposed large enough, Proposition 2 is proved. □

4. Some Properties of $g(n)$

Here, we recall some known properties of $g(n)$ which can be found for instance in [16]. Let us define the arithmetic function ℓ in the following way: ℓ is additive, and, if p is a prime and $k \ge 1$, then $\ell(p^k) = p^k$. It is not difficult to deduce from (1) (cf. [13] or [16]) that

$$g(n) = \max_{\ell(M)\le n} M. \tag{18}$$

Now the relation (cf. [16, p. 139])

$$M \in g(\mathbb{N}) \quad \Longleftrightarrow \quad (M' > M \implies \ell(M') > \ell(M)) \tag{19}$$

easily follows from (18), and shows that the values of the Landau function g are the "champions" for the small values of ℓ. So the methods introduced by Ramanujan (cf. [19]) to study highly composite numbers can also be used for $g(n)$. Indeed M is highly composite, if it is a "champion" for the divisor function d, that is to say if

$$M' < M \quad \implies \quad d(M') < d(M).$$

Corresponding to the so-called superior highly composite numbers, one introduces the set $G : N \in G$ if there exists $\rho > 0$ such that

$$\forall M \ge 1, \quad \ell(M) - \rho \log M \ge \ell(N) - \rho \log N. \tag{20}$$

Equations (19) and (20) easily imply that $G \subset g(\mathbb{N})$. Moreover, if $\rho > 2/\log 2$, let us define $x > 4$ such that $\rho = x/\log x$ and

$$N_\rho = \prod_{p\le x} p^{\alpha_p} = \prod_{p} p^{\alpha_p} \tag{21}$$

with

$$\alpha_p = \begin{cases} 0 & \text{if} \quad p > x \\ 1 & \text{if} \quad \frac{p}{\log p} \le \rho < \frac{p^2-p}{\log p} \\ k \ge 2 & \text{if} \quad \frac{p^k - p^{k-1}}{\log p} \le \rho < \frac{p^{k-1}-p^k}{\log p} \end{cases}$$

then $N_\rho \in G$. With the above definition, since $x \ge 4$, it is not difficult to show that (cf. [11, (5)])

$$p^{\alpha_p} \le x \tag{22}$$

holds for $p \le x$, whence N_ρ is a divisor of the least common multiple of the integers $\le x$. Here we can prove

Proposition 3. *For every prime p, there exists n such that the largest prime factor of $g(n)$ is equal to p.*

Proof. We have $g(2) = 2, g(3) = 3$. If $p \ge 5$, let us choose $\rho = p/\log p > 2/\log 2$. N_ρ defined by (21) belongs to $G \subset g(\mathbb{N})$, and its largest prime factor is p, which proves Proposition 3. □

From Proposition 3, it is easy to deduce a proof of Theorem 1, under the twin prime conjecture. Let $P = p + 2$ be twin primes, and n such that the largest prime factor of $g(n)$ is p. The sequence n_k being defined by (3), we define k in terms of n by $n_k \le n < n_{k+1}$, so that $g(n_k) = g(n)$ has its largest prime factor equal to p. Now, from (18) and (19),

$$\ell(g(n_k)) = n_k$$

and $g(n_k + 2) > g(n_k)$ since $M = \frac{P}{p} g(n_k)$ satisfies $M > g(n_k)$ and $\ell(M) = n_k + 2$. So $n_{k+1} \le n_k + 2$, and Theorem 1 is proved under this strong hypothesis.

Let us introduce now the so-called benefit method. For a fixed $\rho > 2/\log 2$, $N = N_\rho$ is defined by (21), and for any integer M,

$$M = \prod_p p^{\beta_p},$$

one defines the benefit of M:

$$\operatorname{ben}(M) = \ell(M) - \ell(N) - \rho \log M/N. \tag{23}$$

Clearly, from (20), $\operatorname{ben}(M) \ge 0$ holds, and from the additivity of ℓ one has

$$\operatorname{ben}(M) = \sum_p \left(\ell(p^{\beta_p}) - \ell(p^{\alpha_p}) - \rho(\beta_p - \alpha_p) \log p \right). \tag{24}$$

In the above formula, let us observe that $\ell(p^\beta) = p^\beta$ if $\beta \ge 1$, but that $\ell(p^\beta) = 0 \ne p^\beta = 1$ if $\beta = 0$, and, due to the choice of α_p in (21), that, in the sum (24), all the terms are non negative: for all p and for $\beta \ge 0$, we have

$$\ell(p^\beta) - \ell(p^{\alpha_p}) - \rho(\beta - \alpha_p) \log p \ge 0. \tag{25}$$

Indeed, let us consider the set of points $(0,0)$ and $(\beta, p^\beta \log p)$ for β integer ≥ 1. For all p, the piecewise linear curve going through these points is convex, and for a given ρ, α_p is chosen so that the straight line L of slope ρ going through $\left(\alpha_p, \frac{p^{\alpha_p}}{\log p}\right)$ does not cut that curve. The left-hand side of (25), (which is $\operatorname{ben}(Np^{\beta-\alpha_p})$) can be seen as the product of $\log p$ by the vertical distance of the point $\left(\beta, \frac{p^\beta}{\log p}\right)$ to the straight line L, and because of convexity, we shall have for all p,

$$\operatorname{ben}(Np^t) \geq t \operatorname{ben}(Np), \quad t \geq 1 \tag{26}$$

and for $p \leq x$,

$$\operatorname{ben}(Np^{-t}) \geq t \operatorname{ben}(Np^{-1}), \quad 1 \leq t \leq \alpha_p. \tag{27}$$

5. Proof of Theorem 1

First the following proposition will be proved:

Proposition 4. *Let $\alpha < 1/2$, and x large enough such that (10) holds. Let us denote the primes surrounding x by:*

$$\ldots < Q_j < \ldots < Q_2 < Q_1 \leq x < P_1 < P_2 < \ldots < P_i < \ldots .$$

Let us define $\rho = x/\log x, N = N_\rho$ by (21), $n = \ell(N)$. Then for $n \leq m \leq n + 2x^\alpha, g(m)$ can be written

$$g(m) = N \frac{P_{i_1} P_{i_2} \ldots P_{i_r}}{Q_{j_1} Q_{j_2} \ldots Q_{j_r}} \tag{28}$$

with $r \geq 0$ and $i_1 < \ldots < i_r, j_1 < \ldots < j_r, P_{i_r} \leq x + 4x^\alpha, Q_{j_r} \geq x - 4x^\alpha$.

Proof. First, from (18), one has $\ell(g(m)) \leq m$, and from (23) and (18)

$$\operatorname{ben}(g(m)) = \ell(g(m)) - \ell(N) - \rho \log \frac{g(m)}{N} \leq m - n \leq 2x^\alpha \tag{29}$$

for $n \leq m \leq 2x^\alpha$.

Further, let $Q \leq x$ be a prime, and $k = \alpha_Q \geq 1$ the exponent of Q in the standard factorization of N. Let us suppose that for a fixed m, Q divides $g(m)$ with the exponent $\beta_Q = k + t$, $t > 0$. Then, from (24), (25), and (26), one gets

$$\operatorname{ben}(g(m)) \geq \operatorname{ben}(NQ^t) \geq \operatorname{ben}(NQ) \tag{30}$$

and

$$\begin{aligned} \operatorname{ben}(NQ) &= Q^{k+1} - Q^k - \rho \log Q \\ &= \log Q \left(\frac{Q^{k+1} - Q^k}{\log Q} - \rho \right). \end{aligned}$$

From (21), the above parenthesis is nonnegative, and from (10), one gets:

$$\operatorname{ben}(NQ) \geq \log 2 \frac{\sqrt{x}}{\log^4 x}. \tag{31}$$

For x large enough, there is a contradiction between (29), (30) and (31), and so, $\beta_Q \leq \alpha_Q$.

Similarly, let us suppose $Q \leq x$, $k = \alpha_Q \geq 2$ and $\beta_Q = k - t$, $1 \leq t \leq k$. One has, from (24), (25) and (27),

$$\operatorname{ben}(g(m)) \geq \operatorname{ben}(NQ^{-t}) \geq \operatorname{ben}(NQ^{-1})$$

and

$$\begin{aligned} \operatorname{ben}(NQ^{-1}) &= Q^{k-1} - Q^k + \rho \log Q \\ &= \log Q \left(\rho - \frac{Q^k - Q^{k-1}}{\log Q} \right) \geq \log 2 \frac{\sqrt{x}}{\log^4 x} \end{aligned}$$

which contradicts (29), and so, for such a Q, $\beta_Q = \alpha_Q$.

Now, let us suppose $Q \leq x, \alpha_Q = 1$, and $\beta_Q = 0$ for some $m, n \leq m \leq n + 2x^\alpha$. Then

$$\operatorname{ben}(g(m)) \geq \operatorname{ben}(NQ^{-1}) = -Q + \rho \log Q = y(Q)$$

by setting $y(t) = \rho \log t - t$. From the concavity of $y(t)$ for $t > 0$, for $x \geq e^2$, we get

$$\begin{aligned} y(Q) \geq y(x) + (Q - x) y'(x) &= (Q - x) \left(\frac{\rho}{x} - 1 \right) \\ &= (x - Q) \left(1 - \frac{1}{\log x} \right) \geq \frac{1}{2}(x - Q) \end{aligned}$$

and so,

$$\operatorname{ben}(g(m)) \geq \frac{1}{2}(x - Q)$$

which, from (29) yields

$$x - Q \leq 4x^\alpha.$$

In conclusion, the only prime factors allowed in the denominator of $\frac{g(m)}{N}$ are the Q's, with $x - 4x^\alpha \leq Q \leq x$, and $\alpha_Q = 1$.

What about the numerator? Let $P > x$ be a prime number and suppose that P^t divides $g(m)$ with $t \geq 2$. Then, from (26) and (23),

$$\operatorname{ben}(Np^t) \geq \operatorname{ben}(Np^2) = P^2 - 2\rho \log P.$$

But the function $t \mapsto t^2 - 2\rho \log t$ is increasing for $t \geq \sqrt{\rho}$, so that,

$$\operatorname{ben}(NP^t) \geq x^2 - 2x > 2x^\alpha$$

for x large enough, which contradicts (29). The only possibility is that P divides $g(m)$ with exponent 1. In that case, from the convexity of the function $z(t) = t - \rho \log t$, inequality (26) yields

$$\begin{aligned}\operatorname{ben}(g(m)) \geq \operatorname{ben}(NP) = z(P) \geq z(x) + (P-x)z'(x)\\ = (P-x)\left(1 - \frac{1}{\log x}\right) \geq \frac{1}{2}(P-x)\end{aligned}$$

for $x \geq e^2$, which, with (29), implies

$$P - x \leq 4x^{\alpha}.$$

Up to now, we have shown that

$$g(m) = N\frac{P_{i_1}\dots P_{i_r}}{Q_{j_1}\dots Q_{j_s}}$$

with $P_{i_r} \leq x + 4x^{\alpha}, Q_{j_s} \geq x - 4x^{\alpha}$. It remains to show that $r = s$. First, since $n \leq m \leq n + 2x^{\alpha}$, and N belongs to G, we have from (18) and (19)

$$n \leq \ell(g(m)) \leq n + 2x^{\alpha}. \tag{32}$$

Further,

$$\ell(g(m)) - n = \sum_{t=1}^{r} P_{i_t} - \sum_{t=1}^{s} Q_{j_t}$$

and since $r \leq 4x^{\alpha}$, and $s \leq 4x^{\alpha}$,

$$\begin{aligned}\ell(g(m)) - n &\leq r(x + 4x^{\alpha}) - s(x - 4x^{\alpha})\\ &\leq (r-s)x + 32x^{2\alpha}.\end{aligned}$$

From (32), $\ell(g(m)) - n \geq 0$ holds and as $\alpha < 1/2$, this implies that $r \geq s$ for x large enough. Similarly,

$$\ell(g(m)) - n \geq (r-s)x,$$

so, from (32), $(r-s)x$ must be $\leq 2x^{\alpha}$, which, for x large enough, implies $r \leq s$; finally $r = s$, and the proof of Proposition 4 is completed. □

Lemma 1. *Let x be a positive real number, $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_k$ be real numbers such that*

$$b_k \leq b_{k-1} \leq \dots \leq b_1 \leq x < a_1 \leq a_2 \leq \dots \leq a_k$$

and Δ be defined by $\Delta = \sum_{i=1}^{k}(a_i - b_i)$. Then the following inequalities

$$\frac{x+\Delta}{x} \leq \prod_{i=1}^{k} \frac{a_i}{b_i} \leq \exp\left(\frac{\Delta}{x}\right)$$

hold.

Proof. It is easy, and can be found in [16, p. 159]. □

Now it is time to prove Theorem 1. With the notation and hypothesis of Proposition 4, let us denote by B the set of integers M of the form

$$M = N\frac{P_{i_1}P_{i_2}\dots P_{i_r}}{Q_{j_1}Q_{j_2}\dots Q_{j_r}}$$

satisfying

$$\ell(M) - \ell(N) = \sum_{t=1}^{r}(P_{i_t} - Q_{j_t}) \leq 2x^\alpha.$$

From Proposition 4, for $n \leq m \leq 2x^\alpha, g(m) \in B$, and thus, from (18),

$$g(m) = \max_{\substack{\ell(M)\leq m\\ M\in B}} M. \tag{33}$$

Further, for $0 \leq d \leq 2x^\alpha$, define

$$B_d = \{M \in B : \ell(M) - \ell(N) = d\}.$$

I claim that, if $d < d'$ (which implies $d \leq d' - 2$), any element of B_d is smaller than any element of $B_{d'}$. Indeed, let $M \in B_d$, and $M' \in B_{d'}$. From Lemma 1, one has

$$\frac{M}{N} \leq \exp\left(\frac{d}{x}\right) \quad \text{and} \quad \frac{M'}{N} \geq \frac{x+d'}{x} \geq \frac{x+d+2}{x}.$$

Since $d < 2x^\alpha < x$, and $e^t \leq \frac{1}{1-t}$ for $0 \leq t < 1$, one gets

$$\frac{M}{N} \leq \frac{1}{1-d/x} = \frac{x}{x-d}.$$

This last quantity is smaller than $\frac{x+d+2}{x}$ if $(d+1)^2 < 2x+1$, which is true for x large enough, because $d \leq 2x^\alpha$ and $\alpha < 1/2$.

From the preceding claim, and from (33), it follows that, if B_d is non empty, then

$$g(n+d) = \max B_d.$$

Further, since $N \in G$, we know that $n = \ell(N)$ belongs to the sequence (n_k) where g is increasing, and so, $n = n_{k_0}$. If $0 < d_1 < d_2 < \dots < d_s \leq 2x^\alpha$ denote the values of d for which B_d is non empty, then one has

$$n_{k_{0+i}} = n + d_i, 1 \leq i \leq s. \tag{34}$$

Suppose now that $\alpha < 1/2$ and x have been chosen in such a way that (12) and (13) hold. With the notation of Proposition 2, the set $E(x, \alpha)$ is certainly included in the set $\{d_1, d_2, \dots, d_s\}$, and from Proposition 2,

$$s \geq C_2 x^\alpha \tag{35}$$

which implies that for at least one i, $d_{i+1} - d_i \leq \frac{2}{C_2}$, and thus

$$n_{k_0+i+1} - n_{k_0+i} \leq \frac{2}{C_2}.$$

Finally, for $\frac{1}{6} < \alpha < \frac{1}{2}$, Proposition 1 allows us to choose x as wished, and thus, the proof of Theorem 1 is completed. □

With ε very small, and α close to $1/2$, the values of C_1 and C_2 given in Proposition 2 yield that for infinitely many $k's$,

$$n_{k+1} - n_k \leq 20{,}000.$$

To count how many such differences we get, we define

$$\gamma(n) = \mathrm{Card}\{m \leq n : g(m) > g(m-1)\}.$$

Therefore, with the notation (3), we have $n_{\gamma(n)} = n$.

In [16, 162–164], it is proved that

$$n^{1-\tau/2} \ll \gamma(n) \leq n - c\frac{n^{3/4}}{\sqrt{\log n}}$$

where τ is such that the sequence of consecutive primes satisfies $p_{i+1} - p_i \ll p_i^{\tau}$. Without any hypothesis, the best known τ is $> 1/2$.

Proposition 5. *We have $\gamma(n) \geq n^{3/4-\varepsilon}$ for all $\varepsilon > 0$, and n large enough.*

Proof. With the definition of $\gamma(n)$, (34) and (35) give

$$\gamma(n + 2x^{\alpha}) - \gamma(n) \geq s \gg x^{\alpha} \tag{36}$$

whenever $n = \ell(N), N = N_{\rho}, \rho = x/\log x$, and x satisfies Proposition 1. But, from (21), two close enough distinct values of x can yield the same N.

I now claim that, with the notation of Proposition 1, the number of primes p_i between ξ and ξ' such that there is at least one $x \in [p_i, p_{i+1})$ satisfying (8), (9) and (10) is bigger than $\frac{1}{2}(\pi(\xi') - \pi(\xi))$. Indeed, for each i for which $[p_i, p_{i+1})$ does not contain any such x, we get a measure $p_{i+1} - p_i \geq 2$, and if there are more than $\frac{1}{2}(\pi(\xi') - \pi(\xi))$ such $i's$, the total measure will be greater than $\pi(\xi') - \pi(\xi) \sim \xi/\log^2 \xi$, which contradicts Proposition 1.

From the above claim, there will be at least $\frac{1}{2}(\pi(\xi') - \pi(\xi))$ distinct N's, with $N = N_{\rho}, \rho = x/\log x$, and $\xi \leq x \leq \xi'$. Moreover, for two such distinct N, say $N' < N''$, we have from (21), $\ell(N'') - \ell(N') \geq \xi$.

Let $N^{(1)}$ and $N^{(0)}$ the biggest and the smallest of these N's, and $n^{(1)} = \ell(N^{(1)}), n^{(0)} = \ell(N^{(0)})$, then from (36),

$$\gamma(n^{(1)}) \geq \gamma(n^{(1)}) - \gamma(n^{(0)}) \geq \frac{1}{2}\left(\pi(\xi') - \pi(\xi)\right)\xi^{\alpha} \gg \frac{\xi^{1+\alpha}}{\log^2 \xi}. \tag{37}$$

But from (21) and (22), $x \sim \log N_{\rho}$, and from (2),

$$x \sim \log N_{\rho} \sim \sqrt{n \log n} \quad \text{with} \quad n = \ell(N_p)$$

so

$$\xi \sim \sqrt{n^{(1)} \log n^{(1)}}$$

and since α can be choosen in (37) as close as wished of 1/2, this completes the proof of Proposition 5. □

References

1. L. Alaoglu, P. Erdős, "On highly composite and similar numbers", Trans. Amer. Math. Soc. 56, 1944, 448–469.
2. M. Deléglise, J.-L. Nicolas, P. Zimmermann, "Landau's function for one million billions", J. de Théorie des Nombres de Bordeaux, 20, 2008, 625–671.
3. M. Deléglise, J.-L. Nicolas, "Le plus grand facteur premier de la fonction de Landau", Ramanujan J., 27, 2012, 109–145.
4. P. Erdős, "On highly composited numbers", J. London Math. Soc., 19, 1944, 130–133.
5. P. Erdős, "Ramanujan and I", Number Theory, Madras 1987, Editor : K. Alladi, Lecture Notes in Mathematics n^o 1395, Springer-Verlag, 1989.
6. P. Erdős, J.-L. Nicolas, "Répartition des nombres superabondants", Bull. Soc. Math. France, 103, 1975, 65–90.
7. P. Erdős, P. Turan, "On some problems of a statistical group theory", I to VII, Zeitschr. fur Wahrschenlichkeitstheorie und verw. Gebiete, 4, 1965, 175–186; Acta Math. Hung., 18, 1967, 151–163; Acta Math. Hung., 18, 1967, 309–320; Acta Math. Hung., 19, 1968, 413–435; Periodica Math. Hung., 1, 1971, 5–13; J. Indian Math. Soc., 34, 1970, 175–192; Periodica Math. Hung., 2, 1972, 149–163.
8. J. Grantham, "The largest prime dividing the maximal order of an element of S_n", Math. Comp., 64, 1995, 407–410.
9. E. Landau, "Uber die Maximalordung der Permutation gegebenen Grades", Handbuch der Lehre von der Verteilung der Primzahlen, vol. 1, 2nd edition, Chelsea, New-York, 1953, 222–229.
10. J. P. Massias, "Majoration explicite de l'ordre maximum d'un élément du groupe symétrique", Ann. Fac. Sci. Toulouse Math., 6, 1984, 269–281.
11. J. P. Massias, J.-L. Nicolas, G. Robin, "Evaluation asymptotique de l'ordre maximum d'un élément du groupe symétrique", Acta Arithmetica, 50, 1988, 221–242.
12. J. P. Massias, J.-L. Nicolas, G. Robin, "Effective bounds for the Maximal Order of an Element in the Symmetric Group", Math. Comp., 53, 1989, 665–678.
13. W. Miller, "The Maximum Order of an Element of a Finite Symmetric Group", Amer. Math. Monthly, 94, 1987, 497–506.
14. H. L. Montgomery, R. C. Vaughan, "The large sieve", Mathematika, 20, 1973, 119–134.
15. J.-L. Nicolas, "Sur l'ordre maximum d'un élément dans le groupe S_n des permutations", Acta Arithmetica, 14, 1968, 315–332.
16. J.-L. Nicolas, "Ordre maximum d'un élément du groupe de permutations et highly composite numbers", Bull. Soc. Math. France, 97, 1969, 129–191.
17. J.-L. Nicolas, "Ordre maximal d'un élément d'un groupe de permutations", C.R. Acad. Sci. Paris, 270, 1970, 1473–1476.
18. J.-L. Nicolas, "Répartition des nombres largement composés", Acta Arithmetica, 34, 1979, 379–390.

19. S. Ramanujan, “Highly composite numbers”, Proc. London Math. Soc., Series 2, 14, 1915, 347–400; and “Collected papers”, Cambridge at the University Press, 1927, 78–128.
20. S. Ramanujan, “Highly composite numbers, annotated and with a foreword by J.-L. Nicolas and G. Robin”, Ramanujan J., 1, 1997, 119–153.
21. A. Selberg, “On the normal density of primes in small intervals and the difference between consecutive primes”, Arch. Math. Naturvid, 47, 1943, 87–105.
22. S. Shah, “An Inequality for the Arithmetical Function $g(x)$”, J. Indian Math. Soc., 3, 1939, 316–318.
23. H. Siebert, “Montgomery’s weighted sieve for dimension two”, Monatsch., Math., 82, 1976, 327–336.
24. N. J. A. Sloane, “The On-Line Encyclopedia of Integer Sequences”, http://oeis.org. Accessed 12 December 2012.
25. M. Szalay, “On the maximal order in S_n and S_n^*”, Acta Arithmetica, 37, 1980, 321–331.
26. http://math.univ-lyon1.fr/~nicolas/landaug.html.

On Divisibility Properties of Sequences of Integers

András Sárközy

A. Sárközy (✉)
Department of Algebra and Number Theory, Eötvös Loránd University,
Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary
e-mail: sarkozy@cs.elte.hu

Dedicated to Paul Erdős on the occasion of his 80th birthday

1. Introduction

Our first joint paper with Erdős appeared in 1966. It was a triple paper with Szemerédi written on divisibility properties of sequences of integers which is one of Erdős' favorite subjects. Nine further triple papers written on the same subject followed it, and since 1966, we have written altogether 52 joint papers with Erdős. On this special occasion I would like to return to the subject of our very first paper. In Sect. 2, I will give a survey of the related results, while in Sect. 3, I will study a further related problem.

2. Survey of the Results on Divisibility Properties of Sequences of Integers

Throughout this paper, the following notations will be used: $\mathbb{N}$ denotes the set of the positive integers. $c_1, c_2, \ldots$ denote positive absolute constants. If $f(x) = 0(g(x))$, then we write $f(x) \ll g(x)$. If $\mathcal{A} \subset \mathbb{N}$, $n \in \mathbb{N}$, then we write

$$s(\mathcal{A}, n) = \sum_{a \in \mathcal{A}, a \leq n} \frac{1}{a},$$

$$t(\mathcal{A}, n) = \sum_{a \in \mathcal{A}, 2 \leq a \leq n} \frac{1}{a \log a},$$

$$s(\mathcal{A}) = \sum_{a \in \mathcal{A}} \frac{1}{a}$$

and

$$t(\mathcal{A}) = \sum_{a \in \mathcal{A}, 2 \leq a} \frac{1}{a \log a}.$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_15,

For $n \geq 3$ we put

$$\ell(\mathcal{A}, n) = s(\mathcal{A}, n)(\log n)^{-1}$$

and

$$m(\mathcal{A}, n) = t(\mathcal{A}, n)(\log \log n)^{-1}$$

These quantities can be called as "logarithmic density" and "log log density" of $\mathcal{A}$. In fact, both of them can be considered as densities since if the infinite sequence $\mathcal{A} \subset \mathbb{N}$ possesses asymptotic density, then the limit of both $\ell(\mathcal{A}, n)$ and $m(\mathcal{A}, n)$ is equal to the asymptotic density of $\mathcal{A}$.

If $\mathcal{A} \subset \mathbb{N}$ and there are no a, a' with $a \in \mathcal{A}$, $a' \in \mathcal{A}$, $a \neq a'$ and $a|a'$, then $\mathcal{A}$ is said to be primitive.

In 1935, Behrend [6] proved the following result:

Theorem 1. *If $n \in \mathbb{N}, n > n_0, \mathcal{A} \subset \{1, 2, \ldots, n\}$ and $\mathcal{A}$ is primitive, then we have*

$$\ell(\mathcal{A}, n) < c_1(\log \log n)^{-1/2}.$$

The proof of this nice theorem is of a combinatorial nature; the crucial tool in the proof is Sperner's theorem on subsets of finite sets. So that it is an Erdős-type result, except that it is due to Behrend and not to Erdős; however, it is just a matter of a few months that now this result is known as Behrend's theorem and not as Erdős' theorem. The story is the following:

Due to the steadily worsening political atmosphere in Hungary in the early thirties, Erdős was forced to leave the country in 1934. He took a train for Cambridge, and during the travel he felt very depressed since he had to leave all his relatives and friends for the unknown. Thus to cheer himself up, he started to do some mathematics. After a while, he ended up with the result formulated in Theorem 1. Arriving to Cambridge, Davenport and Radó were waiting for him at the station. Erdős told them his new result immediately. They said that, indeed, it was a nice result but unfortunately, Behrend had proved it a few months earlier. Erdős [9] consoled himself soon by proving a series of important new results in combinatorial number theory. Moreover, he had another result on primitive sets not anticipated by anyone. Indeed, he proved that for the "log log density" of a primitive set better estimate can be given, than for the "logarithmic density":

Theorem 2. *If $n \in \mathbb{N}$, $n > n_0$, $\mathcal{A} \subset \{1, 2, \ldots, n\}$ and $\mathcal{A}$ is primitive, then we have*

$$t(\mathcal{A}, n) < c_2$$

(so that $m(\mathcal{A}, n) < c_2(\log \log n)^{-1}$).

This theorem might suggest that Theorem 1 can be improved. This is not so as Pillai [28] pointed out:

Theorem 3. *If $n \in \mathbb{N}$, $n > n_0$, then there is a primitive set $\mathcal{A}$ such that $\mathcal{A} \subset \{1, 2, \ldots, n\}$ and*

$$\ell(\mathcal{A}, n) > c_3(\log\log n)^{-1/2}.$$

If $\mathcal{A} \subset \mathbb{N}$ is an infinite sequence of positive upper logarithmic density, then it cannot be primitive by Theorem 1. Davenport and Erdős [8] proved that having this assumption, more can be said:

Theorem 4. *If $\mathcal{A} \subset \mathbb{N}$ is an infinite set such that*

$$\lim_{n\to+\infty} \sup \ell(\mathcal{A}, n) > 0,$$

then $\mathcal{A}$ contains an infinite subsequence $a_1, a_2, \ldots$ such that $a_i | a_{i+1}$ for $i = 1, 2, \ldots$.

For $n \geq 3$, write

$$G(n) = \max \ell(\mathcal{A}, n)(\log\log n)^{1/2}$$

where the maximum is taken over all primitive sets $\mathcal{A} \subset \{1, 2, \ldots, n\}$. Then by Theorems 1 and 3, for $n > n_0$ we have

$$c_4 < G(N) < c_5.$$

In 1948, Erdős [10] proved a result which implies that for $n > n_0(\epsilon)$ we have

$$G(n) > (2\pi)^{-1/2} - \epsilon.$$

On the other hand, Anderson [2] showed that for $n > n_0(\epsilon)$ we have

$$G(n) < \pi^{-1/2} + \epsilon.$$

In 1967, Erdős, Szemerédi and I [20] determined the best possible value of the constants in Theorems 1 and 3 by proving that we have

$$G(n) = (1 + o(1))(2\pi)^{-1/2}.$$

Moreover, in [21] we showed that for infinite sets, Theorem 1 can be improved:

Theorem 5. *If $\mathcal{A} \subset \mathbb{N}$ is an infinite primitive set, then for $n \to +\infty$ we have*

$$\ell(\mathcal{A}, n) = o((\log\log n)^{-1/2}).$$

Alexander [1] and Erdős, Szemerédi and I [17] sharpened Theorem 4 in various directions. In [18] and [19], Erdős, Szemerédi and I studied the number of divisibility relations up to a certain bound, i.e., the function

$$f(\mathcal{A}, n) = |\{(a, a') : a, a' \in \mathcal{A}, a|a', a' \leq n\}|.$$

In [22], we gave a survey of all these results and our other related works. This survey paper contained many unsolved problems. The two most interesting problems are, perhaps, the following ones:

Problem 1. *Is it true that if $a_1, a_2, \dots$ are real numbers with $1 < a_1 < a_2 < \cdots$, and for $i, j, k \in \mathbb{N}, i \neq j$ we have*

$$|a_i - ka_j| \geq 1,$$

then

$$\left(\sum_{a_i \leq n} \frac{1}{a_i}\right)(\log n)^{-1} < c_6(\log\log n)^{-1/2}$$

and

$$\sum_{a_i \leq n} \frac{1}{a_i \log a_i} < c_7?$$

(This would generalize Theorems 1 and 2, respectively.)

Problem 2. *Is it true that for all $\epsilon > 0$ there is a $K = K(\epsilon)$ such that if $\mathcal{A} \subset \mathbb{N}$ and $\mathcal{A}$ is primitive, then we have*

$$\sum_{a \in \mathcal{A}, K \leq a} \frac{1}{a \log a} < 1 + \epsilon?$$

Haight [24], resp. Erdős and Zhang Zhenxiang [23] have certain partial results, but in their original form both problems are unsolved yet.

Pomerance and Sárközy [29] studied the following problem: if $l(\mathcal{A}, n)$ is large enough to ensure the existence of pairs a, a' with $a, a' \in \mathcal{A}$, $a \neq a'$, $a|a'$, then what can be said about the arithmetic properties of the quotients a'/a (with $a|a'$)? If $\mathcal{P}$ is a finite set of prime numbers, $\mathcal{A} \subset \mathbb{N}$ and there are no a, a' with $a \in \mathcal{A}$, $a' \in \mathcal{A}$, $a \neq a'$, $a|a'$ and $\frac{a'}{a}| \prod_{p \in \mathcal{P}} p$, then $\mathcal{A}$ is said to be $\mathcal{P}$-primitive. In [29] we proved the extension of Theorem 1:

Theorem 6. *If $n \in \mathbb{N}, n > n_0, \mathcal{P}$ is a set of prime numbers not exceeding n with*

$$s(\mathcal{P}) = \sum_{p \in \mathcal{P}} \frac{1}{p} > c_s,$$

$\mathcal{A} \subset \{1, 2, \dots, n\}$ and $\mathcal{A}$ is $\mathcal{P}$-primitive, then we have

$$\ell(\mathcal{A}, n) < c_9(s(\mathcal{P}))^{-1/2}.$$

We showed that, apart from the values of the constants n_0, c_8 and c_9, the theorem is best possible for all n and $\mathcal{P}$, i.e., there exist n_1, c_{10}, c_{11} such that if $n > n_1$ and $\mathcal{P}$ is a set of prime numbers not exceeding n with $s(\mathcal{P}) > c_{10}$, then there is a set $\mathcal{A}$ such that $\mathcal{A} \subset \{1, 2, \dots, n\}$, $\mathcal{A}$ is a $\mathcal{P}$-primitive and

$$\ell(\mathcal{A}, n) > c_{11}(s(\mathcal{P}))^{-1/2}.$$

Moreover, we discussed some consequences of the theorem and some further related problems. The most interesting problem, that we could not settle, is the following one:

Problem 3. *Is it true that if* $k, n \in \mathbb{N}$*,* $n > n_0(k)$*,* $\mathcal{A} \subset \{1, 2, \ldots, n\}$*,* $a \equiv a' \pmod{k}$ *for all* $a, a' \in \mathcal{A}$ *and*

$$\ell(\mathcal{A}, n) > c_{12} k^{-1} (\log \log n)^{-1/2},$$

then there exist integers a, a' *such that* $a, a' \in \mathcal{A}$*,* $a \neq a'$ *and* $a | a'$*?*

For $\mathcal{A} \subset \mathbb{N}$, $n \in \mathbb{N}$ write

$$d(\mathcal{A}, n) = \sum_{a \in \mathcal{A}, a | n} 1,$$

and let

$$D(\mathcal{A}, x) = \max_{n \leq x} d(\mathcal{A}, n).$$

In [13–15] and [16], Erdős and I studied the function $D(\mathcal{A}, x)$.

Further related results have been proved by Anderson [3, 4], Erdős and Sárközy [12], Meijer [26, 27], Sattler [30], Anderson, Cohen and Stothers [5], Cohen [7], Klotz [25] and Erdős [11].

3. A Further Result on $\mathcal{P}$-Primitive Sets

In this section, we will extend Theorem 2 to $\mathcal{P}$-primitive sets (in the same way as Theorem 6 extends Theorem 1 to $\mathcal{P}$-primitive sets).

If $n \in \mathbb{N}, n \geq 3$ and $\mathcal{P}$ is a set of prime numbers not exceeding n, then let

$$L(\mathcal{P}, n) = \max \ell(\mathcal{A}, n)$$

and

$$M(\mathcal{P}, n) = \max m(\mathcal{A}, n)$$

where in both cases the maximum is taken over all $\mathcal{P}$-primitive sets $\mathcal{A} = \{1, 2, \ldots, n\}$. In the special case when $\mathcal{P}$ consists of all the primes not exceeding n, by Theorems 2 and 3 we have

$$M(\mathcal{P}, n) = o(L(\mathcal{P}, n)). \tag{1}$$

We will show that this also holds in the more general case when $\mathcal{P}$ contains all or "almost all" the large primes, and it will be shown that a condition of this type is necessary.

Theorem 7. *There are absolute constants* n_0*,* c_{13} *with the following properties: Assume that* $n \in \mathbb{N}$*,* $n > n_0$ *and* $\mathcal{P}$ *is a set of primes not exceeding* n*. If*

$$\sum_{p \le n, p \notin \mathcal{P}} \frac{1}{p} \ge 100 \tag{2}$$

then let y *denote the smallest positive integer* y *such that*

$$\log\log\log y > \sum_{y < p \le n, p \notin \mathcal{P}} \frac{1}{p}; \tag{3}$$

if the left-hand side of (2) *is* < 100*, then let* $y = 10$*. Assume that* $A \subset \{1, 2, \dots, n\}$ *and* $\mathcal{A}$ *is* $\mathcal{A}$*-primitive. Then we have*

$$t(\mathcal{A}) < c_{13} \log\log y. \tag{4}$$

Note that Theorem 7 is certainly superior to Theorem 6, i.e., (1) holds if

$$\sum_{p \le n, p \notin \mathcal{P}} \frac{1}{2} < \left(\frac{1}{2} - \varepsilon\right) \log\log\log n$$

(namely, in this case y in Theorem 7 satisfies $\log\log y < (\log\log n)^{1/2-\epsilon/2}$). Moreover, Theorem 7 is superior to Theorem 6 in the important special case when $\mathcal{P}$ is of the form

$$\mathcal{P} = \{p : z < p \le n, p \text{ prime}\}. \tag{5}$$

In this special case we obtain

Corollary 1. *There are absolute constants* n_0, c_{14} *such that if* $n > n_0$*,* $1 \le z \le n$*,* $\mathcal{A} \subset \{1, 2, \dots, n\}$ *and*

$$t(\mathcal{A}) > c_{14} \log\log(z + 3) \tag{6}$$

then there are a*,* a' *with* $a \in \mathcal{A}$*,* $a' \in \mathcal{A}$*,* $a \ne a'$*,* $a|a'$ *and* $p(a'/a) > z$*.*

Note that in the $z = 1$ special case we obtain Theorem 2.

Corollary 1 (and thus also Theorem 7) is best possible apart from the value of the constants n_0, c_{14}. Indeed, if we take $\mathcal{A} = \{1, 2, \dots, [z]\}$, then clearly we have

$$t(\mathcal{A}) < c_{15} \log\log(z + 3)$$

and there are no a, a' with $a \in \mathcal{A}$, $a' \in \mathcal{A}$, $a \ne a'$, $a|a'$ and $p(a/a') > z$.

Proof of Theorem 7. Put

$$\mathcal{A}_1 = \{a : a \le y, a \in \mathcal{A}\}, \mathcal{A}_2 = \{a : y < a, a \in \mathcal{A}\}$$

so that

$$t(\mathcal{A}) = t(\mathcal{A}_1) + t(\mathcal{A}_2). \tag{7}$$

Clearly we have

$$t(\mathcal{A}_1) = \sum_{2\le a\le y, a\in\mathcal{A}} \frac{1}{a\log a} \le \sum_{2\le a\le y} \frac{1}{a\log a}$$
$$= (1+o(1))\log\log y < c_{16}\log\log y. \qquad (8)$$

Now we will estimate $t(\mathcal{A}_2)$. Write $P = \prod_{p\in\mathcal{P}} p$, and for $m\in\mathbb{N}$, $m>1$, let $p(m)$ denote the smallest prime factor of m. Consider all the integers of the form

$$ad \text{ where } a\in\mathcal{A}_2,\quad d|P \quad\text{and}\quad p(d)>a \quad\text{or}\quad d=1. \qquad (9)$$

Now we will show that these numbers are distinct. We will prove this by contradiction: assume that

$$ad = a'd', a<a', d|P, d'|P, p(d)>a \text{ or } d=1, \text{ and } p(d')>a' \text{ or } d'=1. \qquad (10)$$

It follows that

$$d'|ad \qquad (11)$$

and either $d'=1$ or $p(d')>a'>a$; in both cases we have

$$(a,d')=1. \qquad (12)$$

It follows from (11) and (12) that $d'd$. Then by (10) we have

$$a\cdot\frac{d}{d'} = a'$$

whence $a|a'$ and $\frac{a'}{a} = \frac{d}{d'}|d$, $d|P$ so that $\frac{a'}{a}|P$ which contradicts the assumption that $\mathcal{A}$ is $\mathcal{P}$-primitive. This proves that, indeed, the numbers of form (9) are distinct. Let $\mathcal{B}$ denote the set of these numbers.

If $b=ad\in\mathcal{B}$, then the prime factors of ad do not exceed n. Thus by Mertens' formula

$$\prod_{p\le x}\left(1-\frac{1}{p}\right) = (1+o(1))c_{17}(\log x)^{-1}, \qquad (13)$$

we have

$$s(\mathcal{B}) = \sum_{ad\in\mathcal{B}}\frac{1}{ad} \le \prod_{p\le n}\left(\sum_{k=0}^{+\infty}\frac{1}{p^k}\right) = \prod_{p\le n}\left(1-\frac{1}{p}\right)^{-1} < c_{18}\log n. \qquad (14)$$

On the other hand, by (13) and the definition of y we have

$$s(\mathcal{B}) = \sum_{ad\in\mathcal{B}}\frac{1}{ad} = \sum_{a\in\mathcal{A}_2}\left(\frac{1}{a}\sum_{\substack{d|\mathcal{P}\\ p(d)>a \text{ or } d=1}}\frac{1}{d}\right) = \sum_{a\in\mathcal{A}_2}\left(\frac{1}{a}\prod_{p\in\mathcal{P},p>a}\left(1+\frac{1}{p}\right)\right)$$

$$= \sum_{a\in\mathcal{A}_2}\left(\frac{1}{a}\prod_{a<p\le n}\left(1+\frac{1}{p}\right)\prod_{a<p\le n, p\notin\mathcal{P}}\left(1+\frac{1}{p}\right)^{-1}\right)$$

$$
\begin{aligned}
&\geq \sum_{a\in\mathcal{A}_2}\left(\frac{1}{a}\prod_{a<p\leq n}\left(1+\frac{1}{p}\right)\prod_{y<p\leq n, p\notin\mathcal{P}}\left(1+\frac{1}{p}\right)^{-1}\right)\\
&> \exp\left(-\sum_{y<p\leq n, p\notin\mathcal{P}}\frac{1}{p}\right)\sum_{a\in\mathcal{A}_2}\left(\frac{1}{a}\prod_{a<p\leq n}\left(1+\frac{1}{p}\right)\right) \qquad (15)\\
&> c_{19}\exp(-\log\log\log y)\sum_{a\in\mathcal{A}_2}\frac{1}{a}\prod_{a<p\leq n}\left(1-\frac{1}{p}\right)^{-1}\\
&> c_{20}(\log\log y)^{-1}\sum_{a\in\mathcal{A}_2, 2\leq a}\frac{1}{a}\frac{\log n}{\log a}\\
&= c_{20}(\log\log y)^{-1}(\log n)t(\mathcal{A}_2) \qquad (16)
\end{aligned}
$$

since

$$1+x<e^x \quad \text{for all} \quad x>0$$

and

$$
\begin{aligned}
\prod_{p\leq x}\left(1+\frac{1}{p}\right) &= \prod_{p\leq x}\left(1-\frac{1}{p^2}\right)\prod_{p\leq x}\left(1-\frac{1}{p}\right)^{-1}\\
&= (1+o(1))c_{21}\prod_{p\leq x}\left(1-\frac{1}{p}\right)^{-1} \qquad \text{for} \quad x\to\infty.
\end{aligned}
$$

It follows from (14) and (16) that

$$t(\mathcal{A}_2) < c_{22}\log\log y. \qquad (17)$$

Finally, (4) follows from (7), (8) and (17), and this completes the proof of Theorem 7. □

Proof of Corollary 1. We apply Theorem 7 with the set $\mathcal{P}$ defined by (5). Then the right-hand side of (3) with z in place of y is 0, thus the number y defined in Theorem 7 satisfies either $y\leq z$ or $y=10$. Thus by Theorem 7, for a $\mathcal{P}$-primitive set we have

$$t(\mathcal{A}) < c_{13}\log\log y \leq c_{13}\log\log\max(z,10) < c_{23}\log\log(z+3).$$

Thus choosing $c_{14}=c_{23}$ in Corollary 1, (6) implies that $\mathcal{A}$ is not $\mathcal{P}$-primitive which completes the proof of Corollary 1. □

Appendix

The paper above appeared in 1997. Since that time, 10 related papers have been published; they will be listed at the end of this Appendix. I will keep the notations and reference numbers (from [1–30]) of the original paper while the reference numbers of the papers published since 1997 will start with [31]. In this appendix my goal is to present a short survey of these recent papers.

In [35] Ahlswede, Khachatrian and Sárközy extended the study of the density of "large" primitive sets from the logarithmic density and "loglog density" to other densities defined in terms of different weight functions. Among others they proved that if the weight function $f : \mathbb{N} \to [0, \infty)$ is "smooth" in a well defined sense, then for $\varepsilon > 0$, $N > N_0(\varepsilon, f)$ there is a primitive set $\mathcal{A} \subset \{1, 2, \dots, N\}$ such that its density with respect to f is greater, than $(1 - \varepsilon)\frac{1}{\log \log N}$, and if a conjecture of Erdős is assumed then this lower bound is essentially best possible.

In [34] Ahlswede, Khachatrian and Sárközy studied "large" primitive sets consisting of squarefree integers. Among others, they proved that considering all the primitive sets $\mathcal{A} \subset \{1, 2, \dots, N\}$ consisting of squarefree integers we have

$$\max_{\mathcal{A}} \sum_{a \in \mathcal{A}} \frac{1}{a} = (1 + o(1)) \frac{6}{\pi^2} \frac{\log N}{(2\pi \log \log N)^{1/2}}$$

which settles a conjecture of Pomerance and Sárközy [29].

In [22] we wrote: "The following problem seems difficult: Let $b_1 < \cdots$ be an infinite sequence of integers. What is the necessary and sufficient condition that there should exist a primitive sequence $a_1 < \cdots$ satisfying $a_n < cb_n$ for every n?" This was one of the favorite problems of Erdős. However, in this original form the problem seems hopelessly difficult. Thus Ahlswede, Khachatrian and Sárközy [32] studied the following milder and more realistic version of the problem: How fast can the counting function $A(x) = |\mathcal{A} \cap [1, x]|$ of an infinite primitive set grow? It follows from Erdős's Theorem 2 above that for a primitive set $\mathcal{A}$ we must have

$$A(x) < \frac{x}{(\log \log x)(\log \log \log x)}$$

for infinitely many $x \in \mathbb{N}$, and we proved in [32] that for $\varepsilon > 0$ there is an infinite primitive set $\mathcal{A}$ with

$$A(x) > \frac{x}{(\log \log x)(\log \log \log x)^{1+\varepsilon}}.$$

Martin and Pomerance [40] improved on this result significantly. Denote the k-fold logarithm of x by $\log_k x$. They proved (in a slightly simplified form): for any integer $k \geq 3$ and every real number $\varepsilon > 0$ there exists a primitive set $\mathcal{A}$ such that

$$A(x) \asymp \frac{x}{\log_2 x \cdots \log_{k-1} x (\log_k x)^{1+\varepsilon}};$$

this estimate leaves a very small gap between the best lower and upper bounds.

Porubsky [41] generalized some of the results on primitive sequences of integers to multiplicative arithmetical semigroups G satisfying J. Knopfnacker's "Axiom A".

For $\mathcal{A} \subseteq \mathbb{N}$ let $Q_{\mathcal{A}}$ denote the set of the integers which can be represented as a quotient a/a' with $a, a' \in \mathcal{A}$, and let $Q_{\mathcal{A}}^{\infty}$ denote the set of the integers which have infinitely many representation in this form. Ahlswede, Khachatrian and Sárközy [31] studied the size properties of $Q_{\mathcal{A}}$ and $Q_{\mathcal{A}}^{\infty}$ under various density assumptions on $\mathcal{A}$.

If $p_1 < p_2 < \cdots < p_t$ are primes, $r < t$, $a = p_1 \dots p_r$ and $b = p_1 \dots p_r p_{r+1} \dots p_t$, then a is said to be a prefix of b. If the set $\mathcal{A} \subseteq \mathbb{N}$ is such that no element of it is a prefix of another then $\mathcal{A}$ is said to be prefix-free. The notion of suffix and suffix-free set is defined similarly. In [33] Ahlswede, Khachatrian and Sárközy studied density properties of prefix-free and suffix-free sets.

Let $y_1, y_2, \dots, y_n, \dots$ be integers greater than 1. Then the products y_1, $y_1 y_2$, …, $y_1 y_2 \cdots y_n$, …are called initial products. Beigböck, Bergelson, Hindman and Strauss [36] proved that if $\mathcal{A} \subset \mathbb{N}$ is "additively large" in a well defined sense then it contains initial products of a given length, resp. infinite sequence of initial products.

In 1938 Erdős [39] estimated the maximal number of positive integers up to N with the property that no one of them divides the product of two others. Chan, Győri and Sárközy [38], resp. Chan [37] extended this problem by studying sets of integers such that no one of them divides the product of k others.

References

1. R. Alexander, Density and multiplicative structure of sets of integers, Acta Arith. 12 (1966/67), 321–332.
2. I. Anderson, On primitive sequences, J. London Math. Soc, 42 (1967), 137–148.
3. I. Anderson, An application of a theorem of de Bruijn, Tengbergen and Kruyswijk, J. Combinatorial Theory 3 (1967), 43–47.
4. I. Anderson, Primitive sequences whose elements have no large prime factors, Glasgow Math. J. 10 (1969), 10–15.
5. I. Anderson, S. D. Cohen and W.W. We Stothers, Primitive polynomial sequences, Mathematika 21 (1974), 239–247.
6. F. Behrend, On sequences of numbers not divisible one by another, J. London Math. Soc. 10 (1935), 42–44.
7. S. D. Cohen, Dense primitive polynomial sequences, Mathematika 22 (1975), 89–91.
8. H. Davenport and P. Erdős, On sequences of positive integers, Acta Arith. 2 (1937), 147–151 and J. Indian Math. Soc. 15 (1951), 19–24.
9. P. Erdős, Note on sequences of integers no one of which is divisible by any other, J. London Math. Soc. 10 (1935), 126–128.
10. P. Erdős, Integers with exactly k prime factors, Ann. Math. 49 (1948), 53–66.

11. P. Erdős, Some extremal problems on divisibility properties of sequences of integers, Fibonacci Quart. 19 (1981), 208–213.
12. P. Erdős and A. Sárközy, On the divisibility properties of sequences of integers, Proc. London Math. Soc. (3) 21 (1970), 97–101.
13. P. Erdős and A. Sárközy, Some asymptotic formulas on generalized divisor functions, I, Studies in Pure Mathematics, To the Memory of Paul Turán, Akadémiai Kiadó, 1983, 165–179.
14. P. Erdős and A. Sárközy, Some asymptotic formulas on generalized divisor functions, II, J. Number Theory 15 (1982), 115–136.
15. P. Erdős and A. Sárközy, Some asymptotic formulas on generalized divisor functions, III, Acta Arith. 41 (1982), 395–411.
16. P. Erdős and A. Sárközy, Some asymptotic formulas on generalized divisor functions, IV, Studia Sci. Math. Hung. 15 (1980), 467–479.
17. P. Erdős, A. Sárközy and E. Szemerédi, On divisibility properties of sequences of integers, Studia Sci. Math, Hung. 1 (1966), 431–435.
18. P. Erdős, A. Sárközy and E. Szemerédi, On the divisibility properties of sequences of integers, I, Acta Arith. 11 (1966), 411–418.
19. P. Erdős, A. Sárközy and E. Szemerédi, On the divisibility properties of sequences of integers, II, 14 (1967/68), 1–12.
20. P. Erdős, A. Sárközy and E. Szemerédi, On an extremal problem concerning primitive sequences, J. London Math. Soc. 42 (1967), 484–488.
21. P. Erdős, A. Sárközy and E. Szemerédi, On a theorem of Behrend, J. Australian Math. Soc. 7 (1967), 9–16.
22. P. Erdős, A. Sárközy and E. Szemerédi, On divisibility properties of sequences of integers, Number Theory, Coll. Math. Soc. J. Bolyai 2 (1970), 35–49.
23. P. Erdős and Zhang Zhenxiang, Upper bound of $\sum 1/(a_i \log a_i)$ for primitive sequences, Proc. Amer. Math. Soc., to appear.
24. J. A. Haight, On multiples of certain real sequences, Acta Arith. 49 (1988), 302–306.
25. W. Klotz, Generalization of some theorems on sets of multiples and primitive sequences, Acta Arith. 32 (1977), 15–26.
26. H. G. Meijer, On the upper asymptotic density of $(0, r)$-primitive sequences, Acta Arith. 25 (1973/74), 191–197.
27. H. G. Meijer, Note on $(0, r)$-primitive sequences, Delft Progress Rep. Ser. F1 (1974/75), no. 3, 82–84.
28. S. Pillai, On numbers which are not multiples of any other in the set, Proc. Indian Acad. Sci. A 10 (1939), 392–394.
29. C. Pomerance and A. Sárközy, On homogeneous multiplicative hybrid problems in number theory, Acta Arithmetica 49 (1988), 291–302.
30. R. Sattler, A theorem concerning $(r_1, r_2, \dots, r_t)$-primitive sequences and an application to $(0, r)$-primitive sequences, Delft Progress Rep. 2 (1976/ 77), no. 1, 23–26.
31. R. Ahlswede, L. H. Khachatrian and A. Sárközy, On the quotient sequence of sequences of integers, Acta Arith. 91 (1999), 117–132.
32. R. Ahlswede, L. H. Khachatrian and A. Sárközy, On the counting function of primitive sets of integers, J. Number Theory 79 (1999), 330–341.
33. R. Ahlswede, L. H. Khachatrian and A. Sárközy, On prefix-free and suffix-free sequences of integers, in: Numbers, Information and Complexity, eds. I Althöfer et al., Kluwer Academic Publishers, Boston, 2000; 1–16.
34. R. Ahlswede, L. Khachatrian and A. Sárközy, On primitive sets of squarefree integers, Periodica Math. Hungar. 42 (2001), 99–115.
35. R. Ahlswede, L. Khachatrian and A. Sárközy, On the density of primitive sets, J. Number Theory 109 (2004), 319–361.

36. M. Beiglböck, V. Bergelson, N. Hindman and D. Straus, Multiplicative structures in additively large sets, J. Combin. Theory Ser. A 113 (2006), 1219–1242.
37. T. H. Chan, On sets of integers, none of which divides the product of k others, European J. Combin. 32 (2011), 443–447.
38. T. H. Chan, E. Győri and A. Sárközy, On a problem of Erdős on integers, non of which divides the product of k others, European J. Combin. 31 (2010), 260–269
39. P. Erdős, On sequences of integers no one of which divides the product of two others and on some related problems, Tomsk. Gos. Univ. Uchen. Zap 2 (1938), 74–82.
40. G. Martin and C. Pomerance, Primitive sets with large counting functions, Publ. Math. Debrecen 79 (2011), 521–530.
41. S. Porubsky, Primitive sequences in arithmetical semigroups, Tatra Mt. Math. Publ. 32 (2005), 85–101.

On Additive Representative Functions

András Sárközy and Vera T. Sós

A. Sárközy (✉)
Department of Algebra and Number Theory, Eötvös Loránd University,
Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary
e-mail: sarkozy@cs.elte.hu

V.T. Sós
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences,
Reáltanoda u. 13-15, H-1053 Budapest, Hungary
e-mail: t.sos.vera@renyi.mta.hu

1. Introduction

In this paper we give a short survey of additive representation functions, in particular, on their regularity properties and value distribution. We prove a couple of new results and present many related unsolved problems.

The study of additive representation functions is closely related to many other topics in mathematics: the first basic questions arose from Sidon's work in harmonic analysis; analytical methods (exponential sums) and combinatorial methods are equally used; Erdős and Rényi introduced probabilistic methods, etc.

Paul Erdős played a dominant role in the advance of this field. As Halberstam and Roth write in their excellent monograph [23] written on sequences of integers:

> Acknowledgements
> Anyone who turns the pages of this book, will immediately notice the predominance of results due to Paul Erdős. In so far as the substance of this book may be said to define a distinct branch of number theory—and its wide range of topics in classical number theory appears to justify this claim — Erdős is certainly its founder. He was the first to recognize its true potential and has been the central figure in many of its developments.
> The authors were indeed fortunate to have the benefit of many discussions with Dr. Erdős. His unique insight and encyclopedic knowledge were, of course, invaluable, but the authors are no less indebted to him for his constant interest and encouragement.

In the last 15 years the authors of this paper and Paul Erdős have written several joint papers on this field (a survey of our joint work is given in [21]). On this very special occasion we have the opportunity to emphasize and

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_16,

appreciate the importance of his contribution to this field, and to thank him for the many invaluable and fruitful discussions with him from which we have learned so much.

2. Notations

The set of integers, non-negative integers, resp. positive integers is denoted by $\mathbb{Z}, \mathbb{N}_0$ and $\mathbb{N}$. $\mathcal{A}, \mathcal{B}, \dots$ denote (finite or infinite) subsets of $\mathbb{N}_0$, and their counting functions are denoted by $A(n), B(n), \dots$ so that, e.g.,

$$A(n) = |\{a : 0 < a \le n, a \in \mathcal{A}\}|.$$

$\mathcal{A}_1 + \mathcal{A}_2 + \dots + \mathcal{A}_k$ denotes the set of the integers that can be represented in the form $a_1 + a_2 + \dots + a_k$ with $a_1 \in \mathcal{A}_1, \dots, a_k \in \mathcal{A}_k$; in particular, we write $\mathcal{A} + \mathcal{A} = 2\mathcal{A} = \mathcal{S}(\mathcal{A})$. For $\mathcal{A} \subset \mathbb{N}$, $\mathcal{D}(\mathcal{A})$ denotes the difference set of the set $\mathcal{A}$, i.e., the set of the positive integers that can be represented in the form $a - a'$ with $a, a' \in \mathcal{A}$. For $\mathcal{A} = \{a_1, a_2, \dots\} \subset \mathbb{N}_0, k \in \mathbb{N}$ we write $k \times \mathcal{A} = \{ka_1, ka_2, \dots\}$.

Representation Functions

For $\mathcal{A} \subset \mathbb{N}_0$, $n \in \mathbb{N}_0$ the number of solutions of the equations

$$a + a' = n \qquad a, a' \in \mathcal{A},$$

$$a + a' = n, \qquad a, a' \in \mathcal{A}, \qquad a \le a'$$

and

$$a + a' = n, \qquad a, a' \in \mathcal{A}, \qquad a < a'$$

is denoted by $r_1(\mathcal{A}, n)$, $r_2(\mathcal{A}, n)$, resp. $r_3(\mathcal{A}, n)$ and are called the additive representation functions belonging to $\mathcal{A}$.

For $g \in \mathbb{N}$, $B_2[g]$ denotes the class of all (finite or infinite) sets $\mathcal{A} \subset \mathbb{N}_0$ such that for all $n \in \mathbb{N}_0$ we have $r_2(\mathcal{A}, n) \le g$, i.e., the equation

$$a + a' = n, \qquad a, a' \in \mathcal{A}, \qquad a \le a'$$

has at most g solutions. The sets $\mathcal{A} \in B_2[1]$ are called Sidon sets.

If $F(n) = O(G(n))$, then we write $F(n) \ll G(n)$.

3. The Representation Function of General Sequences. The Erdős-Fuchs Theorem and Related Results

Erdős and Turán [22] proved in 1941 that for an infinite set $\mathcal{A} \subset \mathbb{N}$, the representation function $r_1(\mathcal{A}, n)$ cannot be a constant from a certain point on. Dirac [6] and Newman proved that the same holds with $r_2(\mathcal{A}, n)$ in place of $r_1(\mathcal{A}, n)$. Since their proof is short and elegant, we present it here:

Let

$$f(x) = \sum_{a \in \mathcal{A}} x^a \qquad (\text{for } x \text{ real}, |x| < 1).$$

If $r_2(\mathcal{A}, n) = k$ for $n > m$, then

$$\frac{1}{2}(f^2(x) + f^2(x^2)) = \sum_{n=0}^{+\infty} r_2(\mathcal{A}, n) x^n = P_m(x) + k\frac{x^{m+1}}{1-x}$$

where $P_m(x)$ is a polynomial of degree $\le m$. If $x \to -1$ from the right, then the right-hand side has a finite limit while the left-hand side tends to $+\infty$. This contradiction proves the theorem.

Moreover, in [22] Erdős and Turán conjectured that their result can be sharpened in the following way: if $\mathcal{A} \subset \mathbb{N}$ and $c > 0$, then

$$\sum_{n=1}^{N} r_1(\mathcal{A}, n) = cN + O(1)$$

cannot hold.

In [12], Erdős and Fuchs proved two theorems one of which sharpens the above mentioned result of Erdős and Turán:

Theorem 1 (**Erdős and Fuchs** [**12**]). *If* $\mathcal{A} = \{a_l, a_2, \dots\} \subset \mathbb{N}, c > 0$, *or* $c = 0$ *and* $a_k < Ak^2$ *(*for $k = 1, 2, \dots$*), and* $i = 1, 2, 3$, *then*

$$\limsup_{N \to +\infty} \frac{1}{N} \sum_{n=0}^{N} (r_i(\mathcal{A}, n) - c)^2 > 0.$$

Their other, better known result (in fact, this is the result known as "the Erdős-Fuchs theorem") proves the conjecture of Erdős and Turán in the following sharper form:

Theorem 2 (**Erdős and Fuchs** [**12**]). *If* $\mathcal{A} \subset \mathbb{N}, c > 0$, *then*

$$\sum_{n=1}^{N} r_1(\mathcal{A}, n) = cN + o(N^{1/4}(\log N)^{-1/2}) \tag{1}$$

cannot hold.

One of the most important problems in number theory is the circle problem, i.e., the estimate of the number of lattice points in the circle $x^2 + y^2 \le N$. Writing

$$\Delta(N) = |\{(x, y) : x, y \in \mathbb{Z}, x^2 + y^2 \le N\}| - \pi N,$$

the problem is to estimate $\Delta(N)$. By a classical result of Hardy and Landau, one cannot have

$$\Delta(N) = o(N^{1/4}(\log N)^{1/4}). \tag{2}$$

The importance of Theorem 2 is based on the fact that the special case $\mathcal{A} = \{1^2, 2^2, \dots\}$ of it corresponds to the circle problem, and the Ω-estimate proved in the much more general Theorem 2 is only by a logarithm power worse than (2).

Theorem 2 has been extended in various directions. Bateman, Kohlbecker and Tull [3] studied the more general problem when the left-hand side of (1) is approximated by an arbitrary "nice" function (instead of cN). Vaughan [40] extended the result to sums of $k(\geq 2)$ terms (see also Hayashi [24]). Richert [29] proved the multiplicative analogue of Theorem 2. Sárközy [34] extended Theorem 2 by giving an Ω-result on the number of solutions of

$$a + b \leq N, \qquad a \in \mathcal{A}, \qquad b \in B.$$

Jurkat showed (unpublished) that the factor $(\log N)^{-1/2}$ on the right-hand side of (1) can be eliminated, and recently, Montgomery and Vaughan [28] published another proof of this result.

Erdős and Sárközy [14, 15] showed that if $f(n) \to +\infty$, $f(n+1) \geq f(n)$ for $n > n_0$ and $f(n) = o\big(\frac{n}{(\log n)^2}\big)$, then

$$\max_{n \leq N} |r_1(\mathcal{A}, n) - f(n)| = o\big((f(N))^{1/2}\big) \tag{3}$$

cannot hold (see also Vaughan [40], Hayashi [24, 25]). Erdős and the authors continued the study of the regularity properties of the functions $r_i(\mathcal{A}, n)$ in [16, 17] and [18], first by studying the *monotonicity properties* of these functions (see also Balasubramanian [2]). In an interesting way, here the three representation functions $r_1(\mathcal{A}, n)$, $r_2(\mathcal{A}, n)$, $r_3(\mathcal{A}, n)$ behave completely differently.

We proved

Theorem 3 (P. Erdős, A. Sárközy, V. T. Sós [17]).

(a) *$r_1(\mathcal{A}, n)$ can be monotone for $n > n_0$ only in the trivial case when $\mathcal{A}$ contains all the positive integers from a certain point on; $A(N) = N - c$ for $N > n_1$.*

(b) *There is an infinite set $\mathcal{A}$ such that $N - A(N) \gg N^{1/3}$ and $r_3(\mathcal{A}, n)$ is monotone increasing for $n > n_0$.*

(c) *If*

$$\lim_{N \to \infty} \frac{N - A(N)}{\log N} = +\infty$$

then $r_2(\mathcal{A}, n)$ cannot be increasing from a certain point on. (See also Balasubramanian [2].)

But we still do not have the answer for

Problem 1. *Does there exist an infinite set $\mathcal{A}$ such that $\mathbb{N} \setminus \mathcal{A}$ is infinite and $r_2(\mathcal{A}, n)$ is increasing from a certain point on?*

As Theorem 6 below shows, it may change the nature of the problem completely if a “thin” set of sums can be neglected. Here we mention two problems of this type:

Problem 2. *Does there exist a set $\mathcal{A} \subset \mathbb{N}$ such that $\mathbb{N} - \mathcal{A}$ is infinite and*

$$r_1(\mathcal{A}; n+1) \geq r_1(\mathcal{A}, n)$$

holds on a sequence of integers n whose density is 1? If such a set exists, then how “dense” can $\mathbb{N} \setminus \mathcal{A}$ be?

Problem 3. *Does there exist an arithmetic function f satisfying $f(n) \to \infty$, $f(n+1) \geq f(n)$ for $n > n_0$, and $f(n) = o\big(\frac{n}{(\log n)^2}\big)$, and a set $\mathcal{A}$ such that*

$$|r_1(\mathcal{A}, n) - f(n)| = o\big((f(n))^{1/2}\big)$$

holds on a sequence of integers n whose density is 1?

Next we studied the following problem: for which sets $\mathcal{A} \subset \mathbb{N}$ is $|r_1(\mathcal{A}, n+1) - r_1(\mathcal{A}, n)|$ bounded? Since we have recently given a survey [21] of these results, thus we do not present further details here.

We complete this section by adding two problems that the first author of this paper could not settle in [34].

Problem 4. *Is it true that if $a_1 < a_2 < \cdots$ and $b_1 < b_2 < \cdots$ are infinite sets of positive integers with*

$$\lim_{k \to +\infty} \frac{a_k}{b_k} = 1$$

and $c > 0$, then

$$|\{(i,j) : a_i + b_j \leq N\}| = cN + O(1)$$

cannot hold?

Problem 5. *Is it true that if $a_1 < a_2 < \cdots$ is an infinite set of positive integers with*

$$a_{k+1} - a_k \gg a_k^{1/2}$$

and $f(n)$ is a “nice” function (say, its second difference $f(n+2) - 2f(n+1) + f(n) \geq 0$) with

$$n \ll f(n) \ll n^{1+\varepsilon},$$

then

$$|\{(i,j) : 0 < |a_i - a_j| \leq N\}| = f(N) + O(1)$$

cannot hold?

(This would cover Dirichlet’s divisor problem in the same way as the Erdős-Fuchs theorem covers the circle problem.)

4. A Conjecture of Erdős and Turán and Related Problems and Results

In 1941 Erdős and Turán [22] formulated the following attractive conjecture:

Conjecture 1 (Erdős and Turán [22]). *If $\mathcal{A} \subset \mathbb{N}$ and $r_1(\mathcal{A}, n) > 0$ for $n > n_0$ (i.e., $\mathcal{A}$ is an asymptotic basis of order 2), then $r_1(\mathcal{A}, n)$ cannot be bounded:*

$$\limsup_{n \to +\infty} r_1(\mathcal{A}, n) = +\infty. \tag{4}$$

This harmlessly looking conjecture proved to be extremely difficult: since 1941 no serious advance has been made. Erdős and Turán formulated an even stronger conjecture:

Conjecture 2 (Erdős and Turán [22]). *If $a_1 < a_2 < \cdots$ is an infinite sequence of positive integers such that for some $c > 0$ and all $k \in \mathbb{N}$ we have $a_k < ck^2$, then (4) holds.*

Erdős and Fuchs [12] remarked that having the same assumptions as in Conjecture 2, the mean square of $r_1(\mathcal{A}, n)$ can be bounded: there are a $c > 0$ and an infinite set $\mathcal{A} \subset \mathbb{N}$ such that $a_k < ck^2$ for all $k \in \mathbb{N}$ and

$$\limsup_{N \to +\infty} \frac{1}{N} \left(\sum_{n=1}^{N} r_1^2(\mathcal{A}, n) \right) < +\infty. \tag{5}$$

Answering a question of Erdős, Ruzsa has proved recently the analogous result in connection with Conjecture 1:

Theorem 4 (Ruzsa, [32]). *There is an infinite set $A \subset \mathbb{N}$ such that $r_1(\mathcal{A}, n) > 0$ for all $n > n_0$ and (5) holds.*

If Conjecture 1 is true, then assuming that $\mathcal{A} \subset \mathbb{N}$, $\mathcal{A}$ is infinite and $r_2(\mathcal{A}, n)$ is bounded, the function $r_2(\mathcal{A}, n)$ must assume the value 0 infinitely often. Erdős and Freud [11] conjectured that having the same assumptions, $r_2(\mathcal{A}, n)$ must assume also the value 1 infinitely often, i.e., there are infinitely many integers $n \in \mathcal{S}(A)$ whose representation in the form

$$a + a' = n, \quad a, a' \in \mathcal{A}, \qquad a \leq a' \tag{6}$$

is unique. This attractive conjecture seems to be true although probably it is very difficult. Moreover, they write "Probably there are "more" integers n with a unique representation of the form (6) than integers n with more than one representation." We will show that this is not so; at least for $\mathcal{A} \in B_2(g)$, $g \geq 3$.

Theorem 5. *For every $g \in \mathbb{N}$, $g \geq 2$ there is an infinite set $\mathcal{A} \subset \mathbb{N}_0$ such that $\mathcal{A} \in B_2[g]$ and for $\varepsilon > 0$, $n > n_0$ we have*

$$|\{n : n \le N, r_2(\mathcal{A}, n) = 1\}| < (1+\varepsilon)\frac{2}{2g-3}|\{n : n \le N, r_2(\mathcal{A}, n) > 1\}|. \quad (7)$$

Proof. Let $\mathcal{E} = \{e_1, e_2, \ldots\}$ be an infinite Sidon set, and define $\mathcal{A}$ by

$$\mathcal{A} = 2g \times \mathcal{E} + \{0, 1, \ldots, g-1\}.$$

We will show that this set $\mathcal{A}$ has the desired properties:

(i) $\mathcal{A} \in B_2[g]$,
(ii) $\mathcal{A}$ satisfies (7).

If $r_2(\mathcal{A}, n) \ge 1$ for some $n \in \mathbb{N}$, i.e., $n \in \mathcal{S}(\mathcal{A})$, then, by the construction of the set $\mathcal{A}$, n can be represented in the form

$$(2ge + i) + (2ge' + j) = 2g(e + e') + (i + j) = n \quad (8)$$

where

$$e, e' \in \mathcal{E}, \quad (9)$$

$$0 \le i, j \le g - 1, \quad (10)$$

$$2ge + i \le 2ge' + j, \quad (11)$$

and $r_2(\mathcal{A}, n)$ is equal to the number of integers e, e', i, j satisfying (8), (9), (10) and (11). It follows from (10) and (11) that

$$e \le e' \quad (12)$$

and

$$0 \le i + j \le 2g - 2. \quad (13)$$

Define the integers u, v by

$$n = 2gu + v, \qquad 0 \le v < 2g. \quad (14)$$

Then it follows from (8), (13) and (14) that

$$e + e' = u \quad (15)$$

and

$$i + j = v \quad (16)$$

(where $v \le 2g - 2$). Since $\mathcal{E}$ is a Sidon set, (12) and (15) determine e and e' uniquely. Thus $r_2(\mathcal{A}, n)$ is equal to the number of pairs (i, j) satisfying (10), (11) and (16). If $e < e'$, then (11) holds automatically, and the number of solutions of (10) and (11) is $v + 1$ for $v \le g - 1$ and $2g - v - 1$ for $g - 1 < v \le 2g - 2$. Denote the set of the integers n that can be represented in the form

$$n = 2g(e + e') + i \qquad (\text{where } e < e',\ e, e' \in \mathcal{E}) \quad (17)$$

with $i = 0$ or $2g-2$ by $\mathcal{K}$, and let $\mathcal{L}$ denote the set of the integers n of form (17) with $1 \le i \le 2g-3$. Then it follows from the discussion above that

$$r_2(\mathcal{A}, n) \begin{cases} = 1 \text{ for } n \in \mathcal{K} \\ > 1 \text{ for } n \in \mathcal{L} \end{cases} \tag{18}$$

and clearly we have

$$K(n) = \frac{2}{2g-3} L(n) + O(1). \tag{19}$$

Finally, if $r_2(\mathcal{A}, n) \ge 1$ and $n \notin \mathcal{K} \cup \mathcal{L}$, then n can be represented in the form

$$n = 2ge + v \qquad \text{with } e \in \mathcal{E}, \quad 0 \le v \le 2g-2;$$

let $\mathcal{M}$ denote the set of the integers n of this form. Clearly,

$$M(n) = o(K(n)). \tag{20}$$

It follows from (18), (19), (20) and

$$\{n : n \in \mathbb{N}, r_2(\mathcal{A}, n) \ge 1\} = \mathcal{K} \cup \mathcal{L} \cup \mathcal{M}$$

that

$$|\{n : n \le N, r_2(\mathcal{A}, n) = 1\}| = (1 + o(1)) \frac{2}{2g-3} |\{n : n \le N, r_2(\mathcal{A}, n) > 1\}|$$

which completes the proof of the theorem. □

By Theorem 5, it is not true that if $r_2(\mathcal{A}, n)$ is bounded, then

$$r_2(\mathcal{A}, n) = 1 \tag{21}$$

holds more often than

$$r_2(\mathcal{A}, n) > 1.$$

On the other hand, we think that (21) must hold for a positive percentage of the elements of $\mathcal{S}(\mathcal{A})$:

Problem 6. *Show that if $\mathcal{A} \subset \mathbb{N}$ is an infinite set such that $r_2(\mathcal{A}, n)$ is bounded, then we have*

$$\limsup_{n \to +\infty} \frac{|\{n; n \le N, r_2(\mathcal{A}, n) = 1\}|}{S(\mathcal{A}, N)} > 0. \tag{22}$$

Note that it could be shown that the lim sup in (22) cannot be replaced by lim inf.

Moreover, if (22) is true, then for sets $\mathcal{A} \in B_2[g]$ one might like to give a lower bound in terms of g for the lim sup in (22). Perhaps Theorem 5 is close to the truth so that this lim sup is $\gg \frac{1}{g}$. The special case $g = 2$ seems to be the most interesting and, perhaps, in this case there is a good chance for a reasonable lower bound:

Problem 7. *Assuming, that $\mathcal{A} \subset \mathbb{N}$ is an infinite set with $\mathcal{A} \in B_2[2]$, i.e., $r_2(\mathcal{A}, n) \le 2$ for all n, give a lower bound for*

$$\limsup_{n \to +\infty} \frac{|\{n : n \le N, r_2(\mathcal{A}, n) = 1\}|}{|\{n : n \le N, r_2(\mathcal{A}, n) = 2\}|}.$$

By Theorem 5, this $\limsup$ *can be ≤ 2; is it true, that it is always ≥ 2?*

By our conjecture formulated in Problem 6, the assumption

$$r_2(\mathcal{A}, n) = O(1) \tag{23}$$

implies that $r_2(\mathcal{A}, n) = 1$ must hold for a positive percentage of the elements of $\mathcal{S}(\mathcal{A})$. First we thought that (23) can be replaced by the weaker condition that $r_2(\mathcal{A}, n)$ is bounded apart from a “thin” set of integers n and still the same conclusion holds. Now we will show that this is not so and, indeed, for every finite set $U \subset \mathbb{N}$ there is a set $\mathcal{A}$ such that, apart from a “thin” set of integers n, $r_2(\mathcal{A}, n)$ assumes only the prescribed values $u \in U$ with about the same frequency.

For $\mathcal{A} \subset \mathbb{N}_0$, $u \in \mathbb{N}$ denote the set of the integers $n \in \mathbb{N}$ with

$$r_2(\mathcal{A}, n) = u$$

by $\mathcal{S}_u(\mathcal{A})$ so that $\mathcal{S}(\mathcal{A}) = \bigcup_{u=1}^{+\infty} \mathcal{S}_u(\mathcal{A})$.

Theorem 6. *Let $k \in \mathbb{N}$ and let $u_1 < u_2 < \ldots < u_k$ be positive integers. Then there is an infinite set $\mathcal{A} \subset \mathbb{N}_0$ such that writing*

$$\mathcal{B} = \mathbb{N} \setminus \left(\bigcup_{i=1}^{k} \mathcal{S}_{u_i}(\mathcal{A}) \right)$$

we have

$$\mathcal{S}_{u_i}(\mathcal{A}, N) = \frac{N}{k} + O(N^\alpha)$$

and

$$B(N) = O(N^\alpha)$$

where $\alpha = \frac{\log 3}{\log 4}$.

(Here $S_{u_i}(\mathcal{A}, N)$ denotes the counting function of $\mathcal{S}_{u_i}(\mathcal{A})$.)

Thus, e.g., there is a set $\mathcal{A}$ such that $r_2(\mathcal{A}, n) = 2$ for all but $O(N^\alpha)$ values of n with $n \le N$.

Proof. The proof will be based on the following lemma:

Lemma 1. *Let $\mathcal{F}$ and $\mathcal{G}$ denote the set of the non-negative integers that can be represented in the form $\sum_{i=0}^{m} \varepsilon_i 2^{2i}$, resp. $\sum_{i=0}^{m} \varepsilon_i 2^{2i+1}$ where $\varepsilon_i = 0$ or 1 for all i, and write $\mathcal{H} = \mathcal{F} \cup \mathcal{G}$. Then* *(i) Every $n \in \mathbb{N}$ has a unique representation in the form*

$$f + g = n, \qquad f \in \mathcal{F}, \qquad g \in \mathcal{G};$$

(ii) $S(\mathcal{F}, N) = O(N^\alpha)$*;*
(iii) $S(\mathcal{G}, N) = O(N^\alpha)$*;*
(iv) We have

$$\big|\{n : n \in \mathbb{N}, r_2(\mathcal{H}, n) > 1\}\big| = O(N^\alpha).$$

Proof. (i) is trivial.

(ii) follows from the fact that if $n \in \mathcal{S}(\mathcal{F})$, then representing n in the form $n = \sum_{i=0}^{m} \varepsilon_i 4^i$ where $\varepsilon_i = 0, 1, 2$ or 3, we have $0 \le \varepsilon_i \le 2$ for all i, i.e., the digit 3 is missing.

(iii) follows from (ii) and $\mathcal{G} = 2 \times \mathcal{F}$.

Finally, (iv) follows from (i), (ii) and (iii), and this completes the proof of the lemma. □

Now we will construct a set $\mathcal{A}$ of the desired properties. Denote the elements of the set $\mathcal{G}$ (defined in Lemma 1) by $(0 =)g_1 < g_2 < \cdots$, write $\mathcal{G}_i = \{g_1, g_2, \dots, g_{u_i}\}$ and $\mathcal{L}_i = k \times (\mathcal{F} + \mathcal{G}_i) + \{i\}$ for $i = 1, 2, \dots k$, and finally, let $\mathcal{A} = \left(\bigcup_{i=1}^{k} \mathcal{L}_i\right) \bigcup (k \times \mathcal{G})$. Clearly, it suffices to show that

(i) If $i \in \{1, 2, \dots, k\}$, $n \in \mathbb{N}$, $n \equiv i \pmod{k}$ and n is large enough (depending on u_i), then n has exactly u_i representations as the sum of an element of $\bigcup_{j=k}^{k} \mathcal{L}_i$ and an element of $k \times \mathcal{G}$;

(ii) For $1 \le i \le j \le k$ we have

$$|\{n : n \le N,\, n \in \mathcal{L}_i + \mathcal{L}_j\}| = O(N^\alpha);$$

(iii) $S(k \times \mathcal{G}, N) = O(N^\alpha)$.

To prove (i), define m by $n = km + i$, and consider a representation of n in the desired form:

$$\ell + kg = n = km + i, \qquad \ell \in \bigcup_{j=1}^{k} \mathcal{L}_j, \quad g \in \mathcal{G}, \tag{24}$$

By the definition of the sets $\mathcal{L}_j$, we have $\ell \in \mathcal{L}_j$ if and only if

$$\ell = k(f + g_t) + j \tag{25}$$

for some $f \in \mathcal{F}, 1 \le t \le u_j$. It follows from (24) and (25) that

$$k(f + g_t + g) + j = km + i. \tag{26}$$

By $1 \le i, j \le k$, this implies that $i = j$. Thus (26) can be written in the equivalent form

$$f + g = m - g_t.$$

By (i) in Lemma 1, for $m > g_{u_i}$ and each of $t = 1, 2, \ldots, u_i$, this equation has exactly one solution in f and g. Again by (i) in Lemma 1, these u_i pairs (f, g) determine distinct solutions (ℓ, g) of (24).

To complete the proof of (i), it remains to show that distinct pairs (ℓ, g), (ℓ', g') satisfying (24) (also with (ℓ', g') in place of (ℓ, g)) determine distinct representations of n if n is large enough, i.e., if

$$\ell + kg = n = \ell' + kg' \tag{27}$$

and n is large, then $\ell \neq kg'$, $\ell' \neq kg$. Indeed, assume that contrary to this statement we have

$$\ell = kg', \quad \ell' = kg. \tag{28}$$

Then by (27) and (28), $\ell + \ell' = n$. Hence

$$\ell \geq n/2 \tag{29}$$

or $\ell' \geq n/2$; we may assume that (29) holds. By (25) and (28) we have

$$\ell = k(f + g_t) + j = kg'. \tag{30}$$

By $1 \leq j \leq k$, it follows that $j = k$. Thus (30) implies

$$f + g_t + 1 = g'. \tag{31}$$

By (25) and (29), we have

$$f \to +\infty \qquad \text{as} \qquad n \to +\infty. \tag{32}$$

It is easy to see that

$$\lim_{x \to +\infty} \min_{\substack{f \in \mathcal{F},\, g \in \mathcal{G}, \\ f, g > x}} |f - g| = +\infty. \tag{33}$$

By (32) and (33), (31) cannot hold for $t \leq u_k$ and large n. This contradiction completes the proof of (i).

To prove (ii), observe that $n \in \mathcal{L}_i + \mathcal{L}_j$ implies that

$$n \in k \times (\mathcal{F} + \mathcal{G}_i) + \{i\} + k \times (\mathcal{F} + \mathcal{G}_j) + \{j\} = k \times \mathcal{S}(\mathcal{F}) + (\{i + j\} + k \times \mathcal{G}_i + k \times \mathcal{G}_j).$$

Here $\{i + j\} + k \times \mathcal{G}_i + k \times \mathcal{G}_j$ is a finite set (in fact, it has at most u_k^2 elements).

Thus (ii) follows from Lemma 1 (ii).

Finally, by $\mathcal{S}(k \times \mathcal{G}) = k \times \mathcal{S}(\mathcal{G})$, (iii) follows from Lemma 1 (ii). □

Remark 1. Let $r_i \in Q^+$, $1 \leq i \leq k$ with $\sum_{i=1}^{k} r_i = 1$. Using the same idea as in the proof of Theorem 6 we can prove the existence of an infinite set $\mathcal{A} \subseteq \mathbb{N}_0$ for which

$$S_{u_i}(\mathcal{A}, N) = r_i N + O(N^\alpha) \qquad 1 \leq i \leq 1$$

with some $0 < \alpha < 1$. It seems likely that an analogous theorem holds with arbitrary given densities $\lambda_i, 1 \leq i \leq k$, in place of r_i. If so, the proof will be more involved.

5. Sidon Sets: The Erdős-Turán Theorem, Related Problems and Results

In 1932 Sidon [36] in connection with his work in Fourier-analysis considered power series of type $\sum_{i=1}^{\infty} z^{a_i}$ when $\big(\sum_{i=1}^{\infty} z^{a_i}\big)^h$ is of bounded coefficients. This led to the investigation of finite and infinite sequences (a_i) with the property that for g fixed the number of solutions

$$a_{i_1} + \cdots + a_{i_k} = n$$

is bounded by g for $n \in N$.

Sidon sequences correspond to the case $h = 2$ and $g = 1$, i.e. $r_2(\mathcal{A}, n) \le 1$. Recall that for $g \in \mathbb{N}$, $B_2(g)$ denotes the class of all (finite or infinite) sets $\mathcal{A} \subset \mathbb{N}_0$ such that for all $n \in \mathbb{N}$ we have $r_2(\mathcal{A}, n) \le g$.

Some specific lines of investigations are the following:

(a) For $\mathcal{A} \in \mathcal{B}_2(g)$ and $\mathcal{A} \subset [1, \dots, N]$ how large can $|\mathcal{A}|$ be? In the infinite case how fast can the counting function $A(n)$ grow?

(b) What can we say about the structure of $\mathcal{A}$ resp. $\mathcal{A} + \mathcal{A}$ if $|\mathcal{A}|$ resp. $A(n)$ is large?

There is an excellent account on this subject in Halberstam-Roth [23] and also a recent survey Erdős-Freud [11].

While there are many results on Sidon sets, much less is known on sets $\mathcal{A} \in B_2[g]$. In particular, let $F(N, g)$ denote the cardinality of the largest set $\mathcal{A} \in B_2[g]$ selected from $\{1, 2, \dots, N\}$. Chowla [5], Erdős [7] and Erdős and Turán [22] gave quite sharp estimates for the cardinality of the largest Sidon set selected from $\{1, 2, \dots, N\}$:

$$N^{1/2} - O(N^{5/16}) < F(N, 1) < N^{1/2} + O(N^{1/4}). \tag{34}$$

On the other hand, very little is known on $F(N, g)$ for $g > 1$. Clearly we have

$$F(N, g) \ge F(N, 1) \qquad \Big(= (1 + o(1))N^{1/2}\Big)$$

for all $g \in \mathbb{N}$. Erdős and Freud [11] showed that $F(N, 2) \ge 2^{1/2} N^{1/2}$. On the other hand, a trivial counting argument gives

$$F(N, g) \le 2g^{1/2} N^{1/2}.$$

Problem 8. *Show that for all $g \in \mathbb{N}$ the limit $\lim_{N \to +\infty} F(N, g) N^{-1/2}$ exists, and determine the value of this limit. In particular, estimate $F(N, 2)$.*

Further, very little is known on sets $\mathcal{A} \in B_2[g]$ and their Sidon subsets. Erdős, resp. Ruzsa (see [7]) studied the size of Sidon sets selected from given sets $\mathcal{A} \in B_2[g]$.

A related problem is the following:

Problem 9. *Is it true that for $g \geq 2$, every Sidon set selected from $\{1, 2, \ldots, N\}$ can be embedded into a much greater set $\mathcal{A} \in B_2[g]$ selected from $\{1, 2, \ldots, N\}$?*

In other words, if $\mathcal{A} \subset \{1, 2, \ldots, N\}$ is a Sidon set, then let $H(\mathcal{A}, N, g)$ denote the cardinality of the greatest set $\mathcal{E}$ such that $\mathcal{E} \in B_2[g]$, $\mathcal{E} \subset \{1, 2, \ldots, N\}$ and $\mathcal{A} \subset \mathcal{E}$. Is it true that writing $K(N, g) = \min(H(\mathcal{A}, N, g) - |\mathcal{A}|)$, where the minimum is taken over all Sidon sets $\mathcal{A}$ selected from $\{1, 2, \ldots, N\}$, we have

$$\lim_{N \to +\infty} K(N, 2) = +\infty?$$

How fast does the function $K(N, g)$ grow in terms of N? Is it true that

$$\lim_{N \to +\infty} (K(N, g+1) - K(N, g)) = +\infty \qquad \text{for all} \qquad g \in \mathbb{N}?$$

A Sidon set $\mathcal{A} \subset \{1, 2, \ldots, N\}$ is said to be *maximal* if there is no integer b such that $b \in \{1, 2, \ldots, N\}$, $b \notin \mathcal{A}$ and $\mathcal{A} \cup \{b\}$ is a Sidon set. (Note that very little is known on the cardinality of maximal Sidon sets; see Problem 15 in [15].) Another problem closely related to Problem 9:

Problem 10. *Does there exist a* maximal *Sidon set such that it can be embedded into a much larger set $\mathcal{E} \in B_2[g]$?*

In other words, let $L(N, g) = \max(H(\mathcal{A}, N, g) - |\mathcal{A}|)$ where $H(\mathcal{A}, N, g)$ is the function defined in Problem 9 and the maximum is taken over all *maximal* Sidon sets selected from $\{1, 2, \ldots, N\}$. Is it true that

$$\lim_{N \to +\infty} L(N, 2) = +\infty?$$

Is it true that

$$\lim_{N \to +\infty} (L(N, g+1) - L(N, g)) = +\infty$$

for all $g \in \mathbb{N}$?

As Sect. 4 also shows, it may change the nature of the problem completely if a "thin" set of sums can be neglected. Several problems of this type are:

Problem 11. *How large set $\mathcal{A}$ can be selected from $\{1, 2, \ldots, N\}$ so that it is an "almost Sidon set" in the sense that*

$$|\mathcal{S}(\mathcal{A})| = (1 + o(1)) \binom{|\mathcal{A}|}{2}? \tag{35}$$

It follows from a construction of Erdős and Freud [11] that there is a set $\mathcal{A}$ such that $\mathcal{A} \subset \{1, 2, \ldots, N\}$, (35) holds and

$$|A| > \left(\frac{2}{\sqrt{3}} + o(1) \right) N^{1/2}, \tag{36}$$

so that $|\mathcal{A}|$ can be much greater than $F(N, 1) = (1 + o(1)) N^{1/2}$ (see (34)).

In the infinite case much less is known than in the finite case. Beyond what follows from (34), Erdős proved

Theorem 7 (**Stöhr** [**38**])**.** *There is an absolute constant $c > 0$, such that for every (infinite) Sidon sequence $\mathcal{A}$*

$$A(n) < c(n/\log n)^{1/2}$$

holds infinitely often.

On the other hand, Krückeberg, improving a result of Erdős, proved in 1961

Theorem 8. *There is an (infinite) Sidon sequence $\mathcal{A}$ such that*

$$A(n) > \frac{1}{\sqrt{2}} n^{1/2}$$

holds infinitely often.

It is not known whether or not the factor $1/\sqrt{2}$ is best possible. The greedy algorithm gives the existence of an (infinite) Sidon sequence for which

$$A(n) > n^{1/3} \quad \text{for all } n.$$

Ajtai, Komlós and Szemerédi improved this [1]: There is a Sidon sequence $\mathcal{A}$ such that

$$A(n) > c(n \log n)^{1/3} \quad \text{for all } n \geq n_0.$$

Weak Sidon Sets

We considered Sidon sets defined by

$$r_2(\mathcal{A}, n) \leq 1 \tag{37}$$

which means that we require

$$x + y \neq u + v \tag{38}$$

for any $x, y, u, v \in \mathcal{A}$ of which at least three are different.

In connection with some particular problems it is more appropriate to consider Sidon sets where we require (38) only for $x, y, u, v \in \mathcal{A}$ where all four are distinct. (So we may have an arithmetic progression of length three, a solution of $x + y = 2u$.)

If

$$r_3(\mathcal{A}; n) \leq 1 \tag{39}$$

holds, $\mathcal{A}$ is called a weak Sidon set.

It is easy to see that the maximum size of Sidon set resp. of a weak Sidon set in $[1, N]$ are asymptotically the same.

A problem of Erdős on the distribution of distances in the plane led us to formulate the following question:

Let $\mathcal{A}^*$ be a weak Sidon set. How large Sidon set $\mathcal{A}$ must be contained by $\mathcal{A}^*$?

Another formulation of the problem is:

Suppose that for $\mathcal{A}^* \subset [1, N]$ any four distinct $a_{i_1}, a_{i_2}, a_{i_3}, a_{i_4} \in \mathcal{A}^*$ determine at least five distinct differences:

$$\left|\{|a_{i_\nu} - a_{i_\mu}|, \quad 1 \le \nu < \mu \le 4\}\right| \ge 5.$$

Let $h(\mathcal{A}^*)$ denote the cardinality of the largest Sidon set $\mathcal{A} \subseteq \mathcal{A}^*$.

Let

$$f(m) = \min_{|\mathcal{A}^*|=m} h(\mathcal{A}^*)$$

If $\mathcal{A}^*$ is a weak Sidon set, then for each $a \in \mathcal{A}^*$ there is at most one pair $b, c \in \mathcal{A}^*$ such that $b + c = 2a$. This implies that

$$f(m) \ge \frac{1}{2} m,$$

Gyárfás and Lehel [27] proved that with some absolute constant $\delta > \frac{1}{100}$

$$\left(\frac{1}{2} + \delta\right) m \le f(m) \le \frac{3}{5} m + 1.$$

Problem 12. *Prove that* $\lim\limits_{m\to\infty} \frac{f(m)}{m}$ *exists and determine the limit.*

It is very probable that a dense weak Sidon set contains a Sidon set of almost the same size:

Problem 13. *Suppose* $\mathcal{A}^* \subset [1, N]$ *and* $m = |\mathcal{A}^*| > \varepsilon N^{1/2}$. *Is it true that*

$$h(\mathcal{A}^*) > \delta(\varepsilon) N^{1/2}$$

where $\delta \to 1$ *if* $\varepsilon \to 1$*?*

Remark 2. The problem of Sidon sets resp. weak Sidon sets is related to anti-Ramsey-type problems.

Consider the complete graph K_N with vertex set $V(K_N) = \{1, \ldots, N\}$ and an edge-coloring $\varphi : [V]^2 \to V$ where $\varphi(a, b) = |a - b|$. A Sidon set $\mathcal{A} \subseteq V$ is the vertex set of a so-called totally multicolored complete subgraph (where *all* the edges have different colors).

A weak Sidon set $\mathcal{A}^* \subseteq V$ corresponds to the vertex set of a complete subgraph where *independent* edges have different colors.

6. Difference-Sets

Above we considered mostly sums $a + a'$. One might like to study the analogues of some of these problems with differences $a - a'$ in place of sums $a + a'$.

Problem 14. *In [19] and [20] we studied the structure of the sum set $\mathcal{S}(\mathcal{A})$ of Sidon sets $\mathcal{A}$. What can be said about the structure of the difference set $\mathcal{D}(A)$ of Sidon sets $\mathcal{A}$; in particular,*

(a) What can be said about the number and length of blocks of consecutive integers in $\mathcal{D}(\mathcal{A})$,
(b) About the size of the gaps between the consecutive elements of $\mathcal{D}(\mathcal{A})$, etc.?

Another closely related problem:
In [19] we studied the solvability of the equation

$$\mathcal{D}(\mathcal{A}) = \mathcal{B}$$

for fixed sets $\mathcal{B} \subset \mathbb{N}$ and, in particular, we gave a quite general sufficient condition for the solvability of this equation and in fact we showed that under quite general circumstances, not only the elements of the difference set $\mathcal{D}(\mathcal{A})$, but also the number of solutions of

$$a - a' = b, \qquad a, a' \in \mathcal{A}$$

(for all $b \in \mathcal{B}$) can be prescribed. The nature of the problem completely changes if we restrict ourselves to Sidon sets $\mathcal{A}$.

Problem 15. *Find possibly general conditions such that for sets $\mathcal{B} \subset \mathbb{N}$ satisfying these conditions, there is a* Sidon set $\mathcal{A}$ *whose difference set is the given set $\mathcal{B}$.*

One might like to see what is the connection between the behavior of sums and differences (see Ruzsa [33] for a related result):

Problem 16. *Consider finite sets $\mathcal{A}$ such that*

$$a - a' = d, \quad a, a' \in \mathcal{A}$$

has at most two solutions for all $d \in \mathbb{N}$ What can be said about the size of the Sidon sets, resp. sets $\mathcal{A} \in B_2[2]$ selected from such a set $\mathcal{A}$?

Problem 17. *Do there exist numbers $\delta > 0$, N_0 such that for $N > N_0$ there is a set $\mathcal{A} \subset \{1, 2, \dots, N\}$ for which*

$$|\mathcal{A}| > (1 + \delta)N^{1/2}$$

and both

$$a - a' = d, \quad a, a' \in \mathcal{A}$$

$$a + a' = n, \quad a, a' \in \mathcal{A}, \quad a \le a'$$

have at most two solutions for all $d \in \mathbb{N}$, $n \in \mathbb{N}$?

7. Generalizations

So far we have studied sums $a+a'$ and differences $a-a'$. Already in these two cases the difficulty of problems and the results can be completely different. It is even more so if we consider the linear form $ca+c'a'$, or more generally $f(a_1,\dots,a_k)=c_1a_1+\dots+c_ka_k$ where $c_i\in\mathbb{Z}$ for $1\le i\le k$ and the c_i's are fixed.

This is indicated already by the following simple but important example.

Example of Ruzsa. Let $\mathcal{A}=\left\{a : a=\sum\limits_{i=0}^{\infty}\varepsilon_i 2^{2i}, \varepsilon_i=0 \text{ or } 1\right\}$. Then for $n\in\mathbb{N}$ the number of solutions

$$a+2a'=n, \qquad a,a'\in\mathcal{A},$$

is 1 for any $n\in\mathbb{N}$.

This shows that the behavior of the representation functions depends very much on the coefficients of the linear form. Here we formulate only a few questions by extending the problems we discussed above.

Problem 18. *For which* $(c_1,\dots,c_k)$ *can the representation-function*

$$R(\mathcal{A},c_1,\dots,c_k;n),$$

counting the number of solutions of $c_1a_1+\dots+c_ka_k=n$ $(a_1,\dots,a_k\in\mathcal{A})$, *be constant for* $n>N_0$*?*

Problem 19. *For which* $(c_1,\dots,c_k)$ *is there an Erdős-Fuchs-type result, analogous to Theorems 1 and 2?*

Problem 20. *For which linear forms is there an Erdős-Sárközy [15]-type result, when* $R_1(\mathcal{A},c_1,\dots,c_k;n)$ *cannot be too close to a "nice" function?*

Problem 21. *When and how the results on the monotonicity of* $r_i(\mathcal{A};n)$ *(see Theorem 3) can be extended to the linear form* $c_1a+\dots+c_ka_k$*?*

One may generalize these problems even further by studying *polynomials* $f(a_1,a_2,\dots,a_k)$. In particular, very little is known on *products* aa'. Erdős [10] estimated the cardinality of sets $\mathcal{A}$ such that $\mathcal{A}\subset\{1,2,\dots,N\}$ and all the products aa' with $a,a'\in\mathcal{A},\ a\le a'$ are distinct. Moreover he [9] studied the multiplicative analogue of the Erdős-Turán conjecture mentioned in Sect. 2. Three further problems involving products are:

Problem 22. *For* $\mathcal{A}\in\mathbb{N}$, $n\in\mathbb{N}$, *let* $s(\mathcal{A},n)$ *denote the number of solutions of the equation*

$$aa'=n, \quad a,a'\in\mathcal{A}, \quad a\le a'.$$

Characterize the regularity properties of this function $s(\mathcal{A},n)$ *analogously as in the papers [14–18] where we discussed the additive analogue of this problem by studying* $r_1(\mathcal{A},n)$, $r_2(\mathcal{A},n)$, $r_3(\mathcal{A},n)$. *In particular, how well can one approximate* $s(\mathcal{A},n)$ *by a "nice" arithmetic function* $f(n)$*?*

Problem 23. *Find a multiplicative analogue of the conjecture of Erdős and Freud mentioned in Sect. 2 and, perhaps, this can be attacked more easily. In other words, is it true that if $\mathcal{A} \subset \mathbb{N}$ is an infinite set such that the function $s(\mathcal{A}, n)$ defined in Problem 15 is bounded, then $s(\mathcal{A}, n) = 1$ for infinitely many values of n?*

Problem 24. *Roth [30, 31], Heath-Brown [26], Szemerédi [39] and others estimated the cardinality of sets $\mathcal{A} \subset \{1, 2, \dots, N\}$ not containing three-term arithmetic progressions. Find a multiplicative analogue of this problem: estimate the cardinality of the largest set $\mathcal{A} \subset \{1, 2, \dots, N\}$ not containing three term geometric progressions, i.e.,*

$$a_1 a_2 = a_3^2, \qquad a_1, a_2, a_3 \in \mathcal{A}$$

implies that $a_1 = a_2 = a_3$. (Note that the square-free integers not exceeding N form a set $\mathcal{A}$ of this property.)

Ramsey-Type Problems

Many of the problems discussed above can be formulated in the following way: if $\mathcal{A}$ is a "dense" set of integers, then an equation of the form

$$f(a_2, a_2, \dots, a_k) = 0 \tag{40}$$

can be solved with $a_1, a_2, \dots, a_k \in \mathcal{A}$. There are several important results of the type where instead of considering solutions $a_1, a_2, \dots, a_k$ belonging to a "dense" set $\mathcal{A}$, we assume that a partition

$$\mathbb{N} = \bigcup_{i=1}^{\ell} \mathcal{A}^{(i)} \qquad (\mathcal{A}^{(i)} \cap \mathcal{A}^{(j)} = \emptyset \quad \text{for} \quad 1 \le i < j \le \ell) \tag{41}$$

of $\mathbb{N}$ is given, and then we are looking for "monochromatic" solutions of (40), i.e., for solutions $a_1, a_2, \dots, a_k$ such that all these a's belong to the same set $\mathcal{A}^{(i)}$; a result of this type can be called a Ramsey-type theorem. In particular, Schur [35] resp. van der Waerden [41] proved that the equation

$$a_1 + a_2 = a_3,$$

resp.

$$a_1 + a_2 = 2a_3, \qquad a_1 \neq a_2$$

has a monochromatic solution for every partition (41) of $\mathbb{N}$. (Indeed, van der Waerden proved the more general theorem that for every $k \in \mathbb{N}$ and every partition (41), there is a monochromatic arithmetic progression of k distinct terms.) It follows from these results that for every partition (41) both equations

$$a_1 a_2 = a_3$$

and

$$a_1 a_2 = a_3^2, \qquad a_1 \neq a_2$$

have monochromatic solutions. (Indeed, in both cases there is a solution of the form $a_1 = 2^{b_1}, a_2 = 2^{b_2}, a_3 = 2^{b_3}$.)

Problem 25. *Characterize the polynomials $f(a_1, a_2, \ldots, a_k)$ such that the Eq.* (40) *has a monochromatic solution for every partition of form* (41) *or, at least, find further polynomials $f(a_1, a_2, \ldots, a_k)$ with this property. In particular, does there exist an integer $m \geq 2$ such that the equation*

$$a_1^2 + a_2^2 + \cdots + a_m^2 = a_{m+1}^2$$

has a monochromatic solution for every partition (41)*?*

8. Probabilistic Methods. The Theorems of Erdős and Rényi

In [36] Sidon asked the following question: Does there exist an $\mathcal{A} \subset \mathbb{N}$ such that $r_1(\mathcal{A}, n) \geq 1$ for all $n > n_0$, and $r_1(\mathcal{A}, n) = O(n^{\varepsilon})$? In 1956 Erdős gave an affirmative answer in the following sharper form:

Theorem 9 (Erdős [8]). *There is an infinite set $\mathcal{A} \subset \mathbb{N}$ such that*

$$c_1 \log n < r_1(\mathcal{A}, n) < c_2 \log n \quad \textit{for } n > n_0.$$

Erdős proved this by a probabilistic argument. In fact, he proved that there are "many" sets $\mathcal{A} \subset \mathbb{N}$ with this property.

In 1960 Erdős and Rényi published an important paper in which, by using probabilistic methods, they proved several results on additive representation functions. The most interesting result is, perhaps, the following theorem:

Theorem 10 (Erdős and Rényi, [13]). *For all $\varepsilon > 0$, there is a $\lambda = \lambda(\varepsilon)$ such that there is an infinite $B_2[\lambda]$ set $\mathcal{A} \subset \mathbb{N}$ with*

$$A(n) > n^{1/2-\varepsilon} \quad \textit{for } n > n_0(\varepsilon).$$

Note that for a $B_2[\lambda]$ set $\mathcal{A}$ we have $A(n) = O_{\lambda}(n^{1/2})$. Thus Theorem 10 provides a quite sharp answer to Sidon's first question, mentioned in Sect. 5.

Remarkably enough, this paper of Erdős and Rényi appeared in the same year as their paper written "On the evolution of random graphs" which had tremendous influence on graph theory and led to one of the most extensively investigated and comprehensive theories in graph theory. (See Bollobás [4].) On the other hand, the paper [13] was nearly unnoticed for about three decades.

The paper [13] of Erdős and Rényi was somewhat sketchy. In their monograph [15] Halberstam and Roth worked out the details. In [16], Erdős and Sárközy extended Theorem 9 by showing that if $f(n)$ is a "nice" function (e.g., combination of the functions n^α, $(\log n)^\beta$, $(\log\log n)^\gamma$) with $f(n) \gg \log n$, then there is an $\mathcal{A} \subset \mathbb{N}$ such that

$$|r_1(\mathcal{A}, n) - f(n)| \ll (f(n)\log n)^{1/2}.$$

(Compare this with their result [14] on Eq. (3).)

The really intensive work in this field started only about 2–3 years ago. Erdős, Nathanson, Ruzsa, Spencer and Tetali have proved several remarkable results. Since their papers have not appeared yet, some of them have not even been written up yet, it would be too early to survey their work here.

Remark 3. Many of the problems in additive number theory are or can be formulated for arbitrary groups, semigroups or for some specified structures, like for set systems. (An independent source for Sidon-type problems is for example coding theory.) We refer to a survey of V.T. Sós [37] on this subject.

Appendix

A1. Introduction

The paper above appeared in 1997. Since that time more than 100 papers have been published on related problems. In this Appendix our goal is to give a short survey of these papers. In order to limit the extent of it we will focus on the most important results, and in the reference list we will present only the records of the most important and most recent papers, and a few survey papers; the references to the further related work can be found in these papers.

A2. Notations

We will keep the notations and the reference numbers of the original paper; thus, e.g., Problem 2 will refer to the second problem in Sect. 3 of the original paper. On the other hand, we will refer to the sections and references in the Appendix by using a prefix A so that, e.g., the second item in the reference list of the Appendix is marked as [43].

A3. The Representation Function of General Sequences. The Erdős-Fuchs Theorem and Related Results

Sárközy [88] proved the following local version of Theorem 1 of Erdős and Fuchs: for all $C > 0$ there are $N_0 = N_0(C)$ and $C_1 = C_1(C)$ so that if $\mathcal{A} \subset \mathbb{N}$ and $N > N_0$, then there is an M with

$$N < M \le N^2 \text{ and } \sum_{n=1}^{M} (R(n) - C)^2 > C_1 M.$$

He also showed that this result is best possible: for all $\varepsilon > 0$ there is an $\mathcal{A} \subset \mathbb{N}$ such that for infinitely many N we have

$$\sum_{n=1}^{M} (R(n) - 2)^2 < \varepsilon M \text{ for all } N < M < \frac{\varepsilon}{136} N^2.$$

Ruzsa [81] proved a "converse" of the Erdős-Fuchs theorem (Theorem 2) by showing that there exists a non-decreasing sequence $\mathcal{A}$ of nonnegative integers such that

$$\sum_{n=1}^{N} r_1(\mathcal{A}, n) = cN + O(N^{1/4} \log N)$$

for some constant $c > 0$.

Tang [93] sharpened Vaughan's result [40] on the extension of the Erdős-Fuchs theorem to k term sums, and later Chen and Tang [46] estimated the constant implied by the ordo notation.

Horváth [68] extended the Erdős-Fuchs theorem further by considering the sum $\mathcal{A}_1 + \mathcal{A}_2 + \cdots + \mathcal{A}_k$ of different sets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$, and later in another paper [64] he sharpened this result.

Let $\mathcal{A} = \{a_1 \le a_2 \le \cdots\}$ be an infinite sequence of nonnegative integers, and write

$$R(\mathcal{A}, x; k) = \left|\left\{\left(a_{i_1}, \ldots, a_{i_k}\right) \in \mathcal{A}^k : a_{i_1} + \cdots + +a_{i_k} \le x\right\}\right|$$

and

$$P(\mathcal{A}, x; k) = R(\mathcal{A}, x; k) - cx.$$

Chen and Tang [45] estimated the mean square of this discrepancy $P(\mathcal{A}, x; k)$.

Lev and Sárközy [74] proved an Erdős-Fuchs-type theorem for finite groups, and they showed that their result is sharp.

Horváth [65] proved the following theorem which is closely related to the first theorem of Erdős and Fuchs (Theorem 1): If $\mathcal{A} = \{a_1, a_2, \ldots\}$ ($a_1 < a_2 < \cdots$) is an infinite sequence of nonnegative integers and d is a positive integer then there is no integer n_0 such that for all $n > n_0$ we have

$$d \le r_3(\mathcal{A}, n) \le d + \left[\sqrt{2d} + \frac{1}{2}\right].$$

Sándor [86] proved a similar theorem, and Chen and Tang [49] extended Horváth's theorem to k term sums and the k term analogues of the other two functions r_1 and r_2.

In our original paper we mentioned the results of Erdős and Sárközy [14, 15] that if the function $f(n)$ satisfies certain assumptions, then (3) cannot hold, and that this theorem is nearly sharp. Horváth [66] extended the first result to k term sums in place of $r_1(\mathcal{A}, n)$, and Kiss [71] proved that Horváth's result is nearly best possible.

In [16] Erdős, Sárközy and T. Sós proved that if $\mathcal{A}$ is an infinite set of positive integers, and, denoting the number of blocks formed by consecutive integers in $\mathcal{A}$ up to N by $B(\mathcal{A}, N)$, we have

$$\lim_{N\to+\infty} \frac{B(\mathcal{A}, N)}{N^{1/2}} = +\infty,$$

then the differences $|r_1(\mathcal{A}, n+1) - r_1(\mathcal{A}, n)|$ cannot be bounded. They also showed that this result is best possible. Kiss extended the theorem to kth differences $|\Delta_k(R(n))|$, and later he also showed [69] that his result is sharp.

In a recent paper Sárközy [89] studied the analogues in $\mathbb{Z}/m\mathbb{Z}$ of the problems considered in [16].

The results of Erdős, Sárközy and T. Sós [17, 18], resp. Balasubramanian [2] on the monotonicity properties of additive representation functions have been extended by Tang [94], Chen and Tang [47, 48], resp. Chen, Sárközy, T. Sós and Tang [50] in various directions. In particular, it is proved in [48] and [50] that if $\mathcal{A}$ is an infinite set of positive integers such that its complement $\mathcal{B} = \mathbb{N}\backslash\mathcal{A}$ satisfies certain simple conditions then $r_2(\mathcal{A}, n)$ cannot be ultimately increasing. However, Problem 1 is still open in its original form.

S. Giri settled the first half of Problem 2 by constructing a set $\mathcal{A}$ of the desired properties (unpublished yet). It might be interesting to study the second half of the problem as well: how dense can $\mathbb{N} \setminus \mathcal{A}$ be for such a set $\mathcal{A}$?

Problems 3–5 are still open.

A4. A Conjecture of Erdős and Turán and Related Problems and Results

Grekos, Haddad, Helou and Pihko [60] proved that if $\mathcal{A}$ is a set of nonnegative integers such that

$$r_1(\mathcal{A}, n) \geq 1 \tag{A4.1}$$

for *every* $n \in \mathbb{N}$ then we have $r_1(\mathcal{A}, n) > 5$ for infinitely many n, and Borwein, Choi and Chu improved this to $r_1(\mathcal{A}, n) > 7$.

Konstantoulas [72] proved that if there is a number n_0 such that if (A4.1) holds for $n > n_0$ then we have $r_1(\mathcal{A}, n) > 5$ for infinitely many n.

By Ruzsa's Theorem 4 there exists an asymptotic basis $\mathcal{A}$ of order 2 such that for $N > N_0$ we have

$$\frac{1}{N}\left(\sum_{n=1}^{N} r_1^2(\mathcal{A}, n)\right) < C$$

for some absolute constant C. In two papers Tang [92] presented explicit values for these constants N_0, C.

For $m \in \mathbb{N}$ let R_m denote the least integer such that there is a set $\mathcal{A} \subset \mathbb{Z}/m\mathbb{Z}$ with $\mathcal{A} + \mathcal{A} = \mathbb{Z}/m\mathbb{Z}$ and $\left|\{(a,b) : a+b = n,\ a, b \in \mathcal{A}\}\right| \leq R_m$ for all $n \in \mathbb{Z}/m\mathbb{Z}$. It follows from Ruzsa's result above that R_m is bounded.

Chen [44] proved the uniform bound $R_m \leq 288$, and Chen and Tang gave better bound for certain m values of special form.

Konyagin and Lev [73] studied and settled the Erdős-Turán problem in infinite Abelian groups. They determined what are the infinite Abelian groups G for which the analogue of the Erdős-Turán conjecture holds and what are the ones for which it fails, and in both cases they provide further information on the number of representations of the elements g of G in the form $a+a' = g$ with a, a' belonging to a basis $\mathcal{A}$ of G.

(See also a paper of Haddad and Helou [62].)

In Sect. 4 we mentioned the conjecture of Erdős and Freud that if $\mathcal{A} \subset \mathbb{N}$ is infinite and $r_2(\mathcal{A}, n)$ is bounded then there are infinitely many n with

$$r_2(\mathcal{A}, n) = 1, \tag{A4.2}$$

and probably there are more integers n satisfying (A4.2) than integers n with

$$r_2(\mathcal{A}, n) > 1.$$

Our Theorem 5 above disproved this second stronger version of the conjecture of Erdős and Freud. Sándor [87] also disproved the weaker version of the conjecture by constructing an infinite set $\mathcal{A}$ of nonnegative integers for which $r_2(\mathcal{A}, n) \leq 3$ for all n and it assumes only the values 0, 2 and 3 infinitely many times. Sándor's construction also disproves the conjecture formulated in our Problem 6 but it does not settle Problem 7. Moreover, in Sándor's construction the counting function $A(n)$ of $\mathcal{A}$ grows slowly: $A(n) = O\big((\log n)^2\big)$. Thus it remains to see whether there exists a set $\mathcal{A}$ such that $A(n) \gg n^c$ for some $c > 0$ and all n, $r_2(\mathcal{A}, n)$ is bounded, and (A4.2) has only finitely many solutions.

A5. Sidon Sets: The Erdős-Turán Theorem, Related Problems and Results

This has been a very intensively studied field in the last 15 years. Since the extent of this Appendix is limited thus we have to restrict ourselves to listing some of the most important papers written on this subject. If the reader wants to know more on the papers written on Sidon sets, then O'Bryant's excellent survey paper [77] can be used, while for more information on large $B_h[g]$ sets one should consult the paper of Cilleruelo, Ruzsa and Vinuesa [51].

In our original paper we mentioned the result of Ajtai, Komlós and Szemerédi [1] on dense infinite Sidon sets: they proved that there is an infinite Sidon set $\mathcal{A}$ with $A(n) \gg (n \log n)^{1/2}$. Ruzsa [83] improved on this significantly by proving that there is an infinite Sidon set $\mathcal{A}$ with $A(n) = n^{\sqrt{2}-1+o(1)}$.

Ruzsa [84] showed that there is a *maximal* Sidon set $\mathcal{A} \subset \{1, 2, \dots, N\}$ with $|\mathcal{A}| \ll (N \log N)^{1/3}$.

Erdős, Sárközy and T. Sós [19, 21] asked whether there is a Sidon set which is also an asymptotic basis of order 3. Deshouillers and Plagne [54] proved in this direction that there is a Sidon set which is also an asymptotic basis of order 7, and Kiss [70] improved on this result by showing that there is a Sidon set which is also an asymptotic basis of order 5.

Answering a question of Sárközy, Ruzsa [82] showed that there is a set $\mathcal{A} \subset \{1, 2, \dots, n\}$ with $|\mathcal{A}| \geq \left(\frac{1}{2} + o(1)\right) n^{1/2}$ which is both additive and multiplicative Sidon set.

Improving on a result of Erdős, Sárközy and T. Sós [19, 20], Spencer and Tetali [91] showed that there exists an infinite Sidon set $\mathcal{A}$ such that any two consecutive elements s_i and s_{i+1} of the sum set $\mathcal{A} + \mathcal{A}$ satisfy $s_{i+1} - s_i < Cs_i^{1/3} \log s_i$ (for $i = 1, 2, \dots$) where C is an absolute constant.

As far as we know Problems 8–12 are still open.

In our original paper we mentioned the Erdős-Turán estimate (34) for the cardinality $F(N, 1)$ of the largest Sidon set selected from $\{1, 2, \dots, N\}$. By (34) we have $F(N, 1) = N^{1/2} + O(N^{5/16})$. We remark that Babai and T. Sós [42] generalized the notion of Sidon set to groups and they studied the size of Sidon sets in groups. Among others, they proved that any finite group G has a Sidon subset of cardinality greater than $c|G|^{1/3}$. This seems to be quite far from being best possible, however, as far as we know it has not been sharpened yet.

A6. Difference-Sets

Some recent results and problems on the connection of sum sets and difference sets are discussed in the survey and problem papers by Martin and O'Bryant [75], Nathanson [76], Ruzsa [85] and Gyarmati, Hennecart and Ruzsa [61].

We do not know about any papers related to Problems 14–17.

A7. Generalizations

Horváth [67] proved partial results related to Problem 18; however, the problem is far from being settled.

On the other hand, we do not know about any papers related to Problems 19–24. In the case of the additive problems the reason of this is probably that the tools used in the special case of sums $a_1 + \dots + a_k$ fail when one tries to extend them to the general case $c_1 a_1 + \dots + c_k a_k$. In the case of the multiplicative problems there does not seem to exist such a barrier, and one would expect that there is a better chance to achieve nontrivial results.

Ramsey-Type Problems

The problems of this type are getting quite popular.

Erdős, Sárközy and T. Sós [59] proved that for any $k \in \mathbb{N}$ and any k-colouring of $\mathbb{N}$, almost all the even numbers have a monochromatic representation in the form $a + a'$ with $a \neq a'$. (This settled a conjecture of Roth.) In a recent paper Borbély [43] extended this result in various directions. (In another paper Erdős and Sárközy [56] also studied the multiplicative analogue of the problem in [59].)

Shkredov [90] proved both density results on the solvability of nonlinear equations of the type

$$f(a_1, \dots, a_n) = 0 \tag{A7.1}$$

over $\mathbb{Z}/p\mathbb{Z}$ and the existence of monochromatic solutions of equations of this type.

Csikvári, Gyarmati and Sárközy [53] also studied both density and Ramsey-type problems involving equations of form (A7.1) over $\mathbb{Z}/m\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{Q}$. Among others they extended Schur's theorem [35] by proving that if $n, k \in \mathbb{N}$ and the prime p is large enough in terms of n and k, then for any k-colouring of $\mathbb{Z}/p\mathbb{Z}$ the Fermat equation

$$x^n + y^n = z^n$$

has a nontrivial monochromatic solution in $\mathbb{Z}/p\mathbb{Z}$. Moreover, they conjectured that for any k colouring of $\mathbb{N}$ the equation

$$a + b = cd, \quad a \neq b \tag{A7.2}$$

has a monochromatic solution, and they proved partial results in this direction. Later Hindman [63] proved this conjecture in a more general form.

P. P. Pach [78] studied the following questions: is it true that if $k \in \mathbb{N}$, and $m \in \mathbb{N}$ is large enough, then the Eqs. (A7.2) and

$$ab + 1 = cd \tag{A7.3}$$

have a "nontrivial" monochromatic solution in $\mathbb{Z}/m\mathbb{Z}$ for any k-colouring of it? He proved that in case of equation (A7.2) the answer is affirmative, while in case of equation (A7.3) one needs further assumptions on the prime factor structure of m to ensure the solvability.

Starting out from a problem of Pomerance and Schinzel, Sárközy asked the following question: is it true that for any r-colouring of the squarefree numbers greater than 1 the equation $ab = c$ has a monochromatic solution? Pomerance and Schinzel [80] proved that the answer is affirmative for $r = 2$, and P. P. Pach [79] also proved this for $r > 2$.

A8. Probabilistic Methods. The Theorems of Erdős and Rényi

Dubickas [55] slightly sharpened Theorem 9 by showing that one can take $c_1 = \varepsilon^2/10$ and $c_2 = 2e + \varepsilon$ in the theorem for any $0 < \varepsilon < 1/2$.

Erdős and Rényi [13] also claimed in their paper that Theorem 10 can be extended from sums of two terms to sums of h terms (for fixed h), i.e., there is a similar theorem on $B_h[\lambda]$ sets in place of $B_2[\lambda]$ sets. However, for $h > 2$ independence issues arise which are not at all easy to handle. This problem was cleared by Vu [95] who gave a complete and correct proof for the following theorem: for $h \in \mathbb{N}$ and $h \geq 2$, and any $\varepsilon > 0$ there is a constant $g = g(\varepsilon)$ and a $B_h[g]$ sequence $\mathcal{A}$ such that $A(x) \gg x^{1/h-\varepsilon}$, and, indeed, one can take $g_h(\varepsilon) \ll \varepsilon^{-h+1}$. (See also the paper [52] of Cilleruelo, Kiss, Ruzsa and Vinuesa.)

We remark that the probabilistic approach is used in many of the papers mentioned in this Appendix.

At the end of Sect. 8 we mentioned a few papers to appear soon; these papers appear as Refs. [57, 91] and [58].

*

We remark that the results described above induce many further problems. In a subsequent paper we will return to some of these problems and also present some related results.

References

1. M. Ajtai, J. Komlós, E. Szemerédi: A dense infinite Sidon sequence, European J. Comb. 2 (1981), 1–11.
2. R. Balasubramanian, A note on a result of Erdős, Sárközy and Sós, Acta Arithmetica 49 (1987), 45–53.
3. P. T. Bateman, E. E. Kohlbecker and J. P. Tull, On a theorem of Erdős and Fuchs in additive number theory, Proc. Amer. Math. Soc. 14 (1963), 278–284.
4. B. Bollobás, Random graphs, Academic, New York, 1985.
5. S. Chowla, Solution of a problem of Erdős and Turán in additive number theory, Proc. Nat. Acad. Sci. India 14 (1944), 1–2.
6. G. A. Dirac, Note on a problem in additive number theory, J. London Math. Soc. 26 (1951), 312–313.
7. P. Erdős, Addendum, On a problem of Sidon in additive number theory and on some related problems, J. London Math. Soc. 19 (1944), 208.
8. P. Erdős, Problems and results in additive number theory, Colloque sur la Théorie des Nombres (CBRM) (Bruxelles, 1956), 127–137.
9. P. Erdős, On the multiplicative representation of integers, Israel J. Math. 2 (1964), 251–261.
10. P. Erdős, On some applications of graph theory to number theory, Publ. Ramanujan Inst. 1 (1969), 131–136.
11. P. Erdős and R. Freud, On Sidon sequences and related problems, Mat. Lapok 1 (1991), 1–44 (in Hungarian).
12. P. Erdős and W. H. J. Fuchs, On a problem of additive number theory, J. London Math. Soc. 31 (1956), 67–73.
13. P. Erdős and A. Rényi, Additive properties of random sequences of positive integers, Acta Arithmetica 6 (1960), 83–110.
14. P. Erdős and A. Sárközy, Problems and results on additive properties of general sequences, I, Pacific J. 118 (1985), 347–357.

15. P. Erdős and A. Sárközy, Problems and results on additive properties of general sequences, II, Acta Math. Hung. 48 (1986), 201–211.
16. P. Erdős, A. Sárközy and V. T. Sós, Problems and results on additive properties of general sequences, III, Studia Sci. Math. Hung. 22 (1987), 53–63.
17. P. Erdős, A. Sárközy and V. T. Sós, Problems and results on additive properties of general sequences, IV, in: Number Theory, Proceedings, Ootacamund, India, 1984, Lecture Notes in Mathematics 1122, Springer-Verlag, 1985; 85–104.
18. P. Erdős, A. Sárközy and V. T. Sós, Problems and results on additive properties of general sequences, V, Monatshefte Math. 102 (1986), 183–197.
19. P. Erdős, A. Sárközy and V. T. Sós, On sum sets of Sidon sets, I, J. Number theory 47 (1994), 329–347.
20. P. Erdős, A. Sárközy and V. T. Sós, On sum sets of Sidon sets, II, Israel J. Math. 90 (1995), 221–233.
21. P. Erdős, A. Sárközy and V. T. Sós, On additive properties of general sequences, Discrete Math. 136 (1994), 75–99.
22. P. Erdős and P. Turán, On a problem of Sidon in additive number theory and some related problems, J. London Math. Soc. 16 (1941), 212–215.
23. H. Halberstam and K. F. Roth, Sequences, Springer-Verlag, New York, 1983.
24. E. K. Hayashi, An elementary method for estimating error terms in additive number theory, Proc. Amer. Math. Soc. 52 (1975), 55–59.
25. E. K. Hayashi, Omega theorems for the iterated additive convolution of a nonnegative arithmetic function, J. Number Theory 13 (1981), 176–191.
26. R. Heath-Brown, Integer sets containing no arithmetic progressions, J. London Math. Soc. 35 (1987), 385–394.
27. A. Gyárfás, Z. Lehel, Linear sets with five distinct differences among any four elements, J. Combin. Theory Ser. B 64 (1995), 108–118.
28. H. L. Montgomery and R. C. Vaughan, On the Erdős-Fuchs theorems, in: A tribute to Paul Erdős, eds. A. Baker, B. Bollobás and A. Hajnal, Cambridge Univ. Press, 1990; 331–338.
29. H.-E. Richert, Zur multiplikativen Zahlentheorie, J. Reine Angew. Math. 206 (1961), 31–38.
30. K. F. Roth, On certain sets of integers, I, J. London Math. Soc. 28 (1953), 104–109.
31. K. F. Roth, On certain sets of integers, II, J. London Math. Soc. 29 (1954), 20–26.
32. I. Z. Ruzsa, A just basis, Monatsh. Math. 109 (1990), 145–151.
33. I. Z. Ruzsa, On the number of sums and differences, Acta Math. Hung. 59 (1992), 439–447.
34. A. Sárközy, On a theorem of Erdős and Fuchs, Acta Arithmetica 37 (1980), 333–338.
35. J. Schur, Über die Kongruenz $x^m + y^m \equiv z^m \pmod p$, Jahresbericht der Deutschen Math. Verein. 25 (1916), 114–117.
36. S. Sidon, Ein Satz über trigonomische Polynome und seine Anwendung in der Theorie der Fourier-Reihen, Math. Annalen 106 (1932), 536–539.
37. V.T. Sós, An additive problem on different structures, 3rd Internat. Comb. Conf., San Francisco 1989. Graph Theory, Comb. Alg. and Appl. SIAM, ed. Y. Alavi, F. R. K. Chung, R. L. Graham, D. F. Hsu (1991), 486–508.
38. A. Stöhr, Gelöste und ungelöste Fragen über Basen der natürlichen Zahlen, J. Reine Angew. Math. 194 (1955), 40–65, 111–140.
39. E. Szemerédi, On a set containing no k elements in an arithmetic progression, Acta Arithmetica 27 (1975), 199–245.
40. R. C. Vaughan, On the addition of sequences of integers, J. Number Theory 4 (1972), 1–16.

41. B. L. van der Waerden, Beweis einer Baudetschen Vermutung, Nieuw Arch. Wisk. 15 (1927), 212–216.
42. L. Babai and V. T. Sós, Sidon sets in groups and induced subgraphs of Cayley graphs, European J. Combin. 6 (1985), 101–114.
43. J. Borbély, On the higher dimensional generalization of a problem of Roth, Integers, to appear.
44. Y.-G. Chen, The analogue of Erdős-Turán conjecture in $\mathbb{Z}_m$, J. Number Theory 128 (2008), 2573–2581.
45. Y.-G. Chen and M. Tang, A generalization of the classical circle problem, Acta Arith. 152 (2012), 279–290.
46. Y.-G. Chen and M. Tang, A quantitative Erdős-Fuchs theorem and its generalization, Acta Arith. 149 (2011), 171–180.
47. Y.-G. Chen and M. Tang, On additive properties of general sequences, Bull. Austral. Math. Soc. 71 (2005), 479–485.
48. Y.-G. Chen and M. Tang, On the monotonicity properties of additive representation functions, II, Discrete Math. 309 (2009), 1368–1373.
49. Y.-G. Chen and M. Tang, Some extension of a property of linear representation functions, Discrete Math. 309 (2009), 6294–6298.
50. Y.-G. Chen, A. Sárközy, V. T. Sós and M. Tang, On the monotonicity properties of additive representation functions, Bull. Austral. Math. Soc. 72 (2005), 129–138.
51. J. Cilleruelo, I. Ruzsa and C. Vinuesa, Generalized Sidon sets, Adv. Math. 225 (2010), 2786–2807.
52. J. Cilleruelo, S. Kiss, I. Z. Ruzsa and C. Vinuesa, Generalization of a theorem of Erdős and Rényi on Sidon sequences, Random Structures Algorithms 37 (2010), 455–464.
53. P. Csikvári, K. Gyarmati and A. Sárközy, Density and Ramsey-type results on algebraic equations with restricted solution sets, Combinatorica 32 (2012), 425–449.
54. J.-M. Deshouillers and A. Plagne, A Sidon basis, Acta Math. Hungar. 123 (2009), 233–238.
55. A. Dubickas, A basis of finite and infinite sets with small representation function, Electron. J. Combin. 19 (2012), Paper 6, 16 pp.
56. P. Erdős and A. Sárközy, On a conjecture of Roth and some related problems, II, in: Number Theory, Proc. of the First Conference of the Canadian Number Theory Association, ed. R. A. Mollin, Walter de Gruyter, Berlin–New York, 1990; 125–138.
57. P. Erdős and P. Tetali, Representations of integers as the sum of k terms, Random Struct. Algorithms 1 (1990), 245–261.
58. P. Erdős, M. B. Nathanson and P. Tetali, Independence of solution sets and minimal asymptotic bases, Acta Arith. 69 (1995), 243–258.
59. P. Erdős, A. Sárközy and V. T. Sós, On a conjecture of Roth and some related problems, I, in: Irregularities of Partitions, eds. G. Halász and V. T. Sós, Algorithms and Combinatorics 8, Springer, 1989; 47–59.
60. G. Grekos, L. Haddad, C. Helou and J. Pihko, On the Erdős-Turán conjecture, J. Number Theory 102 (2003), 339–352.
61. K. Gyarmati, F. Hennecart and I. Z. Ruzsa, Sums and differences of finite sets, Funct. Approx. Comment. Math. 37 (2007), 157–186.
62. L. Haddad and C. Helou, Additive bases representations in groups, Integers 8 (2008), A5, 9 pp.
63. N. Hindman, Monochromatic sums equal to products in $\mathbb{N}$, Integers 11A (2011), Art. 10, 1–10.
64. G. Horváth, An improvement of an extension of a theorem of Erdős and Fuchs, Acta Math. Hungar. 104 (2004), 27–37.

65. G. Horváth, On additive representation function of general sequences, Acta Math. Hungar. 115 (2007), 169–175.
66. G. Horváth, On an additive property of sequences of nonnegative integers, Period. Math. Hungar. 45 (2002), 73–80.
67. G. Horváth, On a property of linear representation functions, Studia Sci. Math. Hungar. 39 (2002), 203–214.
68. G. Horváth, On a theorem of Erdős and Fuchs, Acta Arith. 103 (2002), 321–328.
69. S. Kiss, On a regularity property of additive representation functions, Period. Math. Hungar. 51 (2005), 31–35.
70. S. Z. Kiss, On Sidon sets which are asymptotic bases, Acta Math. Hungar. 128 (2010), 46–58.
71. S. Z. Kiss, On the number of representations of integers as the sum of k terms, Acta Arith. 139 (2009), 395–406.
72. J. Konstantoulas, Laver bounds for a conjecture of Erdős and Turán, Acta Arith., to appear.
73. S. Konyagin and V. T. Lev, The Erdős-Turán problem in infinite groups, in: Additive number theory, 195–202, Springer, New York, 2010.
74. V. F. Lev and A. Sárközy, An Erdős–Fuchs-type theorem for finite groups, Integers 11A (2011), Art. 15, 7 p.
75. G. Martin and K. O'Bryant, Many sets have more sums than differences, in: CRM Proceedings and Lecture Notes, vol. 43, 2007; 287–305.
76. M. B. Nathanson, Sets with more sums than differences, Integers 7 (2007), A5, 24 pp.
77. K. O'Bryant, A complete annotated bibliography of work related to Sidon sequences, Electron. J. Combin. Dynamic Survey 11 (2004), 39.
78. P. P. Pach, Ramsey-type results on the solvability of certain equations in $\mathbb{Z}_m$, Integers, to appear.
79. P. P. Pach, The Ramsey-type version of a problem of Pomerance and Schinzel, Acta Arith., to appear.
80. C. Pomerance and A. Schinzel, Multiplicative properties of sets of residues, Moscow Math.J. Combin. Number Theory 1 (2011), 52–66.
81. I. Z. Ruzsa, A converse to a theorem of Erdős and Fuchs, J. Number Theory 62 (1997), 397–402.
82. I. Z. Ruzsa, Additive and multiplicative Sidon sets, Acta Math. Hungar. 112 (2006), 345–354.
83. I. Z. Ruzsa, An infinite Sidon sequence, J. Number Theory 68 (1998), 63–71.
84. I. Z. Ruzsa, A small maximal Sidon set, Ramanujan J. 2 (1998), 55–58.
85. I. Z. Ruzsa, Many differences, few sums, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 51 (2008), 27–38.
86. C. Sándor, A note on a conjecture of Erdős-Turán, Integers 8 (2008), A30, 4 pp.
87. C. Sándor, Range of bounded additive representation functions, Period. Math. Hungar. 42 (2001), 169–177.
88. A. Sárközy, A localized Erdős–Fuchs theorem, in: Bonner Mathematische Schriften, Nr. 360, Proceedings of the Session in analytic number theory and Diophantine equations (Bonn, January–June 2002), eds. D. R. Heath-Brown and B. Z. Moroz, Bonn, 2003.
89. A. Sárközy, On additive representation functions of finite sets, I (Variation), Period. Math. Hungar., to appear.
90. I. D. Shkredev, On monochromatic solutions of some nonlinear equations in $\mathbb{Z}/p\mathbb{Z}$ (Russian), Mat. Zametki 88 (2010), 603–611.
91. J. Spencer and P. Tetali, Sidon sets with small gaps, in: Discrete probability and algorithms (Minneapolis, MN, 1993), 103–109, IMA Vol. Math. Appl. 72, Springer, New York, 1995.

92. M. Tang, A note on a result of Ruzsa, II, Bull. Austral. Math. Soc. 82 (2010), 340–347.
93. M. Tang, On a generalization of a theorem of Erdős and Fuchs, Discrete Math. 309 (2009), 6288–6293.
94. M. Tang, Some extensions of additive properties of general sequences, Bull. Austral. Math. Soc. 73 (2006), 139–146.
95. V. H. Vu, On a refinement of Waring's problem, Duke Math. J. 105 (2000), 107–134.

Arithmetical Properties of Polynomials

Andrzej Schinzel

A. Schinzel (✉)
Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8,
00-956 Warsaw, Poland
e-mail: schinzel@impan.pl

The present article describes Erdős's work contained in the following papers.

[E1] On the coefficients of the cyclotomic polynomials, Bull. Amer. Math. Soc. 52 (1946), 179–183.

[E2] On the coefficients of the cyclotomic polynomial, Portug. Math. 8 (1949), 63–71.

[E3] On the number of terms of the square of a polynomial, Nieuw Archief voor Wiskunde (1949), 63–65.

[E4] On the greatest prime factor of $\prod_{k=1}^{x} f(k)$, J. London Math. Soc. 27 (1952), 379–384.

[E5] On the sum $\sum_{k=1}^{x} d(f(k))$, ibid. 7–15.

[E6] Arithmetical properties of polynomials, ibid. 28 (1953), 436–425.

[E7] Über arithmetische Eigenschaften der Substitutionswerte eines Polynoms für ganzzahlige Werte des Arguments, Revue Math. Pures et Appl. 1 (1956) No. 3, 189–194.

[E8] On the growth of the cyclotomic polynomial in the interval (0,1), Proc. Glasgow Math. Assoc. 3 (1957), 102–104.

[E9] On the product $\prod_{k=1}^{n}(1-\zeta^{a_i})$, Publ. Inst. Math. Beograd 13 (1959), 29–34 (with G. Szekeres).

[E10] Bounds for the r-th coefficients of cyclotomic polynomials, J. London Math. Soc. (2) 8 (1974), 393–400 (with R. C. Vaughan).

[E11] Prime polynomial sequences, ibid. (2) 14 (1976), 559–562 (with S. D. Cohen, M. B. Nathanson).

[E12] On the greatest prime factor of $\prod_{k=1}^{x} f(k)$, Acta Arith. 55 (1990), 191–200 (with A. Schinzel).

The papers [E1], [E2], [E8] and [E10] concern the coefficients of the cyclotomic polynomial

$$\phi_n(x) = \prod_{d|n}(x^d-1)^{\mu(\frac{n}{d})}.$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_17,

Let A_n denote the largest absolute value of the coefficient of $\phi_n(x)$. It was proved by E. Lehmer in 1936 that

$$A_n = \Omega(n^{1/2}).$$

In [E1] Erdős proved that

$$\log A_n = \Omega((\log n)^{4/3})$$

and in [E2] that

$$\log\log A_n = \Omega\left(\frac{\log n}{\log\log n}\right).$$

In [E8] he gave a simpler proof of the last relation and conjectured that

$$\log\log A_n > c\frac{\log n}{\log\log n}$$

for every $c < \log 2$ and infinitely many n. This conjecture even for $c = \log 2$ has been proved by R. C. Vaughan [11], his result is best possible.

The paper [E10] treats the coefficient $a_r(n)$ of x^r in $\phi_n(x)$. The authors prove that

$$|a_r(n)| < \exp(2\tau^{1/2}r^{1/2} + c_1r^{3/8})$$

and

$$\limsup_{n\to\infty} |a_r(n)| > \exp(c_2(r/\log r)^{1/2}), \quad (r > r_0)$$

where $c_1,\ c_2$ are absolute constants, $c_2 > 0$,

$$\tau = \prod_{p\ \text{prime}} \left(1 - \frac{2}{p(p+1)}\right).$$

The subject is still alive as shows Maier's paper [4].

Somewhat related to the four Erdős's papers discussed is the paper [E9], in which the authors consider the functional

$$M(a_1, a_2, \ldots, a_n) = \max_{|z|=1}\left|\prod_{i=1}^{n}(1 - z^{a_i})\right|$$

($a_1 \le a_2 \le \ldots \le a_n$ are positive integers) and

$$f(n) = \min_{a_1,\ldots,a_n} M(a_1, \ldots, a_n).$$

They prove that $\log f(n) = o(n)$. This result has been sharpened by F. V. Atkinson [1].

In the paper [E3] Erdős considers the sequence $Q(n)$ first studied by A. Rényi. Let $N(f)$ be the number of nonzero coefficients of a polynomial f and let

$$Q(n) = \min N(f^2),$$

where f runs through all polynomials $f \in \mathbb{Q}[x]$ with $N(f) = n$. Rényi proved that $Q(29) < 29$. Erdős proves that $Q(n) \ll n^c$ for a certain $c < 1$ and quotes the conjecture at which he and Rényi independently arrived, namely that

$$\lim_{n\to\infty} Q(n) = \infty.$$

A good value of the constant c has been given by W. Verdenius [12] and the above conjecture has been proved by the writer [7]. The result of [7] has been improved in [9] to the form

$$\mathbb{Q}(n) > 2 + \frac{\log(n-1)}{\log 8} \qquad (n > 1).$$

The subject is still alive as shown in Coppersmith and Davenport's paper [2] and Zannier's paper [14].

The papers [E4] and [E12] concern the greatest prime factor $P(f, x)$ of $\prod_{k=1}^{x} f(k)$, where f is an irreducible polynomial of degree $d > 1$. The first estimate for $P(f, x)$ in the case $f = x^2 + 1$ was given by Chebyshev and his result was extended to all relevant polynomials by T. Nagell in 1921 in the following form

$$P(f, x) > c(f, \epsilon)x(\log x)^{1-\epsilon}$$

for all $\epsilon > 0$ and a suitable $c(f, \epsilon) > 0$.

In [E4] Erdős proved that

$$P(f, x) > x(\log x)^{c(f) \log\log\log x} \text{ for } c(f) > 0,\ x > x_0(f)$$

and asserted that by a much more complicated argument one can show that

$$P(f, x) > x \exp((\log x)^{\delta}) \text{ for } \delta = \delta(f) > 0,\ x > x_l(f). \qquad (*)$$

An attempt made in [E12] to reconstruct the proof of $(*)$ led only to a weaker estimate

$$P(f, x) > x \exp\exp(c(\log\log x)^{1/3}) \text{ for } x > x_2(f)$$

where $c > 0$ is an absolute constant. However G. Tenenbaum [10] has succeeded in proving $(*)$ with any δ less than $2 - \log 4$.

In the paper [E5] Erdős considers the sum

$$S(x) = \sum_{k \le x} d(f(k)),$$

where $d(n)$ is the divisor function and $f \in \mathbb{Z}[x]$ is irreducible. He proves that

$$c_1 x \log x \le S(x) \le c_1 x \log x,$$

where c_1, c_2 are positive constants depending upon f.

The upper estimate, which is much deeper has been considerably generalized by D. Wolke [13] He has replaced $d(n)$ by a multiplicative function and the polynomial f by a quickly growing integer valued function, both functions subject only to mild restrictions. The subject is still alive as shown by Nair's paper [6].

In the paper [E6] Erdős considers values of a polynomial free from k-th powers. It was proved by G. Ricci in 1933 that a primitive irreducible polynomial of degree $d > 1$ represents infinitely many integers free from d-th powers. Erdős improves this as follows: every irreducible polynomial of degree $d > 2$ without a fixed $(d-1)$-th power divisor greater than 1 represents infinitely many integers free from $(d-1)$-th powers. This result has been improved by C. Hooley [3], who has given an asymptotic formula for the number of integers in question and also by M. Nair [5] who has replaced $d-1$ by $[\lambda d]+1$, where $\lambda = \sqrt{2} - \frac{1}{2}$. Hooley gives a tribute to Erdős by saying "It is to the perspicacity of Erdős that we owe our present appreciation of the manifold uses to which sieve methods can be put" (l.c., p. xi).

[E7] is a survey of problems and results, in which in particular [E4], [E5] and [E6] are discussed.

The remaining paper [E11] contains the following theorem. Let $F \in \mathbb{Z}[x]$ be a polynomial of degree $d \geq 2$ such that $F(n) \geq 1$ for all $n \geq 1$. Let $O_F = \{F(n)\}_{n=1}^{\infty}$. Then $F(N)$ is called prime in O_F, if $F(n)$ is not the product of strictly smaller terms in O_F. If $F(x)$ is not of the form $a(bx+c)^d$, then almost all terms of O_F are prime in O_F. The "almost all" is indeed, quantified. For a later, related work, see [8].

References

1. F. V. Atkinson, On a problem of Erdős and Szekeres, Canad. Math. Bull. 4 (1961), 7–12.
2. D. Coppersmith and J. Davenport, Polynomials whose powers are sparse, Acta Arith. 58 (1991), 79–87.
3. C. Hooley, Applications of sieve methods to the theory of numbers, Cambridge University Press 1976.
4. H. Maier, Cyclotomic polynomials with large coefficients, Acta Arith. 65 (1993), 227–235.
5. M. Nair, Power free values of polynomials, Mathematika 26 (1976), 159–183.
6. M. Nair, Multiplicative functions of polynomial values in short intervals, Acta Arith. 62 (1992), 257–269.
7. A. Schinzel, On the number of terms of a power of polynomial, ibid, 49 (1987), 55–70.
8. A. Schinzel and U. Zannier, Distribution of solutions of diophantine equations $f_1(x_1)f_2(x_2) = f_3(x_3)$, where f_i are polynomials, Rend. Sem. Mat. Univ. Padova 87 (1992), 39–68.
9. A. Schinzel and U. Zannier, On the number of non-zero coefficients of a power of a polynomial, Atti Accad. Naz. Lincei, Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 20 (2009), 95–98.
10. G. Tenenbaum, Sur une question d'Erdős et Schinzel II, Inv. Math. 99 (1990), 215–224.

11. R. C. Vaughan, Bounds for the coefficients of cyclotomic polynomials, Michigan Math. J. 21 (1974), 289–295.
12. W. Verdenius, On the number of terms of the square and the cube of polynomials, Indag. Math. 11 (1949), 596–565.
13. D. Wolke, Multiplikative Funktionen auf schnell wachsenden Folgen, J. Reine Angew. Math. 251 (1971), 54–67.
14. U. Zannier, On composite lacunary polynomials and the proof of a conjecture of Schinzel, Invent. Math. 174 (2008), 127–138.

Some Methods of Erdős Applied to Finite Arithmetic Progressions

T. N. Shorey and Robert Tijdeman

T.N. Shorey (✉)
School of Mathematics, Tata Institute of Fundamental Research, Mumbai, 400005, India
e-mail: shorey@math.tifr.res.in

R. Tijdeman
Mathematical Institute, Leiden University, 2300 RA Leiden, The Netherlands
e-mail: tijdeman@math.leidenuniv.nl

Dedicated to P. Erdős

Summary. Since 1934 Erdős has introduced various methods to derive arithmetic properties of blocks of consecutive integers. This research culminated in 1975 when Erdős and Selfridge (Ill J Math 19:292–301, 1975) established the old conjecture that the product of two or more consecutive positive integers is never a perfect power. It is very likely that the product of the terms of a finite arithmetic progression of length at least four is never a perfect power. In the present paper it is shown how Erdős' methods have been extended to obtain results for arithmetic progressions.

1. History Until 1976

In a letter to D. Bernoulli written in 1724 Goldbach argued that the product of three consecutive positive integers is not a square. In 1857 Mlle. A. D. proved that it is not any perfect power. In the same year Liouville showed by use of Bertrand's postulate that $x(x+1)\cdots(x+k-1)$ is not a square or higher power if at least one factor $x, x+1, \ldots, x+k-1$ is prime, or if $k > x-5$. In particular, $k!$ is not a perfect power for $k > 1$. In 1917 Narumi [17] showed that the product of k consecutive integers is not a square for $2 \leq k \leq 202$. In the thirties Szekeres proved that it is not a higher power for $2 \leq k \leq 9$. For more details, see Dickson [1, pp. 679–680], Obláth [18] and Erdős [2, p. 194].

In 1939 Rigge [22] and a few months later Erdős [4] proved that the product of two or more consecutive positive integers is never a square by developing Narumi's proof. They further proved that for fixed $l \geq 3$ the equation

$$x(x+1)\cdots(x+k-1) = y^l \tag{1}$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_18,

has at most finitely many solutions in integers $x > 0, k \geq 2, y \geq 2$. In 1940 Erdős and Siegel jointly proved that there is an absolute constant c such that (1) has no solutions with $k > c$, but this proof was never published. In 1955 Erdős [7] published a different, elementary proof by which c could be computed. The elementary method was developed by Erdős and Selfridge [9] in 1975 for proving that the product of two or more consecutive positive integers is never a cube or a higher power. See Sect. 2 for a sketch of the proof.

The equation (1) *has no solutions in integers* $x > 0, k > 1, l > 1, y > 1$.

Actually the results were more general. Rigge [22] showed that if all prime factors of b are not greater than $\frac{1}{2}k$, then the equation

$$x(x+1)\cdots(x+k-1) = by^l \quad \text{in integers} \quad x > 0, k > 1, l > 1, y > 1 \qquad (2)$$

has no solutions with $l = 2$. Erdős [5] showed that

$$\binom{x+k-1}{k} = y^l \quad \text{in integers} \quad x > k, k > 1, l > 1, y > 1 \qquad (3)$$

is impossible for $l = 3$. Note that equation (3) is the same as equation (2) with $b = k!$ and that it involves no loss of generality to assume that $x > k$, since

$$\binom{x+k-1}{k} = \binom{x+k-1}{x-1}.$$

In 1951 Erdős [6] proved that (3) has no solutions with $k \geq 4$. Observe that $\binom{50}{3} = 140^2$. Erdős and Selfridge actually proved that if $k \geq 3$ and $x + k$ is greater than the least prime $\geq k$, then there is a prime $p \geq k$ which divides the left side of (1) to an order which is not divisible by l. They conjectured that for some $p \geq k$ the order should be one. The results suggest that it may be true that (2) has no solutions with $kl > 6$ for any b composed of prime factors $\leq k$. See Sects. 5, 7 and 8 for results in this direction.

In his 1955 paper Erdős made an extension into a different direction, namely that if from the k integers $x, x+1, \ldots, x+k-1$ less than $(1-\varepsilon)$ $k \log\log k/\log k$ are deleted, the product of the remaining numbers is never an l-th power provided that $\varepsilon > 0$, $k > c(\varepsilon)$, $l > 2$ and $n > k^l$. For $l = 2$ it is allowed to delete $ck/\log k$ numbers. See further Sect. 9.

2. Sketch of the Proof of the Erdős-Selfridge Result for $k \geq 30{,}000$, $l \geq 3$

The proof of Erdős and Selfridge is split into the following cases:

$k \geq 30{,}000$, $l \geq 3$; $4 \leq k < 30{,}000$, $l > 3$; $1{,}000 \leq k < 30{,}000$, $l = 3$; $4 \leq k < 1{,}000$, $l = 3$; $k = 3$, $l \geq 3$; $l = 2$. The latter case is the 1939 result of Rigge and Erdős.

The first case is a quantitative version of Erdős' 1955-result. The proof in this case can be divided in four parts. Suppose

$$x(x+1)\cdots(x+k-1)=y^l \quad \text{in integers} \quad x>0, k>1, l>2, y>1. \qquad (4)$$

Then we write

$$x+j=a_jx_j^l \qquad (0\le j<k)$$

where a_j is not divisible by any l-th power >1. If p divides both a_i and a_j, then p divides $i-j$ and is therefore less than k, It follows that prime factors $\geq k$ only occur in the x_j's and that the a_j's are composed of prime factors $<k$.

The four steps are:

(a) $x>k^l$,
(b) The products $a_ia_j(0\le i<j<k)$ are distinct,
(c) There are $k-\pi(k)$ distinct a_i's with product less than $k!$,
(d) The products $a_ia_j(0\le i<j<k)$ cannot be distinct.

We have a closer look at each step.

Step (a). $x>k^l$.

By Bertrand's postulate there is a prime in $\left[\frac{x+k}{2}, x+k-1\right]$. If $x\le k$, then $\frac{x+k}{2}\ge x$ and this prime divides the left side of (4) to the first power, a contradiction. If $x>k$, then a result of Sylvester, rediscovered by Schur (cf. [2]), states that $x(x+1)\cdots(x+k-1)$ is divisible by some prime greater than k. Such a prime divides only one of the k factors, so $x+k-1\ge(k+1)^l$. Thus $x>k^l$.

Step (b). The products a_ia_j $(0\le i<j<k)$ are distinct. Since $\gcd(x+g, x+i)<k<\sqrt{x}$ for $g\ne i$ by (a) and similarly $\gcd(x+g,x+j)<\sqrt{x}$ for $g\ne j$, we see that $x+g$ does not divide $(x+i)(x+j)$. Thus the products $(x+i)(x+j)$ $(0\le i\le j<k)$ are distinct. Suppose $a_ga_h=a_ia_j$ with $0\le g<h<k, 0\le i<j<k$ and $(x+g)(x+h)>(x+i)(x+j)$. Then

$$(x+g)(x+h)-(x+i)(x+j)=a_ia_j(x_g^lx_h^l-x_i^lx_j^l). \qquad (5)$$

The left side of (5) is smaller than $(x+k)^2-x^2=2kx+k^2<3kx$ by (a). The right side of (5) is larger than

$$a_ia_jl(x_ix_j)^{l-1}>l(a_ix_i^la_jx_j^l)^{(l-1)/l}>lx^{2(l-1)/l}.$$

Thus, by (a) and $l\ge 3$,

$$kx>x^{2(l-1)/l}\ge x^{4/3}>kx$$

a contradiction.

Step (c). There are $k-\pi(k)$ distinct a_i's with product less than $k!$. Since $a_g=a_i$ implies $a_ga_j=a_ia_j$, it follows from (b) that the numbers a_i $(0\le i< k)$ are distinct. For every prime $p<k$, we choose an $f(p)$ in $\{0,1,\ldots,k-1\}$ such that the power of p in a_j for $j=0,\ldots,k-1$ is maximal for $j=f(p)$.

In this way we select at most $\pi(k)$ elements a_j. The total number of factors p in the remaining a_j's is at most

$$\left[\frac{f(p)}{p}\right]+\left[\frac{k-f(p)}{p}\right]+\left[\frac{f(p)}{p^2}\right]+\left[\frac{k-f(p)}{p^2}\right]+\left[\frac{f(p)}{p^3}\right]+\cdots$$
$$\leq\left[\frac{k}{p}\right]+\left[\frac{k}{p^2}\right]+\left[\frac{k}{p^3}\right]+\cdots.$$

By counting the number of factors p in $k!$ we see that

$$k! = \Pi_{p\leq k}p^{[k/p]}.$$

Thus the product of the not selected a_j's is less than $k!$. This argument of Erdős introduced in his 1955 paper has turned out to be fundamental.

Step (d). The products $a_i a_j$ $(0 \leq i < j < k)$ cannot be distinct. Here Erdős and Selfridge apply an elegant graph theoretic lemma to give a quantitative version of a result of Erdős contained in his 1955 paper. A subgraph of a graph is called a rectangle if it comprises two pairs of vertices, with each member of one pair joined to each member of the other.

Lemma 1 ([9, p. 295]). *Let G be a bipartite graph of s white and t black vertices which contains no rectangles. Then the number of edges of G is at most $s+\binom{t}{2}$.*

Proof. Let s_i be the number of white vertices of valency i, so $\sum_{i\geq 1} s_i = s$. Since there are no rectangles, each pair of black vertices is linked to at most one white vertex. A white vertex of valency i corresponds with $\binom{i}{2}$ black pairs. Hence

$$\sum_{i\geq 2} s_i\binom{i}{2} \leq \binom{t}{2}.$$

If E is the number of edges of G, then

$$E=\sum_{i\geq 1} is_i = s+\sum_{i\geq 2}(i-1)s_i \leq s+\sum_{i\geq 2} s_i\binom{i}{2} \leq s+\binom{t}{2}. \qquad \square$$

The lemma is applied as follows. Let $u_1 < u_2 < \cdots < u_s$ and $v_1 < v_2 < \cdots < v_t$ be two sequences of positive integers such that every positive integer up to x can be written in the form $u_i v_j$. If $a_1 < \cdots < a_r < x$ are positive integers such that all the products are distinct, form the bipartite graph G with s white vertices labeled $u_1, \ldots, u_s$ and t black vertices labeled $v_1, \ldots, v_t$ and an edge between u_i and v_j if $u_i v_j = a_m$ for some m. Distinctness of the products $a_i a_j$ ensures that G has no rectangles, so, by Lemma 1,

$$r \leq s+\binom{t}{2}.$$

Using this inequality Erdős and Selfridge show that the product of any $k - \pi(k)$ of the a_i's exceed $k!$ by proving that

$$a_i \geq 3.5694(i - 304) \quad \text{for} \quad i \leq 6{,}993 \tag{6}$$

$$a_i \geq 4.3402(i - 1{,}492) \quad \text{for} \quad i > 6{,}993. \tag{7}$$

For (6) specific sets $\{u_1, \ldots, u_s\}$ and $\{v_1, \ldots, v_t\}$ are constructed with $t = 25$ and $s < \frac{353}{1{,}260}x + 4$, for (7) with $t = 55$ and $s < \frac{2{,}281}{9{,}900}x + 7$. With the inequalities (6) and (7) a routine calculation using Stirling's formula suffices to contradict (c) when $k \geq 30{,}000$.

It is remarkable that Lemma 1 and a variant of it can also be used to derive a simple proof of a result that played a key role in Erdős' 1955-paper. By applying the Cauchy-Schwarz inequality at the last line of the proof of Lemma 1 we obtain

$$E \leq s+\sum_{i\geq 2}(i-1)s_i \leq s+\sqrt{\sum_{i\geq 2}(i-1)^2 s_i}\sqrt{\sum_{i\geq 2} s_i} \leq s+\sqrt{2\binom{t}{2}}\sqrt{s} \leq s+t\sqrt{s}.$$

This yields

Lemma 2. *Under the conditions of Lemma 1 the number of edges of G is at most $s + t\sqrt{s}$.*

Remark 1. *A slight modification of the proof yields the upper bound $\frac{1}{2}s + t\sqrt{s} + \frac{s^{3/2}}{8t}$.*

3. Integers with Distinct Products

Denote by $N(X)$ the maximum number of integers $1 \leq b_1 < b_2 < \cdots b_r \leq X$ so that the products $b_i b_j$ $(1 \leq i < j \leq r)$ are distinct. Since we can take all primes $\leq X$, the number $N(X)$ can be as large as $\pi(X) \sim X/\log X$. Erdős proved the striking fact that $N(X)$ can hardly be larger. In [3], published in 1938, he proved $N(X) < \pi(X) + 8X^{3/4} + X^{1/2}$, in [7] he showed by a different proof that $N(X) < \pi(X) + 3X^{7/8} + 2X^{1/2}$. Finally, in 1968/1969, he proved in [8] that

$$N(X) - \pi(X) \ll n^{3/4}/(\log n)^{3/2}$$

and that there exist sequences such that

$$N(X) - \pi(X) \gg n^{3/4}/(\log n)^{3/2}.$$

We use Lemmas 1 and 2 to give an elegant proof of the following estimate, applying arguments due to Erdős.

Theorem 1. $N(X) < \pi(X) + X^{7/8} + X^{3/4} + X^{1/2}$.

Proof. Let $1 \leq b_1 < b_2 < \cdots b_r \leq X$ be such that all the products $b_i b_j$ $(1 \leq i \leq j \leq r)$ are distinct. We write $b_i = u_i v_i$ where v_i is the greatest divisor of b_i which is not greater than $X^{1/2}$. First of all it is clear that the numbers $u_1 v_1, u_1 v_2, u_2 v_1, u_2 v_2$ cannot all be b's, since $(u_1 v_1)(u_2 v_2) = (u_1 v_2)(u_2 v_1)$.

Next we show if $b_i = u_i v_i$ with $v_i < X^{1/4}$, then u_i must be a prime. For if not, let p be the least prime factor of u_i. If $p < X^{1/4}$ then $p v_i < X^{1/2}$ which contradicts the maximum property of v_i. Since u_i is assumed to be composite we have $p \leq X^{1/2}$. Hence $X^{1/4} \leq p \leq X^{1/2}$. But then $p > v_i$ which again contradicts the maximum property of v_i. Thus u_i must be a prime indeed.

As before we form a bipartite graph G with s white vertices u_i and t black vertices v_i and an edge between u_i and v_i if $b_i = u_i v_i$ for some i. Distinctness of the products $b_i b_j$ ensures that G has no rectangles.

First we count the number of edges in G with u-value greater than $X^{3/4}$. Since $uv \leq X$, the v-value is then less than $X^{1/4}$. We have shown that in this case u is prime. By Lemma 1 we find the following upper bound for the number of edges:

$$E_1 \leq \pi(X) + \frac{1}{2} X^{1/2}.$$

Secondly we count the number of edges in G with u-value in $(1, X^{3/4}]$. The v-value is at most $X^{1/2}$. By Lemma 2 we find for the number of such edges:

$$E_2 \leq X^{3/4} + X^{7/8}.$$

The number $N(X)$ is bounded by $E_1 + E_2 \leq \pi(X) + X^{7/8} + X^{3/4} + X^{1/2}$.□

4. Arithmetic Progressions Composed of Small Primes

We consider the generalisations to finite arithmetic progressions. In step (a) it was shown that x is large compared with k. It sufficed to show that $x(x+1) \cdots (x+k-1)$ was divisible by a prime $p > k$. In this section we consider the greatest prime factor of the product of the terms of a finite arithmetic progression.

Let x, d and k be positive integers. We consider

$$\Delta := \Delta(x, d, k) := x(x+d) \cdots (x+(k-1)d) \tag{8}$$

and in particular $P(\Delta)$, the greatest prime factor of Δ Bertrand's postulate, proved by Chebyshev, states that for any positive integer k the sequence $k+1, k+2, \ldots, 2k$ contains a prime, that is $P(\Delta(k+1, 1, k)) > k$. In 1892 Sylvester [32] generalised this inequality by showing $P(\Delta(x, d, k)) > k$ for $x \geq k + d$. Langevin [14] proved that $P(\Delta(x, d, k)) > k$ for $x > k$.

If $d = 1$, the problem of determining the best lower bound for x is equivalent with the classical problem how large gaps between consecutive primes can be. Using results on this gap problem the condition $x > k$ can be replaced by $x > k/13$ when $k > 118$ by a result of Rohrbach and Weis [RW] and by $x > k^{.548}$ when k is sufficiently large by improvements of the theorems of Hoheisel and Ingham (cf. [21, p. 193]).

For $d = 2, 3, 4$ and 6 Breusch, Molsen, Rohrbach and Weis, and Erdős derived extensions of Bertrand's postulate. References can be found in Moree [16] who extended the results to 54 values of d less than 1,000. Each extension provides an improvement of Langevin's result for that value of d. The authors showed in [28] that

$$P(\Delta(x, d, k)) > k \quad \text{if} \quad d > 1 \quad \text{and} \quad (x, d, k) \neq (2, 7, 3). \tag{9}$$

They applied a sharp upper bound for $\pi(x)$ due to Rosser and Schoenfeld. In [31] they further proved that

$$P(\Delta(x, d, k)) \gg k \min\left(\log\log(x + (k-1)d), \log\left(\frac{x}{k-1} + d\right)\right)$$

and that, for any $\varepsilon > 0$,

$$P(\Delta(x, d, k)) \gg_\varepsilon k \log\log(x + (k-1)d) \quad \text{for} \quad x + (k-1)d > k(\log k)^\varepsilon.$$

These results are based on upper bounds for the solutions for Thue-Mahler equations due to Győry.

The stated results enable us to conclude:

Theorem 2. *The equation*

$$x(x + d) \cdots (x + (k-1)d) = y^l \tag{10}$$

has no solutions in integers $x > 0$, $d > 0$, $k > 1$, $l > 1$, $y > 1$ with $P(y) \leq k$.

Proof. For $d = 1$ Theorem 2 follows from Theorem 1, for $d > 1$ from (9). □

5. Perfect Powers in Products of the Terms of an Arithmetic Progression, I

A natural generalisation of equation (1) is

$$x(x+d) \cdots (x+(k-1)d) = y^l \quad \text{in integers} \quad x > 0, d > 0, k > 2, l > 1, y > 1. \tag{11}$$

Without loss of generality we assume that l is a prime number. Theorem 2 implies that $P(y) > k$. For any numbers x, d and k it is easy to find a positive integer A such that the left side of (11) with x replaced by Ax and d replaced by Ad represents a perfect power. To avoid such solutions we shall assume in the sequel that $\gcd(x, d) = 1$.

Soon after he proved his joint result with Selfridge, Erdős conjectured that (11) implies that k is bounded by an absolute constant and later he conjectured that even $k \leq 3$. The theory on the Pell equation yields that there are infinitely many pairs x, d with $\gcd(x, d) = 1$ such that $x(x+d)(x+2d)$ is a square. On the other hand, Euler [10] (cf. [1, p. 635]) has shown that (11) has no solutions with $k = 4, l = 2$ and Obláth [19, 20] has proved that (11) has no solutions with $(k, l) = (5, 2)$, $(3, 3)$, $(3, 4)$ and $(3, 5)$.

Marszalek [15] showed that k is bounded if d is fixed. More precisely, he proved that (11) implies that $k \leq \exp(C_1 d^{3/2})$ if $l = 2$, $k \leq \exp(C_2 d^{7/3})$ if $l = 3$, $k \leq C_3 d^{5/2}$ if $l = 4$, $k \leq C_4 d$ if $l \geq 5$ with explicitly stated absolute constants C_1, C_2, C_3, C_4.

We shall consider the more general equation

$$x(x+d)\cdots(x+(k-1)d) = by^l \text{ in integers } x > 0, d > 0, k > 2, b > 0, l > 1, y > 1 \tag{12}$$

subject to $\gcd(x, d) = 1$, $P(b) \leq k$, l is prime. One of us conjectured in [33, p. 219] that (12) has only finitely many solutions with $kl > 6$. Shorey [26] applied the theory of linear forms in logarithms to show that (12) with $l \geq 3$ implies that k is bounded by a computable number depending only on $P(d)$. Shorey and Tijdeman [27, 29] derived various improvements. For example, they showed that there exist effectively computable absolute constants C_5, C_6, C_7, C_8, C_9 such that, for $k \geq C_5$,

$$\begin{array}{llll} (a) & d_1 \geq C_6 k^{l-2} & (b) & d_1 \geq k^{C_7 \log\log k} \\ (c) & P(d) \geq C_8 \log k \log\log k & (d) & l^{w(d)} \geq \frac{C_9 k}{\log k} (l \geq 2) \end{array} \tag{13}$$

where d_1 is the maximal divisor of d such that all the prime factors of d_1 are $\equiv 1 \pmod{l}$ and $w(d)$ denotes the number of distinct prime factors of d. Since $d_1 | d$, the results (a) (cf. [27, (2.7)]) and (b) (cf. [29, Theorem 1 and (2)]) provide considerable improvements of Marszalek's estimates. In particular $k \ll d^{\varepsilon}$ for any $\varepsilon > 0$. Inequality (c) (cf. [29, Corollary 1]) shows, for $l > 1$, how k can be bounded in terms of $P(d)$. Inequality (d) (cf. [27, Theorem 1 and Corollary 1]) shows that for fixed $l \geq 2$ the number k is even bounded by a computable number depending only on $w(d)$.

First we shall illustrate the method by showing how the extension of Erdős' ideas leads to a proof of (13.a).

Theorem 3 ([27, (2.7)]). *Let* (12) *hold subject to* $\gcd(x, d) = 1$, $P(b) \leq k$, *l is prime and let d_1 be the maximal divisor of d such that all the prime factors of d_1 are* $\equiv 1 \pmod{l}$. *Then, for k sufficiently large,*

$$d_1 \gg k^{l-2}.$$

Proof. We assume $l > 2$. We compare with the steps in Sect. 2.
Step (a) We show that $x + (k-1)d > k^l$.

Write

$$x + jd = A_j y_j^l \quad (0 \leq j < k) \tag{14}$$

where A_j is composed of primes $\leq k$ and y_j of primes $> k$. Since $P(y) > k$, one of the terms of the arithmetic progression is divisible by an l-th power of a prime $> k$, whence

$$x + (k-1)d > k^l \qquad \text{and} \qquad x + d > k^{l-1}. \tag{15}$$

Step (b) We show that we may assume that the A_j's with $j \geq k/8$ are distinct. Suppose that $A_i = A_j$ for some $i > j > 0$. Then, by (14),

$$(i-j)d = A_j(y_i^l - y_j^l). \tag{16}$$

Since $\gcd(x,d) = 1$, we see that $A_j | (i-j)$ whence $A_j < k$. Hence, by step (a), $y_i > k$ and $y_j > k$. On the other hand, $d | (y_i^l - y_j^l)$. It is well known that every prime factor of

$$\frac{y_i^l - y_j^l}{y_i - y_j} \tag{17}$$

is either l or $\equiv 1 \pmod{l}$ and that l occurs at most to the first power in (17). Consequently

$$y_i - y_j \geq d/ld_1. \tag{18}$$

Now from (16) we derive

$$kd > (i-j)d > A_j^{1/l}(y_i - y_j)l(A_j y_j^l)^{(l-1)/l}.$$

If $j \geq k/8$, then we obtain, by (15),

$$dk > \frac{d}{d_1}\left(x + \frac{k}{8}d\right)^{(l-1)/l} > \frac{d}{8d_1}k^{l-1}$$

which implies $d_1 > \frac{1}{8}k^{l-2}$. Thus we may assume that the numbers A_i with $i \geq k/8$ are distinct.

Back to step (a) We show $x + (k-1)d > k^{l+1}/4$.

If $y_i = 1$, then $x + id$ is composed of primes $\leq k$. By the argument in step (c) of Sect. 2 there are at most $\pi(k)$ selected A_i's and the product of the remaining A_i'S is at most $k!$. Therefore, by (15) and $l > 2$, the number of i with $0 < i < k$ and $y_i = 1$ is bounded by

$$\pi(k) + \frac{\log(k!)}{\log(k^{l-1})} < \frac{2k}{\log k} + \frac{k}{l-1} < \frac{5k}{8}$$

for k large. Let S_0 denote the set of elements A_i with $i \geq k/8$ and $y_i > 1$. Then $|S_0| \geq k/4$. Note that $y_i > k$ for $A_i \in S_0$. Hence

$$x + (k-1)d \geq \max_{A_i \in S_0}(A_i y_i^l) \geq \frac{1}{4}k^{l+1} \tag{19}$$

and

$$x + d \geq \frac{1}{4}k^l. \tag{20}$$

Back to step (b). We show that we may assume that the A_j's with $0 < j < k$ are distinct.

Suppose that $A_i = A_j$, but $i > j > 0$. Then, as in step (b) above, but now with (20) instead of (15),

$$dk > \frac{d}{d_1}(x+d)^{(l-1)/l} > \frac{1}{4}\frac{d}{d_1}k^{l-1},$$

which implies $d_1 > \frac{1}{4}k^{l-2}$.
Note. We now turn to steps (c) and (d) and only later deal with step (b).

The argument given in step (c) of Sect. 2 applies similarly to the A_j's:
Step (c). There exists a subset T of $\{0, 1, \ldots, k-1\}$ consisting of at least $k - \pi(k)$ elements such that $\prod_{j \in T} A_j < k!$.

It follows that the average value of these A_i's is at most $(k/e)^{k/(k-\pi(k))}$ which is asymptotically equal to k. A straightforward application of the box principle would yield that for any η with $0 < \eta < 1$ there are at least ηk numbers $j > 0$ such that $A_j \leq k^{1/\eta}$. This estimate is too rough for applications. The following lemma provides a useful sharpening of this bound when the set $\{A_j\}_{j \in T}$ is large.

Lemma 3 (cf. [27, Lemma 6]). *Let $0 < \eta < 1$. Put $S_1 = \{A_j\}_{j \in T}$. Suppose*

$$|S_1| \geq k - \frac{g^k}{\log k} \tag{21}$$

where $g < (1-\eta)\log k$. Then there exists a subset S_2 of S_1 with at least η_k elements satisfying

$$A_j \leq Ck \quad \textit{for} \quad A_j \in S_2 \tag{22}$$

where $C = \exp((g + \eta + .37)/(1 - \eta - g/\log k))$.

Proof. Let S_2 be the subset of S_1 defined by (22). By steps (b) and (c) we have

$$k! \geq \Pi_{A \in S_1} A \geq (|S_2|)!(Ck)^{|S_1|-|S_2|}.$$

Suppose $|S_2| < \eta k$. Then, by $n! > (n/e)^n$ for $n = 1, 2, \ldots$ and the fact that $(y/x)^y$ is monotonic decreasing in y for $0 < y < x/e$ and (21), we obtain

$$k! \geq \left(\frac{|S_2|}{eCk}\right)^{|S_2|}(Ck)^{|S_1|} \geq \left(\frac{\eta}{eC}\right)^{\eta k}(Ck)^{k-\frac{gk}{\log k}}$$

$$= \left(\frac{\eta^\eta C}{e^\eta C^\eta C^{g/\log k} e^g}\right)^k k^k.$$

Since $\eta^\eta > e^{-.37}$ for $0 < \eta < 1$ and $k^k > k!$, we obtain a contradiction. □

Step (d). The products $A_i A_j (0 < i < j < k)$ cannot be distinct.

By step (b) we know that the numbers A_i $(i > 0)$ are distinct. Hence the set T in Lemma 3 consists of at least $k - 1 - \pi(k) > k - 2k/\log k$ elements for k large. We apply Lemma 3 with $g = 2$ and $\eta = 1/3$. Then we find a set S_2 of at least $k/3$ elements A_i satisfying

$$A_i \leq 60k \qquad \text{for} \quad A_i \in S_2$$

for k large. By (20), $y_i > k$ for $A_i \in S_2$. We write S_3 for the set of all $A_i \in S_2$ with

$$i \geq \frac{k}{9} \quad \text{and} \quad A_i \geq \frac{k}{9}. \tag{23}$$

Then $|S_3| \geq k/10$. It follows from Theorem 1 with $X = 60k$ that the products A_iA_j with $A_i, A_j \in S_3$ cannot be distinct for k large.

Step (b). The products A_iA_j $(A_i, A_j \in S_3)$ are distinct. Since $\gcd(x+id, x+jd)$ divides both $(i-j)x$ and $(i-j)d$ and $\gcd(x,d) = 1$, we have $\gcd(x+id, x+jd) \leq |i-j| < k$ for $i \neq j$. Hence, by (20), $\gcd(x+gd, x+id) < k \leq \sqrt{x}$ for $g \neq i$ and similarly $\gcd(x+gd, x+jd) < \sqrt{x}$ for $g \neq j$. It follows that the products $(x+id)(x+jd)$ $(0 \leq i \leq j < k)$ are distinct.

Suppose there are elements A_g, A_h, A_i, A_j of S_3 satisfying $A_gA_h = A_iA_j$ with $g \neq i$ and $g \neq j$. Without loss of generality we may assume that

$$(x+gd)(x+hd) - (x+id)(x+jd) = A_iA_j((y_gy_h)^l - (y_iy_j)^l) \tag{24}$$

is positive. Since d divides the left side and $\gcd(d, A_iA_j) = 1$, we see that d divides the difference of the l-th powers. Hence, by (18),

$$y_gy_h - y_iy_j \geq \frac{d}{ld_1}.$$

Hence the right side of (24) is at least

$$(A_iA_j)^{1/l}\frac{d}{d_1}((A_iy_i^l)(A_jy_j^l))^{l-1/l}.$$

Since $A_i, A_j \in S_3$, we obtain from (23) the lower bound

$$\left(\frac{k}{9}\right)^{2/l}\frac{d}{d_1}\left(x + \frac{k}{9}d\right)^{2(l-1)/l} \geq \frac{1}{9^2}\frac{d}{d_1}k^{2/l}(x+kd)^{2-2/l}.$$

The left side of (24) is at most $2kd(x+kd)$. Comparing both bounds we obtain, by (19),

$$162d_1 \geq \left(\frac{x+kd}{k}\right)^{1-2/l} \geq \frac{1}{4}k^{l-2}.$$

This completes the proof of Theorem 3. □

6. Some Applications of Brun's Sieve

Erdős has introduced some applications of Brun's sieve which are used in the proofs of (13.b), (13.c), and (13.d), but have also interest in themselves. Let $0 < z < X$. Brun's sieve implies that for any $A > 0$ the number $\Phi(X, w)$ of integers $\leq X$ free of prime factors $\leq w$ satisfies

$$\Phi(X, w) \ll_A X \,\Pi_{p \leq w}(1 - \frac{1}{p}) \quad \text{for} \quad w \leq X^A$$

(cf. [13, p. 68]). Since $\Pi_{p \leq w}(1 - \frac{1}{p}) \ll \frac{1}{\log w}$, we conclude that

$$\Phi(X, w) \ll_A \frac{X}{\log w} \quad \text{for} \quad w \leq X^A. \tag{25}$$

Erdős [5] applies (25) in the following way.

Lemma 4. *Let $b_1, \ldots, b_s$, denote all integers between z and X such that every proper divisor of b_i, is at most z. Then*

$$s \ll \frac{X}{\log(X/z)}.$$

Proof. If $z > \sqrt{X}$, then $b1, \ldots, b_s$ are prime numbers and the result is obvious. We assume $z \leq \sqrt{X}$. Put $y = X/\log(X/z)$. If b_i, is larger than y, then every prime factor of b_i, is greater than y/z. By (25) the number of such b_i is at most

$$\ll \frac{X}{\log(y/z)}.$$

The number of b_i's not exceeding y is at most y. Hence

$$s \ll \frac{X}{\log(X/z)} + \frac{X}{\log(y/z)}.$$

Since $\log(y/z) = \log(X/z) - \log\log(X/z) \geq \frac{1}{2}\log(X/z)$, the result follows. □

Lemma 4 can be used to show that in a dense sequence some numbers have a large common divisor.

Lemma 5 (cf. Erdős [5]). *Let r, z, X be integers with $0 < z < X$. Let $0 < a_1 < a_2 < \cdots < a_r \leq X$ be a sequence of integers. Then there are at least*

$$r - z - \frac{c_1 X}{\log(X/z)} \tag{26}$$

pairs a_i, a_j for which $\gcd(a_i, a_j) > z$ where c_1 is some absolute constant.

Proof. Denote by $b_1, \ldots, b_s$ all integers between z and X having every proper divisor $\leq z$. Then, by Lemma 4,

$$s \le \frac{c_1 X}{\log(X/z)}.$$

Obviously every integer between z and X has a divisor among the b's. Hence there are at least $r - z - s$ pairs a_i, a_j for which $\gcd(a_i, a_j)$ is divisible by a b, whence $\gcd(a_i, a_j) > z$. □

Remark 2. *If $r > z + 2s$, then a better bound is possible. Adjoin to each a_j larger than z one divisor b. If the number of a_i's corresponding to b_i equal t_i, then the number of pairs is $\binom{t_i}{2}$. Hence the total number of pairs is $\sum_{i=1}^{s} \binom{t_i}{2}$. By the Cauchy-Schwarz inequality we obtain*

$$\sum_{i=1}^{s} \binom{t_i}{2} = \frac{1}{2}\sum_{i=1}^{s} t_i^2 - \frac{1}{2}\sum_{i=1}^{s} t_i \ge \frac{1}{2s}\left(\sum_{i=1}^{s} t_i\right)^2 - \frac{1}{2}\sum_{i=1}^{s} t_i.$$

Hence we obtain, since $\sum_{i=1}^{s} t_i \ge r - z > 2s$,

$$\sum_{i=1}^{s} \binom{t_i}{2} \ge \frac{1}{2s}(r-z)^2 - \frac{1}{2}(r-z) = \frac{1}{2s}(r-z)(r-z-s).$$

Thus the bound (26) can be replaced by $\frac{1}{2s}(r-z)(r-z-s)$ where $s = c_1 X/\log(X/z)$.

Another result which can be derived by sieve methods and is used in the proofs of (13) is the Brun-Titchmarsh inequality. Let $\pi(x; k, l)$ denote the number of primes $\le x$ which are $\equiv l \pmod{k}$. Then, for sufficiently large x,

$$\pi(x; k, l) \le \frac{3x}{\phi(k)\log(x/k)} \qquad \text{(cf. [13, p. 110])} \qquad (27)$$

where $\phi(k)$ denotes Euler's indicator function.

7. An Application of Evertse's Bound for the Number of Solutions of the Equation $Ax^l - By^l = dz$

In 1939 Erdős [5] also applied the following special case of Thue's Theorem to prove that for $l > 2$ and $k \ge k_0(l)$ the equation

$$x(x+1)\cdots(x+k-1) = y^l$$

is impossible.

The number of solutions of $Ax^l - By^l = C$, where $l > 2$ and A, B, C are given positive integers, is finite.

Evertse [12], cf. [11], has calculated bounds for the number of solutions of $Ax^l - By^l = C$. Let $R(l, C)$ denote the number of residue classes $Z \pmod{C}$ with $Z^l \equiv 1 \pmod{C}$. Evertse proved that the number of solutions of $|Ax^l - By^l| = C$ in positive integers x, y with $\gcd(x, y) = 1$ is bounded by

$\alpha R(l,C)+\beta$ where $(\alpha,\beta)=(2,4)$ if $l=3$, $(1,3)$ if $l=4$, $(1,2)$ if $l=5$, $(1,1)$ if $l\geq 6$. It turns out that the following estimates, proved similarly, are even more useful for our purpose.

Lemma 6 (Evertse, [12, pp. 17–18]). *The number of solutions of*

$$Ax^l - By^l = dz \quad (A,B,d\in\mathbb{Z}_{>0}, l\in\mathbb{Z}_{>2}, \gcd(A,d)=\gcd(B,d)=1) \quad (28)$$

in integers $x>0$*,* $y>0$*,* z *subject to* $\gcd(x,y)=1$ *and* $0<|z|<d^{(2l/5)-1}$ *is bounded by* $\alpha R(l,d)+\beta$ *where*

$$(\alpha,\beta)=\begin{cases}(3,6)\ \mathit{if}\ l=3\\(2,2)\ \mathit{if}\ l=4\\(2,1)\ \mathit{if}\ l=5,6\\(1,3)\ \mathit{if}\ l=7\\(1,2)\ \mathit{if}\ i\geq 8.\end{cases}$$

It follows from the data at p. 18 of [12] that $R(l,d)=l^{w(d_1)}$ when l is an odd prime, where d_1 is the greatest divisor of d composed of primes $\equiv 1 \pmod{l}$.

We shall demonstrate how Lemma 6 can be applied to (12) We are going to prove (13.d) in the sharpened form

$$l^{w(d_1)}\geq\frac{C_9k}{\log k}\quad (l\geq 5).$$

Without loss we may therefore assume that $l^{w(d_1)}<\frac{.1k}{\log k}$. If $A_i=A_j$, for some $i>j$, then

$$A_j(y_i^l-y_j^l)=(i-j)d.$$

Since $\gcd(A_j,d)=1$, we obtain $A_j|i-j$. Hence

$$y_i^l-y_j^l=zd$$

where $0<z<k\ll d^{1/(l-2)}$, the latter inequality by Theorem 3. Since y_i and y_j are composed of primes $>k$ and are coprime to d, we have $\gcd(y_i,y_j)=1$. Furthermore, for k sufficiently large and $l\geq 3$, $|z|<d^{2l/5}-1$. On applying Lemma 6 we find that the number of pairs (i,j) with $i>j$ and $A_i=A_j$ is at most $\alpha l^{w(d_1)}+\beta$ which is less than $.9k/\log k$ by our supposition.

The argument given in step (c) of Sect. 2 applies to the A_j's. Hence there exists a subset T of $\{0,1,\ldots,k-1\}$ consisting of at least $k-\frac{.9k}{\log k}-\pi(k)$ elements A_j such that $\prod_{j\in T}A_j<k!$. By the Prime number theorem we have $\pi(k)<1.1k/\log k$ for large k. Hence we can apply Lemma 3 with $g=2$ and $\eta=\frac{1}{3}$. As in Step (d) of Sect. 5 we find a set S_2 of at least $k/3$ elements A_i satisfying

$$A_i\leq 60k \quad \text{for} \quad A_i\in S_2$$

for k large. Note that

$$A_i y_i^l - A_j y_j^l = (x+id) - (x+jd) = d(i-j)$$

so that $(y_i, y_j, i-j)$ is a solution of (28) for $A = A_i$, $B = A_j$, $z = i - j < k$. The number of possible pairs A_i, A_j is too large to make a straightforward application of the Box principle to apply Lemma 6. However, we have Lemma 5 at our disposal. We apply this lemma with $r = k/3$, $z = \varepsilon k$, $X = 60k$ and $\{a_j\}_{j=1}^r$ is the ordered set of A_i with $A_i \leq 60k$, where $\varepsilon > 0$ is so small that $60/\log(60/\varepsilon) < \frac{1}{20}$. We obtain that for k sufficiently large there are at least $k/4$ pairs A_i, A_j for which $\gcd(A_i, A_j) > \varepsilon k$. Since $\gcd(A_i, d) = \gcd(A_j, d) = 1$, we can divide by $\gcd(A_i, A_j)$ to arrive at an equation

$$B_i y_i^l - B_j y_j^l = dz_{ij}$$

where B_i, B_j are positive integers bounded by $60/\varepsilon$ and $|Z_{ij}| \leq 1/\varepsilon$. Hence we have at most $(60/\varepsilon)^2$ equations

$$Ax^l - By^l = dz$$

with in total at least $k/4$ solutions (x, y, z). By the Box principle at least one of the equations has at least $(\varepsilon/120)^2 k$ solutions. We know from (13.a) that d is sufficiently large, whence $[z] \leq d^{2l/5-1}$. We conclude from Lemma 6 that

$$\left(\frac{\varepsilon}{120}\right)^2 \frac{k}{\log k} \leq \alpha l^{w(d_1)} + \beta.$$

In this way we derive the following improvement of (13.d):

Theorem 4. *Under the conditions of Theorem 3 we have*

$$l^{w(d_1)} \gg \frac{k}{\log k} \qquad (l \geq 5). \tag{29}$$

For a slightly stronger form of (29) and for a proof not depending on Evertse's result, we refer to [27, Corollary 1].

8. Perfect Powers in Products of the Terms of an Arithmetic Progression, III

In Sect. 5 we have shown how (13.a) can be proved and in Sect. 7 we have derived a refinement of (13.d). We first show how these results can be used to obtain formulas of the form (13.b) and (13.c). Afterwards we say something on the remaining case $l = 2$.

By Theorem 4 we have $l^{w(d_1)} \gg k$. Since every prime factor of d_1 is $\equiv 1 \pmod{l}$, the greatest prime factor $P(d_1)$ of d_1 has to be at least $lw(d_1)$. By using the Brun-Titchmarsh theorem we even find that

$$P(d_1) \gg lw(d_1) \log w(d_1).$$

If $l > (\log k)^2$, then (13.c) holds trivially. If $l \le (\log k)^2$, then

$$P(d_1) \gg l \frac{\log k}{\log l} \log \frac{\log k}{\log l} \gg \log k \log\log k. \qquad (13.c)$$

If $l \ge \log\log k$, then (13.b) follows from (13.a). If $l < \log\log k$, then Theorem 4 implies that, for some constant c,

$$d_1 \ge (cw(d_1))^{w(d_1)}$$

whence

$$\log d_1 \gg \frac{\log k}{\log l} \log\log k \gg \frac{\log k \log\log k}{\log\log\log k}$$

which is slightly weaker than (13.b).

The proofs given until now do not include the case $l = 2$. The proofs for $l = 2$ start in a different way, namely by writing

$$x + id = a_i x_i^2 \qquad (i = 0, 1, \ldots, k-1) \tag{30}$$

where a_i is squarefree (cf. Sect. 2). Hence x_i can contain prime factors $< k$, but also in (30) we have $P(a_i) < k$. It is now much harder to show that the set $\{a_i\}_{i=0}^{k-1}$ has at least $k - ck/\log k$ elements. Because of the fact that the a_i's are squarefree, the product of the a_i's is essentially larger than $[k - ck/\log k]!$ and this lower bound contradicts an upper bound, due to Erdős and Rigge, obtained by saving the powers of 2 and 3 in the product. This contradiction yields the results (13.b), (13.c), and (13.d) for $l = 2$. For details see [27, Sect. 3].

Looking once again at Erdős' conjecture stated in the Abstract, we see that for difference $d > 1$ all results up to now are for length k sufficiently large. It follows from (13.a) and (13.b) that k should be much smaller than d. As long as we cannot disprove that there are arbitrarily large arithmetic progressions consisting of l-th powers there is no hope to prove Erdős' conjecture. However, there is new hope to settle the latter assertion after Wiles' announcement of a proof of Fermat's Last Theorem.

9. Generalisations and Extensions

It will be obvious from the preceding proofs that it is not necessary to require that all terms $x + id$ $(i = 0, 1, \ldots, k-1)$ are almost-powers $A_i y_i^l$. It suffices if there are enough almost-powers among them. As mentioned in Sect. 1 the question how many such numbers are required was first investigated by Erdős in 1955. Shorey [25] completed Erdős' result by showing that also for $l = 2$ the equation

$$(x + d_1)(x + d_2) \cdots (x + d_t) = y^l \tag{31}$$

in integers $x > 0$, $t > 0$, $l > 1$, $y > 1$, $0 \leq d_1 < d_2 < \cdots < d_t < k$ implies that

$$t \leq k - (1-\varepsilon)k\frac{\log\log k}{\log k} \tag{32}$$

for $k > k_0(\varepsilon), \varepsilon > 0$. The proof depends on the finiteness of integral solutions of hyper-elliptic equations under suitable conditions. Moreover, he sharpened Erdős' bound (32) for $l > 2$ considerably. Combining the elementary method of Erdős with the theory of linear forms in logarithms, irrationality measures of Baker for algebraic numbers and the method of Halberstam and Roth on l-free integers, Shorey [24, 25] proved that, for k sufficiently large and $l > 2, t \leq \nu_l k$ with

$$\nu_3 = \frac{47}{56}, \nu_4 = \frac{45}{64}, \nu_l < \frac{2}{3} \quad \text{for} \quad l \geq 5 \quad \text{and} \quad \nu_l \ll^{l-1/11} \quad \text{for } l \text{ sufficiently large.}$$

The authors [30] derived some results for the general equation

$$(x + d_1 d)(x + d_2 d) \cdots (x + d_t d) = by^l \tag{33}$$

where b and d are positive integers with $P(b) \leq k$ and gcd $(x, d) = 1$ and the other unknowns are as above. For example, they showed for any $\varepsilon > 0$ that (33) with $l = 2$ and $k \geq k_1(\varepsilon, w(d))$ implies

$$t \leq k - (1-\varepsilon)k\frac{\log\log\log k}{\log k}$$

and that (33) with prime $l > 2$ and $k \geq k_2(\varepsilon)$ and $l^{w(d)} \ll_\varepsilon kh(k)/\log k$ implies that

$$t \leq k - (1-\varepsilon)\frac{kh(k)}{\log k}$$

where

$$h(k) = \begin{cases} \log\log\log k & \text{if } l = 3 \\ \log\log k & \text{if } l \geq 5. \end{cases}$$

References

1. L.E. Dickson, *History of the Theory of Numbers*, Vol. II, Carnegie Institute, Washington, 1920. Reprint: Chelsea, New York, 1971.
2. P. Erdős, *A theorem of Sylvester and Schur,* J. London Math. Soc. **9** (1934), 282–288.
3. P. Erdős, *On sequences of integers no one of which divides the product of two others and on some related problems*, Mitt. Forsch. Inst. Math. Mech. Univ. Tomsk **2** (1938), 74–82.
4. P. Erdős, *Note on products of consecutive integers*, J. London Math. Soc. **14** (1939), 194–198.

5. P. Erdős, *Note on the product of consecutive integers (II)*, J. London Math. Soc. **14** (1939), 245–249.
6. P. Erdős, *On a diophantine problem*, J. London Math. Soc. **26** (1951), 176–178.
7. P.Erdős, *On the product of consecutive integers (III)*, Indag. Math. **17** (1955), 85–90.
8. P. Erdős, *On some applications of graph theory to number theoretic problems*, Publ. Ramanujan Inst. No. **1** (1968/69), 131–136.
9. P. Erdős & J.L. Selfridge, *The product of consecutive integers is never a power*, Illinois J. Math. **19** (1975), 292–301.
10. L. Euler, Mém. Acad. Sc. St. Pétersbourg 8, années 1817–1818 (1780), **3**; Comm. Arith. II, 411–413.
11. J.-H. Evertse, *On the equation $ax^n - by^n = c$*, Compos. Math. **47** (1982), 289–315.
12. J.-H. Evertse, *Upper Bounds for the Numbers of Solutions of Diophantine Equations*, Mathematical Centre Tracts **168**, Mathematisch Centrum, Amsterdam, 1983.
13. H. Halberstam & H.-E. Richert, *Sieve* Methods, Academic Press, 1974.
14. M. Langevin, *Plus grand facteur premier d'entiers en progression arithmétique*, Sém. Delange-Pisot-Poitou, 18e année, 1976–77, no. **3**,7 pp.
15. R. Marszalek, *On the product of consecutiv e elements of an arithmetic progression*, Monatsh. Math. **100** (1985), 215–222.
16. P. Moree, *Bertrand's postulate for primes in arithmetical progressions*, Computers Math. Applic. 26 (1993), 35–43.
17. S. S. Narumi, *An extension of a theorem of Liouville's*, Tôhoku Math. J. **11** (1917), 128–142.
18. R. Obláth, *Über Produkte aufeinanderfolgender Zohlen*, Tôhoku Math. J. **38** (1933), 73–92.
19. R. Obláth, *Über das Produki fünf aufeinander folgender Zahlen in einer arithmetischen Reihe*, Publ. Math. Debrecen **1** (1950), 222–226.
20. R. Obláth, *Eine Bemerkung über Produkie aufeinander folgender Zahlen*, J. Indian Math. Soc. (N.S.) **15** (1951), 135–139.
21. P. Ribenboim, *The Book of Prime Numb er Records*, Springer-Verlag, 1988.
22. O. Rigge, *Über ein diophantisches Problem*, 9th Congress Math. Scand., Helsingfors 1938, Mercator, 1939, pp. 155–160.
23. H. Rohrbach & J. Weis, *Zum finiten Fall des Bertrandschen Postulates*, J. reine angew. Math. **214/215** (1964), 432–440.
24. T. N. Shorey, *Perfect powers in values of certain polynomials at integer points*, Math. Proc. Cambridge Philos. Soc. **99** (1986), 195–207.
25. T. N. Shorey, *Perfect powers in products of integers from a block of consecutive integers*, Acta Arith. **49** (1987), 71–79.
26. T.N. Shorey, *Some exponential equations, in: New Advances in Transcendence Theory*, A. Baker (ed.), Cambridge University Press, 1988, pp. 217–229.
27. T.N. Shorey & R. Tijdeman, *Perfect powers in products of terms in an arithmetical progression*, Compos. Math. **75** (1990), 307–344.
28. T. N. Shorey & R. Tijdeman, *On the greatest prime factor of an arithmetical progression*, in: A Tribute to Paul Erdős, A. Baker, B. Bollobás & A. Hajnal (eds.), Cambridge University Press, 1990, pp. 385–389.
29. T. N. Shorey & R. Tijdeman, *Perfect powers in products of terms in an arithmetical progression (II)*, Compos. Math. **82** (1992), 119–136.
30. T. N. Shorey & R. Tijdeman, *Perfect powers in products of terms in an arithmetical progression (III)*, Acta Arith. **61** (1992), 391–398.

31. T. N. Shorey & R. Tijdeman, *On the greatest prime factors of an arithmetical progression (III)*, In: Diophantine Approximations and Transcendental Numbers, Luminy 1990, Walter de Gruyter, Berlin etc., 1992, pp. 275–280.
32. J.J. Sylvester, *On arithmetical series*, Messenger Math. **21** (1892), 1–19, 87–120. Collected Mathematical Papers **4** (1912), 687–731.
33. R. Tijdeman, *Diophantine equations and diophantine approximations*, in: Number Theory and Applications, R.A. Mallin (ed.), Kluwer Academic Press, 1989, pp. 215–243.

Sur la non-dérivabilité de fonctions périodiques associées à certaines formules sommatoires

Gérald Tenenbaum

G. Tenenbaum (✉)
Institut Élie Cartan, Université de Lorraine, BP 70239,
54506 Vandœuvre-lès-Nancy Cedex, France
e-mail: gerald.tenenbaum@uni-lorraine.fr

1. Introduction

Les fonctions arithmétiques associées aux systèmes de représentations d'entiers, comme le développement dans une base donnée, satisfont généralement des relations de récurrence qui facilitent considérablement l'étude de leur valeur moyenne. Considérons par exemple la somme des chiffres en base 2, que nous désignons par $\sigma(n)$. On a

$$\sigma(2n) = \sigma(n), \qquad \sigma(2n+1) = \sigma(n) + 1 \qquad (n \geqslant 1), \tag{1}$$

d'où il découle que la fonction sommatoire $S(n) := \sum_{0 \leqslant m < n} \sigma(m)$ satisfait à

$$S(2n) = n + 2S(n), \; S(2n+1) = n + \sigma(n) + 2S(n) \quad (n \geqslant 0). \tag{2}$$

En particulier, si l'on pose $\varphi(n) := S(n) - (n \log\ n)/(2\log 2)$ $(n \geqslant 1)$, la première relation (2) implique $\varphi(2n) = 2\varphi(n)$ pour tout n, de sorte que l'on peut écrire

$$\varphi(n) = nG\left(\frac{\log n}{\log 2}\right), \qquad S(n) = \frac{n \log n}{2\log 2} + nG\left(\frac{\log n}{\log 2}\right), \tag{3}$$

où G est périodique de période 1.

Dans la quasi-totalité des exemples connus, on obtient de même, sans trop de difficulté, une formule du type

$$\sum_{m<n} a_m = P(n) + Q(n)G(\log n) + R(n) \tag{4}$$

pour une fonction arithmétique donnée $\{a_m\}_{m=0}^{\infty}$, où P et Q sont des fonctions régulières — typiquement des combinaisons linéaires de produits de puissances n^{α} et de puissances de logarithmes $(\log n)^{\beta}$—, $R(n)$ est un terme résiduel, souvent périodique et/ou borné, et G est une fonction oscillante périodique à caractère fractal, en général continue.

Trollope [41] a donné, pour la fonction G de la formule (3), une formule explicite impliquant en particulier qu'elle est continue, et partant bornée. Utilisant une méthode plus simple, Delange [10] a généralisé le résultat au cas

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_19,

de la somme des chiffres en base $q \geqslant 2$ quelconque, que nous notons $\sigma_q(n)$, soit

$$\sum_{m<n} \sigma_q(m) = \frac{q-1}{2\log q} n \log n + nG_q\left(\frac{\log n}{\log q}\right), \tag{5}$$

où G_q est continue et 1-périodique. Il montre en outre que G_q n'est nulle part dérivable et détermine son développement de Fourier.

À côté de celles de Trollope et Delange, plusieurs autres techniques sont en fait susceptibles de fournir le calcul explicite de G_q. Nous utilisons au paragraphe 3 une approche assez générale fondée sur l'intégration complexe. Voyons ici, par exemple, comment fonctionne, dans le cas $q = 2$, celle de Brillhart, Erdős et Morton dans [2]. Soit x un nombre réel positif, dont le développement en base 2 est

$$x = \sum_{r \geqslant 0} \varepsilon_r / 2^r,$$

avec $\varepsilon_0 \in \mathbb{N}, \varepsilon_r = 0$ ou 1 pour $r \geqslant 1$, et $\varepsilon_r \neq 1$ pour une infinité de valeurs de r. On pose

$$x_k := \lfloor 2^k x \rfloor = \sum_{0 \leqslant r \leqslant k} \varepsilon_r 2^{k-r}, \qquad \text{et} \qquad T_k := \frac{S(x_k)}{2^k} - \frac{x \log(2^k x)}{2 \log 2}.$$

On a $x_k = 2x_{k-1} + \varepsilon_k$. Grâce à (2), on en déduit par un calcul de routine que

$$T_k - T_{k-1} = \frac{\varepsilon_k \sigma(x_{k-1})}{2^k} - \sum_{r \geqslant k} \frac{\varepsilon_r}{2^{r+1}}, \tag{6}$$

ce qui implique l'existence de $\varphi(x) := \lim_{k\to\infty} T_k$. De plus, par itération puis passage à la limite en k, la relation (6) fournit

$$\varphi(x) - T_0 = \sum_{r \geqslant 1} \frac{\varepsilon_r}{2^r} \{\sigma(x_{r-1}) - \tfrac{1}{2} r\},$$

d'où

$$\varphi(x) = S\big(\lfloor x \rfloor\big) - \frac{x \log x}{2 \log 2} + \sum_{r \geqslant 1} \frac{\varepsilon_r}{2^r} \{\sigma(x_{r-1}) - \tfrac{1}{2} r\} \tag{7}$$

Lorsque $x = n \in \mathbb{N}$, on a $\varepsilon_0 = n, \varepsilon_r = 0 \ \ (r \geqslant 1)$ et l'on retrouve bien (3). De plus, lorsque ξ est un rationnel dyadique positif de dénominateur réduit $2^m \, (m \geqslant 0)$, on a

$$\lim_{x \to \xi^-} \sigma(x_r) = \begin{cases} \sigma(\xi_r) & (r < m) \\ \sigma(\xi_m - 1) + r - m & (r \geqslant m), \end{cases}$$

$$\lim_{x \to \xi^-} \varepsilon_r(x) = \begin{cases} \varepsilon_r(\xi) & (r < m) \\ \varepsilon_r(\xi) - 1 & (r = m). \\ 1 & (r > m) \end{cases}$$

Cela permet de vérifier facilement que φ est continue sur $\mathbb{R}^+$. La fonction G de (3) peut donc être prolongée en une fonction continue sur $\mathbb{R}^+$ par la formule

$$G(u) := \varphi(2^u)/2^u.$$

Puisque les quantités $(\log n)/\log 2$ sont denses modulo 1, la propriété de périodicité observée plus haut est encore valable pour le prolongement.

La littérature abonde en exemples de situations similaires — ainsi qu'on pourra s'en convaincre à la lecture de notre bibliographie, issue de celle rassemblée dans la thèse de Cateland [5]. Désignons, conformément à l'usage, par *q-noyau* d'une suite $\{a_n\}_{n=0}^{\infty}$ l'ensemble des sous-suites

$$\{n \mapsto a_{q^k n+r} : k \geqslant 0,\, 0 \leqslant r < q^k\}.$$

La généralisation naturelle de la propriété (1) est celle des suites q-automatiques, i.e. dont le q-noyau est fini, voire des suites q-régulières, c'est-à-dire dont le q-noyau engendre un module de type fini — cf. [1]. La quasi-totalité des exemples connus relève effectivement de ces deux définitions.

Dans cette note, nous nous intéressons plus particulièrement à la non-dérivabilité des fonctions fractales G apparaissant dans des formules de type (4). Peu de résultats généraux sont disponibles dans cette direction. Les travaux les plus significatifs sont ceux de Dumont-Thomas [11–13] et Cateland [5]. On distingue essentiellement trois types de méthodes : d'une part celles qui exploitent l'expression exacte de $G(x)$, généralement sous la forme d'une série liée à la représentation de x dans un système adéquat, d'autre part celles qui établissent l'existence d'équations fonctionnelles pour G (c'est en particulier la voie explorée par Dumont et Thomas), enfin celles qui utilisent les divers renseignements disponibles sur les coefficients de Fourier de G — ce qui ne fournit en général qu'une preuve de la non-dérivabilité presque-partout, et non partout. Nous nous proposons de développer ici une quatrième approche, sans doute la plus naïve de toutes. Elle consiste à "oublier" la définition explicite de G pour ne retenir que la formule (4): comme Ie membre de gauche est arithmétique, donc irrégulier, il est naturel d'attendre que le membre de droite contienne lui aussi un certain degré d'irrégularité — qui doit alors être nécessairement le fait du terme fractal $G(\log n)$. Il reste ensuite à opérer un "relèvement" des valeurs de la variable, en transportant les propriétés des $G(\log n)$ aux $G(x)$ où x est un nombre réel quelconque. Il est vraisemblable que ce principe puisse être formalisé dans un contexte assez général. Nous nous contentons ici de le mettre en œuvre dans trois cas particuliers importants de la littérature.

Le premier exemple est celui de la suite de Newman-Coquet

$$c_n = (-1)^{\sigma(3n)}. \tag{8}$$

Coquet [7] établit la formule sommatoire

$$\sum_{m<n} c_m = n^\vartheta G_0\left(\frac{\log n}{\log 4}\right) + \tfrac{1}{3}\eta(n), \tag{9}$$

où G_0 est continue et l-périodique, et où l'on a posé

$$\vartheta := \frac{\log 3}{\log 4}, \qquad \eta(n) := \begin{cases} 0 & \text{si } n \text{ est pair,} \\ (-1)^{\sigma(3n-3)} & \text{si } n \text{ est impair.} \end{cases}$$

Nous montrons, directement à l'aide de (9) et sans utiliser l'expression de G_0, le résultat suivant, qui est d'ailleurs implicitement contenu dans la preuve de Coquet de la non-dérivabilité de G_0.

Théorème 1. *La fonction G_0 n'est dérivable pour aucune valeur de $x \in \mathbb{R}$. Plus précisément, on a pour tout $x \in \mathbb{R}$*

$$G_0(x+h) - G_0(x) = \Omega(|h|^\vartheta) \qquad (h \to 0). \tag{10}$$

Nous considérons ensuite la suite de Rudin-Shapiro

$$r_n := (-1)^{e(n)}, \qquad \text{avec } e(n) := \sum_{j\geqslant 0} \varepsilon_j\varepsilon_{j+1} \quad \text{si} \quad n = \sum_{j\geqslant 0} \varepsilon_j 2^j.$$

Brillhart, Erdős et Morton établissent dans [2] la formule sommatoire

$$\sum_{m<n} r_m = \sqrt{n} G_1\left(\frac{\log n}{\log 4}\right), \tag{11}$$

où G_1 est 1-périodique, bornée, et continue sauf aux points $(\log n)/\log 4$, $n \in \mathbb{N}^*$, et prouvent que G_1 n'est dérivable en aucun point x normal en base 4. Dans [12], Dumont et Thomas montrent que cette dernière restriction est inutile. Notre approche directe fonctionne ici très simplement et fournit le résultat suivant.

Théorème 2. *La fonction G_1 n'est dérivable pour aucune valeur de $x \in \mathbb{R}$. Plus précisément, on a, pour tout $x \in \mathbb{R}$,*

$$G_1(x+h) - G_1(x) = \Omega\big(\sqrt{|h|}\big) \qquad (h \to 0). \tag{12}$$

La troisième application concerne les *suites digitales*, introduites par Cateland [5], et qui sont une généralisation de la somme des chiffres en base q. Pour $q \geqslant 2, \ell \geqslant 1$, notons $E(q,\ell)$ l'ensemble des ℓ-uples $\varepsilon = (\varepsilon_0, \ldots, \varepsilon_{\ell-1})$ avec $\varepsilon_j \in \{0, \ldots, q-1\}$ pour tout j. Pour $n \in \mathbb{N}$, on écrit $n = \sum_{j=0}^\infty \varepsilon_j(n)q^j$ le développement de n en base q et l'on pose

$$\varepsilon_k(n) := (\varepsilon_k(n), \ldots, \varepsilon_{k+\ell-1}(n)) \in E(q,\ell) \qquad (k = 0, 1, \ldots).$$

On note encore

$$\chi_k(n;\varepsilon) := \begin{cases} 1 \text{ si } \varepsilon_k(n) = \varepsilon \\ 0 \text{ si } \varepsilon_k(n) \neq \varepsilon \end{cases} \qquad \big(\varepsilon \in E(q,\ell)\big), \quad \varrho(n;\varepsilon) := \sum_{k\geqslant 0} \chi_k(n;\varepsilon),$$

avec la convention $\varrho(n;0) = 1$ pour tout n. La fonction $\varrho(n;\varepsilon)$ est donc égale au nombre d'occurrences du mot ε dans la représentation q-adique de n. Étant donnée une fonction $F : E(q,\ell) \to \mathbb{C}$ telle que $F(0) = 0$, on définit une suite digitale $\{u_F(n)\}_{n=0}^{\infty}$ par la formule

$$u_F(n) = \sum_{k\geqslant 0} F(\varepsilon_k(n)) = \sum_{\varepsilon\in E(q,\ell)} F(\varepsilon)\varrho(n;\varepsilon). \tag{13}$$

On retrouve la suite $\sigma_q(n)$ en choisissant $\ell = 1$ et F égale à l'identité. Cateland a établi, par la méthode de Delange, la formule sommatoire générale

$$\sum_{m<n} u_F(m) = A_F n\log n + nG_F\left(\frac{\log n}{\log q}\right) + \delta_F(n) \tag{14}$$

avec

$$A_F := \frac{1}{q^\ell \log q} \sum_{\varepsilon\in E(q,\ell)} F(\varepsilon),$$

et où G_F est continue et 1-périodique, et δ_F est $q^{\ell-1}$-périodique. Nous donnons une preuve assez simple de ce résultat au paragraphe 3. Notre objectif principal consiste à déduire de (14) le résultat suivant de non-dérivabilité, qui étend optimalement celui de Cateland. Comme nous le verrons, la démonstration, reposant sur le principe énoncé plus haut, est extrêmement simple.

Théorème 3. *Soient $q \geqslant 2$, $\ell \geqslant 1$, $F : E(q,\ell) \to \mathbb{C}$, et $\{u_F(n)\}_{n=0}^{\infty}$ la suite digitale correspondante. Une condition ncessaire et suffisante pour que la fonction 1-périodique G_F associée soit nulle part dérivable est qu'il existe un entier $a \geqslant 1$ tel que $u_F(q^{\ell-1}a) \neq 0$.*

Le résultat de Cateland était conditionnel à l'hypothèse $A_F \neq 0$. Lorsque $u_F(q^{\ell-1}a) = 0$ $(a \geqslant 1), u_F$ est périodique, $A_F = 0$, et la fonction G_F est constante.

L'auteur tient ici à remercier Jean-Paul Allouche pour son aide précieuse lors de la préparation de cet article.

2. Démonstration des Théorèmes 1, 2 et 3

Prouvons d'abord les Théorèmes 1 et 2. Soit $x \in [0,1[$. Nous écrivons le dveloppement 4-adique de 4^x, soit

$$4^x = \sum_{j\geqslant 0} \varepsilon_j/4^j,$$

avec $0 \leqslant \varepsilon_j \leqslant 3$ pour tout j et $\varepsilon_j \neq 3$ pour une infinité d'indices j. Ensuite, nous définissons, pour $k \geqslant 0$, les nombres réels x_k, y_k et l'entier n_k par les formules

$$n_k = 4^{x_k+k} = \sum_{0\leqslant j\leqslant k} \varepsilon_j 4^{k-j}, \qquad n_k+1 = 4^{y_k+k}. \tag{15}$$

En écrivant (9) pour $n = n_k$ et $n = n_k + 1$ et en effectuant la différence, il vient

$$c_{n_k} = (n_k+1)^\vartheta G_0(y_k) - n_k^\vartheta G_0(x_k) + \tfrac{1}{3}(\eta(n_k+1) - \eta(n_k)),$$

d'où

$$n_k^\vartheta\{G_0(y_k) - G_0(x_k)\} - c_{n_k} - \tfrac{1}{3}\{\eta(n_k+1) - \eta(n_k)\} + O\big(n_k^{\vartheta-1}\big). \tag{16}$$

Compte tenu de la définition de $\eta(n)$, il est clair que le second membre est de valeur absolue $\gg 1$. Par ailleurs, il découle immdiatement de (15) que

$$n_k \asymp 4^k \asymp (y_k - x_k)^{-1}. \tag{17}$$

Il suit

$$|G_0(y_k) - G_0(x_k)| \gg (y_k - x_k)^\vartheta.$$

Comme

$$\max\{|x - x_k|, |x - y_k|\} \ll 4^{-k}, \tag{18}$$

cela contredit

$$G_0(x+h) - G_0(x) = o(|h|^\vartheta) \qquad (h \to 0),$$

et partant implique la conclusion requise (10) du Théorème 1.

La situation est encore plus simple pour le Théorème 2. On obtient parallèlement à (16)

$$\sqrt{n_k}\{G_1(y_k) - G_1(x_k)\} = r_{n_k} + O\big(1/\sqrt{n_k}\big), \tag{19}$$

d'où par (17), puisque $|r_{n_k}| = 1$,

$$|G_1(y_k) - G_1(x_k)| \gg \sqrt{y_k - x_k}.$$

Grâce à (18), cela implique (12) et établit ainsi le Théorème 2.

La même approche fonctionne encore pour établir le Théorème 3. L'hypothèse $F \neq 0$ implique $u_F \neq 0$, et, plus précisément, implique l'existence d'un entier $a, 1 \leqslant a < q^\ell$, tel que $u_F(a) \neq 0$. En effet, notant $F^*(h) := F(\varepsilon_0, \ldots, \varepsilon_{j-1})$ pour $h = \sum_{r=0}^{\ell-1} \varepsilon_r q^r$, on a

$$u_F(j) = \sum_{0\leqslant h<q^\ell} \alpha_{jh} F^*(h) \quad (1 \leqslant j < q^\ell)$$

avec $\alpha_{jh} \geqslant 1$ si j est de la forme $j = a + q^s h$ avec $s \geqslant 0$, $a < q^s$, et $\alpha_{jh} = 0$ dans le cas contraire. En particulier, on a $\alpha_{jj} = 1$ pour $1 \leqslant j < q^\ell$ et $\alpha_{jh} = 0$ si $h > j$. La matrice carrée (α_{jh}) est donc triangulaire supérieure, avec des 1 sur la diagonale principale. Par conséquent, elle est inversible et cela établit la propriété indiquée.

Cependant, lorsque $\ell \geqslant 2$, on peut avoir $u_F(q^{\ell-1}a) = 0$ pour tout $a \geqslant 1$ sans que u_F soit identiquement nulle: pour $\ell = q = 2$ et $F(1,0) = -F(0,1) = 1, F(1,1) = 0, u_F$ est la fonction indicatrice des nombres impairs. On peut vérifier facilement que, dans un tel cas, u_F est périodique, $A_F = 0$, et la fonction G_F est constante.

Soit alors $a \geqslant 1$. Nous allons montrer que si G_F est dérivable en un point $x \in [0,1[$ alors $u_F(q^{\ell-1}a) = 0$. On écrit les développements q-adiques

$$a = \sum_{0 \leqslant r \leqslant m(a)} \varepsilon_r(a) q^r, \qquad q^x = \sum_{j \geqslant 0} \varepsilon_j / q^j,$$

et l'on pose

$$L := 2\ell + m(a), \qquad n_k = q^{x_k+L+k} = q^L \sum_{0 \leqslant j \leqslant k} \varepsilon_j q^{k-j}, \qquad n_k + 1 = q^{y_k+L+k}. \tag{20}$$

On a

$$n_k \asymp q^k \asymp (y_k - x_k)^{-1}, \qquad \max\{|x - x_k|, |x - y_k|\} \ll q^{-k}. \tag{21}$$

Ici et dans la suite de cette démonstration les constantes implicites peuvent dépendre de a ou F mais pas de k.

En appliquant (14) avec n et $n+1$ et en faisant la différence, on obtient, lorsque $n \equiv 0 \,(\operatorname{mod} q^{\ell-1})$,

$$\begin{aligned} u_F(n) = A_F \log n + n \left\{ G_F\left(\frac{\log(n+1)}{\log q}\right) - G_F\left(\frac{\log n}{\log q}\right) \right\} \\ + A_F(n+1)\log(1+1/n) + G_F\left(\frac{\log(n+1)}{\log q}\right) + \delta_F(1) - \delta_F(0), \end{aligned} \tag{22}$$

où l'on a tenu compte de la périodicité de δ_F. Substituons $n = n_k$ dans cette relation. La périodicité de G_F nous permet de remplacer $G_F(\log(n+1)/\log q)$ par $G_F(y_k)$ et $G_F(\log n/\log q)$ par $G_F(x_k)$, soit

$$\begin{aligned} u_F(n) = A_F \log n + n\{G_F(y_k) - G_F(x_k)\} \\ + A_F(n+1)\log(1+1/n) + G_F(y_k) + \delta_F(1) - \delta_F(0). \end{aligned} \tag{23}$$

Si G_F est dérivable au point x, on a, lorsque $k \to \infty$,

$$G_F(x_k) - G_F(x) = (x_k - x)G_F'(x) + o(x_k - x) = (x_k - x)G_F'(x) + o(1/n_k),$$

et similairement

$$G_F(y_k) - G_F(x) = (y_k - x)G_F'(x) + o(1/n_k).$$

Il suit

$$n_k\{G_F(y_k) - G_F(x_k)\} = n_k(y_k - x_k)G_F'(x) + o(1) = \frac{G_F'(x)}{\log q} + o(1).$$

En reportant dans (23), on obtient

$$u_F(n_k) = A_F \log n_k + A_F + \frac{G'_F(x)}{\log q} + G(x) + \delta_F(1) - \delta_F(0) + o(1),$$

et donc

$$u_F(n_k) = kA_F \log q + B + o(1), \tag{24}$$

avec $B := A_F\{1 + (x+L)\log q\} + G'_F(x)/\log q + G_F(x) + \delta_F(1) - \delta_F(0)$.

Substituons maintenant $n = n_k + q^{\ell-1}a$ dans (22). Les calculs qui précédent restent valables *mutatis mutandis*, et l'on obtient

$$u_F(n_k + q^{\ell-1}a) = kA_F \log q + B + 0(1). \tag{25}$$

Or il découle immédiatement des définitions de u_F et L que

$$u_F(n_k + q^{\ell-1}a) = u_F(n_k) + u_F(q^{\ell-1}a).$$

Les relations (24) et (25) impliquent donc par différence

$$u_F(q^{\ell-1}a) = o(1), \qquad \text{c'est-à-dire} \quad u_F(q^{\ell-1}a) = 0.$$

Cela termine la démonstration du Théorème 3.

3. Preuve de la formule de Cateland par intégration complexe

Nous nous proposons ici de donner une démonstration de la formule (14) en utilisant la formule de Perron. La démarche, semblable à celle de Flajolet et al. dans [15], possède le double avantage de ne nécessiter que quelques calculs assez simples et de fournir directement les développements de Fourier des fonctions G_F et δ_F.

Au vu de (13), nous pouvons nous restreindre à estimer la valeur moyenne de $\varrho(n, \varepsilon)$ pour $\varepsilon = (\varepsilon_0, \ldots, \varepsilon_{\ell-1}) \in E(q, \ell)$ fixé. Posons $h := \sum_{j=0}^{\ell-1} \varepsilon_j q^j$ et

$$V(n) := \sum_{0 \leqslant m < n} \varrho(m; \varepsilon) = \sum_{k \geqslant 0} \sum_{0 \leqslant m < n} \chi_k(m; \varepsilon). \tag{26}$$

La somme intérieure est égale au nombre des entiers $m \in [0, n-1]$ qui sont de la forme $m = a + q^k h + q^{k+\ell} b$ avec $0 \leqslant a < q^k, b \geqslant 0$. Elle vaut donc

$$\sum_{0 \leqslant a < q^k} \left(1 + \left\lfloor \frac{n - (hq^k + a + 1)}{q^{k+\ell}} \right\rfloor\right) = \sum_{0 \leqslant a < q^k} \int_a^{a+1} \left(1 + \left\lfloor \frac{n - (hq^k + t)}{q^{k+\ell}} \right\rfloor\right) \mathrm{d}t$$

$$= \int_0^{q^k} \left(1 + \left\lfloor \frac{n - (hq^k + t)}{q^{k+\ell}} \right\rfloor\right) \mathrm{d}t = q^k \int_h^{h+1} \left(1 + \left[\frac{n}{q^{k+\ell}} - \frac{t}{q^\ell} \right]\right) \mathrm{d}t.$$

Pour établir la première égalité, nous avons utilisé le fait que l'intégrande du second membre est constant sur chaque intervalle $]a, a+1]$.

Pour évaluer la partie entière de la dernière intégrale, nous introduisons la fonction zêta de Hurwitz, définie, pour chaque valeur du paramètre $\alpha \in]0,1]$, par la formule

$$\zeta(s;\alpha) := \sum_{n\geqslant 0}(n+\alpha)^{-s} \qquad \big(\Re e s > 1\big),$$

et prolongée en une fonction méromorphe dans le plan complexe tout entier ayant pour unique singularité un pôle simple en $s=1$, de résidu 1. On a pour $x>0$

$$1+\lfloor x-\alpha\rfloor = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\zeta(s;\alpha)x^s\frac{\mathrm{d}s}{s} \qquad (c>1),$$

sauf si $x\in\alpha+\mathbb{Z}$, où le membre de droite vaut $\lfloor x-\alpha\rfloor+\frac{1}{2}$. On obtient donc

$$\begin{aligned}V(n) &= \sum_{k\geqslant 0} q^k\int_h^{h+1}\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty}\zeta(s;t/q^\ell)\left(\frac{n}{q^{k+\ell}}\right)^s\frac{\mathrm{d}s\,\mathrm{d}t}{s}\\ &= \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty}\left(\frac{n}{q^\ell}\right)^s\frac{1}{s(1-q^{1-s})}\int_h^{h+1}\zeta(s;t/q^\ell)\ \mathrm{d}t\mathrm{d}s\\ &= \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty}\left(\frac{n}{q^\ell}\right)^s\frac{q^\ell Z(s-1;h/q^\ell)}{s(1-s)(1-q^{1-s})}\mathrm{d}s,\end{aligned}$$

où l'on a posé

$$Z(s;t) := \zeta(s;t+1q^\ell)-\zeta(s;t) = s\int_0^{1/q^\ell}\zeta(s+1;t+u)\mathrm{d}u.$$

Déplaçons maintenant l'abscisse d'intégration vers la gauche jusqu'à l'axe $\Re e\, s=\frac{1}{2}$. La contribution du pôle double en $s=1$ vaut

$$\frac{n}{\log q}\left\{Z(0;h/q^\ell)\left(\log(n/q^\ell)-1+\tfrac{1}{2}\log q\right)+Z'(0,h/q^\ell)\right\}.$$

La contribution des pôles simples $p_k := 1+2\pi ki/\log q$ $(k\neq 0)$ est égale à

$$n\sum_{k\in\mathbb{Z}\smallsetminus\{0\}}\frac{Z(p_k-1;h/q^\ell)}{p_k(1-p_k)\log q}\,\mathrm{e}\left(k\frac{\log n}{\log q}\right) = ng_h\left(\frac{\log n}{\log q}\right) \qquad \text{(disons)},$$

avec la notation traditionnelle $\mathrm{e}(t) := \exp\{2\pi it\}$. La fonction g_h est 1-périodique, et sa série de Fourier, explicitée ci-dessus, est absolument convergente. En particulier, g_h est continue.

On a

$$Z(0;h/q^\ell)=q^{-\ell}, \quad \zeta'(0;\alpha)=\log\left(\frac{\Gamma(\alpha)}{\sqrt{2\pi}}\right), \quad Z'(0;\alpha)=\log\Big(\Gamma(\alpha+q^{-\ell})/\Gamma(\alpha)\Big).$$

Il suit

$$V(n)=\frac{n}{q^\ell\log q}\left\{\log n-1+\log\left(\frac{\Gamma((h+1)/q^\ell)}{\Gamma(h/q^\ell)q^{\ell-1/2}}\right)\right\}+ng_h\left(\frac{\log n}{\log q}\right)+\delta_h(n),\tag{27}$$

avec

$$\delta_h(n)=\frac{1}{2\pi i}\int_{1/2-i\infty}^{1/2+i\infty}\left(\frac{n}{q^\ell}\right)^s\frac{q^\ell Z(s-1;h/q^\ell)}{s(1-s)(1-q^{1-s})}\mathrm{d}s.\tag{28}$$

Nous évaluons $\delta_h(n)$ en faisant appel à l'équation fonctionnelle de la fonction zêta de Hurwitz, soit

$$\zeta(s;\alpha)=\Gamma(1-s)\sum_{r\in\mathbb{Z}\setminus\{0\}}(2r\pi i)^{s-1}e(r\alpha)\quad(\Re e\, s<0),$$

où le logarithme complexe est pris en détermination principale. On en déduit, en posant $\alpha=h/q^\ell$,

$$Z(s-1;\alpha)=-s(1-s)\Gamma(-s)\sum_{r\in\mathbb{Z}\smallsetminus\{0\}}(2r\pi i)^{s-2}\,\mathrm{e}(r\alpha)\{\mathrm{e}(r/q^\ell)-1\}.$$

Reportons dans (28) en développant $1/(q^{1-s}-1)=\sum_{k\geqslant1}q^{-k(1-s)}$. Il vient

$$\delta_h(n)=\sum_{r\in\mathbb{Z}\smallsetminus\{0\}}\frac{q^\ell\mathrm{e}(r\alpha)\{\mathrm{e}(r/q^\ell)-1\}}{-4\pi^2r^2}\sum_{k\geqslant1}\frac{1}{2\pi iq^k}\int_{1/2-i\infty}^{1/2+i\infty}\Gamma(-s)\big(2\pi rniq^{k-\ell}\big)^s\mathrm{d}s.$$

Par la formule de Mellin inverse

$$\frac{1}{2\pi i}\int_{1/2-i\infty}^{1/2+i\infty}\Gamma(-s)x^s\mathrm{d}s=\mathrm{e}^{-x}-1\qquad(x>0)$$

(où le terme -1 provient du pôle de Γ à l'origine), on obtient

$$\delta_h(n)=\sum_{r\in\mathbb{Z}\smallsetminus\{0\}}\sum_{1\leqslant k<\ell}\frac{q^\ell\mathrm{e}(rh/q^\ell)\{\mathrm{e}(r/q^\ell)-1\}}{4\pi^2r^2q^k}\{1-\mathrm{e}(-rnq^{k-\ell})\}$$

Cela implique que $\delta_h(n)$ est bien une fonction $q^{\ell-1}$-périodique de n et, compte tenu de (27), achève ainsi la démonstration.

Ajouté à la seconde édition:

En 1903, Takagi [39] a exhibé un exemple de fonction continue non dérivable plus simple, mais essentiellement de même nature, que celui, bien connu, de Weierstrass. La fonction de Takagi est définie par la formule

$$T(x):=\sum_{n\geqslant0}\frac{\|2^nx\|}{2^n}\qquad(x\in\mathbb{R}),$$

où $\|z\|$ désigne la distance du nombre réel z à l'ensemble des entiers. Cette fonction, qui a récemment suscité un intérêt soutenu dans la littérature (voir notamment [19, 25]) est directement liée à la fonction périodique G apparaissant dans le terme résiduel de la formule (3) de Trollope–Delange.

References

1. J.-P. Allouche & J. Shallit, The ring of k-regular sequences, *Theor. Comp. Sci.* **98** (1992), 163–187.
2. J. Brillhart, P. Erdős & P. Morton, On sums of Rudin-Shapiro coefficients, II, *Pac. J. Math.* **107** (1983), 39–69.
3. J. Brillhart & P. Morton, Über Summen von Rudin-Shapirosehen Koeffizienten, *Ill. J. Math.* **22** (1978), 126–148.
4. L. E. Bush, An asymptotic formula for the average sums of the digits of integers, *Amer. Math. Monthly* **47** (1940),154–156.
5. E. Cateland, Suites digitales et suites k-régulières, *Thèse*, Université de Bordeaux 1, 1992.
6. P. Cheo & S. Yien, A problem on the K-adic representation of positive integers, *Acta Math. Sinica* **5** (1955), 433–438.
7. J. Coquet, A summation formula related to the binary digits, *Invent. Math.* **73** (1983), 107–115.
8. J. Coquet, Power sums of digital sums, *J. Number Theory* **22** (1986), 161–176.
9. J. Coquet & P. van den Bosch, A summation formula involving Fibonacci digits, *J. Number Theory* **22** (1986), 139–146.
10. H. Delange, Sur la fonction sommatoire de la fonction "somme des chiffres", *Ens. Math.* **21** (1975), 31–47.
11. J.-M. Dumont, Formules sommatoires et systèmes de numération liés aux substitutions, *Séminaire de théorie des nombres de Bordeaux* (1987/88), Exposé n° 39.
12. J.-M. Dumont & A. Thomas, Systèmes de numération et fonctions fractales relatifs aux substitutions, *Theor. Camp. Sci.* **65** (1989), 153–169.
13. J.-M. Dumont & A. Thomas, Digital sum problems and substitutions on a finite alphabet, *J. Number Theory* **39** (1991), 351–366.
14. P. Flajolet & L. Ramshaw, A note on Gray code and odd-even merge, *SIAM J. Comp.* **9** (1980), 142–158.
15. P. Flajolet, P. Grabner, P. Kirsehenhoffer, H. Prodinger & R. Tichy, Mellin transforms and asymptotics: digital sums, *Theoret. Comput. Sci.* **123** (1994), no. 2, 291–314.
16. D. M. Foster, Estimates for a remainder term associated with the sum of digits function, *Glasgow Math. J.* **29** (1987), 109–129.
17. P. J. Grabner & R. F. Tichy, Contributions to digit expansions with respect to linear recurrences, *J. Number Theory* **36** (1990), 160–169.
18. R. Girgensohn, Digital sums and functional equations, *Integers* ***11*** (2011), #*A*54.
19. H. Harboth, Number of odd binomial coefficients, *Proc. Amer. Math. Soc.* **63** (1977), 19–22.
20. J. Honkala, On number systems with negative digits, *Ann. Acad. Sci. Fenn., Series A. I. Mathematica* **14** (1989), 149–156.
21. R. E. Kennedy & C. N. Cooper, An extension of a theorem by Cheo and Yien concerning digital sums, *Fibonacci Quarterly* **29** (1991), 145–149.
22. P. Kirschenhoffer, Subblock occurrences in the q-ary representation of n, *Siam J. Alg. Disc. Meth.* **4** (1983), 231–236.
23. P. Kirschenhoffer & H. Prodinger, Subblock occurrences in positional number systems and Gray code representation, *J. Inf. Opt. Sci.* **5** (1984), 29–42.
24. P. Kirschenhoffer & R. F. Tichy, On the distribution of digits in Cantor representations of integers, *J. Number Theory* **18** (1984), 121–134.

25. M. Krüppel, Takagi's continuous nowhere differentiable function and binary digital sums, *Rostock. Math. Kolloq.* **63** (2008), 37–54.
26. G. Larcher & R. F. Tichy, Some number-theoretical properties of generalized sum-of-digit functions, *Acta Arith.* **52** (1989), 183–196.
27. M. D. McIlroy, The number of 1's in binary integers: bounds and extremal properties, *SIAM J. Comput.* **3** (1974), 225–261.
28. L. Mirsky, A theorem on representations of integers in the scale of r, *Scripta Math.* **15** (1949), 11–12.
29. D. J. Newman, On the number of binary digits in a multiple of three, *Proc. Amer. Math. Soc.* **21** (1969), 719–721.
30. A. Pethö & R. F. Tichy, On digit expansions with respect to linear recurrences, *J. Number Theory* **33** (1989), 243–256.
31. H. Prodinger, Generalizing the "sum of digits" function, *SIAM J. Alg. Disc. Meth.* **3** (1982), 35–42.
32. P. Shiu & A. H. Osbaldestin, A correlated digital sum problem associated with sums of three squares, *Bull. London Math. Soc.* **21** (1989), 369–374.
33. A. H. Stein, Exponential sums related to binomial coefficient parity, *Proc. Amer. Math. Soc.* **80** (1980), 526–530.
34. A. H. Stein, Exponential sums of an iterate of the binary sum of digit function, *Indiana Univ. Math. J.* **31** (1982), 309–315.
35. A. H. Stein, Exponential sums of sum-of-digit functions, *Ill. J. Math.* **30** (1986), 660–675.
36. A. H. Stein, Exponential sums of digit counting functions, in : J.-M. De Koninck et C. LeVesque (eds.), *Théorie des Nombres* (Québec 5-18/7/87), 861–868, Walter de Gruyter, Berlin-New York, 1989.
37. K. B. Stolarsky, Digital sums and binomial coefficients, *Notices Amer. Math. Soc.* **22** (1975), A 669, Abstract # 728-A7.
38. K. B. Stolarski, Power and exponential sums related to binomial digit parity, *SIAM J. Appl. Math.* **32** (1977), 717–730.
39. T. Takagi, A simple example of the continuous function without derivative, *Proceedings of the Physico-Mathematical Society of Japan*, ser. II, **1** (1903), 176–177. [Collected Papers of Teiji Takagi (S. Iyanaga, ed.), Springer Verlag, New York 1990].
40. J. R. Trollope, Generalized bases and digital sums, *Amer. Math. Monthly* **74** (1967), 690–694.
41. J. R. Trollope, An explicit expression for binary digital sums, *Math. Mag.* **41** (1968), 21–25.

1105: First Steps in a Mysterious Quest

Gérald Tenenbaum

G. Tenenbaum (✉)
Institut Élie Cartan, Université de Lorraine, BP 70239,
54506 Vandœuvre-lès-Nancy Cedex, France
e-mail: gerald.tenenbaum@univ-lorraine.fr

To Paul Erdős, who held the torch.

During the summer of 1975, I spent a few days with my mother and sister who were on holidays near La Baule. I had just left *École Polytechnique*, and needed some rest after the military service. For 8 months I had been a sub-lieutenant in the $2^{\text{ème}}$*RAMA*, a semi-disciplinary unit based in Vernon, Eure, and felt rather depressed after what had been for me a dreadful experience. For the time being, my main concern was the starting of my research in mathematics. I had regular "night-dreams", and also daydreams, seeing myself "content-free" as a mathematician, working hard but having no ideas—and, of course, no results.

Here I was on the beach, reading Hardy & Wright and trying to make a few notes on a pad in spite of sand and wind. I soon became fascinated by arithmetic functions (a subject completely new to me) and considered it hardly believable that the number $r(n)$ of representations of an integer n as a sum of two squares had such nice and simple properties. On the one hand, I was trying to reconstruct the proof of the main formula, viz

$$r(n) = 4 \prod_{p^\nu \| n,\ p \equiv 1 \pmod 4} (\nu + 1) \prod_{p^\nu \| n,\ p \equiv 3 \pmod 4} \left(\frac{1 + (-1)^\nu}{2} \right),$$

along the lines of the algebraic number theory course I had taken at the University of Paris, writing here and there on the pad: *revise Kummer's Theorem*. On the other hand, I was doing small experiments such as computing explicitly the first integer with a given value for $r(n)$. The smallest n with $r(n) = 32$ is

$$1105 = 5 \times 13 \times 17.$$

There are four genuinely distinct ways of writing 1105 as a sum of two squares, namely

$$1105 = 23^2 + 24^2 = 31^2 + 12^2 = 32^2 + 9^2 = 33^2 + 4^2.$$

Don't you agree that this set of representations is rather odd?

Today, nearly 20 years later (this is odd too, isn't it?), I would probably consider the occurrence of three consecutive squares $(31^2, 32^2, 33^2)$ as an epiphenomenon, and turn my attention to what I would regard as deeper

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_20,

subjects. But, at that time, I had nothing else to bite on and so started right away to try to describe the set of those integers N which have two genuinely distinct representations

$$N = x^2 + y^2 = (x+1)^2 + z^2.$$

This turned out to be very easy, as the reader can imagine, and I soon found out that these N are exactly those which can be written as

$$N = \{q^2 + (q+1)^2\}\{r^2 + (r+1)^2\}$$

with suitable relative integers q, r not equal to 0 or -1. One then retrieves

$$x = r + 2qr + q, \quad y = q + r + 1, \quad z = q - r.$$

For $q = r = 1$, we obtain $N = 25 = 4^2 + 3^2 = 5^2 + 0^2$, and $N = 1105$ corresponds to $q = 6, r = 2$.

Thus, 1105 was not a unique specimen and actually had infinitely many relatives! This was nice and called for generalization. The next, obvious step was to replace the condition that the squares should be consecutive by imposing that they should be squares of integers with prescribed difference n, in other words to search all N which can be written as

$$N = x^2 + y^2 = (x+n)^2 + z^2$$

with $\{x, y\} \neq \pm\{x+n, z\}$. This was not much more difficult, and I ascertained that these integers are exactly described by the formula

$$N = \{q^2 + (q+n)^2\}\{r^2 + (r+n)^2\}/n^2$$

where $2qr \equiv 0 \pmod{n}$ and $qr(q+n)(r+n)(2q+n)(2r+n) \neq 0$. I denoted by $g(n)$ the smallest such N, and left for Bordeaux with my own arithmetical function in my pocket.

At this stage the reader might wonder when Erdős is going to enter the scene. Such expectation, however, is perhaps significant: in one of Maurice Leblanc's best novels, *L'Éclat d'obus*, one does await the hero Arsène Lupin for the major part of the book, but his presence is increasingly felt as ineluctable as the plot is developed—and the fact that one of the chapters is entitled *75 ou 155?* is a purely formal coincidence for which number theorists should give no hasty interpretation.

I arrived in Bordeaux in the early days of September, 1975, and went to see François Dress who, I was told, had worked on sums of squares. In fact , he was interested in Waring's problem and had obtained new bounds for $g(4)$—I mean, of course, *the* g-function, usually used with argument k and defined as the smallest integer such that all numbers are sums of at most $g(k)$ kth powers.[1] This perhaps explains, at least partly, why he wasn't immediately

[1] Balasubramanian, Deshouillers and Dress established in 1986 that $g(4) = 19$, thereby closing up (in a certain sense, which would take us too far to describe here) Waring's classical problem.

crazy about my g-function. To tell the truth, he thought this was a load of piffle, but, not wanting to discourage an enthusiastic young man, he said that, in his opinion, the significance was ε. This g-function wasn't simply nor nicely defined, after all, and he believed that even writing a computer program to tabulate it up to $n = 10^8$ or 10^{10} might not be an easy task. A few days later, I came up with the following formula for $g(n)$. Put $t(x) = \frac{1}{2}(x^2+1)$ and define, for integer n, the functions

$$\rho_1(n) := \max_{d|n,\ d\leqslant\sqrt{n}} d, \qquad \rho_2(n) := \max_{d|n,\ d\geqslant\sqrt{n}} d.$$

(Thus $\rho_1(n)$ and $\rho_2(n)$ are the two divisors of n closest to $\sqrt{n}$.) Then, denoting by p a prime number, we have

$$g(1) = g(2) = 25, \quad g(p) = 5t(p) \quad (p > 2), \quad g(2p) = 10t(p),$$
$$g(n) = t\big(\rho_1(n)\big)t\big(\rho_2(n)\big) \quad (n \text{ odd, not a prime number}),$$
$$g(2n) = 4t\big(\rho_1(n)\big)t\big(\rho_2(n)\big) \quad (n \text{ not a prime number}).$$

I was quite happy with this result and had already applied it to determine the limit points of the set $\{g(n)/n^2 : n \in \mathbb{N}\}$, as well as the asymptotic behaviour of the average $(1/N)\sum_{n\leqslant N} g(n)\ldots$ Dress understood I was not going to give up easily: he erased the whole blackboard, except these new functions ρ_1, ρ_2, which "were defined in less than 14 characters", so not unnatural; moreover, they seemed not to have been studied before...

By this time, about a month had passed, and I had settled in Bordeaux, where the number theory group, created under the impulse of Pisot, was developing in a kind of semi-familial everyday life—with its well-known pleasant and not so pleasant implications. My 'older brother' (and future co-adviser with Dress) was Jean-Marc Deshouillers, with whom I had frequent discussions. He patiently taught me all basic notions in the field of arithmetic functions as well as recent directions of research, and I tried to apply these to my two newcomers.

Everybody had already noticed that these questions where 'Erdős-like'. After a while it became apparent that the hard problem was to evaluate the average order of $\rho_1(n)$. Up to this date, this is still open. Dress had offered 50 francs for a proof or disproof of

$$x^{-1}\sum_{n\leqslant x}\rho_1(n) = o\big(\sqrt{x}\big).$$

After about 2 weeks of struggle with bare hands, I opened Halberstam and Roth, Chap. V, and got in one night the bounds

$$\sqrt{x}/\log x \ll x^{-1}\sum_{n\leqslant x}\rho_1(n) \ll \sqrt{x}/\log\log x.$$

This was something, but obviously insufficient, and I had no idea of what the next step should be.

Fortunately, a meeting on elementary and analytic number theory organized by Richert, Schwarz and Wirsing was scheduled in November, if my memory is right, at Oberwolfach. Deshouillers and Dress were invited and promised to ask Erdős about ρ_1. They came back with an amazing answer: Erdős conjectured that

$$x^{-1}\sum_{n\leqslant x}\rho_1(n)=\sqrt{x}/(\log x)^{\delta+o(1)},$$

with, believe me or not, $\delta = 1-(1+\log\log 2)/\log 2 = 0.08607\ldots$!! He added that he thought the method was similar to that of his 'Russian paper'—the only paper Erdős ever published in the Russian language, although, as he told me later, he does not know any Russian.

My friend Didier Nordon provided a translation and I started to study the paper, with Deshouillers' help. This wasn't an easy job. I remember (and, I am sure, so does Jean-Marc) a five-hour train trip between Paris and Bordeaux during which we tried to understand the notation, sometimes rather obscure, and above all the 'philosophy' of the man who knew about numbers.

Hardy and Ramanujan proved in 1917 that almost all integers n have about log log n prime factors, but it was Erdős who really understood all the possibilities opened up by this theorem. Indeed, Hardy and Ramanujan's proof gives more than the so-called normal order of the function $\Omega(n)$, equal to the total number of prime factors of n, counted with multiplicity. Essentially, the extra information is that the distribution of values of $\Omega(n)$ among integers $n\leqslant x$ is roughly Poisson, with mean and variance $\log\log x$. The peak of the Poisson distribution is very narrow, and the tails are dominated, in first approximation, by single values. Erdős took advantage of this situation by clever splittings of the integers according to their number of prime factors (possibly in a given range), which shed light on otherwise rather intractable problems.

Let us give an example with the following reasoning, typical of his approach. At the same time this will give an idea of the proof of the upper bound in the δ-conjecture—which I could establish, along the lines foreseen by Erdős, by the end of 1975. In Erdős' setting (the one I used at the time), the computations would involve factorials, because of the Poisson probabilities, and the optimization would be rather cumbersome. However, the 'parametric method', introduced in this context by Richard Hall and which we subsequently developed together,[2] simplifies the technicalities a great deal. Suppose we want an upper estimate for the number $S(x)$ of integers $n\leqslant x$ with $\frac{1}{2}\sqrt{x}<\rho_1(n)\leqslant\sqrt{x}$. We give ourselves a parameter $\lambda\in(1,\frac{3}{2})$ and split the integers up to x into two classes, according to whether $\Omega(n)>\lambda\log\log x$ or not. The number of elements of the first class may be bounded, for any $v\in(1,\frac{3}{2})$, by

[2] See our book *Divisors* for an expository text on this subject.

$$\sum_{n\leqslant x} v^{\Omega(n)-\lambda\log\log x} \ll x(\log x)^{v-1-\lambda\log v}.$$

The best choice is clearly $v = \lambda$, so the exponent of log x becomes $-Q(\lambda)$, with

$$Q(\lambda) = \lambda\log\lambda - \lambda + 1 > 0.$$

The above estimate follows from a classical result on non-negative arithmetic functions (see e.g. Halberstam and Richert, 1979) and we only mention that the version needed here may be proved elementarily in a few lines. As for the elements of the second class, the following expression is certainly an upper bound for any $w \in (0,1]$

$$\sum_{n\leqslant x} w^{\Omega(n)-\lambda\log\log x} \sum_{d|n,\frac{1}{2}\sqrt{x}<d\leqslant\sqrt{x}} 1$$
$$= (\log x)^{-\lambda\log w} \sum_{\frac{1}{2}\sqrt{x}<d\leqslant\sqrt{x}} w^{\Omega(d)} \sum_{m\leqslant x/d} w^{\Omega(m)},$$

where, in the right-hand side, we have written $n = md$ and permuted summations. Estimating the inner sum as previously, we obtain that this is

$$\ll x(\log x)^{w-1-\lambda\log w} \sum_{\frac{1}{2}\sqrt{x}<d\leqslant\sqrt{x}} w^{\Omega(d)}/d \ll x(\log x)^{2w-2-\lambda\log w}.$$

The optimal choice is now $w = \frac{1}{2}\lambda$, which does belong to (0,1], and the subsequent exponent of $\log x$ is $-Q(\lambda) + \lambda\log 2 - 1$. Putting our estimates together, we deduce from standard calculus that we must select $\lambda = 1/\log 2$. The absolute value of the exponent becomes $Q(1/\log 2)$, which is equal to Erdős' miraculous δ!

Erdős came to Bordeaux in March 1977. At that time, he taught me about the law of iterated logarithm for prime factors (i.e. the jth prime factor of a normal number is roughly $\exp\exp j$) but in the form of a basic consequence: the largest divisor of n all of whose prime factors do not exceed y is itself usually not much larger than y. The following, not unrelated device also came up in the conversation: if you want to find out whether an integer has a divisor in a given interval $(y, z]$, you should essentially look at those prime factors of n in the range $(z/y, z]$, Of course, these mottos weren't formulated this way, and the light only came gradually. Nevertheless, I learned a great deal during this first visit, and I must say that all other meetings proved just as fruitful for me. It would be unfair not to mention that this unconventional teaching was delivered with extreme patience and kindness.

Thus, Erdős, in the first instance, showed the way into the study of divisors by giving the basic devices; this is invaluable: everybody having worked in analytic number theory knows that the first step is very often the most difficult. But a second help, equally important, came from the conjectures he made. These acted as a permanent measure of the depth of our methods and the quality of our knowledge.

In the first rank of these, I would place the conjecture, made in the late 30s, according to which almost all integers have at least two divisors d, d' with $d < d' \leqslant 2d$. This was precisely the question to challenge the current model for the set of divisors of a normal integer. The point is important; let us be more explicit. According to the law of iterated logarithm the jth prime factor of n, say $p_j(n)$, satisfies $\log p_j(n) \approx \mathrm{e}^j$; hence, the divisors d should be described (assuming n squarefree for simplicity) by the formula

$$\log d \approx \sum_{1 \leqslant j \leqslant \Omega(n)} \varepsilon_j \mathrm{e}^j$$

where the ε_j take the values 0 or 1. However, since $\mathrm{e} > 2$, these quantities have increasing differences, and hence the smallest ratio between consecutive divisors should normally stay away from zero, and indeed should be large on a set of positive density. (This last statement is justified, for instance, by the fact that we can assume all prime factors to be large on a set of positive density.) Thus, the conjecture is really about the effect of the error terms in the law of the iterated logarithm. This explains simultaneously why it is interesting and why it is hard. This problem shares with most of Erdős' questions the feature of being phrased in the simplest form which contains the whole substance of the matter. It was eventually solved in 1983 by Maier and myself, and the proof led to the following evaluation (another conjecture of E.P.) valid for almost all integers

$$\min_{d, d' | n, \ d < d'} d'/d = 1 + (\log n)^{1 - \log 3 + o(1)}.$$

As a matter of fact, it would be hard to imagine a way to solve the initial problem that wouldn't give quantitative results of this kind: the question is so pertinent that only a deeper understanding of the structure of the set of divisors will yield an answer. This experience, and many others of a similar type, explain why I find it sad to hear, from time to time, Erdős' problems qualified as anecdotal or scattered. This criticism comes from occasionally powerful, but always ignorant, people.

Since Erdős' first pioneering articles in the thirties, quite a few mathematicians have taken part in the strange quest which I entered through 1105. It would, of course, take us too far to give here a complete survey of the results that have been obtained, even during the last two decades. Nevertheless, I would like to mention another recollection.

At the Durham Symposium on Analytic Number Theory, organized in 1979 by Halberstam and Hooley, a (gentle) mathematical controversy occurred between Erdős and Montgomery about the normal behaviour of the function

$$G(n) = |\{1 \leqslant j < \tau(n) : d_j(n) | d_{j+1}(n)\}|,$$

where $d_j(n)$ denotes the jth divisor of n. Erdős thought that $G(n) = o(\tau(n))$ for almost all n because this fitted with another conjecture of his, namely that, almost always,

$$\tau^+(n) := |\{k \in [1, \log n/\log 2] : \exists d, d|n, 2^k < d \leqslant 2^{k+1}\}| = o\big(\tau(n)\big).$$

(Note that $G(n) \leqslant \tau^+(n)$ for all n, since $d_{j+i}(n)/d_j(n) \geqslant 2$ whenever $d_j(n)|d_{j+1}(n)$.) The above bound for $\tau^+(n)$ would have trivially implied the d, d'-conjecture mentioned earlier, and indeed it could be regarded as a kind of ultimate consequence of Erdős' conception that the set of divisors could be described by big aggregates with very large gaps between them. Montgomery's argument was probabilistic. Assume for simplicity that n is even, say $n = 2m$. Then at least one half of the divisors $d_j(n)$ divide m; for such a divisor, we have $d_j(n)|d_{j+1}(n)$ unless $d_{j+1}(n) < 2dj(n)$ because $2d_j(n)$ is a divisor of n. However, the intervals $\big(d_j(n), 2d_j(n)\big]$ occupy, on a logarithmic scale, a set of measure $\leqslant (\log 2)\tau(n) = (\log n)^{\log 2+o(1)}$ inside the interval $(1, n]$, of measure log n. Therefore, it is conceivable that the probability that a divisor belongs to one of these intervals tends to 0. As soon as this probability is less than $\frac{1}{2} - c$, we have $G(n) \geqslant c\tau(n)$.

Thus, one can see that an apparently innocent question about a seemingly artificial function turns out to be critical to our understanding of the structure of the divisors. It was well-known that, on a slightly larger scale, the structure was indeed mainly in aggregates, so Erdős' question was really: how strong is this tendency? Is it so strong that it can destroy probabilistic effects even in very short intervals of constant multiplicative length? As a matter of fact, it turned out in this case that Erdős' conjecture was wrong-headed. In a joint work (1981), he and I proved that the density of those n with $G(n) \leqslant c\tau(n)$ tends to 0 with c.[3] It remains that the question was well-posed, and that the answer revealed precious information about the multiplicative structure of the integers, and particularly on the phenomena which involve the conflict between the ordering imposed, on a large scale, by the law of iterated logarithm and the local, random perturbations.

The purpose of the present text is to give a comprehensible (if not comprehensive) account on Erdős' extraordinary stamp on this part of Probabilistic Number Theory. The reader interested in further information may consult, besides the research bibliography, our book with R.R. Hall, *Divisors*. A synthetic approach, in the frame of fractal sets, which provides a complementary description of some models of the divisors is presented in a joint article with Michel Mendès France (1993): the fractal dimension of the set of divisors of a normal integer is log 2. The whole subject has been discovered, developed and guided by Erdős' theorems, intuitions, and conjectures. Let these few lines be a token of admiration to the mathematician, and of gratitude to the man.

[3] But this density is strictly positive, which shows that the tendency on which Erdős based his conjecture is nevertheless quite constraining.

Added at the second edition:

The reader will find an update of many topics discussed in this paper in my paper entitled *Some of Erdős' unconventional problems in number theory, thirty-four years later*, to appear in the acts of the Erdős centennial conference, Budapest, July 2013.

References

1. R. Balasubramanian, J.-M. Deshouillers & F. Dress, Problème de Waring pour les bicarrés, 1 : schéma de la solution, *C.R. Acad. Sci. Paris* **303** (1986), 85–89.
2. R. Balasubramanian, J.-M. Deshouillers & F. Dress, Problème de Waring pour les bicarrés, 2 : résultats auxilaires pour le théorème asymptotique, *C.R. Acad. Sci. Paris* **303** (1986), 161–163.
3. P. Erdős, On an asymptotic inequality in the theory of numbers (Russian), *Vestnik Leningrad Univ.* **13** (1960), 41–49.
4. P. Erdős & G. Tenenbaum, Sur la structure de la suite des diviseurs d'un entier, *Ann. Inst. Fourier* **31**, 1 (1981), 17–37.
5. H. Halberstam & H.E. Richert, On a result of R.R. Hall, *J. Number Theory* (1) **11** (1979), 76–89.
6. H. Halberstam & K.F. Roth, *Sequences*, Oxford University Press (1966).
7. R.R. Hall & G. Tenenbaum, *Divisors*, Cambridge University Press (1988).
8. H. Maier & G. Tenenbaum, On the set of divisors of an integer, *Inventiones Math.* **76** (1984), 121–128.
9. M. Mendès France & G. Tenenbaum, Systèmes de points, diviseurs, et structure fractale, *Bull. Soc. Math. de France, Bull. Soc. Math. de France* **121** (1993), 197–225.

III. Randomness and Applications

Introduction

It would be foolish to try describing Paul Erdős as an applied mathematician, somebody who is looking outside of mathematics for motivation and justification of his activity. Yet we believe that the word application in the title is justified. An essential part of Erdős' personality and success is his broad knowledge and a true feeling of unity of mathematics. This understanding brought him to many of his crucial discoveries and topics. Randomness is a pivotal example of this Erdős approach. With M. Kac he initiated this technique in number theory in 1939 (see the Erdős paper in this volume) and in graph theory he did so in 1946. This technique—the probabilistic or nonconstructive method—is by now one of the universally accepted modern combinatorial techniques and this is also reflected by papers in this section. The origins of the probabilistic method are described in the paper by Spencer while the random graph papers of Erdős and Rényi are described in the Karonski–Ruciński paper (with update). Applications of probabilistic methods have reached virtually all mathematical disciplines as well as many areas of theoretical computer science. The papers by Pyber, Pudlák and Sgall, and Razborov are such examples. This does not exhaust the papers relevant to probabilistic methods in this volume, see for example, the papers in the second volume by Bollobás and Füredi, Kahn, Laczkovich and Ruzsa, and the paper by Cameron (devoted to infinity). All of these papers are related to various aspects of Erdős' work and belong to branches pioneered by him. By now this is well recognized and, there are, in fact, numerous books devoted to these areas, such as: B. Bollobás: Random Graphs, B. Bollobás: Extremal Graph Theory and also E. M. Palmer: Random graphs, P. Erdős, J. Spencer: Probabilistic methods in combinatorics and N. Alon, J. Spencer: The probabilistic method, and P. D. T. A. Elliot: Probabilistic number theory (to name just a few).

Games, Randomness and Algorithms

József Beck*

J. Beck (✉)
Department of Mathematics, Rutgers University, New Brunswick,
NJ 08903, USA
e-mail: jbeck@math.rutgers.edu

1. Tic-Tac-Toe-Like Games

The object of this 50 % survey and 50 % "theorem-proof" paper is to demonstrate recent developments of some of the ideas initiated by Erdős [17, 18], Erdős and Selfridge [20], Erdős and Lovász [19] and Erdős and Chvátal [15].

The story begins with a class of deterministic games of complete information.

The most well-known chapter of combinatorial game theory deals with Nim-like games (a player unable to move loses). A beautiful theory was developed by Bouton, Sprague, Grundy, and recently by Berlekamp, Conway, Guy and others—and of course, Conway's theory of "numbers and games" (see the remarkable book of B-C-G [13]). These ideas can be employed in games in which the positions are composed of several non-interacting simpler games. Then the first thing to do is to associate values (numbers or "generalized numbers") with these parts. Next comes the problem of finding ways of determining the outcome of a sum of games given information only about the values of the separate components.

Here, however, we discuss a quite different branch of combinatorial game theory. This branch is in such an early stage of its developments that the experts didn't even find a right name for the subject: it is frequently called as Tic-Tac-Toe-like Games, or Positional Games, or Pattern Games, or Achievement Games. We prefer the name "Tic-Tac-Toe-like games". In contrast to Nim-like games which start out as composites, or develop into composites in the normal course of play, these games usually remain as single coherent entities throughout play. So the theory of Nim-like games (or Neumann's "theory of games") has little relevance to such "condensed" games.

The best introduction to these games is Chap. 22 *Lines and Squares* of *Winning Ways* by Berlekamp, Conway and Guy. The traditional ideas in Tic-Tac-Toe-like games are the "strategy stealing argument", "pairing strategies" and in general, the trick of "decomposition into non-interacting local games".

* Partially supported by NSF Grant # DMS-9106631.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_21,

Here we make a more systematic use of Ramsey Theory, and emphasize the importance of a relatively new idea what we call the "probabilistic intuition". This leads to very effective *global* weight function strategies, and at the same time, it provides surprisingly accurate predictions to the outcomes of many complex games. We consider Tic-Tac-Toe-like games as a sort of bridge between the two well-established chapters of "Random Graphs and Hypergraphs" and "Ramsey Theory". Also we include some applications in Complexity Theory.

As a warm-up we start with two well-known board games which are won by the first player to complete some kind of winning pattern.

Tic-Tac-Toe (or Noughts and Crosses): The game board is a big square which is partitioned into $9 = 3 \times 3$ smaller squares. Whoever moves first puts a cross in one of the nine small squares. His opponent then puts a nought into any other square and then they alternate nought and cross in the remaining empty squares until one player wins by getting three of his own squares in line. There are precisely eight winning lines.

Hex: The game board is a rhombus of hexagons, the actual size of the board being a matter of agreement between the players. Each player takes a pair of opposite sides, his move is to take an untaken hexagon, and his object is to form a continuous chain between his two sides. Hex was invented by Piet Hein.

Strategy Stealing Argument

A winning strategy for a player is a list of instructions telling the player that if the opponent does this, then he does that so if he follows this strategy, he will win. In a tic-tac-toe type game, one can argue that the first player can achieve at least a draw. This follows by the well-known *strategy stealing argument.*

Suppose otherwise the second player has a winning strategy, and the first player steals it. Then he can use this strategy to win the game. He randomly takes his first move, and then pretends himself as the second player. He reads the instruction to take action. If he is told to take a move that is still available, he takes it. If this move was taken by him before, then he randomly takes another move. The crucial point here is that an extra move only benefits him. In this way, he can guarantee a win. Therefore, the second player cannot have a winning strategy, which implies that the first player can guarantee at least a draw.

Remember, however, when a draw is impossible, the first player wins. It is easy to see that in Hex a draw is impossible. Therefore, whoever plays first wins. We must warn the reader that this is a theoretical result, which assumes that the first player doesn't make mistakes. If he makes a mistake, then the opponent can take advantage of the mistake, and might manage a win.

The reader possibly noticed that the two games above have some differences. In the tic-tac-toe game, both players can take any move, and they also share the same winning sets. In the Hex game, both can take any move, but they have different winning sets.

We define the class of *generalized tic-tac-toe games* as follows. Let there be given a finite set X, which is the *board* of the game, and a family of *winning subsets* of X, say $\mathcal{F} = \{A_1, A_2, \ldots, A_m\}$, $A_i \subseteq X$. The two players alternately take an element from X which is still available. That player who takes a complete winning set first is the winner. This family of games contains the tic-tac-toe (noughts and crosses) game as a particular case.

Since an extra move in a generalized tic-tac-toe game does not harm a player, by the strategy stealing argument, whoever plays first has a drawing strategy.

Theorem 1. *Let X be finite, and $\mathcal{F}$ be an arbitrary family of subsets of X. Then the first player can force at least a draw in the generalized tic-tac-toe game on $(X, \mathcal{F})$.*

However, the "complexity" of finding this (at least) drawing strategy is enormous. Indeed, a strategy for the first player is a function f with domain = "the set of subsequences of the board X" such that the "next move" $f(x_1, y_1, \ldots, x_{i-1}, y_{i-1})$ is always an element of X different from the "previously selected" elements $x_1, y_1, \ldots, x_{i-1}, y_{i-1}$ of X. In a play according to this strategy, the first player determines all his moves by f as follows: Suppose the players alternately picked the elements $x_1, y_1, \ldots, x_{i-1}, y_{i-1}$ of X in this order. Then the first player's ith move is $x_i = f(x_1, y_1, \ldots, x_{i-1}, y_{i-1})$.

Let $|X| = N$. Then the total number of strategies is between

$$2^{2^N} \text{ and } N^{N^N}.$$

The amount of computation demanded by a computer to find an optimal strategy is a plain exponential function of the size of the board. Human brains can sometimes diagnose shortcuts, but we cannot expect substantial shortcuts in general. Even "small" board games are so complex that it is possible that the existing drawing (or winning) strategies may never be found. As a discouraging example, we mention that, Owen Patashnik, of Bell Lab, was the first to find a first-player winning strategy in the 3-dimensional 4 × 4 × 4 tic-tac-toe (or as better called "tic-toc-tac-toe"). His 1977 program required 1,500 h of computing time, and the winning strategy contains 2,929 strategic moves.

We emphasize that the strategy stealing argument is highly nonconstructive. It does not give much help in finding an explicit winning or drawing strategy for the first player. For example, no explicit winning strategy in Hex is known.

When Can the First Player Win?

There is a rather general sufficient condition for the first player's win. Consider a generalized tic-tac-toe game on $(X, \mathcal{F})$. The chromatic number of $\mathcal{F}$ is the least integer r such that the elements of X can be colored with r colors yielding no monochromatic $A \in \mathcal{F}$. If the chromatic number of $\mathcal{F}$ is ≥ 3, then draw is impossible, so by the strategy stealing argument, the first player has a winning strategy.

Theorem 2. *Suppose that the board X is finite, and the family $\mathcal{F}$ of winning sets has chromatic number at least* 3. *Then the first player has a winning strategy in the generalized tic-tac-toe game on* $(X, \mathcal{F})$.

Note that it is a "hard" problem to decide whether a system $(X, \mathcal{F})$ has chromatic numbers ≥ 3. We have to check all the $2^{|X|}$ two-colorings of X.

There is a well-established chapter of Combinatorics, called Ramsey Theory, which is entirely devoted to families $\mathcal{F}$ of chromatic number ≥ 3. We associate games with three famous results of Ramsey Theory.

Ramsey Game

If S is a set, then let $[S]^k$ denote the family of subsets of S containing exactly $k \geq 2$ elements. Following the set-theoretical traditions, we identify the natural number n with the set of its predecessors, that is, $n = \{0, 1, \ldots, n-1\}$. So $[n]^2$ can be regarded as a complete graph with n vertices, that is, $[n]^2 = K_n$. Let $2 \leq n < N$. The board of the game is $[N]^2 = K_N$, and the two players alternately occupy edges of this graph (i.e., elements of $[N]^2$) and that player wins who picks all the edges of a complete subgraph with n vertices (i.e., all the elements of $[S]^2$ for some n-element subset $S \subset N = \{0, 1, \ldots, N-1\}$) first. This game is denoted by $R(N, n) = R_2(N, n)$.

For $k \geq 3$, the game $R_k(N, n)$ is a trivial generalization. The players alternately occupy k-element subsets of N, and the winner is who picked all the elements of $[S]^k$ for some n-element subset $S \subset N$ first.

Van der Waerden Game

The game $W(N, n)$ is played on the board $X = \{0, 1, \ldots, N-1\}$, and the winning sets are the arithmetic progressions of n terms from X.

Hales-Jewett Game

This is a straightforward multidimensional generalization of the game tic-tac-toe. The two players alternately put their marks in the cells of a d-dimensional cube of size $n \times n \times \cdots \times n = n^d$. The winner is the first player to have n of his marks in a line. More precisely, the board of the game $H\ J(d,n)$ is the set of d-tuples

$$X = \{\mathbf{a} = (a_1, a_2, \ldots, a_d) : 0 \le a_j < n \text{ for each } 1 \le j \le d\}\,.$$

The winning sets of $H\ J(d,n)$ are those n-element sequences

$$\left(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(n)}\right)$$

of the board X such that, for each j, the sequence $a_j^{(1)}, a_j^{(2)}, \ldots, a_j^{(n)}$ composed of the jth coordinates is either strictly increasing from 0 to $n-1$, or strictly decreasing from $n-1$ to 0, or constant.

The well-known theorems of Ramsey, Van der Waerden, and Hales and Jewett state that for every integer n there is a finite threshold number $r_k(n)$, $w(n)$ and $h(n)$, respectively, such that the family of winning sets in the Ramsey Game, Van der Waerden Game and Hales-Jewett Game has chromatic number ≥ 3 if $N = r_k(n)$, $N = w(n)$ and $d = h(n)$, respectively. Therefore, by Theorem 2, the first player has a winning strategy in these games. Unfortunately this theorem is rather weak, because either the best upper bound on the threshold number known at present is very poor, or the threshold number is in fact enormous. Consider e.g. the Van der Waerden threshold number $w(n)$. The best upper bound on $w(n)$, due to Shelah, is primitive recursive (Van der Waerden's original argument couldn't even provide a primitive recursive upper bound), but it is still enormous for "pedestrian mathematics". For example, it is an open problem whether $w(n) < \exp_n(n)$ holds, where $\exp_k$ denotes the k-fold iteration of the exponential function $\exp(x) = e^x$. On the other hand, the best known lower bound is plain exponential: $w(n) > 2^n$. The situation is exactly the same in the case of the Hales-Jewett theorem.

In the case of Ramsey theorem, however, the upper and lower bounds are much closer to each other. It is known

$$2^{n/2} < r_2(n) < 4^n, \tag{1}$$

$$2^{n^2/6} < r_3(n) < 2^{2^{4n}}, \tag{2}$$

and in general, for arbitrary $k \ge 3$,

$$\exp_{k-2}(c_k \cdot n) < r_k(n) < \exp_{k-1}(c'_k \cdot n), \tag{3}$$

where c_k and c'_k are absolute constants depending on k only.

It seems to be highly unlikely that the breaking points for the behavior of these Ramsey type games is anywhere close to the behavior of the corresponding Ramsey threshold functions, but no method is known for handling these problems. Note that even the existence of a breaking point is questionable. It may well happen that the first player wins e.g. the Hales-Jewett Game $H\ J(d,n)$, but the $H\ J(d+1,n)$ game is a draw.

Pairing Strategies

By far the most common technique to guarantee a win or a draw is to find a decomposition of the board into disjoint pairs, and when your opponent takes one member from a pair, you take the other one.

Every child "knows" that in the usual tic-tac-toe (i.e., the Hales-Jewett Game $H\ J(2,3)$) the second player can force a draw. A precise mathematical proof of this fact can be found in Chap. 22 of Berlekamp, Conway and Guy [13]. If the board is 4×4 and the object of the game is to find 4-in a-row (i.e., the Hales-Jewett Game $H\ J(2,4)$), then again the second player can force a draw. Exactly the same holds for all the Hales-Jewett Games of the type $H\ J(2,n)$, $n = 5, 6, 7, 8, \ldots$ (see B-C-G [13]). We just give a quick proof of the cases $n = 5$ and 6 by the following pairing strategies: whenever the first player occupies a numbered cell, the second player takes the other cell of the same number.

$$n = 5 : \qquad \begin{bmatrix} 2 & 10 & 5 & 5 & 1 \\ 6 & 9 & 12 & 8 & 9 \\ 6 & 11 & * & 11 & 4 \\ 7 & 10 & 12 & 7 & 4 \\ 1 & 3 & 3 & 8 & 2 \end{bmatrix}$$

If the first player takes the $*$-labeled center, the second player may take any cell, and if the cell he is required to take by the pairing strategy is occupied, he may play anywhere.

$$n = 6 : \qquad \begin{bmatrix} 1 & 13 & 2 & 13 & 3 & 12 \\ 6 & 14 & 5 & 14 & 4 & 12 \\ 7 & 8 & 15 & 9 & 10 & 15 \\ 16 & 3 & 11 & 1 & 16 & 2 \\ 17 & 4 & 11 & 6 & 17 & 5 \\ 7 & 8 & 18 & 9 & 10 & 18 \end{bmatrix}$$

What is more, this last numbering leads to an elegant proof that the unrestricted 9-in-a-row (i.e., the board is an infinite chess board, and the object of the game is to find 9-in-a-row orthogonally or diagonally) is a draw. Cover the infinite board with copies of the 6 by 6 matrix above. The second player can force a draw by always taking the nearest cell with the same number as of the previous play. It is easy to see that the first player can obtain no line longer than 8.

unrestricted 8-in-a-row: It is known that the second player has an explicit drawing strategy. However this strategy is not a pairing strategy. To illustrate the idea, first we outline a second proof of the fact that the unrestricted 9-in-a-row is a second player's draw.

Decomposition into Non-interacting Games

It was first proved by Pollak and Shannon in 1954 that for 9-in-a-row the second player can force a draw, by using the following strategy. Tile the plane with H-shaped heptominos (seven squares): the second player plays ordinary tic-tac-toe in each of these 7-square regions, concentrating on preventing a line of 3 in either a diagonal, or the horizontal or the right vertical.

Using the same idea T.G.L. Zetters (nom de guerre of some Amsterdam combinatorists) recently showed that the second player can even draw 8-in-a-row. Their proof uses a parallelogram-shaped tile of 12 cells, and goes some way towards showing that 7-in-a-row is also a draw.

Next we give a general sufficient condition for the existence of pairing strategies.

Theorem 3. *Consider a generalized tic-tac-toe game on a finite system $(X, \mathcal{F})$. Assume that for any subfamily $\mathcal{G} \subseteq \mathcal{F}$,*

$$\left| \bigcup_{A \in \mathcal{G}} A \right| \geq 2|\mathcal{G}|.$$

Then the second player can force a draw by a pairing strategy.

Proof. By the well-known König-Hall theorem, we can find disjoint 2-element representatives: $h(A) \subset A$ $(A \in \mathcal{F})$, $|h(A)| = 2$, $h(A) \cap h(B) = \emptyset$ whenever A and B are different elements of $\mathcal{F}$. (A technical twist: we in fact apply the König-Hall theorem to the "duplicate" of $\mathcal{F}$, i.e., every $A \in \mathcal{F}$ is taken in two copies.) Now whenever the first player occupies one member from a 2-element representative, then the second player takes the other one. □

Note that there are well-known efficient (i.e. polynomial time) algorithms to actually find the disjoint 2-element representatives.

Corollary 1. *Let $\mathcal{F}$ be an n-uniform system, i.e., $|A| = n$ for every $A \in \mathcal{F}$. Further assume that every $x \in X$ is contained by at most $n/2$ members of $\mathcal{F}$. Then the second player can force a draw by a pairing strategy.*

Proof. Observe that the criterion of Theorem 3 applies here. Indeed, for any subfamily $\mathcal{G} \subseteq \mathcal{F}$, by a standard double-counting argument,

$$n|\mathcal{G}| = \sum_{A \in \mathcal{G}} |A| = \sum_{x \in \bigcup_{A \in \mathcal{G}} A} \ \sum_{A \in \mathcal{G} \ x \in A} 1 \leq \left| \bigcup_{A \in \mathcal{G}} A \right| \cdot \frac{n}{2},$$

and dividing by $n/2$,

$$\left| \bigcup_{A \in \mathcal{G}} A \right| \geq 2|\mathcal{G}|.$$

□

Note that Theorem 3 is very general. It doesn't restrict the global size of $\mathcal{F}$ at all, and in fact it holds for infinite boards as well. In spite of this, Theorem 3 is not very useful in our "condensed" games. Consider e.g. the three Ramsey type games $R_2(N, n)$, $W(N, n)$ and $H\ J(d, n)$. We leave to the reader to check that in the cases of Ramsey Game $R_2(N, n)$ and Van der Waerden Game, Theorem 3 gives a ridiculously weak bound on N (like if $N < const{\cdot}n$ then the game is a draw). For the Hales-Jewett Game $H\ J(d, n)$, in their fundamental paper [21], Hales and Jewett proved, by using Theorem 3 that if

$$n \geq 3^d - 1 \quad (n \text{ odd}) \quad \text{and} \quad n \geq 2^{d+1} - 2 \quad (n \text{ even}) \tag{4}$$

then the game is a draw. They conjectured that the game is always a draw if there are at least twice as many cells as lines (which is a necessary condition for any pairing strategy). How many lines are there? It is easy to see that each winning line in $H\ J(d, n)$ is determined by the two cells which extend the line into the surrounding cube

$$(n+2) \times (n+2) \times \cdots \times (n+2) = (n+2)^d.$$

So the total number of lines is

$$\frac{(n+2)^d - n^d}{2}. \tag{5}$$

Therefore, the Hales-Jewett conjecture is that the game is a draw whenever

$$n^d \geq (n+2)^d - n^d, \tag{6}$$

that is,

$$2 \geq \left(\frac{n+2}{n}\right)^d.$$

Since

$$\left(\frac{n+2}{n}\right)^d = \left(1 + \frac{2}{n}\right)^d \approx e^{2d/n},$$

(6) is equivalent to

$$n \approx \frac{2d}{\log 2} \tag{7}$$

(compare it to the weak bounds in (4)).

2. Weight Function Strategies: A Fake Probabilistic Method

In 1973 a really useful sufficient condition for second-player's draw was found by Erdős and Selfridge [20]. Opposite to the pairing strategy and its variants, where we decompose the position into several non-interacting small *local* games, the Erdős-Selfridge approach is *global*.

Theorem 4. *If*

$$\sum_{A \in \mathcal{F}} 2^{-|A|} < \frac{1}{2}$$

then the second player can force a draw in the generalized tic-tac-toe game on $(X, \mathcal{F})$.

Corollary 2. *If* $\mathcal{F}$ *is* n*-uniform and* $|\mathcal{F}| < 2^{n-1}$*, then the game is a draw.*

The proof of Theorem 4 is based on an exponential weight function technique. It is often called as "derandomization of the first moment method": indeed, the criterion in Theorem 4 ensures that in a standard random 2-coloring of X the expected number of monochromatic members $A \in \mathcal{F}$ is less than 1. (If the expected number of monochromatic sets is less than one, then of course there exists a "good" 2-coloring in the sense that *no* monochromatic set shows up (see Erdős [18]). This gives at least a *chance* for a *drawing strategy*. The message of Theorem 4 is that under the global condition above this drawing strategy really exists. Note, however, that the usual 3×3 tic-tac-toe gives a family of 8 triplets, for which "good" 2-coloring exists, but the second player cannot prevent the first one from completing a winning triplet.)

Proof of Theorem 4. We in fact prove a stronger statement: under the condition of Theorem 4, the second player can prevent his opponent from completing any winning set $A \in \mathcal{F}$. According this, we call the first and second players "Maker" and "Breaker", respectively.

Given any family $\mathcal{G}$ of subsets of X we assign the total weight

$$T(\mathcal{G}) = \sum_{A \in \mathcal{G}} 2^{-|A|}$$

to $\mathcal{G}$. Consider now a play on our family $\mathcal{F}$ in which the points

$$v_1, w_1, v_2, w_2, \ldots \in X$$

were picked by the two players in this order. After Maker's ith move, define the "truncated family" $\mathcal{F}_i$ for $i \geq 1$ as follows. Throw away those sets from

$\mathcal{F}$ which contain any point picked by Breaker ("dead sets"), and from the remaining sets ("survivors") throw away the points picked by Maker, i.e.,

$$\mathcal{F}_i = \{A \setminus \{v_1, \dots, v_i\} : A \in \mathcal{F} \text{ and } A \cap \{w_1, \dots, w_{i-1}\} = \emptyset\}.$$

Maker wins if and only if some of the $\mathcal{F}_i$'s contain the empty set, and since the cardinality of the empty set is zero, in this case $T(\mathcal{F}_i) \geq 2^{-0} = 1$. Thus if $T(\mathcal{F}_i) < 1$ for every $i \geq 1$, then Breaker wins.

We define a strategy for Breaker. Let the weight of a set $A \in \mathcal{F}_i$ be $2^{-|A|}$, and the weight of a point of $\mathcal{F}_i$ be the sum of the weights of the sets in $\mathcal{F}_i$ it belongs to. In his ith move Breaker picks that point of $\mathcal{F}_i$ which is of largest weight. We claim that

$$T(\mathcal{F}_{i+1}) \leq T(\mathcal{F}_i)$$

independently of Maker's $(i+1)$st move. If we prove this the result follows since every set of $\mathcal{F}_1$ contains at most one point of Maker, so by hypothesis $T(\mathcal{F}_1) < 2 \cdot \frac{1}{2} = 1$. Therefore, $T(\mathcal{F}_i) < 1$ for every $i \geq 1$.

We check $T(\mathcal{F}_{i+1}) \leq T(\mathcal{F}_i)$. Right after Breaker's ith move (that is, before Maker's $(i+1)$st move), the sum of the weights of the sets is $T(\mathcal{F}_i) - W$ where W is the weight of the ith point of Breaker. On Maker's next move he doubles the weight of each "surviving" set containing his $(i+1)$st point v_{i+1}, so he adds to $T(\mathcal{F}_i) - W$ no more than the *previous* weight W' of v_{i+1}. But Breaker's ith move was a point of largest weight, so $W \geq W'$. It follows that

$$T(\mathcal{F}_{i+1}) \leq T(\mathcal{F}_i) - W + W' \leq T(\mathcal{F}_i)$$

which was to be proved. □

Note that the strategy above has a probabilistic interpretation: it is a greedy algorithm to minimize certain conditional probabilities—see e.g. Chap. 15 in Alon-Spencer [1]. This weight function method is our most important tool to handle Tic-Tac-Toe-like Games.

Note that Corollary 2 is sharp: the full branches of a binary tree with n levels form an n-uniform family of 2^{n-1} winning sets such that the first player has an easy win in n steps.

In opposition to Theorem 3, Corollary 2 gives quite good results in Ramsey type games. In the Ramsey Game $R_2(N, n)$, the criterion holds if

$$\binom{N}{n} < 2^{\binom{n}{2}-1},$$

so the game

$$R_2(N, n) \text{ is a draw if } N \leq 2^{n/2}. \tag{8}$$

In the other direction, by Theorem 2 and (2),

$$\text{the first player can win the } R_2(N,n) \text{ game if } N \geq 4^n. \tag{9}$$

These are the best results known at present.

Next consider the Van der Waerden Game $W(N,n)$. Since the number of n-term arithmetic progressions in $\{0,1,\ldots,N-1\}$ is less than N^2/n, the criterion applies if $N^2/n < 2^{n-1}$. This holds if

$$N < \sqrt{n} \cdot 2^{\frac{n-1}{2}}, \tag{10}$$

so the game $W(N,n)$ is a draw if (10) is satisfied.

By employing some special properties of arithmetic progressions, we can improve on this bound (see [4]): if

$$N < (2-\varepsilon)^n \text{ and } n > n_0(\varepsilon), \tag{11}$$

then the $W(N,n)$ game is a draw. This is the best known estimation.

Finally consider the Hales-Jewett Game $H\ J(d,n)$. The Erdős-Selfridge criterion holds if

$$\frac{(n+2)^d - n^d}{2} < 2^{n-1},$$

that is, the second player can force a draw whenever $n > const \cdot d \cdot \log d$. This result just falls short of conjecture (7) of Hales and Jewett that the game is a draw if $n \approx 2d/\log 2$. The family of winning sets in $H\ J(d,n)$, however, has an important additional feature: any two winning lines have at most one point in common. Set-systems with this property are called *almost disjoint.* For almost disjoint set-systems the upper bound 2^{n-1} of Corollary 2 can be raised considerably (see [3]).

Theorem 5. *There is an absolute constant $c > 0$ such that for every n-uniform almost disjoint family $\mathcal{F}$ of winning sets, if*

$$|\mathcal{F}| < 4^{n-c\sqrt{n}},$$

then the second player can force a draw in the generalized tic-tac-toe game on $(X,\mathcal{F})$.

Note that this theorem is also sharp as far as the order of magnitude is concerned. Erdős constructed a 3-chromatic n-uniform almost-disjoint set-system $\mathcal{F}^*$ with no more than $n^4 \cdot 4^n$ n-sets (see [19]). In view of Theorem 4, the generalized tic-tac-toe game played on this $\mathcal{F}^*$ is a first-player win.

Though the family of n-term arithmetic progressions in the Van der Waerden game $W(N,n)$ is not almost disjoint, the situation strongly resembles to almost disjointness. This is why we can employ the proof-technique of Theorem 5 to the game $W(N,n)$, and can jump in (10)–(11) from $2^{n/2}$ up to $(2-\varepsilon)^n$.

To settle conjecture (7) about the Hales-Jewett Game $H\ J(d,n)$ (at least for sufficiently large numbers), we just need a slight refinement of Theorem 5. The really important property is the restriction $2^{n-c\sqrt{n}}$ of the local size, the global size of the family of winning sets can be as large as (say) $2^{n^{3/2}}$, and the game is still a draw (see [3]).

Theorem 6. *Let $\mathcal{F}$ be an n-uniform family of almost disjoint subsets of a finite X. Assume that every $x \in X$ is contained in at most $2^{n-c\sqrt{n}}$ members of $\mathcal{F}$, and $|\mathcal{F}| < 2^{n^{3/2}}$. Then the second player has a winning strategy in the generalized tic-tac-toe game on $(X, \mathcal{F})$.*

This result can be interpreted as a sort of "game-theoretic local lemma", at least for almost disjoint systems.

The proofs of Theorems 5–6 are based on the same idea. We split the game into two non-interacting parts, what we call the "BIG GAME" (a shrinking game) and the "small game" (a growing game). The "BIG SETS" are the unions of some "connected" subfamilies of $\mathcal{F}$. The "small sets" are the available parts of those "dangerous" members of $\mathcal{F}$, which are almost entirely occupied by the first player and are not blocked by the second player yet. If the family of "small sets" is "sparse", then the second player can employ Theorem 3 to block them. The "BIG GAME" plays an auxiliary role here: by using the strategy of Theorem 4, the second player can force that the "small game" is always "sparse", so that pairing strategy really works.

It is easy to see that in the Hales-Jewett Game $H\ J(d,n)$, every cell of the board $n \times \cdots \times n = n^d$ is contained by at most $(3^d-1)/2$ winning lines. So Theorem 6 applies here if

$$\frac{3^d-1}{2} \leq 2^{n-c\sqrt{n}} \text{ and } \frac{(n+2)^d-n^d}{2} < 2^{n^{3/2}}. \tag{12}$$

Inequalities (12) trivially hold if

$$n > \left(\frac{\log 3}{\log 2} + \varepsilon\right) d \text{ and } n > n_0(\varepsilon). \tag{13}$$

Observe that (13) is asymptotically better than Hales-Jewett conjecture (7). That is, the second player can force a draw even when pairing strategy simply cannot exist.

3. Maker-Breaker Version

As we have seen in Sect. 1, the second player has no chance to win a generalized tic-tac-toe game against a perfect first player. Then, why does he not just concentrate on preventing his opponent from a win? We thus can name him the Breaker, and the other one the Maker. We can even take this one step further, and consider the Maker-Breaker game even if the Breaker moves first.

On the same system $(X, \mathcal{F})$ we can play the symmetric generalized tic-tac-toe game, and also the asymmetric Maker-Breaker version, where Maker's aim is to pick every element of a winning set $A \in \mathcal{F}$, and Breaker's aim is to prevent Maker from doing so. The winner is the one who achieves his goal (so a draw is impossible).

If Breaker = second-player wins the Maker-Breaker version on a system $(X, \mathcal{F})$, then the same play gives him, as a second player, (at least) a draw in the tic-tac-toe version on $(X, \mathcal{F})$. So from Theorem 2 immediately follows

Corollary 3. *Assume that X is finite, and the family $\mathcal{F}$ of winning sets has chromatic number at least 3. Then Maker = first-player has a winning strategy in the Maker-Breaker version on $(X, \mathcal{F})$.*

The converse is not true. It is possible that Maker = first-player wins the Maker-Breaker version while Breaker = second-player can force a draw in the tic-tac-toe version. This happens for example in the usual 3×3 tic-tac-toe: the original game is a draw but the Maker-Breaker version is a win for Maker = first-player. It is true, however, that the same criterions as in Theorems 4–6 are sufficient for Breaker = second-player to win the Maker-Breaker version (and so they are automatically sufficient for Breaker = first-player to win as well).

The proof of Theorem 4 actually gives the stronger

Theorem 7. *Assume that X is finite, and*

$$\sum_{A \in \mathcal{F}} 2^{-|A|} < \frac{1}{2}.$$

Then Breaker = second-player has a winning strategy in the Maker-Breaker version on $(X, \mathcal{F})$.

Similarly, we have (see [3])

Theorem 8. *Assume that $\mathcal{F}$ is an n-uniform almost disjoint family of subsets of a finite X. Suppose further that every $x \in X$ is contained in at most $2^{n-c\sqrt{n}}$ members of $\mathcal{F}$, and $|\mathcal{F}| < 2^{n^{3/2}}$. Then Maker = second-player has a winning strategy in the Maker-Breaker version on $(X, \mathcal{F})$.*

While playing the tic-tac-toe version on an $(X, \mathcal{F})$, both players have their own threats, and either of them, fending off the other's, may build his own winning set. Therefore, a play is a delicate balancing between threats and counter-threats and can be of very intricate structure even if the system $(X, \mathcal{F})$ of the game is simple.

The Maker-Breaker version is usually simpler. Maker doesn't have to waste valuable moves fending off his opponent's threats. He can simply concentrate on his own goal. This is why we have a surprisingly simple sufficient condition for Maker's win (see [4]), which is usually far better than Corollary 3 (see Theorem 9 below). Note however that, though the Maker-Breaker version is usually simpler, it is still "very hard". For example, Hex is equivalent to a Maker-Breaker version: Maker = first-player wants a connecting chain, and Breaker = second-player simply wants to prevent his opponent from achieving his goal.

Theorem 9. *Assume that $\mathcal{F}$ is n-uniform, and any point in X is contained by less than half of the members of $\mathcal{F}$. Suppose moreover that, fixing any two elements of the board X, no more than d winning sets from $\mathcal{F}$ contain both of them. If*

$$|\mathcal{F}| \geq 2^{n-2} \cdot d \cdot |X|,$$

then Maker has a winning strategy in the Maker-Breaker version on $(X, \mathcal{F})$, and it doesn't matter at all that Marker is the first or the second player.

Proof. We consider the worse case where Maker is the second player. Here Maker "borrows" Breaker's weight function strategy from the proof of Theorem 4.

Given any family $\mathcal{G}$ of subsets of X we assign the total weight

$$T(\mathcal{G}) = \sum_{A \in \mathcal{G}} 2^{-|A|}$$

to $\mathcal{G}$. Consider now a play on our family $\mathcal{F}$ in which the points

$$v_1, w_1, v_2, w_2, \ldots \in X$$

were picked by the two players in this order. After Breaker's ith move, define the "truncated family" $\mathcal{F}_i$ for $i \geq 1$ as follows. Throw away those sets from $\mathcal{F}$ which contain any point picked by Breaker ("dead sets"), and from the remaining sets ("survivors") throw away the points picked by Maker, i.e.,

$$\mathcal{F}_i = \{A \setminus \{w_1, \ldots, w_{i-1}\} : A \in \mathcal{F} \text{ and } A \cap \{v_1, \ldots, v_i\} = \emptyset\}.$$

Maker wins if the last total sum $T(\mathcal{F}_{end})$ is still positive, that is, if at the end of the game there is a "survivor".

We define the winning strategy for Maker = second-player as follows. Let the weight of a set $A \in \mathcal{F}_i$ be $2^{-|A|}$, and the weight of a point of $\mathcal{F}_i$ be the sum of the weights of the sets in $\mathcal{F}_i$ it belongs to. In his ith move Maker picks that point of $\mathcal{F}_i$ which is of largest weight. We claim that

$$T(\mathcal{F}_{i+1}) \geq T(\mathcal{F}_i) - \frac{d}{4}$$

independently of Breaker's $(i+1)$st move. If we prove this, the result follows since every set of $\mathcal{F}_1$ contains at most one point of Breaker, namely v_1, so by hypothesis $T(\mathcal{F}_1) > |\mathcal{F}|/2 \cdot 2^{-n}$. Therefore,

$$T(\mathcal{F}_{end}) > |\mathcal{F}|2^{-n-1} - |X|/2 \cdot d/4 \geq 0,$$

and Maker wins.

We check $T(\mathcal{F}_{i+1}) \geq T(\mathcal{F}_i) - \frac{d}{4}$. Right after Maker's ith move (that is, before Breaker's $(i+1)$st move), he doubles the weight of each "surviving" set containing his ith point w_i, so he adds to $T(\mathcal{F}_i)$ the weight W of his ith point w_i. On Breaker's next move he subtracts the *new* weight of each "surviving" set containing his $i+1$st point v_{i+1}, so he subtracts from $T(\mathcal{F}_i) + W$ the *previous* weight W' of v_{i+1}, and also the *previous* weight of those "surviving" sets which contain *both* w_i and v_{i+1}:

$$W'' = \sum_{\substack{B \in \mathcal{F}_i \\ \{w_i, v_{i+1}\} \subseteq B}} 2^{-|B|}.$$

By hypothesis, $W'' \leq d/4$. Since Maker's ith move was a point of largest weight, so $W \geq W'$. It follows that

$$T(\mathcal{F}_{i+1}) = T(\mathcal{F}_i) + W - W' - W'' \geq T(\mathcal{F}_i) - \frac{d}{4},$$

which was to be proved. □

If $\mathcal{F}$ is almost disjoint (that is, any two different winning sets have at most one common element), then fixing two elements of the board X, no more than one winning set can contain both of them. So we have

Corollary 4. *If $\mathcal{F}$ is an n-uniform almost disjoint system such that any point of X is contained in less than half of the members of $\mathcal{F}$, and $|\mathcal{F}| > 2^{n-2}|X|$, then Maker has a winning strategy in the Maker-Breaker version on $(X, \mathcal{F})$, and it doesn't matter that Maker is the first or the second player.*

To appreciate Theorem 9, we study the Maker-Breaker versions of the Ramsey type games $R_k(N, n)$, $W(N, n)$ and $H\ J(d, n)$. Note that statement (iv) below is a straightforward consequence of Theorem 9 (see [4]).

Theorem 10. *Consider first the Maker-Breaker version of the Ramsey Game $R_2(N, n)$:*

(i) If $N \leq 2^{n/2}$ then Breaker;
(ii) If $N \geq (2+\varepsilon)^n$ then Maker

has a winning strategy. For every $k \geq 3$ there are positive constants c_k and c'_k such that, in the Maker-Breaker version of $R_k(N,n)$:

(iii) If $N \leq 2^{c_k \cdot n^{k-1}}$ then Breaker;
(iv) If $N \geq 2^{c'_k \cdot n^k}$ then Maker

has a winning strategy.

Comparing Theorems 2 and 10 we see that the breaking point for the Maker-Breaker version is within much-much-much closer bounds than that of the tic-tac-toe version. For a general k, the former one is between (roughly) $2^{n^{k-1}}$ and 2^{n^k}, and the latter one is between $2^{n^{k-1}}$ and the tower

$$\exp\left(\exp\left(\cdots\exp(n)\cdots\right)\right)$$

of height k. Moreover Theorem 10(iv) in fact proves that, in the Maker-Breaker version of the Ramsey Game $R_k(N,n)$, Corollary 3 (= Ramsey theory) *fails* to give the true order of magnitude of the breaking point. Indeed, the Ramsey threshold number $r_k(n)$ is known to be bigger than the tower

$$\exp\left(\exp\left(\cdots\exp(n)\cdots\right)\right) \tag{14}$$

of height $k-1$ (see (3)), but Maker = first-player can win around $N = 2^{n^k}$, which is asymptotically much smaller than (14) if $k \geq 4$.

The difference between the Maker-Breaker version and the tic-tac-toe version of the Van der Waerden Game is even more dramatic (see [4]).

Theorem 11. *Let $\varepsilon > 0$ be arbitrary. If N is large enough depending only on ε, then in the Maker-Breaker version of the Van der Waerden Game $W(N,n)$:*

(i) If $N \leq (2-\varepsilon)^n$ then Breaker;
(ii) If $N \geq n^3 \cdot 2^n$ then Maker

has a winning strategy.

By Theorems 2 and 11, the breaking point in the Maker-Breaker version is around 2^n, but in the tic-tac-toe version we cannot even prove the upper bound

$$\exp\left(\exp\left(\cdots\exp(n)\cdots\right)\right),$$

which is a tower of height n.

The situation is very similar in the Hales-Jewett Game: the best known upper bounds in the two versions are an astronomical distance from each other.

Theorem 12. *Let $\varepsilon > 0$ be arbitrary. In the Maker-Breaker version of the Hales-Jewett Game $H\ J(d, n)$:*

(i) If $n < \sqrt{\frac{2d}{\log 2}}$ then Maker;

(ii) If $n > \bigl(\frac{\log 3}{\log 2} + \varepsilon\bigr)d$ then Breaker

has a winning strategy (if n is large enough).

Finally, we return to Hex. Nobody knows an *explicit* winning strategy for Hex, but there is a large class of very similar games for which explicit winning strategies were found.

Lehman's Theory of "the Shannon Switching Game"

The Shannon Switching Game is played on a graph representing an electrical network in which certain nodes are labelled $+$ and some others are labelled $-$. Each edge (begin the game with them drawn in pencil) represents a permissible connexion between the nodes at its ends. Maker, at his move may establish one of these connexions permanently (ink over a penciled edge) and attempts to form a chain between some $+$ node and a $-$ one. His opponent, Breaker may permanently prevent a possible connexion (erase a penciled edge) and tries to separate $+$ from $-$ forever. You can always suppose that there is only one positive node and one negative one by making identifications.

Supposing this, Alfred Lehman has proved that Maker can win as second player if and only if he can find two edge-disjoint trees which each contain all the nodes of some subgraph containing $+$ and $-$. The "only if" part is hard, but there is an easy *explicit* strategy which proves "if": whenever Breaker's move separates one of the trees into two parts, say A and B, Maker makes a move on the other tree joining a vertex of A to one of B.

The game can be generalized to make the winning configurations for Maker just those which contain a specified family $\mathcal{P}$ of sets of edges. (In the original game $\mathcal{P}$ was the family of all paths from $+$ to $-$.) Lehman proves the "only if" part of this theorem by taking $\mathcal{P}$ to be the family of all trees containing every vertex (spanning trees).

If Maker, as second player, has a win in the modified game, then by the strategy stealing argument there must be two edge-disjoint spanning trees. For since an extra move is no disadvantage, both players can play Maker's strategy. If they do this, two spanning trees will be established, using disjoint sets of edges. Conversely, if two such trees exist, our previous strategy for Maker actually wins for him as second player, even in the modified game.

The more detailed part of Lehman's argument establishes that, in a suitable sense, the modified game reduces to the original one.

4. Complex Graph Games and Random Graphs

When a 1-1 game is overwhelmingly in favor of one of the players one can make up for this handicap by allowing the other player to claim many points in a move. For example, when played on a large complete graph K_n, Lehman's game (= Maker wants a spanning tree) is overwhelmingly in favor of Maker. So we allow Breaker to claim many edges per move: let each move of Breaker consist of claiming b previously unclaimed edges. Clearly if b is large with respect to n then Breaker wins, if b is small with respect to n then Maker has a win. In the biased case Lehman's criterion above hopelessly breaks down. However, the following heuristic argument, due to P. Erdős, suggests the game-theoretic threshold point has to come around $n/\log n$.

During a play Maker takes $\approx \frac{n^2}{2b}$ edges. In particular, if $b = \frac{n}{2c\log n}$, then Maker creates a graph with $\approx c \cdot n \cdot \log n$ edges. A well-known theorem from the theory of Random Graphs states that a random graph with n points and $c \cdot n \cdot \log n$ edges is "almost certainly" connected for $c > 1/2$ and "almost certainly" disconnected for $c < 1/2$. It turns out that the game theoretic breaking point does indeed come around $b = n/\log n$; more precisely, it is between $(\log 2 - \varepsilon)n/\log n$ and $(1+\varepsilon)n/\log n$ for all sufficiently large values of n.

Let us return to the random graph with n points and $c \cdot n \cdot \log n$ edges. Around $c = 1/2$ the random graph undergoes a remarkable change: If $c < 1/2$ then it has plenty of isolated points; if $c > 1/2$ then it contains a Hamiltonian cycle (which trivially implies connectivity). The fundamental difference between connectivity and the existence of Hamiltonian cycles is that the former property can be efficiently (i.e., the "worst case" running time is a polynomial function of the input data) proved if it holds and efficiently disproved if it does not, but nobody knows an efficient way of disproving that a graph contains Hamiltonian cycles. We can overcome this technical difficulty, and are able to show that the "probabilistic intuition" works: the breaking point for the Hamiltonian Cycle Game comes around $b = n/\log n$. More precisely, if $b = (\frac{\log 2}{27} - \varepsilon)\frac{n}{\log n}$ then Maker can build up a Hamiltonian cycle of his own.

In general, we shall examine the following class of graph games. Two players, Breaker and Maker, with Breaker going first, play on the complete graph K_n of n vertices in such a way that Breaker claims $b(\geq 1)$ previously unselected edges per move and Maker claims one previously unselected edge per move. Maker wins if he claims all the edges of some graph from a family of prescribed subgraphs of K_n. Otherwise Breaker wins, that is, Breaker simply wants to prevent Maker from doing his job.

Let *Clique*$(n; b, 1; r)$ denote the game where Maker wants a complete subgraph of r vertices (from his own edges of course). Denote by *Connect* $(n; b, 1)$ and *Hamilt*$(n; b, 1)$ the games where Maker's goal is to select a spanning tree (i.e. a connected subgraph of K_n) and a Hamiltonian cycle of

K_n, respectively. It is well-known that the largest clique in a random graph of n vertices and of parameter $p = 1/2$ has approximately $2 \log n / \log 2$ vertices with probability tending to one as n tends to infinity. In view of Erdős' heuristic, this suggests that Maker wins the fair game $\mathit{Clique}(n; 1, 1; r)$ if r is around $\log n$. The following result is just a reformulation of Theorem 10(i)–(ii) (see Erdős-Selfridge [20] and Beck [4]).

Theorem 13. *Breaker has a winning strategy in the fair game*

$$\mathit{Clique}(n; 1, 1; 2 \log n / \log 2).$$

On the other hand, Maker has a winning strategy in

$$\mathit{Clique}(n; 1, 1; (1 - \varepsilon) \log n / \log 2)$$

if n is large enough depending on $\varepsilon > 0$ only.

The next result is due to Erdős-Chvátal [15] and Beck [5, 7].

Theorem 14. *We have:*

(i) If $b > (1 + \varepsilon) n / \log n$

then Breaker has a winning strategy in $\mathit{Connect}(n; b, 1)$ if n is large enough.

(ii) If $b > (\log 2 - \varepsilon) n / \log n$

then Maker has a winning strategy in $\mathit{Connect}(n; b, 1)$ if n is large enough.

(iii) If $b > (\log 2/27 - \varepsilon) n / \log n$

then Maker has a winning strategy in $\mathit{Hamilt}(n; b, 1)$ if n is large enough.

Next we point out further instances of this exciting analogy between the evolution of random graphs and biased graph games (see [10]). The basic idea is that Maker's graph possesses some fundamental properties of random graphs (mostly "expandability" type properties) provided Maker uses his best possible strategy.

Let us begin with a trivial observation: if $b = 2n$ then Breaker can easily prevent Maker even from getting a path of two edges (Breaker blocks the two endpoints of Maker's edge). If $b = \varepsilon n$, $\varepsilon > 0$ constant, then, in view of Theorem 14(i), Breaker can force the disconnectivity of Maker's graph. In fact, Breaker can force at least $\frac{\varepsilon}{2} e^{-\frac{1}{\varepsilon}} n$ isolated points in Breaker's graph, and it goes as follows (the following argument is a straightforward adaptation of the Erdős-Chvátal proof of Theorem 14(i)). Breaker proceeds in two stages. In the first stage, he will claim all the edges of some clique K_m^* with $m = \varepsilon n/2$ vertices such that none of the Maker's edges has an endpoint in this K_m^*. In the second stage, he will claim all the remaining edges incident with at least $\frac{\varepsilon}{2} e^{-\frac{1}{\varepsilon}} n$ vertices of K_m^* thereby forcing at least $\frac{\varepsilon}{2} e^{-\frac{1}{\varepsilon}} n$ isolated points in Maker's graph.

The first stage lasts no more than $m = \varepsilon n/2$ moves, and goes by a simple induction by m. During his first $i-1$ $(1 \le i \le m)$ moves, Breaker has created a clique K_{i-1}^* with $i-1$ vertices such that none of the Maker's edges has an endpoint in K_{i-1}^*. At this moment there are $i-1 < \varepsilon n/2$ Maker's edges, hence there are at least two vertices u, v in the complement of $V(K_{i-1}^*)$ which are incident with none of the Maker's edges. On his ith move, Breaker claims edge $\{u,v\}$ and all the edges joining u and v to the vertices of K_{i-1}^*, thereby enlarging K_{i-1}^* by two vertices. Then Maker can kill one vertex from this clique K_{i+1}^* by claiming an edge incident with that vertex. Nevertheless, a clique of i vertices still "survives".

In the second stage, Breaker has $m = \varepsilon n/2$ *pairwise disjoint* edge-sets: for every $u \in V(K_m^*)$, the edges joining u to all vertices in the complement of $V(K_m^*)$. It is easy to see that Breaker can completely occupy at least $e^{-\frac{1}{\varepsilon}}m$ of these m disjoint edge-sets by the simple rule that he has the same (or almost the same) number of edges from all the "surviving" edge-sets at any time.

The first result says that Maker is able to build up a cycle of length $> (1 - e^{-\frac{1}{200\varepsilon}})n$. That is, if Breaker claims εn edges per move, then Maker has an "almost Hamiltonian cycle" in the sense that the complement is "exponentially small" (the constant factor of 200 in the exponent is of course very far from the best possible—here we don't make any effort to find the optimal, or at least nearly optimal, constants).

Theorem 15. *If* $0 < \varepsilon < 1/200$ *and Breaker selects* εn *edges per move then Maker can build up a cycle of length* $> (1 - e^{-\frac{1}{200\varepsilon}})n$ *on a board* K_n.

Note that if Maker just wants a path P_m of $m = (1 - \text{const}\sqrt{\varepsilon})n$ edges, then he can do it in the *shortest way*: in m moves. Indeed, Maker can employ the following simple greedy strategy: he keeps extending his path by adding that available point (as a new endpoint) which has minimum degree in Breaker's graph. Trivial calculation shows that this greedy procedure doesn't terminate in $m = (1 - \mathit{const} \cdot \sqrt{\varepsilon})n$ moves.

Observe that if $\varepsilon = \frac{1}{200 \log n}$ then Theorem 15 gives the order of magnitude of Theorem 14(iii).

Next question: what can we say about Maker's largest degree (i.e. Maker's largest star)? Obviously Breaker can prevent Maker from having a degree larger than $2N/b$ on a board K_N: if Maker selects an edge $\{u,v\}$ then Breaker occupies $b/2$ edges from u and $b/2$ edges from v.

For simplicity restrict ourselves to the fair case (i.e. $b = 1$). As far as I know this problem is due to Erdős (oral communication). László Székely [27] proved (by using Lemma 3 in Beck [4]) that Breaker can prevent a star of size $\frac{n}{2} + \mathit{const} \cdot \sqrt{n \log n}$. In the other direction, we can prove that Maker can achieve a star of size $\frac{n}{2} + \mathit{const} \cdot \sqrt{n}$. If the complete graph K_n is replaced with the complete *bipartite* graph $K_{n,n}$ then we obtain the following interesting "row-column game".

Theorem 16. *Consider the game where the board is an $n \times n$ chessboard. Breaker and Maker alternately select one previously unselected cell. Breaker marks his cells blue and Maker marks his cells red. Maker's object is to achieve at least $\frac{n}{2} + k$ ($k \geq 1$) red cells in some line (= row or column). If $k = \frac{\sqrt{n}}{32}$ then Maker has a winning strategy.*

Note that the proof of Theorem 16 is based on a game theoretic second moment method. This explains the "standard random fluctuation" of size $\sqrt{n}$.

Theorem 16 is in sharp contrast with the chessboard type alternating two-coloring, where the discrepancy in every line is 0 or 1 depending on the parity of n.

Note that a straightforward modification of the proof of Theorem 16 gives the above-mentioned case where the board is K_n.

Proof of Theorem 15. Given a simple and undirected graph G, and an arbitrary subset S of the vertex-set $V(G)$ of G, denote by $\Gamma_G(S)$ the set of vertices in G adjacent to at least one vertex of S. Let $|S|$ denote the number of elements of a set S.

The following lemma is essentially due to Pósa [26] (a weaker version was earlier proved by Komlós-Szemerédi [22]). A trivial corollary of the lemma is that an expander graph has a long path.

Lemma 1. *Let G be a non-empty graph, $v_0 \in V(G)$ and consider a path $P = (v_0, v_1, \ldots, v_m)$ of maximum length which starts from v_0. If $(v_i, v_m) \in G (1 \leq i \leq m-1)$ then we say that the path $(v_0, \ldots, v_i, v_m, v_{m-1}, \ldots, v_{i+1})$ arises by Pósa-deformation from P. Let $\mathrm{end}(G, P, v_0)$ denote the set of all endpoints of paths arising by repeated Pósa-deformation from P (the starting point v_0 is always fixed). Assume that for each vertex-set $S \subset V(G)$ with $|S| \leq k$, $|\Gamma_G(S) \setminus S| \geq 2|S|$. Then $|end(P, G, v_0)| \geq k+1$.*

In order to use Lemma 1, we need

Lemma 2. *Under the hypothesis of Theorem 15, Maker can guarantee, that right after Breaker occupied $\frac{1}{20}\binom{n}{2}$ edges, Maker's graph satisfies the following property. Property A: For any vertex-set S of K_n with $\frac{1}{3}e^{-\frac{1}{200\varepsilon}}n \leq |S| \leq n/4$, $|\Gamma_G(S) \setminus S| \geq 2|S| + e^{-\frac{1}{200\varepsilon}}n$ where G is Maker's graph right after Breaker claimed $\frac{1}{20}\binom{n}{2}$ edges.*

Proof. We employ a general theorem concerning hypergraph games. Let $\mathcal{H}$ be a hypergraph with vertex-set $V(\mathcal{H})$ and hyperedge-set $E(\mathcal{H})$, and let $p \geq 1$ and $q \geq 1$ be integers. A($\mathcal{H}; p, 1; q$)-game is a game on $\mathcal{H}$ in which two players, **I** and **II**, select p and 1 previously unselected vertices per move from $V(\mathcal{H})$. The game proceeds until $\frac{1}{q}|V(\mathcal{H})|$ vertices has been selected by **I**. **II** wins if he occupies at least one vertex from every hyperedge $A \in E(\mathcal{H})$; otherwise

I wins. In [7] we proved the following generalization of the Erdős-Selfridge criterion: If

$$\sum_{A\in E(\mathcal{H})} 2^{-|A|/pq} < \frac{1}{2} \tag{15}$$

then **II** has a winning strategy in the $(\mathcal{H}; p, \mathbf{1}; q)$-game.

In order to employ (13), we introduce some hypergraphs. Let m be an integer satisfying $\frac{1}{3}e^{-\frac{1}{200\varepsilon}}n \leq m \leq n/4$, and let $\mathcal{H}(n;m)$ be the set of all complete $m\times(n-3m-e^{-\frac{1}{200\varepsilon}}n+1)$ bipartite subgraphs of K_n. The "vertices" of $\mathcal{H}(n;m)$ are the edges of K_n.

Now to ensure property A, in view of (15) with $p = b = \varepsilon n$ and $q = 20$, it is enough to check the following inequality:

$$\sum_{m=\frac{1}{3}e^{-\frac{1}{200\varepsilon}}n}^{n/4} \binom{n}{m}\binom{n-m}{2m+e^{-\frac{1}{200\varepsilon}}n-1} 2^{-m(n-3m-e^{-\frac{1}{200\varepsilon}}n+1)/20\varepsilon n} < \frac{1}{2}. \tag{16}$$

A standard calculation shows that (16) holds, and Lemma 2 follows from (15) and (16). □

Now we are ready to complete the proof of Theorem 15. We show that if Maker uses the strategy in Lemma 2, then $H =$"Maker's graph at the end" contains a cycle of $\left(1-e^{-\frac{1}{200\varepsilon}}\right)n$ edges.

Let G = "Maker's graph right after Breaker occupied $\frac{1}{20}\binom{n}{2}$ edges". Assume that there exists a vertex-set $S_1 \subset V(K_n)$ with $|S_1| \leq \frac{1}{3}e^{-\frac{1}{200\varepsilon}}n$ such that $|\Gamma_G(S_1)\setminus S_1| < 2|S_1|$. Throwing away the vertices $\Gamma_G(S_1)\cup S_1$ from G, we get a new graph G_1. Again assume that there exists a vertex-set $S_2 \subset V(G_1)$ with $|S_2| \leq \frac{1}{3}e^{-\frac{1}{200\varepsilon}}n$ such that $|\Gamma_{G_1}(S_2)\setminus S_2| < 2|S_2|$. Throwing away the vertices $\Gamma_{G_1}(S_2)\cup S_2$ from G_1, we get a new graph G_2, and so on. This truncation procedure terminates (say) in t steps: $G_t = G_{t+1} = \cdots$. That is, for any vertex-set $S \subset V(G_t)$ with $|S| \leq \frac{1}{3}e^{-\frac{1}{200\varepsilon}}n$,

$$|\Gamma_{G_t}(S)\setminus S| \geq 2|S|. \tag{17}$$

We claim

$$|V(G_t)| > \left(1-e^{-\frac{1}{200\varepsilon}}\right)n. \tag{18}$$

Indeed, otherwise there is an index $i(\leq t)$ such that at the ith stage of the truncation, the union set $S = S_1\cup\cdots\cup S_i$ *first* satisfies $\frac{1}{3}e^{-\frac{1}{200\varepsilon}}n \leq |S|$, so $\frac{1}{3}e^{-\frac{1}{200\varepsilon}}n \leq |S| \leq \frac{2}{3}e^{-\frac{1}{200\varepsilon}}n < n/4$ and

$$|\Gamma_G(S) \setminus S| < 2|S|,$$

which contradicts property A in Lemma 2.

It follows from (17), (18) and property A that for every vertex-set $S \subset V(G_t)$ with $|S| \leq n/4$,

$$|\Gamma_{G_t}(S) \setminus S| \geq 2|S|. \tag{19}$$

It immediately follows from (19) that G_t is a *connected* graph. We are going to show that Maker can build up a Hamiltonian cycle on the vertex-set $V(G_t)$.

Let P be a path in G_t of maximum length. Inequality (19) ensures that the truncated graph G_t satisfies the condition of Pósa's lemma with $k = n/4$, so (see Lemma 1) $|\operatorname{end}(G_t, P, v_0)| > n/4$ where v_0 is one of the endpoints of P.

Let $\operatorname{end}(G_t, P, v_0) = \{x_1, x_2, \ldots, x_k\}$ $(k > n/4)$, and denote by $P(x_i)$, $1 \leq i \leq k$ a path arising from P by a sequence of Pósa-deformations (see Lemma 1), v_0 is fixed and having other endpoint x_i. By Lemma 1, for every $x_i \in \operatorname{end}(G_t, P, v_{\mathrm{U}})$,

$$|\operatorname{end}(G_t, P(x_i), x_i)| > n/4. \tag{20}$$

Let

$$\operatorname{close}(G_t, P) = \{(x_i, y) : x_i \in \operatorname{end}(G_t, P, v_0), y \in \operatorname{end}(G_t, P(x_i), x_i)\}.$$

By (20) we have $|\operatorname{close}(G_t, P)| > (n/4)^2/2 = n^2/32$. Since at this moment Breaker's graph contains $\frac{1}{20}\binom{n}{2}$ edges, there must exist a previously unselected edge e_1 in $\operatorname{close}(G_t, P)$. Let e_1 be Maker's next move. Then Maker's graph $G_t^{(1)} = G_t \cup \{e_1\}$ contains a cycle of length $|P|$. Moreover, $G_t^{(1)} = G_t \cup \{e_1\}$ is *connected*, thus either $|P| = |V(G_t)|$, and we have a Hamiltonian cycle in the truncated vertex-set, or $G_t^{(1)}$ contains a *longer* path (i.e. a path of length $\geq |P| + 1$).

Let P_1 be a path of maximum length in $G_t^{(1)}$. Repeating the argument above, we obtain that $|\operatorname{close}(G_t^{(1)}, P_1)| > n^2/32$. Since at this moment Breaker's graph contains $\frac{1}{20}\binom{n}{2} + \varepsilon n < n^2/32$ edges, there must exist a previously unselected edge e_2 in $\operatorname{close}(G_t^{(1)}, P_1)$. Let e_2 be Maker's next move. Then Maker's graph $G_t^{(2)} = G_t \cup \{e_1, e_2\}$ contains a cycle of length $|P_1|$. Moreover, $G_t^{(2)} = G_t \cup \{e_1, e_2\}$ is *connected*, thus either $|P_1| = |V(G_t)|$, and we have a Hamiltonian cycle in the truncated vertex-set, or $G_t^{(2)}$ contains a *longer* path (i.e. a path of length $\geq |P_1| + 1$). By repeated application of this procedure, in less than n moves (so the required inequality $\frac{1}{20}\binom{n}{2} + n \cdot \varepsilon n < n^2/32$ holds), Maker's graph will certainly contain a Hamiltonian cycle in the truncated vertex-set $V(G_t)$. Theorem 15 follows. □

Proof of Theorem 16 (a "fake second moment" method). Consider a play according to the rules. Let $x_1, x_2, \ldots, x_i$ be the blue cells in the chessboard selected by Breaker in his first i moves, and let $y_1, y_2, \ldots, y_{i-1}$ be the red cells selected by Maker in his first $(i-1)$ moves, and the question is how to find Maker's optimal ith move y_i. Write

$$X_i = \{x_1, x_2, \ldots, x_i\} \text{ and } Y_{i-1} = \{y_1, y_2, \ldots, y_{i-1}\}.$$

Let A be a line (= row or column) of the $n \times n$ chessboard, and introduce the the "weight":

$$w_i(A) = \left\{ |A \cap Y_{i-1}| - |A \cap X_i| + \frac{\sqrt{n}}{4} \right\}^+$$

where

$$\{\alpha\}^+ = \begin{cases} \alpha, \text{ if } \alpha > 0; \\ 0, \text{ otherwise.} \end{cases}$$

Let y be an arbitrary unselected cell, and write

$$w_i(y) = w_i(A) + w_i(B)$$

where A and B are the row and the column containing y.

Here is Maker's winning strategy: at his ith move he selects that previously unselected cell y for which the maximum of the "weights"

$$\max_{y \text{ unselected}} w_i(y)$$

is attained.

The following total sum is a sort of "variance":

$$T_i = \sum_{2n \text{ lines } A} (w_i(A))^2 .$$

The idea of the proof is to study the behaviour of T_i as $i = 1, 2, 3, \ldots$ and to show that T_{end} is "large".

Remark 1. *The more natural "symmetric" total sum*

$$\sum_{2n \text{ lines } A} (|Y_{i-1} \cap A| - |X_i \cap A|)^2$$

is useless because it can be large if in some line Breaker overwhelmingly dominates. This is exactly the reason why we had to introduce the "shifted and truncated weight" $w_i(A)$.

First we compare T_i and T_{i+1}, that is, we study the effects of the cells y_i and x_{i+1}. We distinguish two cases.

Case 1: the cells y_i and x_{i+1} determine four different lines.
Case 2: the cells y_i and x_{i+1} determine three different lines.

In Case 1, an easy analysis shows that

$$T_{i+1} \geq T_i + 1 \tag{21}$$

except for the "unlikely situation" when $w_i(y_i) = 0$. Indeed,

$$w_i(y_i) = w_i(A) + w_i(B) \geq w_i(x_{i+1}) = w_i(C) + w_i(D),$$

and so

$$T_{i+1} = T_i + 2w_i(y_i) - 2w_i(x_{i+1}) + \{2 \text{ or } 1 \text{ or } 0\} \geq T_i + \{2 \text{ or } 1 \text{ or } 0\}$$

where

$$\{2 \text{ or } 1 \text{ or } 0\} = \begin{cases} 2, \text{ if } w_i(A) > 0, w_i(B) > 0; \\ 1, \text{ if } \max\{w_i(A), w_i(B)\} > 0, \min\{w_i(A), w_i(B)\} = 0; \\ 0, \text{ if } w_i(A) = w_i(B) = 0. \end{cases}$$

Even if the "unlikely situation" occurs, we have at least equality: $T_{i+1} = T_i$. Because y_i was a cell of maximum weight, for x_{i+1}, and for every other unselected cell x, $w_i(x) = 0$.

Similarly, in Case 2,

$$T_{i+1} \geq T_i + 1 \tag{22}$$

except for the following "unlikely situation": $w_i(B) = 0$ where A is the line containing both y_i and x_{i+1}, and B is the other line containing y_i. Even if this "unlikely situation" occurs, we have at least equality: $T_{i+1} = T_i$. Because y_i was a cell of maximum weight, it follows that $w_i(C) = 0$ where C is the other line containing x_{i+1}, and similarly, for every other unselected cell x in line A, $w_i(D_x) = 0$ where D_x is the other line containing x.

If i is an index for which the "unlikely situation" in Case 1 occurs, let unsel(i) denote the set of all unselected cells after Breaker's ith move. Similarly, if i is an index for which the "unlikely situation" in Case 2 occurs, let unsel(i, A) denote the set of all unselected cells after Breaker's $(i + 1)$st move in line A containing both y_i and x_{i+1}, including y_i and x_{i+1}.

If the "unlikely situation" occurs in less than $3n^2/10$ moves (i.e. in less than 60 % of the total time), then we are trivially done. Indeed, then by (21) and (22),

$$T_{end} = T_{n^2/2} \geq \frac{n^2}{5}.$$

Since T_{end} is a sum of $2n$ terms, we have

$$\max_{2n \text{ lines } A} \left(w_{n^2/2}(A)\right)^2 \ge \frac{n^2/5}{2n} = \frac{n}{10}.$$

That is, for some line A,

$$w_{n^2/2}(A) = \left\{ |A \cap Y_{n^2/2-1}| - |A \cap X_{n^2/2}| + \frac{\sqrt{n}}{4} \right\}^+ \ge \sqrt{n/10}$$

where

$$\{\alpha\}^+ = \begin{cases} \alpha, \text{ if } \alpha > 0; \\ 0, \text{ otherwise.} \end{cases}$$

So

$$|A \cap Y_{n^2/2-1}| - |A \cap X_{n^2/2}| \ge \sqrt{n/10} - \frac{\sqrt{n}}{4} > \frac{\sqrt{n}}{16},$$

and Theorem 16 follows.

If the "unlikely situation" in Case 1 occurs in more than $n^2/10$ moves (i.e. in more than 20 % of the time), then let i_0 be the first time when this happens. Clearly

$$|\operatorname{unsel}(i_0)| > 2n^2/10 = n^2/5.$$

It follows that there are at least $(n^2/5)/n = n/5$ distinct columns D containing (at least one) element of unsel(i_0) each. So $w_i(D) = 0$ for at least $n/5$ columns D, that is,

$$|D \cap X_i| - |D \cap Y_{i-1}| \ge \frac{\sqrt{n}}{4}$$

for at least $n/5$ columns D. Therefore, after Breaker's i_0th move,

$$\sum_{n \text{ columns } D} \{|D \cap X_i| - |D \cap Y_{i-1}|\}^+ > \frac{n}{5}\frac{\sqrt{n}}{4}. \tag{23}$$

Since

$$1 + \sum_{n \text{ columns } D} \{|D \cap Y_{i-1}| - |D \cap X_i|\}^+ = \sum_{n \text{ columns } D} \{|D \cap X_i| - |D \cap Y_{i-1}|\}^+,$$

by (23),

$$\sum_{n \text{ columns } D} \{|D \cap Y_{i-1}| - |D \cap X_i|\}^+ \ge \frac{n^{3/2}}{20}.$$

Since the number of terms on the left-side is less than $n - n/5 = 4n/5$, after Breaker's i_0th move we have,

$$\max_D \{|D \cap Y_{i-1}| - |D \cap X_i|\} > \frac{n^{3/2}/20}{4n/5} = \frac{\sqrt{n}}{16}.$$

Obviously Maker can keep this advantage of $\sqrt{n}/16$ for the rest of the game, and again Theorem 16 follows.

Finally, we study the case when the "unlikely situation" of Case 2 occurs for at least $n^2/5$ moves (i.e., for at least 40 % of the time). Without loss of generality, we can assume that there are at least $n^2/10$ "unlikely" indices i when the line A containing both y_i and x_{i+1} is a *row*. We claim that there is an "unlikely" index i_0 when

$$|\operatorname{unsel}(i_0, A)| \geq n/5. \tag{24}$$

Indeed, by choosing y_i and x_{i+1}, in each "unlikely" move the set $\operatorname{unsel}(i, A)$ is decreasing by 2, and because we have n rows, the number of "unlikely" indices i when $\operatorname{unsel}(i, A) < n/5$ is altogether less than $n \cdot \frac{n/5}{2} = n^2/10$.

Now we can finish just like before. We recall that $w_{i_0}(D) = 0$ for those columns D which contain some cell from $\operatorname{unsel}(i_0, A)$ (here A is the row containing both y_{i_0} and x_{i_0+1}). So by (24), $w_{i_0}(D) = 0$ for at least $n/5$ columns D, that is,

$$|D \cap X_i| - |D \cap Y_{i-1}| \geq \frac{\sqrt{n}}{4}$$

for at least $n/5$ columns D. Therefore, after Breaker's i_0th move,

$$\sum_{n \text{ columns } D} \{|D \cap X_i| - |D \cap Y_{i-1}|\}^+ > \frac{n}{5}\frac{\sqrt{n}}{4}. \tag{25}$$

Since

$$1 + \sum_{n \text{ columns } D} \{|D \cap Y_{i-1}| - |D \cap X_i|\}^+ = \sum_{n \text{ columns } D} \{|D \cap X_i| - |D \cap Y_{i-1}|\}^+,$$

by (25),

$$\sum_{n \text{ columns } D} \{|D \cap Y_{i-1}| - |D \cap X_i|\}^+ \geq \frac{n^{3/2}}{20}.$$

Since the number of terms on the left-side is less than $n - n/5 = 4n/5$, after Breaker's i_0th move we have,

$$\max_D \{|D \cap Y_{i-1}| - |D \cap X_i|\} > \frac{n^{3/2}/20}{4n/5} = \frac{\sqrt{n}}{16}.$$

Obviously Maker can keep this advantage of $\sqrt{n}/16$ for the rest of the game, and again Theorem 16 follows. The proof is complete. □

5. Algorithms and Complexity

Perhaps the most interesting applications of game-theoretic ideas are in theoretical computer science. The following is an illustration of this.

Let $\mathbb{N}$ denote, as usual, the set of natural numbers, and let $K_\infty = [\mathbb{N}]^2$ be the set of pairs, that is, an explicit representation of the *infinite complete graph* on the set of natural numbers.

Suppose we are given a 2-coloration (red and blue) of the edges of $K_\infty = [\mathbb{N}]^2$. The problem is to *find* a monochromatic complete subgraph of k vertices, that is, a monochromatic copy of K_k. The well-known Ramsey's theorem implies the *existence* of such a subgraph. At each step we may ask the color of an arbitrary edge and we are interested in the minimum number of questions necessary. The standard proof of Ramsey's theorem, the so-called *ramification* method, guarantees that we can find a monochromatic complete subgraph of size k within 4^k steps (= questions). Here is an outline of the procedure. In the first $2^{2k-2}-1$ steps we test a star, e.g. we ask for the colors of the edges $\{1,2\},\{1,3\},\ldots,\{1,2^{2k-2}\}$. Let $\{1,r\}$, $r \in R$ and $\{1,b\}$, $b \in B$ be the sets of red and blue edges, respectively. The sum of the cardinalities of R and B is precisely $2^{2k-2}-1$, therefore one of them, say R has cardinality at least 2^{2k-3}. Choosing an arbitrary element $s_2 \in R$ we ask the colors of the $2^{2k-3}-1$ edges $\{s_2,r\}, r \in R(r \neq s_2)$ and so on. Finally we obtain a sequence $s_1 = 1, s_2, \ldots, s_{2k-1}$ of vertices such that the color of the edge $\{s_i,s_j\}, i<j$ depends on index i only. Let $c(i) \in \{red, blue\}$ be this color. Then the complete subgraphs induced by the vertex-sets $\{s_i : c(i)\ red\} \cup \{s_{2k-1}\}$ and $\{s_j : c(j)\ blue\} \cup \{s_{2k-1}\}$ are monochromatic. Since one of them has size at least k, we are done. This procedure needs at most $\sum_{i=1}^{2k-2}(2^i-1) < 4^k$ steps (= questions).

Many years ago we proved that no method can perform much faster than this ramification procedure (see [6]).

Theorem 17. *An algorithm which, for any 2-coloration of K_∞, will determine a monochromatic complete subgraph with $k(\geq 3)$ vertices requires more than $2^{k/2}$ steps. Here a step means asking the color of a single edge.*

Theorem 17 was suggested by a well-known result of Erdős [17] stating that there exists a 2-coloration of the complete graph of $const \cdot k2^{k/2}$ vertices which does not contain a monochromatic K_k. The proof of Theorem 4 provides an *efficient algorithmic* proof of Erdős' theorem which was originally proved by a *probabilistic* argument.

Consider the following game-theoretic equivalent of the problem above. Two players, I and II, are playing on the "board" $K_\infty = [\mathbb{N}]^2$. On each move, Player I selects a previously unselected edge of $K_\infty = [\mathbb{N}]^2$ and Player II colors it by red or blue. Player I wins if there exists a monochromatic copy of K_k in the graph selected by him during the play; otherwise Player II wins. That is, the aim of Player II is simply to prevent Player I from achieving his goal. Observe that Theorem 17 is equivalent to the statement:

Player II has a strategy such that I is unable to win within $2^{k/2}$ moves.

Recently Noga Alon found a very simple proof of Theorem 17. We briefly describe his elegant argument. We want to show that Player II can force $r_2(k)/2$ questions (= moves), where $r_2(k)$ is the maximum cardinality of a complete 2-colored graph with *no* monochromatic K_k. This gives, by Erdős' bound $r_2(k) > const \cdot k2^{k/2}$, the lower bound in Theorem 17 (in fact, a sightly better bound with an extra factor of k).

Put $n = r_2(k)$ and let Player II *fix* a 2-coloration of the edges of $K_n = \{1, 2, \ldots, n\}^2$ with *no* monochromatic K_k. We are going to show that Player II can make sure that after $n/2$ moves the colored part of $K_\infty = \mathbb{N}^2$ will be *isomorphic* to a subgraph of his own fixed 2-colored $K_n = \{1, 2, \ldots, n\}^2$ with no monochromatic K_k. So Player I cannot win within $n/2$ moves.

The strategy of Player II goes as follows: let $i_1, i_2, i_3, \ldots, i_l$ be the *vertices* that appear in the questions of Player I (clearly $l \le 2(n/2) = n$, if at most $n/2$ questions have been asked), and assume this is an enumeration of these vertices according to the order they appear in the questions of Player I. Define $f(i_j) = j$ (observe that Player II knows $f(i_j)$ as soon as vertex number i_j is asked, he does not have to wait to the next questions of Player I). Now, when Player I asks for the color of $\{i_j, i_s\}$ Player II answers by telling the color of the edge $\{j, s\}$ in his own fixed 2-colored $K_n = \{1, 2, \ldots, n\}^2$ with no monochromatic K_k. This will guarantee what is needed: up to $n/2$ moves the colored graph we have will be isomorphic by f to a graph with no monochromatic K_k.

Alon's argument proves that any decision tree algorithm needs $r_2(k)/2$ many queries in worst case. By employing the Erdős-Selfridge "derandomization" of Erdős' probabilistic proof for $r_2(k) > const \cdot k2^k$, we get a simple weight-function algorithm (= strategy), and so we can essentially reduce the complexity of the adversary's strategy. The "danger" of this is that the Erdős-Selfridge lower bound might turn out to be much smaller than $r_2(k)$.

Note that our original proof for Theorem 17 was a weight-function strategy, too, but it was more complicated.

Next we study the *parallel matching complexity* of the problem above, that is, we are interested in the minimum number of steps necessary to find a monochromatic K_n, where a step means asking the colors of many point-disjoint edges at once. First we formulate a

Conjecture 3. *A parallel matching algorithm which, for any 2-coloration of the edges of K_∞, will determine a monochromatic K_n requires more than c^n steps (i.e., exponential time), where $c > 1$ is an absolute constant and a step means asking the colors in an arbitrary matching at once.*

This conjecture was suggested by the following easy corollary of the Lovász Local Lemma (see [19]).

Theorem 18. *If G is an arbitrary graph (finite or infinite) with maximum degree at most $2^{n/2}$, then the edges of G can be 2-colored such that there is no monochromatic K_n.*

The remarkable feature of the Local Lemma is that it guarantees the existence of a "needle in a haystack". Indeed, if the number $|G|$ of edges of G is a super-exponential function of n, then the probability of "success" for the standard random 2-coloring is an exponentially small function of $|G|$. So the Local Lemma does not supply even an efficient randomized algorithm.

Recently we managed to convert some of the applications of the Local Lemma into polynomial time sequential algorithms at the cost of a weaker constant factor in the exponent (see [8]). The algorithmic version of Theorem 18 goes as follows (note that its proof heavily uses game-theoretic ideas, especially "derandomization" and the trick of "BIG GAMES" and "small games").

Theorem 19. *Let G be a finite graph with maximum degree at most $2^{n/96}$. Then there is a deterministic sequential algorithm with running time $|G|^{const}$ which produces a 2-coloring of the edges of G such that there is no monochromatic K_n.*

A natural approach to attack the Conjecture above is to try to convert the algorithmic proof of Theorem 19 to a game-theoretic strategy. However, some surprising (or natural) technical obstacles prevented us from realizing this heuristic.

On the other hand, we could prove the following interesting partial result: the Conjecture above holds if a *step* means asking the colors in a matching of size $\leq 2^{n^2/10}$ *at once* (see [9]).

Theorem 20. *A parallel matching algorithm which, for any 2-coloration of the edges of K_∞, will determine a monochromatic K_n requires more than $const \cdot 2^{n/40}$ steps (i.e., exponential time), where a step means asking the colors in an arbitrary matching of size at most $2^{n^2/10}$ at once.*

We emphasize that $2^{n^2/10}$ is *super-exponential*, that is, $2^{n^2/10}$ is asymptotically much bigger than 4^n (which is the number of edges necessary to get a monochromatic K_n via the "ramification" method described above). This justifies our intuition that independent edges do not help too much to build a monochromatic K_n up.

It is worth to note that, in contrast to the Conjecture (and to Theorem 20), where the parallel *matching* complexity is *exponential*, the parallel *star* complexity of the Ramsey theorem is *linear*. Indeed, the ramification method requires the testing of at most $2n-1$ *stars* (each having at most exponentially many edges).

Consider the following game-theoretic equivalent of Theorem 20. Two players, I and II, are playing on the "board" $K_\infty = [\mathbb{N}]^2$. On each move, Player I selects $m = 2^{n^2/10}$ previously unselected point-disjoint edges of $K_\infty = [\mathbb{N}]^2$ (i.e., an m-matching) and Player II colors it by red and blue. Player I wins if there exists a monochromatic copy of K_n in the graph selected by him during the play; otherwise Player II wins. That is, the aim of Player II is simply to prevent Player I from achieving his goal. Theorem 20 is equivalent to the statement: Player II has a strategy such that Player I is unable to win this m-matching game within $const \cdot 2^{n/40}$ moves.

References

1. N. Alon, and J. Spencer, *The Probabilistic Method*, Academic Press, New York, 1992.
2. J. Baumgartner, F. Galvin, R. Laver and R. McKenzie, Game theoretic versions of partition relations, in: *Colloq. Math. Soc. János Bolyai "Infinite and Finite sets"* Keszthely, Hungary, 1973, 131–135.
3. J. Beck, On positional games, *Journal of Combinatorial Theory, ser. A* **30** (1981), 117–133.
4. J. Beck, Van der Waerden and Ramsey type games, *Combinatorica* **2** (1981), 103–116.
5. J. Beck, Remarks on positional games - Part I, *Acta Math. Acad. Sci. Hungarica* (1–2) **40** (1982), 65–71.
6. J. Beck, There is no fast method for finding monochromatic complete subgraphs, *Journal of Combinatorial Theory, ser. B* **34** (1983), 58–64.
7. J. Beck, Random graphs and positional games on the complete graph, *Annals of Discrete Math.* **28** (1985), 7–13.
8. J. Beck, An algorithmic approach to the Lovász Local Lemma. I., *Random Structures and Algorithms* **2** (1991), 343–365.
9. J. Beck, Parallel matching complexity of Ramsey's theorem, manuscript (1992).
10. J. Beck, Deterministic graph games and a probabilistic intuition, manuscript (1993).
11. J. Beck and L. Csirmaz, Variations on a game, *Journal of Combinatorial Theory, ser. A* **33** (1982), 297–315.
12. C. Berge, Sur les jeux positionelles, *Cahiers Centre Études Rech. Opér.* **18** (1976).
13. E. R. Berlekamp, J. H. Conway and R. K. Guy, *Winning Ways*, Academic Press, London, 1982.
14. B. Bollobás, *Random Graphs*, Academic Press, London, 1985.
15. V. Chvátal and P. Erdős, Biased positional games, *Annals of Discrete Math.* **2** (1978), 221–228.
16. V. Chvátal, V. Rödl, E. Szemerédi and W. T. Trotter, The Ramsey number of a graph with bounded maximum degree, *Journal of Combinatorial Theory, series B* **34** (1983), 239–243.
17. P. Erdős, Some remarks on the theory of graphs, *Bull. Amer. Math, Soc.* **53** (1947), 292–294.
18. P. Erdős, On a combinatorial problem, I, *Nordisk. Mat. Tidskr.* **11** (1963), 5–10.
19. P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, in: *Infinite and Finite Sets* (eds.: A. Hajnal et al.), *Colloq. Math. Soc. J. Bolyai*, **11**, North-Holland, Amsterdam, 1975, 609–627.

20. P. Erdős and J. Selfridge, J, On a combinatorial game, *Journal of Combinatorial Theory, ser. A* **14** (1973), 298–301.
21. A. W. Hales and R. I. Jewett, On regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222–229.
22. J. Komlós and E. Szemerédi, Hamilton cycles in random graphs, in: *Proc. of the Combinatorial Colloquium in Keszthely*, Hungary, 1973, 1003–1010.
23. A. Lehman, A solution to the Shannon switching game, *SIAM Journ. Appl. Math.* **12** (1964),687–725.
24. L. Lovász, *Combinatorial Problems and Exercises*, North-Holland and Akadémia Kiadó, 1979.
25. C. St. J. A. Nash-Williams, Edge-disjoint spanning trees of finite graphs, *Journ. London Math. Soc.* **36** (1961), 445–450.
26. L. Pósa, Hamilton circuits in random graphs, *Discrete Math.* **14** (1976), 359–364.
27. L. A. Székely, On two concepts of discrepancy in a class of combinatorial games, in: *Colloq. Math. Soc. János Bolyai 37 "Finite and Infinite Sets"* Eger, Hungary, 1981, North-Holland, 679–683.
28. W. T. Tutte, On the problem of decomposing a graph into n connected factors, *Journ. London Math. Soc.* **36** (1961), 221–230.

On Some Hypergraph Problems of Paul Erdős and the Asymptotics of Matchings, Covers and Colorings

Jeff Kahn

J. Kahn (✉)
Department of Mathematics and RUTCOR, Rutgers University,
New Brunswick, NJ 08903, USA
e-mail: jkahn@math.rutgers.edu

1. Introduction

This article summarizes progress on several old hypergraph problems of Paul Erdős and a few questions to which they led. Quite unexpectedly, there turned out to be substantial connections between the problems under discussion, surely some indication (if any were needed) that Erdős' questions were the "right" ones. Here's a quick synopsis.

The story basically begins about 10 years ago, with Vojta Rödl's beautiful proof [80] of the "Erdős-Hanani" Conjecture (Sect. 4). His proof was based on a powerful "semirandom" or "guided-random" approach. (I wish there were a better name for this.) A similar method had earlier been used in a less precise context by Ajtai, Komlós and Szemerédi [1] and Komlós, Pintz and Szemerédi [70]. Substantial extensions of Rödl's work were subsequently achieved by Frankl and Rödl [38], Pippenger (see [87] or [42]), and Pippenger and Spencer [77] (see Sects. 4 and 6).

Most of the work described in this paper had its beginnings in attempts to apply these ideas to prove a nonlinear lower bound on the function $n(r)$ of Erdős and Lovász discussed in Sect. 3. In the event, $n(r)$ turned out to be linear, though discovering this would certainly not have been possible if the results of those initial attempts (see Sect. 5) had not suggested where—or at least where *not*—to look for examples.

In the meantime, an understanding of the above-mentioned results, particularly [77], had led to a proof of the "asymptotic correctness" of the well-known Erdős-Faber-Lovász Conjecture (Sect. 2), which proof led eventually to a much stronger result (Theorem 12) on the asymptotic behavior of the list-chromatic index for hypergraphs; and further efforts to prove $n(r)/r \to \infty$ had suggested the conjecture which eventually became the main result of Sect. 5. (Theorem 9), and led in its turn to the investigations mentioned in Sects. 7 and 8.

In this paper we mainly try to give an overview of these developments and connections, with discussion of proofs limited to hints at most. More detailed accounts of some of the material—especially more serious discussions of the

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_22,

semirandom method—may be found in [42] and [60]. (I should also say that various bits and pieces of this article are borrowed from [60–62].)

Terminology

Throughout we use $\mathcal{H}$ to denote a hypergraph on vertex set V. For further hypergraph background see, e.g., [42] or [9].

The *degree* (in $\mathcal{H}$) of $x \in V$ is the number of edges of $\mathcal{H}$ containing x, and is denoted $d_{\mathcal{H}}(x)$ or simply $d(x)$. Similarly, $d(x, y)$ denotes the number of edges containing both of the vertices x, y and $d(X)$ the number of edges containing all vertices of $X \subseteq V$. We write $D(\mathcal{H})$ for the largest degree in $\mathcal{H}$. A hypergraph is *D-regular* if each of its vertices has degree D.

A hypergraph is *intersecting*, resp. *simple* (or *nearly-disjoint*, but we won't use this), if any two of its edges have at least, resp. at most, one vertex in common.

For $X, Y \in \mathcal{H} \cup V$, the *distance* from X to Y, denoted $\Delta(X, Y)$, is the least m for which there exists a sequence $X = X_0, \ldots, X_m = Y$ from $\mathcal{H} \cup V$ such that for each i, X_{i-1} is an element of X_i or vice versa.

A *matching* of $\mathcal{H}$ is a collection of pairwise disjoint edges, and the size of a largest such collection, denoted $\nu(\mathcal{H})$, is the *matching number* of $\mathcal{H}$. We write $\mathcal{M}(\mathcal{H})$ for the set of matchings of $\mathcal{H}$.

A *vertex cover* (clearer would be "cover of edges by vertices") of $\mathcal{H}$ is a set of vertices meeting every edge of $\mathcal{H}$, while an *edge cover* is a collection of edges whose union is V. Either of these may be shortened to "cover" if there seems no danger of confusion. The *vertex* and *edge cover numbers* of $\mathcal{H}$ are the minimum sizes of its vertex and edge covers, and are denoted $\tau(\mathcal{H})$ and $\rho(\mathcal{H})$.

Each of ν, τ, ρ has a fractional counterpart, obtained by regarding the object in question as the solution of an integer program and taking the linear relaxation thereof. Thus a *fractional (edge) cover*—the only one of the three needed here—is a function $t : \mathcal{H} \to \mathbf{R}^+$ satisfying

$$\sum_{A \ni x} t(A) \geq 1 \quad \forall x \in V, \tag{1}$$

and the *fractional (edge) cover number* is

$$\rho^*(\mathcal{H}) = \min\{\sum t(A) : t \text{ a fractional edge cover of } \mathcal{H}\}.$$

We also say that $t : \mathcal{H} \to \mathbf{R}^+$ is a *fractional tiling* if equality holds in (1).

The *chromatic index* (or *edge coloring number*) of $\mathcal{H}$, denoted $\chi'(\mathcal{H})$, is the least t for which there is a *"coloring"* $\sigma : \mathcal{H} \to [t]$ which is *proper* in the usual sense that $\sigma(A) \neq \sigma(B)$ whenever A, B are distinct, nondisjoint edges. Equivalently, $\chi'(\mathcal{H})$ is the least size of a collection of matchings whose union is $\mathcal{H}$. We also write $\phi(\mathcal{H})$ for the greatest size of a collection of pairwise disjoint covers contained in $\mathcal{H}$.

These too have fractional versions, of which we only need the *fractional chromatic index* of $\mathcal{H}$, denoted (unfortunately) $\chi'^*(\mathcal{H})$, and defined as the minimum value of

$$\sum_{M\in\mathcal{M}} f(M)$$

over $f:\mathcal{M}\to\mathbf{R}^+$ satisfying

$$\sum_{A\in M\in\mathcal{M}} f(M)\geq 1 \qquad \forall A\in\mathcal{H}.$$

Finally we need to say a little about asymptotic notation. For nonnegative f,g we use $f\sim g$ and $f\lesssim g$ for "$f/g\to 1$" and "$\limsup f/g\leq 1$", with limits taken as some relevant parameter tends to infinity. We also write $f=_\varepsilon g$ for $(1+\varepsilon)^{-1}<f/g<1+\varepsilon$. As usual we use $f=O(g)$, $f=o(g)$ and $f=\omega(g)$ for (respectively) $\sup(f/g)<\infty$, $f/g\to 0$ and $f/g\to\infty$.

We adopt the "uniformity convention" of [77], viz: any limiting statement involving one or more free variables ranging over vertices, edges or hypergraphs is understood to hold uniformly with respect to all possible choices of these variables, as some specified numerical parameter tends to infinity. (See Theorem 7 for a first instance of this.)

2. The Erdős-Faber-Lovász Conjecture

To avoid trivialities, hypergraphs in this section are assumed to have no singleton edges.

The celebrated Erdős-Faber-Lovász Conjecture may be stated as follows (see [53]):

Conjecture 1. *Any simple hypergraph $\mathcal{H}$ on n vertices has chromatic index at most n.*

Erdős has for many years listed this as one of his "three favorite combinatorial problems" (the other two being the Δ-system problem of Erdős and Rado, and the problem of Erdős and Lovász described in Sect. 3), and currently offers \$500 for its resolution (see, e.g., [29]).

Notice first of all that the Conjecture is sharp in the case $\mathcal{H}$ is either

(a) A projective plane or degenerate projective plane (the latter being the hypergraph with vertex set $\{0,1,\ldots,n-1\}$ and edge set $\{\{0,1\},\ldots,\{0,n-1\},\{1,\ldots,n-1\}\}$), or
(b) A complete graph on n vertices, n odd.

(Sufficiently minor modifications of (b) also give equality.)

On the other hand:

(c) For intersecting $\mathcal{H}$, Conjecture 1 is just the de Bruijn-Erdős Theorem [20], which says that if $|A \cap B| = 1$ for all distinct $A, B \in \mathcal{H}$, then $|\mathcal{H}| \leq n$ (with equality only for $\mathcal{H}$ as in (a)).

(d) For graphs, Conjecture 1 is contained in Vizing's Theorem [90] stating that the chromatic index of a simple graph of maximum degree D is at most $D+1$. (Of course this special case—Vizing's Theorem for complete graphs—is easily proved directly. On the other hand, as observed, e.g., by Meyniel (unpublished), Berge [10] and Füredi [41], it seems likely that the bound in Conjecture 1 can be replaced by $\max_{x \in V} |\cup_{A \ni x} A|$, which for graphs reduces to Vizing's Theorem in full.)

Graphs and intersecting hypergraphs are in some sense the extreme cases of Conjecture 1. One of the problem's most appealing aspects is that it has proved so intractable despite being manageable at these extremes, and apparently less accurate between them.

Bounds

The history of results on Conjecture 1 is rather brief, surely more an indication of the difficulty of the problem than of any lack of attempts to resolve it. The first significant progress was made by P. Seymour, who showed

Theorem 1 (**[84]**). *If $\mathcal{H}$ is simple on n vertices, then $\nu(\mathcal{H}) \geq |\mathcal{H}|/n$, with equality only in the cases (a) and (b).*

Note this is immediate from Conjecture 1. An intermediate statement was conjectured in [84] and proved in [68]:

Theorem 2. *If $\mathcal{H}$ is simple on n vertices, then $\chi'^* \leq n$.*

Equivalently (by LP-duality),

$$\forall f : \mathcal{H} \to \mathbf{R}^+ \exists M \in \mathcal{M}(\mathcal{H}) \text{ such that}$$
$$\sum\{f(A) : A \in M\} \geq n^{-1} \sum\{f(A) : A \in \mathcal{H}\}. \quad (2)$$

(So taking $f \equiv 1$ we recover Theorem 1.) The proof of Theorem 2 turned out to be much simpler than that of Theorem 1 because it was possible to exploit properties of a worst f in (2).

It seems to have been noticed by several people that a greedy coloring of edges of $\mathcal{H}$ in any nonincreasing size order requires at most $2n-3$ colors. In the absence of edges of size 2 this bound shrinks to about $3n/2$, and Chang and Lawler [21] showed how to modify the greedy procedure to achieve the same bound (precisely, $\chi'(\mathcal{H}) \leq \lceil 1.5n - 2 \rceil$) in general. That Conjecture 1 is at least asymptotically correct was subsequently proved in [58]:

Theorem 3. *If $\mathcal{H}$ is simple on n vertices, then $\chi'(\mathcal{H}) < n + o(n)$.*

The proof of this is based on the "semirandom" method discussed below (see especially Sect. 6), and actually gives $\chi'(\mathcal{H}) < (1+o(1)) \max_{x\in V} |\cup_{A\ni x} A|$ (c.f. (d) above).

Digression: Borsuk and Larman

There's at least a formal similarity between Conjecture 1 and the following problem of Larman [73]. (A hypergraph is *t-intersecting* if any two of its edges share at least t vertices.)

Conjecture 2. *If $\mathcal{H}$ is a t-intersecting hypergraph on n vertices, then $\mathcal{H} = \mathcal{H}_1 \cup \cdots \cup \mathcal{H}_n$ with each $\mathcal{H}_i$ $(t+1)$-intersecting.*

This is motivated by, and for uniform $\mathcal{H}$ is a special case of "Borsuk's Conjecture" that every bounded set in $\mathbf{R}^d$ is the union of $d+1$ sets of smaller diameter ([19]; see [17, 24, 46] for further discussion).

Conjecture 2 and Borsuk's Conjecture were recently disproved in [65]. (We again come back to Erdős. The disproof is a simple application of a Theorem of Frankl and Wilson [39] which has its roots in the de Bruijn-Erdős Theorem and Fisher's inequality [36]. Erdős was also one of the first to suggest that Borsuk's Conjecture might be false [30].)

The case $t = 1$ of Conjecture 2 remains open (and interesting). Here Füredi and Seymour (see [31, 68]) proposed the stronger conjecture that one may use $\mathcal{H}_i$'s of the form $\{A \in \mathcal{H} : A \supseteq \{x, y\}\}$ for appropriate vertex pairs $\{x, y\}$. This too turns out to be false [64], though a simple disproof would still be welcome. (Curiously, the random construction of [64] takes just a few lines to describe, but as of now about 20 pages to justify.)

3. A Problem of Erdős and Lovász

In a seminal paper [33], Erdős and Lovász pose the problem of estimating, for positive integer r,

$$n(r) := \min\{|\mathcal{H}| : \mathcal{H} \text{ } r\text{-uniform, intersecting, with } \tau(\mathcal{H}) = r\}.$$

That is, with how few intersecting r-edges can one force $\tau = r$? While the conditions here may at first glance seem a little arbitrary, notice that we must require "intersecting," or some substitute, to make the question nontrivial, and that once we assume "r-uniform, intersecting," we are just asking that τ be as large as possible subject to these conditions. Thus the Erdős-Lovász problem is a quite natural way of making concrete the vague question, how can one economically force large cover number in a hypergraph with large edges?

Erdős and Lovász showed (writing $\mathcal{P}_r$ for any projective plane of order $r-1$)

$$n(r) \geq 8r/3 - 3 \text{ for all } r, \tag{3}$$

and

$$n(r) \leq 4r^{3/2} \log r \text{ if there exists a } \mathcal{P}_r, \tag{4}$$

the second inequality being an immediate consequence of

Theorem 4 ([33]). *If $\mathcal{H}$ is a set of $m \geq 4r^{3/2} \log r$ random lines from $\mathcal{P}_r$, then $Pr(\tau(\mathcal{H}) = r) \to 1 (r \to \infty)$.*

They also conjectured that the correct rate of growth here should be $r \log r$. This was shown in [59]:

Theorem 5. *If $\mathcal{H}$ is a set of $m \geq 22r \log r$ random lines from $\mathcal{P}_r$, then $Pr(\tau(\mathcal{H}) = r) \to 1 (r \to \infty)$.*

Of course this also gives the corresponding improvement in (4). The correct value of m here is probably about $3r \log r$; see [59] or [60] for a precise statement.

The problem from Erdős' "list of three" was to decide whether

$$n(r) = O(r). \tag{5}$$

This was done in [61]. The answer—that (5) *is* true—was probably not what most people expected. (Certainly it wasn't what the author expected.)

We don't have space to go into the construction here, but want to mention that one ingredient is the work of Chowla, Erdős and Straus [23] on the existence of large sets of mutually orthogonal Latin squares. See also the discussion following Theorem 9 for a small additional hint at what's involved.

The constant in (5) is so far not very good. Quite surprisingly the best lower bound is still (3), though I feel quite certain this could be improved somewhat via the ideas of [57, 66] discussed in Sect. 5.

Meyer's Problem

In connection with $n(r)$, let us just briefly mention a related problem of similar vintage due to J.-C. Meyer [76]. Meyer defined

$$m(r) = \min\{|\mathcal{H}| : \mathcal{H} \text{ a maximal intersecting}, r - \text{uniform hypergraph}\}$$

and conjectured that $m(r) \geq r^2 - r + 1$ (projective planes being the obvious examples). This was disproved by Füredi [40], who showed

$$m(r) \leq 3r^2/4 \text{ if there exists a } \mathcal{P}_{r/2+1}. \tag{6}$$

On the other hand, despite a fair amount of subsequent effort, it remains quite unclear how $m(r)$ ought to grow. As of now the best results are (from [11, 18] and [25] respectively)

$$m(r) \leq r^5 \ \forall r, \tag{7}$$

$$m(r) < r^2/2 + O(r) \text{ if there exists a } \mathcal{P}_r, \tag{8}$$

and

$$m(r) \geq 3r \text{ for } r \geq 4. \tag{9}$$

(The lower bound is a slight improvement on $m(r) \geq 8r/3 - 3$, which follows from (3), since trivially $m(r) \geq n(r)$. See also [60] for a proposed construction for $m(r) = o(r^2)$ when there exists a $\mathcal{P}_r$.)

While the examples for $n(r)$ described above don't seem to give anything for $m(r)$, they seem to me strongly to suggest the truth of

Conjecture 3. *$m(r) = O(r)$.*

4. The Erdős-Hanani Conjecture and Asymptotics of Packing and Covering Problems

Both Theorem 3 and, in a sense, the proof of (5) had their roots in yet another Erdős problem, the so-called "Erdős-Hanani Conjecture" of 1963, and in Rödl's beautiful and seminal proof thereof. Here and in the next two sections we outline work which grew out of Rödl's Theorem. As mentioned earlier, much of this material, and in particular the powerful "semirandom" approach underlying it all, was discussed at some length in [42, 60], so we will be pretty brief here, especially as regards the proofs.

The Erdős-Hanani Conjecture

For positive integer t, say a family $\mathcal{F}$ of subsets of a set V is a *t-packing* (resp. *t-cover*) if each t-subset of V is contained in at most (resp. at least) one member of $\mathcal{F}$. For $2 \leq t < k < v = |V|$, let $P(v,k,t)$ (resp. $C(v,k,t)$) denote the size of a largest t-packing (resp. smallest t-cover) of k-sets in V.

Erdős and Hanani [32] proved that the obvious bounds

$$P(v,k,t) \leq \binom{v}{t}\binom{k}{t}^{-1} \leq C(v,k,t) \tag{10}$$

are asymptotically tight for $t = 2$ and any fixed k, and conjectured the same result for every t and k. This is Rödl's Theorem:

Theorem 6 ([80]). *For every fixed t and k,*

$$P(v,k,t) \sim \binom{v}{t}\binom{k}{t}^{-1} \sim C(v,k,t).$$

(This is an asymptotic version of a well-known conjecture in the theory of block designs which states that the bounds (10) are exact for large enough v satisfying the obvious necessary conditions

$$\binom{k-i}{t-i} \Bigm| \binom{v-i}{t-i} \quad \text{for } 0 \le i \le t-1.$$

For $t = 2$ this was proved by R. M. Wilson in the early 1970s [93], but for $t \ge 3$ a proof still appears remote.)

In other language, Theorem 6 gives the asymptotics of the matching and edge cover numbers of the hypergraph $\mathcal{H} = \left\{ \binom{K}{t} : K \in \binom{V}{k} \right\}$ on the vertex set $\binom{V}{t}$. In fact, as shown by P. Frankl and Rödl [38], Rödl's Theorem is just one instance of a remarkably general packing and covering phenomenon for hypergraphs with bounded edge sizes. An even stronger and cleaner version of their Theorem was proved by N. Pippenger (unpublished; for the original proof see [87] or [42]):

Theorem 7. *Let k be fixed and $\mathcal{H}$ a k-uniform D-regular hypergraph on n vertices satisfying*

$$d(x,y) < o(D) \textit{ for all distinct vertices } x,y. \tag{11}$$

Then

$$\nu(\mathcal{H}) \sim n/k \sim \rho(\mathcal{H}).$$

(The Frankl-Rödl Theorem differs from Theorem 7 in requiring an explicit bound (roughly $D/(\log|V|)^3$) on pairwise degrees. Incidentally, Theorem 7 is our first use of the "uniformity convention" (see Terminology): limits are taken as $D \to \infty$, with convergence uniform in x, y and $\mathcal{H}$.)

The "semirandom" Approach

Joel Spencer remarks in [87] that the Erdős-Hanani Conjecture always seemed a natural candidate for a probabilistic proof. And the proof *was* probabilistic. . .

A natural way to try to prove Theorem 7, say for matchings, would be as follows. Let $\mathcal{H}_0 = \mathcal{H}$, $M_0 = \emptyset$, and for $i = 1, \ldots$ do

(i) Choose A_i uniformly at random from $\mathcal{H}_{i-1}$,
(ii) Set $M_i = M_{i-1} \cup \{A_i\}$ and $\mathcal{H}_i = \{A \in \mathcal{H}_{i-1} : A \cap A_i = \emptyset\}$.

When $\mathcal{H}_i = \emptyset$ we stop and take M_i to be our matching.

Most likely this procedure does work (e.g., in the sense that the random matching it produces has expected size asymptotic to n/k), but I don't think anyone knows how to show this at the moment. Rödl's fundamental insight (translated to the proof of Theorem 7) was that we *can* do the analysis if at each stage we choose a small but fixed (positive) proportion of the desired matching M, rather than just one edge.

To say this a little more precisely, we switch from matchings to covers for a while. Thus we want to show $\rho(\mathcal{H}) < (1+\delta)n/k$ for $\delta > 0$ fixed, $\mathcal{H}$ as in Theorem 7, and sufficiently large D.

Fix $\varepsilon > 0$ small relative to δ. Let $\mathcal{H}_0 = \mathcal{H}$, $V_0 = V$, and iterate the following procedure for i running from 1 to about $\varepsilon^{-1}\log(1/\delta)$. Let $\mathcal{K}_i$ be a random subset of $\mathcal{H}_{i-1}$ chosen according to

$$Pr(A \in \mathcal{K}_i) = \varepsilon/D_{i-1}$$

(for D_i see below), these events mutually independent. Set

$$V_i = V_{i-1} \setminus \bigcup\{A : A \in \mathcal{K}_i\}, \qquad \mathcal{H}_i = \{A \in \mathcal{H}_{i-1} : A \subseteq V_i\}.$$

After the specified number of iterations we add to $\cup\mathcal{K}_i$ one edge containing x for each $x \in V$ not covered by $\cup\mathcal{K}_i$, and claim this (usually) gives the desired cover. Of course what needs to be shown is that $|\cup \mathcal{K}_i|$ is typically about n/k, while $|V \setminus \cup\{A : A \in \cup\mathcal{K}_i\}|$ is small relative to n.

The key to the success of this approach is that we can understand how various relevant quantities—$|\mathcal{K}_i|, |V_i|, |\mathcal{H}_i|$, and especially *degrees* in $\mathcal{H}_i$—*ought* to evolve, and, moreover, show that they *do* typically evolve approximately as they ought. In particular, each "residual" hypergraph $\mathcal{H}_i$ will be close to regular, meaning most of its vertices will have degree close to some (predictable) D_i.

To see why this might be true, suppose we fix $x \in V_{i-1}$ and condition on $\{x \in V_i\}$ (that is on $\{x \in A \in \mathcal{H}_{i-1} \Rightarrow A \notin \mathcal{K}_i\}$). Then writing X_A for the indicator of $\{A \in \mathcal{H}_i\}$, $d_{\mathcal{H}_i}(x) = \sum\{X_A : x \in A \in \mathcal{H}_{i-1}\}$ is usually the sum of about D_{i-1} Bernoulli random variables whose expectations are, because of the approximate regularity of $\mathcal{H}_{i-1}$ and (11), about $(1 - \varepsilon/D_{i-1})^{(k-1)D_{i-1}}$. Moreover, again because of (11), there is considerable independence among these random variables, enough to enable us to say (via Chebyshev's inequality) that $d_{\mathcal{H}_i}(x)$ is likely to be close to its expectation.

Actual implementation of this rough description requires considerable care. In particular, it does take some thought to convince oneself that the various estimates don't deteriorate unacceptably over the specified number of iterations; but we won't go into this here.

For the number of iterations, note that the "natural" value of $Pr(x \in V_i | x \in V_{i-1})$ is about $(1-\varepsilon/D_{i-1})^{D_{i-1}} \approx e^{-\varepsilon}$, so that $\varepsilon^{-1}\log(1/\delta)$ iterations should reduce the number of vertices to about $\delta|V|$.

Remarks.

1. A technical but important point is that, if n is large relative to D we cannot guarantee that *all* degrees in $\mathcal{H}_i$ are close to D_i. It was precisely in the handling of this point that Pippenger improved on [38].
2. For the matching portion of Theorem 7 we may dispense with the final augmentation of $\cup\mathcal{K}_i$ and simply take our matching M to consist of all isolated edges of $\cup\mathcal{K}_i$. The number of edges which this removes from

$\mathcal{K}_i$ should be about $\varepsilon|\mathcal{K}_i|$, an acceptable loss. Actually the two parts of Theorem 7 are easily seen to be equivalent, but for the proof of Theorem 10 below one wants a procedure for generating a nicely behaved *random* matching; see Theorem 11. The procedure just described—essentially that of [77]—is an improved version of Pippenger's original proof designed to accomplish this.

5. Fractional Versus Integer

As stated earlier, the starting point for most of the work discussed here was the realization that something like Theorem 7 could be used to try to prove $n(r)/r \to \infty$. In this section we give a little indication of this connection and sketch the work (other than [61]) that evolved most directly from this attempt.

Connection with $n(r)$

For $t : \mathcal{H} \to \mathbf{R}^+$, let $t(\mathcal{H}) = \sum\{t(A) : A \in \mathcal{H}\}$, define $\bar{t} : 2^V \to \mathbf{R}^+$ by

$$\bar{t}(A) = \sum\{t(B) : B \supseteq A\},$$

and set

$$\alpha_i(t) = \max\{\bar{t}(W) : W \subseteq V, |W| = i\}.$$

For example, if $\mathcal{H}$ is r-regular, then for the fractional tiling $t \equiv 1/r$ we have $\alpha_2(t) = r^{-1}\max\{d(x,y)\}$ and (11) (with r replacing D) is equivalent to

$$\alpha_2(t) \to 0, \tag{12}$$

so that Theorem 7 is contained in

Theorem 8 ([**57**])**.** *Let k be fixed, $\mathcal{H}$ a k-bounded hypergraph, and $t : \mathcal{H} \to \mathbf{R}^+$ a fractional cover. Then*

$$\rho(\mathcal{H}) \lesssim t(\mathcal{H}) \quad (\alpha_2(t) \to 0).$$

(A similar result holds for matchings, but we confine ourselves here to covers.)

To see the connection with the Erdős-Lovász problem, we dualize: $n(r)$ is the least number of vertices in an r-regular hypergraph satisfying

$$d(x,y) > 0 \text{ for all distinct vertices } x, y \tag{13}$$

and having edge cover number r. Thus the following consequence of Theorem 8 is relevant.

Corollary 1. *Suppose $\mathcal{H}$ is r-regular with at most cr vertices, c fixed, $d(x,y) > 0$ for all $x, y \in V$ and*

$$\max\{d(x,y) : x, y \in V, x \neq y\} = o(r).$$

Then $\rho(\mathcal{H}) < (c/(c+1) + o(1))r$.

Or undualized:

Corollary 2. *Suppose* $\mathcal{H}$ *is* r*-uniform, intersecting, of size at most* cr, c *fixed, and satisfies*

$$\max\{|A \cap B| : A, B \in \mathcal{H}, A \neq B\} = o(r). \tag{14}$$

Then $\tau(\mathcal{H}) < (c/(c+1) + o(1))r$.

After a preliminary step which eliminates large edges, the connection between Theorem 8 and Corollary 1 is provided by the observation that if $\mathcal{H}$ is r-regular with $n \leq cr$ vertices and satisfies (13), then the function $t : \mathcal{H} \to \mathbf{R}^+$ given by

$$t(A) = |A|/(n + r - 1) \tag{15}$$

is a fractional cover of total weight

$$\sum_{A \in \mathcal{H}} t(A) = nr/(n + r - 1) \approx nr/(n + r) \leq cr/(c + 1). \tag{16}$$

Notice also that larger pairwise degrees—corresponding to larger intersection sizes in the original formulation—will tend to give even smaller fractional cover number, suggesting that the best hope for proving (5) should indeed be via something akin to the projective plane based constructions of Theorems 4 and 5. But the above results say that one cannot prove (5) with $\mathcal{H}$'s in which all edge intersections are small.

This seemed for quite a while to support the opinion that $n(r)/r \to \infty$. But the correct lesson, very roughly, was that one should allow a few strategically placed small sets X with large $d(X)$. This, it turns out, can be done in such a way that the value of the fractional cover doesn't shrink too much, but we lose Theorem 8 entirely.

The proof of Theorem 8 is similar to that of Theorem 7. At each stage we have some fractional tiling t_{i-1} of the remaining hypergraph $\mathcal{H}_{i-1}$ which guides the choice of $\mathcal{K}_i$: we take each $A \in \mathcal{H}_{i-1}$ to be in $\mathcal{K}_i$ with probability $\varepsilon t_{i-1}(A)$.

It's then necessary to update t_{i-1} in addition to V_{i-1} and $\mathcal{H}_{i-1}$. A nice bonus of the more general framework is that, because we are not restricted to uniform hypergraphs, the difficulty mentioned in Remark 1 at the end of Sect. 4 here essentially takes care of itself. Our random procedure will produce a hypergraph $\mathcal{G} \subseteq \mathcal{H}_{i-1}$ and approximate fractional tiling s. We can then (with high probability) replace $\mathcal{G}$ by some $\mathcal{H}_i \leq \mathcal{G}$ (meaning each edge of $\mathcal{H}_i$ is contained in an edge of $\mathcal{G}$) and s by a fractional tiling t_i of $\mathcal{H}_i$ such that $t_i(\mathcal{H}_i) \approx s(\mathcal{G})$. In particular, we begin each iteration with a fractional tiling, the fractional analogue of a *regular* $\mathcal{H}_{i-1}$.

Local Behavior

The work in [66] again grew out of attempts to push the approach of Theorem 8 to prove $n(r)/r \to \infty$. The general idea was that it should be possible to relax (12) to a requirement that "locally"—that is, on small sets of vertices—one can find ordinary (integer) covers which mimic the fractional cover t. A way to make this precise is as follows.

For $t : \mathcal{H} \to \mathbf{R}^+$ and any $X \subseteq V$, define $t|_X : 2^X \to \mathbf{R}^+$ by

$$t|_X(A) = \sum\{t(B) : B \cap X = A\}.$$

Write $MP(X)$ for the *matching polytope* of X:

$$MP(X) = \text{conv}\{1_M : M \text{ a matching of } 2^X\}.$$

Denote by $b(t)$ the largest b such that for any $X \subseteq V$ with $|X| \leq b$ we have $t|_X \in MP(X)$. In place of (12) we then require that $b(t) \to \infty$. (Note this is weaker than (12).)

Suppose for example that $V(\mathcal{H})$ is partitioned into triples, and that we allow $\bar{t}(\{x, y\})$ to be large when x, y are in the same triple and take each edge of $\mathcal{H}$ to meet each triple in either 0 or 2 vertices. Then $b(t) = 2$ and it's more or less typical (e.g., take $\mathcal{H}$ regular and uniform) for $\rho(\mathcal{H})$ to be about $(4/3)\rho^*(\mathcal{H})$, reflecting the fact that we have $\rho(\Gamma) = (4/3)\rho^*(\Gamma)$ for the underlying graph Γ of pairs for which $\bar{t}$ is allowed to be large.

But—this was the starting point—the fractional covers (15) arising in connection with $n(r)$ cannot look like this, and in fact $b(t)$ does tend to be large in situations of interest for $n(r)$. (To see what's meant here, assume most pairwise degrees in $\mathcal{H}$ are 1—if they're not then we gain substantially in (16)—and use the fact that t in (15) is given by $t(A) = (n + r - 1)^{-1} \sum_{x \in V} 1_{\{A \ni x\}}$ to show that for X not too large, $t|_X$ is (usually, approximately) in $MP(X)$.)

At any rate, it turns out that "$b(t) \to \infty$" is the correct relaxation of (12), provided we at least insist that α_3 be small:

Theorem 9 ([66]). *Let k be fixed, $\mathcal{H}$ a k-bounded hypergraph, and $t : \mathcal{H} \to \mathbf{R}^+$ a fractional tiling. Then*

$$\rho(\mathcal{H}) \lesssim t(\mathcal{H}) \quad (\alpha_3(t) \to 0, b(t) \to \infty).$$

This implies for example that in Theorem 8 we could replace $\alpha_2(t)$ by

$$\max\{\bar{t}(\{x, y\}) : x, y \in X \text{ or } \ x, y \in Y\}$$

where $X \cup Y$ is a partition of V. That is, allowing $\bar{t}$ to be large on the edges of a bipartite graph doesn't create obstructions to good cover behavior, reflecting the fact that fractional and integer edge cover numbers coincide for bipartite graphs.

(Though this has yet to be checked, Theorem 9 probably implies that Corollary 2 holds even if we relax (14) to the analogous condition on 3-wise intersections.)

If we allow $\alpha_3(t)$ to be large, then the situation changes completely. For instance, the dual, $\mathcal{H}$, of a random cubic graph (as in [8, 12, 13]) is a 3-uniform, 2-regular hypergraph which typically has no short cycles, yet has ρ substantially greater than $|V|/3$. Thus the fractional tiling $t \equiv 1/2$ has large $b(t)$, yet $\rho(\mathcal{H})$ is much larger than $t(\mathcal{H})$.

I think it's fair to say that it's this phenomenon that lies at the heart of the construction of [61]: there the "small sets X with large $d(X)$" mentioned following (16) comprise a hypergraph with properties akin to those described in the preceding paragraph.

Theorem 9 was conjectured in [60]. The proof turned out to be both harder and much more interesting than originally anticipated, involving, centrally, some understanding of the behavior of so-called "normal" distributions on the set of matchings of a graph. This connection is sketched a little in Sect. 7. The questions raised in [66] also led, if somewhat tangentially, to the work on random matchings outlined in Sect. 8.

6. Chromatic and List-Chromatic Indices

It was suggested by Füredi [88] that the hypotheses of Theorem 7 might guarantee the existence not just of one good matching or cover, but of a decomposition of the entire hypergraph into matchings or covers which are good on average. This was proved by Pippenger and Spencer.

Theorem 10 (**[77]**). *Under the hypotheses of Theorem 7,*

$$\chi'(\mathcal{H}) \sim D(\mathcal{H}) \sim \phi(\mathcal{H}).$$

(As noted in [77], the second assertion of Theorem 10 follows from the first; here we restrict our attention to χ'.)

Theorem 10 is based on an elegant variant of Theorem 7, which we state for future reference.

Theorem 11. *For every $\varepsilon > 0$ and k there are $\delta > 0$ and t so that if $\mathcal{H}$ is a k-uniform, D-regular hypergraph on V with*

$$d(x, y) < \delta D \qquad \forall x, y \in V,$$

then there is a probability distribution p on the set $\mathcal{M}$ of matchings of $\mathcal{H}$ satisfying

(a) $\sum\{p(M) : A \in M \in \mathcal{M}\} =_\varepsilon 1/D \qquad \forall A \in \mathcal{H}$,
(b) *For M chosen according to p, and $A \in \mathcal{H}$, the event $\{A \in M\}$ is independent of the events $\{\{B \in M\} : \Delta(A, B) > t\}$.*

In fact what's shown in [77] is that the distribution obtained from the random procedure sketched in Sect. 4 has these properties. (The present ε and δ are not those used earlier). Of course for the expected size, say μ, of a random matching drawn from a distribution satisfying (a), we have $\mu =_\varepsilon n/k$. This (letting $\varepsilon \to 0$) yields Theorem 7, and says that matchings of the desired size are plentiful in some sense.

The proof of Theorem 10 is in the same vein as that of Theorem 7. Rather than cover vertices by edges, we must cover edges by matchings. Again we proceed in stages, choosing at each stage enough random matchings to cover a small but constant fraction of the surviving edges. Here, in contrast to the earlier situation, it is far from obvious what ought to be meant by "random matchings"; but matchings drawn from distributions as in Theorem 11 work very nicely (except that we should relax D-regularity in the theorem to something like "$d(x) =_\delta D \ \forall x \in V$").

The point of (b)—I think this the nicest new idea here—is that it supports use of the Lovász local lemma ([33] or e.g. [4]) to say that our small sets of matchings have positive probability of being well-behaved at *every* vertex. (Here and again in Theorems 12 and 13, application of the local lemma requires much stronger concentration assertions than are given by Chebyshev's inequality. These are achieved via martingales: the so-called "Azuma-Hoeffding" inequality ([7, 54], or e.g. [14, 75]) in the present instance, and extensions thereof for the later results.)

List Colorings

The final (for now; see Conjecture 7) development in the direction we've been discussing was an extension of Theorem 10 to list-colorings which was conjectured in [58] and proved in [62].

Recall that the *list-chromatic index*, $\chi'_l(\mathcal{H})$, of $\mathcal{H}$ is the least t such that if $S(A)$ is a set ("list") of size t for each $A \in \mathcal{H}$, then there exists a proper coloring σ of $\mathcal{H}$ with $\sigma(A) \in S(A)$ for each $A \in \mathcal{H}$. Of course χ'_l is always at least χ', so Theorem 10 is contained in

Theorem 12 (**[62]**). *Let k be fixed and $\mathcal{H}$ a k-bounded hypergraph of maximum degree D satisfying* (11). *Then*

$$\chi'_l(\mathcal{H}) \sim D \qquad (D \to \infty).$$

List-colorings have recently been getting a lot of attention. Here we just want to give enough background on list-chromatic indices of graphs to put the graphic case of Theorem 12 in context. But see [3] for a survey of recent results, and, e.g., [47, 48, 62] for more on what's touched on here. Let us also mention that we are, as usual, in Erdős territory: the study of list colorings was initiated by Vizing in [92] and, independently, Erdős, Rubin and Taylor in [34].

The following central problem, now called the "list-chromatic" or "list coloring" conjecture, seems to have been proposed several times, probably first by Vizing in 1975 (see, e.g., [22, 47, 48] for more on this story).

Conjecture 4. *For every multigraph G, $\chi'_l(G) = \chi'(G)$.*

The case $G = K_{n,n}$ was proposed by J. Dinitz in about 1978 (see [28]) in the context of Latin squares. This version is particularly appealing, and seems to have provided much of the initial stimulus for western interest in such questions.

Conjecture 5 (Dinitz Conjecture). *Suppose that for $1 \leq i, j \leq n$, $S_{i,j}$ is a set of size n. Then there is a partial Latin square $(s_{i,j})_{1\leq i,j\leq n}$ with $s_{i,j} \in S_{i,j}$ for all i, j.*

(A *partial Latin square* of order n is an $n \times n$ array of symbols with the property that no symbol appears more than once in any row or column.)

Let us just quickly mention that there are natural extensions of these problems to vector spaces and matroids. For example, the following was proposed in [62] as a common generalization of the Dinitz Conjecture and a conjecture of G.-C. Rota [55] (see [55, 62] for more in this direction).

Conjecture 6. *Let V be an n-dimensional vector space and suppose that for $1 \leq i, j \leq n$, $S_{i,j}$ is a basis of V. Then there exist $s_{i,j} \in S_{i,j}$ for $1 \leq i, j \leq n$ such that each of the sets $\{s_{i,j} : j = 1, \ldots, n\}$, $\{s_{i,j} : i = 1, \ldots, n\}$ is a basis of V.*

For simple G, Conjecture 4 together with Vizing's Theorem would imply that $D(G) \leq \chi'_l(G) \leq D(G) + 1$, while Theorem 12 says that $D(G)$ is at least the right asymptotic value. This improved several earlier bounds (e.g., [15, 16, 22]), the best of which was

$$\chi'_l(G) < 7D(G)/4 + o(D(G))$$

due to Bollobás and Hind [16].

We haven't had space in this very brief summary to discuss a beautiful algebraic approach to list colorings which was introduced by Alon and Tarsi in [5] and has had several important consequences ([27, 37] or [3]). Recently J. Janssen [56] used this approach to give a simple and elegant proof that $\chi'_l(K_{n,n+1}) = n + 1$ (which in particular says that the Dinitz Conjecture is not off by more than 1), and then Häggkvist and Janssen [48], taking [56] as a starting point, showed, inter alia,

$$\chi'_l(G) < D + O(D^{2/3} \log D) \tag{17}$$

for simple bipartite G of maximum degree D. At this writing they can also show $\chi'_l(K_n) \leq n + 1$, and expect that their method will extend to give (17) for nonbipartite graphs.

Semirandom Again

Theorem 12 was conjectured in [58], where a much more limited extension of Theorem 10 was used to prove Theorem 3. (Derivation of Theorem 3 from Theorem 12 is left as a nice exercise for the reader; or see [60].) While it was nice to see the asymptotic correctness of Conjecture 1, it's my (perhaps not majority) opinion that the most important thing to come out of [58] was the right conjecture along these lines, namely what became Theorem 12.

The special case of Theorem 12 proved in [58] requires a good understanding of [77] but not too much in the way of new ideas. Theorem 12, on the other hand, for some time seemed beyond reach, an opinion influenced in part by the apparent difficulty of even the graphic case, and in part by the fact that the Pippenger-Spencer proof clearly would not extend.

The basic idea of the eventual proof is actually quite natural, though a little strange in that it initially seems doomed to failure. Here's a thumbnail sketch in the "standard" case that all the $S(A)$'s are the same. (The general case is not essentially different, but involves some fiddling to keep the relevant parameters on track.)

We color the hypergraph in stages. At each stage we tentatively assign each as yet uncolored edge A a random color from its current list of legal colors. In some (most) cases, the color tentatively assigned to A will also be assigned to one or more edges meeting A. Such edges A are simply returned to the pool of uncolored edges. The remaining edges (those not involved in such "collisions") are permanently colored with their tentative colors and removed from the hypergraph. We then modify the lists of legal colors (mainly meaning that we delete from $S(A)$ all colors already assigned to edges which meet A) and repeat the process.

Martingale Concentration results together with the Lovász local lemma are used to show that this procedure can be repeated many times, leaving after each stage a hypergraph and modified lists, of legal colors which are reasonably well-behaved. (Finding the correct definition of "well-behaved" is crucial.) Eventually our control here does deteriorate, but by the time this happens the degrees in the remaining hypergraph are small relative to the (minimum) number of colors still admissible at an edge, and the remaining edges can be colored greedily.

The strange feature alluded to above is that the lists of legal colors initially shrink much faster than the degrees. (Roughly, when the degrees have shrunk to βD, with β not too small, the lists will have size about $\beta^k D$ if edges are of size k.) This at first seems unpromising, since we are accustomed to thinking of degree as a trivial lower bound on chromatic index. What saves us here—this is perhaps the central idea of the proof—is that the lists $S^i(A)$ (of legal colors for A at the end of stage i) tend to evolve fairly independently, except where obviously dependent. So for example, for a color γ which through stage i has not been permanently assigned to any edge meeting $A \cap B$ (that is, we condition on this being so), the probability that γ belongs to $S^i(B)$ is not much affected by its membership or nonmembership in $S^i(A)$.

Implementation of this idea is reasonably delicate; still, I think the proof demonstrates considerable flexibility for the "guided-random" approach (beyond what was already apparent from the results discussed above), and expect to see further applications in the near future.

For example, about a year ago, J. H. Kim [69] used a similar method to make significant progress on Vizing's old problem [91] of upper bounds for the chromatic number of a triangle-free graph G of maximum degree D, proving

Theorem 13. *If G is a graph of maximum degree D and girth at least 5, then $\chi_l(G) < (1+o(1))D/\log D$.*

(with the list-chromatic *number* χ_l defined in the obvious way). Here even for large girth the best previous upper bound was about $D/2$ due to Kostochka [71], though it seems reasonable to expect, particularly in view of [1, 2, 86], that the the bound of Theorem 13 remains valid for triangle-free G.

(Recently, R. Häggkvist told me that A. Johansson and S. McGuiness, had just (independently) proved Kim's result—following the method of Theorem 12 as described in [60]—and believed that for girth 4 they could show $\chi(G) = O(D/\log D)$ and $\chi_l(G) = o(D)$.)

A Conjecture

Before closing this section, we mention one more (important) problem which recalls the "fractional vs. integer" theme of Sect. 5.

Conjecture 7. *For fixed k and k-bounded hypergraph $\mathcal{H}$,*

$$\chi'_l(\mathcal{H}) \sim \chi'(\mathcal{H}) \sim \chi'^*(\mathcal{H}).$$

This goes far beyond Theorem 12, giving in effect—LP's being regarded as "tractable" problems—a complete understanding of the asymptotics of chromatic and list-chromatic indices of k-bounded hypergraphs, even if we abandon assumptions such as (11) entirely. Note it contains Theorem 12 via Theorem 11, since the existence of a distribution p on $\mathcal{M} = \mathcal{M}(\mathcal{H})$ satisfying

$$\sum\{p(M) : A \in M \in \mathcal{M}\} \sim 1/D \qquad \forall A \in \mathcal{H}$$

is the same as $\chi'^*(\mathcal{H}) \sim D$.

Even the very special case of Conjecture 7,

$$\text{for multigraphs } G, \chi'(G) \sim \chi'^*, \tag{18}$$

is open, and of considerable interest. (For multigraphs χ'^* is given by Edmonds' Matching Polytope Theorem ([26] or, e.g., [82]). More precise versions of (18) were proposed by Goldberg (see [89]), Andersen [6], Seymour [83] and again Goldberg [45]. The most important results on chromatic indices of multigraphs are those of Shannon [85] and Vizing [90]; see also [35]. For χ'_l of a multigraph G, the current upper bound is $9D(G)/5$, due to Hind [52].)

For one way of attacking (18) see Question 1 below.

7. Normal Distributions

As mentioned earlier, a central role in the proof of Theorem 9 is played by so-called "normal" distributions on the set of matchings of a graph. In this section we say what these are and try to give some idea of what they have to do with Theorem 9. Then in Sect. 8 we describe some recent results for the special case of uniform distribution.

Let $G = (V, E)$ be a graph and $\mathcal{M} = \mathcal{M}(G)$ the set of matchings of G. For $M \in \mathcal{M}$, $v \in V$, we write $v \prec M$ if v is contained in some edge of M.

A *normal distribution* on $\mathcal{M}$ is a probability distribution $p = p_\lambda$ derived from some $\lambda : E \to \mathbf{R}^+$ according to

$$w(M) = \prod_{A \in M} \lambda_A,$$

$$p(M) = w(M) / \sum_{M' \in \mathcal{M}} w(M').$$

For p a probability distribution on $\mathcal{M}$, $M \in \mathcal{M}$ chosen according to p and $F_i \in E$, set $p_{F_1,\ldots,F_t} = Pr(F_1, \ldots, F_t \in M)$. We call the probabilities p_F for $F \in E$ the *marginals* of p.

Let $f : E \to \mathbf{R}^+$. Edmonds' Matching Polytope Theorem says (though not in this language) that there is a probability distribution on $\mathcal{M}$ with marginals f iff

$$\sum_{F \ni v} f(F) \leq 1 \qquad \forall v \in V \tag{19}$$

and

$$\sum \{f(F) : F \subseteq W\} \leq \lfloor |W|/2 \rfloor \qquad \forall W \subseteq V. \tag{20}$$

The *matching polytope*, $MP(G)$, is the set of such f's. For normal distributions the analogous characterization was observed by Rabinovich, Sinclair and Wigderson [78]:

Theorem 14. *There exists a normal distribution with marginals f if and only if the inequalities* (19) *and* (20) *are strict for all $v \in V$ and $W \subseteq V$.*

We say that distribution p on $\mathcal{M}$ has the property $Ed(\delta)$ if its marginal distribution f is in $(1 - \delta)MP(G)$, or equivalently, if the inequalities (19) and (20) hold even when their right hand sides are multiplied by $(1 - \delta)$. A crucial ingredient of the proof of Theorem 9 is a somewhat more technical version of

Lemma 1. *For all δ, $\varepsilon > 0$, and l there exists D such that if p has $Ed(\delta)$ and if $F_1, \ldots, F_l \in E$ are pairwise at distance at least D in G, then*

$$p_{F_1,\ldots,F_l} =_\varepsilon p_{F_1} \cdots p_{F_l}.$$

Thus, requiring that the marginals of p stay well away from the boundary of $MP(G)$ guarantees a fair amount of (approximate) independence among the events $\{F \in M\}$.

Theorem 9 is proved by applying Theorem 8 to a sort of contraction of a randomly generated subhypergraph of the given hypergraph $\mathcal{H}$. To give some idea of the relevance of Lemma 1, we make some simplifying, and slightly vague, assumptions. (The actual proof uses some preliminary reductions to arrive at similar, but somewhat weaker assumptions.)

Suppose Γ is a set of pairs—thought of as a graph—from V such that $\max\{\bar{t}(\{x,y\}) : \{x,y\} \notin \Gamma\}$ is small, and such that each $A \in \mathcal{H}$ is the union of some collection of edges of Γ, called the *parts* of A, which are pairwise far apart in Γ. Under the latter assumption, the restriction $\bar{t}|_\Gamma$ is a fractional tiling; moreover, it's not hard to see that if $b(t)$ is large enough then $f := (1-\vartheta)\bar{t}|_\Gamma \in (1-\delta)MP(\Gamma)$ for appropriate ϑ, δ, both small positive constants.

Write $F \prec A$ if F is a part of A. Let p be the normal distribution with marginals f (as shown in [78], p is unique), and let M be chosen according to p. By the preceding remark, Lemma 1 applies to p. Define a new hypergraph $\mathcal{H}^*$ on vertex set V^*, and $t^* : \mathcal{H}^* \to \mathbf{R}^+$ by

$$V^* = \{F^* : F \in M\},$$

$$\mathcal{H}^* = \{A^* : A \in \mathcal{H},\ \text{all parts of } A \text{ are in } M\}$$

(with $F^* \in A^*$ iff $F \prec A$), and

$$t^*(A^*) = \prod_{F \prec A} f(F)^{-1} t(A) \qquad \forall A^* \in \mathcal{H}^*.$$

We prove Theorem 9 by showing that typically (using "$\approx$" and "$\lesssim$" only qualitatively in what follows)

$$\rho(\mathcal{H}) \lesssim \rho(\mathcal{H}^*) \lesssim t^*(\mathcal{H}^*) \approx t(\mathcal{H}).$$

Note for example that if the various events $\{F \in M\}$ were mutually independent, then we'd have $E[t^*(\mathcal{H}^*)] = t(\mathcal{H})$.

Let us just say a little about the middle inequality, which is the heart of the matter. To prove it, we intend to apply a mild generalization of Theorem 8 to the pair $(\mathcal{H}^*, t^*)$. The main point (of the whole business) is to show that t^* is (usually and in an appropriate sense) an approximate fractional tiling of $\mathcal{H}^*$. This goes roughly as follows.

Suppose we condition on $\{F \in M\}$ (for some $F \in \Gamma$) and consider $\bar{t^*}(F^*) = \sum\{t^*(A^*) : F^* \in A^* \in \mathcal{H}^*\}$. This has (conditional) expectation

$$\sum_{A\succ F} Pr(A^* \in \mathcal{H}^* | F \in M) \prod_{G\prec A} f(G)^{-1} t(A) \approx f(F)^{-1} \sum_{A\succ F} t(A)$$

$$= f(F)^{-1}\bar{t}(F) = (1-\vartheta)^{-1} \approx 1,$$

since according to Lemma 1 and our simplifying assumptions,

$$Pr(A^* \in \mathcal{H}^* | F \in M) = Pr(G \in M \ \forall F \neq G \prec A | F \in M) \approx \prod_{F\neq G\prec A} f(G).$$

Moreover—we leave the reader to ponder this nicest point—Lemma 1 enables us to use Chebyshev's inequality to show that $\bar{t^*}(F^*)$ (again conditioned on $\{F \in M\}$) is usually close to its mean. (There's a slight lie here: we should add an assumption to the effect that for $\{x,y\} \in \Gamma$, $\bar{t}(\{x,y\})$ is not too small.)

Thus, typically, $\bar{t^*}(F^*)$ will be *close* to 1 for *most* $F^* \in V^*$, which is basically what we want.

8. Random Matchings

For the rest of our discussion we consider uniform distribution on $\mathcal{M}(G)$ (so the normal distribution corresponding to $\lambda \equiv 1$). The main results here, Theorems 17 and 18, are surprisingly strong, and show that the distributions in question are in some respects much nicer than seems to have been previously realized.

The material of this section is a little tangential to the topics of Sects. 2–6, but I include it here for two reasons. First, I do think of it as growing out of the earlier work. Second, understanding the extent to which the results described here extend to k-bounded hypergraphs would greatly enhance our understanding of these objects, and would in some cases (see in particular Conjecture 8 and Question 1) have specific consequences for questions considered above.

Let G be a graph and let M be drawn uniformly at random from $\mathcal{M}(G)$. Mainly following [74, p. 341], we set $\xi = \xi(G) = |M|$ and $p_k(G) = Pr(\xi = k)$, and let $\mu = \mu(G)$ and $\sigma = \sigma(G)$ denote respectively the mean and standard deviation of ξ.

In what follows we deal with a sequence $\{G_n\}$ of (*simple*) graphs. We abbreviate $\xi(G_n)$, $\mu(G_n)$ and $\sigma(G_n)$ to ξ_n, μ_n and σ_n, and in addition, set $|V(G_n)| = v_n$, $|E(G_n)| = e_n$, $D(G_n) = D_n$ and $\nu(G_n) = \nu_n$. To avoid trivialities we always assume $v_n \to \infty$ $(n \to \infty)$.

In 1981 C. Godsil [43] gave sufficient conditions for asymptotic normality of the distributions $\{p_k(G_n)\}$:

Theorem 15. *If $D_n^2/e_n \to 0$ then the distribution $\{p_k(G_n)\}_{k\geq 0}$ of matching sizes in G_n is asymptotically normal.*

That is, for each $x \in \mathbf{R}$

$$Pr(\frac{\xi_n - \mu_n}{\sigma_n} < x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \quad (n \to \infty).$$

The same conclusion was obtained by Ruciński [81] under a weaker hypothesis:

Theorem 16. *If $\nu_n/D_n \to \infty$ then $\{p_k(G_n)\}_{k\geq 0}$ is asymptotically normal.*

The main result of [63] is a *necessary and sufficient* condition:

Theorem 17. *The distribution $\{p_k(G_n)\}_{k\geq 0}$ is asymptotically normal if and only if*

$$\nu_n - \mu_n \to \infty \ (n \to \infty). \tag{21}$$

Let $p(G,x)$ denote the probability generating function of the sequence $\{p_k(G)\}$,

$$p(G,x) = \sum_{k=0}^{\nu(G)} p_k(G) x^k.$$

A fundamental fact, proved in [51] (see also [50]) and [72], is

$$\text{for every } G, p(G,x) \text{ has real roots.} \tag{22}$$

The significance of this for our discussion was first noticed by L. Harper [49] in his proof of asymptotic normality of the sequence $\{S(n,k)/B_n\}_{k\geq 1}$ (with $S(n,k)$ the Stirling number of the second kind and B_n the Bell number). Harper's neat observation is that, given (22), a necessary and sufficient condition for asymptotic normality of $\{p_k(G_n)\}_{k\geq 0}$ is that

$$\sigma_n \to \infty. \tag{23}$$

Thus Godsil and Ruciński just need to prove (23) under their respective hypotheses (in both cases the key is the fact, shown in [51], that if $p(G,x_0) = 0$, then $|x_0| \geq 4(D(G) - 1)$), while proving Theorem 17 amounts to bounding $\nu - \mu$ as a function of σ. (The current proof gives $\nu - \mu = O(\sigma^8)$; curiously, one can have $\nu - \mu$ as large as $\Omega(n^6)$, which seems likely to be the truth.)

Having said this, let us stress that, as illustrated by the next result, Theorem 17 is by no means a matter of replacing one unverifiable condition by another. (We write δ_n for the minimum degree of G_n.)

Corollary 3. *Each of the following implies asymptotic normality.*

(a) $\nu_n/D_n \to \infty$
(b) $\delta_n = (1 - o(1))D_n$
(c) $\nu_n > (1 - o(1))v_n/2$

(The reader might try to see why each of (a)–(c) implies (21).)

Thus we recover Theorems 15 and 16, and for example have asymptotic normality for any sequence of regular graphs—note Theorems 15 and 16 do not apply to sequences of regular graphs for which degree grows in proportion to the number of vertices—or graphs with perfect matchings. The latter includes Harper's theorem (again, see [43] or [79, p. 213] for the connection). Slightly unbalanced complete bipartite graphs show that (c) can't be relaxed to "$\nu_n > (1-\varepsilon)v_n/2$" if $\varepsilon > 0$ is fixed.

We don't have space to say much about the proof of Theorem 17. As of now it is not very easy, though there are some special cases—e.g., sequences of regular graphs or sequences as in Theorems 15 and 16—for which the methods of [63] give fairly simple proofs. Though there's no concrete connection between these methods and the guided-random approach of earlier sections, they do have in common a reliance on having a lot of approximate independence in the relevant probability spaces. For the earlier results, this independence is usually derived from something like (11). For Theorem 17, and also for Theorem 9 (see Lemma 1), the required independence derives from the following simple fact.

Let M be a random matching drawn from a normal distribution on a graph G, and for vertex x, set $p(x) = Pr(x \prec M)$. For $x \in V$, $W \subseteq V(G)$, set $\mu(W) = \mu_G(W) = \sum_{w\in W} p(w)$ and $\mu(W|\bar{x}) = \mu_{G-x}(W \setminus \{x\})$. (Thus $\mu(W)$ is the expected number of vertices of W covered by M, while $\mu(W|\bar{x})$ is the same number conditioned on $\{x \not\prec M\}$.)

Lemma 2. *If $x \notin W$, then $\big|\mu(W) - \mu(W|\bar{x})\big| \le p(x)$.*

In particular, conditioning on $\{x \not\prec M\}$ changes $\mu(W)$ by at most 1.

As noted above, Theorem 17 provides the first proof of (23) for sequences of regular graphs. In fact, as shown even more recently in [67], the values of μ and σ^2 for a regular graph are remarkably well determined just by degree and number of vertices:

Theorem 18. *For any d-regular simple graph G,*

(a) $v(G) - 2\mu(G) \sim v(G)/\sqrt{d}$,
(b) $\sigma^2(G) \sim v(G)/(4\sqrt{d})$

(limits taken as $d \to \infty$).

Actually (a) is a consequence of the finer

$$p(\bar{x}) := Pr(x \not\prec M) \sim d^{-1/2} \quad \forall x \in V(G).$$

(It's worth stressing once more the use of the uniformity convention: the rates of convergence in the last three assertions depend on nothing but d.)

The proof of Theorem 18 is again based in part on martingale concentration results, in this case for martingales related to random self-avoiding walks on the graphs in question. (The connection takes a little too long to discuss here, but see [44] for an indication of the relation between matchings and self-avoiding walks.)

Let us close with a conjecture and a question which take us back to some of the topics discussed earlier. (Though we won't pursue the subject here, it would also be of considerable interest to understand to what extent Theorem 17 continues to hold for k-bounded hypergraphs; see [63] for some speculation as to what might be true in this direction.)

Conjecture 8. *For fixed k and simple, k-uniform, D-regular $\mathcal{H}$,*

$$p(\bar{v}) \sim D^{-1/k} \quad \forall v \in V(\mathcal{H}).$$

If we relax "simple" to (11), then the conclusion should be $p(\bar{x}) \to 0$, which, like Theorem 11, would say that the value of ν predicted by Theorem 7 is actually the *average* size of an appropriately defined *random* matching. Given the sophistication of the distribution of Theorem 11 (and the difficulty of Theorem 7 itself), it would be extremely interesting if uniform distribution accomplished the same thing.

Finally, an affirmative answer to the following would imply (18) (in the same way that Theorem 11 implies Theorem 10), and would also be very interesting in its own right.

Question 1. *Is it true that for each $\varepsilon > 0$ there exist t and D_0 such that for every $D \geq D_0$ and D-regular multigraph G with $\chi'^*(G) = D$, there is a probability distribution p on $\mathcal{M} = \mathcal{M}(G)$ satisfying*

(a) $p_A =_\varepsilon 1/D \quad \forall A \in E(G)$, and

(b) For M chosen according to p, and $A \in \mathcal{H}$, the event $\{A \in M\}$ is independent of the events $\{\{B \in M\} : \Delta(A, B) > t\}$?

Added in Proof

Conjecture 7 for bipartite multigraphs, so in particular the Dinitz Conjecture (Conjecture 5), was proved by Fred Galvin around the end of 1993 [F. Galvin, The list chromatic index of a bipartite multigraph, *J. Combinatorial Th. (B)* **63** (1995), 153–158].

Anders Johansson [A. Johansson, An improved upper bound on the choice number for triangle free graphs, manuscript, 1994], again along the lines of the proof of Theorem 12, proved $\chi_l(G) = O(D/\log D)$ for triangle-free G (compare Theorem 13).

Another major application of the semirandom method discussed in Sects. 4 and 6 was given in [J.H. Kim, The Ramsey number $R(3, t)$ has order of magnitude $t^2/\log t$, *Random structures and Algorithms* **7** (1995), 173–207].

Joel Spencer [J . Spencer, Asymptotic packing via a branching process, *Random structures and Algorithms* **7** (1995), 167–172.] showed that the "natural" proof suggested after Theorem 7 does indeed work.

A simpler proof of Theorem 8, based on the [77] proof of Theorem 7, was given in [J. Kahn, A linear programming perspective on the Frankl-Rödl-Pippenger Theorem, *Random Structures and Algorithms* **8** (1996), 149–157].

A proof of (18), based on normal distributions and the approximate independence results of [66], was given in [J. Kahn, Asymptotics of the chromatic index for multigraphs, *J. Combinatorial Th. (B)* **68** (1996), 233–254].

The list-coloring version of (18) (so Conjecture 7 for multigraphs) was proved in [J. Kahn, Asymptotics of the list-chromatic index for multigraphs, *Random Structures & Algorithms* **17** (2000), 117–156].

Several of the above results, together with further applications of the ideas sketched in Sect. 6, are discussed in detail in [M. Molloy and B. Reed, *Graph Colouring and the Probabilistic Method*, Springer, Berlin, 2002.]

References

1. M. Ajtai, J. Komlós and E. Szemerédi, A dense infinite Sidon sequence, *Europ. J. Combinatorics* **2** (1981), 1–11.
2. M. Ajtai, J. Komlós and E. Szemerédi, A note on Ramsey numbers, *J. Combinatorial Th.* (A) **299** (1980), 354–360.
3. N. Alon, Restricted colorings of graphs, *Surveys in Combinatorics, 1993* (Proc. 14th British Combinatorial Conf.), Cambridge Univ. Press, Cambridge, 1993.
4. N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley, New York, 1992.
5. N. Alon and M. Tarsi, Colorings and orientations of graphs, *Combinatorica* **12** (1992), 125–134.
6. L.D. Andersen, On edge-colourings of graphs, *Math. Scand.* **40** (1977), 161–175.
7. K. Azuma, Weighted sums of certain dependent random variables, *Tokuku Math. J.* **19** (1967), 357–367.
8. E. A. Bender and E. R. Canfield, The asymptotic number of labelled graphs with given degree sequences, *J. Combinatorial Th.* (A) **24** (1978), 296–307.
9. C. Berge, *Hypergraphs: Combinatorics of Finite Sets*, North Holland, Amsterdam, 1989.
10. C. Berge, On the chromatic index of a linear hypergraph and the Chvátal conjecture, pp. 40–44 in *Combinatorial Mathematics: Proc. 3rd Int'l. Conf* (New York, 1985), *Ann. N. Y. Acad. Sci.* **555**, New York Acad. Sci., New York, 1989.
11. A. Blokhuis, More on maximal intersecting families of finite sets, *J. Combinatorial Th.* (A) **44** (1987), 299–303.
12. B. Bollobás, A probabilistic proof of an asymptotic formula for the number of regular graphs, *Europ. J. Combinatorics* **1** (1980), 311–316.
13. B. Bollobás, The independence ratio of regular graphs, *Proc. Amer. Math. Soc.* **83** (1981), 433–436.
14. B. Bollobás, Martingales, isoperimetric inequalities and random graphs, in *Combinatories*, A. Hajnal, L. Lovász and V.T. Sós Eds., Colloq. Math. Soc. János Bolyai **52**, North Holland, 1988.
15. B. Bollobás and A. J. Harris, List-colourings of graphs, *Graphs and Combinatorics* **1** (1985), 115–127.
16. B. Bollobás and H. Hind, A new upper bound for the list chromatic number, *Discrete Math.* **74** (1989), 65–75.

17. V. Boltjansky and I. Gohberg, *Results and Problems in Combinatorial Geometry*, Cambridge University Press, Cambridge, 1985.
18. E. Boros, Z. Füredi and J. Kahn, Maximal intersecting families and affine regular polygons, *J. Combinatorial Th.* (A) **52** (1989), 1–9.
19. K. Borsuk, Drei Sätze über die n-dimensionale euklidische Sphäre, *Fundamenta Math.* **20** (1933), 177–190.
20. N. G. de Bruijn and P. Erdős, A combinatorial problem, *Indagationes Math.* **8** (1946), 461–467.
21. W. I. Chang and E. Lawler, Edge coloring of hypergraphs and a conjecture of Erdős, Faber, Lovász, *Combinatorica* **8** (1989), 293–295.
22. A. Chetwynd and R. Häggkvist, A note on list-colorings, *J. Graph Th.* **13** (1989), 87–95.
23. S. Chowla, P. Erdős and E. Straus, On the maximal number of pairwise orthogonal Latin squares of a given order, *Can. J. Math.* **12** (1960), 204–208.
24. H. Croft, K. Falconer and R. Guy, *Unsolved Problems in Geometry*, Springer-Verlag, New York, 1991.
25. S. J. Dow, C. A. Drake, Z. Füredi and J. A. Larson, A lower bound for the cardinality of a maximal family of mutually intersecting sets of equal size, *Congressus Numerantium* **48** (1985), 47–48.
26. J. Edmonds, Maximum matching and a polyhedron with 0, 1-vertices, *J. Res. Nat. Bureau of Standards* (B) **69** (1965), 125–130.
27. M.N. Ellingham and L. Goddyn, List edge colourings of some regular 1-factorable multigraphs, **Combinatorica 16** (1996), 343–352.
28. P. Erdős, Some old and new problems in various branches of combinatorics, *Congressus Numerantium* **23** (1979), 19–37.
29. P. Erdős, On the combinatorial problems which I would most like to see solved, *Combinatorica* **1** (1981), 25–42.
30. P. Erdős, My Scottish book "problems", pp. 35–43 in *The Scottish Book. Mathematics from the Scottish Café* (R.D. Mauldin, ed.), Birkhäuser, 1981.
31. P. Erdős, Some of my old and new combinatorial problems, pp. 35–46 in *Paths, Flows and VLSI-Layout* (B. Korte, L. Lovász, H.-J. Promel and A. Schrijver, eds.), Springer, Berlin, 1990.
32. P. Erdős and H. Hanani, On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10** (1963), 10–13.
33. P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, *Coll. Math. Soc. J. Bolyai* **10** (1974), 609–627.
34. P. Erdős, A. Rubin and H. Taylor, Choosability in graphs, *Congressus Numerantium* **26** (1979), 125–157.
35. S. J. Fiorini and R. J. Wilson, *Edge Colourings of Graphs*, Research Notes in Mathematics **16**, Pitman, London, 1977.
36. R. A. Fisher, An examination of the different possible solutions of a problem in incomplete blocks, *Ann Eugenics* **10** (1940), 52–75.
37. H. Fleischner and M. Stiebitz, A solution to a colouring problem of P. Erdős, *Disc. Math.* **101** (1992), 39–48.
38. P. Frankl and V. Rödl, Near-perfect coverings in graphs and hypergraphs, *Europ. J. Combinatorics* **6** (1985), 317–326.
39. P. Frankl and R.M. Wilson, Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), 357–368.
40. Z. Füredi, On maximal intersecting families of finite sets, *J. Combinatorial Th.* (A) **28** (1980), 282–289.
41. Z. Füredi, The chromatic index of simple hypergraphs, *Graphs and Combinetorics* **2** (1986), 89–92.
42. Z. Füredi, Matchings and covers in hypergraphs, *Graphs and Combinatorics* **4** (1988), 115–206.

43. C. D. Godsil, Matching behavior is asymptotically normal, *Combinatorica* **1** (1981), 369–376.
44. C.D. Godsil, Matchings and walks in graphs, *J. Graph Th.* **5** (1981), 285–297.
45. M. K. Goldberg, Edge-colorings of multigraphs: recoloring technique, *J. Graph Th.* **8** (1984), 123–127.
46. B. Grünbaum, Borsuk's problem and related questions, *Proc. Symp. Pure Math.* **7** (*Convexity*), Amer. Math. Soc., 1963.
47. R. Häggkvist and A. Chetwynd, Some upper bounds on the total and list chromatic numbers of multigraphs, *J. Graph Th.* **16** (1992), 503–516.
48. R. Häggkvist and J. C. M. Janssen, On the list-chromatic index of bipartite graphs, manuscript, 1993.
49. L.H. Harper, Stirling behavior is asymptotically normal, *Ann. Math. Stat.* **38** (1967), 410–414.
50. O. J. Heilmann and E.H. Lieb, Monomers and dimers, *Phys. Rev. Letters* **24** (1970), 1412–1414.
51. O.J. Heilmann and E.H. Lieb, Theory of monomer-dimer systems, *Comm. Math. Physics* **25** (1972), 190–232.
52. H.R. Hind, Restricted edge-colourings, Doctoral thesis, Peterhouse College, Cambridge, 1988.
53. N. Hindman, On a conjecture of Erdős, Faber and Lovász about n-colorings, *Canadian J. Math.* **33** (1981), 563–570.
54. W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Amer. Stat. Assoc.* **27** (1963), 13–30.
55. R. Huang and G.-C. Rota, On the relations of various conjectures on Latin squares and straightening coefficients, manuscript, 1993.
56. J.C.M. Janssen, The Dinitz problem solved for rectangles, *Bull. Amer. Math. Soc.* **29** (1993), 243–249.
57. J. Kahn, On a theorem of Frankl and Rödl, in preparation.
58. J. Kahn, Coloring nearly-disjoint hypergraphs with $n + o(n)$ colors, *J. Combinatorial Th. (A)* **59** (1992), 31–39.
59. J. Kahn, On a problem of Erdős and Lovász: random lines in a projective plane, *Combinatorica* **12** (1992), 417–423.
60. J. Kahn, Recent results on some not-so-recent hypergraph matching and covering problems, pp. 305–353 in *Extremal Problems for Finite Sets, Visegrád, 1991*, Bolyai Soc. Math. Studies **3**, 1994.
61. J. Kahn, On a problem of Erdős and Lovász II: $n(r) = O(r)$, *J. Amer. Math. Soc.* **7** (1994), 125–143.
62. J. Kahn, Asymptotically good list-colorings, *J. Combinatorial Th. (A)*, **73** (1996), 1–59.
63. J. Kahn, A normal law for matchings, *Combinatorica* **20** (2000), 339–391.
64. J. Kahn and G. Kalai, A problem of Füredi and Seymour on covering intersecting families by pairs, *J. Combinatorial Th.* **68** (1994), 317–339.
65. J. Kahn and G. Kalai, A counterexample to Borsuk's Conjecture, *Bull. Amer. Math. Soc.* **29** (1993), 60–62.
66. J. Kahn and M. Kayll, Fractional vs. integral covers in hypergraphs of bounded edge size, *J. Combinatorial Th.* A **78** (1997), 199–235.
67. J. Kahn and J.H. Kim, Random matchings in regular graphs, *Combinatorica* **18** (1998), 201–226.
68. J. Kahn and P. D. Seymour, A fractional version of the Erdős-Faber-Lovász Conjecture, *Combinatorica* **12** (1992), 155–160.
69. J. H. Kim, On Brooks' Theorem for sparse graphs, *Combinatorics, Probability and Computing* **4** (1995), 97–132.
70. J. Komlós, J. Pintz and E. Szemerédi, A lower bound for Heilbronn's problem, *J. London Math. Soc.* **25** (1982), 13–24.

71. A. V. Kostochka, Degree, girth and chromatic number, pp. 679–696 in *Combinatorics, Keszthely 1976*, A. Hajnal and V. T. Sós Eds., Colloq. Math. Soc. János Bolyai **18**, North Holland, 1978.
72. H. Kunz, Location of the zeros of the partition function for some classical lattice systems, *Phys. Lett. (A)* (1970), 311–312.
73. D. G. Larman, in *Convexity and Graph Theory* (Rosenfeld and Zaks, eds.), *Ann. Discrete Math.* **20** (1984), 336.
74. L. Lovász and M.D. Plummer, *Matching Theory*, North Holland, Amsterdam, 1986.
75. C.J.H. McDiarmid, On the method of bounded differences, pp. 148–188 in *Surveys in Combinatorics 1989, Invited Papers at the 12th British Combinatorial Conference*, J. Siemons Ed., Cambridge Univ. Pr., 1989.
76. J.-C. Meyer, pp. 285–286 in *Hypergraph Seminar* (Berge and Ray-Chaudhuri, eds.), Springer, Berlin-Heidelberg-New York, 1974.
77. N. Pippenger and J. Spencer, Asymptotic behavior of the chromatic index for hypergraphs, *J. Combinatorial Th.* (A) **51** (1989), 24–42.
78. Y. Rabinovich, A. Sinclair and A. Wigderson, Quadratic Dynamical Systems, *Proc. 33rd IEEE Symposium on Foundations of Computer Science* (1992), 304–313.
79. J. Riordan, *An Introduction to Combinatorial Analysis*, Wiley, New York, 1958.
80. V. Rödl, On a packing and covering problem, *Europ. J. Combinatorics* **5** (1985), 69–78.
81. A. Ruciński, The behaviour of $\binom{n}{k,,\ldots,,k,n-ik}c^i/i!$ is asymptotically normal, *Discrete Math.* **49** (1984), 287–290.
82. A. Schrijver, *Theory of Linear and Integer Programming*, Wiley, Chichester, 1986.
83. P. D. Seymour, Some unsolved problems on one-factorizations of graphs, in *Graph Theory and Related Topics* (Bondy and Murty, eds.), Academic Press, New York, 1979.
84. P. D. Seymour, Packing nearly-disjoint sets, *Combinatorica* **2** (1982), 91–97.
85. C. E. Shannon, A theorem on coloring the lines of a network, *J. Math. Phys.* **28** (1949), 148–151.
86. J. B. Shearer, A note on the independence number of triangle-free graphs, *Discrete Math.* **46** (1983), 83–87.
87. J. Spencer, Lecture notes, M.I.T., 1987.
88. J. Spencer, personal communication.
89. B. Toft, 75 graph-colouring problems, pp. 9–35 in *Graph Colourings* (R. Nelson and R.J. Wilson, eds.), Wiley, New York, 1990.
90. V. G. Vizing, On an estimate of the chromatic class of a p-graph (in Russian), *Diskret. Analiz* **3** (1964), 25–30.
91. V. G. Vizing, Some unsolved problems in graph theory (Russian), *Uspehi Mat. Nauk.* **23** (1968), 117–134. English translation in *Russian Math. Surveys* **23** (1968), 125–141.
92. V. G. Vizing, Coloring the vertices of a graph in prescribed colors (in Russian), *Diskret. Analiz No 29 Metody Diskret. Anal. v Teorii Kodov i Shem* (1976), 3–10, 101 (MR58 #16371).
93. R. M. Wilson, An existence theory for pairwise balanced designs I–III, *J. Combinatorial Th.* (A) **13** (1972), 220–273, **18** (1975), 71–79.

The Origins of the Theory of Random Graphs

Michał Karoński and Andrzej Ruciński

M. Karoński (✉)
Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland
Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA
e-mail: karonski@amu.edu.pl

A. Ruciński
Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA
Department of Discrete Mathematics, Adam Mickiewicz University,
61-614, Poznan, Poland
e-mail: rucinski@amu.edu.pl

1. Introduction

The origins of the theory of random graphs are easy to pin down. Undoubtedly one should look at a sequence of eight papers co-authored by two great mathematicians: Paul Erdős and Alfred Rényi, published between 1959 and 1968:

[ER59] *On random graphs I*, Publ. Math. Debrecen **6** (1959), 290–297.
[ER60] *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. **5** (1960), 17–61.
[ER61a] *On the evolution of random graphs*, Bull. Inst. Internat. Statist. **38**, 343–347.
[ER61b] *On the strength of connectedness of a random graph*, Acta Math. Acad. Sci. Hungar. **12** (1961), 261–267.
[ER63] *Asymmetric graphs*, Acta Math. Acad. Sci. Hung. **14**, 295–315.
[ER64] *On random matrices*, Publ. Math. Inst. Hung. Acad. Sci. **8** (1964), 455–461.
[ER66] *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hung. **17** (1986), 359–368.
[ER68] *On random matrices II*, Studia Sci. Math. Hung. **3** (1968), 459–464.

Our main goal is to summarize the results, ideas and open problems contained in those contributions and to show how they influenced future research in random graphs.

For us it was a great adventure to return to the roots of the theory of random graphs, and to find out again and again, how far-reaching the impact of Erdős and Rényi's work on the field is. The reader will find in our paper many quotations from their original papers (*always in italics*). We use this

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_23,

convention to let them speak directly and to preserve their special insightful style and way of thinking and stating the problems. Starting from there we lead the reader through the literature, including the most current one, trying to show how the ideas of Erdős and Rényi developed, how much time, skills and effort to solve some of their most challenging open problems was needed. Finally, to add some "salt and pepper" to our presentation, full of admiration and respect, we point out to a few false statements and oversimplifications of proofs, which have been found in their monumental legacy by the next generations of random graph theorists.

2. The First Question: Connectivity

Although the notion of a random graph appeared in connection to the probabilistic method already in the Erdős paper [25] (see J. Spencer's article in this volume), it was forgotten for a decade until Paul Erdős and Alfred Rényi published a series of papers entirely devoted to properties of random graphs. The model of a random graph they exclusively investigated was the uniform one. Here is how they defined it: "*Let $E_{n,N}$ denote the set of all graphs having n given labeled vertices and N edges. A random graph $\Gamma_{n,N}$ can be defined as an element of $E_{n,N}$ chosen at random, so that each of the elements of $E_{n,N}$ have the same probability to be chosen, namely $1/\binom{\binom{n}{2}}{N}$.*" (In this paper we adopt the original notation $\Gamma_{n,N}$.)

They were aware of existing results about other models of random graphs. In particular, they acknowledge in a footnote to [ER61a] that E. N. Gilbert [36] studied the connectedness of what we call today the binomial model, where "*We may decide with respect to each of the $\binom{n}{2}$ edges, whether they should form part of the random graph considered or not, the probability of including a given edge being $p = N/\binom{n}{2}$ for each edge and the decisions concerning different edges being independent.*" (In this paper we shall denote this model by $\Gamma_{n,p}$.) In [ER61a] they mention that the investigations of the binomial model can be reduced, due to a conditional argument they attribute to Hajek, to that of $\Gamma_{n,N}$. However, they did not formulate any equivalence theorem (these appeared much later in [14] and [59]) and occasionally stated the binomial counterparts of their theorems without proofs or repeated their proofs step by step.

Apparently they were not aware of the result of Gilbert and of the binomial model at all when they wrote their first paper on random graphs,"On random graphs I". The question addressed there was that of connectedness of a random graph. In fact, according to a remark in [ER59], this problem was tried and partially solved already in 1939, when P. Erdős and H. Whitney, in an unpublished work: "*proved that if $N > \left(\frac{1}{2}+\varepsilon\right) n\log n$ where $\varepsilon > 0$ then the probability of $\Gamma_{n,N}$ being connected tends to 1 if $n \to \infty$, but if $N < \left(\frac{1}{2}-\varepsilon\right) n\log n$ with $\varepsilon > 0$ then the probability of $\Gamma_{n,N}$ being connected, tends to 0 if $n \to \infty$.*"

In the first "official" paper on random graphs, Erdős and Rényi refined the above result as their (partial) answer to questions 1–3 from the following list of problems they posed.

1. *What is the probability of* $\Gamma_{n,N}$ *being completely connected?*
2. *What is the probability that the greatest connected component (subgraph) of* $\Gamma_{n,N}$ *should have effectively* $n-k$ *points?* $(k=0,1,\ldots)$
3. *What is the probability that* $\Gamma_{n,N}$ *should consist of exactly* $k+1$ *connected components?* $(k=0,1,\ldots)$
4. *If the edges of a graph with n vertices are chosen successively so that after each step every edge which has not yet been chosen has the same probability to be chosen as the next, and if we continue this process until the graph becomes completely connected, what is the probability that the number of necessary steps* ν *will be equal to a given number* l*?*

Note that in problem 4 Erdős and Rényi describe a genuine random graph process, whose advanced analysis could be carried over only two decades later.

Before turning to the proofs, they recall a recursive formula and a generating function for the number $C(n,N)$ of connected graphs on n labeled vertices and with N edges, due to Riddell and Uhlenbeck, and also Gilbert. But immediately they comment that neither of them "*. . . helps much to deduce the asymptotic properties of* $C(n,N)$*. In the present paper we follow a more direct approach.*"

We now present the first result on random graphs and its proof in a slightly modified form. The idea of the proof, however, remains unchanged. In the 1959 paper only the middle part of the theorem below was stated explicitly. The other two follow by letting $c=c_n$ tend to $+\infty$ or $-\infty$, respectively.

Theorem 2.1 ([**ER59**]).

$$P(\Gamma_{n,N} \textit{ is connected }) \to \begin{cases} 0 & \textit{if } \frac{N}{n}-\frac{1}{2}\log n \to -\infty \\ e^{-e^{-2c}} & \textit{if } \frac{N}{n}-\frac{1}{2}\log n \to c \\ 1 & \textit{if } \frac{N}{n}-\frac{1}{2}\log n \to \infty. \end{cases}$$

Proof. For convenience we switch to the binomial model, shortening the original argument a lot, and, at the same time, avoiding a harmless error in the proof of "*the rather surprising Lemma*" of [ER59], pointed out by Godehardt and Steinbach [37].

To make this argument formal, assume that $2np-\log n-\log\log n\to\infty$ but $np=O(\log n)$. Thus, *almost surely* (i.e., with probability tending to 1 as $n\to\infty$), there are no isolated edges in $\Gamma_{n,p}$. What remains to be shown is that there are no components of size $3\le k\le\frac{n}{2}$ either. To this end consider the random variable X counting such components. Then, bounding the probability that a given set of k vertices spans a connected subgraph by $k^{k-2}p^{k-1}$, and using the inequality $np > \frac{1}{2}\log n$, we obtain

$$\begin{aligned} Exp(X) &\le \sum_{k=3}^{n/2} \binom{n}{k} k^{k-2}p^{k-1}(1-p)^{k(n-k)} < \sum_k \left(\frac{en}{k}\right)^k k^{k-2}p^{k-1}e^{-(n-k)pk} \\ &\le \frac{1}{p}\sum_{k=3}^{\sqrt{n}} \frac{1}{k^2}\left(\frac{enp}{e^{(n-\sqrt{n})p}}\right)^k + \frac{1}{p}\sum_{k\ge\sqrt{n}}^{n} \frac{1}{n}\left(\frac{enp}{e^{np/2}}\right)^k \\ &= O\left(\frac{n}{\log n}\frac{\log^3 n}{n^{3/2}}\right) + \frac{1}{\log n}\left(\frac{e\log n}{2n^{1/4}}\right)^{\sqrt{n}} = o(1). \end{aligned}$$

Hence, almost surely there are no components outside the largest one other than isolated vertices (Erdős and Rényi say that such a graph is of type A) and the threshold for connectedness coincides with that for disappearance of isolated vertices, i.e., for $2np - \log n - \log\log n \to \infty$

$$P(\Gamma_{n,p} \text{ is connected }) = P(\delta(\Gamma_{n,p}) > 0) + o(1).$$

Erdős and Rényi found the limiting value of $P(\delta(\Gamma_{n,p}) > 0)$ by inclusion-exclusion. Nowadays a standard approach is by the method of moments which serves to show that the number of isolates is asymptotically Poisson. They used that method in the 1960 paper in a more general setting where components isomorphic to a given graph G were considered. We shall return to this later.

Answering question 4, they gave a somewhat oversimplified proof of the fact that

$$\lim_{n\to\infty} P\left(\frac{\nu - \frac{1}{2}n\log n}{n} < x\right) = e^{-e^{-2x}} . \qquad \Box$$

Erdős and Rényi conclude the 1959 paper as follows. "*The following more general question can be asked: Consider the random graph $\Gamma_{n,N(n)}$ with n possible vertices and $N(n)$ edges. What is the distribution of the number of vertices of the greatest connected component of $\Gamma_{n,N(n)}$ and the distribution of the number of its components? What is the typical structure of $\Gamma_{n,N(n)}$ (in the sense in which, according to our Lemma, the typical structure of $\Gamma_{n,N(n)}$ is that it belongs to type A)? We have solved these problems in the present paper only in the case $N(n) = \frac{1}{2}n\log n + cn$. We shall return to the general case in another paper [8].*" ([8] = [ER60] on our reference list.)

As far as connectedness is concerned, in the 1961 paper Erdős and Rényi go on and find the threshold for r-connectivity of $\Gamma_{n,p}$ for every natural r. "*If G is an arbitrary non-complete graph, let $c_p(G)$ denote the least number k such that by deleting k appropriately chosen vertices from G (...) the resulting graph is not connected. (...) Let $c_e(G)$ denote the* *least number l such that by deleting l appropriately chosen edges from G the resulting graph is not connected.*" A graph is *r-connected* if no removal of r or less vertices can disconnect it. When the random graph becomes almost surely r-connected? Theorem 2.1 revealed an interesting feature of random graphs. Namely, quite often trivial necessary conditions become asymptotically sufficient in the sense that for a typical, large graph their fulfillment guaranties that the property in question holds. Due to Theorem 2.1 this is the case of connectedness versus the nonexistence of isolated vertices. For r-connectedness such natural necessary condition is that the minimum degree (denoted in [ER61b] by $c(G)$) must be at least r. Otherwise removing the vertices adjacent to a vertex of minimum degree would disconnect the graph. Erdős and Rényi showed in 1961 that in the range $\frac{1}{2}n \log n \leq N \leq n \log n$ this is the only way one can disconnect the random graph $\Gamma_{n,N}$ by removing the smallest possible number of vertices. A *minimal cutset* is a set of vertices whose removal makes the graph disconnected but no proper subset of that set has this property. For $2 \leq k \leq \frac{n-1}{2}$ let $\mathcal{A}_k$ be the event that there is in $\Gamma_{n,N}$ a minimal cutset of size s, $1 \leq s \leq r-1$, which leaves the second largest component of size k. Arguing similarly as in the proof of Theorem 2.1, they proved that $P(\bigcup_{k\geq 2} \mathcal{A}_k) = o(1)$, meaning that, almost surely, if $\Gamma_{n,N}$ is not r-connected then the only reason for that is the presence of vertices of degree less than r. The method of moments (again, in the inclusion-exclusion cover-up) gives that, for $N(n) = \frac{1}{2}n \log n + \frac{r}{2} n \log \log n + an + o(n)$, their number is asymptotically Poisson. We thus arrived at the main result of the 1961 paper.

Theorem 2.2 (**[ER61a]**). *If we have*

$$N(n) = \frac{1}{2}n \log n + \frac{r}{2} n \log \log n + an + o(n)$$

where a is a real constant and r a non-negative integer, then

$$\lim_{n\to\infty} P(c_p(\Gamma_{n,N(n)}) = r) = 1 - \exp\left(-\frac{e^{-2a}}{r!}\right), \tag{3}$$

further

$$\lim_{n\to\infty} P(c_e(\Gamma_{n,N(n)}) = r) = 1 - \exp\left(-\frac{e^{-2a}}{r!}\right) \tag{4}$$

and

$$\lim_{n\to\infty} P(c(\Gamma_{n,N(n)}) = r) = 1 - \exp\left(-\frac{e^{-2a}}{r!}\right). \tag{5}$$

In a proceeding remark they promise: "*The statement (5) of Theorem 2.2 gives information about the minimal valency of points of $\Gamma_{n,N}$. In a forth-*

coming note we shall deal with the same question for larger ranges of N (when $c(\Gamma_{n,N})$ tends to infinity with n), further with the related question about maximal valency of points of $\Gamma_{n,N}$.” This promise was never fulfilled. The only trace of their interest in the vertex degrees of a random graph can be found in the description of the last phase of the evolution of $\Gamma_{n,N}$ in [ER61a]: “*Phase 5. consists of the range $N(n) \sim (n \log n)w(n)$ where $w(n) \to \infty$. In this range the whole graph is not only almost surely connected, but the orders of points are almost surely asymptotically equal. Thus the graph becomes in this phase ‘asymptotically regular’.*” The proof of that statement can be found in the last section of [ER60]. A very careful analysis of vertex degrees in a random graph is due to Bollobás [10, 11] and can be found also in his book [14].

3. Subgraphs: The Beginning of a Theory

After having written their paper on connectivity of a random graph Erdős and Rényi decide to write a long paper addressing several properties of random graphs. That seminal paper was preceded by an extended abstract [ER61a], where they outlined the main goals of the theory to be born. “*Our main goal is to show (...) that the evolution of a random graph shows very clear-cut features. The theorems we have proved belong to two classes. The theorems of the first class deal with the appearance of certain subgraphs (e.g., tress, cycles of a given order etc.) or components, or other local structural properties, and show that for many types of local structural properties A a definite ‘threshold’ $A(n)$ can be given, so that if $\frac{N(n)}{A(n)} \to 0$ for $n \to \infty$ then the probability that the random graph $\Gamma_{n,N(n)}$ has the structural property A tends to 0 for $n \to \infty$, while for $\frac{N(n)}{A(n)} \to \infty$ for $n \to \infty$ the probability that the random graph $\Gamma_{n,N(n)}$ has the structural property A tends to 1 for $n \to \infty$. (...) The theorems of the second class are of similar type, only the properties A considered are not of a local character, but global properties of the graph $\Gamma_{n,N(n)}$ (e.g., connectivity, total number of components, etc.).*” The existence of a threshold in all cases they considered was a rather surprising fact for Erdős and Rényi. Only three decades later it was proved by Bollobás and Thomason [19] that, as a consequence of the Kruskal-Katona inequality, every monotone property (family) of random subsets of a set has a threshold in the above sense.

In the same abstract they comment that their proofs are “*... completely elementary, and are based on the asymptotic evaluation of combinatorial formulae and on some well-known general methods of probability theory ....*”

The first theorem of the major paper [ER60] established the threshold for the existence of a subgraph of a given type for a broad class of subgraphs. “*If a graph has n vertices and N edges, we call the number $\frac{2N}{n}$ the ‘degree’ of the graph (as a matter of fact $\frac{2N}{n}$ is the average degree of the vertices of G.) If a graph G has the property that G has no subgraph having a larger degree than G itself, we call G a* balanced *graph.*”

Theorem 3.1 ([**ER60**]). *Let $k \geq 2$ and l $(k-1 \leq l \leq \binom{k}{2})$ be positive integers. Let $\mathcal{B}_{k,l}$ denote an arbitrary not empty class of connected balanced graphs consisting of k points and l edges. The threshold function for the property that the random graph considered should contain at least one subgraph isomorphic with some element of $\mathcal{B}_{k,l}$ is $n^{2-\frac{k}{l}}$.*

Among special cases they mention trees, connected unicyclic graphs, cycles, complete graphs and complete bipartite graphs all of which are balanced. Over 20 years later, Bollobás [9] generalized this theorem to arbitrary (not only balanced) graphs. He, however, used a rather complicated method. In 1985, to a great surprise to all involved, Ruciński and Vince [73] found out that the original proof of Erdős and Rényi which was based on the second moment method can be easily adapted to cover all graphs as well. We now state that result in the binomial model.

Theorem 3.2 ([**9**]). *For an arbitrary graph G with at least one edge,*

$$\lim_{n\to\infty} P(G \subset \Gamma_{n,p}) = \begin{cases} 0 & \textit{if } p = o(n^{-1/m_G}) \\ 1 & \textit{if } n^{-1/m_G} = o(p), \end{cases}$$

where $m_G = \max_{H\subseteq G} d_H$ and $d_G = \frac{|E(G)|}{|V(G)|}$.

A crucial role in the Ruciński-Vince proof of Theorem 3.2 is played by the quantity $\Phi_G = \min_{H\subseteq G} Exp(X_H)$. In fact, the inequalities

$$1 - \Phi_G \leq P(G \not\subset \Gamma_{n,p}) \leq c_1/\Phi_G$$

obtained in that proof have been strengthened to exponential bounds

$$e^{-c_2\Phi_G} \leq P(G \not\subset \Gamma_{n,p}) \leq e^{-c_3\Phi_G},$$

where the L-H-S follows by the FKG inequality and the R-H-S is a special case of a recent inequality from [42].

As far as the asymptotic distributions of subgraph counts are concerned, Erdős and Rényi treated in [ER60] only trees and cycles. For trees of order k they established a limiting Poisson distribution on the threshold $N \sim cn^{\frac{k-2}{k-1}}$. They observed that the same result holds for isolated trees, since in this range almost surely all k-vertex trees are isolated (i.e., are components of the random graph). They also found another Poisson threshold for isolated trees at $N = \frac{1}{2k} n \log n + \frac{k-1}{2k} n \log\log n + cn + o(n)$, beyond which isolated trees die out (swallowed by the giant component on its way to absorb all the vertices of the random graph). They also established an asymptotic normality of the number of isolated trees of order k (after suitable standardization) in the whole range of N between the two thresholds. As observed by A. Barbour in [5], the proof given by Erdős and Rényi was not correct and in the range $N \sim cn$, $c \neq 1/2$, the standardization was not right. However, using another method Barbour showed that indeed the asymptotic normality holds in the entire range in question. For cycles and isolated cycles they established a Poisson distribution (different in each case) at $N \sim cn$ and observed that contrary to isolated trees, "*… the probability that $\Gamma_{n,N}$ contains an isolated cycle of order k never approaches* 1." A similar result was proved for connected unicyclic graphs. All these results were obtained by the method of moments based on a fact from probability theory that for all distributions which are uniquely determined by their moments (Poisson and normal are such) the convergence of all moments of a sequence of random variables to the moments of that distribution implies convergence in distribution [8, Theorem 30.2]. Erdős and Rényi prove this fact as a lemma just for the Poisson distribution, although they use it also for the normal distribution. At the end of the paper, in a remark added in proof, they acknowledge that N. V. Smirnov proved this lemma already in 1939.

They conclude their investigations of local properties of random graphs with the comment: "*Similar results can be proved for other types of subgraphs, e.g., complete subgraphs of a given order. As however these results and their proofs have the same pattern as those given above we do not dwell on the subject any longer and pass to investigate global properties of the random graph $\Gamma_{n,N}$.*" In 1979, K. Schürger, a former Ph.D. student of Erdős, proved similar results for complete subgraphs [74] and a few years later Karoński [47] extended them to so called k-trees, a common generalization of trees and complete graphs. All these particular cases led to a general result for all strictly balanced graphs. A graph is *strictly balanced* if every proper subgraph has its degree strictly smaller than the graph itself. Let us denote $d_G = \frac{|E(G)|}{|V(G)|}$ and recall that X_G is the number of copies of G in a random graph $\Gamma_{n,p}$. The following result was proved independently in [9] and [48].

Theorem 3.3 ([**9**, **48**]). *If G is a strictly balanced graph and $np^{d_G} \to c > 0$ then X_G converges to the Poisson distribution with expectation $\frac{c^v}{aut(G)}$.*

If a graph G is balanced but not strictly balanced then the limiting distribution of X_G on the threshold, i.e. when $p = \Theta(n^{-1/d_G})$, becomes quite involved. Although, in principle, as shown by Bollobás and Wierman [20], it can be computed, there is no nice closed formula. For example, when G is a disjoint union of 2 triangles then the limit distribution is that of the random variable $\binom{Y}{2}$, where Y is Poisson. When G is the triangle with a pendant edge, the limit is $\sum_{i=1}^{Z} Y_i$, where all random variables involved are independent and Poisson. When G is the triangle with two pendant edges hanging at the same vertex then X_G converges to the distribution of $\sum_{i=1}^{Z} \binom{Y_i}{2}$, where again all random variables are independent Poisson. One more example: if G is the triangle with a path of length 2 hanging at one of it vertices, then the limit distribution is that of $\sum_{i=1}^{\sum_{j=1}^{U} W_j} Y_i$, where all random variables are independent Poisson. We can only hope that so far the reader is convinced that a pattern does indeed exist.

If G is nonbalanced, then the expectation of X_G tends to infinity and one has to normalize. It turns out that there is a nonrandom sequence $a_n(G) \to \infty$ such that the asymptotic distribution of $\frac{X_G}{a_n(G)}$ coincides with that of X_H, where H is the largest subgraph of G for which $d_H = m_G$. Clearly, H is balanced and we are back to the balanced case. The sequence $a_n(G)$ is equal to the expected number of extensions of a given copy of H to a copy of G in the random graph $\Gamma_{n,p}$. For details see [71, page 292].

Beyond the threshold, i.e., when $np^{m_G} \to \infty$, X_G converges after standardization to the standard normal distribution as long as $n^2(1-p) \to \infty$. (For bigger p X_G is either Poisson or degenerate, according to the formula $X_G \sim \binom{n}{v}\frac{v!}{aut(G)} - c_n(G)Z$, where Z is the binomial random variable counting edges in the complement of $\Gamma_{n,p}$ and $c_n(G)$ is the number of copies of G in K_n containing a fixed edge. For details see [70].) This result was supplemented by the rate of convergence in [7]. It was shown there that the total variation distance between standardized X_G and the standard normal distribution can be bounded by $O(\frac{1}{\sqrt{\Phi_G}})$ as long as $p \not\to 1$ and by $O(\frac{1}{n\sqrt{1-p}})$ otherwise. Recall that $\Phi_G \to \infty$ if and only if $np^{m_G} \to \infty$.

A variant of the small subgraph problem is one when we only count induced subgraphs of $\Gamma_{n,p}$ which are isomorphic to G (*induced copies*). Let Y_G count them. Then, denoting $v = |V(G)|$ and $l = |E(G)|$, $Exp(Y_G) = Exp(X_G)(1-p)^{\binom{v}{2}-l}$, and as long as $p \to 0$ there is no substantial difference in the limiting distribution of X_G and Y_G. For p constant, however, interesting things may happen. First of all, in contrast to X_G, the variance of Y_G may drop below the order of n^{2v-2}. It does so when $Exp(I|J_{12}) = Exp(I)$, i.e., when $p = l/\binom{v}{2}$, where I is the indicator of the event that there is an induced copy of G in $\Gamma_{n,p}$ on the vertex set $\{1, \ldots, v\}$ and J_{ij} is the indicator that the edge ij is present in $\Gamma_{n,p}$. But if $Var(Y_G) = \Theta(n^{2v-3})$ then still Y_G is asymptotically normal, and only when the variance drops further down to the order of n^{2v-4} the distribution of standardized Y_G becomes nonnormal (the convolution of normal and χ^2 distributions). It is a purely combinatorial question when $Var(Y_G) = \Theta(n^{2v-4})$. For the higher terms to cancel out one needs that $Exp(I|J_{12}, J_{13}, J_{23}) = Exp(I)$, or, equivalently, that in addition to $p = l/\binom{v}{2}$, the proportion $t_3 : t_2 : t_1 : t_0 = p^3 : 3p^2q : 3pq^2 : q^3$ is satisfied, where t_i is the number of induced subgraphs of G isomorphic to the graph with 3 vertices and i edges. For $p = \frac{1}{2}$, an example of a graph satisfying these requirements is the wheel on 8 vertices, i.e. the graph obtained from the 7-cycle by joining a new vertex to every vertex of the cycle. For some time it was an open question if such abnormal cases take place for every rational p. A positive answer to that puzzle is due to combined efforts of Janson, Kratochvíl, Kärrman and Spencer [41, 45, 49].

The random variables X_G and Y_G are examples of sums of random variables with only few dependent summands. In particular, the summands forming Y_G are dependent only if the sets corresponding to the indices intersect (on at least 2 vertices, in fact). The reason is that the property of the vertex set we are after depends only on the presence and absence of the edges within the set. The situation changes when we move to the properties depending also on the pairs with one endpoint in the set. Then all summands are mutually dependent, but most just weakly. We have already encountered such a case when studying the number of components of $\Gamma_{n,p}$ which are isomorphic to a given graph G. Clearly this property requires that there is no edge with one endpoint in the set of vertices of a copy of G. Another example of such "semi-induced" property is the notion of *a maximal clique.* This is a complete subgraph not contained in any bigger complete subgraph of a graph. For a vertex set to span a maximal clique one needs that no other vertex is adjacent to all the vertices of the set. In [6] the limiting distribution of the number of maximal k-cliques was investigated. It was proved that for $k \geq 2$ there are two Poisson thresholds for the existence of maximal k-cliques and the phase of asymptotic normality between them. Finally, there are characteristics which lead to sums of random variables indexed by vertex sets, which each depend on the presence or absence of all the edges in $\Gamma_{n,p}$. An example of this is the number of copies of G disjoint from all other copies of G in $\Gamma_{n,p}$. Here even the expectation is difficult to obtain, and the limiting normal distribution is still beyond ones reach.

4. Phase Transition

Sections 4–9 of [ER60] are devoted to global properties of random graphs. The proofs follow the same pattern. First, the expectation of the quantity in question is asymptotically evaluated. Then, using Markov's and Chebyshev's inequality (the first and the second moment method, resp.) the asymptotics of the quantities themselves are derived. As a summary of these results we quote here how Erdős and Rényi characterize the process of the evolution of a random graph in the paper presented to the International Statistical Institute meeting in Tokyo in 1961 [ER61a]:

"If n is fixed large positive integer and n is increasing from 1 to $\binom{n}{2}$, the evolution of $\Gamma_{n,N}$ passes through five clearly distinguishable phases. These phases correspond to ranges of growth of the number N of edges, these ranges being defined in terms of the number n of vertices.

- **Phase 1** *corresponds to the range $N(n) = o(n)$. For this phase it is characteristic that $\Gamma_{n,N(n)}$ consists almost surely (i.e. with probability tending to 1 as $n \to +\infty$) exclusively of components which are trees. (...)*
- **Phase 2** *corresponds to the range $N(n) \sim cn$ with $0 < c < 1/2$. (...) In this range almost surely all components of $\Gamma_{n,N(n)}$ are either trees or components consisting of an equal number of edges and vertices, i.e. components containing exactly one cycle. (...) In this phase though not all, but still almost all (i.e. $n - o(n)$) vertices belong to components which*

are trees. The mean number of components is $n - N(n) + O(1)$*, i.e. in this range by adding a new edge the number of components decreases by* 1*, except for the finite number of steps.*

- **Phase 3** *corresponds to the range* $N(n) \sim cn$ *with* $c \geq 1/2$*. When* $N(n)$ *passes the threshold* $n/2$*, the structure of* $\Gamma_{n,N(n)}$ *changes abruptly. As a matter of fact this sudden change of the structure of* $\Gamma_{n,N(n)}$ *is the most surprising fact discovered by the investigation of the evolution of random graphs. While for* $N(n) \sim cn$ *with* $c < 1/2$ *the greatest component of* $\Gamma_{n,N(n)}$ *is a tree and has (with probability tending to* 1 *as* $n \to +\infty$*) approximately* $\frac{1}{\alpha}\left(\log n - \frac{5}{2}\log\log n\right)$ *vertices, where* $\alpha = 2c - \log 2c$*, for* $N(n) \sim n/2$ *the greatest component has (with probability tending to* 1 *as* $n \to +\infty$*) approximately* $n^{2/3}$ *vertices and has rather complex structure. Moreover for* $N(n) \sim cn$ *with* $c > 1/2$ *the greatest component of* $\Gamma_{n,N(n)}$ *has (with probability tending to* 1 *as* $n \to +\infty$*) approximately* $G(c)n$ *vertices, where*

$$G(c) = 1 - \frac{1}{2c}\sum_{k=1}^{+\infty} \frac{k^{k-1}}{k!}\left(2ce^{-2c}\right)^k$$

  (clearly $G(1/2) = 0$ *and* $\lim_{c\to+\infty} G(c) = 1$*).*

  Except this "giant" component, the other components are all relatively small, most of them being trees, the total number of vertices belonging to components, which are trees being almost surely $n(1 - G(c)) + o(n)$ *for* $c \geq 1/2$*. (...)*

  The evolution of $\Gamma_{n,N(n)}$ *in Phase 3. may be characterized by that the small components (most of which are trees) melt, each after another, into the giant component, the smaller components having the larger chance of "survival"; the survival time of a tree of order* k *which is present in* $\Gamma_{n,N(n)}$ *with* $N(n) \sim cn$*,* $c > 1/2$ *is approximately exponentially distributed with mean value* $n/2k$*.*
- **Phase 4** *corresponds to the range* $N(n) \sim cn\log n$ *with* $c \leq 1/2$*. In this phase the graph almost surely becomes connected. (...)*
- **Phase 5** *consists of range* $N(n) \sim (n\log n)\omega(n)$ *where* $\omega(n) \to +\infty$*. In this range the whole graph is not only almost surely connected, but the orders of all points are almost surely asymptotically equal. Thus the graph becomes in this phase "asymptotically regular". "*

Erdős and Rényi in their fundamental paper [ER60] gave a fairly complete "big picture" of the evolution of a random graphs. However many fascinating questions were left unanswered. For example, how did the giant component grow so rapidly, what is the nature of the "double jump" of its size: from $O(\log n)$ when $c < 1/2$ to $\Theta(n^{2/3})$ when $c = 1/2$ and finally being of the order of n when $c > 1/2$?

Often we say that a random graph goes through the phase transition at $c = 1/2$ due to an obvious resemblance of this period of its evolution to the physical phenomena of changing the state, for example, from liquid to solid. Here a random graph changes abruptly its state from a loose collection of small components being trees and unicyclic to solid single giant component dominating its structure.

The critical moment of the phase transition was unresolved until the milestone paper of Béla Bollobás [13] who revealed the mechanism of the formation of the giant component. He also focused the attention, for the first time, on the nature of the phase transition phenomena, investigating this critical moment of the evolution and looking at the beginning of so called *supercritical phase*. He asked what is the typical structure of a random graph $\Gamma_{n,N}$ when $N(n) = \frac{1}{2}n + s$, where $s = o(n)$. In particular he proved that the largest component is almost surely unique once $s \geq 2(\log n)^{1/2}n^{2/3}$ and its size $L_1(\Gamma_{n,N})$ is approximately $4s$ while the size of the second largest component $L_2(\Gamma_{n,N})$ is much smaller.

Bollobás gave a good lead to what we might consider as the proper magnification if we want to get undistorted picture of the phase transition while looking at the neighborhood of the "critical point" $n/2$. Due to later results of Łuczak [58], combined with those of Kolchin [51], we know that the correct parametrization is

$$N(n) = \frac{1}{2}n + \lambda n^{2/3}.$$

When $\lambda \to -\infty$ then $\Gamma_{n,N}$ consists of many components of the same size as the largest one, which is still very small and consists roughly of $\frac{n^2}{2s^2}\log(s^3/n^2)$ vertices, and the large components are unable to "swallow" each other and therefore are forced to hunt for smaller query. Hence large components grow absorbing only small ones and no clear favorite to win the race for the giant emerges. As the number of edges $N(n)$ increases, the number of contestants decreases. When $\lambda = \mathit{constant} < 0$ the probability that two specified large components will form a new component is bounded away from zero, but still too small to ensure the creation of unique giant component. At the same time, a big gap between the orders of large and small components arises which prevents the creation of new large components from the small ones. Next, as soon as $\lambda \to \infty$, all large components almost "instantly" merge together and a unique large component emerges. This component is still not giant, it has barely over $n^{2/3}$ vertices, but it will continue to absorb other components, first the largest ones, rapidly becoming giant.

The next result of Łuczak [58] gives a clear picture of the sizes $L_i(\Gamma_{n,N})$ of the ith largest components during the phase transition of $\Gamma_{n,N}$. Here and throughout the paper the abbreviation a.s. stands for 'almost surely', a phrase whose precise meaning was explained in the description of Phase 1. above.

Theorem 4.1 ([58]). *Let k be natural number and $sn^{-2/3} \to \infty$ but $s = o(n)$.*

(i) If $N = n/2 - s$ then for every $i = 1, 2, \dots, k$ and every real r

$$\lim_{n\to\infty} P\left(L_i(\Gamma_{n,N}) < \frac{n^2}{2s^2}\left(\log\frac{s^3}{n^2} - \frac{5}{2}\log\log\frac{s^3}{n^2} + r\right)\right) = \sum_{j=0}^{i-1}\frac{\lambda^j}{j!}e^{-\lambda},$$

where $\lambda = \lambda(r) = 2/\sqrt{\pi}e^{-r}$.

Moreover, a.s. the ith largest component of $\Gamma_{n,N}$ is a tree for $i = 1, 2, \dots, k$ and $\Gamma_{n,N}$ contains no component with more edges than vertices.

(ii) Let $N = n/2 + s$ and let s' be the unique positive solution of the equation

$$\left(1 - \frac{2s'}{n}\right)e^{\frac{2s'}{n}} = \left(1 + \frac{2s'}{n}\right)e^{-\frac{2s'}{n}}.$$

Then a.s.

$$\left|L_1(\Gamma_{n,N}) - \frac{2(s+s')n}{n+2s}\right| < \omega(n)\frac{n}{\sqrt{s}}$$

and so

$$|L_1(\Gamma_{n,N}) - 4s| < \omega(n)\frac{n}{\sqrt{s}} + O\left(\frac{s^2}{n}\right).$$

Moreover, for every $i =, 2, \dots, k$ and every real r

$$\lim_{n\to\infty} P\left(L_i(\Gamma_{n,N}) < \frac{n^2}{2s^2}\left(\log\frac{s^3}{n^2} - \frac{5}{2}\log\log\frac{s^3}{n^2} + r\right)\right) = \sum_{j=0}^{i-1}\frac{\lambda^j}{j!}e^{-\lambda},$$

where $\lambda = \lambda(r) = 2/\sqrt{\pi}e^{-r}$.

Furthermore a.s. the ith largest component of $\Gamma_{n,N}$, $i = 2, 3, \dots, k$, is a tree and no component of $\Gamma_{n,N}$, except for the largest one, contains more edges than vertices.

To study the critical "interval" when the phase transition takes place, i.e., when $N(n) = \frac{1}{2}n + \lambda n^{2/3}$ and $\lambda \to \mp\infty$, requires very sophisticated and delicate tools. Janson, Knuth, Łuczak and Pittel in their extensive, almost 140 pages long, study [40] applied machinery of generating functions with great success. They were able to analyze the structure of evolving graphs (and multigraphs) when edges are added one at a time and at random, with great precision, mainly looking and so called excess and deficiency of a graph. To give the reader a taste of their results let us quote the following theorem.

Theorem 4.2 ([40]). *The probability that a random graph or multigraph with n vertices and $\frac{1}{2}n + O(n^{1/3})$ edges has exactly r bicyclic components (i.e., components with exactly two cycles), and no components of higher cyclic order, is*

$$\left(\frac{5}{18}\right)^r \sqrt{\frac{2}{3}} \frac{1}{(2r)!} + O(n^{-1/3}).$$

They also study the following fascinating problem: What is the probability that the component which during the evolution becomes the first "complex" component (i.e., the first component with more than one cycle) is the only complex component which emerges during the whole process? So they ask what is the probability that the first bicyclic component is the "seed" for the giant one. They prove that it happens quite often indeed.

Theorem 4.3 ([40]). *The probability that an evolving graph or multigraph on n vertices never has more than one complex component throughout its evolution approaches $\frac{5\pi}{18} \approx 0.8727$ as $n \to \infty$.*

5. Planarity and Chromatic Number

In a paper of such an enormous length one can likely find less rigorous claims. One of such things happened in the paper [ER60] in relation to the question when a random graph $\Gamma_{n,N}$ is planar.

Since trees and components with exactly one cycle are planar, Erdős and Rényi easily deduced from their findings about early stages of the evolution of a random graph, that when $c < 1/2$ then the probability that $\Gamma_{n,N}$ is planar tends to 1. Now, to support the claim that when c passes $1/2$ the graph becomes non-planar they used the argument that $\Gamma_{n,N}$ contains an induced cycle with d diagonals. Although their claim (Theorem 8a on page 51) regarding the distribution of the number of such cycles is incorrect, as it was pointed out later by Łuczak and Wierman [63], their intuition was perfect and the following result is indeed true.

Theorem 5.1 ([63]). *Let us suppose that $N \sim cn$. If $c < 1/2$ the probability that the graph $\Gamma_{n,N}$ is planar is tending to* 1 *while for $c > 1/2$ this probability tends to* 0.

Such a behavior of a random graph shows the fundamental difference in its typical structure before and after the phase transition. Now, thanks to the contribution of Łuczak, Pittel and Wierman [64], we have more detailed knowledge about planarity of a random graph, also during the phase transition.

Theorem 5.2 ([64]). *Let $\epsilon = \epsilon(n) \to 0$ as $n \to \infty$. Then $\Gamma_{n,p}$ is:*

(i) *a.s. planar, when* $p = (1-\epsilon)/n, \epsilon^3 n \to \infty$*;*

(ii) *Planar with probability tending to* $a(\lambda)$*,* $0 < a(\lambda) < 1$*, as* $n \to \infty$*, when* $p = (1+\epsilon)/n$*, where* $\epsilon^3 n \to \lambda$ *and* $-\infty < \lambda < \infty$ *is a constant;*

(iii) *a.s. non-planar, when* $p = (1+\epsilon)/n, \epsilon^3 n \to \infty$*.*

In the final section of the paper [ER60] Erdős and Rényi collected unsolved problems. One of them is closely related to planarity: *Another interesting question is: what is the threshold for the appearance of a "topological complete graph of order k", i.e., of k points such that any two of them can be connected by a path and these paths do not intersect. For* $k > 4$ *we do not know the solution.* The solution was found many years later by Ajtai, Kómlos, and Szemerédi [2].

Another problem mentioned there turned out to be one of the central and most challenging questions of the theory. Erdős and Rényi asked "*what will be the chromatic number of* $\Gamma_{n,N}$ *?*" What they knew then about this important graph invariant was limited to facts which can be deduced from general results regarding the evolutionary process. Here is what they were able to conclude : "*Clearly every tree can be colored by 2 colors, and thus by Theorem 4a almost surely* $Ch(\Gamma_{n,N}) = 2$ *if* $N(n) = o(n)$*. As however the chromatic number of a graph having an equal number of vertices and edges is equal to* 2 *or* 3 *according whether the only cycle contained in such graph is of even or odd order, it follows from Theorem 5e that almost surely* $Ch(\Gamma_{n,N}) \leq 3$ *for* $N(n) \sim nc$ *with* $c < 1/2$*. For* $N(n) \sim n/2$ *we have almost surely* $Ch(\Gamma_{n,N}) \geq 3$*. As a matter of fact, in the same way, as we proved Theorem 5b, one can prove that* $\Gamma_{n,N}$ *contains for* $N(n) \sim n/2$ *almost surely a cycle of odd order. It is an open problem how large* $Ch(\Gamma_{n,N})$ *is for* $N(n) \sim n/2$ *with* $c > 1/2$*.*"

This question remained open for next 30 years, and was answered, for large c, by Łuczak in [57]. He proved that the chromatic number $\chi(\Gamma_{n,p})$ behaves as follows.

Theorem 5.3 ([**57**]). *Let* $np = c$ *and* $\epsilon > 0$ *be fixed. Suppose* $c_\epsilon \leq c + o(n)$ *for sufficiently large constant* c_ϵ*. Then*

$$P\left(\frac{c}{2\log c} < \chi(\Gamma_{n,p}) < (1+\epsilon)\frac{c}{2\log c}\right) \to 1 \quad as \quad n \to \infty.$$

Although the original question was posed for sparse random graphs the ideas leading to the proof came from investigations of the chromatic number of dense random graphs. The first step toward the solution was made by Matula [66, 67] and Bollobás and Erdős [16] who discovered high concentration of the size of the largest independent set in $\Gamma_{n,p}$ around $2\log_b n$, where $b = 1/(1-p)$ and edge probability p is a constant. It suggested that the respective lower bound for $\chi(\Gamma_{n,p})$ should be $n/(2\log_b n)$. Only a few years later, Grimmett and McDiarmid published a paper [38] in which they showed that a greedy algorithm, which assigns colors to vertices of a random graph sequentially, in such a way that a vertex gets the first available color, needs, with high probability, approximately $n/\log_b n$ colors to produce a proper coloring of $\Gamma_{n,p}$. It established an upper bound for the chromatic number of dense random graph, twice as large as the lower bound. Grimmett and McDiarmid conjectured that the lower bound sets, in fact, the correct order of magnitude for $\chi(\Gamma_{n,p})$. The right tool to settle this conjecture was delivered by Shamir and Spencer [76]. They proved that the chromatic number of $\Gamma_{n,p}$ is sharply concentrated in an interval of length of order $n^{1/2}$ but, what perhaps was more important then their result itself, they introduced to the theory of random graphs a new powerful technique based on concentration measure of martingales, known in the probabilistic literature as Hoeffding-Azuma inequality. But it was Béla Bollobás who showed how the potential of martingale approach can be utilized to solve long standing conjecture. In his paper [15] he proved the following theorem.

Theorem 5.4 ([**15**]). *Let* $0 < p < 1$ *be fixed and* $b = 1/(1-p)$. *Then for every* $\epsilon > 0$

$$P(\frac{n}{2\log_b n} < \chi(\Gamma_{n,p}) < (1+\epsilon)\frac{n}{2\log_b n}) \to 1 \quad as \quad n \to \infty.$$

Later on Matula and Kucera [68] gave an alternative proof of the above theorem, using the second moment and "expose and merge" algorithmic approach. Łuczak's proof of Theorem 5.3 is in fact an ingenious blend of the martingale and "expose and merge" techniques.

The chromatic number of a random graph is a random variable, the distribution of which should be highly concentrated. It is easy to notice (see above) that if $p = o(n^{-1})$ then $\chi(\Gamma_{n,p})$ is 2 (not counting the case when the edge probability is of the order smaller then n^{-2} and therefore, with high probability the graph is empty). One can also show that when $p \sim cn^{-1}$, $O < c < 1$ then $P(\chi(\Gamma_{n,p}) = 2) \to a$ and $P(\chi(\Gamma_{n,p}) = 3) \to 1 - a$, where $a = e^{c/2}((1-c)/(1+c))^{1/4}$. The last probabilities are simply the same as the probabilities that $\Gamma_{n,p}$ has or does not have an odd cycle. Such a behavior of a random variable χ has been confirmed, for small edge probabilities only, by Łuczak. He proved in [61] that if $p < n^{-5/6-\epsilon}$ then the chromatic number, as expected, takes on at most two values.

6. Asymmetric Graphs

Another interesting topic originated from a joint paper by Erdős and Rényi in the peak of their cooperation in early 1960s [ER63]. Here is how they describe their goals: "*We shall call (...) a graph* symmetric*, if there exists a non-identical permutation of its vertices, which leaves the graph invariant. By other words, a graph is called symmetric if the group of its automorphisms has degree greater than 1. A graph which is not symmetric will be called* *asymmetric. The degree of symmetry of a symmetric graph is evidently measured by the degree of its group of automorphisms. The question which led us to the results contained in the present paper is the following: how can we measure the degree of asymmetry of an asymmetric graph?*"

They answer the last question in what follows: "*Evidently any asymmetric graph can be made symmetric by deleting certain of its edges and by adding certain new edges connecting its vertices. We shall call such a transformation of the graph its symmetrization. For each symmetrization of the graph let us take the sum of the number of deleted edges – say r – and the number of new edges – say s –; it is reasonable to define the degree of asymmetry $A[G]$ of a graph G, as the minimum of $r+s$ where the minimum is taken over all possible symmetrizations of the graph G. (...) The question arises: how large can be the degree of asymmetry of a graph of order n (i.e., a graph which has n vertices)? We shall denote by $A(n)$ the maximum of $A[G]$ for all graphs G of order $n(n=2,3,\dots)$.*"

They first notice that $A(2)=A(3)=A(4)=A(5)=0$ while $A(6)=1$. In general, a rather straightforward deterministic argument leads to the following result.

Theorem 6.1 ([**ER63**]).

$$A(n) \leq \left\lfloor \frac{n-1}{2} \right\rfloor.$$

To find the lower bound for $A(n)$ Erdős and Rényi use a non-constructive argument, i.e., they show via the probabilistic method that there exists a certain graph on n vertices with the degree of asymmetry at least $n(1-\epsilon)/2$, $0<\epsilon<1$.

Theorem 6.2 ([**ER63**]). *Let us choose at random a graph Γ having n given vertices so that all possible $2^{\binom{n}{2}}$ graphs should have the same probability to be chosen. Let $\epsilon>0$ be arbitrary. Let $P_n(\epsilon)$ denote the probability that by changing not more than $\frac{n(1-\epsilon)}{2}$ edges of Γ it can be transformed into a symmetric graph. Then we have*

$$\lim_{n\to\infty} P_n(\epsilon)=0.$$

Corollary 6.1. *For any ϵ with $0<\epsilon<1$ there exists an integer $n_0(\epsilon)$ depending only on ϵ, such that for every $n>n_0(\epsilon)$ there exists a graph G of order n with $A[G]>n(1-\epsilon)/2$.*

Indeed, for large n, Theorem 6.2 shows that almost every graph is a counterexample to the hypothesis that its symmetrization is possible with less than $\frac{n}{2}(1-o(1))$ edges.

Hence, if we combine Theorem 6.1 and Corollary 6.1 we see that

$$\lim_{n\to\infty} \frac{A(n)}{n} = \frac{1}{2}.$$

After showing that almost all labeled simple graphs are asymmetric, Erdős and Rényi turned their attention to graphs with a prescribed number of edges. First they noticed that since almost every tree has a cherry, i.e., a pair of pendant vertices adjacent to a common neighbor, therefore almost every tree on n vertices is symmetric. Furthermore they proved that any connected graph of order n having n edges is either symmetric or its asymmetry is one and gave the following bound.

Theorem 6.3 ([**ER63**]). *If a graph G of order n has $N = \lambda n$ edges* ($0 < \lambda < (n-1)/2$) *then*

$$A[G] \leq 4\lambda \left(1 - \frac{2\lambda}{n-1}\right).$$

Erdős and Rényi went further in their investigations. Let us quote a few more lines from their paper [ER63]. "*Another interesting question is to investigate the asymmetry or symmetry of a graph for which not only the number of vertices but also the number of edges N is fixed, and to ask that if we choose one of these graphs at random, what is the probability of its being asymmetric. We have solved this question too, and have shown that if $N = \frac{n}{2}(\log n + \omega(n))$, where $\omega(n)$ tends arbitrarily slowly to $+\infty$ for $n \to +\infty$, then the probability that a graph with n vertices and N edges chosen at random (so that any such graph has the same probability $\binom{\binom{n}{2}}{N}^{-1}$ to be chosen) should be asymmetric, tends to 1 for $n \to +\infty$. This and some further results will be published in another forthcoming paper.*"

Unfortunately the announced paper has never been published! Several years later this problem and the analogous one for unlabeled graphs was attacked again by Wright [78].

Consider graphs $\Gamma_{n,N}$ and $U_{n,N}$ picked at random from the families of all labeled and unlabeled graphs on n vertices and with $N = N(n)$ edges, respectively. Here is the result of Wright.

Theorem 6.4 ([**78**]). *If $\omega(n) = (2N(n)/n) - \log n \to \infty$ then $\Gamma_{n,N}$ and $U_{n,N}$ are almost surely asymmetric while when $\omega(n) \leq 0$ then they are almost surely symmetric.*

Later Łuczak [56] gave precise results about the structure of the automorphism group $Aut(\Gamma_{n,N})$ of a random graph $\Gamma_{n,N}$. He studied the symmetry of the largest component $L_1(n, N)$ of this random graph. What he found was that when $N(n) = \frac{1}{2}n\alpha(n)$ then there exists a constant d such that for $\alpha(n) \geq d$ almost surely $Aut(L_1(n,N)$ is isomorphic to some product of symmetric groups. From this result he was able to deduce the following strengthening of the "labelled" part of Theorem 6.4.

Theorem 6.5 ([**56**]). *Let* $N = \frac{n}{2}(\log n + \omega(n))$.

(i) If $\omega(n) \to -\infty$ *then* $|Aut(\Gamma_{n,N})| \to \infty$ *a.s.*
(ii) If $\omega(n) \to c$ *then*

$$\lim_{n\to\infty} P(|Aut(\Gamma_{n,N})| = 1) = e^{\lambda}(1+\lambda)$$

$$\lim_{n\to\infty} P(|Aut(\Gamma_{n,N})| = k!) = \frac{\lambda^k}{k!}e^{-\lambda}$$

for $k = 2, 3, \dots,$ *where* $\lambda = e^{-c}$ *and* c *is a constant.*
(iii) If $\omega(n) \to \infty$ *then* $|Aut(\Gamma_{n,N})| = 1$ *a.s.*

7. Perfect Matchings

The last three papers Erdős and Rényi wrote on the subject of random graphs were devoted to the existence of 1-factors. In [ER64] and [ER68] they coped with the relatively easier case of random bipartite graphs. In both papers they consequently emphasized the matrix terminology. "*In the present paper we deal with certain random* 0-1 *matrices. Let* $\mathcal{M}(n,N)$ *denote the set of all* n *by* n *square matrices among the elements of which there are exactly* N *elements* ($n \leq N \leq n^2$) *equal to* 1, *all the other elements are equal to* 0. *The set* $\mathcal{M}(n,N)$ *contains clearly* $\binom{n^2}{N}$ *such matrices; we consider a matrix* M *chosen at random from the set* $\mathcal{M}(n,N)$, *so that each element of* $\mathcal{M}(n,N)$ *has the same probability* $\binom{n^2}{N}^{-1}$ *to be chosen. We ask how large* N *has to be, for a given large value of* n, *in order that the permanent of the random matrix* M *should be different from zero with probability* $\geq \alpha$, *where* $0 < \alpha < 1$. *(...) A second way to formulate the problem is as follows: we shall say that two elements of a matrix are in independent position if they are not in the same row and not in the same column. Now our question is to determine the probability that the random matrix* M *should contain* n *elements which are all equal to 1 and pairwise in independent position.*"

The result they prove resembles that for the connectedness (compare Theorem 2.1).

Theorem 7.1 ([**ER66**]). *Let* $P(n,N)$ *denote the probability of the event that the permanent of the random matrix* M *is positive. Then if*

$$N(n) = n\log n + cn + o(n)$$

where c is any real constant, we have

$$\lim_{n\to\infty} P(n, N(n)) = e^{-2e^{-c}}.$$

Finally, they also mention graphs: "*This result can be interpreted also in the following way, in terms of graph theory. Let* $\Gamma_{n,N}$ *be a bichromatic random graph containing* n *red and* n *blue vertices, and* N *edges which are chosen at random among the* n^2 *possible edges connecting two vertices having different color (so that each of the* $\binom{n^2}{N}$ *possible choices has the same probability). Then* $P(n, N)$ *is equal to the probability that the random graph* $\Gamma_{n,N}$ *should contain a factor of degree* 1*, i.e.,* $\Gamma_{n,N}$ *should have a subgraph which contains all vertices of* $\Gamma_{n,N}$ *and* n *disjoint edges, i.e.,* n *edges which have no common endpoint.*" (They seem not to use the name 'perfect matching' at all.)

As far as the proof is concerned, "*Besides elementary combinatorial and probabilistic arguments similar to that used by us in our previous work on random graphs (...) our main tool in proving our results is the well-known theorem of D. König, which is nowadays well known in the theory of linear programming, according to which if* M *is an* n *by* n *matrix, every element of which is either* 0 *or* 1*, then the minimal number of lines (i.e., rows or columns) which contain all the* 1*-s, is equal to the maximal number of* 1*-s in independent position. As a matter of fact, for our purposes we need only the special case of this theorem, proved already by Frobenius (1917), concerning the case when the maximal number of ones in independent positions is equal to* n *(...). According to the theorem of Frobenius-König* $1 - P(n, N)$ *is equal to the probability that there exists a number* k *such that there can be found* k *rows and* $n-k-1$ *columns of* M *which contain all the ones (*$0 \le k \le n-1$*).*" The rest of the proof is devoted to showing that this is very unlikely for $N(n)$ given. It is interesting to notice that Erdős and Rényi never mention Hall's theorem, which is equivalent to Frobenius but far more popular in combinatorics nowadays.

The 1968 paper is a straightforward extension of the 1964 result, where it is shown that setting

$$N(n) = n \log n + (r-1)n \log\log n + n\omega(n)$$

where $\omega(n)$ tends arbitrarily slowly to infinity then almost surely the bichromatic random graph contains r disjoint 1-factors. The only new element of the proof is the observation that if there are no r disjoint 1-factors then there is a way to delete some edges so that no vertex looses more than $r-1$ from its degree and the resulting subgraph contains no 1-factor at all. Then again the theorem of Frobenius is used.

The most involved of the three papers about 1-factors is that from 1966, where an ordinary (not bichromatic) random graph $\Gamma_{n,N}$ is considered. The reason is that the theorem of Tutte describing the structure of graphs which admit 1-factors is more complex than its counterpart in the bipartite case. "*It should be added that the problem investigated in the present paper is much more difficult than the corresponding problem for even graphs solved in [5]. Thus for instance in [5] we made use of the well known theorem of D. König; the corresponding tool in the present paper is the much deeper theorem of Tutte mentioned above.*" ([5] = [ER 64])

The result of that paper says that the threshold for containing a 1-factor coincides with that for disappearance of isolated vertices, and thus also with that for connectivity (see Theorem 2.1). The proof is long and tedious and involves a weaker version of Tutte's theorem ignoring the parity of components.

Erdős and Rényi make also the following claim: "*If* $N = \frac{1}{2}n\log n + O(n)$, *as mentioned above, with probability near to* 1, $\Gamma_{n,N}$ *consists of a connected component and a certain number of isolated points. With the same method (...) one can prove that if the connected component of* $\Gamma_{n,N}$ *consists of an even number of points, it has with probability near* 1 *a factor of degree one. As the proof of this result is almost the same (...) we do not go into the details.*"

The above mentioned result was proved (in a strengthened form) by Bollobás and Thomason [18]. In order to quote that result let us extend the notion of a perfect matching by saying that a graph satisfies property $\mathcal{PM}$ if there is a matching covering all but at most one of the nonisolated vertices. It is known that, switching to the binomial model, as soon as $2np - \log n - \log\log n \to \infty$, there are only isolated vertices outside the giant component. However, the main obstacle for the property $\mathcal{PM}$ is the presence of a pair (at least two such pairs when the number of nonisolates is odd) of vertices of degree 1 adjacent to the same vertex (called, as we already mentioned, 'a cherry'). The expected number of cherries is

$$3\binom{n}{3}p^2(1-p)^{2(n-3)} < n^3p^2e^{-2np+6p} = o(1)$$

if $2np - \log n - 2\log\log n \to \infty$. Again, a trivial necessary condition becomes almost surely sufficient.

Theorem 7.2 ([**18**]). *Let* $y_n = 2np - \log n - 2\log\log n \to \infty$. *Then*

$$P(\Gamma_{n,p} \in \mathcal{PM}) \to \begin{cases} 0 & \text{if } y_n \to -\infty \\ e^{-\frac{1}{8}e^{-c}} & \text{if } y_n \to c \\ 1 & \text{if } y_n \to \infty. \end{cases}$$

The proof, again, was based on Tutte's theorem. Years later Łuczak and Ruciński proposed an alternative approach, via Hall's Theorem, invented in [65] to attack a more general question. For a given graph G, a perfect G-matching of a graph is a spanning subgraph which is a disjoint union of copies of G. For $G = K_2$ this is the ordinary notion of a 1-factor.

In [65] it was shown that for every nontrivial tree T, the threshold is the same as that for disappearance of isolated vertices.

Theorem 7.3 (**[65]**). *For every tree T on t vertices and with at least one edge, assuming n is divisible by t,*

$$P(\Gamma_{n,p} \textit{ has a perfect } T\textit{-matching }) \to \begin{cases} 0 & \textit{if } np - \log n \to -\infty \\ e^{-e^{-c}} & \textit{if } np - \log n \to c \\ 1 & \textit{if } np - \log n \to \infty. \end{cases}$$

The threshold for arbitrary G is not known in general. Some partial results are contained in [4] and [72].

Coming back to the original papers of Erdős and Rényi, the last of them is concluded by the following problem: "*does a random graph $\Gamma_{n,N}$ where n is even and*

$$N = \frac{1}{2} n \log n + \frac{r-1}{2} n \log\log n + \omega(n) n$$

where $\omega(n) \to \infty$, contain at least r disjoint factors of degree one with probability tending to 1 *for $n \to \infty$?*"

Shamir and Upfal [77] answered this question in the positive. Given a map f of $V(G)$ into the set of non-negative integers, define an f-factor of G as a spanning subgraph of G in which the degree of vertex x is $f(x)$.

Theorem 7.4 (**[77]**). *If*

$$p = \frac{1}{n}(\log n + (r-1)\log\log n + \omega(n),$$

$r \geq 1$, $\lim_{n\to\infty} \omega(n) = \infty$ *and* $1 \leq f(x_i) \leq r$, $\sum_{i=1}^{n} f(x_i)$ *even, then* $\Gamma_{n,p}$ *has an f-factor, almost surely.*

Although f-factors are characterized by Tutte's theorem, Shamir and Upfal chose an alternative approach using an algorithmic technique (introduced to random graphs by Pósa) of augmentation of sub-factors by alternating paths. In fact, the answer to the last question of Erdős and Rényi does not follow directly from the above result (not every r-factor has a 1-factorization) but from the proof. In 1985 Bollobás and Frieze [17] strengthened this answer by proving that almost surely in the random graph process of adding edges one by one, as soon as the minimum degree becomes r, there are $\lfloor r/2 \rfloor$ disjoint hamiltonian cycles plus a disjoint perfect matching if r is odd.

The next problem we would like to mention cannot be directly attributed to Erdős and Rényi. Here is how Erdős describes their omission [3, Appendix B]. "*When Rényi and I developed our theory of random graphs, we thought of extending our study for hypergraphs. We mistakenly thought that* *all (or most) of the extensions would be routine and we completely overlooked the following beautiful question of Shamir. (...) Shamir asked how many triples must one choose on* $3n$ *elements so that with probability bounded away from zero one should get* n *vertex disjoint triples. Shamir proved that* $n^{3/2}$ *triples suffice, but the truth may very well be* $n^{1+\epsilon}$ *or even* $cn \log n$*. The reason for the difficulty is that Tutte's theorem seem to have no analogy for triple systems or more generally for hypergraphs.*" The result mentioned by Erdős belongs, in fact, to J. Schmidt-Pruzan and E. Shamir [75]. In 1995, Frieze and Janson in [35] pushed the bound down to $n^{4/3}$.

Fortunately, Erdős and Rényi did not overlook some other important problems which stimulated the research in the theory of random graphs over the years. One such problem was the threshold for existence of a Hamiltonian cycle in a random graph. They, in fact, asked only: *for what order of magnitude of* $N(n)$ *has* $\Gamma_{n,N(n)}$ *with probability tending to* 1 *a Hamilton-line (i.e., a path which passes through all vertices).* This problem was first tried by Pósa [69] and Korshunov [55] and finally solved by Kómlos and Szemerédi [54] and, in a stronger form, by Bollobás [12]. They proved that the threshold for Hamiltonian cycle coincides with that of disappearance of all vertices of degree 0 and 1.

8. Update for the Second Edition

We wrote this paper back in 1995. In this second edition of the volume we decided to leave the original text intact except for a few obvious corrections and the proofs of Theorems 3.2 and 3.3 which have been deleted entirely. However, several new developments have occurred afterward. Here we would like to mention some of them along with a couple of earlier results omitted in the first edition. Needles to say, our choice is quite subjective. For more thorough treatment of random graphs we refer the reader to the monograph [43] published in 2000.

In relation to connectivity, one should note that an old result of Łuczak [60] states that the k-core of a random graph, for p large enough, is $a.s.$ empty or k-connected. It implies that $\Gamma_{n,p}$ is a.s. $c(\Gamma_{n,p})$-connected for the ranges of N larger than those in Theorem 2.2.

In the domain of small subgraphs of random graphs there has been an intense study of the so called *upper tail* of the random variable X_G counting copies of a given graph G in $\Gamma_{n,p}$. As far as the *lower tail* is concerned, whose special case is the probability $P(X = 0)$ discussed briefly after Theorem 3.2, the asymptotic order of magnitude of the logarithm of $P(X \leq (1-\epsilon)EX)$ has been determined in [39] to be $-\Phi_G$. The exponent in the upper tail, $P(X \geq (1+\epsilon)EX)$, is of a smaller order of magnitude which is still to be determined. In [44] general lower and upper bounds were obtained which differ only by a logarithmic factor. Very recently DeMarco and Kahn [21] have found the right threshold for cliques and formulated the "right" conjecture for the general case.

In Sect. 5, the threshold for topological cliques found in [2] has been sharpened (see [62], a remark after Corollary 18). A significant result about the chromatic number of a random graph appeared in [1]. Achlioptas and Naor found therein an explicit two-point limiting distribution of $\chi(\Gamma_{n,p})$, where $p = d/n$, for every $d > 0$, strengthening a theorem from [61] mentioned at the end of Sect. 5.

The most acclaimed result in random graph theory which appeared after 1995 is, without doubt, a solution to the celebrated Shamir problem posed in Sect. 7. After some initial attempts (Krivelevich [52, 53] and Kim [50]), in 2008 Johansson, Kahn, and Vu [46] published a complete solution to both, the hypergraph Shamir problem and to its random graph counterpart (triangle-factors), receiving for their achievement the prestigious Fulkerson Prize. Quite recently in a series of papers, Dudek, Frieze, Loh, and Speiss [22–24, 34] obtained thresholds for the hamiltonicity of random uniform hypergraphs. In the hardest case of so called loose Hamilton cycles they incorporated in their proofs the result on perfect matchings from [46].

Acknowledgements We would like to thank Tomasz Łuczak for his invaluable help in updating this paper. We are also grateful to Steve Butler for turning our obsolete amstex file from 1995 into a modern latex file.

References

1. D. Achlioptas and A. Naor, *The two possible values of the chromatic number of a random graph*, Ann. of Math. **162**(2) (2005), no. 3, 1335–1351.
2. M. Ajtai, J. Komlós and E. Szemerédi, *Topological complete subgraphs in random graphs*, Studia. Sci. Math. Hungar. **14** (1979), 293–297.
3. N. Alon and J. Spencer, *The Probabilistic Method*, 1992, Wiley.
4. N. Alon and R. Yuster, *Threshold functions for H-factors*, Combinatorics, Probability and Computing **2** (1993), 137–144.
5. A.D. Barbour, *Poisson convergence and random graphs*, Math. Proc. Cambr. Phil. Soc. **92** (1982), 349–359.
6. A.D. Barbour, S. Janson, M. Karoński and A. Ruciński, *Small cliques in random graphs*, Random Structures Alg. **1** (1990), 403–434.
7. A.D. Barbour, M. Karoński and A.Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, J. Comb. Th.-B **47** (1989), 125–145.
8. P. Billingsley, *Probability and Measure*, 1979, Wiley.
9. B. Bollobás, *Threshold functions for small subgraphs*, Math. Proc. Cambr. Phil. Soc. **90** (1981), 197–206.
10. B. Bollobás, *Vertices of given degree in a random graph*, J. Graph Theory **6** (1982), 147–155.

11. B. Bollobás, *Distinguishing vertices of random graphs*, Annals Discrete Math. **13** (1982), 33–50.
12. B. Bollobás, *Almost all regular graphs are Hamiltonian*, Europ. J. Combinatorics **4** (1983), 97–106.
13. B. Bollobás, *The evolution of random graphs*, Trans. Amer. Math. Soc. **286** (1984), 257–274.
14. B. Bollobás, *Random Graphs*, Academic Press, London, 1985.
15. B. Bollobás, *The chromatic number of random graphs*, Combinatorica **8** (1988), 49–55.
16. B. Bollobás and P. Erdős, *Cliques in random graphs*, Math. Proc. Cambr. Phil. Soc. **80** (1976), 419–427.
17. B. Bollobás and A. Frieze, *On matchings and hamiltonian cycles in random graphs*, in: Random Graphs '83, Annals of Discrete Mathematics **28** (1985), 1–5.
18. B. Bollobás and A. Thomason, *Random graphs of small order*, Annals of Discrete Math. **28** (1985), 47–98.
19. B. Bollobás and A. Thomason, *Threshold functions*, Combinatorica **7** (1987), 35–38.
20. B. Bollobás and J.C. Wierman, *Subgraph counts and containment probabilities of balanced and unbalanced subgraphs in a large random graph*, in: Graph Theory and Its Applications: East and West (Proc. 1st China-USA Intern. Graph Theory Conf.), Eds. Capobianco et al., Annals of the New York Academy of Sciences **576** (1989), 63–70.
21. R. DeMarco and J. Kahn, *Tight upper tail bounds for cliques* **41** (2012), 469487.
22. [DFl] A. Dudek and A. Frieze, *Loose Hamilton Cycles in Random k-Uniform Hypergraphs* Electronic Journal of Combinatorics, **18** (2011) P48.
23. A. Dudek and A. Frieze, *Tight Hamilton Cycles in Random Uniform Hypergraphs*, Random Structures Alg., to appear.
24. A. Dudek, A. Frieze, P.-S. Loh and S. Speiss, *Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs* , Electronic Journal of Combinatorics **19** (2012), P44.
25. P. Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294.
26. P. Erdős and A. Rényi, *On random graphs I*, Publ. Math. Debrecen **6** (1959), 290–297.
27. P. Erdős and A. Rényi *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. **5** (1960), 17–61.
28. P. Erdős and A. Rényi, *On the evolution of random graphs*, Bull. Inst. Internat. Statist. **38** (1961), 343–347.
29. P. Erdős and A. Rényi, *On the strength of connectedness of a random graph*, Acta Math. Acad. Sci. Hungar. **12** (1961), 261–267.
30. P. Erdős and A. Rényi, *Asymmetric graphs*, Acta Math. Acad. Sci. Hung. **14** (1963), 295–315.
31. P. Erdős and A. Rényi, *On random matrices*, Publ. Math. Inst. Hung. Acad. Sci. **8** (1964), 455–461.
32. P. Erdős and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hung. **17** (1966), 359–368.
33. P. Erdős and A. Rényi *On random matrices II*, Studia Sci. Math. Hung. **3** (1968), 459–464.
34. A. Frieze *Loose Hamilton Cycles in Random 3-Uniform Hypergraphs* Electronic Journal of Combinatorics **17** (2010) N28
35. A. Frieze and S. Janson, *Perfect Matchings in Random s-Uniform Hypergraphs* Random Structures and Algorithms **7** (1995) 41–57.

36. E. N. Gilbert, *Random graphs*, Annals of Mathematical Statistics **30** (1959), 1141–1144.
37. E. Godehardt and J. Steinebach, *On a lemma of P. Erdős and A. Rényi about random graphs*, Publ. Math. **28** (1981), 271–273.
38. G.R. Grimmett and C.J.H. McDiarmid, *On colouring random graphs*, Math. Proc. Cambr. Phil. Soc. **77** (1975), 313–324.
39. S. Janson, *Poisson approximation for large deviations*, Random Structures & Algorithms **1** (1990), 221–229.
40. S. Janson, D.E. Knuth, T. Łuczak and B. Pittel, *The birth of the giant component*, Random Structures & Algorithms **4** (1993), 233–358.
41. S. Janson and J. Kratochvil, *Proportional graphs*, Random Structures & Algorithms **2** (1991), 209–224.
42. S. Janson, T. Łuczak and A.Ruciński, *An exponential bound for the probability of nonexistence of a specified subgraph of a random graph*, in: Proceedings of Random Graphs '87, Wiley, Chichester, 1990, 73–87.
43. S. Janson, T. Łuczak and A. Ruciński, *Random Graphs* Wiley, (2000).
44. S. Janson, K. Oleszkiewicz and A. Ruciński, *Upper tails for subgraph counts in random graphs*, Israel J. Math. **141** (2004), 61–92.
45. S. Janson and J. Spencer, *Probabilistic constructions of proportional graphs*, Random Structures & Algorithms **3** (1992), 127–137.
46. A. Johansson, J. Kahn and V. Vu, *Factors in random graphs*, Random Structures and Algorithms **33** (2008), 1–28.
47. M. Karoński, *On the number of k-trees in a random graph*, Prob. Math. Stat., **2** (1982), 197–205.
48. M. Karoński and A. Ruciński, *On the number of strictly balanced subgraphs of a random graph*, in: Graph Theory, Łagów 1981, Lecture Notes in Math. 1018, Springer-Verlag, 1983, 79–83.
49. J. Kärrman, *Existence of proportional graphs*, J. Graph Theory **17** (1993), 207–220.
50. J. H. Kim, *Perfect matchings in random uniform hypergraphs*, Random Struct. Algorithms **23**(2) (2003), 111–132
51. V.F. Kolchin, *On the limit behavior of a random graph near the critical point*, Theory Probability Its Appl. **31** (1986), 439–451.
52. M. Krivelevich, *Perfect fractional matchings in random hypergraphs*, Random Structures and Algorithms **9** (1996), 317334.
53. M. Krivelevich, *Triangle factors in random graphs*, Combinatorics, Probability and Computing **6** (1997), 337347.
54. J. Komlós and E. Szemerédi, *Limit distributions for the existence of Hamilton cycles*, Discrete Math. **43** (1983), 55–63.
55. A.D. Korshunov, *A solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs*, Metody Diskr. Anal. **31** (1977), 17–56.
56. T. Łuczak, *The automorphism group of random graphs with a given number of edges*, Math. Proc. Camb. Phil. Soc. **104** (1988), 441–449.
57. T. Łuczak, *On the chromatic number of sparse random graphs*, Combinatorica **10** (1990), 377–385.
58. T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Structures & Algorithms **1** (1990), 287–310.
59. T. Łuczak, *On the equivalence of two basic models of random graphs*, in: Proceedings of Random Graphs '87, Wiley, Chichester, 1990, 151–157.
60. T. Łuczak, *Size and connectivity of the k-core of a random graph*, Discrete Math. **91** (1991) 61–68.
61. T. Łuczak, *A note on the sharp concentration of the chromatic number of a random graph*, Combinatorica **11** (1991), 295–297.

62. T. Łuczak, *The phase transition in a random graph*, Combinatorics, Paul Erdős is Eighty, vol.2 (Dezsä Miklós, Vera T.Sós, Tamás Szõnyi, eds.), Budapest, 1996, Bolyai Society Mathematical Studies 2, 399–422.
63. T. Łuczak and J.C. Wierman, *The chromatic number of random graphs at the double-jump threshold*, Combinatorica **9** (1989), 39–49.
64. T. Łuczak, B. Pittel and J.C. Wierman, *The structure of a random graph at the double-jump threshold*, Trans. Am. Math. Soc. **341** (1994), 721–728.
65. T. Łuczak and A. Ruciński, *Tree-matchings in random graph processes*, SIAM J. Discr. Math **4** (1991), 107–120.
66. D. W. Matula, *The employee party problem*, Notices Amer. Math. Soc. **19** (1972), A-382
67. D. W. Matula, *The largest clique size in a random graph*, Tech. Rep. Dept. Comput. Sci., Southern Methodist Univ., Dallas, 1976.
68. D. W. Matula and L. Kucera, *An expose-and-merge algorithm and the chromatic number of a random graph*, in: Random Graphs '87 (J. Jaworski, M. Karoński and A. Ruciński, eds.), John Wiley & Sons, New York, 1990, 175–188.
69. L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), 359–364.
70. A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Prob. Th. Rel. Fields **78** (1988), 1–10.
71. A. Ruciński, *Small subgraphs of random graphs: a survey*, in: Proceedings of Random Graphs '87, Wiley, Chichester, 1990, 283–303.
72. A. Ruciński, *Matching and covering the vertices of a random graph by copies of a given graph*, Discrete Math. **105** (1992), 185–197.
73. A. Ruciński and A. Vince, *Balanced graphs and the problem of subgraphs of random graphs*, Congres. Numerantium **49** (1985), 181–190.
74. K. Schürger, *Limit theorems for complete subgraphs of random graphs*, Period. Math. Hungar. **10** (1979), 47–53.
75. J. Schmidt and E. Shamir, *A threshold for perfect matchings in random d-pure hypergraphs*, Discrete Math. **45** (1983), 287–295.
76. E. Shamir and J. Spencer, *Sharp concentration of the chromatic number of random graphs $G_{n,p}$*, Combinatorica **7** (1987), 121–129.
77. E. Shamir and E. Upfal, *On factors in random graphs*, Israel J. Math.**39** (1981), 296–302.
78. E. M. Wright, *Asymmetric and symmetric graphs*, Glasgow Math. J. **15** (1974), 69–73.

An Upper Bound for a Communication Game Related to Time-Space Tradeoffs

Pavel Pudlák and Jiří Sgall

P. Pudlák (✉)
Academy of Sciences, Institute of Mathematics, Prague, Czech Republic
e-mail: pudlak@math.cas.cz

J. Sgall
Computer Science Institute of Charles University, Prague, Czech Republic

Summary. We prove an unexpected upper bound on a communication game proposed by Jeff Edmonds and Russell Impagliazzo [2, 3] as an approach for proving lower bounds for time-space tradeoffs for branching programs. Our result is based on a generalization of a construction of Erdős, Frankl and Rödl [5] of a large 3-hypergraph with no 3 distinct edges whose union has at most 6 vertices.

1. Introduction

Suppose that we have two vectors u and v of length k. We want to decide whether $u = v$, but our access to the bits is very limited—at any moment we can see at most one bit of each pair of the bits u_i and v_i. You can imagine the corresponding bits to be written on two sides of a card, so that we can see all the cards but only one side of each card. After every flip we can write down some information, but the memory is not reusable—after the next flip we have to use new memory. We are charged for every bit of memory that we use and for every time we flip one or more cards.

It seems natural to suppose that if we flip the cards only a few times, we need a lot of memory. We prove an unexpected upper bound on the amount of memory needed; our bound is asymptotically tight if the number of card flips is constant. Our result is based on a construction of Erdős, Frankl and Rödl [5], whose special case is a large (in the number of edges) 3-hypergraph (a system of 3-element sets) with no 3 distinct edges whose union has at most 6 vertices, and on a previous result of Rusza and Szemerédi [8]. This special case of their construction corresponds to the case when only two flips of the cards are allowed. Our main idea is a geometric interpretation of the construction of a large set with no three element arithmetic progression due to Behrend [1] and its generalization to higher dimensions. We discuss this connection in Sect. 4.

In fact, Jeff Edmonds and Russell Impagliazzo proposed this game as a tool for proving lower bounds in complexity of boolean functions and proved that a reasonable lower bound on the sum of the number of flips and the

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_24,

number of bits of memory used during the game would imply a new lower bound for branching programs [2, 3]. We give a simple protocol with $O(\log n)$ probes which uses only $O((\log n)^2)$ bits of memory. We discuss this connection and protocol in Sect. 5.

First we describe the game more precisely and state some easy facts in Sect. 2, and prove our upper bound in the communication game setting in Sect. 3.

2. The Game

Formally we describe the game as follows. Each input x is divided into some pieces (substrings) $x_1, \ldots, x_r$ in a fixed way (i.e., it is given which coordinate belongs to which piece). Each piece corresponds to a single card. Let $\Pi_1, \ldots, \Pi_T$ be a sequence of subsets of $[1, r]$ and let the inputs u and v be given. Then $\Pi_t(u, v)$ denotes the input consisting of the pieces x_i defined as $x_i = v_i$ if $i \in \Pi_t$ and $x_i = u_i$ if $i \notin \Pi_t$. These input vectors are called *probes*, and each of them corresponds to a choice of a visible side for each card. The protocol is described by the T probes $\Pi_1, \ldots, \Pi_T$ and a function $F(u)$ on the input vectors. If $F(u) = F(\Pi_1(u, v)) = \cdots = F(\Pi_T(u, v))$, the protocol answers "$u = v$", if not, the answer is "$u \neq v$". Let B be the number of bits of memory that we need, i.e., the maximal length of $F(u)$ over all inputs u. We are interested in the dependency of B on T and k. In the context of timespace tradeoff, the most important quantity is the minimal possible $B + T$, which corresponds to the total communication.

This corresponds to a protocol in which we first write down some information about u, and then after each of T flips we just check whether the current probe is consistent with that information. It is obvious that if we discover inconsistency, the vectors are different. Thus a protocol is correct if no two distinct vectors pass the test. In fact, here the use of memory is even more restricted than in the version described in the introduction—effectively the first time the protocol writes down arbitrary information $F(u)$ but then after each flip we write down only a single bit indicating consistency of the current probe. It is easy to show that this restriction increases the amount of memory by at most the maximum of B and T, see [2].

We can describe the set of probes in another equivalent way, more convenient for our proof. For each $i \leq r$ let $V_i \subseteq [1, T]$ denote the set of indices of the probes such that $t \in V_i$ if and only if the ith piece of $\Pi_t(u, v)$ is v_i (as opposed to u_i in other probes). We can assume that all sets V_i are distinct, because the pieces which appear in identical sets of probes can be joined (we can use one card instead of two cards that are always flipped together). Also, for every i, V_i is nonempty, as at least one probe has to look at v_i in a correct protocol. This means that $r \leq 2^T - 1$.

In case of $T = 1$, the game is just the usual communication game with one player having access to u and the other player to v; it is easy to see that k bits of memory are needed to test equality in that case. Similarly, it is easy to see that B must be at least the length of any piece u_i. If we fix all other pieces to be all zeros, then every probe is either u_i or v_i, and a correct protocol needs enough memory to distinguish any two distinct inputs u_i. This shows that $B \geq k/r \geq k/(2^T - 1)$.

We describe the best previously known protocol; it uses $B = \lceil k/T \rceil$ bits of memory [2]. Set $r = T$ and divide the input into r pieces of the same length (we assume without loss of generality that k is divisible by r). Set $\Pi_t = \{t\}$, or equivalently $V_i = \{i\}$ (i.e., each probe looks at just one piece of v and the rest comes from u). Set $F(u) = u_1 \oplus u_2 \oplus \cdots \oplus u_r$ to be the bitwise parity of the pieces. If the protocol answers "$u = v$", we know that $F(u) = F(\Pi_1((u, v)) = v_1 \oplus u_2 \oplus \cdots \oplus u_r$ and hence $u_1 = v_1$. The same argument is valid for other pieces, hence $u = v$ and the protocol is correct.

It is easy to prove that no protocol in which the function F is linear (over $GF(2)$) can be better. The equation $F(u) = F(\Pi_1(u, v)) = \cdots = F(\Pi_T(u, v))$ translates into a system of BT linear equations with $2k$ unknowns. If the protocol is correct, then the points in the k-dimensional subspace defined by $u = v$ are the only solutions of the system of equations, and hence there have to be at least k equations. This gives $B \geq k/T$.

Jeff Edmonds conjectured that this is in fact optimal even for non-linear protocols, i.e., $B = \Omega(k/T)$ for every protocol [2]. We disprove this conjecture. In fact, we prove that for constant T the easy lower bound is much closer to the truth as our protocol needs only $k/(2^T - 1) + O(\sqrt{k})$ bits of memory. If the number of probes is not bounded it is possible to achieve $B + T = O((\log k)^2)$ using a very simple protocol. We discuss this protocol and its consequences in Sect. 5.

3. The Upper Bound

Theorem 1. *For each parameter d and for each T there exists a protocol with T probes such that the number of bits of memory B is at most*

$$\left(1 + \frac{T}{d}\right) \left\lceil \frac{k}{2^T - 1} \right\rceil + (T + 2d + 1 + \log k) 2^{2T-1}$$

Corollary 1. *For T constant and $d = \sqrt{k}$ the bound is $k/(2^T - 1) + O(\sqrt{k})$.*

Proof of Theorem 1. First we present the **protocol**.

Each input vector is divided into $r = 2^T - 1$ pieces $u_1, \ldots, u_r$ of the same length $l = \lceil k/r \rceil$. We define the probes by taking the sets V_i, $i = 1, \ldots, r$ to be all nonempty subsets of $[1, T]$. (In other words, the probes intersect as much as possible and the pieces are all of the same size—we have seen that this is necessary in a good protocol.)

We represent each u_i as a real vector of dimension $\lceil l/d \rceil$, where d is the parameter from the statement of the theorem, as follows. We partition u_i, a string of l bits, into $\lceil l/d \rceil$ substrings of d bits. Each coordinate is one of these substrings (in a given order) interpreted as an integer from $[0, 2^d - 1]$. From now on we abuse the notation and by u_i and v_i we mean the vectors described above, interpreted as points in the Euclidean space $\mathbf{R}^{\lceil l/d \rceil}$.

Now we are ready to describe the function $F(u)$. Let $\|x\|$ denote the Euclidean norm of a vector. Let u_0 denote the center of gravity of $u_1, \dots, u_r$, i.e., $u_0 = (\sum_{i=1}^{r} u_i)/r$. The function $F(u)$ consists of the concatenation of u_0 and all the distances $\|u_i - u_j\|$, $0 \le i < j \le r$.

As always, we first examine u and write down $F(u)$ and then for each probe we check if the value of F is equal to $F(u)$. This finishes the description of the protocol.

Let us compute the number of bits of $F(u)$. Instead of communicating u_0, we can communicate the vector ru_0, as r is a scalar constant. Its coordinates are integers from $[0, r(2^d - 1)]$, hence $d + \log r$ bits are sufficient for each coordinate, a total of $(d + \log r)\lceil l/d \rceil \le (1 + T/d)\lceil k/r \rceil + T + d$ bits for all coordinates. Instead of each distance we communicate its square multiplied by r^2. This is a non-negative integer bounded by $r^2(2^{2d} - 1)\lceil l/d \rceil < r^2 2^{2d} l < 2r2^{2d}k$, hence it can be represented by at most $T + 2d + 1 + \log k$ bits, a total of $(T + 2d + 1 + \log k)\binom{r+1}{2}$ for all distances. This gives the bound in the theorem.

To prove the **correctness** of the protocol, we need to prove that if two inputs u and v have the same value of F for u and all probes $\Pi_1, \dots, \Pi_T$, then $u = v$. The intuitive idea is that for a given piece v_i we have sufficient information about its distances from u_j, $j \ne i$, to conclude that $v_i = u_i$.

We need a simple geometric lemma, which we prove later. Recall that a point $x \in \mathbf{R}^n$ is affinely dependent on points $x_1, \dots, x_r \in \mathbf{R}^n$ if it can be written as their linear combination $\sum_{i=1}^{r} \alpha_i x_i$ such that $\sum_{i=1}^{r} \alpha_i = 1$.

Lemma 1. *Let $x, x_1, \dots, x_r$ be points in Euclidean space $\mathbf{R}^n$ such that x is affinely dependent on $x_1, \dots, x_r$. Let $y \in \mathbf{R}^n$ be a point satisfying $\|x - x_i\| = \|y - y_i\|$ for all $i = 1, \dots, r$. Then $x = y$.*

Now we finish the proof of the theorem using this lemma. We can assume that the pieces of input are indexed in such a way that $V_i \supseteq V_j$ implies $i \le j$, i.e., in the reverse topological order with respect to inclusion of the sets V_i. (This means for example that the piece v_1 appears in all probes.)

We prove by induction on $i = 1, \dots, r$ that $u_i = v_i$. Suppose that we are proving the induction step for i, i.e., we have to prove that $u_i = v_i$. We want to use the lemma with $x = u_i$, $\{x_1, \dots, x_r\} = \{u_0, \dots, u_r\} - \{u_i\}$, and $y = v_i$. From the construction we know that x is affinely dependent on the remaining points. We need to prove that $\|u_i - u_j\| = \|v_i - u_j\|$ for all $j = 0, \dots, r$, $j \ne i$. We distinguish three cases.

First, let $j > i$. Take any $t \in V_i - V_j$. By the assumption about the indexing of the pieces we know that such t exists. This means that $(\Pi_t)_i = v_i$ and $(\Pi_t)_j = u_j$ (we use $(\Pi_t)_i$ as a shorthand for $(\Pi_t(u, v))_i$, i.e., the point that represents the ith piece of the vector $\Pi_t(u, v)$). Now using the fact that $F(u) = F(\Pi_t(u, v))$ we get $\|u_i - u_j\| = \|(\Pi_t)_i - (\Pi_t)_j\| = \|v_i - u_j\|$.

The second case is $0 < j < i$. Now take any $t \in V_i$. By induction assumption we already know that $u_j = v_j$, hence $(\Pi_t)_j = u_j$. As in the previous case, we get $\|u_i - u_j\| = \|(\Pi_t)_i - (\Pi_t)_j\| = \|v_i - v_j\| = \|v_i - u_j\|$.

In the last case, $j = 0$, take again any $t \in V_i$. We know that $u_0 = (\Pi_t)_0$ because u_0 is a part of $F(u)$. The rest is again the same.

We have established the assumptions of the lemma, and its application finishes the induction step. We can conclude that $u = v$, hence the theorem holds. □

Proof of Lemma 1. First we prove that the vector $y - x$ is orthogonal to $x_i - x_j$ for every i and j. Using the assumption of the lemma we have

$$\begin{aligned} 0 &= \frac{1}{2}(\|y - x_i\|^2 - \|x - x_i\|^2 + \|x - x_j\|^2 - \|y - x_j\|^2) \\ &= -y^T x_i + x^T x_i - x^T x_j + y^T x_j = (x - y)^T (x_i - x_j). \end{aligned}$$

This means that x is the projection of y on the affine subspace generated by $\{x_1, \ldots, x_r\}$. Using Pythagoras theorem, the projection of a point outside the subspace is always closer to any point in that subspace, hence y has to be identical with x. □

Let us point out that our protocol works for any r, as long as every two pieces are distinguished by some probe. The choice of $r = 2^T - 1$ is done to optimize the bound.

We could have saved some communication in the protocol. Instead of taking the center of $u_1, \ldots, u_r$ it is possible to communicate the shift from u_r to the center of $u_1, \ldots, u_{r-1}$ and then to use this center instead of u_0 for the distances. If we do this, it is not necessary to communicate the distances from u_r at all. It is also not necessary to communicate the distances to the center, as they can be computed from the other distances. But all these savings still leave us with $\Theta(r^2)$ distances to communicate.

4. The Connection with Extremal Problems

In this section we describe how the communication game can be translated into an extremal problem about hypergraphs and how that problem is related to extremal problems about arithmetic progressions.

Suppose that the input is divided into pieces $u_1, \ldots, u_r$ as before, and let $\mathcal{U}_1, \ldots, \mathcal{U}_r$ be the sets of all possible values for each piece. We assume for simplicity that all pieces have the same size, and denote $m = |\mathcal{U}_1| = \cdots = |\mathcal{U}_r| = 2^{k/r}$. Let G be the complete r-partite r-hypergraph on these sets of vertices. This means that each edge is a set of r points, exactly one from each $\mathcal{U}_i$. Each edge of G naturally corresponds to some input vector $(u_1, \ldots, u_r)$.

The function $F(u)$ from the protocol corresponds to a coloring of the hypergraph by 2^B colors. The condition that the protocol is correct means that for no $u = (u_1, \dots, u_r)$, $v = (v_1, \dots, v_r)$, all the edges corresponding to u, $\Pi_1(u,v), \dots, \Pi_T(u,v)$ have the same color. In other words, some specific patterns (or subhypergraphs) are not allowed to be monochromatic. (We get more patterns, because some pieces of u and v may be equal and we get degenerate versions of the original pattern.)

Note that we always need at least m colors, as the edges $(u_1, \dots, u_{T-1}, u_T)$ and $(u_1, \dots, u_{T-1}, u'_T)$ must always have different color, since this is a degenerate version of the prohibited pattern. (This is really just a translation of the trivial lower bound into the new language, because this degenerate pattern just corresponds to the case in which we change just one component of u.) This also means that every hypergraph without a prohibited pattern has at most m^{r-1} edges (out of m^r possible).

To prove a lower bound for the communication game it would be sufficient to prove that any hypergraph with too many edges necessarily contains a prohibited pattern. For the upper bound we not only need to find a large set with no prohibited pattern, but to decompose the complete hypergraph into a small number of such sets.

This kind of problems—to find a maximal size of a structure without a given pattern—is well-studied in extremal combinatorics, so it is not surprising that at least the simplest cases of our problem have been studied. Most of the information about it that we present now is from the survey Graham and Rödl [6] and the paper by Erdős, Frankl and Rödl [5]. These papers also contain simple proofs for some of the results that we mention below.

Let us look at the case $T = 2$ and $r = 3$. Now the prohibited pattern are the 3 edges (u_1, u_2, u_3), (u_1, v_2, v_3) and (v_1, v_2, u_3). The degenerate version of this pattern are any two edges that differ in a single point. We are now interested in the maximal number of edges of a hypergraph that does not contain any of these patterns.

For this case Rusza and Szemerédi [8] proved a slightly better bound than the trivial one $O(m^2)$ mentioned above, namely they proved that the number of edges is $o(m^2)$. This is proved using Szemerédi's regularity lemma [9], and unfortunately does not give a good bound for "o".

This problem is actually related to a problem of finding a large set of numbers which contains no arithmetic progression of length 3, as was noticed first in [8]. Suppose that we have a set $A \subseteq [0, (m-1)/2]$ with no arithmetic progression of length 3. Then we construct a hypergraph without a prohibited pattern by taking $\mathcal{U}_1 = \mathcal{U}_2 = \mathcal{U}_3 = [0, (m-1)]$ and putting in all edges of the form $(u, u+a, u+2a)$ for $a \in A$ and u arbitrary (the addition is taken modulo m), i.e., all arithmetic progressions with one element from each set and modulus from A.

Obviously no degenerate prohibited pattern can appear, because if two arithmetic progressions have two points identical, the third is identical as well. A little checking shows that the non-degenerate prohibited pattern corresponds exactly to the situation where the moduli a, a', and a'' are an arithmetic progression of length 3. So, if we have a large set A, we have a large hypergraph. How large can A be? The best known bounds on the size of such A are

$$\frac{m}{2^{-O(\sqrt{\log m})}} < |A| < \frac{m}{(\log m)^{\Omega(1)}}.$$

The lower bound is a classical result from Behrend [1], the upper bound is due to Heath-Brown [7]. Improving these bounds is considered to be a very hard problem.

To have a small coloring, we need to decompose the interval $[0, m]$ into a small number of such sets, but that turns out to be easy. We also need to color the edges that are not arithmetic progressions, but that is trivial by using a new set of $m/|A|$ colors for edges of the form $(u, u + a, u + 2a + c)$, for every constant c, a total of $m^2/|A|$ colors. Thus the construction based on the largest known set $|A|$ gives us a coloring by $m2^{O(\sqrt{\log m})}$ colors, which corresponds to communicating $k/3 + O(\sqrt{k})$ bits in our game.

Our upper bound for the communication game is based on this construction, translated into the geometric language, so that it can be generalized into a higher dimension. In our construction, the arithmetic progression of three points is replaced by our two points and their geometric center. In particular our protocol gives a construction of large r-hypergraphs without certain prohibited patterns which generalizes the well-known case of 3-hypergraphs with no 3 distinct edges whose union has at most 6 vertices.

5. Connection to Time-Space Tradeoffs

This communication game was proposed by J. Edmonds and R. Impagliazzo [2] as a tool for proving lower bounds in complexity of boolean functions. We shall briefly describe the kind of results that one could possibly obtain without going into details in order to motivate our combinatorial result, for more information about this connection see [3].

A *branching program* is an oriented acyclic graph with one source, two sinks and each vertex, which is not a sink, having outdegree 2; the edges are labelled by variables and negated variables so that for each vertex we have a variable and its negation at the two outgoing edges; the sinks are labelled by *accept* and *reject*. An input vector determines a unique path from the source to a sink, the label at the sink determines, if the vector is accepted or not. The reason for introducing this special kind of a circuit is that the logarithm of the minimal number of vertices of a branching program is a natural measure of *space* (also called capacity) needed for computing a boolean function. This is because we can think of a vertex in the branching program as a configuration of the memory of a computational device. Similarly, the maximal length of a path from the source to a sink corresponds to *time*, however this measure of complexity is interesting only if combined with some restriction on the size of the branching program.

Since proving nontrivial lower bounds to a single measure, such as space, which we have described above, seems to be a very hard task, it is natural to try to prove lower bounds for combined measures. The branching program model of computation is an ideal combinatorial setting for proving a lower bound for the combined measure *time* × *space*. Nevertheless, so far we have only the trivial lower bound $n \log n$ (for an explicitly defined n-variable boolean function).

Edmonds and Impagliazzo showed [3, 4] that if we could prove a lower bound $f(k)$ on the total number of bits of memory plus the total number of flips, we could prove a lower bound of $n\sqrt{f(n)}/\log n$ for time-space product for oblivious branching programs for the function of *element distinctness*.

However, the following simple protocol discovered by Russell Impagliazzo and the authors shows that it is possible to test the equality using only $O(\log k)$ probes and communicating $O(\log k)$ bits about each probe. This protocol can be converted into a protocol of the form used in Sects. 2 and 3 that uses $O((\log k)^2)$ bits of memory.

We think of u and v as 0-1 vectors in real vector space $\mathbf{R}^k$. We compute the Euclidean distance of u and v and check if it is 0. To compute the distance, $u{\cdot}u+v{\cdot}v-2(u{\cdot}v)$, we compute $u{\cdot}u$ and $v{\cdot}v$ each using one probe and $\log k$ bits of communication. Then we compute the product $(\sum_{i=1}^{k} u_i)(\sum_{i=1}^{k} v_i)$ using the same probes and additional $2\log k$ bits of communication. To compute the desired inner product $u \cdot v$ we need to subtract the sum of terms $u_i v_j$ for $i \neq j$. This is easily done using $2\log k$ probes—choose them so that each of the crossterms can be computed by one of them, and for each probe sum all of these terms assigned to it.

This protocol shows that a lower bound for element distinctness cannot be proved using this communication game. A more general game for which the above protocol cannot be used and thus seems as a feasible approach to time-space lower bounds was proposed by Edmonds and Impagliazzo in [4].

Acknowledgements We would like to thank Russell Impagliazzo and Jeff Edmonds for many fruitful discussions. We are grateful to Vojta Rödl for pointing to us the literature on the related extremal problems.

References

1. F.A. Behrend, *On sets of integers which contain no three in arithmetic progression*, Proc. Nat. Acad. Sci. 23 (1946), 331–332.
2. J. Edmonds and R. Impagliazzo *Towards time-space lower bounds on branching programs*, manuscript.
3. J. Edmonds and R. Impagliazzo *About time-space bounds for st-connectivity on branching programs*, manuscript.

4. J. Edmonds and R. Impagliazzo *A more general communication game related to time-space tradeoffs*, manuscript.
5. P. Erdős, P. Frankl, and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph*, Graphs and Combinatorics 2, 113–121 (1986)
6. R.L. Graham and V. Rödl, *Numbers in Ramsey Theory*, In: Surveys in combinatorics (ed. I. Anderson), London Mathematical Society lecture note series 103, pp. 111–153,1985.
7. D.R. Heath-Brown, *Integer sets containing no arithmetic progressions*, preprint, 1986.
8. I.Z. Rusza and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Coli. Math. Soc. Janos Bolyai 18 (1978), pp. 939–945.
9. E. Szemerédi, *Regular partitions of graphs*, In : Proc. Coloq. Int. CNRS, Paris, CNRS, 1976, 399–401.

How Abelian is a Finite Group?

Lásló Pyber*

L. Pyber (✉)
Mathematical Institute of the Hungarian Academy of Sciences, P.O.B. 127,
Budapest, H-1364, Hungary
e-mail: pyber@renyi.hu

Summary. The first paper with the above title was written by Erdős and Straus. Here we solve one of the problems considered there by proving that every group of order n contains an abelian subgroup of order at least $2^{\varepsilon\sqrt{\log n}}$ for some $\varepsilon > 0$. This result is essentially best possible.

We also give a quick survey of recent developments in related areas of group theory which were greatly stimulated by questions of Erdős.

1. Introduction and a Survey

Large Abelian Subgroups

One of the most ancient questions of group theory is: does every infinite group contain an infinite abelian subgroup?

As a byproduct of their work on the Burnside problem P. S. Novikov and S. I. Adian (see [1]) obtained a negative answer to this question. More recently E. Rips [62] and A. Yu Ol'shanskii (see [58]) have independently constructed Tarski Monsters, i.e., infinite groups all of whose proper non-trivial subgroups have order p where p is some large prime.

On the other hand as shown independently by P. Hall and C. R. Kulatilaka [32] and M. I. Kargapolov [40] the answer is positive for locally finite groups. Various results of similar flavour have been established since (see e.g., [74]). Recently A. Mann [48] suggested that the "dual" class of groups should also be considered: does every infinite residually finite group contain an infinite abelian subgroup?

(A group is called residually finite if the intersection of its finite index normal subgroups is trivial.)

The finite analogue of the above question goes back to Jordan (see [46]). G. A. Miller [46] was the first to publish a complete proof of the following well known result: a group of order p^α (p prime) contains an abelian subgroup of order at least $p^{\sqrt{2\alpha}}$.

* Research partially supported by the Hungarian National Science Foundation Grants No. 4267 and No. 1909.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_25,

It was proved by P. Erdős and E. G. Straus [26] that in general a group of order n contains an abelian subgroup of order roughly $\log n$.

Erdős suggested that this estimate should be improved (in fact he put a copy of their paper in my letter-box).

Our main result here is the following.

Theorem 1. *Every group of order n contains an abelian subgroup of order at least $2^{\varepsilon\sqrt{\log n}}$ for some $\varepsilon > 0$.*

(Throughout this paper log denotes logarithm to the base 2.)

The proof uses an easy consequence of the Classification Theorem of finite simple groups.

In the opposite direction Ol'shanskii [57] (see also [14]) has shown the existence of groups of order p^α without abelian subgroups of order $\geq p^{\sqrt{8\alpha}}$.

Note that despite much effort, no explicit construction of such p-groups is known. Indeed Ol'shanskii's proof is one of the first applications of the random method in finite group theory. Ol'shanskii considers certain random Higman groups, i.e., p-groups with a central elementary abelian Frattini subgroup (describing such a group is essentially equivalent to giving two $GF(p)$ vector-spaces V and W together with an alternating bilinear function $V \times V \to W$.)

For some recent applications of this method see [4, 29, 49, 50].

The Commuting Graph

We can associate with an arbitrary group G its commuting graph $\Gamma = \Gamma(G)$: the vertices of Γ are the elements of G and two vertices g, h are joined by an edge if and only if g and h commute as elements of G. The maximal cliques of Γ correspond to maximal abelian subgroups of G, the dominating vertices are exactly the elements of the center $Z(G)$ of G etc.

Answering a question of Erdős, B. H. Neumann [54] (see also [28]) proved that if $\Gamma(G)$ contains no infinite independent set, then there is a finite bound on the cardinality of independent subsets of Γ. Moreover he proved that in this case $|G/Z(G)|$, the index of the center is bounded in terms of $\alpha = \alpha(\Gamma)$, the maximal size of an independent set of Γ.

It was proved in [60] that in fact we have $|G/Z(G)| \leq c^\alpha$ for some (rather large) constant c.

Let us denote by $a(G)$ the minimal number of abelian subgroups covering G (note that $a(G)$ is the chromatic number of the complementary graph $\overline{\Gamma(G)}$). It is obvious that $\alpha(\Gamma(G)) \leq a(G) \leq |G/Z(G)|$.

Confirming a conjecture of Erdős and Straus [26] D. R. Mason [51] proved that for a group of order n we have $a(G) \leq \left[\frac{n}{2}\right] + 1$.

As noted in [26] I. M. Isaacs has shown that $a(G) \leq (\alpha!)^2$. Isaacs also observed that for extra special 2-groups we have $a(G) \geq 2^{\alpha/2}$ (see [6] for details).

Erdős [23] suggested that the upper bound should be improved. Such an improvement is a consequence of the above mentioned result from [60] namely we have $a(G) \le c^{\alpha}$.

In the opposite direction as R. Baer observed (see [53]) if $a(G)$ is finite then $|G/Z(G)|$ is finite. M. J. Tomkinson [69] proved that in fact we have $|G/Z(G)| \le 2a(G)^{\log a(G)}$ and suggested that perhaps $|G/Z(G)| \le (a(G)-1)^2$ is true.

In response to a question of Erdős, V. Faber, R. Laver and R. McKenzie [28] showed that if $a(G) = \kappa$ is an infinite cardinal then $|G/Z(G)|$ is still bounded in terms of $a(G)$. Using a Partition Theorem of Erdős, A. Hajnal and R. Rado [25] and Tomkinson [69] improved this estimate to $|G/Z(G)| \le 2^{2^{\kappa}}$. Tomkinson suggests that perhaps even $|G/Z(G)| \le 2^{\kappa}$ is true (this would be best possible [28]).

Given a class χ of groups we can associate with an arbitrary group G a graph $\Gamma_{\chi}(G)$ as follows: the vertices of Γ_{χ} are the elements of G and two vertices g, h are joined by an edge if and only if $\langle g, h\rangle \in \chi$. Thus if χ is the class of abelian groups then $\Gamma_{\chi}(G)$ is the commuting graph.

J. C. Lennox and J. Wiegold [45] suggested that the questions of Erdős should also be considered for the graphs $\Gamma_{\chi}(G)$ where χ is taken to be the class of finite groups, soluble groups, polycyclic groups etc. For finitely generated groups G they obtained certain analogues of B. H. Neumann's result mentioned above. Further results in this direction were obtained in [15, 30, 39, 71, 76].

As Lennox and Wiegold noted [45] when considering infinite groups it is necessary to restrict attention to finitely generated soluble groups, for otherwise these analogues are no longer true. For example M. R. Vaughan-Lee and Wiegold [73] constructed infinite perfect groups in which every two-generator subgroup is nilpotent of bounded class (see [45] for more details concerning similar examples).

However, such monsters are necessarily infinite. Indeed by a well known corollary of J. G. Thompson's characterisation of minimal simple groups if every 2-generator subgroup of a finite group G is soluble then G itself is soluble.

Furthermore, as Tomkinson [69] proved if a group G has an irredundant covering by n subgroups $H_1, H_2, \dots, H_n$ then $\left|G : \bigcap_{i=1}^{n} H_i\right| \le n!$. This shows for example that if G is a finite group which can be covered by m soluble subgroups then $|G/\operatorname{Sol}(G)|$, the index of the maximal soluble normal subgroup of G is bounded in terms of m.

So it seems to be reasonable to investigate the properties of the graphs $\Gamma_{\chi}(G)$ for arbitrary finite groups.

Conjugacy Classes

A group G is said to be a BFC group if its conjugacy classes are finite and of bounded size, it is called an m-BFC group if the least such upper bound is m.

One of B. H. Neumann's discoveries [52] was that in a BFC group the commutator subgroup G' is finite. Wiegold [75] conjectured that for an m-BFC group we have $|G'| \leq m^{\frac{1}{2}(1+\log m)}$. This conjecture was confirmed for nilpotent groups by Vaughan-Lee [72]. In general P. M. Neumann and Vaughan-Lee [56] obtained the slightly weaker inequality $|G'| \leq m^{\frac{1}{2}(3+5\log m)}$ (see [17] for some improvements). The proof of the exponential bound for the index of the centre [60] rests upon this result via the following observation: in a finite group G every conjugacy class has size at most $4\alpha^2(\Gamma(G))$.

Further results concerning BFC groups are to be found in [36, Chap. VIII/9], [47, 56]. A group which has finite conjugacy classes is called an FC group. For the extensive theory of these groups we refer the reader to [63, 68].

Note that a group G is m-BFC exactly if $\delta = \frac{|G|}{m} - 1$ holds for the minimal degree δ of the commuting graph $\Gamma(G)$. The maximal degree of Γ is clearly equal to $|G| - 1$. For soluble groups E. A. Bertram [7] started investigations concerning the largest order of the centralizer of a non-central element. This number is $\Delta_2 - 1$ where Δ_2 denotes the second largest degree of Γ.

Confirming a conjecture of Bertram Isaacs [38] and J. Cossey [19] independently proved that for G soluble we have $\Delta_2 \geq |G|^{\frac{1}{2}}$. T. Kepka and M. Niemenmaa [41] proved using a classical result of R. Brauer and K. A. Fowler [13] that $\Delta_2 \geq |G|^{\frac{1}{4}}$ for an arbitrary finite group G. While the bound for soluble groups is sharp, the result of Kepka and Niemenmaa can probably be improved.

Let us denote by $A_t(G)$ the number of ordered t-tuples of pairwise commuting elements of G.

Another problem considered by Erdős and Straus [26] is that of obtaining lower bounds for $A_t(G)$. This was originally suggested by Linnik.

As observed earlier by Erdős and Turán [27] we have $A_2(G) = |G|k(G)$ where $k(G)$ is the number of conjugacy classes of G (note that the number of edges in $\Gamma(G)$ is $\frac{1}{2}|G|(k(G) - 1)$). This latter invariant has been investigated repeatedly.

Answering a question of Frobenius, E. Landau [44] proved in 1903 that for a given k there are only finitely many groups having k conjugacy classes. Brauer [12] noted that Landau's argument implies $k(G) \geq \log\log|G|$ and proposed the problem of finding substantially better bounds.

The first such bound was established in [61]: every group G of order $n \geq 4$ satisfies $k(G) \geq \varepsilon \frac{\log n}{(\log\log n)^8}$ for some $\varepsilon > 0$.

On the other hand very recently L. G. Kovács [42] constructed groups of order n with roughly $(\frac{\log n}{\log \log n})^2$ conjugacy classes for certain numbers n (cyclic extensions of the p-groups constructed in [43]).

Groups with a very large number of conjugacy classes were characterised by P. M. Neumann [55].

As the above discussion shows we know a lot about $A_2(G)$. Erdős and Straus [26] proved that $A_t(G)/A_{t-1}(G) \to \infty$ as $|G| \to \infty$. However it is not clear how fast $A_t(G)$ should grow when $t \geq 3$. For some partial results see [26] and the final section of this paper.

For infinite groups G. Higman, B. H. Neumann and Hanna Neumann [34] have shown that every torsion-free group can be embedded in a torsion-free group with just two conjugacy classes.

Furthermore Rips [62] has constructed countably infinite groups with exactly 3 non-conjugate subgroups. In contrast S. Shelah [66] proved that every group of cardinality $\aleph_1$ has at least $\aleph_1$ non-conjugate subgroups.

A few words about the rest of this paper. In Sect. 2 we establish some general results on large subgroups of finite groups. Section 3 contains the proof of our main result concerning abelian subgroups. In the last section we offer some comments on some related problems. Our notation will mainly follow that of [67].

2. Large Subgroups

We use the classification of finite simple groups via the following.

Proposition 1. *Let G be a nonabelian simple group and S a Sylow subgroup of largest order in G. If G is an alternating group then we have $|S|^{2 \log \log |S|} \geq |G|$ and in all other cases we have $|S|^5 \geq |G|$.*

Proof. For the alternating groups, the sporadic groups and for the Tits simple group our statement follows by inspection. For the groups of Lie type we have in fact $|S|^3 \geq |G|$ as noted in [4]. □

It would be most interesting and useful to show the existence of "large" soluble subgroups of simple groups without using the Classification Theorem. It is curious to note that A. Chermak and A. Delgado [18] obtained an elementary proof of a result that goes the other way around: if G is a nonabelian simple group then for every abelian subgroup A we have $|A|^2 \leq |G|$.

Following Wielandt we call a subgroup H *intravariant* in the group G if every image of H under an automorphism of G is conjugate in G to H. For example any Sylow subgroup of G is intravariant.

We are going to use the following very special case of a theorem of S. A. Chunikhin [67, Chap. 5, Theorem 3.17].

Proposition 2. *Let $G = G_0 \rhd G_1 \ldots \rhd G_2 = \mathbb{1}$ be a normal series with each G_i normal in G. Let $\overline{G_i} = G_{i-1}/G_i$ denote the factors. Suppose for each i that $\overline{H_i}$ is an intravariant subgroup of $\overline{G_i}$. Then there is a subgroup H of G such that all the nonabelian composition factors of H occur as composition factors of some of the groups $\overline{H_i}$ and $|H| \geq \prod |\overline{H_i}|$.*

Let us say that a group G is an *alternating type* group if all the nonabelian composition factors of G are alternating of degree at least 8 (we exclude alternating factors of smaller degree for technical reasons).

Corollary 1. *Let G be an arbitrary finite group. Then*

(a) G contains a soluble subgroup S such that

$$|G| \leq |S|^{2 \log\log |S|}.$$

(b) G contains an alternating type subgroup H such that $|G| \leq |H|^5$.

Proof. Let $G_0 \rhd G_1 \ldots \rhd G_r = \mathbb{1}$ be a chief series of G. Then for each i the factor group $\overline{G_i} = G_i/G_{i-1}$ is a direct product of isomorphic simple groups. If G_i is abelian we set $\overline{G_i} = \overline{S_i}$, otherwise let $\bar{S}_i$ denote a Sylow subgroup of largest order in $\overline{G_i}, \bar{S}_i$ is a soluble intravariant subgroup of $\overline{G_i}$ and by Proposition 1 we have $|\overline{G}_i| \leq |\bar{S}_i|^{2 \log\log |\overline{S_i}|}$. Now (a) follows from Proposition 2.

The proof of (b) is analogous. □

By a result of Dixon [20] a solvable subgroup of the symmetric group Sym(n) has order at most $24^{\frac{n-1}{3}}$. This shows that the "log log" factor in Corollary 1 (a) cannot be dispensed with.

A folklore consequence [33, 48] of a well known result of P. P. Pálfy [59] and T. R. Wolf [77] is that if G is a soluble group then $|F(G)|^\alpha \geq |G|$ holds for the Fitting subgroup $F(G)$ of G ($F(G)$ is the unique maximal nilpotent normal subgroup) where $\alpha = 2.24399\ldots$. Moreover H. Heineken [33] proved that a soluble group G has a nilpotent subgroup H such that $|H|^3 \geq |G|$.

By Miller's theorem a nilpotent group of order $n = \prod_{i=1}^{t} p_i^{\alpha_i}$ contains an abelian subgroup of order at least $\prod_{i=1}^{t} p_i^{\sqrt{2\alpha_i}}$ (it is not clear whether this holds for say soluble groups).

When combined the above results immediately yield that an arbitrary group of order n contains an abelian subgroup of order at least $2^{\varepsilon \sqrt{\frac{\log n}{\log\log n}}}$. To erase the "log log" factor here we have to consider abelian subgroups of alternating type groups in more detail.

Before that we would like to point out another application of the above simple ideas (and the Classification Theorem).

As Jordan proved there is a function $J(n)$ such that whenever G is a finite subgroup of $GL(n, \mathbb{C})$ then G has an abelian normal subgroup A with

$|G : A| \leq J(n)$. The best upper bound obtained for $J(n)$ by elementary means is roughly $2^{\frac{n^2}{\log n}}$ (see [37, 64]).

As the symmetric group $\mathrm{Sym}(n+1)$ has a faithful linear representation of degree n we know that $J(n) \geq (n+1)!$.

Here we observe the following.

Proposition 3. *$J(n) \leq (n!)^{20}$ for n sufficiently large.*

Proof. Let G be a linear group of degree n. By [37, Theorem 14.16] there is an abelian normal subgroup A of G such that $|H : H \cap A| \leq 12^n$ holds for every abelian subgroup H of G.

Let S be a solvable subgroup of maximal order in G (clearly $S \supseteq A$). By a result of L. Dornhoff (see [22]) a finite soluble subgroup of $GL(n, \mathbb{C})$ contains an abelian normal subgroup H of index less than c^n for some $c < 9$. Now $|S/A| \leq |S : H||H : H \cap A| \leq 100^n$. Using Corollary 1 (a) we obtain that $|G : A| \leq (100^n)^{2 \log\log(100^n)} \leq (n!)^{20}$ for n sufficiently large. □

Close to sharp estimates for $J(n)$ appear in unpublished work of B. Weisfeiler (personal communication from G. R. Robinson).

3. The Proof of the Theorem

Let G be a group of alternating type of order n. Clearly all alternating factors of G have degree less than $\ell = \lceil \log n \rceil$. We define the numbers ℓ_i by $\ell_i = \lceil \log^{\overset{i}{\smile}} n \rceil$ (where $\log^{\overset{i}{\smile}}$ denotes the i-th iterated logarithm of n) for $i = 1, \ldots t$ where t is the largest natural number such that $\log^{\overset{i}{\smile}} n > 6$ and set $\ell_{t+1} = 6$ and $\ell_{t+2} = 4$. Note that $2^{\ell_{i+1}} \geq \ell_i$ holds for $i = 1, \ldots, t+1$.

We will need the following corollary of results due to Bercov [5].

Lemma 1. *Suppose that a minimal normal subgroup of a finite group G is a direct product of (isomorphic) alternating groups of degree at least 7. Then N has a complement C in G (i.e., we have $NC = G$ and $N \cap C = \mathbb{1}$).*

Lemma 2. *Let G be a group of alternating type. There exists a series of subgroups $G_0, G_1, \ldots, G_t$ of G such that*

(a) G_0 is soluble and its order is the product of the orders of all abelian composition factors of G.
(b) For $i = 1, \ldots, t$ the nonabelian composition factors of G_i are alternating groups. The number of all such factors of degree r is the same as in G for $\ell_i > r \geq \ell_{i+1}$ and otherwise it is zero.
(c) $|\mathrm{Sol}(G_i)|$ divides $|G_0|$.
(d) $\mathrm{Sol}(G_i) = F(G_i)$ for every i.

Proof. The existence of a series of subgroups satisfying (a), (b) and (c), follows from Lemma 1 by an obvious induction argument.

Suppose now that $G_0, G_1, \ldots, G_t$ is a series of subgroups of G satisfying (a), (b) and (c) such that the orders of the subgroups $G_1, \ldots, G_t$ are minimal. We claim that this series satisfies (d) as well.

By a well known observation [35, p. 269] if G is a group, N a normal subgroup of G such that no proper subgroup of G has G/N as a factor group then N is contained in the Frattini subgroup $\Phi(G)$ of G. Applying this to G_i and $\mathrm{Sol}(G_i)$ we see that $\mathrm{Sol}(G_i) \subseteq \Phi(G_i) \subseteq F(G_i)$ by the minimality of G_i for every i. □

We also need the following useful result [3].

Lemma 3. *Let G be a permutation group of degree d. If G has no alternating composition factors of degree $> D$ $(D \geq 6)$ then $|G| \leq D^{t-1}$.*

Next we prove a crucial technical lemma.

Lemma 4. *Let the group H be a direct product of the alternating groups $Alt(d_i)$ $(i = 1, \ldots, k)$ with $r \leq d_i < 2^r$ for some $r \geq 8$. Let A be a subgroup of $\mathrm{Aut}(H)$ such that A has no alternating composition factors of degree $\geq 2^r$. Then*

(a) $|A| \leq |H|^2$.
(b) For any $p \leq r$ (p prime) H contains an elementary abelian p-subqroup of order at least

$$|H|^{\frac{\log p}{2rp}}.$$

Proof. A permutes the k factors of H. The kernel of this action K is a subgroup of $\prod_{i=1}^{k} \mathrm{Sym}(d_i)$ and therefore we have $|K| \leq |H| 2^t$ (here we used the fact that $\mathrm{Aut}(\mathrm{Alt}(d)) = \mathrm{Sym}(d)$ for $d \geq 7$). The factor group A/K is a subgroup of $\mathrm{Sym}(k)$ and by Lemma 3 its order is less than $(2^r)^t$. On the other hand we have $|H| \geq (\frac{r!}{2})^t \geq (2r+1)^t$ and (a) follows.

Clearly it is sufficient to prove (b) for the case when $H = \mathrm{Alt}(d)$ is an alternating group (i.e., $t = 1$). In this case H contains an elementary abelian p-subgroup of order $p^{\lfloor \frac{d}{p} \rfloor}$.

If $d \geq 2p$ then $2p\lfloor \frac{d}{p} \rfloor \geq d$ and $r \geq \log d$ imply $2rp\lfloor \frac{d}{p} \rfloor \geq d \log d$. Therefore $p^{\lfloor \frac{d}{p} \rfloor} \geq (d^d)^{\frac{\log p}{2r\,p}} > |\,\mathrm{Alt}(d)|^{\frac{\log p}{2r\,p}}$.

If $d < 2p$ then we have $p^{\frac{2r\,p}{\log p}} = (2^r)^{2p} > |\,\mathrm{Alt}(d)|$ as required. □

Next we will list some basic properties of the generalized Fitting subgroup $F^*(G)$ (see [2, Chap. 11]).

Let us recall that a group H is called *quasisimple* if it is a perfect central extension of a simple group L, i.e, if $H = H'$ and $L \cong H/Z(H)$. It is well known that if L is an alternating group of degree at least 8 then $|Z(H)|$ divides 2 [67, I Chap. 3, Theorem 2.22].

The components of a group G are its subnormal quasisimple subgroups. The subgroup $E(G)$ is generated by the components of G and in fact it is a central product of the components.

The generalised Fitting subgroup is $F^*(G) = E(G) \cdot F(G)$ (this again turns out to be a central product). The most significant fact about $F^*(G)$ is that $C_G(F^*(G)) = Z(F^*(G))$. This means that $G/Z(F^*(G))$ acts faithfully on $F^*(G)$ by conjugation.

We also need two easy observations about $\mathrm{Sol}(G)$: if C is a normal subgroup of G then $\mathrm{Sol}(C) = \mathrm{Sol}(G) \cap C$ and if $C \subseteq \mathrm{Sol}(G)$ then $\mathrm{Sol}(G/C) = \mathrm{Sol}(G)/C$.

Let $\mathrm{alt}(G)$ denote the product of the orders of all (nonabelian) alternating composition factors of G.

The following lemma contains a major part of the proof of our Theorem.

Lemma 5 (Main). *Let us fix an index $i, 1 \leq i \leq t$, and consider the group G_i (as in Lemma 2) and an odd prime $p \leq \ell_{i+1}$. Suppose that for $C = C_{G_i}(O_p(G_i))$ we have* $\mathrm{alt}(C) \geq x$. *Then G contains an abelian p subgroup of order at least $x^{\frac{\log p}{4p\ell_{i+1}}}$.*

Proof. The centraliser C is a normal subgroup of G_i therefore $\mathrm{Sol}(C) = \mathrm{Sol}(G_i) \cap C$. By Lemma 2 (d) we have $\mathrm{Sol}(G_i) = F(G_i)$ therefore $\mathrm{Sol}(C)$ is the centraliser of $O_p(G_i)$ in $F(G_i) = O_p(G_i) \times O_{p'}(G_i)$, i.e., it is $Z(O_p(G_i)) \times O_{p'}(G_i)$.

Consider the factor group $\tilde{C} = C/O_{p'}(G_i)$. If $\tilde{C}$ has an abelian p-subgroup P then clearly C has an abelian p-subgroup isomorphic to P. Therefore it is sufficient to prove that $\tilde{C}$ has a large abelian p-subgroup.

Now $\mathrm{Sol}(\tilde{C}) = \mathrm{Sol}(C)/O_{p'}(G_i) \cong Z(O_p(G_i))$. It follows that $\mathrm{Sol}(\tilde{C}) = F(\tilde{C})$ and in fact $F(\tilde{C}) = Z(\tilde{C})$ (for by definition $Z(O_p(G_i))$ is contained in the center of C).

Consider $F^*(\tilde{C}) = E(\tilde{C}) \cdot F(\tilde{C})$. As the nonabelian composition factors of $\tilde{C}$ are all alternating groups of degree at least 8 the center of $E(\tilde{C})$ is a 2-group. On the other hand $Z(E(\tilde{C}))$ is contained in the p-group $\mathrm{Sol}(\tilde{C})$, i.e, $Z(E(\tilde{C})) = \mathbb{1}$ and $E(\tilde{C})$ is a direct product of alternating groups.

$\tilde{C}$ acts by conjugation on $F^*(\tilde{C})$ and as noted above the kernel of this action is $Z(F^*(\tilde{C}))$. As we have $F(\tilde{C}) = Z(\tilde{C})$ it follows that $\tilde{C}/Z(F^*(\tilde{C}))$ acts faithfully on $E(\tilde{C})$. It is also clear that $|\tilde{C}/Z(F^*(\tilde{C}))| \geq \mathrm{alt}(C) \geq x$.

Using Lemma 4 (a) we obtain that $|E(\tilde{C})| \geq \sqrt{x}$ and by Lemma 4 (b) $E(\tilde{C})$ contains an abelian p-subgroup of order at least $x^{\frac{\log p}{4p\ell_{i+1}}}$. The statement of the lemma follows. □

We need two more folklore observations.

Proposition 4. *Let P be a p-group, p an odd prime. Suppose that* $\mathrm{Aut}(P)$ *has an elementary abelian section (factor group of a subgroup) of order 2^α. Then $|P| \geq p^\alpha$.*

Proof. Let S denote the group of automorphisms of P which stabilise $P/\Phi(P)$. It is well known that S is a p-group, therefore $\mathrm{Aut}(P)/S$ has an elementary abelian section of order 2^{α}. However, $\mathrm{Aut}(P)/S$ has a natural embedding into $GL(r,p)$ where $|P/\Phi(P)| = p^r$. From the structural description of the Sylow 2-subgroups of $GL(r,p)$ [16] the inequality $\alpha \le r$ follows immediately. □

Proposition 5. *There exists an absolute constant c such that for any n and $1 \le i \le t$ the product of the primes p, $\ell_{i+1} \ge p > \ell_{i+2}$, is at least $2^{c\ell_{i+1}}$.*

Proof. This follows, e.g., from the estimates in [65] using $2^{\ell_{i+2}} \ge \ell_{i+1}$. □

Proof of Theorem 1. Let G be a group of alternating type. Take a subgroup G_i such that $x_i = \mathrm{alt}(G_i) \ge 2^{8\lceil \log\log n\rceil^2 \sqrt{\log n}}$.

Suppose first that for some prime p, $\ell_{i+1} \ge p > \ell_{i+2}$, we have $\mathrm{alt}(C_{G_i}(O_p(G_i)) \ge \sqrt{x}_i$. By Lemma 5 G_i and therefore G contains an abelian subgroup of order at least

$$x_i^{\frac{1}{8p\ell_{i+1}}} \ge 2^{\sqrt{\log n}}.$$

On the other hand if $\mathrm{alt}(C_{G_i}(O_p(G_i))) \le \sqrt{x_i}$ then for $A = G_i/C_{G_i}(O_p(G_i)))$ we have $\mathrm{alt}(A) \ge \sqrt{x_i}$ and $A \subseteq \mathrm{Aut}(O_p(G_i))$. Consider the factor group $\tilde{A} = A/\mathrm{Sol}(A)$ and $H = \mathrm{Soc}(\tilde{A})$. It is well known that $\mathrm{Soc}(\tilde{A})$ is a product of nonabelian simple groups and that $\tilde{A}$ has an embedding into $\mathrm{Aut}(H)$. Now it is clear that $|\tilde{A}| \ge \sqrt{x_i}$ and that $\tilde{A}$ and H satisfy the conditions of Lemma 4 with $r = \ell_{i+1}$. Therefore A has a section which is an elementary abelian 2-group of order at least $x_i^{\frac{1}{8\ell_{i+1}}}$. By Proposition 4 it follows that

$$|O_p(G_i)| \ge p^{\frac{\log x_i}{8\ell_{i+1}}}.$$

If this latter inequality holds for all primes p, $\ell_{i+1} \ge p > \ell_{i+2}$, then using Proposition 5 we obtain that

$$\prod_{\ell_{i+1} \ge p > \ell_{i+1}} |O_p(G_i)| \ge (2^{c\ell_{i+1}})^{\frac{\log x_i}{8\ell_{i+1}}} \ge x_i^{\frac{c}{8}}.$$

Suppose now that G contains no abelian subgroup of order greater than $2^{\sqrt{\log n}}$. It is clear that t is less than, say, $10 \log\log n$. Therefore the product of the x_i for which $x_i \le 2^{8\lceil \log\log n\rceil^2 \sqrt{\log n}}$ is less than $2^{80\lceil \log\log n\rceil^2 \sqrt{\log n}}$.

Using Lemma 2 (b) and (c) we obtain that $|G_0| \ge \left(\frac{\mathrm{alt}(G)}{2^{8\lceil \log\log n\rceil^3 \sqrt{\log n}}}\right)^{\frac{c}{8}}$. By Lemma 2 (a) this implies that for n sufficiently large $|G_0|^{2+\frac{8}{c}} \ge \mathrm{alt}(G)\ |G_0| = |G|$.

By the result of Heineken [33] mentioned in Sect. 2 this implies that G contains a nilpotent subgroup of order at least $|G|^{\frac{1}{3(2+\frac{8}{c})}}$.

Using Miller's theorem we obtain that G contains an abelian subgroup of order at least $2^{\varepsilon\sqrt{\log n}}$ for some $\varepsilon > 0$.

This completes the proof for alternating type groups G and by Lemma 2 (a) it follows that the Theorem holds for arbitrary finite groups. □

4. Odds and Ends

By a classical result of P. Hall [31] a p-group P of order p^α has derived length less than $\log\alpha$. This implies that P has an abelian section of order at least $|P|^{\frac{1}{\log\alpha}}$. Combining this observation with the results of Sect. 2 we obtain that an arbitrary group of order n has an abelian section of order at least $n^{\frac{c}{(\log\log n)^2}}$ for some $c > 0$. Perhaps much more is true.

Problem

Does there exist an $\varepsilon > 0$ such that every soluble group G has a class 2 subgroup of order at least $|G|^\varepsilon$?

A group G is said to have property P_m if for each m element subset $\{x_1, \ldots, x_m\}$ of G there exists a permutation $\pi \in \mathrm{Sym}(m)$ such that $x_1 \cdot \ldots \cdot x_m = x_{\pi(1)} \cdot \ldots \cdot x_{\pi(m)}$. Let $\mathrm{per}(G)$ denote the smallest m such that G has property P_m.

Much is known about groups with $\mathrm{per}(G)$ small [9–11]. For example R. Brandl [11] proved that $|G/F(G)|$ is bounded in terms of $\mathrm{per}(G)$.

In the opposite direction it follows from some crude estimates (see, e.g., [9]) that if H is a subgroup of G then $\mathrm{per}(G) \leq |G : H|(|H'| + 1)$. Therefore if G has an abelian section of order x then $\mathrm{per}(G)$ is less than, say, $\frac{2(G)}{x}$. For a group G of order n this gives us $\mathrm{per}(G) \leq n^{1-\frac{C}{(\log\log n)^2}}$ (this sharpens a result of Brandl [11]).

It should be possible to improve this estimate to say $\mathrm{per}(G) \leq n^{\frac{1}{2}}$ for n sufficiently large. In fact no examples with $\mathrm{per}(G)$ significantly larger than $\log n$ seem to be known (note however that $\mathrm{per}(\mathrm{Sym}(m)) \geq m$ [9]).

Let us make now a few comments concerning the behaviour of the function $A_t(G)$ (recall that $A_t(G)$ is the number of pairwise commuting t-tuples of G). Denote by $A_t(n)$ the minimum of $A_t(G)$ for all groups G of order n. We will indicate how one can give a lower bound for $A_t(n)$ when n is a prime-power.

Proposition 6. *For every integer $t \geq 1$ there exists a constant c_t such that $A_t(p^\alpha) \geq \lfloor c_t\alpha^{t-1}\rfloor p^\alpha$.*

Sketch of the proof. Take a group P of order p^α. Define a series of elements $g_1, \ldots, g_\alpha$ as follows. If we are given $g_1, \ldots g_{i-1}$ then consider the subgroup P_{i-1} they generate (P_{i-1} turns out to be a normal subgroup of P of order p^{i-1}). If $\tilde{g}_i$ is an element of order p in the center of P/P_{i-1} then any element of the corresponding coset of P_i can be taken as g_i.

It is clear from the definition that $g_1, \ldots, g_\alpha$ are pairwise non-conjugate elements and g_i has at most p^i conjugates.

Now it is obvious that $A_t(P) = \sum_{g \in P} A_{t-1}(C_P(g))$. As $|P : C_P(g_i)|$ is the number of conjugates of g_i it follows that $\frac{A_t(P)}{|P|} \geq \sum_{i=1}^{\alpha} \frac{A_{t-1}(C_P(g_i))}{|C_P(g_i)|}$ and that $|C_P(g_i)| \geq p^{\alpha - i}$.

By induction we obtain that $A_t(p^\alpha) \geq p^\alpha \sum_{i=1}^{\alpha} \lfloor c_{t-1}(\alpha - i)^{t-2} \rfloor$ and our statement follows easily. □

It would be most interesting to find an infinite series of p-groups P_i of order p^{α_i}, p fixed $\alpha_i \to \infty$ for which $\frac{A_2(P_i)}{|P_i|} \leq c\alpha_i$ for some constant c. (Note again that $A_2(P_i)/|P_i|$ is roughly the same as $k(P_i)$.)

It is not clear how fast $A_t(n)$ should grow in general for $t \geq 3$. However, we observe that $(\frac{A_t(n)}{n})^{\frac{1}{t-1}}$ can at least be much smaller than n.

Proposition 7. *$A_t(\mathrm{Sym}(m)) \leq cm!3^{\frac{m\,t}{3}}$ for some $c > 0$.*

Proof. As Dixon [21] observed the number of maximal abelian subgroups of $\mathrm{Sym}(m)$ is less than $m!$ and the maximal order of an abelian subgroup is roughly $3^{\frac{m}{3}}$. Our statement follows. □

As noted in Sect. 1 if a group G contains at most n pairwise non-commuting elements then $|G/Z(G)| \leq c^n$ for some large c and therefore G can be covered by c^n abelian subgroups. Erdős [24] suggested that the value of c in [60] should be further improved.

We would like to point out that such an improvement (and an essential simplification of the proof in [60]) could be obtained by solving a problem of Tomkinson [69] in the special case of abelian p-groups. Tomkinson suggests that if a group G has an irredundant covering by n subgroups $H_1, \ldots, H_n$ then $|G : (H_1 \cap \ldots \cap H_n)| \leq c_0^n$ should hold (and that c_0 should be small). Let us see how the proof of this conjecture for abelian groups G would help us.

It is proved in [69] that if $\{x_1, \ldots, x_n\}$ is a set of pairwise non-commuting elements of a group G having maximal size then the centralisers $C_G(x_1), \ldots, C_G(x_n)$ form an irredundant covering of G and their intersection coincides with $Z(G)$. Suppose now that P is a class 2 p-group (i.e., $P' \leq Z(P)$) and that Tomkinson's conjecture holds for the abelian group $G/Z(G)$. The subgroups $C_P(x_i)/Z(P)$ form an irredundant covering of $P/Z(P)$ with intersection $\mathbb{1}$, i.e., we would have $|P/Z(P)| \leq c_0^n$.

By some preliminary results in [60] if $|P/Z(P)| \leq c_0^n$ holds for class 2 p-groups P then for an arbitrary group G we have $|G/Z(G)| \leq c_0^n 2^{100(\log n)^4}$.

For a better understanding of the above problems it would also be useful to consider their analogues for Lie rings.

References

1. S. I. Adian, The Burnside Problem and Identities in Groups, Ergebnisse der Math. vol. 95 Springer, Berlin (1979).
2. M. Aschbacher, Finite Group Theory, Univ. Press, Cambridge (1986).
3. L. Babai, P. J. Cameron and P. P. Pálfy, On the orders of primitive groups with restricted nonabelian composition factors, J. Algebra 79 (1982), 161–168.
4. L. Babai, A. J. Goodman and L. Pyber, On faithful permutation representations of small degree, Comm. in Algebra 21 (1993), 1587–1602.
5. R. Bercov, On groups without abelian composition factors, J. Algebra 5 (1967), 106–109.
6. E. A. Bertram, Some applications of graph theory to finite groups, Discrete Math. 44 (1983), 31–43.
7. E. A. Bertram, Large centralizers in finite solvable groups, Israel J. Math. 47 (1984), 335–344.
8. E. A. Bertram, Lower bounds for the number of conjugacy classes in finite solvable groups, Isr. J. Math. 75 (1991), 243–255.
9. R. D. Blyth and D. J. S. Robinson, Recent progress on rewritability in groups, in Group Theory, Proc. Singapore Group Theory Conference 1987 (eds. K. N. Cheng and Y. K. Leong) Walter de Gruyter Berlin, New York (1988), 77–85.
10. R. D. Blyth and D. J. S. Robinson, Insoluble groups with P_8, J. Pure Appl. Algebra 72 (1991), 251–263.
11. R. Brandl, General bounds for permutability in finite groups, Arch. Math. 53 (1989), 245–249.
12. R. Brauer, Representations of finite groups, in Lectures in modern mathematics, Vol 1. (ed. T. L. Saaty) John Wiley and Sons, New York (1963).
13. R Brauer and K. A. Fowler, On groups of even order, Ann of Math. (2) 62 (1955), 565–583.
14. J. Buhler, R. Gupta and J. Harris, Isotropic subspaces for skewforms and maximal abelian subgroups of p-groups, J. Algebra 108 (1987), 269–279.
15. M. A. Brodie and L. C. Kappe, Finite coverings by subgroups with a given property, Glasgow Math. J. 35 (1993), 179–188.
16. R. Carter and P. Fong, The Sylow 2-subgroups of the finite classical groups, J. Algebra 1 (1964), 139–151.
17. M. Cartwright, The order of the derived group of a BFC-group: J. London Math. Soc. (2) 30 (1984), 227–243.
18. A. Chermak and A. Delgado, A measuring argument for finite groups, Proc. Amer. Math. Soc. 107 (1989), 907–914.
19. J. Cossey, Finite soluble groups have large centralisers, Bull. Aust. Math. Soc. 35 (1987), 291–298.
20. J. D. Dixon, The Fitting subgroup of a linear solvable group, J. Austral. Math. Soc. 7 (1967), 417–424.
21. J. D. Dixon, Maximal abelian subgroups of the symmetric groups, Can. J. Math. XXIII (1971),426–438.
22. L. Dornhoff, Group representation theory, Part A, Dekker, New York (1972).
23. P. Erdős, On some problems in graph theory, combinatorial analysis and combinatorial number theory, in Graph theory and Combinatorics, Acad. Press, London (1984), 1–17.
24. P. Erdős, Some of my favourite unsolved problems, in A Tribute to Paul Erdős (eds. A. Baker, B. Bollobás and A. Hajnal), Cambridge Univ, Press (1990), 467–478.

25. P. Erdős, A. Hajnal and R. Rado, Partition relations for cardinal numbers, Acta Math. Acad. Sci. Hungar. 16 (1965), 93–196.
26. P. Erdős and E. G. Straus, How abelian is a finite group?, Linear and Multilinear Algebra 3, Gordon and Breach (1976),307–312.
27. P. Erdős and P. Turán, On some problems of statistical group-theory, IV, Acta Math. Hungar. 19 (1968), 413–435.
28. V. Faber, R. Laver and R. McKenzie, Coverings of groups by abelian subgroups, Canad. J. Math. 30 (1978), 933–945.
29. A. J. Goodman, The edge-orbit conjecture of Babai, JCT (B) 57 (1993), 26–35.
30. J. R. J. Groves, A conjecture of Lennox and Wiegold concerning supersoluble groups, J. Austral. Math. Soc. (A) 35 (1983), 218–220.
31. P. Hall, A contribution to the theory of groups of prime power order, Proc. London Math. Soc. (2) 36 (1933), 29–95.
32. P. Hall and C. R. Kulatilaka, A property of locally finite groups, Proc. London Math. Soc. (3) 16 (1966), 1–39.
33. H. Heineken, Nilpotent subgroups of finite soluble groups, Arch. Math. 56 (1991), 417–423.
34. G. Higman, B. H. Neumann and Hanna Neumann, Embedding theorems for groups, J. London Math. Soc. 24 (1949), 247–254.
35. B. Huppert, Endliche Gruppen I, Springer, Berlin, 1967.
36. B. Huppert and N. Blackburn, Finite Groups II, Springer, Berlin, Heidelberg, New York (1981).
37. I. M. Isaacs, Character theory of finite groups, Acad. Press, New York (1976).
38. I. M. Isaacs, Solvable groups contain large centralizers, Israel J. Math. 55 (1986), 58–64.
39. L-C. Kappe, Finite coverings by 2-Engel groups, Bull. Austral. Math. Soc. 38 (1988), 141–150.
40. M. I. Kargapolov, On a problem of O. J. Schmidt, Sibirsk. Math, Z. 4 (1963), 232–235.
41. T. Kepka and M. Niemenmaa, On conjugacy classes in finite loops, Bull. Austral. Math. Soc. 38 (1988), 171–176.
42. L. G. Kovács, unpublished.
43. L. G. Kovács and C. R. Leedham-Green, Some normally monomial p-groups of maximal class and large derived length, Quart. J. Math. Oxford (2) 37 (1986), 49–54.
44. E. Landau, Über die Klassenzahl der binären quadratischen Formen von negativer Diskriminant, Math. Ann. 56 (1903), 260–270.
45. J. C. Lennox and J. Wiegold, Extension of a problem of Paul Erdős on groups, J. Austral. Math. Soc. (A) 31 (1981), 459–463.
46. G. A. Miller, On an important theorem with respect to the operation groups of order p^α, p being any prime number, Messenger of Math. 27 (1898), 119–121.
47. I. D. Macdonald, Some explicit bounds in groups, Proc. London Math. Soc. (3) 11 (1969), 23–56.
48. A. Mann, Some applications of powerful p-groups, Proc. Groups St. Andrews 1989, Cambridge (1991), 370–385.
49. G. Zh. Mantashyan, The number of generators of finite p-groups and nilpotent groups without torsion and dimensions of associative rings and Lie algebras (in Russian) Matematika 6 (1988), 178–186, Zbl. Math. 744.20034.
50. U. Martin, Almost all p-groups have automorphism group, a p-group, Bull. Amer. Math. Soc. 15 (1986), 78–82.
51. D. R. Mason, On coverings of groups by abelian subgroups, Math. Proc. Cambridge Phil. Soc. 83 (1978), 205–209.
52. B. H. Neumann, Groups covered by permutable subsets, J. London Math. Soc. 29 (1954), 236–248.

53. B. H. Neumann, Groups covered by finitely many cosets, Publ. Math. Debrecen 3 (1954), 227–242.
54. B. H. Neumann, A problem of Paul Erdős on groups, J. Austral. Math. Soc. 21 (1976), 467–472.
55. P. M. Neumann, Two combinatorial problems in group theory, Bull. London Math. Soc. 21 (1989), 456–458.
56. P. M. Neumann and M. R. Vaughan-Lee, An essay on BFC groups, Proc. London Math. Soc. (3) 35 (1977), 213–237.
57. A. Yu. Ol'shanskii, The number of generators and orders of abelian subgroups of finite p-groups, Math. Notes 23 (1978), 183–185.
58. A. Yu. Ol'shanskii, Geometry of defining relations in groups, Kluwer, Dordrecht (1991).
59. P. P. Pálfy, A polynomial bound for the orders of primitive solvable groups, J. Algebra 77 (1982), 127–137.
60. L. Pyber, The number of pairwise non-commuting elements and the index of the center in a finite group, J. London Math. Soc. (2) 35 (1987), 287–295.
61. L. Pyber, Finite groups have many conjugacy classes, J. London Math. Soc. (2) 46 (1992), 239–249.
62. E. Rips, Generalized small cancellation theory and applications II (unpublished).
63. D. J. S. Robinson, Finiteness, Solubility and Nilpotence, in Group Theory essays for Philip Hall (eds. K. W. Gruenberg and J. E. Roseblade) Acad. Press, London (1984), 159–206.
64. G. R. Robinson, On linear groups, J. Algebra 131 (1990), 527–534.
65. J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, Illinois J. Math. 6 (1962), 64–94.
66. S. Shelah, On the number of non-conjugate subgroups, Algebra Universalis 16 (1983), 131–146.
67. M. Suzuki, Group Theory I, II, Springer, New York, 1986.
68. M. J. Tomkinson, FC-groups, Research notes in mathematics 96, Pitman, London (1984).
69. M. J. Tomkinson, Groups covered by abelian subgroups, Proc. Groups St. Andrews 1985, Cambridge (1986), 332–334.
70. M. J. Tomkinson, Groups covered by finitely many cosets or subgroups, Comm. in Algebra 15 (1987), 845–859.
71. M. J. Tomkinson, Hypercentre-by-finite groups, Publ. Math. Debrecen 40 (1992), 313–321.
72. M. R. Vaughan-Lee, Breadth and commutator subgroups of p-groups, J. Algebra 32 (1974), 278–285.
73. M. R. Vaughan-Lee and J. Wiegold, Countable locally nilpotent groups of finite exponent with no maximal subgroups, Bull. London Math. Soc. 13 (1981), 45–46.
74. E. I. Zelmanov, On the Restricted Burnside Problem, in Proc. Int. Congress of Math. Kyoto, Japan 1990, Springer, Tokyo (1991), 1479–1489.
75. J. Wiegold, Groups with boundedly finite classes of conjugate elements, Proc. Roy. Soc. London (A) 238 (1956), 389–401.
76. J. S. Wilson, Two-generator conditions for residually finite groups, Bull. London Math. Soc. 23 (1991), 239–248.
77. T. R. Wolf, Solvable and nilpotent subgroups of $GL(n, q^m)$, Canad. J. Math. 34 (1982), 1097–1111.

On Small Size Approximation Models

Alexander A. Razborov

A.A. Razborov (✉)
Steklov Mathematical Institute, Vavilova 42, 117966, GSP–1, Moscow, Russia
e-mail: razborov@cs.uchicago.edu

Summary. In this paper we continue the study of the method of approximations in Boolean complexity. We introduce a framework which naturally generalizes previously known ones. The main result says that in this framework there exist approximation models providing in principle exponential lower bounds for almost all Boolean functions, and such that the number of testing functionals *is only singly exponential in the number of variables.*

1. Introduction

Proving superpolynomial lower bounds on the complexity of explicitly given Boolean functions is one of the most challenging tasks of the modem complexity theory. Its importance stems from the fact that such bounds could be easily translated into similar bounds for Turing models and, thus, would lead to resolving central open questions in Complexity Theory like $P \overset{?}{=} NP$ or $NC \overset{?}{=} P$.

At the moment, however, we have succeeded in proving desired bounds only for rather restrictive models. A substantial part of these bounds was obtained via a general scheme originally proposed in [16, 17] and called afterwards *the method of approximations*:

- On the monotone circuit size—[1, 12, 14–17];
- For bounded-depth circuits with modular gates—[2, 11, 18];
- For switching-and-rectifier networks (= nondeterministic branching programs)—[19];
- For $\oplus$-branching programs (see [7] for definitions)—[6].

The reader willing to learn more about these and related results or about the general perspective of the field is referred to the survey paper [3].

Concrete approximation models which have appeared in the literature can be naturally subdivided into two large groups.

Models from the first group use inputs of the original function as their error tests. Such are models from [1, 2, 11, 12, 14–18]. We will call the method based on models of this kind the *pure approximation method.*

Other models use as error tests specially designed functionals, every functional being attached to a single input. These models were studied,

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_26,

and sometimes actually used in [4–6, 8, 9, 19]. See [13] for an extended survey; following this source, we will call the corresponding method the *fusion method*. The same word "fusion" will be also used for functionals and models.

The most interesting question is, of course, to which extent the approximation method might be useful in proving lower bounds for unrestricted circuits. To that end, it was shown in [9] that the pure approximations can not prove bounds greater than $O(n^2)$ or, more precisely, $O(n \cdot n_0)$, where $n_0 \leq n$ is the number of essential variables of our function. Since every fusion model with N functionals can always be considered as a pure approximation model with $n_0 := n$ and $n := n + \lceil \log_2 N \rceil$ (see [9, Claim 2.5]), it follows that lower bounds provable by any such fusion model never exceed $O(n^2 + n \log N)$.

On the other hand, in [9] for every function f a fusion model was exhibited which, at least in principle, provides tight, up to a polynomial, lower bounds on the circuit size of f. The number of fusing functionals involved in that model was triply exponential in n, and it was also remarked in [9] that it can be decreased to doubly exponential. The resulting model, however, is somewhat artificial.

More natural fusion model with the number of fusing functionals being only doubly exponential in n has been found in [5]. Their model is universal for nondeterministic circuit size, hence it still can prove exponential lower bounds for almost all Boolean functions. Note that. in view of the above-mentioned limitation $O(n^2 + n \log N)$, this is roughly optimal for fusion models.

In this paper we study the question whether there exists a natural version of the method of approximations in which proving exponential lower bounds is possible (again, in principle) with the number of fusion functionals being only singly exponential in the number of variables. We indeed find such a framework generalizing both pure approximations and fusion models. In fact, our framework is obtained by cleaning the underlying idea of approximations from the prejudice of attaching error tests to particular input strings which is characteristic for previous models.

More exactly, we show that for every integer-valued function $t = t(n)$ in the range $n \leq t(n) \leq \frac{2^n}{3n}$ there exists an approximation model $\mathfrak{M}$ (in our framework) with $O(t^3 \log^2 t)$ error tests such that for almost all Boolean functions f, $\rho(f, \mathfrak{M}) \geq t$, where $\rho(f, \mathfrak{M})$ is the distance between f and $\mathfrak{M}$ (Theorem 2).

The main motivation for this work comes from [10], where I put forward the thesis that the right theory capturing the kind of machinery existing in Boolean complexity at the moment is the second-order system V_1^1. This system can freely talk of those approximation models in which the number of error tests is bounded by $2^{O(n)}$ (and thus error tests can be represented by first order objects). Hence, unlike previous models, the models considered in this paper are within the reach of V_1^1 in terms of size. It should be noted, however, that gaining in size we lose in the constructibility. Indeed, our proof heavily relies upon Erdös probabilistic argument, and in order to carry it over in V_1^1 we need an explicit construction.

2. Definition of Approximation Models

Throughout the paper, F_n stands for the set of all Boolean functions in n variables. Let $P_n \rightleftharpoons \{x_1, \dots, x_n, \neg x_1, \dots, \neg x_n\} \subseteq F_n$ be the set of input functions.

Let $\mathcal{F}$ be a finite set of arbitrary nature. We define an *approximation model* $\mathfrak{M}$ as a subset $\mathfrak{M} \subseteq \mathfrak{F}_n \times \mathcal{P}(\mathcal{F})$ such that

$$P_n \times \{\emptyset\} \subseteq \mathfrak{M} \tag{1}$$

supplied with two binary operations $\wedge$, $\vee$ which are consistent with the projection onto F_n. In other words, we require

$$f(m_1 * m_2) = f(m_1) * f(m_2);\ m_1, m_2 \in \mathfrak{M}, \tag{2}$$

where $* \in \{\wedge, \vee\}$, and we once and for all have fixed notation $f(m)$ for denoting the projection of $m \in \mathfrak{M}$ onto the first coordinate F_n. Similarly, we will denote the projection onto $\mathcal{P}(\mathcal{F})$ by $\mathcal{F}(m)$ so that $m = \langle f(m), \mathcal{F}(m) \rangle$.

Now we give a set of definitions which is routine for the method of approximations. Namely, let

$$\delta_*(m_1, m_2) \rightleftharpoons \mathcal{F}(m_1 * m_2) \setminus (\mathcal{F}(m_1) \cup \mathcal{F}(m_2));\ m_1, m_2 \in \mathfrak{M},$$

$$\Delta \rightleftharpoons \{\delta_*(m_1, m_2) \mid * \in \{\wedge, \vee\}; m_1, m_2 \in \mathfrak{M}\},$$

$$\rho(f, \mathfrak{M}) \rightleftharpoons \min\left\{ t \,\middle|\, \exists m \in \mathfrak{M}\ \exists \delta_1, \dots, \delta_t \in \Delta \left(f(m) = f \;\&\; \mathcal{F}(m) \subseteq \bigcup_{i=1}^{t} \delta_i \right) \right\}. \tag{3}$$

The intuitive idea behind this is that if the real circuit computes some function f at a node u, then the approximating circuit must compute at u some $m \in \mathfrak{M}$ with $f(m) = f$ (due to (2)). Now, all tests $F \in \mathcal{F}(m)$ have already found "an error", that is $\mathcal{F}(m) \subseteq \bigcup_v \delta_v$, where the union is extended over all nodes v lying below u, and $\delta_v \in \Delta$ naturally corresponds to the node v. For the reader familiar with previous analogous statements, this should serve as a self-sufficient proof of the following

Theorem 1. *For every $f \in F_n$ and every approximation model $\mathfrak{M}$, we have $\rho(f, \mathfrak{M}) \leq \mathcal{C}(f)$, where $C(f)$ is the minimal possible size of a circuit over the basis $\wedge$, $\vee$ with inputs from P_n computing f.*

We conclude this section by showing that our new framework generalizes both pure approximations and the fusion method.

Example 1. *Let $\langle \mathcal{M}, \bar{\wedge}, \bar{\vee} \rangle$ be a legitimate model [9, Sect. 2]. Here $P_n \subseteq \mathcal{M} \subseteq F_n$ and $\bar{\wedge}, \bar{\vee}$ are arbitrary binary operations on $\mathcal{M}$. Recall that for*

$\bar g, \bar h \in \mathcal{M}$ and $* \in \{\wedge, \vee\}$, the subsets $\delta_*^+(\bar g, \bar h)$, $\delta_*^-(\bar g, \bar h)$ of $\{0,1\}^n$ are defined as follows:

$$\begin{aligned} \delta_*^+(\bar g, \bar h) &\rightleftharpoons (\bar g * \bar h) \setminus (\bar g \mathbin{\bar *} \bar h), \\ \delta_*^-(\bar g, \bar h) &\rightleftharpoons (\bar g \mathbin{\bar *} \bar h) \setminus (\bar g * \bar h) \end{aligned}$$

(we identify a Boolean function with its set of ones). For $f \in F_n$, the distance *$\rho(f, \mathcal{M})$ is the minimal t for which there exist $\bar f, \bar g_1, \ldots, \bar g_t, \bar h_1, \ldots, \bar h_t \in \mathcal{M}$ and $*_1, \ldots, *_t \in \{\wedge, \vee\}$ such that*

$$\begin{aligned} f \setminus \bar f &\subseteq \bigcup_{i=1}^t \delta_{*_i}^+(\bar g_i, \bar h_i), \\ \bar f \setminus f &\subseteq \bigcup_{i=1}^t \delta_{*_i}^-(\bar g_i, \bar h_i). \end{aligned}$$

The quantity $\rho(f, \mathcal{M})$ provides a lower bound on the circuit size of f.

Take now two disjoint copies B_+^n, B_-^n of $\{0,1\}^n$, and let $\mathcal{F} \rightleftharpoons B_+^n \cup B_-^n$. Consider the product $F_n \times \mathcal{M}$ of two $\{\wedge, \vee\}$-algebras, and embed it into $F_n \times \mathcal{P}(\mathcal{F}) \approx F_n \times \mathcal{P}(B_+^n) \times \mathcal{P}(B_-^n)$ as follows:

$$\begin{aligned} \pi : F_n \times \mathcal{M} &\to F_n \times \mathcal{P}(B_+^n) \times \mathcal{P}(B_-^n), \\ \langle f, \bar f \rangle &\mapsto \langle f, f \setminus \bar f, \bar f \setminus f \rangle. \end{aligned}$$

We let $\mathfrak{M} \rightleftharpoons \mathrm{im}(\pi)$ and endow $\mathfrak{M}$ with the structure of $\{\wedge, \vee\}$-algebra induced from $F_n \times \mathcal{M}$. Note that (1) is implied by $P_n \subseteq \mathcal{M}$.

Assume that $m_1, m_2 \in \mathfrak{M}$; $m_1 = \pi(g, \bar g)$, $m_2 = \pi(h, \bar h)$. Representing $\delta_*(m_1, m_2)$ in the form $\delta_*^+(m_1, m_2) \cup \delta_*^-(m_1, m_2)$, where $\delta_*^\circ(m_1, m_2) \subseteq B_\circ^n$, we have:

$$\begin{aligned} \delta_*^+(m_1, m_2) &= ((g * h) \setminus (\bar g \mathbin{\bar *} \bar h)) \setminus (g \setminus \bar g \cup h \setminus \bar h) \\ &= ((g * h) \setminus (g \setminus \bar g \cup h \setminus \bar h)) \setminus (\bar g \mathbin{\bar *} \bar h) \\ &\subseteq (\bar g * \bar h) \setminus (\bar g \mathbin{\bar *} \bar h) = \delta_*^+(\bar g, \bar h) \end{aligned}$$

and similarly for $\delta_*^-(m_1, m_2)$. Noting that the condition

$$\exists m \in \mathfrak{M}\ \exists \delta_1, \ldots, \delta_{\mathrm{t}} \in \Delta \left(f(m) = f \,\&\, \mathcal{F}(m) \subseteq \bigcup_{\mathrm{i}=1}^t \delta_i \right) \tag{4}$$

from (3) can be rewritten in the form

$$\exists \bar f \in \mathcal{M}\ \exists *_i \in \{\wedge, \vee\}\ \exists m_i^{(1)}, m_i^{(2)} \in \mathfrak{M}$$
$$\left(f \setminus \bar f \subseteq \bigcup_{i=1}^t \delta_{*_i}^+ \left(m_i^{(1)}, m_i^{(2)} \right) \& \, \bar f \setminus f \subseteq \bigcup_{i=1}^t \delta_{*_i}^- \left(m_i^{(1)}, m_i^{(2)} \right) \right),$$

we immediately see that $\rho(f, \mathcal{M}) \leq \rho(f, \mathfrak{M})$. In other words, every legitimate model in the sense of [9] can be simulated in our framework.

Example 2. *Let us now turn to the fusion method. In fact, we might first apply the construction from [9] to get a pure approximation model, and then the construction from Example 1. Things become much more transparent, however, if we combine the two steps into one. Recall some necessary definitions [9, 13].*

Let $f \in F_n$ be a fixed *function, $U \rightleftharpoons f^{-1}(0)$, and $V \rightleftharpoons f^{-1}(1)$. Let $\Omega_f \subseteq \{0,1\}^{\mathcal{P}(U)}$ consist of those functionals on $\mathcal{P}(U)$ which satisfy the following two conditions:*

1. *F is monotone,*
2. *There exists a (uniquely determined) $z(F) \in V$ such that for all $x_i^\epsilon \in P_n$,*

$$F(x_i^\epsilon|_U) = x_i^\epsilon(z(F)).$$

Here, as usual, $x_i^1 \rightleftharpoons x_i$ and $x_i^0 \rightleftharpoons (\neg x_i)$.

Note that these two conditions imply $F(\emptyset) = 0$ and $F(U) = 1$.

For $\bar{g}, \bar{h} \in \{0,1\}^U$ we say that the pair $(\bar{g}, \bar{h})$ covers *$F \in \Omega_f$ if $F(\bar{g}) = F(\bar{h}) = 1$ and $F(\bar{g} \wedge \bar{h}) = 0$. The minimal number of pairs needed to cover the whole Ω_f is denoted by $\rho(f)$ and provides a lower bound on the circuit size of f which is tight up to a polynomial.*

Define now the mapping

$$\begin{aligned} \pi : F_n &\to \mathcal{P}(\Omega_f), \\ g &\mapsto \{F \mid g(z(F)) = 1 \,\&\, F(g|_U) = 0\}. \end{aligned}$$

Note that $\pi(g) = \emptyset$ when $g \in P_n$, and $\pi(f) = \Omega_f$.

Let $\mathcal{F} \rightleftharpoons \Omega_f$. We take the diagonal mapping $\theta : F_n \to F_n \times \mathcal{P}(\Omega_f)$; $g \mapsto \langle g, \pi(g)\rangle$, denote $\mathfrak{M} \rightleftharpoons \mathrm{im}(\theta)$ and endow $\mathfrak{M}$ with the induced structure of $\{\wedge, \vee\}$-algebra. Equation (1) is implied by the remark above.

Now, $\delta_\vee(\theta(g), \theta(h)) = \emptyset$ due to the monotonicity of every $F \in \Omega_f$. If $F \in \delta_\wedge(\theta(g), \theta(h))$ then $g(z(F)) = h(z(F)) = 1$ and $F((g \wedge h)|_U) = 0$. Since $F \notin \pi(g)$ and $F \notin \pi(h)$, we have $F(g|_U) = F(h|_U) = 1$. Hence the pair $(g|U, h|U)$ covers F. As (4) in our case simplifies to $\exists \delta_1, \ldots, \delta_t \in \Delta(\mathcal{F} \subseteq \bigcup_{i=1}^t \delta_i)$, we see that $\rho(f) \leq \rho(f, \mathfrak{M})$.

Another version of the fusion method using $GF(2)$-affine functionals instead of monotone functionals was proposed in [5, 6]. It can also be embedded into our framework if we consider approximation models over the basis $\{\wedge, \oplus\}$ rather than over $\{\wedge, \vee\}$.

3. Main Result

In this section we prove the following:

Theorem 2. *Let $t = t(n)$ be an integer-valued function in the range $n \leq t(n) \leq \frac{2^n}{3n}$. Then there exists an approximation model $\mathfrak{M} \subseteq \mathfrak{F}_n \times \mathcal{P}(\mathcal{F})$, where $|\mathcal{F}| \leq O(t^3 \log^2 t)$, such that $\rho(f, M) \geq t(n)$ for almost all junctions $f \in F_n$.*

Proof. Let $\ell \rightleftharpoons \lfloor 20t^3 \ln^2 t \rfloor$ and $S \rightleftharpoons \binom{\ell}{t}$. Fix a set $\mathcal{F}$ of cardinality ℓ.

For a subset $\mathfrak{M}$ of $F_n \times \mathcal{P}(\mathcal{F})$ and $\mathcal{F}_0 \subseteq \mathcal{F}$, we let

$$\mathfrak{M}(\mathcal{F}_o) \rightleftharpoons \{m \in \mathfrak{M} \mid \mathcal{F}(m) \subseteq \mathcal{F}_o\}$$

and

$$w_{\mathfrak{M}}(\mathcal{F}_0) \rightleftharpoons |\mathfrak{M}(\mathcal{F}_o)|.$$

Let also

$$w_{\mathfrak{M}} \rightleftharpoons \ln(\mathbf{E}[e^{w_{\mathfrak{M}}(\mathcal{F}_o)}]),$$

where $\mathcal{F}_o \subseteq \mathcal{F}$ is a random subset of cardinality t.

We are going to define by induction on k a sequence

$$P_n \times \{\emptyset\} = \mathfrak{M}_o \subseteq \mathfrak{M}_1 \subseteq \ldots \subseteq \mathfrak{M}_\ell \subseteq \ldots \subseteq \mathfrak{F}_n \times \mathcal{P}(\mathcal{F}) \tag{5}$$

along with binary operations $\wedge_k, \vee_k : \mathfrak{M}_{\ell-1} \times \mathfrak{M}_{\ell-1} \to \mathfrak{M}_\ell$ maintaining the following properties:

1. If $k \leq k'$ then $\wedge_{k'}|_{\mathfrak{M}_{\ell-1} \times \mathfrak{M}_{\ell-1}} = \wedge_k$ and $\vee_{k'}|_{\mathfrak{M}_{\ell-1} \times \mathfrak{M}_{\ell-1}} = \vee_k$;
2. $f(m_1 *_k m_2) = f(m_1) * f(m_2)$ for $m_1, m_2 \in \mathfrak{M}_{\ell-1}$;
3. For every $m_1, m_2 \in \mathfrak{M}_{\ell-1}$ and $* \in \{\wedge, \vee\}$,
$$|\mathcal{F}(m_1 *_k m_2) \setminus (\mathcal{F}(m_1) \cup \mathcal{F}(m_2))| \leq 1;$$
4. For every $m \in \mathfrak{M}_\ell \setminus \mathfrak{M}_{\ell-1}$, $|\mathcal{F}(m)| \geq \min(k, \ell)$;
5. $w_{\mathfrak{M}_\ell} \leq 2(n+k)$.

Base $k = 0$ is obvious.

Inductive step. Assume that $\mathfrak{M}_o, \mathfrak{M}_1, \ldots, \mathfrak{M}_{\ell-1}, \mathfrak{M}_\ell$ and $\wedge_k, \vee_k : \mathfrak{M}_{\ell-1} \times \mathfrak{M}_{\ell-1} \to \mathfrak{M}_\ell$ are already defined. Then we randomly extend $\wedge_k, \vee_k$ to $\wedge_{k+1}, \vee_{k+1} : \mathfrak{M}_\ell \times \mathfrak{M}_\ell \to \mathfrak{F}_n \times \mathcal{P}(\mathcal{F})$ as follows. For $(m_1, m_2) \in (\mathfrak{M}_\ell \times \mathfrak{M}_\ell) \setminus (\mathfrak{M}_{\ell-1} \times \mathfrak{M}_{\ell-1})$ we let

$$m_1 *_{k+1} m_2 \rightleftharpoons (f(m_1) * f(m_2), \mathcal{F}(m_1) \cup \mathcal{F}(m_2) \cup \{\mathbf{F}_*(m_1, m_2)\}),$$

where $\mathbf{F}_*(m_1, m_2)$ is chosen at random from $\mathcal{F} \setminus (\mathcal{F}(m_1) \cup \mathcal{F}(m_2))$ if $\mathcal{F}(m_1) \cup \mathcal{F}(m_2) \neq \mathcal{F}$ and arbitrarily otherwise. All $\mathbf{F}_*(m_1, m_2)$ are assumed to be independent.

After this we let

$$\mathfrak{M}_{\ell+1} \rightleftharpoons \mathfrak{M}_\ell \cup \mathrm{im}(\wedge_{k+1}) \cup \mathrm{im}(\vee_{k+1}). \tag{6}$$

Properties 1–3 readily follow from definitions, and 4 follows from the inductive assumption. We are going to show that 5 (with $k := k+1$) also takes place with a non-zero probability.

We may assume that $k + 1 \leq t$ since otherwise property 5 follows from 4 and the inductive assumption. For simplicity we will abbreviate $w_{\mathfrak{M}_i}(\mathcal{F}_0)$ and $w_{\mathfrak{M}_i}$ to $w_i(\mathcal{F}_0)$, w_i respectively.

Let us first fix some $\mathcal{F}_0 \subseteq \mathcal{F}$ of cardinality t and estimate $\mathbf{E}[e^{\mathbf{w_{k+1}}(\mathcal{F}_0)}]$ for this particular $\mathcal{F}_0$. Denote the set

$$\{(m_1, m_2, *) \mid (m_1, m_2) \in (\mathfrak{M}_\ell(\mathcal{F}_o) \times \mathfrak{M}_\ell(\mathcal{F}_o)) \setminus (\mathfrak{M}_{\ell-1} \times \mathfrak{M}_{\ell-1}), * \in \{\wedge, \vee\}\}$$

by A. Then $|A| \leq 2w_k^2(\mathcal{F}_0)$ and

$$\mathfrak{M}_{\ell+1}(\mathcal{F}_0) = \mathfrak{M}_\ell(\mathcal{F}_o) \cup \{m_1 *_{k+1} + m_2 \mid (m_1, m_2, *) \in \mathfrak{U} \,\&\, \mathbf{F}_*(m_1, m_2) \in \mathcal{F}_o\}.$$

Hence

$$w_{k+1}(\mathcal{F}_0) \leq w_k(\mathcal{F}_0) + \sum_{(m_1,m_2,*)\in A} \xi_*(m_1, m_2), \tag{7}$$

where $\xi_*(m_1, m_2)$ is the indicator function of the event $\mathbf{F}_*(m_1, m_2) \in \mathcal{F}_0$.

All $\xi_*(m_1, m_2)$ are, however, independent. Therefore (7) gives us the estimate

$$\begin{aligned}\mathbf{E}[e^{w_{k+1}}(\mathcal{F}_0)] &\leq e^{w_k(\mathcal{F}_0)} \cdot \prod_{(m_1,m_2,*)\in A} \mathbf{E}[e^{\xi_a st(m_1,m_2)}] \\ &\leq e^{w_k(\mathcal{F}_0)} \cdot \left(1 + \frac{t(e-1)}{\ell}\right)^{2w_k^2(\mathcal{F})} \\ &\leq e^{w_k(\mathcal{F}_0) + \frac{4t}{\ell} w_k^2(\mathcal{F}_0)}.\end{aligned}$$

Averaging this inequality over $\mathcal{F}_0$, we have

$$\mathbf{E}[e^{w_{k+1}(\mathcal{F}_0)}] \leq \mathbf{E}[e^{w_k(\mathcal{F}_0) + \frac{4t}{\ell} w_k^2(\mathcal{F}_0)}].$$

Now we fix a particular choice of $\mathfrak{M}_{\ell+1}$, with the property

$$e^{w_{k+1}} = \mathbf{E}[e^{w_{k+1}(\mathcal{F}_0)}] \leq \mathbf{E}[e^{w_k(\mathcal{F}_0) + \frac{4t}{\ell} w_k^2(\mathcal{F}_0)}]. \tag{8}$$

We will show that this implies the desired inequality $w_{k+1} \leq 2n + 2k + 2$ if $k + 1 \leq t$.

Let us denote $e^{w_k}(\mathcal{F}_0)$ by $\theta_k(\mathcal{F}_0)$. Then the inductive assumption can be rewritten in the form

$$\mathbf{E}[\theta_k(\mathcal{F}_0)] \leq e^{2(n+k)}, \tag{9}$$

and (8)—in the form

$$e^{w_{k+1}} \leq \mathbf{E}[\theta_k(\mathcal{F}_0) \cdot e^{\frac{4t}{\ell} \ln^2 \theta_k(\mathcal{F}_0)}]. \tag{10}$$

The function $x \cdot a^{\ln^2 x}$, where $a = e^{\frac{4t}{\ell}}$ is, however, convex on $[1, \infty)$. Hence, under the condition (9), the right-hand side of (10) achieves its maximal value when $\theta_k(\mathcal{F}_0)$ takes on $(S-1)$ times the value 1, and the remaining time—the value $S \cdot e^{2(n+k)} - S + 1 \leq S \cdot e^{2(n+k)}$. This gives us the estimate

$$e^{w_{k+1}} \leq \frac{S-1}{S} + e^{2(n+k)} \cdot e^{\frac{4t}{\ell}(\ln S + 2(n+k))^2}.$$

Finally,

$$w_{k+1} \leq \ln(1+e^{2(n+k)} \cdot e^{\frac{4t}{\ell}(\ln S+2(n+k))^2}) \leq 1+2(n+k)+\frac{4t}{\ell}(\ln S+2(n+k))^2$$

and it is easy to see that $\frac{4t}{\ell}(\ln S+2(n+k))^2 \leq \frac{4t}{\ell}(\ln S+2(n+t))^2 \leq 1$ due to our choice of parameters.

When we have the desired sequence (5), the rest is easy. We let $\mathfrak{M} \rightleftharpoons \bigcup_{\ell \geq o} \mathfrak{M}_\ell$. Property 1 ensures that we can glue together the partial operations $\wedge_k, \vee_k$ to endow $\mathfrak{M}$ with a natural structure of $\{\wedge, \vee\}$-algebra. Property 2 gives us (2), and Property 3 lets us to conclude that $\forall \delta \in \Delta$, $|\delta| \leq 1$. Hence, if $\rho(f, \mathfrak{m}) \leq t$ for some $f \in F_n$ then $\exists m \in \mathfrak{M}(f(m) = f \,\&\, |\mathcal{F}(m)| \leq t)$. Due to Property 4, we may replace here $\mathfrak{M}$ by $\mathfrak{M}_t$.

However, the total number of $m \in \mathfrak{M}_t$ that $|\mathcal{F}(m)| \leq t$ does not exceed

$$\sum_{\substack{\mathcal{F}_0 \subseteq \mathcal{F} \\ |\mathcal{F}_0|=t}} w_t(\mathcal{F}_0).$$

Since $e^{w_t \mathcal{F}_0} \leq S \cdot e^{w_t} \leq S \cdot e^{2(n+t)}$ by Property 5, we have that this number is bounded from above by $S(\ln S+2(n+t)) \leq o(2^{2^n})$. The theorem follows. ■

4. Conclusion

The most interesting open question is, of course, whether the proof of Theorem 2 can be made constructive. The connection with V_1^1 mentioned in Introduction suggests the following specific form of this question.

Can we find a good approximation model $\mathfrak{M}$ such that, as a subset of $F_n \times \mathcal{P}(\mathcal{F})$, it is recognizable in polynomial time, and the operations $\wedge, \vee$ are polynomially time computable? Here "good" means "such that $\rho(f_n, \mathfrak{M}) \geq n^{w(1)}$ for some choice of $f_n \in F_n$", and "polynomial" means "polynomial in 2^n".

References

1. N. Alon and R. Boppana. The monotone circuit complexity of Boolean functions. *Combinatorica*, 7(1):1–22, 1987.
2. D. A. Barrington. A note on a theorem of Razborov. Technical report, University of Massachusetts, 1986.
3. R. B. Boppana and M. Sipser. The complexity of finite functions. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science, vol. A (Algorithms and Complexity)*, chapter 14, pages 757–804. Elsevier Science Publishers B.V. and The MIT Press, 1990.
4. M. Karchmer. On proving lower bounds for circuit size. In *Proceedings of the 8th Structure in Complexity Theory Annual Conference*, pages 112–118, 1993.

5. M. Karchmer and A. Wigderson. Characterizing non-deterministic circuit size. In *Proceedings of the 25th Annual ACM Symposium on the Theory of Computing*, pages 532–540, 1993.
6. M. Karchmer and A. Wigderson. On span programs. In *Proceedings of the 8th Structure in Complexity Theory Annual Conference*, pages 102–111, 1993.
7. C. Meinel. *Modified Branching Programs and Their Computational Power, Lecture Notes in Computer Science*, 370. Springer-Verlag, New York/Berlin, 1989.
8. K. Nakayama and A. Maruoka. Loop circuits and their relation to Razborov's approximation model. Manuscript, 1992.
9. A. Razborov. On the method of approximation. In *Proceedinqs of the 21st ACM Symposium on Theory of Computing*, pages 167–176, 1989.
10. A. Razborov. Bounded Arithmetic and lower bounds in Boolean complexity. To appear in the volume *Feasible Mathematics II*, 1993.
11. R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th ACM Symposium on Theory of Computing*, pages 77–82, 1987.
12. É. Tardos. The gap between monotone and nonmonotone circuit complexity is exponential. *Combinatorica*, 8:141–142, 1988.
13. A. Wigderson. The fusion method for lower bounds in circuit complexity. In *Combinatorics, Paul Erdos is Eighty.* 1993.
14. A. E. Andreev. Ob odnom metode poluqeni ninih ocenok slonosti individualnyh monotonnyh funkci. DAN *CCCP*, 282(5):1033–1037, 1985. A.E. Andreev, On a method for obtaining lower bounds for the complexity of individual monotone functions. *Soviet Math. Dokl.* 31(3):530–534, 1985.
15. A. E. Andreev. Ob odnom metode poluqeni ffektivnyh ninih ocenok monotonno slonosti. Algebra ì logika, 26(1):3–21, 1987: A.E. Andreev, On one method of obtaining effective lower bounds of monotone complexity. *Algebra i logika*, 26(1):3–21, 1987. In Russian.
16. A. A. Razborov. Ninie ocenki monotonno slo nosti nekotoryh bulevyh funkci. DAN *CCCP*, 281(4):798–801, 1985. A. A. Razborov, Lower bounds for the monotone complexity of some Boolean functions, *Soviet Math. Dokl.*, 31:354–357, 1985.
17. A. A. Razborov. Ninie ocenki monotonno slonosti logiqeskogo permanenta. *Mame*m *3a*m., 37(6):887–900, 1985. A. A. Razborov, Lower bounds of monotone complexity of the logical permanent function, *Mathem. Notes of the Academy of Sci. of the USSR*, 37:485–493, 1985.
18. A. A. Razborov. Ninie ocenki razmera shem ograniqenno glubiny v polnom bazise, soderawem funkci logiqeskogo sloeni. *Mame*m *3a*m., 41(4):598- 607, 1987. A. A. Razborov, Lower bounds on the size of bounded-depth networks over a complete basis with logical addition, *Mathem. Notes of the Academy of Sci. of the USSR*, 41(4):333–338, 1987.
19. A. A. Razborov. Ninie ocenki slonosti realizacii simmetriqeskih bulevyh funkci kontaktno-ventilnymi shemami. Matem, 3am., 48(6):79–91, 1990. A. A. Razborov, Lower bounds on the size of switching-and-rectifier networks for symmetric Boolean functions, *Mathem. Notes of the Academy of Sci. of the USSR.*

The Erdős Existence Argument

Joel Spencer

J. Spencer (✉)
Courant Institute, New York University, New York, NY 10012, USA
e-mail: spencer@cims.nyu.edu

Summary. The Probabilistic Method is now a standard tool in the combinatorial toolbox but such was not always the case. The development of this methodology was for many years nearly entirely due to one man: Paul Erdős. Here we reexamine some of his critical early papers. We begin, as all with knowledge of the field would expect, with the 1947 paper Erdős P (1947) Some remarks on the theory of graphs. Bull Amer Math Soc 53:292–294 giving a lower bound on the Ramsey function $R(k,k)$. There is then a curious gap (certainly *not* reflected in Erdős's overall mathematical publications) and our remaining papers all were published in a single ten year span from 1955 to 1965.

1. 1947: Ramsey $R(k,k)$

Let us repeat the key paragraph nearly verbatim. Erdős defines $R(k,l)$ as the least integer so that given any graph G of $n \geq R(k,l)$ vertices then either G contains a complete graph of order k or the complement G' contains a complete graph of order l.

Theorem 1. *Let $k \geq 3$, then*

$$2^{k/2} < R(k,k) \leq \binom{2k-2}{k-1} < 4^{k-1}.$$

Proof. The second inequality was proved by Szekeres thus we only consider the first one. Let $N \leq 2^{n/2}$. Clearly the number of graphs of N vertices equals $2^{N(N-1)/2}$. (We consider the vertices of the graph as distinguishable.) The number of different graphs containing a complete graph of order k is less than

$$\binom{N}{k}\frac{2^{N(N-1)/2}}{2^{k(k-1)/2}} < \frac{N^k}{k!}\frac{2^{N(N-1)/2}}{2^{k(k-1)/2}} < \frac{2^{N(N-1)/2}}{2} \qquad (*)$$

since by a simple calculation for $N \leq 2^{k/2}$ and $k \geq 3$

$$2N^k < k!2^{k(k-1)/2}.$$

But it follows immediately from $(*)$ that there exists a graph such that neither it nor its complementary graph contains a complete subgraph of order k, which completes the proof of the Theorem. □

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_27,

Erdős used a counting argument above, in the more modern language we would speak of the random graph $G \sim G(n,p)$ with $p = \frac{1}{2}$. The probability that G contains a complete graph of order k is less than

$$\binom{N}{k} 2^{-k(k-1)/2} < \frac{N^k}{k!} 2^{-k(k-1)/2} < \frac{1}{2}$$

(calculations as in the original paper) and so the probability that G or G' contains a complete graph is less than one so that with positive probability G doesn't have this property and therefore there exists a G as desired. Erdős has related that after lecturing on his result the probabilist J. Doob remarked "Well, that's very nice but it really is a counting argument." For this result the proofs are nearly identical, the probabilistic proof having the minor advantage of avoiding the annoying $2^{N(N-1)/2}$ factors. Erdős writes interchangeably in the two styles. As the methodology has progressed the probabilistic ideas have become more subtle and today it is quite rare to see a paper written in the counting style. We'll take the liberty of translating Erdős's later results into the more modern style.

The gap between $2^{k/2}$ and 4^k for $R(k,k)$ remains one of the most vexing problems in Ramsey Theory and in the Probabilistic Method. All improvements since this 1947 paper have been only to smaller order terms so that even today $\lim R(k,k)^{1/k}$ could be anywhere from $\sqrt{2}$ to 4, inclusive. Even the existence of the limit has not been shown!

2. 1955: Sidon Conjecture

Let S be a set of positive integers. Define $f(n) = f_S(n)$ as the number of representations $n = x + y$ where x, y are distinct elements of S. We call S a *basis* if $f(n) > 0$ for all sufficiently large n. Sidon, in the early 1930s, asked if there existed "thin" bases, in particular he asked if for all positive ε there existed a basis with $f(n) = O(n^{\varepsilon})$. Erdős heard of this problem at that time and relates that he told Sidon that he thought he could get a solution in "a few days". It took somewhat longer. In 1941 Erdős and Turán made the stronger conjecture that there exists a basis with $f(n)$ bounded from above by an absolute constant—a conjecture that remains open today. In 1955 Erdős [1] resolved the Sidon conjecture with the following stronger result.

Theorem 2. *There exists S with $f_S(n) = \Theta(\ln n)$.*

Proof. The proof is probabilistic. Define a random set by $\Pr[x \in S] = p_x$, the events being mutually independent over integers x, setting

$$p_x = K\left(\frac{\ln x}{x}\right)^{1/2}$$

where K is a large absolute constant. (For the finitely many x for which this is greater than one simply place $x \in S$.) Now $f(n)$ becomes a random variable. For each $x < y$ with $x + y = n$ let I_{xy} be the indicator random variable for $x, y \in S$. Then we may express $f(n) = \sum I_{xy}$. From Linearity of Expectation

$$E[f(n)] = \sum E[I_{xy}] = \sum p_x p_y \sim K' \ln n$$

by a straightforward calculation.

Lets write $\mu = \mu(n) = E[f(n)]$. The key ingredient is now a *large deviation* result. One shows, say, that

$$\Pr[f(n) < \frac{1}{2}\mu] < e^{-c\mu}$$

$$\Pr[f(n) > 2\mu] < e^{-c\mu}$$

where c is a positive absolute constant, not dependent on n, K or μ. This makes intuitive sense: as $f(n)$ is the sum of mutually independent rare indicator random variables it should be roughly a Poisson distribution and such large deviation bounds hold for the Poisson. Now pick K so large that K' is so large that $c\mu > 2 \ln n$. Call n a failure if either $f(n) > 2\mu$ or $f(n) < \mu/2$. Each n has probability less than $2n^{-2}$ failure probability. By the Borel-Cantelli Lemma (as $\sum n^{-2}$ converges) almost surely there are only a finite number of failures and so almost surely this random S has the desired properties. □

While the original Erdős proof was couched in different, counting, language the use of large deviation bounds can be clearly seen and, on this count alone, this paper marks a notable advance in the Probabilistic Method.

3. 1959: High Girth, High Chromatic Number

Tutte was the first to show the existence of graphs with arbitrarily high chromatic number and no triangles, this was extended by Kelly to arbitrarily high chromatic number and no cycles of sizes three, four or five. A natural question occurred—could graphs be found with arbitrarily high chromatic number and arbitrarily high girth—i.e., no small cycles. To many graph theorists this seemed almost paradoxical. A graph with high girth would locally look like a tree and trees can easily be colored with two colors. What reason could force such a graph to have high chromatic number? As we'll see, there is a global reason: $\chi(G) \geq n/\alpha(G)$. To show $\chi(G)$ is large one "only" has to show the nonexistence of large independent sets.

Erdős [2] proved the existence of such graphs by probabilistic means. Fix l, k, a graph is wanted with $\chi(G) > l$ and no cycles of size $\leq k$. Fix $\varepsilon < \frac{1}{k}$, set $p = n^{\varepsilon-1}$ and consider $G \sim G(n,p)$ as $n \to \infty$. There are small cycles, the expected number of cycles of size $\le k$ is

$$\sum_{i=3}^{k} \frac{(n)_i}{2i} p^i = \sum_{i=3}^{k} O\big((np)^i\big) = o(n)$$

as $k\varepsilon < 1$. So almost surely the number of edges in small cycles is $o(n)$. Also fix positive $\eta < \varepsilon/2$. Set $\lfloor u = n^{1-\eta} \rfloor$. A set of u vertices will contain, on average, $\mu \sim u^2p/2 = \Omega(n^\alpha)$ edges where $\alpha = 1 + \varepsilon - 2\eta > 1$. Further, the number of such edges is given by a Binomial Distribution. Applying large deviation results, the probability of the u points having fewer than half their expected number of edges is $e^{-c\mu}$. As $\alpha > 1$ this is smaller than exponential, so $o(2^{-n})$ so that almost surely *every* u points has at least $\mu/2$ edges. We need only that $\mu/2 > n$.

Now Erdős introduces what is now called the Deletion Method. This random graph G almost surely has only $o(n)$ edges in small cycles and every u vertices have at least n edges. Take a specific graph G with these properties. Delete all the edges in small cycles giving a graph G^-. Then certainly G^- has no small cycles. As fewer than n edges have been deleted every u vertices of G^-, which had more than n edges in G, still has an edge. Thus the independence number $\alpha(G^-) \le u$. But

$$\chi(G^-) \ge \frac{n}{\alpha(G^-)} \ge \frac{n}{u} \sim n^\eta$$

As n can be arbitrarily large one can now make $\chi(G^-) \ge k$, completing the proof.

The use of counting arguments became a typographical nightmare. Erdős considered all graphs with precisely m edges where $m = \lfloor n^{1+\varepsilon} \rfloor$. He needed that almost all of them had the property that every u vertices (u as above) had more than n. The number of graphs failing that for a given set of size u was then

$$\begin{aligned}\sum_{i=1}^{n} \binom{\binom{u}{2}}{i}\binom{\binom{n}{2}-\binom{u}{2}}{m-i} &< (n+1)\binom{\binom{u}{2}}{n}\binom{\binom{n}{2}-\binom{u}{2}}{m} \\ &< u^{2n}\binom{\binom{n}{2}-\binom{m}{2}}{m} < \binom{\binom{n}{2}}{m}u^{2n}\left(1-\frac{\binom{u}{2}}{\binom{n}{2}}\right)^m \\ &< \binom{\binom{n}{2}}{m}u^{2n}\left(1-\frac{u^2}{n^2}\right)^m < \binom{\binom{n}{2}}{m}u^{2m}e^{-mu^2/n^2}.\end{aligned}$$

Now the number of possible choices for the u points is

$$\binom{n}{u} < n^u < u^n$$

and so the number of graphs without the desired property is

$$\binom{\binom{n}{2}}{m} u^{3n} e^{-n^{1+\varepsilon-2\eta}} = o\left(\binom{\binom{n}{2}}{m}\right)$$

as desired. Today, with large deviation results assumed beforehand, the proof can be given in one relatively leisurely page.

Many consider this one of the most pleasing applications of the Probabilistic Method as the result seems not to call for probability in the slightest and earlier attempts had been entirely constructive. The further use of large deviations and the introduction of the Deletion Method greatly advanced the Probabilistic Method. And, most important, the theorem gives an important truth about graphs. In a rough sense the truth is a negative one: chromatic number cannot be determined by local considerations only.

4. 1961: Ramsey $R(3, k)$

Ramsey Theory was one of Paul Erdős's earliest interests. The involvement can be dated back to the winter of 1932/33. Working on a problem of Esther Klein, Erdős proved his famous result that in every sequence of $n^2 + 1$ real numbers there is a monotone subsequence of length $n + 1$. At the same time, and for the same problem, George Szekeres rediscovered Ramsey's Theorem. Both arguments appeared in their 1935 joint paper [10]. Bounds on the various Ramsey functions, particularly the function $R(l, k)$, have fascinated Erdős ever since. We have already spoken of his 1947 paper on $R(k, k)$. In his 1961 paper Erdős [3] proves

$$R(3, k) > c\frac{k^2}{\ln^2 k}.$$

The upper bound $R(3, k) = O(k^2)$ was already apparent from the original Szekeres proof so the gap was relatively small. Only in 1994 was the correct order $R(3, k) = \Theta\left(\frac{k^2}{\ln k}\right)$ finally shown.

Erdős shows that there is a graph on n vertices with no triangle and no independent set of size x where $x = \lceil An^{1/2} \ln n \rceil$, and A is a large absolute constant. This gives $R(3, x) > n$ from which the original statement follows easily. We'll ignore A in our informal discussion. He takes a random graph $G(n, p)$ with $p = cn^{-1/2}$. The probability that some x-set is independent is at most

$$\binom{n}{x}(1 - p)^{x(x-1)/2} < [ne^{-p(x-1)/2}]^x$$

which is *very* small. Unfortunately this G will have *lots* ($\Theta(n^{3/2})$) of triangles. One needs to remove an edge from each triangle without making any of the x-sets independent.

The Erdős method may be thought of algorithmically. Order the edges $e_1, \ldots, e_m$ of $G \sim G(n,p)$ arbitrarily. Consider them sequentially and reject e_i if it would make a triangle with the edges previously accepted, otherwise accept e_i. The graph G^- so created is certainly triangle free. What about the sets of x vertices. Call a set S of x vertices *good* (in G, not G^-) if it contains an edge e which cannot be extended to a triangle with third vertex outside of S. Suppose S is good and let e be such an edge. Then S cannot be independent in G^-. If e is accepted we're clearly OK. The only way e could be rejected is if e is part of a triangle e, e_1, e_2 where the other edges have already been accepted. But then e_1, e_2 must (as S is good) lie in S and again S is not independent.

Call S bad if it isn't good. Erdős shows that almost always there are no bad S. Lets say something occurs with high probability if its failure probability is $o\big(1/\binom{n}{x}\big)$. It suffices to show that a given $S = \{1, \ldots, x\}$ is good with high probability. This is the core of the argument. We expose (to use modern terminology) G in two phases. First we examined the pairs $\{s,t\}$ with $s \in S$, $t \notin S$. For each $t \notin S$ let $d(t)$ be the number of edges to S. Set

$$Z = \sum_{t \notin S} \binom{d(t)}{2}.$$

Each $d(t)$ has Binomial Distribution $B(x,p)$ and so expectation $xp = \Theta(\ln n)$ so that one can get fairly easily $E[Z] = \Theta(n \ln^2 n)$. Note this is the same order as x^2. It is definitely not easy to show that for appropriate A, c (Erdős takes $c = A^{-1/2}$ and A large) that $Z < \frac{1}{2}\binom{x}{2}$ with high probability. The requirement "with high probability" is quite severe. But note, at least, that this is a pure probability statement. Lets accept it and move on. Call a pair $\{i,j\} \subset S$ soiled if it lies in a triangle with third vertex outside of S. At most Z pairs are soiled so with high probability at least $\frac{1}{2}\binom{x}{2}$ pairs are unsoiled. Now we expose the edges of G inside S. If any of the unsoiled pairs are in G then G is good and so the failure probability is at most

$$(1-p)^{\frac{1}{2}\binom{x}{2}} < e^{-\Omega(px^2)} = o\left(\binom{n}{x}^{-1}\right)$$

and so G is good with high probability.

Sounds complicated. Well, it is complicated and it is simultaneously a powerful application of the Probabilistic Method and a technical *tour de force*. The story has a coda: the Lovász Local Lemma, developed in the mid-1970s, gave a new sieve method for showing that a set of bad events could simultaneously not hold. This author applied it to the random graph $G(n,p)$ with $p = cn^{-1/2}$ with the bad events being the existence of the various potential triangles and the independence of the various x-sets. The conditions of the Local Lemma made for some calculations but it was relatively straightforward to duplicate this result. Still, the ideas behind this proof, the subtle extension of the Deletion Method notion, are too beautiful to be forgotten.

5. 1962: No Local Coloring

With his 1957 paper previously discussed Erdős had already shown that chromatic number cannot be considered simply a local phenomenon. With this result he puts the nail in the coffin.

Theorem 3 ([4]). *For any $k \geq 3$ there is an $\varepsilon > 0$ so that the following holds for all sufficiently large n: There exists a graph G on n vertices which cannot be k-colored and yet the restriction of G to any εn vertex subgraph can be 3-colored.*

Often probabilistic theorems are best understood as negative results, as counterexamples to natural conjectures. A priori, for example, one might conjecture that if every, say, $n/(\ln n)$ vertices could be 3-colored then G could be 4-colored. This theorem disproves that conjecture.

We examine the random graph $G \sim G(n, p)$ with $p = c/n$. As in the 1957 paper

$$\Pr[\alpha(G) \geq x] < \binom{n}{x}(1-p)^{\binom{x}{2}} < [(ne/x)e^{-p(x-1)/2}]^x.$$

When c is large and, say, $x = 10n(\ln c)/c$, the bracketed quantity is less than one so the entire quantity is $o(1)$ and a.s. $\alpha(G) \leq x$ and so $\chi(G) \geq c/(10 \ln c)$. Given k Erdős may now simply select c so that, with $p = c/n$, $\chi(G) > k$ a.s.

Now for the local coloring. If some set of $\leq \varepsilon n$ vertices cannot be 3-colored then there is a minimal such set S with, say, $|S| = i \leq \varepsilon n$. In the restriction $G|_S$ every vertex v must have degree at least 3—otherwise one could 3-color $S - \{v\}$ by minimality and then color v differently from its neighbors. Thus $G|_S$ has at least $3i/2$ edges. The probability of G having such an S is bounded by

$$\sum_{i=4}^{\varepsilon n} \binom{n}{i}\binom{i}{2} 3i/2 p^{3i/2} \leq \sum_{i=4}^{\varepsilon n} \left[\frac{ne}{i}\left(\frac{ei}{3}\right)^{3/2}\left(\frac{c}{n}\right)^{3/2}\right]^i$$

employing the useful inequality $\binom{a}{b} \leq (\frac{ea}{b})^b$. Picking $\varepsilon = \varepsilon(c)$ small the bracketed term is always less than one, the entire sum is $o(1)$, a.s. no such S exists, and a.s. every εn vertices may be 3-colored.

Erdős's monumental study with Alfred Rényi "On the Evolution of Random Graphs" [8] had been completed only a few years before. The behavior of the basic graph functions such as chromatic and clique number were fairly well understood throughout the evolution. The argument for local coloring required a "new idea" but the basic framework was already in place.

6. 1963/4: Coloring Hypergraphs

Let $A_1, \ldots, A_m$ be n-sets in an arbitrary universe Ω. The family $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ is 2-colorable (Erdős used the term "Property B") if there is a 2-coloring of the underlying points Ω so that no set A_i is monochromatic. In 1963 Erdős gave perhaps the quickest demonstration of the Probabilistic Method.

Theorem 4 ([5]). *If $m < 2^{n-1}$ then $\mathcal{A}$ is 2-colorable.*

Proof. Color Ω randomly. Each A_i has probability 2^{1-n} of being monochromatic, the probability some A_i is monochromatic is then at most $m2^{1-n} < 1$ so with positive probability no A_i is monochromatic. Take that coloring. □

In 1964 Erdős [6] showed this result was close to best possible.

Theorem 5. *There exists a family $\mathcal{A}$ with $m = cn^2 2^n$ which is not 2-colorable.*

Here Erdős turns the original probability argument inside out. Before the sets were fixed and the coloring was random, now, essentially, the coloring is fixed and the sets are random. He sets $\Omega = \{1, \ldots, u\}$ with u a parameter to be optimized later. Let $A_1, \ldots, A_m$ be random n-sets of Ω. Fix a coloring χ with a red points and $b = u - a$ blue points. As A_i is random

$$\Pr[\chi(A_i) \text{ constant}] = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{u}{n}} \leq \frac{2\binom{u/2}{n}}{\binom{u}{n}}.$$

The second inequality, which follows from the convexity of $\binom{x}{n}$, indicates that it is the equicolorings that are the most troublesome. As the A_i are independent

$$\Pr[\text{no } A_i \text{ monochromatic}] \leq \left[1 - \frac{2\binom{u/2}{n}}{\binom{u}{n}}\right]^m.$$

Now suppose

$$2^u \left[1 - \frac{2\binom{u/2}{n}}{\binom{u}{n}}\right]^m < 1$$

The expected number of χ with no A_i monochromatic is less than one. Therefore there is a choice of $A_1, \ldots, A_m$ for which no such χ exists, i.e., $\mathcal{A}$ is not 2-colorable. Solving, one may take

$$m = \left\lceil \frac{u \ln 2}{-\ln\left[1 - 2\binom{u/2}{n}/\binom{u}{n}\right]} \right\rceil$$

Estimating $-\ln(1 - \varepsilon) \sim \varepsilon$ this is roughly $cu\binom{u}{n}/\binom{u/2}{n}$. This leads to an interesting calculation problem (as do many problems involving the

Probabilistic Method!)—find u so as to maximize m. The answer turns out to be $u \sim n^2/2$ at which value $m \sim (e \ln 2) n^2 2^{n-2}$.

Erdős has defined $m(n)$ as the least m for which there is a family of n-sets which cannot be 2-colored. His results give $\Omega(2^n) = m(n) = O(n^2 2^n)$. Beck has improved the lower bound to $\Omega(n^{1/3} 2^n)$ but the actual asymptotics of $m(n)$ remain elusive.

7. 1965: Unrankable Tournaments

Let T be a tournament with players $1, \ldots, n$ (each pair play one game and there are no ties) and σ a ranking of the players, technically a permutation on $\{1, \ldots, n\}$. Call game $\{i, j\}$ a nonupset if i beats j and $\sigma(i) < \sigma(j)$; an upset if i beats j but $\sigma(j) < \sigma(i)$. The fit $f(T, \sigma)$ is the number of nonupsets minus the number of upsets. One might have thought—in preprobabilistic days!—that every tournament T had a ranking σ with a reasonably good fit. With J.W. Moon, Erdős [7] easily destroyed that conjecture.

Theorem 6. *There is a T so that for all σ*

$$f(T, \sigma) \le n^{3/2} (\ln n)^{1/2}.$$

Thus, for example, there are tournaments so that under any ranking at least 49% of the games are upsets. Erdős and Moon take the random tournament, for each pair $\{i, j\}$ one "flips a fair coin" to see who wins the game. For any fixed σ each game is equally likely to be upset or nonupset and the different games are independent. Thus $f(T, \sigma) \sim S_m$, where $m = \binom{n}{w}$ and S_m is the number of heads minus the number of tails in m flips of a fair coin. Large deviation theory gives

$$\Pr[S_m > \alpha] < e^{-\frac{\alpha^2}{2m}}.$$

One now uses *very* large deviations. Set $\alpha = n^{3/2} (\ln n)^{1/2}$ so that the above probability is less than $n^{-n} < 1/n!$. This super small probability is used because there are $n!$ possible σ. Now with positive probability no σ has $f(T, \sigma) > \alpha$. Thus there is a T with no σ having $f(T, \sigma) > \alpha$.

The use of extreme large deviations has become a mainstay of the Probabilistic Method. But I have a more personal reason for concluding with this example. Let $g(n)$ be the least integer so that every tournament T on n players has a ranking σ with $f(T, \sigma) \ge g(n)$. Then $g(n) \le n^{3/2} (\ln n)^{1/2}$. Erdős and Moon showed $g(n) > cn$, leaving open the asymptotics of $g(n)$. In my doctoral dissertation I showed $g(n) > c_1 n^{3/2}$ and later (but see de la Vega [11] for the "book proof") that $g(n) < c_2 n^{3/2}$. Though at the time I was but an ε Paul responded with his characteristic openness and soon [9] I had an Erdős number of one. Things haven't been the same since.

References

1. P. Erdős, Problems and results in additive number theory, in *Colloque sur la Théorie des Nombres (CBRM)*, Bruxelles, 1955, 127–137.
2. P. Erdős, Graph Theory and Probability, *Canad. J. Math.* 11 (1959), 34–38.
3. P. Erdős, Graph Theory and Probability II., *Canad. J. Math.* 13 (1961), 346–352.
4. P. Erdős, On circuits and subgraphs of chromatic graphs, *Mathematika* 9(1962), 170–175.
5. P. Erdős, On a combinatorial problem I., *Nordisk. Mat. Tidskr.* 11 (1963), 5–10.
6. P. Erdős, On a combinatorial problem II., *Acta. Math. Acad. Sci. Hungar.* 15 (1964), 445–447.
7. P. Erdős and J. W. Moon, On sets of consistent arcs in a tournament, *Canad. Math. Bull.* 8 (1965), 269–271.
8. P. Erdős and A. Rényi, On the evolution of random graphs, *Mat. Kutató Int. Közl.* 5 (1960), 17–60.
9. P. Erdős and J. Spencer, Imbalances in k-colorations, *Networks* 1 (1972), 379–385.
10. P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* 2 (1935), 463–470.
11. W. F. de la Vega, On the maximal cardinality of a consistent set of arcs in a random tournament, *J. Combinatorial Theory, Series B* 35 (1983), 328–332.

IV. Geometry

Introduction

Erdős' love for geometry, and elementary or discrete geometry in particular, dates back to his beginnings. The Erdős–Szekeres paper has been influential and certainly helped to create discrete geometry as we know it today. But Erdős also put geometry to the service of other branches, giving definition to various geometrical graphs and proving bounds on their chromatic and independence numbers. We are happy to include papers by Moshe Rosenfeld, Pavel Valtr, Janos Pach, Jiří Matoušek and, in particular, a paper by Miklós Laczkovich and Imre Ruzsa on the number of homothetic sets. While the paper of Peter Fishburn is closely related to Erdős' favorite theme, the papers of N. G. de Bruijn (on Penrose tiling) and J. Aczél and L. Losonczi (on functional equations) cover broader related aspects.

In 1995/6, when the content of these volumes was already crystallizing, we asked Paul Erdős to isolate a few problems, both recent and old, for each of the eight main parts of this book. To this part on Geometry Theory he contributed the following collection of problems and comments.

Erdős in his own words

Let $x_1, \dots, x_n$ be n points in the plane, not all on a line, and join every two of them. Thus we get at least n distinct lines. This follows from Gallai–Sylvester but also from a theorem of de Bruijn and myself.

My most striking contribution to geometry is no doubt my problem on the number of distinct distances. This can be found in many of my papers on combinatorial and geometric problems.

Hickerson, Pach and I proved that on the unit sphere one can find n points for which the distance $\sqrt{2}$ can occur among $n^{1/3}$ pairs. Perhaps this is best possible. In fact, there are $n^{1/3}$ points at distance 1 from every other point. For every $0 < \alpha < 2$ there are n points so that for every point there are $\log^* n$ other points at distance α—again we do not know if this is best possible.

Purdy and I proved (using an idea of Kárteszi) that there are n points in the plane with no three on a line for which the unit distance occurs at least $cn \log n$ times. We have no nontrivial upper bound. If the points are in 3-space the unit distance can occur $n^{4/3}$ times (Hickerson, Pach and myself) but if we also assume that no four are on a plane, we can do no better than $cn \log n$.

Szekeres and I proved, that if $\binom{2n-4}{n-2} + 1$ points are given in a plane no three on a line, then we can always select among them n points which are the vertices of a convex n-gon. Probably $2^{n-2} + 1$ is the correct value—we proved that 2^{n-2} is not enough. This problem (which was due to E. Klein, i.e., Mrs. Szekeres) had a great influence.

*****

So much for Paul Erdős in 1996. He stated it here explicitly: the distance problem is highlighted as the most important of his problems in geometry. We are very fortunate that we can include in this part the paper by Larry Guth which surveys the solution of this problem. His article also describes the rich and fertile ground of Erdős problems in geometry.

Let us remark that some of the other advances are described in the other chapters, for example the development related to the Erdős–Szekeres Theorem is part of the chapter by Graham and Nešetřil in the Ramsey theory chapter in the second volume. We cannot be exhaustive as the number of geometrical problems and results discussed by Erdős was large. One topic which is missing in this part is clearly incidence problems. However they are partially covered in the number theory part and are also part of Guth's article.

The progress in combinatorial and discrete geometry has been spectacular. So much so that Matoušek speaks about 2010 as *annus mirabilis*:

J. Matoušek, The Dawn of an Algebraic Era in Discrete Geometry, Euro CG 2011, Morschach.

Extension of Functional Equations

János Aczél* and László Losonczi

J. Aczél (✉)
Department of Pure Mathematics, University of Waterloo, Waterloo,
ON N2L 3G1, Canada
e-mail: jdaczel@math.uwaterloo.ca

L. Losonczi
Department of Mathematics, Kuwait University, P.O.Box 5969,
Safat 13060, Kuwait
e-mail: losi@math.klte.hu

1. Introduction

Extension theorems are common in various areas of mathematics. In topology continuous extensions of continuous functions are studied. In functional analysis one is interested mainly in linear extensions of linear operators preserving continuity or some other properties like bounds or norm. In algebra extensions of homomorphisms and isomorphisms are investigated. The latter can be considered as extensions of functional equations.

In the area of functional equations one is interested in extending functional equations, i.e., either extending functions satisfying a functional equation (or a system of equations) on a "restricted" set to functions which satisfy the same equation (or system of equations) on the "maximal" set or showing that given functions satisfying a functional equation (or a system of equations) on a restricted set satisfy the same equation (or system) on a "larger" set.

The first extension theorem for the Cauchy functional equation is in the 1965 joint paper [AE 65] of Erdős with Aczél who proved that, if a function $f :]0, \infty[\to \mathbb{R}$ satisfies the Cauchy functional equation for all positive values of x and y, that is

$$f(x+y) = f(x) + f(y) \qquad (x, y \in]0, \infty[),$$

then there exists a unique function $F : \mathbb{R} \to \mathbb{R}$ such that $F(x) = f(x)$ if $x \in]0, \infty[$ (i.e. F is an extension of f) and F satisfies the Cauchy equation for all values of x, y:

$$F(x+y) = F(x) + F(y) \quad (x, y \in \mathbb{R}). \tag{1}$$

* Research supported by the Natural Sciences and Engineering Research Council of Canada grants nr. OGP002972 and CPG0164211

1991 Mathematics Subject Classification: 39 B 22, 39 B 52

Key words and phrases: functional equation, extension of functional equation

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_28,

In the same paper they showed that there exist no Hamel bases of the set of nonnegative numbers, i.e., there does not exist any set B' whose elements are nonnegative and any nonnegative number is representable as a linear combination of a finite number of elements of B' with nonnegative rational coefficients in a unique way.

In 1960 in the problems section of Colloquium Mathematicum Erdős [E 60] raised the following problem. If $f(x+y) = f(x) + f(y)$ is satisfied for almost all (x, y) in the plane (in the sense of plane Lebesgue measure) does there exist a function F satisfying the Cauchy equation everywhere (i.e. satisfying (1)) such that $f(x) = F(x)$ almost everywhere (in the sense of the linear Lebesgue measure)?

The Aczél-Erdős theorem and the problem of Erdős (solved independently by Jurkat [J 65] and de Bruijn [B 66]) became the starting points of two different directions of the extension theory of functional equations, which by now is quite extensive (Kuczma [K 78] quoted already in 1978 more than 100 related papers; see also [PR 95]).

The aim of this paper is to show how the above result and problem of Paul Erdős influenced the development of the extension theory of functional equations.

2. Extensions of Homomorphisms

Let G, H be (multiplicatively written) groups and S be a subsemigroup of S. Let further $f : S \to H$ be a homomorphism of S into H. Under what conditions can f be extended in a unique way to a homomorphism of G? With $G = H =$ the additive group of $\mathbb{R}$ and $S =$ the additive group of positive real numbers, we get back to the Aczél- Erdős extension problem. The general problem was treated by Aczél et al. [ABDKR 71] by proving several theorems which under different assumptions on S gave a solution. The first one is closely related to the extension theorem in [AE 65].

Suppose that for every element $x \in G$, different from the unit element, either $x \in S$ or $x^{-1} \in S$ (or both). Then every homomorphism $f : S \to H$ can be extended uniquely to a homomorphism $F : G \to H$ of G into H.

This can be proved by defining F by

$$F(x) := \begin{cases} E & \text{if } x = e \\ f(x) & \text{if } x \in S \\ f(x^{-1})^{-1} & \text{if } x^{-1} \in S \end{cases}$$

where e, E are the unit elements of G, H, respectively.

A unique extension also exists when S generates the Abelian group G.

Indeed, in this case $G = S \cdot S^{-1} = \{xy^{-l} \mid x \in S, y \in S\}$ and by

$$F(xy^{-1}) := f(x)f(y)^{-1} \quad (x, y \in S)$$

F is well defined on G (because $xy^{-1} = uv^{-l}$ $(x, y, u, v \in S)$ implies $xv = uy$) and supplies the unique extension of f.

A more elaborate result is the following.

Suppose that S is a subsemigroup of the group G such that

$$G = S \cdot S^{-1} \cdot S \cdot S^{-1} = \{xy^{-l}uv^{-l} \mid x, y, u, v \in S\} \tag{2}$$

and f is a homomorphism of S into H. The map f can be extended to a homomorphism of G into H if and only if

$$x, y, z, u, v, w \in S,\ xy^{-1}z = uv^{-1}w \textit{ implies } f(x)f(y)^{-1}f(z) = f(u)f(v)^{-1}f(w).$$

When the extension exists, then it is unique.

There are further results in this vein in [ABDKR 71]. Martin [M 77] generalized these results to the case when G can be represented in a product form similar to (2) but with an arbitrary number of factors.

Necessary and sufficient conditions for the extension were found by Osondu [O 78] in the case when S generates G:

Let S be a subsemigroup of a group G which is generated by S and f a homomorphism of S into some group H. Then f can be extended to a homomorphism of G into H if and only if f satisfies the following condition.

For every positive integer n and for every $s_i \in S$, $\epsilon_i \in \{-1, 1\}$ $(i = 1, 2, \ldots, n)$

$$\prod_{i=1}^{n} s_i^{\epsilon_i} = e \Longrightarrow \prod_{i=1}^{n} f(s_i)^{\epsilon_i} = E$$

holds, where e, E are the unit elements of G, H respectively.

In this result the necessary and sufficient condition seems to be too "close" to the statement itself. A more detached condition may be desirable.

3. Extensions of the Cauchy and Pexider Equation and Their Applications

We introduce some terminology (see Daróczy-Losonczi [DaL 67]). Let D be a nonempty subset of $\mathbb{R}^2$. We write

$$D_x := \{x \mid \exists_y : (x, y) \in D\},$$
$$D_y := \{y \mid \exists_x : (x, y) \in D\},$$
$$D_{x+y} := \{x + y \mid (x, y) \in D\}.$$

Then D_x, D_y are the projections of D onto the x-axis, y-axis, respectively, and D_{x+y} is the projection of D, parallel to the line $x+y = 0$, onto the x-axis. Let further $D' := D_x \cup D_y \cup D_{x+y}$. We say that a function f is *additive on the set D* (or satisfies the Cauchy equation on D) if $f : D' \to \mathbb{R}$ and

$$f(x+y) = f(x) + f(y) \quad ((x, y) \in D) \tag{3}$$

holds. A function $F : \mathbb{R} \to \mathbb{R}$ is called an (*additive*) *extension* of a function f additive on D (we say also that F extends the Cauchy equation from D to $\mathbb{R}^2$) if F satisfies the Cauchy equation on $\mathbb{R}^2$ and $F(x) = f(x)$ for $x \in D'$.

We saw above that there exists a unique (additive) extension from $\mathbb{R}_+^2$ ($\mathbb{R}_+$ is the set of positive reals) to $\mathbb{R}^2$ (take $S = \mathbb{R}_+$). As Daróczy and Losonczi [DaL 67] showed, *there always exists a unique (additive) extension from a neighborhood of* $(0,0) to \mathbb{R}^2$. They took the neighborhood to be circular but it turned out [A 83] that "hexagonal neighborhoods" of $(0,0)$

$$H_r := \{(x,y) \mid x, y, x+y \in]-r, r[\}$$

are more convenient. Indeed, if

$$f(x+y) = f(x) + f(y) \text{ for } (x,y) \in H_r \tag{4}$$

then define for *any* $t \in \mathbb{R}$

$$F(t) := nf\left(\frac{t}{n}\right) (n \in \mathbb{N},\, \frac{t}{n} \in]-r, r[).$$

This definition is unambiguous since, by (4),

$$nf\left(\frac{t}{n}\right) = nmf\left(\frac{t}{nm}\right) = mnf\left(\frac{t}{mn}\right) = mf\left(\frac{t}{m}\right) \text{ for } \frac{t}{n}, \frac{t}{m} \in]-r, r[.$$

Clearly $F(x) = f(x)$ for $x \in]-r, r[$ (choose $n = 1$). The function F, so defined, is additive on $\mathbb{R}^2$ since, for arbitrary $u, v \in \mathbb{R}$, there exists an $n \in \mathbb{N}$ such that $\frac{u}{n}, \frac{v}{n}, \frac{u+v}{n} \in]-r, r[$, thus, again by (4),

$$F(u+v) = nf\left(\frac{u}{n} + \frac{v}{n}\right) = nf\left(\frac{u}{n}\right) + nf\left(\frac{v}{n}\right) = F(u) + F(v).$$

This $F : \mathbb{R} \to \mathbb{R}$, extending (4) from H_r to $\mathbb{R}^2$, is unique because, if also $\bar{F} : \mathbb{R} \to \mathbb{R}$ were additive on $\mathbb{R}^2$ and would satisfy $\bar{F}(x) = f(x)$ for $x \in]-r, r[$ then, for arbitrary $t \in \mathbb{R}$ choosing $n \in \mathbb{N}$ again so that $\frac{t}{n} \in]-r, r[$, we get

$$\bar{F}(t) = \bar{F}\left(n\frac{t}{n}\right) = n\bar{F}\left(\frac{t}{n}\right) = nf\left(\frac{t}{n}\right) = F(t) \text{ for all } t \in \mathbb{R}$$

as asserted.

However there may not exist an additive extension even in some very simple cases. Let, for instance f be defined by

$$f(x) = \begin{cases} 2x+1 & \text{for } x \in]2,3[\\ 2x+3 & \text{for } x \in]4,5[\\ 2x+4 & \text{for } x \in]6,8[\end{cases} . \tag{5}$$

This function is additive on $D =]2,3[\times]4,5[$ but no additive extension to $\mathbb{R}^2$ exists. Indeed, such an extension F would be continuous on $]2,3[$, thus (see e.g. [A 66]) there would exist a constant c such that $F(x) = cx$ for all $x \in \mathbb{R}$ which contradicts $F(x) = f(x)$ for $x \in D' =]2,3[\cup]4,5[\cup]6,8[$. Nevertheless, we see from (5), that $F(x) = 2x$ is "nearly" an extension of this f: it differs from f on D_x, D_y and D_{x+y} just by a constant each and is additive everywhere. Such functions are called *quasi-extensions.*

To be exact, if

$$f(x+y) = f(x) + f(y) \quad ((x,y) \in D) \tag{3}$$

holds and there exist constants α, β and a function F, additive on $\mathbb{R}^2$, such that

$$\begin{aligned} f(x) &= F(x) + \alpha \qquad \text{for all } x \in D_x, \\ f(y) &= F(y) + \beta \qquad \text{for all } y \in D_y, \\ f(z) &= F(z) + \alpha + \beta \quad \text{for all } z \in D_{x+y} \end{aligned} \tag{6}$$

then F is an (additive) quasi-extension of f (from D to $\mathbb{R}^2$). These formulas are reminiscent of the general solution

$$\begin{aligned} f(x) &= F(x) + \alpha \\ g(y) &= F(y) + \beta \\ h(z) &= F(z) + \alpha + \beta \end{aligned} \tag{7}$$

(F is an arbitrary additive function on $\mathbb{R}^2$, α and β are arbitrary constants) of the Pexider equation

$$h(x+y) = f(x) + g(y) \quad ((x,y) \in \mathbb{R}^2).$$

Since in (5) or (6) (and also in (3)) the three occurences of f are anyway defined on possibly different intervals, this is conducive to consider (3) as a *Pexider equation*

$$h(x+y) = f(x) + g(y) \quad ((x,y) \in D) \tag{8}$$

with f, g, h defined on D_x, D_y, D_{x+y}, respectively. As it turns out, these have an extension (not only quasi-extension) to $\mathbb{R}^2$ from the above set $D =]2,3[\times]4,5[$ and from every open connected set (region) D.

In fact, the basic result concerning quasi-extensions of the Cauchy equation and extensions of the Pexider equation is that *from any open connected set (region) D in $\mathbb{R}^2$ there exists a unique quasi-extension of the Cauchy equation* (3) *to* $\mathbb{R}^2$ (Daróczy-Losonczi [DaL 67]) *and a unique extension of the Pexider equation* (8) *to* $\mathbb{R}^2$ (Rimán [R 76], Aczél [A 85], and Radó-Baker [RB 87]). The former follows from the latter. Both can be proved by taking for D first a hexagon, like H; above, but its centre shifted from the origin to (a,b), showing, for Eq. (8), by substitutions that

$$h(u+v+a+b) - h(a+b) = f(u+a) - f(a) + g(v+b) - g(b), \tag{9}$$

where (u,v) is in the hexagon H_r around the origin. This implies (putting $v=0$ or $u=0$, respectively) that

$$h(t+a+b)-h(a+b)=f(t+a)-f(a)=g(t+b)-g(b) \quad (t\in]-r,r[).$$

Defining the function Φ by equating $\Phi(t)$ to this common value on $]-r,r[$, (9) shows that Φ is additive on H_r so, by what we proved above, has a unique (additive) extension F to $\mathbb{R}^2$. The functions given by

$$\begin{aligned}\tilde{f}(x)&=F(x)-F(a)+f(a)\\ \tilde{g}(y)&=F(y)-F(b)+g(b)\\ \tilde{h}(z)&=F(z)-F(a+b)+h(a+b)\end{aligned}\tag{10}$$

$(x,y,z\in\mathbb{R})$, and only these, extend the Pexider equation from the (a,b)-centred hexagon to $\mathbb{R}^2$. The proof that *the Pexider equation has a unique extension from any open connected region D to* $\mathbb{R}^2$ can be completed by covering D by a sequence of hexagons such that each hexagon is contained in D and any two consecutive hexagons have nonempty intersection and by applying the above extension process to each hexagon. Of course, (10) has the form (7) of the general solution of Pexider's equation on $\mathbb{R}^2$.

The general solution of the Pexider equation (8) on D is obtained by replacing, in (10), $-F(a)+f(a)$ and $-F(b)+g(b)$ by two arbitrary constants and $-F(a+b)+h(a+b)$ by their sum, and restricting $\tilde{f},\tilde{g},\tilde{h}$ to D_x, D_y, D_{x+y} respectively. The function F, additive on $\mathbb{R}^2$, which we have just determined, is also the unique quasi-extension of f in (3) from D to $\mathbb{R}^2$, in the sense (6). If two of the sets $D_x\bigcap D_y$, $D_x\bigcap D_{x+y}$, $D_y\bigcap D_+x+y$ are nonempty, then the quasi-extension is an extension of the Cauchy equation.

We note that the concept of extension and quasi-extension can be defined similarly for more general (in particular some closed) domains, ranges (groups, vector spaces) and equations. In their paper [RB 87], quoted above, Radó and Baker had a real or complex topological vector space as domain of $\tilde{f}$, $\tilde{g}$, $\tilde{h}$ and an Abelian group as range. We mention a few further results.

Let G and H be Abelian divisible groups and let X be a subset of G having the properties

(i) For any $x\in X$ and for all rational numbers $\lambda\in]0,1[$ we have $\lambda_x\in X$;
(ii) For any pair $(x,y)\in X^2$ there exists an integer n, which may depend upon x and y, such that $(x+y)/n\in X$.

Suppose that $f:X\to H$ is additive on $D:=\{(x,y)\mid x\in X,\ y\in X,\ x+y\in X\}$. Then there exists an additive extension $F:G\to H$ of f. Moreover F is unique if, and only if, the subgroup generated by X is G. (Ng [N 74] and Dhombres-Ger [DG 78].)

Székelyhidi [S 72] found the general representation of functions which are *additive on an open set $D\subset\mathbb{R}^2$*. He also proved [S 81] an extension theorem for the equation $\Delta_y^{n+1}f(x)=0$ where Δ_y is the usual difference operator defined by $\Delta_y f(x) := f(x+y) - f(x)$. Let again $D \subset \mathbb{R}^2$, n a positive integer and for $k = 0, 1, \ldots, n+1$ let

$$D_k := \{x + (n+1-k)y \mid (x,y) \in D\}.$$

The result by Székelyhidi is the following. *Let $D \subset \mathbb{R}^2$ be an open and connected set with $(0,0) \in D$. Let n be a positive integer and $f : \bigcup_{k=0}^{n+1} D_k \to \mathbb{R}$ be a function such that*

$$\Delta_y^{n+1} f(x) = \sum_{k=0}^{n+1} \binom{n+1}{k} (-1)^k f[x + (n+1-k)y] = 0 \quad ((x,y) \in D).$$

Then there exists an extension $F : \mathbb{R} \to \mathbb{R}$ of f such that $\Delta_y^{n+1} F(x) = 0$.

The Levi-Civitá equation

$$f(x+y) = \sum_{k=1}^{n} g_k(x) h_k(y) \tag{11}$$

is a common generalization of the Cauchy and Pexider equation and the sine and cosine equations.

Recently the second author [L 85a, L 90] proved an extension theorem which corresponds to the Aczél-Erdős result for Eq. (11).

Let the functions $f, g_k, h_k :]0, \infty[\to \mathbb{C}$ ($k = 1, \ldots, n$) satisfy the functional equation (11) *for all positive x, y and assume that the functions $g_1, \ldots, g_n$ and $h_1, \ldots, h_n$ are linearly independent on $]0, \infty[$. Then there exists a unique set of functions $F, G_k, H_k : \mathbb{R} \to \mathbb{C} (k = 1, \ldots, n)$ such that*

$$F(x+y) = \sum_{k=1}^{n} G_k(x) H_k(y) \quad (x, y \in \mathbb{R})$$

and

$$\begin{aligned} F(x) &= f(x) \\ G_k(x) &= g_k(x) \quad (x \in]0, \infty[; \quad k = 1, \ldots, n) \\ H_k(x) &= h_k(x) \end{aligned}$$

hold. If $f, g_k, h_k (k = 1, \ldots, n)$ are continuous or measurable on $]0, \infty[$ then so are F, G_k, H_k ($k = 1, \ldots, n$) on $\mathbb{R}$. The proof is based on the following observation. It is known ([A 66] pp. 201–203) that the functions $F_1, \ldots, F_n$, defined on a set A with values in an arbitrary field, are linearly independent on A if, and only if, there exist elements $a_1, \ldots, a_n \in A$ (necessarily distinct) such that $\det(F_i(a_j))_{i,j=l}^n \neq 0$. Thus by the linear independence of the functions $g_1, \ldots, g_n$ there exist $x'_1, \ldots, x'_n \in]0, \infty[$ such that $\det(g_i(x'_j))_{i,j=1}^n \neq 0$. Without loss of generality we may suppose that $x'_1 = \min\{x'_1, \ldots, x'_n\}$. Let $x_k = x'_k - x'_1 (k = 1, \ldots, n)$ and substitute $x + x_k$ for $x (k = 1, \ldots, n)$ in (11). We obtain a system of equations which can be written as

$$\mathcal{F}(x+y) = \mathcal{G}(x)\mathcal{H}(y) \quad (x, y \in]0, \infty[) \tag{12}$$

where, for $x \in]0, \infty[$ $(x_1 = 0)$,

$$\mathcal{F}(x) = (f(x+x_1), \ldots, f(x+x_n))^T$$

$$\mathcal{G}(x) = \begin{pmatrix} g_1(x+x_1) \ldots g_n(x+x_1) \\ \vdots \qquad\qquad \vdots \\ g_1(x+x_n) \ldots g_n(x+x_n) \end{pmatrix}$$

$$\mathcal{H}(x) = (h_1(x), \ldots, h_n(x))^T$$

and T denotes transposition. Extending the Pexider equation (12) we obtain an extension of (11).

4. Almost Everywhere Additive Functions

The problem of Erdős, mentioned in the introduction, was inspired by Hartman [H 61] who proved that, if $f(x+y) = f(x) + f(y)$ is satisfied for $(x, y) \in A \times A$, where the complement of A has (linear) measure zero, then there exists a function $F : \mathbb{R} \to \mathbb{R}$ satisfying the Cauchy equation everywhere and $F(x) = f(x)$ almost everywhere on. $\mathbb{R}$ Actually Hartman's paper [H 61] appeared later than [E 60] as a partial solution to the Erdős problem [E 60]. Jurkat [J 65] and de Bruijn [B 66] (independently) solved the problem in the affirmative, showing that, *if $f(x+y) = f(x) + f(y)$ holds for all $(x, y) \in D$ where the Lebesgue measure of $\mathbb{R}^2 \setminus D$ is zero, then there exists a function $F : \mathbb{R} \to \mathbb{R}$ satisfying the Cauchy equation everywhere and $F(x) = f(x)$ a.e., that is, for all $x \in D_1$ where $\mathbb{R} \setminus D_1$ is of Lebesgue measure zero.*

Is this F an (additive) extension of f (in the sense of Daróczy-Losonczi)?

The answer is *no*. Counter example (by V. Zinde-Walsh, simplified by C.T. Ng and J. Aczél, see [A 80, AD 89]):

$$D := \{(x, y) \mid x, y, x+y \neq 1, 2\} \cup \{(1, 1)\},$$

i.e., D is obtained from $\mathbb{R}^2$ by removing from it the lines $x = 1$, $x = 2$, $y = 1$, $y = 2$, $x+y = 1$, $x+y = 2$ except the point (1,1). Clearly $\mathbb{R}^2 \setminus D$ is of Lebesgue measure zero and $D' = \mathbb{R}$. Let

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R} \setminus \{1, 2\}, \\ 1 & \text{if } x = 1, \\ 2 & \text{if } x = 2. \end{cases}$$

This function is additive on D (we have $f(1+1) = f(2) = 2 = 1 + 1 = f(1) + f(1)$ and elsewhere, for $(x, y) \in D$, we get $f(x) = f(y) = f(x+y) = 0$) but f is obviously not additive on $D' = \mathbb{R}$.

Questions similar to Erdős's problem can be asked for other functional equations and inequalities.

Let $I \subset \mathbb{R}$ be an open interval. A function $f : I \to \mathbb{R}$ is called *almost convex* if

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} \quad ((x,y) \in I \times IM),$$

where $M \subset I \times I$ is of planar Lebesgue measure zero. Kuczma [K 70] proved that, *if $f : I \to \mathbb{R}$ is almost convex, then there exists a unique convex function $F : I \to \mathbb{R}$ such that $F(x) = f(x)$ a.e. in I.*

Ger [G 71, G 75] noticed that "equality almost everywhere" can also be introduced in an axiomatic way.

A nonempty family $\mathcal{I}^k$ of subsets of the k-dimensional euclidean space $\mathbb{R}^k$ is called *a Linearly Independent Proper Ideal* (abbreviated as LIPI) if

(a) A, $B \in \mathcal{I}^k$ implies $A \cup B \in \mathcal{I}^k$,
(b) $A \in \mathcal{I}^k, B \subset A$ implies $B \in \mathcal{I}^k$,
(c) $\mathbb{R}^k \notin \mathcal{I}^k$,
(d) For every real number $\alpha \neq 0, \beta \in \mathbb{R}^k$, and $A \in \mathcal{I}^k$ we have $\alpha A + \beta \in \mathcal{I}^k$.

It can be easily seen that the family $\mathcal{L}_0^k$ of all subsets of $\mathbb{R}^k$ with Lebesgue measure zero is a LIPI. Similarly, the families $\mathcal{F}^k$ and $\mathcal{L}_f^k$ of all sets of the first category in $\mathbb{R}^k$ and of all sets having finite Lebesgue measure in $\mathbb{R}^k$, are also LIPIs.

We say that two LIPIs $\mathcal{I}^2$ and $\mathcal{I}^1$ are *conjugate* if for every $M \in \mathcal{I}^2$ there exists a set $U \in \mathcal{I}^1$ such that, for every $x \in U$, the set

$$V_x := \{y \mid (x,y) \in M\}$$

belongs to $\mathcal{I}^1$. In view of Fubini's theorem, the LIPIs $\mathcal{L}_0^2$, $\mathcal{F}^2$, $\mathcal{L}_f^2$ and $\mathcal{L}_0^1$, $\mathcal{F}^1$, $\mathcal{L}_f^1$ are, respectively, conjugate.

Ger [G 71] proved the following result. *Let $\mathcal{I}^2$ and $\mathcal{I}^1$ be conjugate linearly invariant proper ideals. If $f : \mathbb{R} \to \mathbb{R}$ satisfies the equation $\Delta_y^{n+1} f(x) = 0$ for all $(x,y) \in \mathbb{R}^2 M$ where $M \in \mathcal{I}^2$ then there exists exactly one function $F : \mathbb{R} \to \mathbb{R}$ and a set $U \in \mathcal{I}^1$ such that $\Delta_y^{n+1} F(x) = 0$ holds for all $(x,y) \in \mathbb{R}^2$ (i.e., F is a polynomial function) and $F(x) = f(x)$ for all $x \notin U$.*

Ger [G 75] found similar results for the Mikusiński equation

$$f(x+y)[f(x+y) - f(x) - f(y)] = 0$$

and for the Pexider equation

$$f(x+y) = g(x) + h(y)$$

5. Some Applications

The above extension theorems have important and amusing applications. An example is that of "aggregated allocations" see [AKNW 83, ANW 84, AW 81].

Suppose that there is a certain amount s of quantifiable goods (raw materials, energy, money for scientific projects, grants etc.) is to be allocated (completely) to $m(\geq 3)$ projects. For this purpose a committee of n assessors is formed. The j-th assessor recommends that the amount x_{ij} should be allocated to the j-th project. If he (she) can count, then $\sum_{j=1}^{m} x_{ij} = s$. Now the recommendations should be aggregated into a consensus allocation by a chairman or external authority. Again the allocated consensus amounts should *add up* to s. It is supposed (this is a restriction) that the aggregated allocation for the j-th project depends only on the recommended allocations to *that* project, that both the individual and the aggregated allocations be *non-negative* (!) and, if all assessors recommend rejection, then the consensus allocation to that project should be 0 ('*consensus on rejection*'). If we write the individual allocations for the j-th project as a vector $\mathbf{x}_j = (x_{1j}, x_{2j}, \ldots, x_{nj})$ and denote the aggregated allocation by $f_j(\mathbf{x}_j)$ then these conditions mean the following:

$$f_j : [0,s]^n \to [0,s], \quad f_j(\mathbf{0}) = 0 \quad (j = 1, 2, \ldots, m)$$

and

$$\sum_{j=1}^{m} \mathbf{x}_j = s\mathbf{1} \Longrightarrow \sum_{j=1}^{m} f_j(\mathbf{x}_j) = s \tag{13}$$

where $\mathbf{1} = (1, 1, \ldots, 1)$, $s\mathbf{1} = (s, s, \ldots, s)$, and $\mathbf{0} = 0\mathbf{1} = (0, 0, \ldots, 0)$. With $\mathbf{x}_1 = s\mathbf{1}$, $\mathbf{x}_2 = \ldots = \mathbf{x}_m = \mathbf{0}$ we get $f_1(s\mathbf{1}) = s$ and, since the subscript 1 has no privileged role, $f_j(s\mathbf{1}) = s$ for all $j = 1, 2, \ldots, m$ ('*consensus on overwhelming merit*'). Now we substitute into (13) $\mathbf{x}_1 = \mathbf{z} = (z_1, z_2, \ldots, z_n)$, $\mathbf{x}_3 = s\mathbf{1} - \mathbf{z}$, $\mathbf{x}_2 = \mathbf{x}_4 = \ldots = \mathbf{x}_m = \mathbf{0}$:

$$f_1(\mathbf{z}) = s - f_3(s\mathbf{1} - \mathbf{z}) \quad (\mathbf{z} \in [0,s]^n).$$

The substitution of $\mathbf{x}_1 = \mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{x}_2 = \mathbf{y} = (y_1, y_2, \ldots, y_n)$, $\mathbf{x}_3 = s\mathbf{1} - \mathbf{x} - \mathbf{y}$, $\mathbf{x}_4 = \ldots = \mathbf{x}_m = \mathbf{0}$ gives

$$f_1(\mathbf{x}) + f_2(\mathbf{y}) = s - f_3(s\mathbf{1} - \mathbf{x} - \mathbf{y}) = f_1(\mathbf{x} + \mathbf{y}), \tag{14}$$

whenever $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x} + \mathbf{y}$ are all in $[0,s]^n$. Putting here $\mathbf{x} = \mathbf{0}$ we get $f_2(\mathbf{y}) = f_1(\mathbf{0}) + f_2(\mathbf{y}) = f_1(\mathbf{y})$ for all $\mathbf{y} \in [0,s]^n$. Again the subscripts $1, 2$ have no privileged role, so

$$f_1 = f_2 = \ldots = f_m = f \quad \text{on } [0,s]^n$$

and (14) reduces to the Cauchy equation

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \quad ((\mathbf{x}, \mathbf{y}) \in D), \tag{15}$$

where $D := \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x}, \mathbf{y}, \mathbf{x} + \mathbf{y} \in [0,s]^n\}$. All solutions of (15) can be extended from D (even though it is closed) to $\mathbb{R}^n \times \mathbb{R}^n$. Applying the known result that every n-place additive function, nonnegative on an n-dimensional interval, is of the form of an inner product $\langle \mathbf{a}, \mathbf{x} \rangle$ ($\mathbf{a}$ is a constant vector with nonnegative components a_i), we have

$$f_j(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n} a_i x_i \quad \left(a_i > 0, \sum_{i=1}^{n} a_i = 1; j = 1, 2, \dots, m \right).$$

We have found that the *weighted arithmetic mean* is the general solution of our allocation problem.

Many applications concern *local solutions* of functional equations. The functional equation

$$f(x + y - xy) + f(xy) = f(x) + f(y) \quad (x, y \in]0, 1[, [0, 1], \text{ or } \mathbb{R})$$

was introduced by Hosszú. The most complicated case is the one where $f :]0, 1[\to \mathbb{R}$ and the equation holds for $x, y \in]0, 1[$. The *general solution* is (Lajkó [La 74])

$$f(x) = A(x) + b \quad (x \in]0, 1[)$$

where $A : \mathbb{R} \to \mathbb{R}$ is an arbitrary additive function (on $\mathbb{R}^2$) and b is an arbitrary constant. The proof is based on the fact that the function $x \to f(x + 1/2) - f(1/2)$ is additive on the "triangle-like" set

$$D = \{(x, y) \mid -\frac{1}{2} < x < 0, \frac{1}{2} + \frac{1}{2x - 1} < y < \frac{1}{2} + 1\}.$$

A is obtained as the quasi-extension of this function.

The extension theorem for the Levi-Cività equation (together with the extension theorem of Daróczy-Losoncai) can be applied to the solution of *functional equations of sum form* (see [L 85a, L 91]).

Kiesewetter [Ki 65] proved that the "arctan equation"

$$f(x) + f(y) = f\left(\frac{x + y}{1 - xy}\right) \quad (x, y \in \mathbb{R},\ 1 - xy \neq 0) \tag{16}$$

has no continuous solutions except $f(x) = 0$ *(*$x \in \mathbb{R}$*)*. Crstici, Muntean and Vornicescu [CMV 83] found the local solutions of the arctan equation valid on the set $D_1 := \{(x, y) \in \mathbb{R}^2 \mid xy < 1\}$ satisfying various regularity conditions (continuity at one point or boundedness or measurability on an interval etc.). The general local solution on D_1 was found by Muntean and Vornicescu [MV 83].

The second author proved [L 85b, L 90] that *the general solution of* (16) *is the function*

$$f(x) = A(\arctan x) \quad (x \in \mathbb{R})$$

where $A : \mathbb{R} \to \mathbb{R}$ *is an arbitrary additive function (on* $\mathbb{R}^2$*) periodic with period* π*.*

This clearly implies $f(\tan r\pi) = 0$ for all rational numbers r, explaining Kiesewetter's result. In [L 85b, L 90, MV 83] again the extension theorem of Daróczy and Losonczi was the main tool of the proof.

Finally we mention that the result of de Bruijn and Jurkat can be applied to find all bounded multiplicative linear functionals on the complex Banach algebra of Lebesgue integrable functions (under convolution) [A 80].

Reference

[A 66] J. Aczél, *Lectures on Functional Equations and Their Applications*, Academic Press, New York-London, 1966 [Mathematics in Science and Engineering, Vol. 19].

[A 80] J. Aczél, *Some good and bad characters I have known and where they led. (Harmonic analysis and functional equations)*, In: 1980 Seminar on Harmonic Analysis. [Canad. Math. Soc. Conf. Proc., Vol. 1]. Amer Math. Soc., Providence, RI, 1981, pp. 177–187.

[A 83] J. Aczél, *Diamonds are not the Cauchy extensionist's best friend*, C. R. Math. Rep. Acad. Sci. Canada 5 (1983), 259—264.

[A 85] J. Aczél, it 28. Remark, Report of Meeting. The Twenty-second International Symposium on Functional Equations (December 16-December 22, 1984, Oberwolfach, Germany). Aequationes Math. 29 (1985), p. l01.

[ABDKR 71] J. Aczél, J. A. Baker, D. Z. Djoković, P1. Kannappan and F. Radó, *Extensions of certain homomorphisms of subsemigroups to homomorphisms of groups*, Aequationes Math. 6 (1971), 263–271.

[AD 89] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, Cambridge-New York-New Rochelle-Melbourne-Sydney, 1989.

[AE 65] J. Aczél and P. Erdős, *The non-existence of a Hamel-basis and the general solution of Cauchy's functional equation for nonnegative numbers*, Publ. Math. Debrecen 12 (1965), 259–265.

[AKNW 83] J. Aczél, P1. Kannappan, C. T. Ng and C. Wagner, *Functional equations and inequalities in 'rational group decision making'*, In: General Inequalities 3 (Proc. Third Internat. Conf. on General Inequalities, Oberwolfach, 1981). Birkh auser, Basel-Boston-Stuttgart, 1983, pp. 239–243.

[ANW 84] J. Aczél, C. T. Ng and C. Wagner, *Aggregation theorems for allocation problems*, SIAM J. Alg. Disc. Meth. 5 (1984), 1–8.

[AW 81] J. Aczél and C. Wagner, *Rational group decision making generalized: the case of several unknown functions*, C. R. Math. Rep. Acad. Sci. Canada 3 (1981), 139–142.

[B 66] N. G. de Bruijn, *On almost additive functions*, Colloquium Math. 15 (1966), 59–63.

[CMV 83] B. Crstici, I. Muntean and N. Vornicescu, *General solution of the arctangent functional equation*, Anal. Numér, Théor. Approx. 12 (1983), 113–123.

[DaL 67] Z. Daróczy and L. Losonczi, *Uber die Erweiterung der auf einer Punktmenge additiven Funktionen*, Publ. Math. Debrecen 14 (1967), 239–245.

[DG 78] J. Dhombres and R. Ger, Conditional Cauchy equations, Glas. Mat. Ser. III. 13 (33) (1978), 39–62.

[E 60] P. Erdős, *P 310*, Colloquium Math. 7 (1960), 311.

[G 71] R. Ger, *On almost polynomial functions*, Colloquium Math. 24 (1971),95–101.

[G 75] R. Ger, *On some functional equations with a restricted domain*, Fundamenta Math. 89 (1975), 95-l01.

[H 61] S. Hartman, *A remark about Cauchy's equation*, Colloquium Math. 8 (1961), 77–79.

[J 65] W. B. Jurkat, *On Cauchy's functional equation*, Proc. Amer. Math. Soc. 16 (1965), 683–686.

[Ki 65] H. Kiesewetter, *Uber die arc tan-Funktionalgleichung, ihre mehrdeutigen stetigen Losunqeti und eine nichtstetige Gruppe*, Wiss. Z. Friedrich-Schiller-Univ. Jena Math.-Natur. 14 (1965), 417–421.

[K 70] M. Kuczrna, *Almost convex functions*, Colloquium Math. 21 (1970), 279–284.

[K 78] M. Kuczma, *Functional equations on restricted domains*, Aequationes Math. 18 (1978), 1–35.

[La 74] K. Lajkó, *Applications of extensions of additive functions*, Aequationes Math. 11 (1974), 68–76.

[L 85a] L. Losonczi, *An extension theorem*, Aequationes Math. 28 (1985), 293–299.

[L 85b] L. Losonczi, Remark 32: *The general solution of the arc tan equation*, In: Proc. Twenty-third Internat. Symp. on Functional Equations (Gargnano, Italy, June 2–11, 1985). Univ. of Waterloo, Centre for Information Theory, Waterloo, Ont., 1985, pp. 74–76.

[L 90] L. Losonczi, *Local solutions of functional equations*, Glasnik Mat. 25 (45) (1990), 57–67.

[L 91] L. Losonczi, *An extension theorem for the Levi-Cività. functional equation and its applications*, Grazer Math. Ber. 315 (1991), 51–68.

[M 77] S. C. Martin, *Extensions and decompositions of homomorphisms of semigroups*, Manuscript, University of Waterloo, Ont., 1977.

[MV 83] I. Muntean and N. Vornicescu, *On the arctangent functional equation*, (Roumanian), Seminarul "Theodor Angheluta", Cluj-Napoca, 1983, pp. 241–246.

[N 74] C. T. Ng, *Representation for measures of information with the branching property*, Inform. and Control 25 (1974), 45–56.

[O 78] K. E. Osondu, *Extensions of homomorphisms of a subsemigroup of a group*, Semigroup Forum 15 (1978), 311–318.

[PR 95] L. Paganoni and J. R atz, *Conditional functional equations and orthogonal additivity*, Aequationes Math. 50 (1995), 134–141.

[RB 87] F. Radó and J. A. Baker, *Pexider's equation and aggregation of allocations*, Aequationes Math. 32 (1987), 227–239.

[R 76] J. Rimán, *On an extension of Pexider's equation*, Zbornik Radova Mat. Inst. Beograd N. S. 1(9) (1976), 65–72.

[S 72] L. Székelyhidi, *The general representation of an additive function on an open point set*, (Hungarian), Magyar Tud. Akad. Mat. Fiz. Oszt. K ozl. 21 (1972), 503–509.

[S 81] L. Székelyhidi, *An extension theorem for a functional equation*, Publ. Math. Debrecen 28 (1981), 275–279.

Remarks on Penrose Tilings

N. G. de Bruijn

N.G. de Bruijn (Deceased) (✉)
Department of Mathematics and Computer Science, Technische Universiteit
Eindhoven, 5600 MB Eindhoven, The Netherlands

1. Introduction

1.1

This paper will cover some details on Penrose tilings presented in lectures over the years but never published in print before. The main topics are: (i) the characterizability of Penrose tilings by means of a local rule that does not refer to arrows on the edges of the tiles, and (ii) the fact that the Ammann quasigrid of the inflation of a Penrose tiling is topologically equivalent to the pentagrid that generates the original tiling.

The fact that any Penrose tiling is the topological dual of the Ammann quasigrid of the inflation was first noticed by Socolar and Steinhardt [9]. They presented it in the equivalent form that the topological dual of the Ammann quasigrid of a Penrose tiling is the *deflation* of that tiling.

The Ammann quasigrid of the inflation of a Penrose tiling can also be defined as the union of the central lines of the stacks of that tiling. Therefore I refer to that union as the *central grid* of the tiling. It can be defined independently of the original notion of Ammann grid. Actually the definition of the central grid can also be given for other kinds of tilings where there is no obvious definition of an Ammann grid and no obvious definition of deflation.

The paper is intended to be readable more or less independently of previous ones, at least in the sense that all relevant notions will be explained in the paper itself.

1.2

My geometric terminology will use the notion of shapes and 1-shapes in the plane. A *shape* is an equivalence class of figures under similarity transform. Similarity includes multiplications (with respect to a point), shifts and rotations, but no reflections with respect to a line. In terms of complex numbers, this similarity transform just means linear transformation. A 1-*shape* is defined similarly, but without multiplications. That means that similarity is replaced by congruence. In other words, figures with the same 1-shape have the same shape and the same size.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_29,

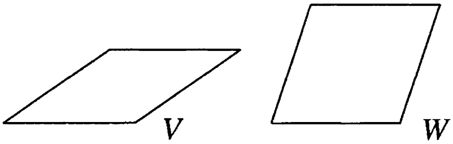


Fig. 1 The 1-shapes V (thin rhomb) and W (thick rhomb)

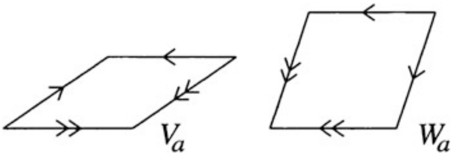


Fig. 2 Penrose's arrowed rhombs V_a (thin) and W_a (thick)

Throughout this paper I shall use the two 1-shapes V and W, pictured in Fig. 1. They are rhombs, with all edges having unit length. The acute angles of V are 36°, and those of W are 72°. V will be called the *thin rhomb* and W the *thick rhomb*. The word "rhomb" will always refer to a V or a W.

The *arrowed rhombs* V_a and W_a (the subscript a stands for "arrowed") are obtained from V and W by putting single and double arrows on the edges in the way depicted in Fig. 2. They will be called the *Penrose rhombs*.

1.3

A *Penrose tiling* is a tiling of the plane by V_a's and W_a's with the property that two tiles always have either nothing, or a vertex, or a full edge in common; in the latter case they are called *direct neighbors* and it is required that along the common edge they have the same kind of arrow in the same direction. Figure 3 shows a fragment of such a tiling.

Various procedures for obtaining all Penrose tilings are known, like

(i) Penrose's use of deflation, in combination with a (non-constructive) selection argument (see [4] for an exposition).
(ii) The use of deflation, in combination with "updown-generation" (see [5]).
(iii) Forming the dual of a "pentagrid" and providing the arrows afterwards (see [1]).

In [5] the relation between (ii) and (iii) was studied in detail.

Somewhat more complicated, but nevertheless interesting and promising, are ways to take the Ammann bars (see [7, 9]) as the basis of the Penrose tilings.

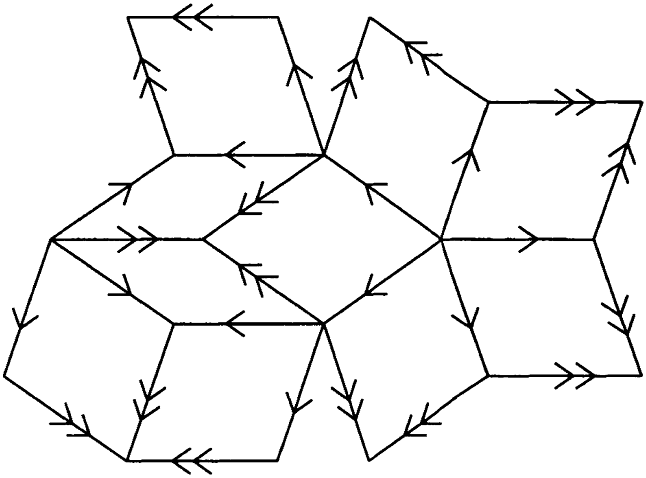


Fig. 3 A piece of a Penrose tiling

1.4

One of the topics to be treated in the present paper is the question how one can determine whether a tiling of the plane by V's and W's can be provided with arrows so as to form a Penrose tiling with V_a's and W_a's. It will be shown (Sect. 2) that this can be settled by inspecting, for every rhomb in the unarrowed tiling, the figure formed by its four direct neighbors.

Section 2 will pay attention to the question of a tiling give by the *vertices* only.

2. Arrow-Free Characterization of Penrose Tilings

2.1

The term *unarrowed rhomb tiling* will mean a tiling of the plane with V's and W's, with the condition that neighboring tiles have a full edge in common.

Any Penrose tiling turns into an unarrowed rhomb tiling by omitting the arrows on the edges.

For any unarrowed rhomb tiling, any way to attach an arrow to every edge such that it becomes a Penrose tiling, will be called an *arrowing* of the tiling, and whenever such an arrowing exists the tiling is called *arrowable.* Not every unarrowed rhomb tiling is arrowable. A simple counterexample is the doubly periodic tiling by thick rhombs only.

It will be shown that the condition for an unarrowed rhomb tiling to be arrowable can be put into a form that does not refer to arrows any more.

2.2

Consider a figure formed by an unarrowed rhomb t and four neighboring rhombs t_1, t_2, t_3, t_4, each one of these having one edge in common with t. It is required that there is no overlap between any two of these rhombs (apart from possible common edges). Let me call such a figure a *cross*.

If to a cross formed by t, t_1, t_2, t_3, t_4 one adds a number of further rhombs $t_5, \ldots, t_k$ which all have a vertex in common with t, such that there is no overlap between any two of $t, t_1, \ldots, t_k$, and such that the areas around the vertices of t are entirely covered, then that figure will be called an *extended cross*.

2.3

An *arrowing* of a cross or an extended cross is a way to attach an arrow (either single or double) to every edge in the figure such that each one of its rhombs becomes a Penrose rhomb.

Not every cross can be arrowed. And a cross that can be arrowed is not necessarily extendable to an extended cross that can be arrowed.

A cross will be called *perfect* if it is extendable to an extended cross that permits an arrowing.

2.4

It is not hard to make a list of all possibilities for perfect crosses. Starting with a thin rhomb t, trying all possible sets of direct neighbors t_1, t_2, t_3, t_4, and investigating whether they have at least one arrowable extension, one finds 6 possibilities, pairwise related by 180° rotation, leading to 3 different 1-shapes P_1, P_2, P_3. Similarly, if starting from a thick rhomb, one gets to 8 possibilities, which are pairwise related by 180° rotation, so there are 4 different 1-shapes P_4, P_5, P_6, P_7.

For the notion "1-shape of a perfect cross" I shall also use the term *neighborhood pattern*. Figure 4 gives them all.

P_1 and P_2 form a pair, in the sense that the mirror image of P_1 has the same 1-shape as P_2. In the same sense P_5 and P_6 form a pair. The others, P_3, P_4 and P_7, are symmetric: the axis of symmetry is the short diagonal of t in P_3 and the long diagonal of t in P_4 and P_7.

2.5

Each one of the seven neighborhood patterns can be arrowed in exactly one way. These arrowed patterns are called $P_{1a}, \ldots, P_{7a}$, and are shown in Fig. 5.

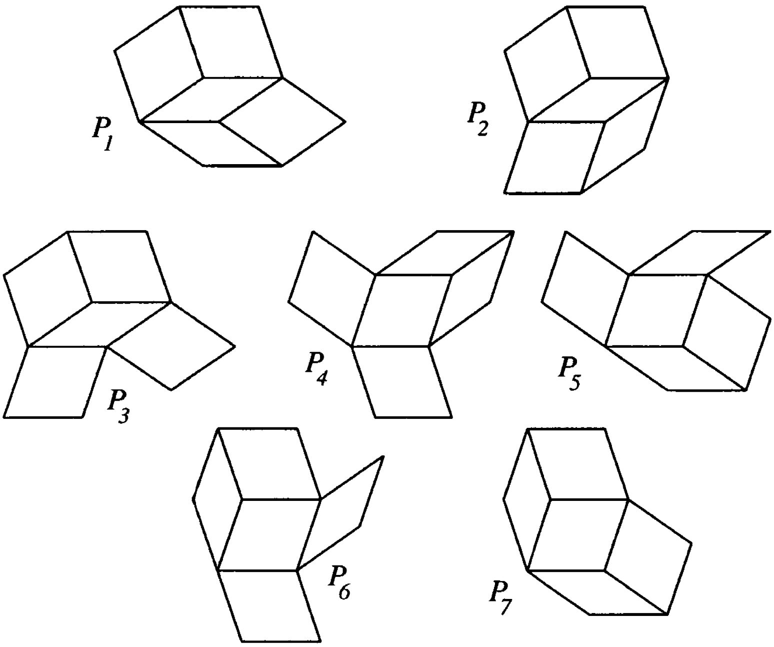


Fig. 4 The neighborhood patterns $P_1, \dots, P_7$

2.6

At this stage it becomes possible to answer the question which unarrowed rhomb tilings can be arrowed.

I shall say that an unarrowed rhomb tiling satisfies the cross condition if for each one of its rhombs the cross formed by that rhomb and its four direct neighbors is one of the seven 1-shapes $P_1, \dots, P_7$.

Theorem 1. *(i) If an unarrowed rhomb tiling of the plane is arrowable then it satisfies the cross condition.*

(ii) If an unarrowed rhomb tiling of the plane satisfies the cross condition then it can be arrowed in exactly one way.

Proof. (i) This is little more than the fact that the list of neighborhood patterns provided in Sect. 2 is exhaustive. If a cross is a part of an arrowable rhomb tiling, then it also has an extension to an extended cross that lies in that tiling, and any arrowing of the whole rhomb tiling implies an arrowing of that extended cross.

(ii) Start from an unarrowed rhomb tiling that satisfies the cross condition. So if t is one of the rhombs in the tiling then it is the central rhomb of a cross of one of the types $P_1, \dots, P_7$, and that uniquely defines an

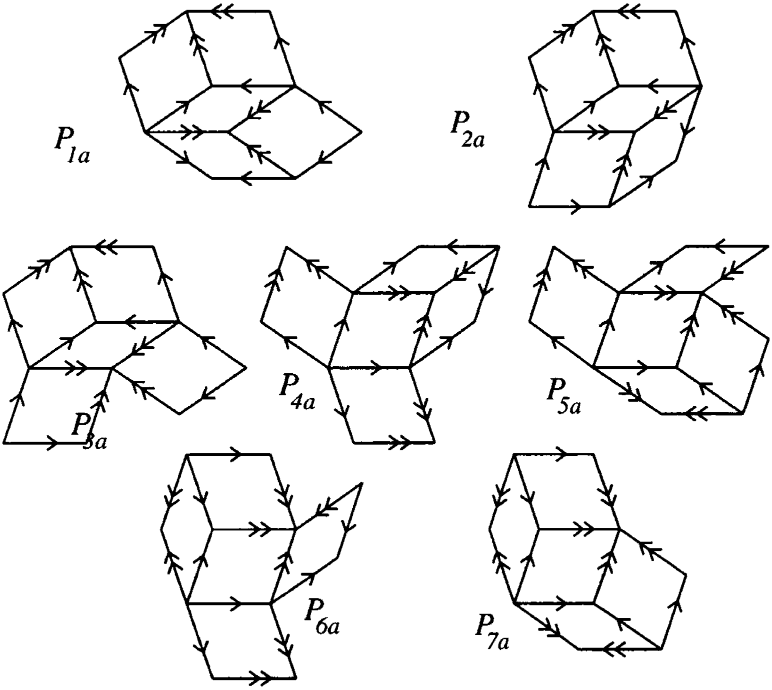


Fig. 5 The *arrowed* neighborhood patterns $P_{1a}, \ldots, P_{7a}$

arrowing of t according to Fig. 5. So the arrows on t are imposed by the cross of t. They will be called the *imposed arrows* of t.

In this way an arrowing is prescribed for every rhomb in the tiling, but it is the question whether neighbors get the same imposed arrow on their common edge. To that end it has to be shown that if in the tiling two rhombs s, t have an edge in common, then along that edge the arrow of s (imposed by the cross of s) is the same as the arrow of t (imposed by the cross of t). This can be checked without having the full tiling available.

If one of s, t is thin and the other one is thick, then uniqueness of arrowing holds already for the figure consisting of that s and t only. That arrowing on that figure is imposed by the cross of s as well as by the one of t, so in particular the imposed arrow on the common edge is the same in both cases.

The next case is that s and t are both thin. The cases where the cross of a thin rhomb contains a second thin rhomb are the P_1 and P_2 of Fig. 4. In both cases s and t have a thick rhomb u as a common neighbor in the way depicted in Fig. 6a (possibly with s and t interchanged). Just by observing the pair s, u it is seen that the cross of s imposes a double arrow on the common edge of s and t, pointing to the right. With the pair t, u the same result is obtained for the imposed arrow of t.

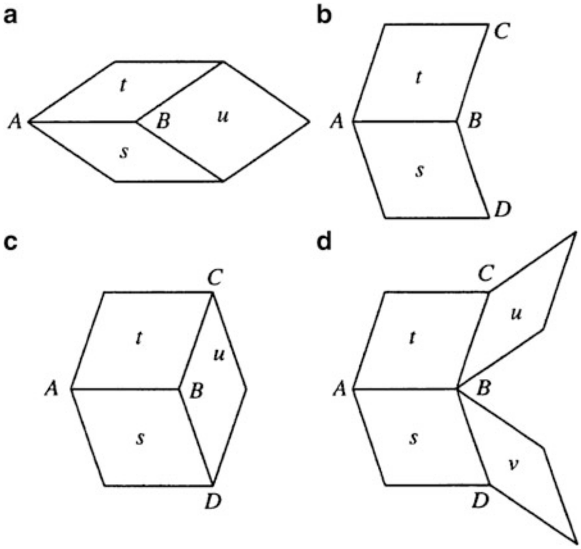


Fig. 6 The arrowing of s imposed by the cross of s and the arrowing of t imposed by the cross of t lead to the same *arrow* on the common edge

The remaining case is that both s and t are thick. This is shown in Fig. 6b (possibly with s and t interchanged). It is a part of a tiling in which the cross condition holds for both s and t. Since the cross of s contains t it cannot contain a thick rhomb u pasted to the edge BC: such a configuration formed by s, t, u does not occur in any of the crosses of s shown in Fig. 4. The same argument shows that the tiling does not contain a thick rhomb pasted to BD.

The only possibilities to paste thin rhombs to BC and BD are shown in Fig. 6c, d.

In Fig. 6c, the only possible arrowing of the pair s, u has a double arrow from B to A, whence that double arrow is imposed by the cross of s. The same argument applied to the pair t, u leads to the same imposed arrow of t.

In Fig. 6d the pair t, u shows that the imposed arrow of t on AB is a single one, pointing to the right. Inspection of the pair s, v shows the same imposed arrow of s on AB.

This completes the proof of the theorem. □

2.7

As a corollary of the theorem of Sect. 2 it can be noted that an unarrowed rhomb tiling can be arrowed in at most one way. This does not require the whole theorem: uniqueness of arrowing holds already for all arrowable extended crosses.

2.8

Every arrowable tiling contains infinitely many copies of each one of $P_1, \ldots, P_7$. Inspection of an arbitrary Penrose tiling shows occurrences of all these perfect crosses, and it is known that any finite configuration occurring in any Penrose tiling occurs infinitely often in any other one.

2.9

An unarrowed rhomb tiling is completely determined by its set of vertices, irrespective of whether the tiling is arrowable or not. In order to see this, it suffices to check that in an unarrowed rhomb tiling any two vertices have distance 1 if and only if they are connected by an edge of a rhomb of the tiling.

It is also easy to see that two vertices in an unarrowed rhomb tiling have distance between 0 and 1 if and only if they are the end-points of the short diagonal of a thin rhomb.

The shortest distance exceeding 1 is $2 \sin 36° (= 1.17557)$. It is the length of the short diagonal in a thick rhomb, but this distance may occur in a different way between vertices in a pair of adjacent thin rhombs.

3. Deflation and Inflation

3.1

This section gives some information about inflation and deflation of Penrose tilings. It is intended as a kind of motivating background, but will not be used in the rest of the paper.

Penrose's idea of inflation and deflation was very central in the discovery of his arrowed rhomb tilings. *Deflation* is a certain way to get from a Penrose tiling to a new tiling with smaller pieces ($\frac{1}{2}(-1 + \sqrt{5})$ times the original ones), and *inflation* is the inverse operation, leading to a tiling with bigger pieces. For a description of the process see [1, 6–8]. It is somewhat easier to describe the deflation as an operation on rhombus *halves*. The half thin rhomb is obtained by cutting a thin rhomb along its short diagonal, the half thick rhomb by cutting a thick rhomb along its long diagonal. For these rhomb halves the deflation is a matter of *subdivision* (see [5]). This is shown in Fig. 7.

In Sect. 2 of [5] it is explained that the idea of inflation comes naturally by observing how in an arrowed rhombus tiling pieces can be grouped together. Those groups form the pieces of the inflation.

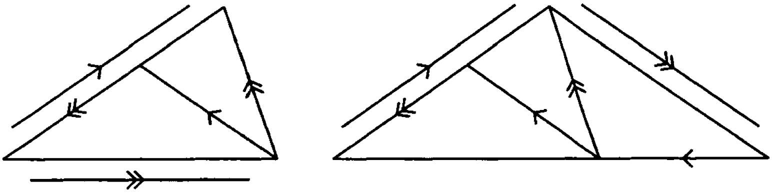


Fig. 7 The deflation of the half thin and the half thick rhomb. The full figure on the *left* is a half thin rhomb, with *arrows* drawn at the outside. It is subdivided into a half thin and a half thick rhomb, with *arrows* indicated along the edges themselves. These two small pieces belong to the deflation. Similarly the full figure on the *right* is a half thick rhomb, and here the deflation has three pieces: a half thin rhomb and two half thick rhombs. There are of course two shapes of half thin rhombs, which are each other's mirror image, and similarly there are two shapes of half thick rhombs. For the mirror images of the half rhombs shown in the figure one has to take the mirror image subdivision.
When this subdivision operation is carried out for every half thin rhomb and every half thick rhomb of a Penrose tiling, the smaller pieces fit together to a Penrose tiling with smaller pieces

3.2

If deflation is applied to a tiling of a *finite* portion of the plane, and if the result is enlarged by a factor $\frac{1}{2}(1+\sqrt{5})$ afterwards, it becomes a tiling of a bigger part of the plane with pieces of the original size. Infinite repetition of this process and application of an infinite selection principle leads, albeit in an inconstructive way, to Penrose tilings of the full plane (see [4, 6, 8]). But the idea of inflation and deflation can also be used in a completely different and very constructive way for the production of *all* infinite Penrose tilings. The method is *updown generation*, a process that is controlled by taking an arbitrary infinite path in a particular finite automaton (see [5]).

4. Duality

4.1

The key method in [1] was the production of tilings by means of their topological dual. I spend a few words here as a short independent introduction.

4.2

In any Penrose tiling the rhombs are arranged in what I shall call *stacks*. Consider an arbitrary edge e of an arbitrary rhomb r of the tiling. Two rhombs of the tiling are called e- *neighbors* if they have a common edge that is parallel to e. Two rhombs of the tiling are called e-*related* if they can be connected by a chain of e-neighbors. The stack generated by r and e is the set of all rhombs which are e-related to r.

Every rhomb of the tiling belongs to exactly two stacks, and any two stacks have exactly one rhomb in common, unless their e's are parallel.

The idea of a stack can at once be generalized to tilings by parallelotopes in higher dimensional spaces. In that case there can be stacks of various dimensions. In the two-dimensional case the stacks are one-dimensional, but in three dimensions there are two-dimensional stacks that can be considered as unions of one-dimensional stacks.

4.3

Consider a Penrose tiling and a stack generated by some r and e. In any rhomb of the stack connect the mid-points of the edges parallel to e by a straight line segment. These line segments form an infinite broken line that can be called the *back bone* of the stack.

The union of the back bones of all stacks is called the *skeleton* of the tiling. In a skeleton one has *curves* (the back bones), *points* (the intersection points of the curves) and *mazes* (the connected components of the complement of the union of all curves). These curves, points and mazes can be deformed topologically without disturbing their relation to the tiling. On anyone of the curves the order of the intersection points with other curves keeps reflecting the order in which the corresponding stack is intersected by other stacks. The topological deformations do not disturb the duality between skeleton and tiling. In that duality curves correspond to stacks, points to rhombs, mazes to rhomb vertices.

4.4

If some topological deformation of a skeleton is given, one is not yet able to reconstruct the tiling. First one has to know, for any curve of the skeleton, the direction and the length of the edge e that was used in the discussion of the corresponding stack. Once all these are known, the Penrose tiling is determined up to a parallel shift in the plane. That is, the Penrose tiling deprived of its arrows. But the latter is no great loss: since one started from a Penrose tiling in the first place, one knows that a consistent system of arrows exists, and Sect. 2 takes care of the uniqueness of the arrowing.

4.5

In [1] it was shown that the skeletons are topologically equivalent to so-called pentagrids, but this had to be taken with a grain of salt in so-called singular cases. The pent agrids are parametrized by real numbers $\gamma_0, \ldots, \gamma_4$ satisfying $\gamma_0 + \ldots + \gamma_4 = 0$. If j is one of the numbers $0, \ldots, 4$ and $\zeta = e^{2\pi i/5}$ then the set of all complex numbers z satisfying $\mathrm{Re}(z\zeta^{-j}) + \gamma_j \in \mathcal{Z}$ is a grid of parallel lines in the complex plane. The superposition of these five grids is called the *pentagrid* generated by $\gamma_0, \ldots, \gamma_4$. The pentagrid is called *singular* if it contains three lines passing through one point.

Key results of [1] are the following. Any non-singular pentagrid is the dual of a Penrose tiling. The singular pentagrids do not have a dual in the ordinary sense, but by infinitesimal perturbations they turn into grids that do have a dual. There can be various perturbations with different effect, and the result is that a singular grid corresponds either to 2 or to 10 different Penrose tilings. With these extra arrangements for the singular cases, *all* Penrose tilings are obtained from pentagrids.

The perturbations will get some more attention in Sect. 6 below.

4.6

The idea of producing a tiling as a topological dual of the superposition of a number of parallel grids can be extended to general classes of tilings of spaces of arbitrary dimension by means of parallelotopes. See [2] for a proof that under fairly general conditions the dual covers the whole space uniquely.

4.7

It can be concluded from Sect. 4 that the skeleton of a Penrose tiling is topologically equivalent to some regular pentagrid or to some perturbed singular pentagrid. But remarkably, there is always a second and completely different straight line grid with the topological structure of the skeleton. This matter will be treated in Sects. 5 and 6, where it is also shown that there are not just these two: there are infinitely many straight line grids with the same topology. The pentagrid can be deformed continuously in such a way that it always keeps the same topology and always remains a superposition of five straight line grids, with the central grid as final result. And one can even get beyond that.

5. Ammann Bars and Central Bars

5.1

An attractive way to introduce the *Ammann bars* of a Penrose tiling is to consider both the thin and the thick arrowed rhomb as a billiard table and to choose particular billiard ball tracks on them. These tracks can already be shown on the rhomb halves, displayed in Fig. 8. The tracks have been chosen in such a way that when fitting pieces together the segments combine to

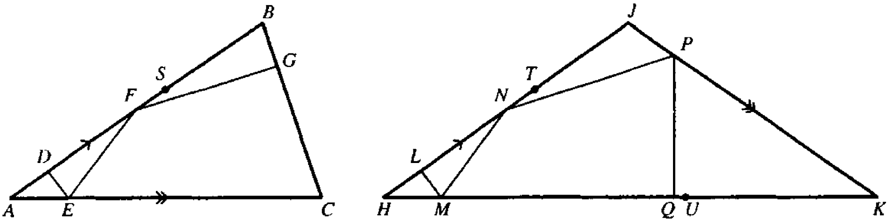


Fig. 8 The billiard ball tracks on the *arrowed* half rhombs. The angles at D, G, L and Q are 90°. The lines EF and FG form equal angles with AB, which is expressed by saying that the billiard ball bounces at the edge AB. At E, M, N and P there are similar cases of bouncing. Finally $AD = HL$, $AF = HN$ and $JP = AE = HM$. In the mirror images of these half rhombs one of course takes the mirror image tracks

infinite straight lines. These are called the Ammann bars after their discoverer R. Ammann (see [7, 9]).

The grid formed by these bars, called *Ammann quasigrid* in [9], has some of the properties of the *pentagrid* used to produce the Penrose tiling in [1]. Both are superpositions of five parallel grids, making angles which are multiples of 72°. In the pentagrid the parallel grids are all equidistant, but in the Ammann quasigrid they show two different distances. On the other hand, the Ammann quasigrid looks much simpler than the pentagrid since its meshes show only a small finite number of shapes. A pentagrid contains infinitely many different mesh shapes, and arbitrarily small ones.

It was observed by J.E.S. Socolar and P.J. Steinhardt (see [9]) that dualization of the Ammann quasigrid of a Penrose tiling leads exactly to the deflation of that tiling.

A quite simple suggestive argument can be given for this statement, although it might not be easy to turn it into a formal proof. It is explained by Fig. 9. The straight line billiard track segments can be continuously deformed into the segments in Fig. 10. The points D, E, G, L, M, P can stay where they were, but F, N and Q move. In the mirror images of these half rhombs the corresponding things are done correspondingly. If in each rhomb of a Penrose tiling the original billiard ball track is deformed continuously into the new track, then the union of all segments shows the picture of a grid deforming itself by bending its bars, all the time keeping the same topology. In the final situation (the one of Fig. 9) each mesh contains exactly one vertex of the deflation, and that is exactly the one produced by the dualization.

5.2

The Socolar-Steinhardt phenomenon can also be expressed like this: the Ammann bars of the *inflation* of a Penrose tiling form a quasigrid that is

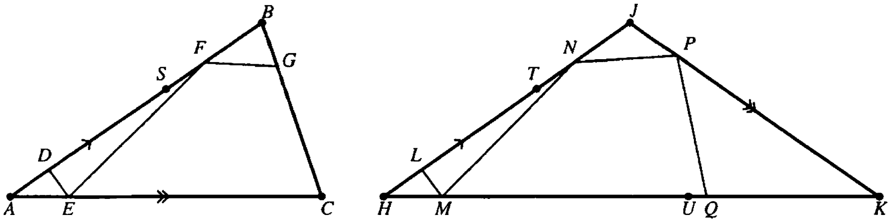


Fig. 9 The deformed billiard ball tracks on the half thin and the half thick rhomb. In the mirror images of these half rhombs mirror image tracks have to be taken. The vertices of the deflation in the above half rhombs are $A, B, C, S, H, K, J, U, T$. Having these tracks in all half rhombs of a Penrose tiling results in a set of mazes where each maze contains exactly one deflation vertex, and in the duality that vertex is exactly the vertex corresponding to that maze

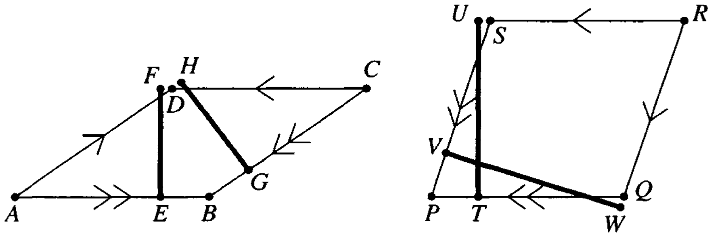


Fig. 10 The basic segments of the central bars. In the thin rhomb $ABCD$ (with *double arrows* AB and CB) the segments are EF and GH, in the thick rhomb $PQRS$ (with *double arrows* QP and SP) they are TU and VW. There are right angles at E, G, T and V. The points F, H, U, W lie on the extensions of the segments CD, AD, RS, RQ, respectively. The lengths of the bar segments are $EF = GH = \sin 36°$, $TU = VW = \sin 72°$. The positions are determined by $EB = GB = TP = VP = 0.25$. Moreover $DF = DH = SU = QW = (-2 + \sqrt{5})/4 = 0.059017$

the topological dual of that tiling itself. Let me call those Ammann bars of the inflation the *central* bars of the tiling itself. It will be shown that they can be defined independently of the billiard ball construction, and that a proof for the duality can be given without reference to the argument of Sect. 5.

The name "central" was chosen because of the the central position of these bars in the stacks of the tiling (see Sect. 6).

The central bars can be found by comparing Figs. 7 and 8. The double arrows of the deflation are from S to A, from C to B, from T to H and from U to J. These are all intersected orthogonally by the billiard ball track. The distance from the intersection point to the end-point of the double arrow is just one fourth of the length of the arrow. This shows that the central bars of a Penrose tiling have to contain the segments EF, GH, TU, VW indicated in Fig. 10.

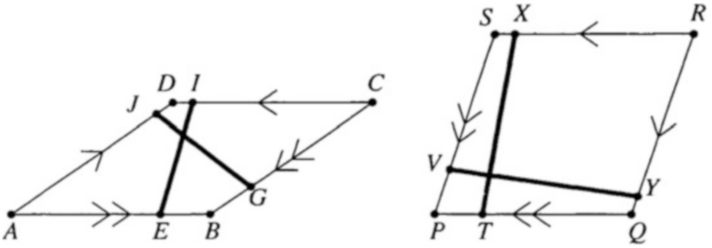


Fig. 11 The segments of the deformed bars EI, GJ in the thin rhomb and TX, VY in the thick rhomb, determined by $DI = DJ = SX = QY = 0.1$. Any maze formed by these deformed bars contains exactly one vertex of the Penrose tiling, and that is exactly the vertex corresponding to the maze in the duality

It has to be admitted that some of the rhombs contain further pieces of central bars, but in order to build the full central grid it suffices to consider those pieces of Fig. 10. Since

$$EB = GB = TP = VP, \quad CF = AH = RU = RW \tag{1}$$

it is obvious that in a stack the segments in the direction of the stack fit together to a full straight line. It is easy to show directly that (1) implies $EB = GB = TP = VP = \frac{1}{4}$, $CF = AH = RU = RW = \frac{1}{4}(2 + \sqrt{5})$, and therefore the central bars can be introduced without any reference to the billiard ball tracks.

The segments of Fig. 10 can be transformed into those of Fig. 11 by continuous deformation. Taking the segments of Fig. 11 in all rhombs of a Penrose tiling one gets something that is topologically equivalent to the skeleton of the tiling (that was defined in Sect. 4 by connecting mid-points of the edges).

The transition from Figs. 10 to 11 does not affect the topology of the grid. This is not completely trivial, since (in contrast to what what was described in Sect. 5) the operations take place partly *outside* the rhombs. So it has to be made sure that the corresponding operations in neighboring rhombs do not interfere.

In Sect. 6 the topological equivalence of pentagrid and central grid will be shown in quite a different way.

6. Algebraic Proof of the Topological Equivalence

6.1

The various grids to be considered are generalizations of the one described in Sect. 4. They are characterized by real numbers $\omega_{j,k}$, where j runs through the set $\{0, 1, 2, 3, 4\}$ and k through the set $\mathcal{Z}$ of all integers. The grid determined by these ω's is the superposition of $\Gamma_{\omega,0}, \ldots, \Gamma_{\omega,4}$, where each $\Gamma_{\omega,j}$ is is a grid of parallel lines in the complex plane. The line with index k in $\Gamma_{\omega,j}$ is the set of all complex numbers z for which $\mathrm{Re}(z\zeta^{-j}) = \omega_{j,k}$ (as before, $\zeta = e^{2\pi i/5}$).

So the pentagrid described in Sect. 4 with its five real parameters $\gamma_0, \ldots, \gamma_4$ satisfying $\gamma_0 + \ldots + \gamma_4 = 0$ is the case where $\omega_{j,k} = k - \gamma_j$.

6.2

In [1] it was indicated how the pentagrid defines a Penrose tiling. Singular grids cannot be used directly: they first have to get an infinitesimal deformation in order to admit proper dualization (see [1], Sect. 12).

A few words may be devoted to the meaning of the word "infinitesimal" here. If in a singular pentagrid a grid line, a vertical one, say, is moved over a small finite distance to the right, then triple intersection points may have been avoided in a big finite range only. For a given line in a pentagrid and a given large positive number R there exists a small positive ε such that within the circular disk given by $|z| < R$ all triple intersections are avoided by shifting the line over a distance δ with $0 < \delta < \varepsilon$. But outside the range R such shifts may cause other intersection points, and alter the topology. The topology of the grid obtained by an infinitesimal shift to the right can be defined as the limit of the topology obtained within $|z| < R$ by means of sufficiently small shifts of the grid line.

But of course, perturbation of a pentagrid is not just a perturbation of a single line, but of the set of the five γ's. I now introduce an index p that describes which infinitesimal perturbation has to be taken. It has ten possible values: $0, \ldots, 9$, but in the majority of the singular cases there are only two that have a different effect, and in the regular cases all ten have the same effect. In [1], Sect. 9, it was explained that the complex number $\xi = \sum_{j=0}^{4} \gamma_j \zeta^{2j}$ is an essential parameter for the Penrose tiling. It is related to the real numbers μ_j, to be used in this section, given by

$$\mu_j = \gamma_j - (\gamma_{j-1} + \gamma_{j+1})\tau^{-1},$$

for $j = 0, \ldots, 4$, where, as before, $\tau = 2\,\mathrm{Re}\,\zeta = \frac{1}{2}(-1+\sqrt{5})$, and j is taken mod 5 (so $\gamma_5 = \gamma_0$, etc.). The μ's can be derived from ξ by $\mathrm{Re}(\xi\zeta^{-2j}) = (1-\frac{1}{2}\tau)\mu_j$.

It was shown in [1], Sect. 11 that a pentagrid is singular if and only if for one of the j's there is an α of the form $(1-\zeta)(n_0 + n_1\zeta + n_2\zeta^2 + n_3\zeta^3 + n_4\zeta^4)$ with integers $n_0, \ldots, n_4$ such that $\mathrm{Re}((\xi - \alpha)\zeta^{-j}) = 0$. It is not hard to show that this condition is equivalent to the following one: for at least one value of j there is an integer k such that $(k - \mu_j)\tau$ is an integer.

Now consider infinitesimal perturbations of ξ, given by a positive infinitesimal $\mathrm{d}w$ and a perturbation parameter p (one of the values $0, \ldots, 9$). The perturbation of ξ is $d\xi = e^{p\pi i/5}dw$. And perturbations of the γ's that produce this $\mathrm{d}\xi$ can be taken as (cf. [1], formula (9.2)): $\mathrm{d}\gamma_j = (2/5)\,\mathrm{Re}(\zeta^{-2j}\mathrm{d}\xi)$.

The perturbations of the μ's turn out to be

$$\mathrm{d}\mu_j = (1 - \frac{1}{2}\tau)^{-1}\,\mathrm{Re}(e^{(p-4j)\pi i/5})\mathrm{d}w.$$

6.3

The lines of a pentagrid correspond to stacks of the Penrose tiling, and for each stack there is a central bar according to Sect. 5. These central bars can be evaluated in terms of the μ's. This will be explained in Sect. 6. The result is that the line with index k in the j-th subgrid of the pentagrid leads to a central bar given by $\mathrm{Re}(z\zeta^{-j}) = \beta_{j,k}$, where, at least in the case of a non-singular grid, $\beta_{j,k} = (1 + \frac{1}{2}\tau^{-1})(\kappa + \tau\lceil(\kappa - \mu_j)\tau\rceil - \frac{1}{2}\tau)$. ($\lceil x\rceil$ is the usual notation for the least integer $\geq x$). In singular cases the value of $\lceil(\kappa - \mu_j)\tau\rceil$ can be affected by perturbation of μ_j if $(\kappa - \mu_j)\tau$ is an integer. The effect can be described by means of the notations $\lceil x\rceil_+$ and $\lceil x\rceil_-$,intended as $\lceil x + \mathrm{d}w\rceil$ and $\lceil x - \mathrm{d}w\rceil$, respectively, where $\mathrm{d}w$ is a positive infinitesimal. This just means

$$\lceil x\rceil_+ = \lfloor x\rfloor + 1, \quad \lceil x\rceil_- = \lceil x\rceil$$

for all real values of x. With this notation the result for the central bar becomes

$$\beta_{j,k} = (1 + \frac{1}{2}\tau^{-1})(\kappa + \tau\lceil(\kappa - \mu_j)\tau\rceil_{\varphi(p,j)} - \frac{1}{2}\tau) \tag{2}$$

where $\varphi(p, j)$ stands for $+$ or $-$ according to whether $\mathrm{Re}(e^{(p-4j)\pi i/5})$ is < 0 or > 0.

In the non-singular cases the $\varphi(p, j)$ can be ignored, of course. But it should be noted that it is not so easy to see from the γ's whether a case is singular or not. If one does not want to bother one might just take the $\varphi(0, j)$ as a standard, but it is dangerous to omit the $\varphi(p, j)$ altogether. In the case $\gamma_0 = \cdots = \gamma_4 = 0$ with its ten different perturbations the formula $\beta_{j,k} = (1 + \frac{1}{2}\tau^{-1})(\kappa + \tau\lceil\kappa\tau\rceil - \frac{1}{2}\tau)$ would definitely *not* represent the central bars of a Penrose tiling.

6.4

Here are some details about the derivation of the expression (2) for the central bar of the stack corresponding to a given line of a pentagrid. For simplicity, non-singularity will be assumed. Moreover, the (unessential) restriction is made that $j = 0$, which makes it possible to talk in terms of left and right. Moreover, the letter j becomes available for other purposes.

So the grid line is (with some integer k) $\mathrm{Re}(z) = k - \gamma_0$. The meshes directly to the left and directly to the right of this line produce the vertices of the rhombs of the stack. According to [1], Sect. 5, the vertices are derived from the meshes as follows. Take any point z in the mesh, and form the integers $K_j(z) = \lceil \mathrm{Re}(z\zeta^{-j}) + \gamma_j \rceil$; then the vertex is $\sum_{j=0}^{4} K_j(z)\zeta^j$. With a real variable t the points of the grid line are represented as $(k - \gamma_0) + it$. So the vertices corresponding to the meshes directly to the left of the line are given by $M(t) = k + \sum_{j=1}^{4} \lceil u_j \rceil \zeta^j$, where $u_j = \mathrm{Re}(((k - \gamma_0) + it)\zeta^{-j}) + \gamma_j$. It will be shown that the real part of $M(t)$ takes only four different values if t varies. That real part is $k + (\lceil u_1 \rceil + \lceil u_4 \rceil)\,\mathrm{Re}(\zeta) + (\lceil u_2 \rceil + \lceil u_3 \rceil)\,\mathrm{Re}(\zeta^2)$. Obviously $\lceil u_1 \rceil + \lceil u_4 \rceil = \lceil u_1 + u_4 \rceil + q$, where q is either 0 or 1, and similarly $\lceil u_2 \rceil + \lceil u_3 \rceil = \lceil u_2 + u_3 \rceil + r$, with r either 0 or 1. With the abbreviation

$$A = \lceil (k - \gamma_0)\tau + \gamma_1 + \gamma_4 \rceil \tau/2 - \lceil -(k - \gamma_0)\tau^{-1} + \gamma_2 + \gamma_3 \rceil \tau^{-1}/2$$

this leads to

$$\mathrm{Re}(M(t)) = k + A + q\,\mathrm{Re}(\zeta) + r\,\mathrm{Re}(\zeta^2).$$

It gives the four different horizontal coordinates of the left end-points of the horizontal edges in the stack. The right end-points are obtained by adding 1. So all the vertices involved here are lying on eight vertical lines. The figure formed by those eight lines has the central bar as an axis of symmetry, and that is why the name “central” was chosen. Since $\mathrm{Re}(\zeta) + \mathrm{Re}(\zeta^2) = -\frac{1}{2}$ the average of the eight horizontal coordinates is $k + A + \frac{1}{4}$, whence the central bar can be represented as $\mathrm{Re}(z) = k + A + \frac{1}{4}$.

From the arrowing of the horizontal vertices in the stack it is easy to derive that this central bar cuts the doubly arrowed horizontal edges at a point $\frac{1}{4}$ from the end, exactly in accordance with Fig. 10.

These calculations did not yet use the condition that $\sum \gamma_j = 0$. Tilings without that condition were mentioned in [3] in connection with a riffle shuffle card trick, and it is actually the riffle shuffle arrangement in a stack that guarantees that there are only eight different horizontal coordinates.

But $\sum \gamma_j = 0$ gives a simplification:

$$-(k - \gamma_0)\tau^{-1} + \gamma_2 + \gamma_3 = \mu_0\tau - k\tau - k,$$

and since non-singularity is assumed, this cannot be an integer. If x is not an integer one has $-\lceil -x \rceil = \lceil x \rceil - 1$, which leads to the following formula for the central bar:

$$\mathrm{Re}(z) = (1 + \frac{1}{2}\tau^{-1})(k + \tau\lceil (k - \mu_0)\tau \rceil - \frac{1}{2}\tau).$$

For the singular cases the result (2) can now be derived by obvious limit operations.

6.5

The topological equivalence of pentagrid and central grid will now be established directly, on the basis of (2). Figure 12 presents an illustration. It looks reasonable to enlarge the pentagrid by a factor 5/2 when comparing

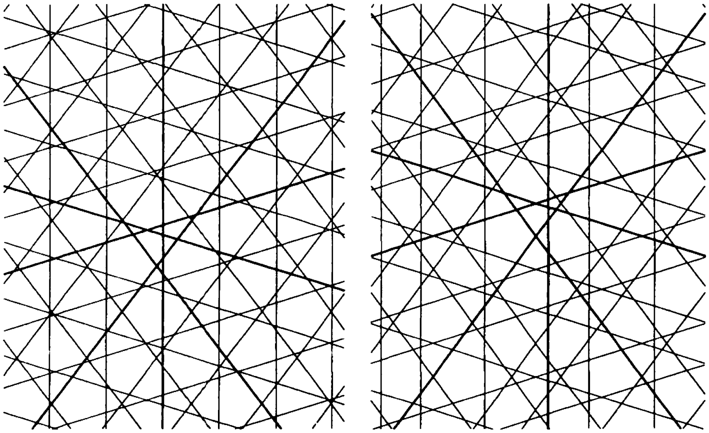


Fig. 12 On the *left* there is a piece of the pentagrid with parameters $\gamma_0 = 0.2$, $\gamma_1 = 0.4$, $\gamma_2 = 0.3$, $\gamma_3 = -0.8$, $\gamma_4 = -0.1$, on the *right* a corresponding piece of the central grid with the same parameters. The scale of the picture on the *left* is 5/2 times the one on the *right*. In order to facilitate the comparison of the topologies, the lines with $k_j = 0$ in the j-th subgrid are thicker than the others

it to the central grid (it is the same factor as in [1], Sect. 5, formula above (5.3)), but for the topology of the grids such factors make no difference at all.

The method is as follows. Consider two grids of the form of Sect. 6, given as $\mathrm{Re}(z\zeta^{-j}) = \alpha_{j,k}$ and $\mathrm{Re}(z\zeta^{-j}) = \beta_{j,k}$. It is assumed that for each j the $\alpha_{j,k}$ and the $\beta_{j,k}$ increase with k. And it is assumed that nowhere in the grids three lines pass through a point. In the topological correspondence the k-th line of the j-th subgrid of the α-grid will correspond to the k-th line of the j-th subgrid of the β-grid, for all j and k.

If p, q, r are straight lines forming a triangle, then they determine an orientation: seen from the inside of the triangle the circular order p, q, r is either clockwise or counter-clockwise. This gives the criterion for the topological equivalence in the grid: for any three lines in the α-grid the orientation has to be the same as for the corresponding three lines in the β-grid.

There are two kinds of triangles here. One is of the kind formed with values j, $j-1$, $j+1$, the other one with j, $j+2$, $j+3$ (j taken mod 5). The orientation is a simple matter of determinants. The result is as follows. The topological equivalence of the two grids is guaranteed if for all j and for all integers p, q, r the combination $a_{j+1,p} + a_{j-1,q} - \tau\alpha_{j,r}$ has the same sign as $\beta_{j+1,p} + \beta_{j-1,q} - \tau\beta_{j,r}$, and $\alpha_{j+2,p} + \alpha_{j-2,q} + \tau^{-1}\alpha_{j,r}$ has the same sign as $\beta_{j+2,p} + \beta_{j-2,q} + \tau^{-1}\beta j, r$.

Since no three lines pass through a point none of these expressions is zero.

It is this very explicit condition that has to be verified for the regular pentagrid and the corresponding central grid. It will be established in

Sects. 6–6 as a theorem on inequalities that can be understood independently of the previous sections. In order to get rid of the factor $(1 + \frac{1}{2}\tau^{-1})$ those sections work with θ instead of β, where $\beta = (1 + \frac{1}{2}\tau^{-1})\theta$. And for simplicity, Sects. 6–6 restrict themselves to $j = 0$; the other cases are completely similar.

6.6

If the conditions mentioned in Sect. 6 are satisfied for α and β then they are obviously also satisfied for α and $\omega(\lambda)$, where $0 \leq \lambda \leq 1$, and $\omega(\lambda)$ is obtained by linear interpolation:

$$\omega(\lambda)_{j,k} = (1 - \lambda)\alpha_{j,k} + \lambda\beta_{j,k}.$$

This means that the the α-grid can be deformed continuously into the β-grid without ever changing the topology. It is even possible to push the λ beyond 1. The lower bound $\frac{1}{2}\tau$ in Theorem 2 (Sect. 6) has the effect that the topology of the $\omega(\lambda)$-grid remains the same over the interval $0 \leq \lambda < (1 - \frac{1}{2}(1 + \frac{1}{2}\tau^{-1})\tau)^{-1}$.

Section 4 discussed infinitesimal perturbations for the interpretation of the dual of a singular pentagrid. The same thing can now be achieved with the $\omega(\lambda)$, with small positive but not infinitesimal values of λ.

6.7

The Sects. 6–6 do not make use of anything said in the previous sections.

Let $\gamma_0, \ldots, \gamma_4$ be real numbers with $\gamma_0 + \ldots + \gamma_4 = 0$. For $j = 0, \ldots, 4$, $k_j \in \mathcal{Z}$, the real numbers $\alpha_{j,k}$ and $\theta_{j,k}$ are defined by

$$\alpha_{j,k} = k - \gamma_j, \quad \theta_{j,k} = k + \tau\lceil(k - \mu_j)\tau\rceil - \frac{1}{2}\tau,$$

where $\mu_j = \gamma_j - (\gamma_{j-1} + \gamma_{j+1})\tau^{-1}$, $\gamma_5 = \gamma_0$ and $\tau = \frac{1}{2}(-1 + \sqrt{5})$.

Let $k_0, \ldots, k_4$ be integers, and abbreviate

$$H = \alpha_{1,k_1} + \alpha_{4,k_4} - \tau\alpha_{0,k_0}, \qquad L = \theta_{1,k_1} + \theta_{4,k_4} - \tau\theta_{0,k_0},$$

$$K = \alpha_{2,k_2} + \alpha_{3,k_3} + \tau^{-1}\alpha_{0,k_0}, \quad M = \theta_{2,k_2} + \theta_{3,k_3} + \tau^{-1}\theta_{0,k_0}.$$

With these abbreviations the following theorem can be proved:

Theorem 2. *If $H \neq 0$ then $L/H \geq \frac{1}{2}\tau$. If $K \neq 0$ then $M/K \geq \frac{1}{2}\tau$.*

6.8

Here is the proof of the first part of the theorem. H and L can be expressed like this:

$$H = k_1 + k_4 - \tau k_0 + \tau\mu_0,$$
$$L = k_1 + k_4 - \tau k_0 + \tau\lceil(k_1-\mu_1)\tau\rceil + \tau\lceil(k_4-\mu_4)\tau\rceil - \tau^2\lceil(k_0-\mu_0)\tau\rceil - \tau + \tfrac{1}{2}\tau^2.$$

If x and y are real numbers then $\lceil x\rceil + \lceil y\rceil - \lceil x+y\rceil$ is either 0 or 1. So

$$\lceil(k_1-\mu_1)\tau\rceil + \lceil(k_4-\mu_4)\tau\rceil = \lceil(k_1+k_4)\tau - (\mu_1+\mu_4)\tau\rceil + p$$

with $p = 0$ or $p = 1$. Moreover

$$(k_0-\mu_0)\tau = k_1 + k_4 - H,$$

and since $(\mu_1+\mu_4)\tau = -\mu_0$,

$$(k_1+k_4)\tau - (\mu_1+\mu_4)\tau = H\tau^{-1} + k_0 - k_1 - k_4.$$

So there is a simple expression for L in terms of H and p:

$$L = \tau\lceil H\tau^{-1}\rceil - \tau^2\lceil -H\rceil - \tau + \frac{1}{2}\tau^2 + p\tau.$$

Now first assume $H > 0$. Use $p \geq 0$, $-\lceil -H\rceil \geq 0$, and note that if $0 < c < 1$ then $\lceil x\rceil \geq 1 - c + cx$ for all $x > 0$. The special case $x = H\tau^{-1}$, $c = \frac{1}{2}\tau$ leads to $L/H \geq \frac{1}{2}\tau$.

Next assume $H < 0$. Use $p \leq 1$, $\lceil -H\rceil \geq 1$, and note that if $0 < c < 1$ then $\lceil x\rceil \leq cx + c$ for all $x < 0$. With $x = H\tau^{-1}$, $c = \frac{1}{2}\tau$ it follows that $L/H \geq \frac{1}{2}\tau$.

6.9

The proof of the second part of the theorem is similar. K and M can be expressed as follows: $K = k_2 + k_3 + k_0\tau^{-1} - \tau\mu_0$ and

$$M = k_2 + k_3 + k_0\tau^{-1} + \tau\lceil(k_2-\mu_2)\tau\rceil + \tau\lceil(k_3-\mu_3)\tau\rceil + \lceil(k_0-\mu_0)\tau\rceil - \tau - \frac{1}{2}.$$

Note that

$$\lceil(k_2-\mu_2)\tau\rceil + \lceil(k_3-\mu_3)\tau\rceil = \lceil(k_2+k_3)\tau - (\mu_2+\mu_3)\tau\rceil + q$$

where q is either 0 or 1. Moreover $(k_0-\mu_0)\tau = K - k_0 - k_2 - k_3$, and since $\mu_2 + \mu_3 = \mu_0\tau$,

$$(k_2+k_3)\tau - (\mu_2+\mu_3)\tau = K\tau - k_0.$$

So M can be expressed in terms of K and q:

$$M = \tau\lceil K\tau\rceil + \lceil K\rceil + \tau q - \tau - \frac{1}{2}.$$

If $K > 0$ it can be used that $\lceil K\rceil \geq 1$, whence $M \geq \tau\lceil K\tau\rceil - \tau + \frac{1}{2}$. If $0 < c < 1$ then $\lceil x\rceil \geq 1 - c + cx$ for all $x > 0$. With $x = K\tau$, $c = \frac{1}{2}\tau^{-1}$ this leads to $M/K \geq \frac{1}{2}\tau$. If $K < 0$ one can use the inequalities $q \geq 0$, $\lceil K\rceil \leq 0$ and $\lceil x\rceil \leq cx + c$ ($x < 0$, $0 < c < 1$) with $x = K\tau$, $c = \frac{1}{2}\tau^{-1}$. This again leads to $M/K \geq \frac{1}{2}\tau$, and that finishes the proof of Theorem 2.

References

1. N. G. de Bruijn, Algebraic theory of Penrose's non-periodic tilings of the plane. Kon. Nederl. Akad. Wetensch. Proc. Ser. A 84 (= Indagationes Mathematicae 43), 38–52 and 53–66 (1981). Reprinted in: P. J. Steinhardt and Stellan Ostlund: The Physics of Quasicrystals, World Scientific Publ., Singapore, New Jersey, Hong Kong.
2. N.G. de Bruijn, Dualization of multigrids. In: Proceedings of the International Workshop Aperiodic Crystals, Les Houches 1986. Journal de Physique, Vol. 47, Colloque C3, supplement to nr. 7, July 1986, pp. 9–18.
3. N. G. de Bruijn, A riffle shuffle card trick and its relation to quasicrystal theory. Nieuw Archief Wiskunde (4) 5 (1987) 285–301.
4. N. G. de Bruijn, Symmetry and quasisymmetry. In: Symmetrie in Geistes- und Naturwissenschaft. Herausg. R. Wille. Springer Verlag 1988, pp. 215–233.
5. N. G. de Bruijn, Updown generation of Penrose tilings, Indagationes Mathematicae, N.S., 1, pp. 201–219 (1990).
6. Martin Gardner, Mathematical games. Extraordinary nonperiodic tiling that enriches the theory of tiles. Scientific American 236 (1) 110–121 (Jan. 1977).
7. Branko Grünbaum and G. C. Shephard. Tilings and patterns. New York, W.H. Freeman and Co. 1986.
8. R. Penrose. Pentaplexity. Mathematical Intelligencer vol 2 (1) pp. 32–37 (1979).
9. J. E. S. Socolar and P. J. Steinhardt. Quasicrystals. II. Unit cell configurations. Physical Rev. B Vol. 34 (1986), 617–647. Reprinted in: P.J. Steinhardt and Stellan Ostlund: The Physics of Quasicrystals, World Scientific Publ., Singapore, New Jersey, Hong Kong.

Distances in Convex Polygons

Peter Fishburn

P. Fishburn (⊠)
Lucent Technologies Bell Laboratories, Murray Hill, NJ 07974, USA

Summary. One of Paul Erdős's many continuing interests is distances between points in finite sets. We focus here on conjectures and results on intervertex distances in convex polygons in the Euclidean plane. Two conjectures are highlighted. Let $t(x)$ be the number of different distances from vertex x to the other vertices of a convex polygon C, let $T(C) = \Sigma t(x)$, and take $T_n = \min\{T(C) : C \text{ has } n \text{ vertices}\}$. The first conjecture is $T_n = \binom{n}{2}$. The second says that if $T(C) = \binom{n}{2}$ for a convex n-gon, then the n-gon is regular if n is odd, and is what we refer to as bi-regular if n is even. The conjectures are confirmed for small n.

1. Introduction

Let $n_2 = \lfloor n/2 \rfloor$, the integer part of $n/2$ for $n \geq 3$. We begin with three conjectures about every convex n-gon in $\mathbb{R}^2$.

C1. Its vertices determine at least n_2 different distances.
C2. Some vertex has at least n_2 distances to the other vertices.
C3. Each of at least n_2 vertices has at least n_2 distances to the other vertices.

C1 was stated by Erdős in 1946 [3] and settled affirmatively by Altman in 1963 [1]. C2, another old conjecture of Erdős, is open. C3 was suggested by recent work with Erdős. It might be false, but we have no evidence of this.

My purpose here is to explore two conjectures related to C1–C3. We preface them with results for C1 and C2. Throughout, a convex polygon is unique up to similarity transformations (translation, uniform rescaling, rotation around a point, reflection around a line). R_n denotes the regular n-gon, and $R_{n+1} - 1$ is the n-gon whose vertices are n of the vertices of R_{n+1}. With $d(x, y)$ the distance between x and y, a *length-k run* in a convex n-gon is a sequence $x_0, x_1, \ldots, x_k$ of successively adjacent vertices for which

$$d(x_0, x_1) < d(x_0, x_2) < \cdots < d(x_0, x_k).$$

Theorem 1. *C1 is true. Suppose convex n-gon C has exactly n_2 different intervertex distances. If n is odd then $C = R_n$. If n is even and $n \geq 8$ then $C \in \{R_n, R_{n+1} - 1\}$.*

Altman [1] proves all but the final sentence, which is proved in [8]. At $n = 6$, a third hexagon besides R_6 and $R_7 - 1$ has exactly three intervertex

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_30,

distances (A_6, defined below); at $n = 4$, two other quadrilaterals besides R_4 and $R_5 - 1$ have exactly two intervertex distances.

The next theorem, from [6], gives the best result known toward C2. It implies that some vertex has at least $n/3$ different distances to the others. It is not known if there is $\lambda > 1$ such that every convex n-gon ($n \geq 4$) has a vertex with at least $\lfloor \lambda n/3 \rfloor$ distances to the other vertices.

Theorem 2. *For every $n \geq 4$, every convex n-gon has a run of length $\lfloor (n+3)/3 \rfloor$, and some convex n-gons have no run of length $\lfloor (n+3)/3 \rfloor + 1$.*

The proof uses the following observation in Moser [12].

Lemma 1 (**Moser's**). *Every convex n-gon has a circumscribed circle with a sub-semicircular arc ending in vertices such that the region enclosed by this arc and the chord between its endpoints contains at least $\lfloor (n-1)/3 \rfloor$ other vertices. Beginning from either endpoint, these vertices produce a run of length at least $\lfloor (n+2)/3 \rfloor$.*

Figure 1 illustrates Moser's lemma and some special polygons, including the bi-regular polygons A_4, A_6 and A_8.

Convex $2k$-gon A_{2k} is called *bi-regular* because it is formed from two regular polygons. A_4 is composed of two equilateral triangles with a common side. To form A_{2k} for $k \geq 3$, begin with a fixed copy R_k^0 of R_k with maximum intervertex distance 1. On each ray from the center of R_k^0 that bisects a side, add a vertex that has maximum distance 1 to those of R_k^0. The vertices of A_{2k} are the k of R_k^0 and the added k on the perpendicular bisectors of the sides of R_k^0.

Let $t(x)$ be the number of distances from vertex x to the other vertices of a convex n-gon C. The t-sequence of C, which we view cyclically, is $(t(x_1), t(x_2), \ldots, t(x_n))$ with $x_1, x_2, \ldots, x_n$ the vertices in clockwise or counterclockwise succession. Let $T(C) = t(x_1) + t(x_2) + \cdots + t(x_n)$. We focus on

$$T_n = \min\{T(C) : C \text{ is a convex } n\text{-gon}\}.$$

For odd n, $T(R_n) = n[(n-1)/2] = \binom{n}{2}$ with t-sequence $((n-1)/2, \ldots, (n-1)/2)$. For even n,

R_n has t-sequence $(n/2, n/2, \ldots, n/2)$;

A_n has t-sequence $(n/2 - 1, n/2, n/2 - 1, n/2, \ldots, n/2 - 1, n/2)$,

so

$$T(R_n) = n^2/2 \text{ and } T(A_n) = \binom{n}{2} < n^2/2.$$

Since we have no intimations of smaller values for $T(C)$, we consider two conjectures. The first was introduced in [4].

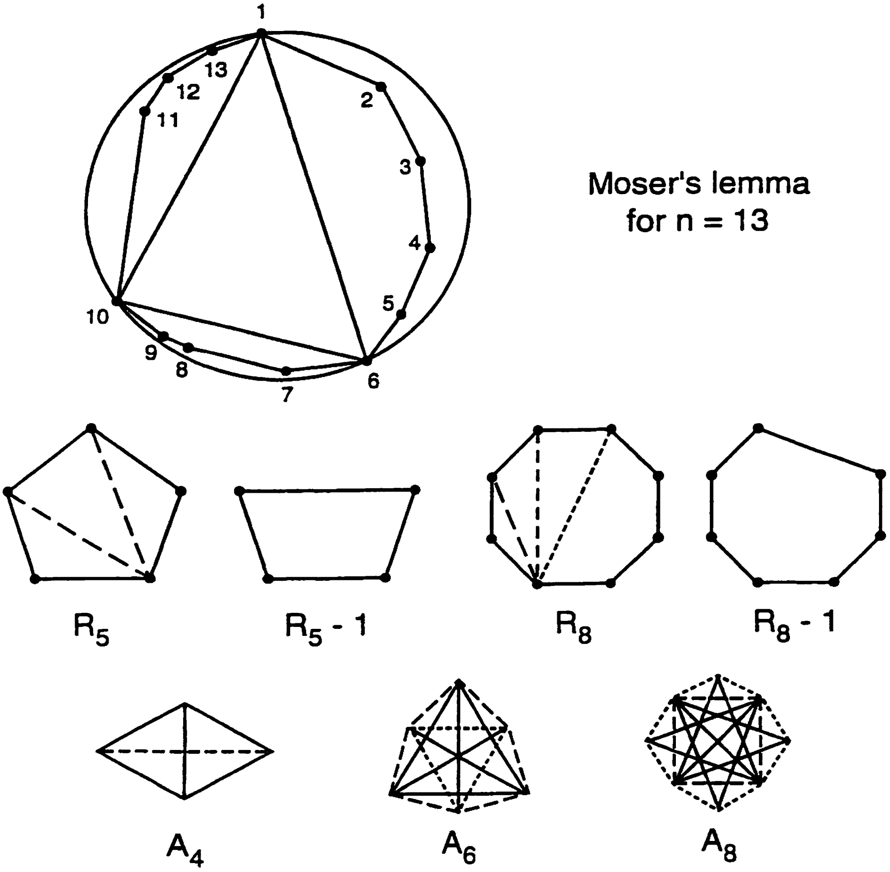


Fig. 1

C4. $T_n = \binom{n}{2}$ for all $n \geq 3$.

C5. For all $n \geq 3$, if $T(C) = \binom{n}{2}$ for a convex n-gon C, then $C = R_n$ if n is odd, and $C = A_n$ if n is even.

There is no apparent implication between these conjectures. If C4 is true, polygons other than R_n and A_n might realize T_n. If C5 holds, there might be other polygons that have $T(C) < \binom{n}{2}$. Because C4 implies that the minimum average $t(x)$ is $(n-1)/2$, it strengthens C2 but is apparently independent of C3. And C5 does not obviously imply either C2 or C3.

Suppose the preceding conjectures are all true. When n is odd, R_n is the unique realizer for T_n and, by Theorem 1, for the minimum n_2 of C1. But it clearly does not minimize the number of vertices with $t(x) \geq n_2$ in regard to C3 for $n \geq 5$. When n is even and $n \geq 8$, the realizers R_n and $R_{n+1} - 1$ for n_2 in C1 differ from the unique realizer A_n for T_n. Moreover, since A_n has exactly $n/2$ vertices with $t(x) \geq n/2$, it gives the tightest possible realization of C3.

Evidence for C4 and C5 is provided in the next two sections. Section 2 sketches proofs which show that C4 is true for $n \leq 8$, and Sect. 3 outlines proofs which verify C5 for $n \leq 7$. Although some of the proofs are case-intensive, it is hoped that their ideas will encourage refinements or extensions that will settle the conjectures fully.

Section 4 offers further remarks on C2–C5, then discusses a set of distance problems for convex polygons motivated by the conjecture of Erdős and Moser which says that the maximum number of pairs of vertices $\{x, y\}$ in a convex n-gon for which $d(x, y) = 1$ is bounded above by cn for some constant c.

2. Evidence for C4

Theorem 3. C4 is true for $n \leq 8$.

Since $t(x) \geq 1$ for all x, $T_3 \geq 3$, and $T_3 = 3$ is uniquely realized by R_3. When $t(x) = 1$ for some x, the other $n - 1$ vertices lie on a subsemicircular arc with center at x. Equal spacing along this arc yields the following lemma.

Lemma 2. *Suppose $t(x) = 1$. If y is adjacent to x for convex n-gon C then $t(y) \geq n - 2$. Moreover,*

$$T(C) \geq (3n^2 - 8n + 9)/4 \text{ if } n \text{ is odd};$$

$$T(C) \geq (3n^2 - 8n + 8)/4 \text{ if } n \text{ is even}.$$

It follows that the minimizing t-sequence for $n = 4$ is $(1, 2, 1, 2)$. Thus $T_4 = 6$, which is uniquely realized by A_4.

For $n \geq 5$, Lemma 2 implies that $T(C) > \binom{n}{2}$ when $t(x) = 1$ for some vertex, so we assume henceforth that $\min t(x) \geq 2$. The minimizing t-sequence at $n = 5$ is then $(2, 2, 2, 2, 2)$, with $T_5 = 10$ realized by R_5. The next section shows that R_5 is the only pentagon for which $T(C) = 10$.

The next lemma will be used to verify $T_6 = 15$.

Lemma 3. *If x and y are adjacent vertices of convex n-gon C and $t(x) = t(y) = 2$, then $n \leq 5$.*

Proof. Let the hypotheses of the lemma hold. Assume without loss of generality that $x = (0, 0)$, $y = (1, 0)$, and the other vertices of C lie above the abscissa. We have $d(x, y) = 1$. Let d_x and d_y be the second intervertex distances for x and y respectively. Since the other vertices lie at intersection points of the x-centered circles of radii 1 and d_x and the y-centered circles of radii 1 and d_y, $n \leq 6$. To obtain $n = 6$, all four intersections above the abscissa must occur, so assume this. Then $\max\{d_x, d_y\} < 2$.

It is easily seen that convexity is violated at $n = 6$ if either $d_x = d_y$ or $\min\{d_x, d_y\} < 1$. The latter is illustrated in Fig. 2a. Assume henceforth that $1 < d_x < d_y < 2$ as pictured in Fig. 2b. Point q has $d(x, q) = d(y, q) = 1$. The other three intersection points are p, z and v.

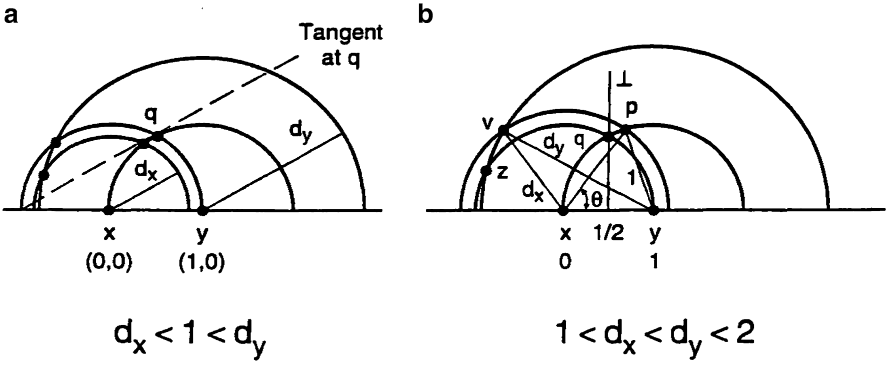


Fig. 2 $t(x) = t(y) = 2$

We prove that convexity is violated at $n = 6$ by showing that q lies below line segment vp. With d_x fixed, the best chance for convexity occurs when d_y is near 2 [z near $(-1, 0)$] since this moves v down the d_x circle. We prove that vp intersects the perpendicular bisector $\perp$ of xy above q when $d_y = 2$.

Fix $d_y = 2$ and let $d = d_x$. Let s be p's horizontal component. Since $\cos\theta = (1 + d^2 - 1)/2d = d/2$ and $\cos\theta = s/d$, $s = d^2/2$. Then p's vertical component is $[d^2 - (d^2/2)^2]^{1/2}$. Similar computations for v give

$$p = (d^2/2, [d^2 - (d^2/2)^2]^{1/2})$$
$$v = (-(3 - d^2)/2,\ [d^2 - \{(3 - d^2)/2\}^2]^{1/2}).$$

Line segment $\alpha p + (1 - \alpha)v$ has first component $\alpha(d^2/2) - (1 - \alpha)(3 - d^2)/2$, which equals $1/2$ for intersection with $\perp$ when $\alpha = (4 - d^2)/3$. Let h be the height at which $\alpha p + (1 - \alpha)v$ intersects $\perp$. Then

$$h = \{d(4 - d^2)^{3/2} + (d^2 - 1)[4d^2 - (3 - d^2)^2]^{1/2}\}/6.$$

Since $q = (1/2, \sqrt{3}/2)$, our claim that vp intersects 1 above q is

$$(4 - d^2)^{3/2}d + (d^2 - 1)[4d^2 - (3 - d^2)^2]^{1/2} > 3\sqrt{3}.$$

Let $u = d^2 - 1$, $0 < u < 3$. Then, after squaring both sides of the preceding inequality, it reduces to

$$(3 - u)^{3/2}\sqrt{8 + 7u - u^2} > (9 - 8u + u^2)\sqrt{u}.$$

This holds when $u \geq 2$ since its right side is then negative. Suppose $u < 2$. We square sides and cancel identical terms to get $216 - 27u > 81u$, i.e., $u < 2$, so the inequality holds for all $0 < u < 3$. □

It follows from $\min t(x) \geq 2$ and Lemma 3 that the smallest possible $T(C)$ at $n = 6$ occurs uniquely for the t-sequence $(2, 3, 2, 3, 2, 3)$. Since this is realized by A_6, $T_6 = 15$.

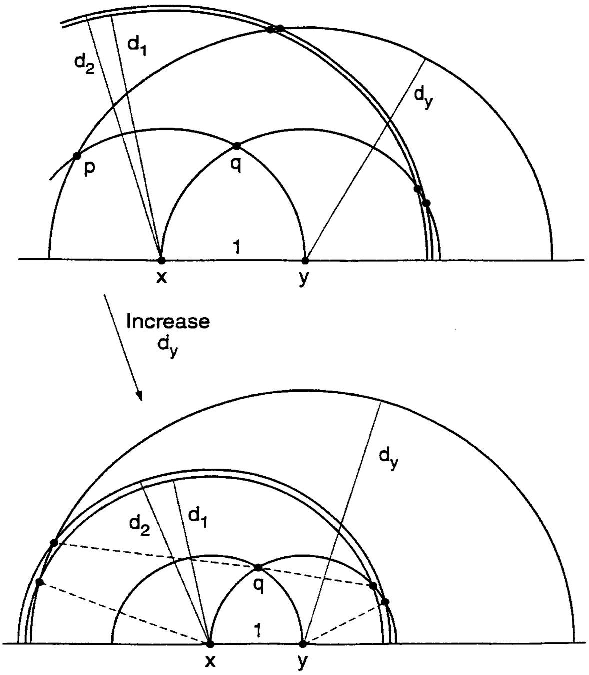


Fig. 3 $t(x) = 3$, $t(y) = 2$

Consider $n = 7$. Since R_7 has t-sequence $(3, 3, \ldots, 3)$, $T_7 \leq 21$. If adjacent vertices never have t-counts of 2 and 3 but $t(x) = 2$ for some x, then $T \geq 22$, as with $(2, 4, 2, 4, 3, 3, 4)$. Hence $T \leq 20$ is conceivable only if there are adjacent vertices with $t(x) = 3$ and $t(y) = 2$. These t-counts are possible at $n = 7$, but only under special circumstances.

Lemma 4. *Suppose $t(x) = 3$ and $t(y) = 2$ for adjacent vertices of C with $d(x, y) = 1$. Let $d_1 < d_2$ be the second and third distances for x; let d_y be the second distance for y. Then $n \leq 7$, and $n \leq 6$ if* $\min\{d_1, d_y\} < 1$.

Proof (outline). Lemma 3 requires $n \leq 5$ if we omit the third x distance. Since $t(y) = 2$, at most two more vertices are added by x's third distance, so $n \leq 7$. If $d_y < 1$, straightforward geometric arguments yield a convexity violation if $n > 6$. If $d_y > 1$ but $d_1 < 1$, convexity also forces $n \leq 6$. When $1 < d_2$ for this case, we use the argument in the latter part of the proof of Lemma 3. □

Figure 3 illustrates $1 < \min\{d_1, d_y\}$ for Lemma 4. The top diagram has all possible circle intersections ($n \leq 8$). As drawn, q must be removed for convexity at $n = 7$, but p could be removed instead if d_y were a bit smaller. By increasing d_y we lose intersection point p, but can still have a convex heptagon with the seven remaining points.

Modulo minor changes in d_1, d_2 and d_y, Fig. 3 shows the only ways we can have a convex heptagon when there are contiguous t-counts 2 and 3. To have $T \leq 20$, the five vertices besides x and y must have a t sum no greater than 15, or an average of $t \leq 3$ per vertex. It is easily checked that no other vertex has $t = 2$, and some clearly have $t \geq 4$, so in fact $T(C) \geq 22$ for these heptagons. We conclude that $T_7 = 21$, which is attainable only with t-sequence $(3, 3, \ldots, 3)$.

Finally, consider $n = 8$ where A_8 with t-sequence $(3, 4, 3, 4, 3, 4, 3, 4)$ gives $T_8 \leq 28$. By Lemmas 2 and 3, the minimum sum of adjacent t-counts for a convex octagon is 6, with $(t(x), t(y))$ either $(4, 2)$ or $(3, 3)$. In the first case, addition of a d_3 circle centered at x on Fig. 3 will give a $T(C)$ for $n = 8$ at least in the low 30's. A similar analysis for $(3, 3)$ yields a similar result. I omit details. It follows that $T_8 = 28$.

At the present time I know of no convex octagon for which $28 < T(C) < 32$. We have $T(R_8) = 32$, but R_8 is not the only octagon with $T = 32$. An example is O_1 on Fig. 6 in [9]. Its t-sequence is $(2, 5, 4, 5, 2, 5, 4, 5)$.

3. Evidence for C5

Theorem 4. C5 is true for $n \leq 7$.

The preceding section verifies this for $n \leq 4$ and shows for $n \in \{5, 6, 7\}$ that the unique t-sequences that have $T(C) = \binom{n}{2}$ are

$$\begin{aligned} &(2, 2, 2, 2, 2) && \text{for } n = 5 \\ &(2, 3, 2, 3, 2, 3) && \text{for } n = 6 \\ &(3, 3, 3, 3, 3, 3, 3) && \text{for } n = 7. \end{aligned}$$

These t-sequences are assumed throughout this section. We also let m denote the number of different intervertex distances in a convex polygon.

Suppose $n = 5$. If $m = 2$, we have R_5 by Theorem 1. If $m = 3$, Theorem 2 and Fig. 2 in [9] show that $t(x) \geq 3$ for some vertex. Suppose $m = 4$. Denote the four distances by d_1 through d_4 and label each vertex with its two distances. Suppose without loss of generality that one is labeled $d_1 d_2$ and another $d_1 d_3$. Then each of the other three vertices has (d_1 or d_2) and (d_1 or d_3), and to obtain a fourth distance one of these has label $d_1 d_4$, so each of the remaining two also has (d_1 or d_4). Then, because $t(x) = 2$ for all vertices, all five have label d_1, and this yields the contradiction that one of the other labels is used on only one vertex. A similar contradiction holds when $m > 4$, so R_5 is the only pentagon with t-sequence $(2, 2, 2, 2, 2)$.

Let $n = 6$. Label the vertices 1 through 6 in succession clockwise and let jk denote $d(j, k)$. By Theorem 2 there is a run of length 3. Assume for definiteness that $12 = a < 13 = b < 14 = c$: see Fig. 4a, where circled numbers give the t-sequence. By Moser's construction [12] there is

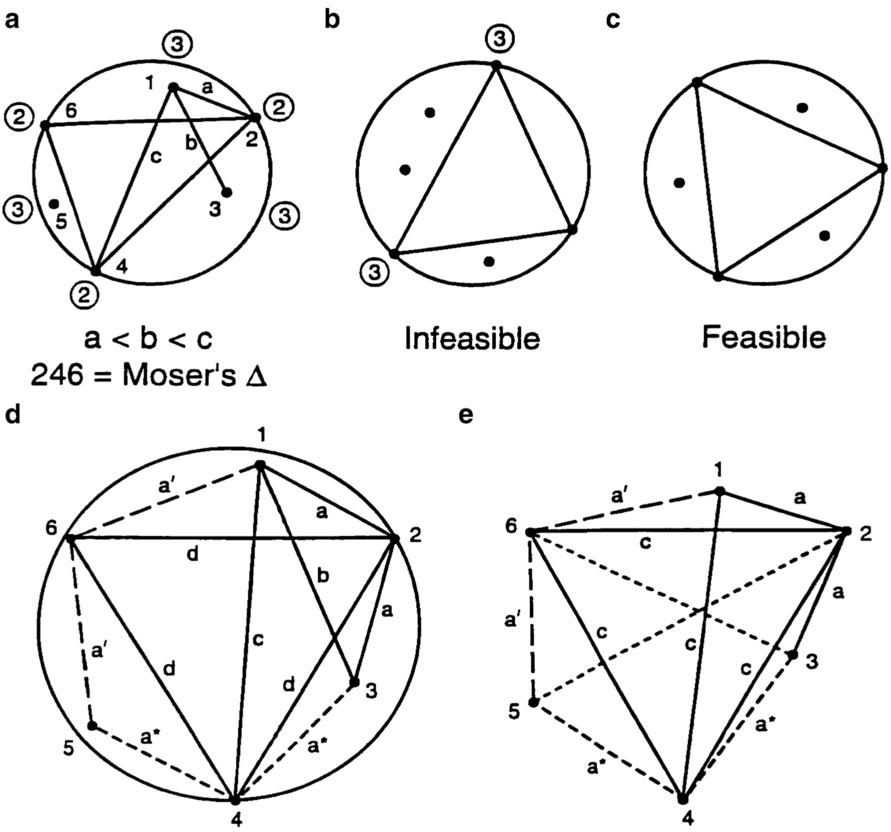


Fig. 4 $n = 6$, $(t(1), \ldots, t(6)) = (3, 2, 3, 2, 3, 2)$

a minimum-diameter circumscribed circle that contains three vertices that give a triangle with each interior angle no greater than $\pi/2$: see Figs. 4b, c. If one of the resulting subsemicircular sectors contains two of the other vertices, its end vertices both have $t \geq 3$, a contradiction to the t-sequence pattern (2,3,2,3,2,3). We therefore have the arrangement of Fig. 4c, where all three vertices of the noted triangle have either $t = 2$ or $t = 3$. If they have $t = 3$, we get a contradiction, for each of other three vertices has $t = 2$, which would force us back to Fig. 4b or to the conclusion that some vertex lies outside the circle. As a consequence, Moser's construction implies the arrangement shown in Fig. 4a.

Because the sectors outside the sides of Moser's triangle are subsemicircular, $21 < 26$ and $23 < 24$. Since $t(2) = 2$, we have $23 = 21 = a$ and $26 = 24$. By analogy, $43 = 45$, $42 = 46$, $65 = 61$ and $64 = 62$. Since $26 = 24 = 46$, Moser's triangle is equilateral: see Fig. 4d. Let $d = 26$. We have $43 = 45 < d = 42 = 46$, $c = 41$ and $t(4) = 2$. A simple geometric argument implies that $c = d$: if $c = 43$ with 1 as pictured, 3 would lie outside the circle. This brings us to Fig. 4e. For reasons similar to those just given for $c = d$, we have $25 = 63 = c$. Therefore triangles 124 and 326 are congruent, so the line through 1 and 3 is parallel to the line through 6 and 4, and as a consequence $a' = a^*$. Similarly, $a' = a$, so all sides of the hexagon have the same length. These and the six c diagonals define A_6.

My proof that R_7 is the only convex heptagon C with t-sequence $(3, 3, \ldots, 3)$ is unreasonably long. A shorter proof is needed. The main steps in mine are as follows:

1. Show that $C = R_7$ if all sides are equally long.
2. Prove that $d(x, z) > \max\{d(x, y), d(y, z)\}$ whenever x, y and z are consecutive vertices.

Label the points 1 through 7 clockwise in such a way that Moser's lemma for $n = 7$ gives 1 through 4 in a subsemicircular sector of a circumscribed circle. Set $12 = a < 13 = b < 14 = c$.

3. Prove that $17 = a$, $16 = b$ and $15 = c$.
4. Apply the result just proved in a sequence of steps (the first is $43 < 42 < 41 \Rightarrow 45 = 43$, $42 = 46$, $41 = 47 = c$) to conclude that $C = R_7$.

The main labor involves eliminating other possibilities in step 3.

4. Discussion

Conjecture C4, which claims that T_n equals $\binom{n}{2}$, evolved from work on C2, and in turn suggested conjectures C3 and C5. It has been featured here because its appealing form and global character may suggest approaches that prove it as well as its parent C2.

We noted earlier that, in view of A_n, C3 is formulated as tightly as possible for even n. This is not true for odd n. At least four vertices of a convex pentagon have $t \geq 2$, and at least five vertices of a convex heptagon have $t \geq 3$. A small challenge is to give a convincingly tight alternative to C3 for odd n.

Another set of conjectures for distances in convex polygons is based on multiplicity vectors. The *multiplicity vector* of n-gon C is $r(C) = (r_1(C), r_2(C), \ldots, r_m(C))$ where m is the number of different intervertex distances and $r_i(C)$ is the number of times the ith most-frequent distance occurs, ties resolved arbitrarily. Thus $r_1(C) \geq r_2(C) \geq \cdots \geq r_m(C) \geq 1$ and $\sum r_i(C) = \binom{n}{2}$.

Let $r_i(n) = \max\{r_i(C) : C \text{ is a convex } n\text{-gon}\}$. Erdős and Moser [7] conjecture that $r_1(n) < cn$ for some constant c. The best general bounds we are aware of are $2n - 7 \leq r_l(n) \leq \pi n(2\log_2 n - 1)$, due to Edelsbrunner and Hajnal [2] and Füredi [11] respectively.

It is known [5] that $r_2(n) = n$ for $5 \leq n \leq 8$ and that $r_2(25) > 25$. We do not know the smallest n at which the second most-frequent distance can exceed n, nor do we have a very good idea of the growth rate of $r_2(n)/n$.

It is not known if $r_3(n) > n$ for some n.

We conjecture [4, 5] that $\Sigma[r_i(C)]^2$ is maximized uniquely over all convex n-gons at $C = R_n$, except for $n \in \{4, 6, 8\}$. The nonregular maximizers for the exceptional cases are identified in [9].

Many years ago Danzer (see [1]) disproved an Erdős conjecture [3] by constructing a convex 9-gon in which each vertex has three others equidistant from it. Fishburn and Reeds [10] constructs a convex 20-gon in which each vertex has three others distance 1 from it. Erdős and Fishburn [4] conjecture that there is no convex n-gon in which every vertex has distance 1 to four other vertices. If true, then $r_1(n) \leq 3n - 6$.

References

1. E. Altman, On a problem of P. Erdős, Amer. Math. Monthly 70 (1963) 148–157.
2. H. Edelsbrunner and P. Hajnal, A lower bound on the number of unit distances between the vertices of a convex polygon, J. Combin. Theory A 56 (1991) 312–316.
3. P. Erdős, On sets of distances of n points, Amer. Math. Monthly 53 (1946) 248–250.
4. P. Erdős and P. Fishburn, Multiplicities of interpoint distances in finite planar sets, Discrete Appl. Math. (to appear).
5. P. Erdős and P. Fishburn, Intervertex distances in convex polygons, Discrete Appl. Math. (to appear).
6. P. Erdős and P. Fishburn, A postscript on distances in convex n-gons, Discrete Comput. Geom. 11 (1994) 111–117.
7. P. Erdős and L. Moser, Problem 11, Canad. Math. Bull. 2 (1959) 43.
8. P. Fishburn, Convex polygons with few intervertex distances, DIMACS report 92–18 (April 1992), AT&T Bell Laboratories, Murray Hill, NJ.
9. P. Fishburn, Convex polygons with few vertices, DIMACS report 92–17 (April 1992), AT&T Bell Laboratories, Murray Hill, NJ.
10. P. C. Fishburn and J. A. Reeds, Unit distances between vertices of a convex polygon, Comput. Geom.: Theory and Appls. 2 (1992) 81–91.
11. Z. Füredi, The maximum number of unit distances in a convex n-gon, J. Combin. Theory A 55 (1990) 316–320.
12. L. Moser, On the different distances determined by n points, Amer. Math. Monthly 59 (1952) 85–91.

Unexpected Applications of Polynomials in Combinatorics

Larry Guth

L. Guth (✉)
Department of Mathematics, MIT, Cambridge MA 02139, USA
e-mail: lguth@math.mit.edu

In the last 6 years, several combinatorics problems have been solved in an unexpected way using high degree polynomials. The most well-known of these problems is the distinct distance problem in the plane. In [Erdős46], Erdős asked what is the smallest number of distinct distances determined by n points in the plane. He noted that a square grid determines $\sim n(\log n)^{-1/2}$ distinct distances, and he conjectured that this is sharp up to constant factors. Recently, an estimate was proven that is sharp up to logarithmic factors.

Theorem 1 ([Guth–Katz11], building on [Elekes–Sharir10]). *For any n point set in the plane, the number of distinct distances is $\geq cn(\log n)^{-1}$.*

The main new thing in the proof is the use of high-degree polynomials. This new technique first appeared in Dvir's paper [Dvir09], which solved the finite field Nikodym and Kakeya problems. Experts had considered these problems very difficult, but the proof was essentially one page long. The method has had several other applications. The joints problem was resolved in [Guth–Katz10]. The argument was simplified and generalized in [KSS10] and [Quilodrán10], leading to another one page proof. A higher-dimensional generalization of the Szemerédi-Trotter theorem was proven in [Solymosi–Tao12]. And several fundamental theorems in incidence geometry were reproved in the paper [KMS12].

The new trick in these proofs can be summarized as follows. We want to understand some finite set S in a vector space. We consider a minimal degree (non-zero) polynomial that vanishes on the set S. Then we use this polynomial to study the problem. This strategy is somewhat surprising because the statements of the problems often involve only points and lines. The joints problem and the finite field Nikodym problem can be solved in a page each using high degree polynomials but seem very difficult to solve without polynomials. Why polynomials play such a crucial role in these problems is somewhat mysterious.

The point of this essay is to explain how these new methods work and to reflect on them philosophically. The main theme is the connection between combinatorics and algebra (polynomials).

Here is an outline of the essay.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_31,

We begin by giving two detailed examples of the polynomial method: the finite field Nikodym problem and the joints problem. This is the subject of Sect. 1: Examples of the polynomial method.

Once we've seen a couple examples of this method, we're going to work on understanding "where it comes from". In Sect. 2, we discuss where the method comes from historically. We discuss related arguments from other areas of mathematics. Polynomials are fundamental mathematical objects, and there are many different perspectives about them. Section 2 is called 'Perspectives on polynomials'. We will see perspectives about polynomials coming from number theory, coding theory, and differential geometry. Each of these perspectives helps to understand why polynomials are useful in these combinatorial problems.

In Sect. 3, we describe the new results in incidence geometry proven with polynomials, and we put them in perspective in the field. We recall the Szemerédi-Trotter theorem—a central result in the field—and discuss why the problem is difficult. We discuss one of the important methods in the field—the cutting method of [CEGSW90]. The Szemerédi-Trotter theorem involves lines in the plane. More generally, it is interesting to try to study k-dimensional objects in n-dimensional space. There are new challenges in higher dimensions. In particular, there is a new difficulty in dealing with objects of codimension > 1, such as lines in $\mathbb{R}^3$ or 2-planes in $\mathbb{R}^4$. We take some time to explain why this type of problem is hard to understand using previous methods. The distinct distance problem appears at first sight (and second and third. . .) as a problem about circles in the plane, but Elekes found a way to rephrase it as a problem about curves in three dimensions. In particular, we will meet two theorems about lines in $\mathbb{R}^3$ that are closely connected to the distinct distance problem and that illustrate the difficulties of incidence geometry in codimension > 1.

In Sect. 4, we explain how polynomials can be used to study incidence geometry. Section 4 is called 'Combinatorial structure and algebraic structure'. We will explain the main ideas in the proofs of the two theorems at the end of Sect. 3. More broadly, we will try to explain the mechanisms why a configuration with a lot of combinatorial structure is forced to have a special polynomial structure.

This essay is for a volume on the mathematics of Paul Erdős. Erdős's ideas influenced the work we describe in many ways. He posed the distinct distance problem in [Erdős46]. This paper was one of the first papers in incidence geometry, perhaps the first, and the problem has shaped many ideas in the field. I am a big admirer of hard problems that are simple to state. The most exciting—in my opinion—is a simply stated problem that is hard *for a new reason.* I think Erdős's distance problems are such problems. They helped create and guide a whole field of math. Mathematicians working in incidence geometry have made a great effort to clarify the nature of the difficulty of these problems, and then to find methods to deal with these difficulties. We describe here one chapter of this story.

1. Examples of the Polynomial Method

Because some of the arguments are so short, I think the best introduction to the polynomial method is to look at some proofs. We give detailed sketches of two proofs, and then we will step back and talk about them.

1.1 The Main Ingredients

There are two basic facts about polynomials which are the main ingredients in the arguments. If $\mathbb{F}$ is a field, let $\mathrm{Poly}_D(\mathbb{F}^n)$ be the space of polynomials over $\mathbb{F}$ with degree $\leq D$ and n variables. $\mathrm{Poly}_D(\mathbb{F}^n)$ is a vector space over $\mathbb{F}$.

Proposition 1.1. *The vector space* $\mathrm{Poly}_D(\mathbb{F}^n)$ *has dimension* $\binom{D+n}{n} \geq D^n/n!$.

Proof. A basis is given by the monomials $x_1^{D_1}, \ldots, x_n^{D_n}$ with $D_1 + \cdots + D_n \leq D$. By the 'stars and stripes' argument, the number of monomials is $\binom{D+n}{n}$.

As a corollary, we can estimate the degree of a polynomial that vanishes at prescribed points.

Corollary 1.2 (**Parameter counting**). *If* $S \subset \mathbb{F}^n$ *is a finite set, then there is a non-zero polynomial that vanishes on* S *with degree* $\leq n|S|^{1/n}$.

In rough terms, when we choose a polynomial in $\mathrm{Poly}_D(\mathbb{F}^n)$, we have $\binom{D+n}{n}$ parameters at our disposal. As long as $\binom{D+n}{n} > S$, we have enough parameters to arrange a non-zero polynomial that vanishes at every point of S. Linear algebra makes this heuristic rigorous.

Proof. We let $\mathrm{Fcn}(S, \mathbb{F})$ be the vector space of functions from S to $\mathbb{F}$. Restricting polynomials to the set S gives a (linear) map $\mathrm{Poly}_D(\mathbb{F}^n) \to \mathrm{Fcn}(S, \mathbb{F})$. There is a non-zero polynomial of degree $\leq D$ vanishing on S if and only if this linear map has a non-trivial kernel. As long as the dimension of the domain is bigger than the dimension of the range, the linear map does have a non-trivial kernel. The dimension of the domain is $\binom{D+n}{n}$, and the dimension of the range is $|S|$. By a brief computation, we can always choose $D \leq n|S|^{1/n}$ so that $\binom{D+n}{n} > |S|$.

The second main fact is that a non-zero polynomial in one variable cannot have more zeroes than its degree. A little more generally, we have the following.

Lemma 1.3 (**Vanishing lemma**). *If* L *is a line in a vector space and* P *is a polynomial of degree* $\leq D$*, and if* P *vanishes at* $D+1$ *points of* L*, then* P *vanishes on* L.

With little more than these tools, we will solve two hard problems about how lines intersect in vector spaces.

1.2 The Nikodym Problem in Finite Fields

Let $\mathbb{F}_q$ be a finite field with q elements. A set $N \subset \mathbb{F}_q^n$ is called a Nikodym set if for each point $x \in \mathbb{F}_q^n$, there is a line L so that $L \setminus \{x\} \subset N$. The question is, "how big does a Nikodym set need to be?" The paper [Dvir09] proves that a Nikodym set needs to have at least $c_n q^n$ elements—it needs to contain a definite fraction of the points in $\mathbb{F}_q^n$.

The history. The problem above is a finite-field adaptation for a problem in Euclidean geometry. A set $N \subset [0,1]^n$ is called a Nikodym set if for each $x \in [0,1]^n$, there is a line L so that N contains $L \cap [0,1]^n \setminus \{x\}$. The main question is "how big does a Nikodym set need to be?" Nikodym proved in the 20s that there are Nikodym sets of measure 0. The Nikodym conjecture says that the (Hausdorff or Minkowski) dimension of a Nikodym set is always n. (This roughly means that the δ neighborhood of a Nikodym set must contain nearly δ^{-n} δ-boxes.)

The Nikodym conjecture is a major open question in harmonic analysis. From our brief description, it's not at all clear why the problem is considered important. The Nikodym problem turns out to have connections to fundamental problems in Fourier analysis and PDE, including the restriction problem. The restriction problem was raised by Stein in the 1960s, and it has played a major role in Fourier analysis ever since then. The Nikodym conjecture is a close cousin of the more well-known Kakeya conjecture. The connection between these geometrical questions and problems in Fourier analysis and PDE is described in [Laba08] and [Tao01].

Mathematicians have put a lot of effort into the Nikodym and Kakeya problems but remain far from a complete solution. Because the problems seem difficult, analysts have begun working on a variety of cousins and model problems that may shed some light back on the original problems. In [Wolff99], Wolff proposed looking at the finite field analogues of these questions. Proving that the Minkowski dimension of a Nikodym set in $[0,1]^n$ is at least α is analogous to proving that a Nikodym set in $\mathbb{F}_q^n$ has $\gtrsim q^\alpha$ elements. In particular, Dvir's theorem is analogous to the Nikodym conjecture.

The proof of the finite field Nikodym conjecture. Let us assume that $N \subset \mathbb{F}_q^n$ is a Nikodym set with $< (10n)^{-n} q^n$ elements. Let P be a non-zero polynomial that vanishes on N with minimal degree.

1. By parameter counting, the degree of P is $\leq n|N|^{1/n} < q-1$.
2. By the vanishing lemma, $P(x) = 0$ at every point $x \in \mathbb{F}_q^n$. To see this, consider the line L given by the definition of the Nikodym set. We know that $x \in L$ and that $|L \cap N| \geq q-1$. So P vanishes on $q-1$ points of L, and since $\deg(P) < q-1$, P must vanish on all of L.
3. Once we know that P vanishes at every point (and that $\deg(P) < q-1$), it's not hard to show that all the coefficients of P are zero. In other words, P is the zero polynomial and we have a contradiction.

The Kakeya problem. The Nikodym problem is a close cousin of the more well-known Kakeya problem. A Kakeya set in $\mathbb{R}^n$ is a set containing a unit line segment in each direction. The Kakeya conjecture says that any Kakeya set in $\mathbb{R}^n$ must have dimension n. A Kakeya set in $\mathbb{F}_q^n$ is a set containing a line "in every direction". More precisely, a Kakeya set contains a translate of any line in $\mathbb{F}_q^n$. By a small modification of the argument above, [Dvir09] proves that any Kakeya set in $\mathbb{F}_q^n$ contains $\geq c(n)q^n$ points.

The influence. This proof shocked the harmonic analysis community. Analysts exchange stories about where they were when they heard about it. In [Erdős], Erdős told a story about how hard it is to judge the difficulty of a problem. This is the most dramatic example that I have personally encountered. The Nikodym and Kakeya and restriction problems are closely connected, notoriously difficult problems of analysis. I believe that the finite field version was considered roughly as difficult as the original version until it was proven in one page. (To be fair, I should also say that the finite field version was only open for about 10 years, and it was much less studied than the original problem.)

After the shock, people tried to adapt the new method to the original Nikodym and Kakeya problems in Euclidean space. So far, not much has been proven this way. It remains to be seen whether these methods will lead to progress in harmonic analysis. But the polynomial method has had a lot of influence in combinatorics. In this section we give one more example: the joints problem.

1.3 The Joints Problem

Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^n$. (The case $n = 3$ is a good case to keep in mind.) A joint is a point that lies in n lines of $\mathfrak{L}$ with linearly independent tangent directions. In other words, if the lines of $\mathfrak{L}$ thru x do not all lie in a hyperplane, then x is a joint. The problem is, how many joints can we make with L lines? The joints theorem says that the number of joints is $\lesssim L^{\frac{n}{n-1}}$. This number is sharp up to constant factors. For example, consider S hyperplanes in general position. Any $n-1$ hyperplanes intersect in a line, giving $L = \binom{S}{n-1}$ lines. Any n hyperplanes intersect in a point, and each of these points is a joint for our set of lines. So the number of joints is $\binom{S}{n} \sim L^{\frac{n}{n-1}}$.

The history. The joints problem was posed by Chazelle, Edelsbrunner, Guibas, Pollack, Seidel, Sharir, and Snoeyink in [CEGPSSS92]. They thought of the problem as a model problem for some difficult (still open) problems connected with computer vision. The original problem was in three dimensions. The best known bound before the polynomial method was that the number of joints is $\lesssim L^{1.62}$, [Feldman–SharirS05]. The paper [Guth–Katz10] proved the joints conjecture in three dimensions using the polynomial method.

The papers [KSS10] and [Quilodrán10] simplified the proof and generalized the result to any dimension.

The proof. We will prove the following main lemma: In any arrangement of lines in $\mathbb{R}^n$ with J joints, one of the lines contains $\lesssim J^{1/n}$ joints. The theorem follows from this main lemma by elementary counting. Given L lines and J joints, we remove the lines one at a time, using the main lemma to find an unpopular line to take out. Each time we remove a line, at most $J^{1/n}$ joints disappear. Therefore, $J \lesssim LJ^{1/n}$, and rearranging gives the theorem.

To prove the main lemma, we let P be a non-zero polynomial of minimal degree that vanishes on all the joints.

1. By parameter counting, the degree of P is $\lesssim J^{1/n}$.
2. If a line l contains $> \deg(P)$ joints, then P vanishes on the whole line. So it suffices to find a line $l \in L$ so that P is not identically zero on l.
3. If P vanishes on all of the lines of $\mathfrak{L}$ going thru a joint x, then ∇P vanishes at x. This is because $\nabla P(x)$ vanishes in the direction tangent to each line, and the span of the tangent directions is all of $\mathbb{R}^n$. So if P vanishes on all the lines in $\mathfrak{L}$, then each partial derivative $\partial_j P$ vanishes at each joint. We know that P is not constant, so one of these partial derivatives is non-zero, and it has degree $< \deg(P)$. This contradicts the definition of P as having minimal degree.

The influence. Starting from these two proofs, this little trick with high degree polynomials has become a major tool in incidence geometry. It has helped resolve several old problems and led to new proofs and perspectives about fundamental theorems. We will discuss the resulting ideas in Sects. 3–4.

1.4 Why Polynomials?

The proofs of the finite field Nikodym conjecture and the joints conjecture feel like the "right" proofs to me because they are so short and because the problems seemed very difficult before. But the proofs still seem a little mysterious to me. These are questions about points and lines, and yet it seems to be crucially important to use high degree polynomials to understand them. Is it really much harder to prove these results without using high degree polynomials? If so, why are polynomials so connected with these problems? I have been thinking about these questions and discussing them with people for several years. In this essay, I will share the observations that I know. I still wish I understood the questions better.

If we play around with questions about how lines intersect in $\mathbb{R}^3$, then we will come to an important example that involves a degree 2 algebraic surface. Let's try a few questions, beginning very naively. If $\mathfrak{L}$ is a set of lines, an intersection point is a point that lies in at least two lines.

Question 1. *Given L lines in $\mathbb{R}^3$, how many intersection points can there be?*

There can be at most $\binom{L}{2}$ intersection points, since any two lines intersect at most once. This upper bound is sharp. If all the lines lie in a plane, and if they lie in general position within the plane, then there will be $\binom{L}{2}$ distinct intersection points.

Perhaps a set of lines in space can have many intersection points only by clustering in a plane? We can probe this issue with the following question.

Question 2. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with ≤ 10 lines in any plane. How many intersection points can there be?*

Remarkably, there can still be $\sim L^2$ intersection points. Here we come to a crucial example involving a degree 2 algebraic surface. The surface is defined by the equation $z = xy$. This surface contains a lot of lines. For any number $b \in \mathbb{R}$, let H_b be the "horizontal" line $(x, b, bx), x \in \mathbb{R}$. For any number $a \in \mathbb{R}$, let V_a be the vertical line $(a, y, ay), y \in \mathbb{R}$. The horizontal lines and the vertical lines both lie in the surface $z = xy$. The horizontal line H_b and the vertical line V_a intersect at (a, b, ab). Let $\mathfrak{L}$ consist of $L/2$ horizontal lines and $L/2$ vertical lines. These lines intersect at $L^2/4$ distinct points, so $\mathfrak{L}$ has $\gtrsim L^2$ intersection points. The intersection of a plane with the surface $z = xy$ is a degree 2 algebraic curve, and so it contains at most two lines. Therefore, any plane contains ≤ 2 lines of $\mathfrak{L}$.

(This degree 2 surface is an example of a regulus. Reguli play an important role in the approach to the joints problem in [CEGPSSS92].)

Although Question 2 is about points and lines, the key examples do not just involve linear objects (lines, planes, etc.)—they also involve algebraic surfaces. This example gives one motivation why polynomials play a role in incidence problems about lines and points.

Let's follow our investigation a bit further. Lines may have many intersection points by clustering in a plane or in a degree 2 surface. Let's forbid both types of clustering.

Many Intersections Problem. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with ≤ 10 lines in any plane or degree 2 surface. How many intersection points can there be?*

This time, there will be far less than L^2 intersection points. The methods of [CEGPSSS92] show that the number of intersection points is $\lesssim L^{5/3}$. Using the polynomial method, the paper [Guth–Katz11] shows that the number of intersection points is $\lesssim L^{3/2}$. This estimate plays a role in the distinct distance estimate, and we will discuss it more later.

The many intersection problem is a significant open problem. The best current upper bound on the number of intersection points is $\sim L^{3/2}$. The examples I know all have $\lesssim L$ intersection points.

We can get a little perspective on this problem by naive parameter counting. The set of lines in $\mathbb{R}^3$ is a 4-dimensional manifold. If we choose L lines, we are choosing $4L$ parameters. In fact, there is no real loss in generality in assuming that each line is given by a graph $x = az+b, y = cz+d$. So we can specify L lines by $4L$ real parameters $a_1, \ldots, a_L, b_1, \ldots, b_L$, etc. The condition that line i intersects line j can be described by one algebraic equation in the parameters $a_i, b_i, c_i, d_i, a_j, b_j, c_j, d_j$. If we want our lines to have I intersection points, then we need to solve I equations in $4L$ variables. This naive parameter counting suggests that getting significantly more than $4L$ intersection points requires some kind of structure or coincidence. A bit more rigorously, I believe that if we replace the set of lines by a "generic" 4-parameter set of curves in $\mathbb{R}^3$, then no arrangement will have more than $4L$ intersection points.

Here is the philosophical question behind the many intersections problem. Morally, any arrangement with more than $4L$ intersection points exists only because of some special structure in the set of lines. Now what special structures could the set of lines have? There is some structure from linear algebra. There is also some structure from polynomials and algebraic geometry. Are there any other 'special structures' of the set of lines in $\mathbb{R}^3$?

In summary, some important examples in incidence geometry come from algebraic surfaces. It is interesting to ask whether *all the examples* come from algebraic surfaces. The polynomial method gives an approach to prove this type of statement in some cases. The main goal of Sect. 4 is to explain how this works.

2. Perspectives on Polynomials

In this section, we explore how this polynomial trick is connected to other parts of math. We will consider three other areas. The areas are diophantine problems in number theory, error-correcting codes in computer science, and surface area estimates in differential geometry. These areas give different perspectives on what makes polynomials special and useful functions.

2.1 There Are Lots of Polynomials: Thue's Work on Diophantine Approximation

Let's begin with a warmup problem. What is the smallest possible degree of a non-zero polynomial $P \in \mathbb{R}[x, y]$ that vanishes at the million points $(j, 2^j)$ where j is an integer in the range $[1, 10^6]$?

The first approach one might try is to write down polynomials that vanish at the prescribed points. For example, we might try $\prod_{j=1}^{10^6}(x - j)$ or $\prod_{j=1}^{10^6}(y - 2^j)$. Either of these options has degree 10^6. We might try to craft a more clever formula that improves the degree. I don't know how to write down any explicit formula with degree $\leq 10^5$. But the optimal degree is less than 1,500. This follows by parameter counting, as in Sect. 1.1. The dimension of $\text{Poly}_{1,498}(\mathbb{R}^2)$ is $\binom{1{,}500}{2} > 10^6$, and so there is a non-zero polynomial of degree $\leq$ 1,498 vanishing at all million points. This type of situation appeared in Thue's work on diophantine equations and approximation. Here is Thue's central result.

Theorem 2.1 (**Thue 1909**). *Suppose that β is an irrational algebraic number of degree $d > 2$. If p/q is any rational number, then*

$$|\beta - p/q| \geq c(\beta) q^{-\frac{d}{2}-1.01}.$$

As an immediate corollary, Thue proved that a huge class of diophantine equations in two variables have only finitely many integer solutions. For example, the following equations have only finitely many integer solutions.

1. $x^3 - 2y^3 = 6$.
2. $x^4 + 11xy^3 + 17y^4 = 29$.
3. $x^5 + 2x^2y^3 + 9y^5 = 9$.

Thue's corollary can be stated as follows:

Corollary 2.2. *If $P(x, y) \in \mathbb{Z}[x, y]$ is a homogeneous polynomial of degree $d \geq 3$ which is irreducible, and if n is an integer, then the equation $P(x, y) = n$ has only finitely many integer solutions $(x, y) \in \mathbb{Z}^2$.*

Thue's result was dramatically more general than any previous theorem about diophantine equations. To get a sense of how Thue's diophantine approximation result implies the corollary, consider equation 1 above. Dividing through by y^3 we get $(x/y)^3 - 2 = 6|y|^{-3}$. This formula shows that x/y is a very good rational approximation of $2^{1/3}$. With a little manipulation, it follows that $|2^{1/3} - (x/y)| \leq 100|y|^{-3}$. In contrast, Thue's theorem on diophantine approximation says that $|2^{1/3} - (x/y)| \geq c|y|^{-2.51}$. Comparing these inequalities, we see that $|y|$ is uniformly bounded, and then it follows that there are only finitely many solutions.

We will give a very partial sketch of Thue's proof, and we will see how it connects with our warmup question about polynomials.

Before Thue, the main theorem about diophantine approximation was Liouville's theorem.

Theorem 2.3 (Liouville 1840s). *If β is an algebraic number of degree $d > 1$, and p/q is any rational number, then*

$$|\beta - p/q| \geq c(\beta) q^{-d}.$$

The idea of the proof is simple and we describe it here. By assumption, β is a root of a degree d polynomial $Q(x) \in \mathbb{Z}[x]$. Since d is the minimal degree of such a polynomial, it's not hard to check that $Q(p/q)$ is non-zero. But $Q(p/q)$ is a rational number with denominator q^d, and so $|Q(p/q)| \geq q^{-d}$.

But $Q(\beta) = 0$, and since Q is minimal, it's not hard to check that $Q'(\beta) \neq 0$. So $|Q(p/q)|$ has the same order of magnitude as $|\beta - p/q|$, and we see that $|\beta - p/q| \geq c(\beta)q^{-d}$. In rough terms, the polynomial Q "protects" β from rational approximations because $Q(\beta) = 0$ but $Q(p/q)$ cannot be too small.

Liouville's theorem is not strong enough to prove finiteness for any diophantine equation. When $d = 2$, Liouville's theorem is optimal, but for any $d > 2$, Thue was able to improve the exponent $-d$. Any improvement of the exponent in Liouville's theorem implies the finiteness corollary. In other words, once we know that $|\beta - p/q| \geq c(\beta)q^{-d+\epsilon}$ for any $\epsilon > 0$, then Thue's finiteness result follows.

Thue had the idea to use other polynomials besides just Q in order to protect β. Looking for other polynomials of one variable doesn't turn up anything, but Thue had the remarkable idea to use polynomials of two variables. If $P(x, y) \in \mathbb{Z}[x, y]$ is a polynomial of two variables that vanishes (maybe to high order) at (β, β), then P can "protect" β from pairs of good rational approximations $(p_1/q_1, p_2/q_2)$. To prove that $|2^{1/3} - (x/y)| \geq c|y|^{-2.51}$, Thue requires an infinite sequence of auxiliary polynomials $P_j(x, y) \in \mathbb{Z}[x, y]$ which vanish at $(2^{1/3}, 2^{1/3})$ to different orders. Each of these polynomials protects $(2^{1/3}, 2^{1/3})$ from rational approximations $(p_1/q_1, p_2/q_2)$ in certain ranges, and working all together they provide enough protection to prove Thue's theorem.

Thue carefully by hand crafted this infinite sequence of polynomials $P_j(x, y)$. He was able to construct the desired polynomials by hand when β is a dth root of a rational number. He became stuck trying to generalize his method to other algebraic numbers, because he didn't know how to construct the auxiliary polynomials. The problem of looking for these auxiliary polynomials is similar to our warmup problem. At a certain point, Thue gave up trying to craft the polynomials he needed. Instead, he proved that they must exist by counting parameters, essentially as we did above.

At the 1974 ICM, Schmidt gave a lecture [Schmidt74] on Thue's work and its influence in number theory. He wrote,

> The idea of asserting the existence of certain polynomials rather than explicitly constructing them is the essential new idea in Thue's work. As Siegel [1970] points out, a study of Thue's papers reveals that Thue first tried hard to construct the polynomials explicitly (and he actually could do so in case β^d is rational).

This idea reminds me of the probabilistic method. Thue proved that his auxiliary polynomials exist using the pigeon-hole principle. No one knows how to give an explicit formula for these polynomials, but there are so many polynomials that some of them are guaranteed to work.

Thue's wonderful argument has many similarities to the proofs in Sect. 1. All the arguments have the following general outline. First, by counting parameters, we find a polynomial that vanishes at certain places. Second, we use basic facts about polynomials to understand what the polynomial does at other places. Polynomials work in these arguments because they have a combination of rigidity and flexibility. Polynomials obey rigid properties like the vanishing lemma, which make them useful in the second step. On the other hand, there are lots of polynomials, which make them rather flexible in the first step. It's somewhat remarkable that such a large space of functions obeys such rigid properties.

2.2 The Resilience of Polynomials: Polynomials in Coding Theory

The two main ingredients in the proofs of finite field Nikodym and joints are the parameter counting lemma and the vanishing lemma. This team of ingredients appeared together earlier in the theory of error-correcting codes. Dvir has a background in coding theory, and this circle of ideas may have influenced his proof of the finite field Nikodym conjecture.

Let $\mathbb{F}_q$ be a finite field with q elements, and let $\text{Poly}_D(\mathbb{F}_q)$ be the vector space of all polynomials over $\mathbb{F}_q$ of degree $\leq D$. Because of the vanishing lemma, any two polynomials in $\text{Poly}_D(\mathbb{F}_q)$ can only agree at $\leq D$ points. As long as D is much less than q, any two polynomials in $\text{Poly}_D(\mathbb{F}_q)$ look very different from each other. This makes them interesting tools for building error correcting codes.

Here is a typical situation in coding theory. Q is a polynomial over $\mathbb{F}_q$ of degree $\leq q/1{,}000$. We want to transmit or save Q, but the data gets corrupted, and instead we end up with a function $F : \mathbb{F}_q \to \mathbb{F}_q$. Suppose we know that $F(x) = Q(x)$ for at least $(51/100)q$ values of x. Is it possible to recover Q from F?

It follows immediately from the vanishing lemma that Q can be recovered from F in theory. Suppose that Q_1 and Q_2 are polynomials of degree $\leq q/1{,}000$ that agree with F for $\geq (51/100)q$ values of x. Then $Q_1 - Q_2$ vanishes for at least $(2/100)q$ values of x, and so $Q_1 - Q_2$ is zero by the vanishing lemma. Hence there is only one polynomial $Q \in \text{Poly}_{q/1{,}000}(\mathbb{F}_q)$ consistent with the data F.

But there's a deeper question that remains: can we recover Q from F *in a practical way*? The argument above tells us that we can find Q by testing all the polynomials of degree $\leq q/1{,}000$—but the length of this procedure is more than exponential in q. In the mid-1980s, Berlekamp and Welch gave a polynomial-time algorithm to recover Q from F [BW86].

We consider the graph of F: the set $\{(x, y) \in \mathbb{F}_q^2 | F(x) = y\}$. This graph looks like a cloud of points. Inside the cloud of points a certain algebraic structure is hidden: most of the points lie on the graph of Q. How can we search out this algebraic structure hidden in the cloud of points?

The main idea of the algorithm is to find the lowest degree non-zero polynomial $P(x, y)$ that vanishes on the graph of F. On the one-hand, we can find an optimal P with an efficient algorithm. On the other hand, this optimal P uncovers the hidden algebraic structure in the cloud of points: looking at the zero set of P, the graph of Q jumps off the page.

We begin by explaining how to find this optimal P. This discussion is closely connected to the parameter counting argument in Sect. 1. Suppose we want to check whether there is a non-zero polynomial of degree $\leq D$ that vanishes on a set $S \subset \mathbb{F}_q^2$. Let $\mathrm{Poly}_D(\mathbb{F}_q^2)$ denote the space of all the polynomials with degree $\leq D$. Let $\mathrm{Fcn}(S, \mathbb{F}_q)$ be the vector space of all the functions from the set S to $\mathbb{F}_q$. This is a vector space of dimension $|S|$. Let $R : \mathrm{Poly}_D(\mathbb{F}_q^2) \to \mathrm{Fcn}(S, \mathbb{F}_q)$ be the restriction map which restricts each polynomial to the set S. The map R is a linear map between vector spaces, and it's not hard to write down an explicit matrix for it. The basic operations of linear algebra can be done in polynomial time. We can check whether R has a non-trivial kernel, and if it does we can find a non-zero element in the kernel. Doing this for each degree D, we find in polynomial time a non-zero polynomial P that vanishes on the graph of F and has minimal degree.

In the discussion so far, we treated the variables x and y on equal terms. Berlekamp and Welch actually treat them differently. This makes sense if we look back at the problem we're trying to attack. We're hoping to find the graph of Q, which is defined by $y - Q(x) = 0$. This defining equation has degree 1 in y and high degree in x. In order to adapt to the problem, it turns out to be a good idea to use polynomials $P(x, y)$ of degree 1 in y and high degree in x. From now on we just consider polynomials $P(x, y) = P_0(x) + yP_1(x)$. By the same linear algebra argument, we can find such a polynomial P which vanishes on the graph of F, and where $\max(\deg(P_0), \deg(P_1))$ is as small as possible.

We can also give an estimate for this degree. If we consider P_0, P_1 of degree $\leq D$, then we get a vector space of polynomials of dimension $2D + 2$. We want to find a polynomial that vanishes on the graph of F, which has q points. As long as $2D + 2 > q$, such a polynomial is guaranteed to exist by parameter counting. Therefore, we know that the degree of P_0, P_1 is $\leq q/2$.

Let's summarize. We found a polynomial $P(x, y) = P_0(x) + yP_1(y)$ which vanishes on the graph of F, where the degrees of P_0 and P_1 are as small as possible and definitely $\leq q/2$. This polynomial will help us to unlock the information hidden in F.

The key point is that P vanishes on the graph of Q! This follows in a few simple steps.

1. We know $P = 0$ on the graph of F. In other words, $P(x, F(x)) = 0$ for all x.
2. But we know that F usually agrees with Q. So $P(x, Q(x)) = 0$ for at least $(51/100)q$ values of x.
3. But $P(x, Q(x)) = P_0(x) + Q(x)P_1(x)$ is a polynomial in x of degree $\leq q/2 + q/1{,}000 < (51/100)q$.
4. By the vanishing lemma, $P(x, Q(x))$ is the zero polynomial!

We have proven that $P(x, Q(x)) = P_0(x) + Q(x)P_1(x)$ is identically zero. Hence $Q(x)P_1(x) = -P_0(x)$. We know P_0 and P_1, and now we can recover Q by doing polynomial division. This is the Berlekamp–Welch algorithm.

There is a more visual way of explaining how to recover Q, which makes the graph of Q jump off the page. We let the set of errors be $E := \{x \in \mathbb{F}_q | F(x) \neq Q(x)\}$. Adding a few more lines to the argument above, one can prove that the zero set of our polynomial P is the union of the graph of Q and a vertical line $x = e$ at each error $e \in E$. Looking at the zero set of P, the set of errors is immediately visible, together with a large chunk of the graph of Q. From this large chunk of the graph of Q, we can quickly recover Q itself.

Computer scientists working on error-correcting codes found a new set of questions about polynomials, very different from questions that pure mathematicians have considered. Working on these questions gave new perspectives about polynomials. Writing about coding theory in [Sudan95], Sudan referred to the resilience of polynomials: we can significantly distort the polynomial Q, but the information in Q survives. There is a lot more work on polynomials and coding theory. Some of it is described in [Sudan95] and in [Trevisan04]. The parameter counting lemma and the vanishing lemma continue to be important players.

2.3 Efficiency of Polynomials: Polynomials in Geometry

The last step of the proof of the distinct distance problem was influenced by ideas about polynomials in differential geometry. The overarching idea is that polynomials are geometrically efficient.

We begin with an older result about the efficiency of complex polynomials. The zero sets of complex polynomials are minimal surfaces. Let's formulate a precise result. We identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ and equip it with the standard Euclidean metric. Let P be a complex polynomial $\mathbb{C}^n \to \mathbb{C}$. Let $Z(P)$ denote the zero set of P. If the zero set of P does not contain any critical points of P, then $Z(P)$ is a submanifold of real dimension $2n - 2$.

Theorem 2.4 (**[Federer69]**). *Suppose that $P : \mathbb{C}^n \to \mathbb{C}$ is a complex polynomial, and that $F : \mathbb{R}^{2n} \to \mathbb{R}^2$ is a smooth function, so that $P = F$ outside of the unit ball $B^{2n} \subset \mathbb{R}^{2n} = \mathbb{C}^n$. Also, suppose that $Z(P)$ and $Z(F)$ don't contain any critical points, which implies that they are both manifolds. Then*

$$\mathrm{Vol}_{2n-2}\, Z(P) \cap B^{2n} \leq \mathrm{Vol}_{2n-2}\, Z(F) \cap B^{2n}.$$

This theorem says that complex algebraic surfaces do not waste any volume.

In this section, we will be interested in analogous results for real polynomials. Initially, it seems that there can be no such result. The Weierstrass approximation theorem says that any continuous function on a compact subset of $\mathbb{R}^n$ can be C^0 approximated by real polynomials. This basically means that real polynomials have no special properties at all.

But if we slightly shift the question, there is an interesting theorem discovered only in the last 10 years. Instead of focusing on one polynomial at a time, we focus on the space $\mathrm{Poly}_D(\mathbb{R}^n)$, the space of all polynomials of degree $\leq D$. Individual polynomials may be wasteful with volume, but we will see that the space $\mathrm{Poly}_D(\mathbb{R}^n)$ is efficient with volume. This follows from two results, one old and one new.

Proposition 2.5. *If P is a non-zero polynomial in $\mathrm{Poly}_D(\mathbb{R}^n)$, then*

$$\mathrm{Vol}_{n-1} Z(P) \cap B^n \leq C(n)D.$$

This is a classical result. Because P is a degree D polynomial, a line can intersect $Z(P)$ at most D times unless the whole line lies in $Z(P)$. The Crofton formula describes how the volume of a hypersurface can be reconstructed in terms of the number of intersections between the surface and all of the lines in space. When we plug our estimate on the intersection numbers into the Crofton formula, it follows that the volume of $Z(P) \cap B^n$ is $\leq C(n)D$.

Now comes the new result. Gromov compared $\mathrm{Poly}_D(\mathbb{R}^n)$ with other vector spaces of the same dimension and saw that $\mathrm{Poly}_D(\mathbb{R}^n)$ has approximately the smallest zero sets.

Theorem 2.6 ([Gromov03], see also [Guth09]). *If W is a vector space of continuous functions $B^n \to \mathbb{R}$, and if $\dim W = \dim \mathrm{Poly}_D(\mathbb{R}^n)$, then there exists $F \in W$ so that*

$$\mathrm{Vol}_{n-1} Z(F) \cap B^n \geq c(n)D.$$

The proof uses a result from topology, but in some ways it is similar to the proof of finite field Nikodym or joints. A leading role is played by the fact that $\dim \mathrm{Poly}_D(\mathbb{R}^n) \sim D^n$.

The contribution from topology is the Stone–Tukey ham sandwich theorem. The original ham sandwich theorem says that given three finite volume sets in $\mathbb{R}^3$, there is a plane that bisects all three. This theorem was proven by Banach in the late 1930s. Stone and Tukey generalized the result. For one thing, they generalized it to higher dimensions, but they did much more than that. They realized that the argument does not apply only to perfectly flat planes but also to many other families of surfaces. Stone and Tukey figured out the right way to formulate the theorem, making it much more general. The formulation is based on functions instead of hypersurfaces.

We say that a continuous function F bisects a finite volume set U if the subset of U where $F > 0$ has half the volume of U and the subset where $F < 0$ has half the volume of U.

Theorem 2.7 (Stone–Tukey 1942). *Suppose W is a vector space of continuous functions on a domain $\Omega \subset \mathbb{R}^n$, so that for every non-zero $F \in W$, $Z(F)$ has measure 0. Let $U_1, \ldots, U_N \subset \Omega$ be finite volume sets, where $N < \dim W$. Then there is a non-zero $F \in W$ which bisects each U_i.*

We can now sketch the proof of Gromov's theorem. If there is a non-zero function $F \in W$ so that $Z(F)$ has positive (n-dimensional!) measure, then it has infinite $(n-1)$-dimensional volume, and we are done. So we can assume that each $Z(F)$ has measure 0, and we can apply the Stone–Tukey ham sandwich theorem. Let U_i be $\sim D^n$ disjoint balls in B^n each of radius $\sim D^{-1}$. We choose a non-zero function $F \in W$ that bisects each ball. A classical result in geometry says that a surface bisecting a ball needs to have a certain minimal volume. In fact, the smallest bisecting surface is a disk through the center of the ball.

Bisection lemma. *If a hypersurface bisects $B^n(r)$, then it has volume at least $c(n)r^{n-1}$.*

Therefore, $Z(F) \cap U_i \geq c(n)D^{-(n-1)}$. And $Z(F) \cap B^n \gtrsim D^n D^{-(n-1)} = D$. This finishes the sketch of Gromov's estimate.

These ideas from geometry/topology give a new twist to the polynomial method. Using linear algebra, we can find a non-zero polynomial $P \in \mathrm{Poly}_D(\mathbb{R}^n)$ that vanishes on a set of points $p_1, \ldots, p_N$ as long as $N < \dim \mathrm{Poly}_D(\mathbb{R}^n)$. This fact plays a key role in the solutions of the finite field Nikodym problem and the joints problem. Now using the Stone–Tukey theorem from topology, we can find a non-zero polynomial $P \in \mathrm{Poly}_D(\mathbb{R}^n)$ that bisects some sets $U_1, \ldots, U_N$ as long as $N < \dim \mathrm{Poly}_D(\mathbb{R}^n)$. The proof of the distinct distance estimate uses this new twist. We will explain how to use it in Sect. 4.

3. Some Methods and Problems in Incidence Geometry

In this section, we describe the impact of the polynomial method in incidence geometry. We begin by recalling some important results and methods in the subject. Then we will come to the new applications of the polynomial method. We will try to motivate these results, and we will discuss why they are hard to prove with previous methods.

This section motivates the results, and in the next section, we will discuss the proofs of these results.

3.1 Incidence Theory in the Plane

Suppose that $\mathfrak{L}$ is a set of lines in the plane. Let $S_r(\mathfrak{L})$ be the set of r-rich points: the set of points that lie in $\geq r$ lines of $\mathfrak{L}$. One of the basic questions in the field is, “for a given number of lines and a given number of r, how big can $S_r(\mathfrak{L})$ be?” This question was answered in a fundamental theorem of Szemerédi and Trotter.

Theorem 3.1 ([ST83]). *If $\mathfrak{L}$ is a set of L lines in the plane, then $|S_r(\mathfrak{L})| \lesssim L^2 r^{-3} + L r^{-1}$.*

This theorem is a central result of incidence geometry.

The first estimates about this problem exploit the following basic fact:

Basic Fact. *Two lines intersect in at most one point.*

Using just this fact and doing some counting arguments, we get some basic estimates. We call these estimates ‘basic’ because they follow just from the basic fact above.

Basic estimate 1. $|S_r(\mathfrak{L})| \lesssim L^2 r^{-2}$.

At each point of $S_r(\mathfrak{L})$, there are $\binom{r}{2}$ pairs of lines intersecting. In total, there are only $\binom{L}{2}$ pairs of lines, and each pair only intersects once. Therefore, $|S_r(\mathfrak{L})| \leq \binom{L}{2}\binom{r}{2}^{-1} \sim L^2 r^{-2}$. Another short counting argument gives the following further estimate.

Basic estimate 2. *If $r \geq 2L^{1/2}$, then $|S_r(\mathfrak{L})| \lesssim L/r$.*

These estimates are not as strong as the conclusion of the theorem. For example, if $r = L^{1/2}$, then the theorem says that $|S_r(\mathfrak{L})| \lesssim L^{1/2}$, but the basic estimates give only $\lesssim L$.

There is a crucial example in the story showing that a proof of the Szemerédi–Trotter theorem requires some quite different ideas. The example involves lines over finite fields. Let $\mathbb{F}_q$ denote the finite field with q elements. Let $\mathfrak{L}$ be the set of q^2 non-vertical lines $y = mx + b$, $m, b \in \mathbb{F}_q$. Each point of $\mathbb{F}_q^2$ lies in q different lines of $\mathfrak{L}$. So we have $|S_q(\mathfrak{L})| = q^2$. Since $q = L^{1/2}$, we have $|S_{L^{1/2}}(\mathfrak{L})| = L$. For L lines in $\mathbb{R}^2$, the Szemerédi–Trotter theorem gives the much better bound $|S_{L^{1/2}}(\mathfrak{L})| \lesssim L^{1/2}$. Now it is still true in $\mathbb{F}_q^2$ that two lines intersect in at most one point. Therefore, we cannot possibly prove the Szemerédi–Trotter theorem just by exploiting the fact that two lines intersect in at most one point.

The main philosophical issue in the proof is to figure out what other information about lines in $\mathbb{R}^2$ we can use. We need to use something that is true in $\mathbb{R}^2$ but false in $\mathbb{F}_q^2$. There are several approaches to the problem, and in some way they all use the topology of the plane.

3.2 The Cutting Method

The cutting method was introduced by Clarkson, Edelsbrunner, Guibas, Sharir, and Welzl in [CEGSW90]. They used the method to give an elegant proof of the Szemerédi–Trotter theorem. They were also able to prove incidence geometry results in higher dimensions. We will discuss this more below. Cutting plays a crucial role in the later applications of the polynomial method.

We illustrate the cutting method by describing the main idea of the proof of the Szemerédi–Trotter theorem. The proof is a divide-and-conquer argument. We cut the plane into pieces using D red lines. Here $D \ll L$ is a parameter we can play with, and the D red lines don't have to be lines from $\mathfrak{L}$. The complement of the red lines consists of convex polygonal cells. The idea is that we use the basic estimates for the points and lines in each cell, and then sum up the pieces. This idea works well as long as the lines of $\mathfrak{L}$ and the points of $S_r(\mathfrak{L})$ are evenly distributed among the cells.

Let's be a little more precise about what we may hope for. The D red lines cut the plane into $\sim D^2$ cells. If the points were evenly distributed among the cells, we would have the following:

Equidistribution 1. *Each cell contains $\lesssim |S_r(\mathfrak{L})| D^{-2}$ points of $S_r(\mathfrak{L})$.*

Now a line may enter at most $D + 1$ cells, because it can only cross each red line once. Since there are $\sim D^2$ cells, each line enters only a small fraction of the cells. If the lines were evenly distributed among the cells, we would have the following

Equidistribution 2. *Each open cell intersects $\lesssim LD^{-1}$ lines of $\mathfrak{L}$.*

If we are allowed to choose any D, and find D red lines that evenly distribute $S_r(\mathfrak{L})$ and $\mathfrak{L}$, then using the basic estimates in each cell and adding the results we get the conclusion of the Szemerédi–Trotter theorem. In fact, we don't need to evenly distribute both $S_r(\mathfrak{L})$ and $\mathfrak{L}$—either one will suffice. We state this precisely as a proposition.

Proposition 3.2. *Let $\mathfrak{L}$ be a set of L lines in the plane and fix some r. Let $i = 1$ or 2. Suppose that for any $1 \leq D \leq L$, we can find D lines cutting the plane into $\sim D^2$ cells so that Equidistribution i holds. Then $|S_r(\mathfrak{L})| \lesssim L^2 r^{-3} + Lr^{-1}$.*

The proof of this result is just a calculation. When I first did this calculation, I thought I had understood the main idea of the proof of Szemerédi–Trotter. Getting the points or lines to evenly distribute among the cells seemed like a minor point to me. My wrong intuition went like this: if I just put down the dividing lines without thinking too much, then the points wouldn't have a reason to concentrate in any particular cells, so they would probably end up pretty evenly distributed. With a little more experience, I think that this intuition was totally wrong.

Here's an alternate perspective. If I choose D red lines, then I have $2D$ real parameters at my disposal. I would like each of D^2 cells to contain $\sim |S_r(\mathfrak{L})|/D^2$ points of $S_r(\mathfrak{L})$. I am trying to satisfy $\sim D^2$ conditions. In essence, I have $2D$ variables, and I am hoping to solve D^2 equations. Without other information, this is a plan that sounds unlikely to work.

Here's an example of a set of points which is impossible to equidistribute. Take any set of points lying on a closed convex curve in the plane. Each red line intersects the curve in at most 2 points. Therefore, D red lines cut the curve into $\leq 2D$ pieces. It follows that most of the $\sim D^2$ cells contain no points of the set.

This divide-and-conquer plan actually does work, driven by one further crucial idea. The crucial idea is to choose the D red lines independently at random from among the lines of $\mathfrak{L}$. If we do this, the lines of $\mathfrak{L}$ interact with the red lines in a good way, and we get something close to Equidistribution 2. We briefly give intuition why this may work. Suppose that we first randomly pick $D/10$ red lines from the lines of $\mathfrak{L}$ and look at the resulting cells. If one of these cells contains $\geq 100LD^{-1}$ lines of $\mathfrak{L}$, then it is very likely that one of them will be chosen among the next $D/10$ red lines, and the cell will get cut into smaller pieces. Cells intersecting more than $100LD^{-1}$ lines have a brief half-life, and this suggests that at the end of the process almost all cells will intersect $\lesssim LD^{-1}$ lines of $\mathfrak{L}$. This gives (a bit of) the flavor of the random line argument. We have left out some important details. The cutting method involves some further care, and the random cutting needs to be refined a little. But choosing a random subset of D lines from $\mathfrak{L}$ is a crucial first step.

3.3 Problems in Higher Dimensions

Generalizations of the Szemerédi–Trotter theorem are a central subject of incidence geometry. One natural direction is to work in higher dimensions. Instead of lines in the plane, we can consider k-planes in $\mathbb{R}^n$. Some of the proofs of the Szemerédi–Trotter theorem are very planar, and it is difficult to generalize them to $\mathbb{R}^n$ for $n \geq 3$. For example, [Székely97] gives a beautiful proof of the theorem using crossing numbers of graphs. This proof generalizes to a huge variety of problems in the plane, but it seems very difficult to generalize it to higher dimensions. The cutting method was invented partly in order to attack higher-dimensional problems.

Let's summarize how to adapt the method to higher dimensions. The general divide-and-conquer strategy still makes sense. To divide $\mathbb{R}^n$ into cells, we need D red hyperplanes instead of D red lines. They divide $\mathbb{R}^n$ into $\sim D^n$ cells. If we have some kind of equidistribution, we still get interesting estimates. Moreover, if we are studying a set of $(n-1)$-dimensional planes in $\mathbb{R}^n$, then we can randomly choose D hyperplanes from our set, and we get some type of equidistribution. The objects don't necessarily have to be planes—we can also study codimension 1 spheres, paraboloids or other shapes.

But if we are studying k-planes in $\mathbb{R}^n$ for $k < n-1$, then there is a major difficulty: k-planes do not divide $\mathbb{R}^n$ into cells. If we try to choose $(n-1)$-planes so that the k-planes are equidistributed among the cells, we cannot use the key random trick above. We are stuck with $\sim D$ parameters hoping to satisfy $\sim D^n$ conditions. Moreover, there are examples of arrangements of k-planes in $\mathbb{R}^n$ where no arrangement of hyperplanes gives equidistribution. These examples generalize the set of points on a convex curve described above.

In summary, there is a major obstacle in dealing with objects of codimension > 1. The joints problem is one of the simplest incidence problems in codimension > 1. That's one reason the joints problem is interesting and important. Following the joints theorem, it looks reasonable to use the polynomial method to attack other incidence problems in codimension > 1. We will see a number of results in this direction.

Before the polynomial method, I only know of one sharp estimate about incidences in codimension > 1. This is Toth's complex generalization of the Szemerédi–Trotter theorem [Toth03]. If $\mathfrak{L}$ is a set of L complex lines in $\mathbb{C}^2$, Toth proved that $|S_r(\mathfrak{L})| \lesssim L^2 r^{-3} + L r^{-1}$—the same estimate as for real lines in $\mathbb{R}^2$. From the point of view of topology, $\mathbb{C}^2$ is homeomorphic to $\mathbb{R}^4$ and the complex lines are homeomorphic to $\mathbb{R}^2$, and so in a topological sense the codimension is 2. Toth's proof is adapted from the first proof of Szemerédi and Trotter, and it is technically difficult.

In his work on the complex problem, Toth raised the following question. Suppose that $\mathfrak{L}$ is a set of k-planes in $\mathbb{R}^{2k}$, and that any two k-planes of $\mathfrak{L}$ intersect in ≤ 1 point. (In other words, we forbid two k-planes to contain a common line.) Is it still true that the number of r-rich points is $\lesssim L^2 r^{-3} + L r^{-1}$. This is a bold higher-dimensional generalization of the Szemerédi-Trotter theorem (and it also includes the complex version of the Szemerédi-Trotter theorem). Recently, Solymosi and Tao proved Toth's conjecture up to a factor of L^ϵ using the polynomial method.

Theorem 3.3 (**[Solymosi–Tao12]**). *If $\mathfrak{L}$ is a set of L k-planes in $\mathbb{R}^{2k}$, and if any two planes of $\mathfrak{L}$ intersect in ≤ 1 point, then for any $\epsilon > 0$, the number of r-rich points of $\mathfrak{L}$ is $\leq C(\epsilon) L^\epsilon (L^2 r^{-3} + L r^{-1})$.*

3.4 Distance Problems in the Plane

There are many deep open problems in incidence geometry even for curves in the plane. One example is the unit distance problem (which Erdős's posed in [Erdős46] alongside the distinct distance problem). It asks, given n points in the plane, how many pairs of points can have distance 1? In all known examples, the number of unit distances is $\lesssim n^{1+\epsilon}$. (In a square grid with a well-chosen spacing, the number of unit distances is slightly superlinear, but $\lesssim n^{1+\epsilon}$ for any $\epsilon > 0$.) The paper [SST] gives the best currently known bound: the number of unit distances is $\lesssim n^{4/3}$. This bound is closely connected with the Szemerédi–Trotter theorem. The unit distance problem is analogous to the Szemerédi–Trotter problem with unit circles in place of lines.

The reason for the difficulty seems to be the following. If we replace unit circles by "unit parabolas" (parabolas of the form $y = x^2 + ax + b$), then the bound $n^{4/3}$ is tight. To improve the $n^{4/3}$ bound, we need to find and use a property that is true for unit circles and false for unit parabolas. There's no clear candidate for this property or how to use it.

The distinct distance problem can also be phrased as a problem about circles in the plane, and it is difficult for similar reasons.

Elekes found a completely different way of thinking about the distinct distance problem, connecting it to problems in higher codimension like the ones we discussed in the last section.

3.5 Partial Symmetries

Suppose G is a group acting on a space X. If $P \subset X$ is a finite set, then we can look at the symmetries of P under the group action. We define

$$G(P) := \{g \in G \text{ such that } g(P) = P\}.$$

Elekes started a study of partial symmetries. A partial symmetry of P is a group element that maps a large chunk of P to another large chunk of P. More precisely we define the r-rich partial symmetries by

$$G_r(P) := \{g \in G \text{ such that } |g(P) \cap P| \geq r\}.$$

It's interesting to try to understand the size and structure of $G_r(P)$ in different situations. Elekes realized that this natural problem is closely connected to the distinct distance problem and to the incidence geometry of curves in 3-dimensional space. In these connections, the group G is the group of orientation-preserving rigid motions of the plane.

Conjecture 3.4 ([Elekes–Sharir10]). *If P is a finite set in the plane, and $r \geq 2$, then*

$$|G_r(P)| \lesssim |P|^3 r^{-2}.$$

(If P is a square grid, then this bound is tight up to a constant factor for all $2 \leq r \leq |P|/10$.)

Elekes and Sharir proved this conjecture for $r = 3$ using the polynomial method. Nets Katz and I proved the conjecture in [Guth–Katz11].

This conjecture is closely related to the distinct distance problem. Elekes realized that if a set P has few distinct distances, then it must have lots of partial symmetries. We sketch the reason. Let $Q(P)$ be the set of distance quadruples, defined as follows.

$$Q(P) := \{(p_1, q_1, p_2, q_2) | dist(p_1, q_1) = dist(p_2, q_2)\}.$$

If there are few distinct distances, then it stands to reason that there will be many pairs of points at the same distance. By a Cauchy–Schwarz argument, one gets $|d(P)||Q(P)| \gtrsim |P|^4$, where $d(P)$ is the number of distinct distances of the set P. So if there are few distinct distances, then $|Q(P)|$ will be large.

On the other hand, each quadruple in $Q(P)$ suggests a partial symmetry of P. For each quadruple of $Q(P)$, there is a unique rigid motion $g \in G$ so that $g(p_1) = p_2$ and $g(q_1) = q_2$. The rigid motion takes two points of P to two other points of P, so it belongs to $G_2(P)$. In this way, we get a map $E : Q(P) \to G_2(P)$. We want to use this map to count $Q(P)$. If the map E were injective, we would have $|Q(P)| \le |G_2(P)|$, which in turn is $\lesssim |P|^3$. The map E is actually not injective. If $|g(P) \cap P| = r$, then the preimage $E^{-1}(g)$ has size $\sim r^2$, because there are $\binom{r}{2}$ pairs of points in $g(P) \cap P$, and each pair yields a distance quadruple. Based on this observation, it's straightforward to relate $Q(P)$ and $G_r(P)$:

$$|Q(P)| \sim \sum_{r=2}^{|P|} r|G_r(P)|.$$

Plugging in the Elekes–Sharir conjecture gives $|Q(P)| \lesssim \sum_{r=2}^{|P|} |P|^3 r^{-1} \sim |P|^3 \log |P|$, and so $|d(P)| \gtrsim |P|/\log |P|$. So the Elekes–Sharir conjecture implies the new bound for the distinct distance problem.

The next observation of Elekes is that understanding the size of $|G_r(P)|$ is an incidence geometry problem where the background is the group G instead of Euclidean space. Instead of lines in $\mathbb{R}^3$, we consider the following special curves in G. For any two points $p_1, p_2 \in \mathbb{R}^2$, define

$$S_{p_1,p_2} := \{g \in G \text{ such that } g(p_1) = p_2\}.$$

These curves are natural objects from the point of view of the group structure of G. The curves S_{p_1,p_1} are 1-dimesional subgroups of G, and the curves S_{p_1,p_2} are their cosets.

For a finite set $P \subset \mathbb{R}^2$, let $S(P)$ denote the $|P|^2$ curves $\{S_{p_1,p_2}\}_{p_1,p_2 \in P}$. Next, we observe that a group element g is in $G_r(P)$ if and only if g lies in $\ge r$ of the curves of $S(P)$. This follows directly from the definition. If g is in $G_r(P)$, then it means that $g : P_1 \to P_2$ bijectively, where P_1 and P_2 are subsets of P with size r. For each point $p_1 \in P_1$, we have $g \in S_{p_1,g(p_1)}$, so g lies in r curves of $S(P)$. The converse direction is similar. So we can redefine $G_r(P)$ in the following way:

$$G_r(P) = \{g | g \text{ lies in } \ge r \text{ curves of } S(P)\}.$$

Understanding the size of $G_r(P)$ is closely analogous to understanding the number of r-rich points of a set of lines in $\mathbb{R}^3$. In particular, both problems involve objects of codimension 2, and they involve the difficulties discussed in Sect. 3.3.

In the future, mathematicians may consider the incidence theory of subgroups and cosets inside of a Lie group G by working intrinsically inside of G. For the time being, we are much more comfortable in Euclidean space, and we choose coordinates on G so that we get a problem about curves in Euclidean space. In the particular case of our curves $S(P)$ in our group

G, there is a good choice of coordinates where the curves become straight lines in $\mathbb{R}^3$. In these coordinates, the Elekes–Sharir conjecture reduces to the following two theorems about straight lines in $\mathbb{R}^3$. The theorems were proven in [Guth–Katz11].

Theorem A. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with $\leq L^{1/2}$ lines in any plane or degree 2 surface. Prove that the number of intersection points of lines of $\mathfrak{L}$ is $\lesssim L^{3/2}$.*

In particular, this theorem gives the best known estimate on the many intersection problem that we discussed in Sect. 1.

Theorem B. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ with $\leq L^{1/2}$ lines in any plane. For $3 \leq r \leq L^{1/2}$, prove that the number of r-rich intersection points of $\mathfrak{L}$ is $\lesssim L^{3/2} r^{-2}$.*

I like to think of this theorem as a generalization of the Szemerédi–Trotter theorem to lines in $\mathbb{R}^3$. There are probably many generalizations of that theorem to higher dimensions. Toth's conjecture is one generalization, and Theorem B is another generalization with a different flavor.

3.6 Conclusion

Studying incidence geometry problems in codimenson > 1 presents particular challenges. The polynomial method is the most effective tool currently available for studying these problems. The simplest case is the case of lines in $\mathbb{R}^3$. For lines in $\mathbb{R}^3$, the joints theorem and Theorems A and B give a good picture of what we now understand. The many intersections problem is a good example of what we still don't understand. In higher dimensions, the Solymosi–Tao result on the Toth conjecture is the main example of what we now know. This result is remarkable (partly) because it works with arbitrarily high dimensions and arbitrarily high codimensions.

Several problems can be transformed into incidence geometry problems in higher codimension. We have seen that the distinct distance problem in the plane and the partial symmetries of plane sets are both related to the incidence structure of lines in $\mathbb{R}^3$.

In the next section we will describe how to attack these problems using high-degree polynomials, extending the ideas from the proofs of finite field Nikodym and joints.

4. Combinatorial Structure and Algebraic Structure

In this section, we will discuss the proofs of Theorems A and B. The proofs are based on the polynomial method, and the key point is the connection between combinatorial structure and algebraic structure.

We saw earlier that lines in $\mathbb{R}^3$ may have many intersection points by clustering into either a plane or a degree 2 surface. Theorem A is a (partial) converse to this observation: more than $L^{\frac{3}{2}+\epsilon}$ intersection points may be formed *only if* the lines cluster into a plane or a degree 2 surface. Theorem A says that a certain combinatorial structure forces a certain algebraic structure. Our goal in this section is to explore how combinatorial structure can force algebraic structure.

We will see two different mechanisms how combinatorial structure can force algebraic structure. We begin by considering what we mean by algebraic structure.

4.1 Algebraic Structure for Finite Sets

If $X \subset \mathbb{R}^n$, let $\deg(X)$ be the smallest degree of a non-zero polynomial that vanishes on X. We have seen that for a finite set X, $\deg(X) \lesssim |X|^{1/n}$. Of course particular finite sets can have much lower degree. For instance, any subset of a plane has degree 1. For generic sets, the $|X|^{1/n}$ bound is sharp. So a generic finite set has $\deg(X) \sim |X|^{1/n}$. Any set with degree significantly smaller than $|X|^{1/n}$ has non-trivial algebraic structure.

There is a similar discussion for finite unions of lines. If X is a union of L lines in $\mathbb{R}^n$, then $\deg(X) \lesssim L^{\frac{1}{n-1}}$. The proof is straightforward, so we sketch it here. Suppose that D is a degree so that $(D+1)L < \dim \mathrm{Poly}_D(\mathbb{R}^n)$. Then we can choose a non-zero polynomial of degree $\le D$ that vanishes at $D+1$ points on each of the L lines. By the vanishing lemma, this polynomial vanishes on each line. A short calculation shows that we can choose $D \lesssim L^{\frac{1}{n-1}}$. In summary, any union of L lines has degree $\lesssim L^{\frac{1}{n-1}}$. If a union of L lines has degree significantly smaller than this, then it has some non-trivial algebraic structure.

With this definition of algebraic structure, we can begin to explore how combinatorial structure forces algebraic structure.

Proposition 4.1 (**Degree reduction**). *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$. Suppose that each line contains $\ge A$ (distinct) intersection points with other lines of $\mathfrak{L}$. Then the degree of the union of the lines is $\le 10^5 L/A$.*

(If $A \le L^{1/2}$, then the conclusion of the proposition is worthless, because every set of L lines has degree $\le 4L^{1/2}$ anyway. But if A is much bigger than $L^{1/2}$, then the lines have non-trivial algebraic structure.)

Here is the idea of the proof. We saw above that for any L' lines of $\mathfrak{L}$, there is a non-zero polynomial which vanishes on those lines with degree $\le 10(L')^{1/2}$. We let $\mathfrak{L}' \subset \mathfrak{L}$ be a subset of L' random lines of $\mathfrak{L}$, and we consider the polynomial P that vanishes on them. If A and L' are large enough, then this polynomial has to vanish on many other lines. Let l be another line of $\mathfrak{L}$. If l intersects the lines of $\mathfrak{L}'$ at $> \deg(P)$ points, then P will vanish on l also. The expected number of intersection points between l

and the lines of $\mathfrak{L}'$ is $A(L'/L)$. Whenever $A(L'/L) > 100(L')^{1/2}$, the expected number of intersection points is $> 10 \deg(P)$. In this situation, the polynomial P will vanish on the vast majority of the lines of $\mathfrak{L}$. Choosing L' optimally, we get a polynomial of degree $\leq 10^5 L/A$ that vanishes on most of the lines of $\mathfrak{L}$. (And with a little extra technique, we can get a polynomial that vanishes on all of the lines of $\mathfrak{L}$.)

I think this proposition is fundamental to the polynomial method. It shows that a set of lines with a lot of intersections must have an algebraic structure. This algebraic structure is an important clue to try to understand such sets of lines. Once we know that the set of lines has a non-trivial algebraic structure, it's natural to try to use algebra and algebraic geometry to understand the set better.

4.2 Ruled Surfaces

The proof of Theorem A is based on the theory of ruled surfaces. An algebraic surface $Z(P) \subset \mathbb{R}^3$ is called ruled if each point of $Z(P)$ lies in a line in $Z(P)$. If is called doubly ruled if each point of $Z(P)$ lies in two different lines in $Z(P)$. There is a classification of doubly ruled surfaces, and in particular the following result is relevant for us.

Proposition 4.2. *A doubly ruled algebraic surface $Z(P) \subset \mathbb{R}^3$ is a union of planes and degree 2 surfaces.*

Theorem A is a discrete analogue of this proposition from algebraic geometry. To try to make the analogy as close as possible, we state a small variation of Theorem A.

Theorem A$'$. *Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$, and that each line contains $\geq 10^{10} L^{1/2}$ intersection points with other lines of $\mathfrak{L}$. Then the lines of $\mathfrak{L}$ are contained in a union of $10^{-5} L^{1/2}$ planes and degree 2 surfaces.*

In this analogy, the set of intersection points of the lines of $\mathfrak{L}$ is a 'discrete approximation of a surface'. Each of these points lies in two lines of $\mathfrak{L}$, and each line of $\mathfrak{L}$ contains many points of our 'discrete surface'. The hypothesis is that we have a kind of 'discrete doubly ruled surface', and the conclusion is that $\mathfrak{L}$ is contained in a union of planes and degree 2 surfaces.

The degree reduction argument is a first step to prove Theorem A$'$. It tells us that the lines of $\mathfrak{L}$ are contained in the zero set of a polynomial P of degree $\leq 10^{-5} L^{1/2}$. This is the right bound for the degree, but we still have to prove that the polynomial factors into polynomials of degree 1 and 2. We have to understand better the structure of the polynomial P. We will see that the combinatorial structure of the lines of $\mathfrak{L}$ is connected with the geometric structure of $Z(P)$. We explain the connection in the next subsection.

4.3 Contagious Structures

Suppose that l is a line in $Z(P)$. If there are $> \deg(P)$ critical points of P on l, then every point of l is critical. The property of being critical is 'contagious'.

Let's give another example of a contagious property. Suppose now that $l \subset Z(P)$ and that each point of l is non-critical. A regular point x in $Z(P)$ is called flat if the curvature of $Z(P)$ vanishes at x—equivalently if there is a plane thru x which is tangent to $Z(P)$ to second order. If the line l contains more than $3\deg(P)$ flat points, then every point on the line is flat. So being flat is also a contagious property.

These properties are contagious because they are described by (other) polynomials. A point is critical if and only if $\partial_1 P$, $\partial_2 P$, and $\partial_3 P$ all vanish. These partial derivatives have degree $\leq \deg(P)-1$. It follows by the vanishing lemma that being critical is contagious. With a little more work, being flat is also described by polynomials. For any polynomial P, there exists a finite list of polynomials SP with degree $\leq 3\deg(P)$, and a (regular) point $x \in Z(P)$ is flat if and only if $SP(x) = 0$. It doesn't take that much work to construct SP, and then we see that being flat is contagious too.

To see how to use contagious properties, we will begin by discussing triple intersection points, because the method is a little easier. Suppose that $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$ and each line contains $\geq 10^{10} L^{1/2}$ triple intersection points. By degree reduction, these lines lie in $Z(P)$ for a polynomial P of degree $\leq 10^{-5} L^{1/2}$. Since the number of triple points on each line is much more than the degree, any contagious property of the triple points will spread to all of the lines.

Triple intersection points indeed have interesting properties. If x lies in 3 lines in $Z(P)$ and the lines are not coplanar, then x is a critical point of P, as we saw in the proof of the joints theorem. On the other hand, if x is not a critical point and x lies in three lines of $Z(P)$, then x is a flat point. The three lines must lie in the tangent plane of $Z(P)$, and then the tangent plane hugs $Z(P)$ along three lines, which forces it to be tangent to $Z(P)$ to second order. Anyway, every triple intersection point is either critical or flat. Since these properties are contagious, every point in the union of the lines of $\mathfrak{L}$ must be either critical or flat.

Contagious properties don't just spread from points to lines. If there are many lines with a contagious property, then it can spread to a whole surface. This follows from the following version of Bezout's theorem.

Theorem 4.3. *If P and Q are polynomials in three variables, and if they have no common factor, then $Z(P) \cap Z(Q)$ contains at most* $\deg(P) \cdot \deg(Q)$ *lines. In particular, if P is irreducible and $Z(P) \cap Z(Q)$ contains* $> \deg(P) \cdot \deg(Q)$ *lines, then Q vanishes on $Z(P)$.*

Remember that our L lines lie in $Z(P)$ where $\deg(P) \leq 10^{-5} L^{1/2}$. Each of the lines is either critical or flat. Suppose for a moment that they are all flat. (The critical case is similar.) For the sake of exposition, let's also assume that P is irreducible. The number of flat lines is $L \geq 10^{10}(\deg(P))^2$. Each polynomial of SP vanishes on these lines. The degree of each polynomial of SP is $\leq 3\deg(P)$. By the Bezout theorem, SP vanishes on $Z(P)$. This means that every point of $Z(P)$ is flat. Then it follows that $Z(P)$ is a plane.

In general, the polynomial P may be reducible and there may be several components, but a similar argument shows that they are all planes. We have sketched the proof of the following result, which essentially appears in [EKS11].

Theorem 4.4 ([**EKS11**]). *If $\mathfrak{L}$ is a set of L lines in $\mathbb{R}^3$, and each line contains $\geq 10^{10}L^{1/2}$ triple intersection points, then the union of the lines is contained in $\leq 10^{-5}L^{1/2}$ planes.*

This theorem is the case $r = 3$ of Theorem B.

It is harder to understand intersection points than triple intersection points. The problem is that if x lies in two lines in $Z(P)$, then it doesn't imply that x is either critical or flat. It's not clear right away if there is another contagious property that we can use instead.

To approach this question, let's step back and try to understand where contagious properties come from. We can build contagious properties by looking at polynomials in P and the derivatives of P. If RP is a polynomial of degree $\leq C$ in P and its derivatives, then $RP(x)$ is a polynomial in x of degree $\leq C\deg(P)$. We can use any such RP in place of SP in the argument above. Algebraic geometry helps understand what geometric properties of $Z(P)$ at a point x can be described by some polynomial equations in P and its derivatives. In short there are a lot of contagious properties.

We give one more example. A point $x \in Z(P)$ is called flecnodal if and only if there is a non-zero vector v so that P vanishes in the direction v to third order at x. It's not immediately obvious that being flecnodal is contagious, but it is. There is a polynomial FP, called the flecnode polynomial, of degree at most $11\deg(P)$, and a point $x \in Z(P)$ is flecnodal if and only if $FP(x) = 0$. This polynomial and this result were discovered by Salmon in the 1800s.

Stepping back from the details, we can describe the moral of the proof of Theorem A$'$. If x lies in two lines in $Z(P)$, it leads to some equations about P and the derivatives of P at x. These equations are all contagious, and so they end up holding at every point of $Z(P)$. So all the points of $Z(P)$ have a lot in common with the intersection points. After working out the details, it follows that *every point of $Z(P)$ lies in two lines in $Z(P)$*. The surface $Z(P)$ is doubly ruled. By the classification of doubly ruled surfaces, $Z(P)$ is a union of planes and degree 2 surfaces. We also know that the degree of P is $\leq 10^{-5}L^{1/2}$. Hence all the lines of $\mathfrak{L}$ lie in $\leq 10^{-5}L^{1/2}$ planes and degree 2 surfaces.

4.4 Polynomial Cell Decompositions

Theorem B involves a combination of all of the difficulties we have encountered in this essay so far. It is a problem about lines in $\mathbb{R}^3$, so the codimension is > 1. This suggests that the proof needs to use high degree polynomials. We saw in the last section how to prove the case $r = 3$ with the polynomial method, and I don't have any idea how to approach the problem without it. But for large r, Theorem B is false over finite fields like the Szemerédi–Trotter theorem. This suggests that the proof needs to use the topology of $\mathbb{R}^3$.

The proof of Theorem A does not generalize to Theorem B. It breaks down in the very first step: the degree reduction argument does not work.

The proof of Theorem B involves a combination of (almost) all of the methods that we've discussed in this essay. The key step is to build cell decompositions using polynomial surfaces, combining the cutting method and the polynomial method.

Instead of cutting space with D hyperplanes, we cut space with a degree D polynomial. A degree D polynomial surface has many good features in common with a union of D hyperplanes. The complement of D hyperplanes consists of $\sim D^n$ components. The complement of a degree D polynomial surface consists of $\lesssim D^n$ components, and there are $\sim D^n$ components in many examples. We will call these components cells. In each case, a line can only enter at most $D + 1$ cells.

The union of D hyperplanes is a special case of a degree D polynomial surface, but there are many more polynomial surfaces. Using polynomial surfaces gives us much more flexibility, and we have a better chance to prove equidistribution. Recall that we would like some equidistribution among $\sim D^n$ cells, which means we are trying to achieve $\sim D^n$ conditions. Choosing D hyperplanes gives us $\sim D$ degrees of freedom. But choosing a degree D polynomial surface gives us $\sim D^n$ degrees of freedom. Having so much more freedom, it looks more realistic to get equidistribution. Here is a precise result about building cell decompositions with polynomial surfaces.

Lemma 4.5 (**Polynomial cell decomposition lemma**). *If $\mathfrak{S}$ is any finite set in $\mathbb{R}^n$, and if $D \geq 1$ is any integer, then there is a non-zero polynomial $P \in \mathrm{Poly}_D(\mathbb{R}^n)$ so that each component of the complement of $Z(P)$ contains $\leq C(n)|\mathfrak{S}|D^{-n}$ points of $\mathfrak{S}$.*

We should make an important caveat right away. The lemma does not say that all the points of $\mathfrak{S}$ are in the complement of $Z(P)$. Some or even all the points of $\mathfrak{S}$ could lie in $Z(P)$.

The proof of the cell decomposition lemma is based on the Stone–Tukey ham sandwich theorem, which we discussed in Sect. 2.3. The ham sandwich theorem allows us to cut a bunch of sets in half. By using it repeatedly, we can cut our set of points into halves, then quarters, then eighths... Here is a detailed sketch.

1. The ham sandwich theorem says that given N finite volume open sets, we can choose a polynomial of degree $\lesssim N^{1/n}$ that bisects all of them.

We are dealing with finite sets, which have volume zero. Suppose that we have N finite sets $S_1, \ldots, S_N$. We let U_j be the ϵ-neighborhood of S_j. We apply the theorem to U_j and take the limit as ϵ goes to zero. In this way we get the following more combinatorial result.

2. If $S_1, \ldots, S_N$ are finite sets, then there is a polynomial P of degree $\lesssim N^{1/n}$ so that $P > 0$ on at most half the points of S_j and $P < 0$ on at most half the points of S_j. (Remark: P might vanish on some or even all of the points of S_j.)
3. We have a set S that we want to divide into 2^J fairly even pieces. Pick a plane that bisects S. Then pick a surface that bisects each half, leaving us with four sets of cardinality at most $|\mathfrak{S}|/4$. Next pick a surface that bisects each of these four sets. Continuing in this way, we have cut $\mathfrak{S}$ into 2^J pieces of cardinality at most $|\mathfrak{S}|2^{-J}$ by a union of J algebraic hypersurfaces. The degrees of these hypersurfaces are bounded by step 2, and adding up we get a total degree $\lesssim 2^{J/n}$ as desired.

Next we discuss how to use the polynomial cell decomposition lemma. We consider an arrangement of lines $\mathfrak{L}$, and we let $\mathfrak{S}$ be the set of r-rich points. We build a polynomial cell decomposition. If all the points of $\mathfrak{S}$ lie in the cells, then we can proceed by a divide-and-conquer argument as in the cutting method. We know that each cell has the same number of points of $\mathfrak{S}$, and we know the number of lines that enter an average cell. In each cell, we can use a more elementary method to count r-rich points. Adding up the contributions from all of the cells, we see that the number of r-rich points is $\lesssim L^{3/2}r^{-2}$—the conclusion of Theorem B.

This is not a complete proof of Theorem B. It may happen that most or all of the points of $\mathfrak{S}$ lie in $Z(P)$, and then the argument breaks down. Here is a slightly more optimistic way of looking at the situation.

The polynomial cell decomposition argument gives a second, completely different mechanism by which combinatorial structure forces algebraic structure. If $\mathfrak{L}$ is a set of L lines with significantly more than $L^{3/2}r^{-2}$ r-rich points, then the argument above shows that almost all of the r-rich points lie in $Z(P)$ for a polynomial P of surprisingly low degree. Since there are many r-rich points on each line, it follows that the lines lie in $Z(P)$ also, and the conclusion is that the degree of $\mathfrak{L}$ is far below $L^{1/2}$. The combinatorial structure of having many r-rich points forces algebraic structure.

Once the set of lines has algebraic structure, the rest of the proof of Theorem B is similar to the proof of Theorem A, using contagious properties.

The polynomial cell decomposition has had several other applications. The paper [Solymosi–Tao12] uses it to prove the higher-dimensional generalization of the Szemerédi–Trotter theorem. The paper [KMS12] uses it to give new proofs and perspectives on several fundamental theorems of incidence geometry.

4.5 Final Summary

The proofs we have been studying get off the ground by proving that arrangements with a lot of combinatorial structure must have unexpectedly low degree. We have seen two mechanisms to find these unexpectedly low degree polynomials. One mechanism is the degree reduction lemma. This lemma is proven by combining the parameter counting argument and the vanishing lemma. It's based on the proof of the finite field Nikodym conjecture and recovery algorithms for error-correcting codes. The second mechanism is the polynomial cell decomposition method. This mechanism is based on the polynomial method, but also on the cutting method and surface area estimates from differential geometry.

Once we know that the arrangement we are studying lies in the zero set of a polynomial of unexpectedly low degree, then it's natural to try to use that polynomial to study the set. The contagious structures are one tool to do that.

References

BW86. E. Berlekamp and L. Welch, *Error correction of algebraic block codes.* US Patent Number 4,633,470. 1986.

CEGPSSS92. B. Chazelle, H. Edelsbrunner, L. Guibas, R. Pollack, R. Seidel, M. Sharir, and J. Snoeyink, *Counting and cutting cycles of lines and rods in space*, Computational Geometry: Theory and Applications, 1(6) 305–323 (1992).

CEGSW90. K.L. Clarkson, H. Edelsbrunner, L. Guibas, M Sharir, and E. Welzl, Combinatorial Complexity bounds for arrangements of curves and spheres, Discrete Comput. Geom. (1990) 5, 99–160.

Dvir09. Z. Dvir, *On the size of Kakeya sets in finite fields*, J. Amer. Math Soc. (2009) 22, 1093–1097.

Erdős46. P. Erdős, *On sets of distances of n points*, Amer. Math. Monthly (1946) 53, 248–250.

Erdős. P. Erdős, *Some of my favorite problems and results*, in *The Mathematics of Paul Erdős*, Springer, 1996.

EKS11. Gy. Elekes, H. Kaplan, and M. Sharir, *On lines, joints, and incidences in three dimensions*, Journal of Combinatorial Theory, Series A (2011) 118, 962–977.

Elekes–Sharir10. Gy. Elekes and M. Sharir, *Incidences in three dimensions and distinct distances in the plane*, Proceedings 26th ACM Symposium on Computational Geometry (2010) 413–422.

Federer69. H. Federer, *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153 Springer-Verlag New York Inc., New York 1969.

Feldman–SharirS05. S. Feldman and M. Sharir, *An improved bound for joints in arrangements of lines in space*, Discrete Comput. Geom. (2005) 33, 307–320.

Gromov03. M. Gromov, *Isoperimetry of waists and concentration of maps.* Geom. Funct. Anal. 13 (2003), no. 1, 178–215.

Guth–Katz10. L. Guth and N. Katz, *Algebraic methods in discrete analogs of the Kakeya problem.* Adv. Math. 225 (2010), no. 5, 2828–2839.

Guth–Katz11. L. Guth and N. Katz, *On the Erdős distinct distance problem in the plane*, arXiv:1011.4105.

Guth09. *Minimax problems related to cup powers and Steenrod squares.* Geom. Funct. Anal. 18 (2009), no. 6, 1917–1987.

KMS12. H. Kaplan, J. Matoušek, and M. Sharir, *Simple proofs of classical theorems in discrete geometry via the Guth–Katz polynomial partitioning technique.* Discrete Comput. Geom. 48 (2012), no. 3, 499–517.

KSS10. H. Kaplan, M. Sharir, and E. Shustin, *On lines and joints*, Discrete Comput Geom (2010) 44, 838–843.

Laba08. I. Laba, *From harmonic analysis to arithmetic combinatorics.* Bull. Amer. Math. Soc. (N.S.) 45 (2008), no. 1, 77–115.

Quilodrán10. R. Quilodrán, *The joints problem in* $\mathbf{R^n}$, Siam J. Discrete Math, Vol. 23, 4, p. 2211–2213.

Schmidt74. Schmidt, Wolfgang M. *Applications of Thue's method in various branches of number theory.* Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974), Vol. 1, pp. 177–185. Canad. Math. Congress, Montreal, Que., 1975.

Solymosi–Tao12. J. Solymosi and T. Tao, *An incidence theorem in higher dimensions.* Discrete Comput. Geom. 48 (2012), no. 2, 255–280

SST. J. Spencer, E. Szemerédi, and W. Trotter, *Unit distances in the Euclidean plane.* Graph theory and combinatorics (Cambridge, 1983), 293–303, Academic Press, London, 1984.

Sudan95. M. Sudan, *Efficient checking of polynomials and proofs and the hardness of approximation problems*, ACM Distinguished Thesees, Springer 1995.

Székely97. L. Székely, *Crossing numbers and hard Erdős problems in discrete geometry.* Combin. Probab. Comput. 6 (1997), no. 3, 353–358.

ST83. E. Szemerédi and W. T. Trotter Jr., *Extremal Problems in Discrete Geometry*, Combinatorica (1983) 3, 381–392.

Tao01. T. Tao, *From rotating needles to stability of waves: emerging connections between combinatorics, analysis, and PDE.* Notices Amer. Math. Soc. 48 (2001), no. 3, 294–303.

Toth03. C. Toth, *The Szemerédi-Trotter theorem in the complex plane.* aXiv:math/0305283, 2003.

Trevisan04. L. Trevisan, *Some applications of coding theory in computational complexity.* Complexity of computations and proofs, 347–424, Quad. Mat., 13, Dept. Math., Seconda Univ. Napoli, Caserta, 2004.

Wolff99. T. Wolff. *Recent work connected with the Kakeya problem.* Prospects in mathematics (Princeton, NJ, 1996). pages 129–162, 1999.

The Number of Homothetic Subsets

Miklós Laczkovich* and Imre Z. Ruzsa

M. Laczkovich (✉)
Department of Analysis, Eötvös Loránd University, Múzeum krt. 6–8,
Budapest, H-I088 Hungary
Department of Mathematics, University College London, London,
WC1E 6BT, UK
e-mail: laczk@cs.elte.hu

I.Z. Ruzsa
Mathematical Institute of the Hungarian Academy of Sciences,
Budapest, Pf. 127, H-1364 Hungary
e-mail: ruzsa@renyi.hu

Summary. We investigate the maximal number $S(P,n)$ of subsets of a set of n elements homothetic to a fixed set P. Elekes and Erdős proved that $S(P,n) > cn^2$ if $|P| = 3$ or the elements of P are algebraic. For $|P| \geq 4$ we show that $S(P,n) > cn^2$ if and only if every quadruple in P has an algebraic cross ratio. Moreover, there is a sequence S_n of numbers such that $S(P,n) \asymp S_n$ whenever $|P| = 4$ and the cross ratio of P is transcendental.

AMS Subject Classification: primary 52ClO, secondary 05D99.

Let $\mathbf{C}$ denote the set of complex numbers. We say that the sets $A, B \subset \mathbf{C}$ are homothetic, and write $A \sim B$, if $B = a \cdot A + b = \{ax + b : x \in A\}$ with suitable $a, b \in \mathbf{C}$, $\neq 0$. If P and A are finite subsets of $\mathbf{C}$ then let

$$s(P,A) \overset{\text{def}}{=} |\{X \subset A : X \sim P\}|,$$

where $|H|$ denotes the cardinality of H. In [2] G. Elekes and P. Erdős investigated the behaviour of the sequence

$$S(P,n) \overset{\text{def}}{=} \max_{|A|=n} s(P,A).$$

It is easy to see that $S(P,n) \leq 2n(n-1)$ holds for every finite P and $n \in \mathbf{N}$. Elekes and Erdős proved that the order of magnitude of $S(P,n)$ is close to n^2 for every P; namely

$$S(P,n) \geq c \cdot n^{2-b\cdot\log^{-a} n}$$

holds for every $n \geq |P|$ with positive constants a, b, and c depending on P but not on n. They also showed that

* Supported by Hungarian National Foundation for Scientific Research, Grant T 7582.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_32,

$$S(P,n) \geq c \cdot n^2 \quad (n \geq |P|) \tag{1}$$

if $|P| = 3$ or if the elements of P are algebraic, and asked whether or not this is true for every finite P. In this paper we answer this question in the negative, and characterize the sets satisfying (1).

Our characterization will be given in terms of the cross ratio and projective equivalence. The cross ratio of the distinct complex numbers a, b, c, d is defined by

$$(a;b;c;d) \overset{\text{def}}{=} \frac{c-a}{c-b} : \frac{d-a}{d-b}.$$

A map $f : A \to B$ $(A, B \subset \mathbf{C})$ is said to be projective if it preserves the cross ratios of the quadruples of A. Two sets are projective equivalent if there is a projective bijection between them. Note that P and P' are projective equivalent whenever $|P| = |P'| \leq 3$. Our main result is the following.

Theorem 1. *For every finite P the following are equivalent.*

(i) There is a positive constant c such that (1) holds.
(ii) The cross ratio of every quadruple of P is algebraic.
(iii) There is a set P' such that P and P' are projective equivalent, and the elements of P' are algebraic.

First we prove the implication (ii) $\Longrightarrow$ (iii). Suppose that the set $P = \{a_1, \ldots, a_k\}$ satisfies (ii), where $a_1, \ldots, a_k$ are distinct complex numbers and $k \geq 4$ (if $k \leq 3$ then the statement is obvious). Let $f(x) = (px+q)/(rx+s)$ $(x \in \mathbf{C})$, where p, q, r, s are chosen such that $ps - qr \neq 0$, $ra_i + s \neq 0$ for $i = 1, \ldots, k$, and $f(a_i)$ is algebraic for $i = 1, 2, 3$. Let $P' = f(P)$. Since f is projective, P and P' are projective equivalent. For the same reason, the cross ratio of the numbers $f(a_1), f(a_2), f(a_3), f(a_i)$ is algebraic for every $4 \leq i \leq k$, Since the first three of these numbers are algebraic, so is $f(a_i)$. Thus the elements of P' are algebraic, and hence P satisfies (iii).

The implication (iii) $\Longrightarrow$ (i) is an immediate consequence of the following theorem and of the result of Elekes and Erdős stating that (1) holds if the elements of P are algebraic.

Theorem 2. *Let $|P| = |P'| = k$, and suppose that P and P' are projective equivalent. Then there are positive constants c_1 and c_2 depending only on k such that*

$$c_1 \cdot S(P,n) \leq S(P',n) \leq c_2 \cdot S(P,n) \tag{2}$$

for every $n \geq k$.

Lemma 1. *For every $Q \subset \mathbf{C}$, $|Q| = k$ and $n \geq k$ we have $S(Q,kn) \leq k^{2k} S(Q,n)$.*

Proof. Let $E \subset \mathbf{C}$ be such that $|E| = kn$ and $s(Q,E) = S(Q,kn)$. Let N denote the number of pairs (C,D) such that $C \subset D \subset E$, $C \sim Q$ and $|D| = n$.

Clearly, $N = S(Q, kn)\binom{kn-k}{n-k}$. On the other hand, as each D contains at most $S(Q, n)$ C's, we have $N \le \binom{kn}{n} S(Q, n)$. This gives

$$\frac{S(Q, kn)}{S(Q, n)} \le \frac{\binom{kn}{n}}{\binom{kn-k}{n-k}} = \frac{kn(kn-1)\dots(kn-k+1)}{n(n-1)\dots(n-k+1)} < k^{2k}. \qquad \square$$

Proof of Theorem 2. The statement of the theorem is obvious if $k = 1$ or $k = 2$, so that we may assume $k \ge 3$. Let a, b, c be distinct elements of P, let f be a projective bijection from P onto P', and let $f(a) = a'$, $f(b) = b'$, $f(c) = c'$. We can suppose that $a = a' = 0$ and $b = b' = 1$, since otherwise we replace P by $(P - a)/(b - a)$ and P' by $(P' - a')/(b' - a')$. This replacement will not affect the projective equivalence of the sets P and P', nor will it change the values of $S(P, n)$ and $S(P', n)$.

Let $A \subset \mathbf{C}$ be such that $|A| = n$ and $s(P, A) = S(P, n)$. If $X \subset A$ and $X \sim P$ then there are elements $x, y \in X$ such that $X = \{x + p(y - x) : p \in P\}$. Therefore, if T denotes the set of all pairs $(x, y) \in A \times A$ satisfying $x + p(y - x) \in A$ for all $p \in P$, then we have $|T| \ge S(P, n)$. Let $\lambda = c(c' - 1) \cdot (c'(c - 1))^{-1}$, and let μ be a number to be fixed later. We put

$$B = \{x + p'(\lambda y + \mu - x) : (x, y) \in T, p' \in P'\}.$$

If $x \ne \lambda y + \mu$, then the set $U_{x,y} = \{x + p'(\lambda y + \mu - x) : p' \in P'\}$ is similar to P'. We can select the number μ in such a way that $x \ne \lambda y + \mu$ for every $(x, y) \in A \times A$, and the sets $U_{x,y}$ $((x, y) \in A \times A)$ are distinct. Fixing such a μ it follows that

$$s(P', B) \ge |T| \ge S(P, n). \tag{3}$$

For every $p \in P$ we denote $p' = f(p)$,

$$A_p = \{x + p(y - x) : (x, y) \in T\}, \quad B_p = \{x + p'(\lambda y + \mu - x) : (x, y) \in T\}.$$

and

$$\phi_p(x + p(y - x)) = x + p'(\lambda y + \mu - x) \quad ((x, y) \in T).$$

First we show that ϕ_p is a well-defined map from A_p onto B_p. To this end we have to prove that if $(x, y) \in T$, $(u, v) \in T$ and

$$x + p(y - x) = u + p(v - u), \tag{4}$$

then

$$x + p'(\lambda y + \mu - x) = u + p'(\lambda v + \mu - u). \tag{5}$$

If $p = p' = 0$ or $p = p' = 1$ then the implication (4) $\Longrightarrow$ (5) is clear. If $p = c$, $p' = c'$, then (4) implies $(1 - c)x + cy = (1 - c)u + cv$. Therefore, by the definition of λ,

$$x + c'(\lambda y + \mu - x) = (1-c')x + \lambda c' y + c'\mu = \frac{1-c'}{1-c}((1-c)x + cy) + c'\mu =$$

$$\frac{1-c'}{1-c}((1-c)u + cv) + c'\mu = (1-c')u + \lambda c' v + c'\mu = u + c'(\lambda v + \mu - u),$$

which gives (5). Finally suppose $p \in P \setminus \{0,1,c\}$. Since f is projective, we have $(0;1;c;p) = (0;1;c';p')$ and thus $\lambda = p(p'-1) \cdot (p'(p-1))^{-1}$. Then a computation identical to the one above shows that (4) implies (5) in this case as well. This proves that the map ϕ_p is well-defined, and it is clear that $\phi_p(A_p) = B_p$.

It follows from the definition of T that $A_p \subset A$, and hence $|A_p| \leq n$. Thus we have $|B_p| \leq n$, and then $|B| \leq \sum_{p\in P} |B_p| \leq kn$. Combining with (3), this gives $S(P',\ kn) \geq S(P,n)$. Then, by Lemma 1, we obtain $S(P',n) \geq k^{-2k} S(P,n)$. Interchanging the roles of P and P' we get the other inequality of (2). □

In order to prove the implication (i) $\Longrightarrow$ (ii) of Theorem 1, we may assume that $|P| = 4$. Indeed, if (i) holds for P then it holds for every four-element subset of P as well, and if (ii) holds for every four-element subset of P, then it also holds for P.

Let α be a transcendental number, and denote

$$S_n = S(\{0,1,2,\alpha\},n) \quad (n \geq 4).$$

The value of S_n does not depend on the choice of α. Indeed, if β is another transcendental number, then there is a field-automorphism σ of $\mathbf{C}$ such that $\sigma(\alpha) = \beta$. For every $X \subset \mathbf{C}$, $|X| = 4$ we have $X \sim \{0,1,2,\alpha\}$ if and only if $\sigma(X) \sim \{0,1,2,\beta\}$ and this easily implies that $S(\{0,1,2,\alpha\},n) = S(\{0,1,2,\beta\},n)$ for every $n \geq 4$.

Theorem 3. *There are positive absolute constants c_1 and c_2 such that for every quadruple $P \subset \mathbf{C}$, if the cross ratio of the elements of P is transcendental, then*

$$c_1 \cdot S_n \leq S(P,n) \leq c_2 \cdot S_n$$

for every $n \geq 4$.

Proof. Let $P = \{a,b,c,d\}$ and $(a;b;c;d) = \alpha$. We put $\beta = 2/(2-\alpha)$ and $P' = \{0,1,2,\beta\}$. Since $(0;1;2;\beta) = \alpha$, the sets P and P' are projective equivalent. Consequently, by Theorem 2, there are absolute constants $c_1, c_2 > 0$ such that $c_1 \cdot S(P',n) \leq S(P,n) \leq c_2 \cdot S(P',n)$ for every $n \geq 4$. Also, since β is transcendental, we have $S(P',n) = S_n$ by the remark preceding the theorem. □

Now, in order to prove the implication (i) $\Longrightarrow$ (ii) of Theorem 1, it is enough to show that $S_n = o(n^2)$ as $n \to \infty$. Indeed, suppose this is true, and let $P \subset \mathbf{C}$ be a four-element set for which (i) holds but (ii) does not. Then, by Theorem 3, $c_2 \cdot S_n \geq S(P,n) \geq c \cdot n^2$ for every $n \geq 4$, a contradiction.

The rest of the paper will be devoted to the proof of $S_n = o(n^2)$. In the sequel we denote

$$S(n,c) = S(\{0,1,2,c\},n) \quad (c \in \mathbf{C} \setminus \{0,1,2\}, n \in \mathbf{N}).$$

Lemma 2. *For every n we have $S(n,c) \geq S_n$ for all, but a finite number of $c \in \mathbf{C}$.*

Proof. Let α be transcendental and let $A = \{a_1,\dots,a_n\} \subset \mathbf{C}$ be such that $s(\{0,1,2,\alpha\},A) = S_n$. This means that there is a set I of quadruples of indices such that $|I| = S_n$ and for every $(i,j,k,m) \in I$ we have

$$a_i - 2a_j + a_k = 0 \quad \text{and} \quad (\alpha-1)a_i - \alpha a_j + a_m = 0. \tag{6}$$

Let

$$V = \left\{ \sum_{i=1}^{n} r_i \cdot a_i : r_i \in \mathbf{Q}(\alpha), \quad i = 1,\dots,n \right\},$$

then V is a linear space over the field $\mathbf{Q}(\alpha)$. Let $B = \{b_1,\dots,b_d\}$ be a basis of V, and let

$$a_i = \sum_{j=1}^{d} r_{ij} \cdot b_j \quad (i = 1,\dots,n) \tag{7}$$

be representations with $r_{ij} \in \mathbf{Q}(\alpha)$ for every $i = 1,\dots,n$ and $j = 1,\dots,d$. Since the coefficients of a_i, a_j, a_k, a_m in the equations (6) belong to $\mathbf{Q}(\alpha)$, and $b_1,\dots,b_d$ are linearly independent over $\mathbf{Q}(\alpha)$, it follows that substituting the representations (7) into the equations (6) we obtain identities. In other words, for every choice of the variables $x_1,\dots,x_d$, the numbers

$$c_i = \sum_{j=1}^{d} r_{ij} \cdot x_j \quad (i = 1,\dots,n) \tag{8}$$

will satisfy the equations

$$c_i - 2c_j + c_k = 0 \text{ and } (\alpha-1)c_i - \alpha c_j + c_m = 0 \tag{9}$$

for every $(i,j,k,m) \in I$. The right-hand sides of (8), as linear forms, are different, because for $x_j = b_j$ they have different values (namely $a_1,\dots,a_n$). Then $x_1,\dots,x_d$ can be chosen to be integers such that the values of the corresponding $c_1,\dots,c_n$ are different. Indeed, for every $i_1 \neq i_2$, the set $\{(x_l,\dots,x_d) : c_{i_1} = C_{i_2}\}$ is a hyperplane, and $\mathbf{Z}^d$ cannot be covered by finitely many hyperplanes. Clearly, if the x_j's are integers then $c_i \in \mathbf{Q}(\alpha)$ for every $i = 1,\dots,n$.

We have proved that there are distinct elements $c_i \in \mathbf{Q}(\alpha)$ satisfying the equations (9). For every $i = 1,\dots,n$ there is a rational function R_i with rational coefficients such that $c_i = R_i(\alpha)$. Thus we have

$$R_i(\alpha) - 2R_j(\alpha) + R_k(\alpha) = 0 \quad \text{and} \quad (\alpha-1)R_i(\alpha) - \alpha R_j(\alpha) + R_m(\alpha) = 0$$

for every $(i,j,k,m) \in I$. Since α is transcendental, the rational functions $R_i(t)-2R_j(t)+R_k(t)$ and $(t-1)R_i(t)-tR_j(t)+R_m(t)$ must be identically zero for every $(i,j,k,m) \in I$. Therefore, whenever the numbers $R_1(c),\dots,R_n(c)$ are defined (that is, c is not a root of the denominator of $R_1 \cdot \ldots \cdot R_n$), and are distinct, then the set $A_c = \{R_1(c),\dots,R_n(c)\}$ contains S_n subsets similar to $\{0,1,2,c\}$. The rational functions R_i are distinct, since they have different values at α. Then $|A_c| = n$ for all but a finite number of c's, and for such a c we have $S(n,c) \ge S(\{0,1,2,c\}, A_c) \ge S_n$. □

We remark that, in fact, $S(n,c) = S_n$ holds for all but finitely many $c \in \mathbf{C}$. As we shall not need this result, we omit the proof. What we need is the fact that $S_n \le S(n,k)$ for every $k \in \mathbf{N}$, $k > k_0(n)$. This implies

$$\limsup_{n\to\infty} \frac{S_n}{n^2} \le \limsup_{n\to\infty} \left(\limsup_{k\to\infty} \frac{S(n,k)}{n^2} \right), \tag{10}$$

and hence, in order to prove $S_n = o(n^2)$, it is enough to show that the right-hand side of (10) is zero. In the next theorem we prove somewhat more.

Theorem 4.

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{S(n,k)}{n^2} = 0.$$

Let $z, q_1, \dots, q_d \in \mathbf{Z}$ and $X_1,\dots,X_d \in \mathbf{N}$. We shall say that the set

$$R(z,q_1,\dots,q_d;X_1,\dots,X_d) = \left\{ z + \sum_{i=1}^{d} x_i q_i : x_i \in \mathbf{Z}, 0 \le x_i < X_i (i=1,\dots,d) \right\}$$

*is a d-*dimensional arithmetical progression of size $\prod_{i=1}^{d} X_i$. *We shall need the following theorem proved by G. A. Freiman in [3–5] and I. Z. Ruzsa in [6].*

Theorem 5. *If $A \subset \mathbf{Z}$, $|A| = n$, and $|A+A| \le cn$, then A is contained in a d-dimensional arithmetical progression of size not exceeding $c'n$, where d and c' only depend on c.*

Let $A \subset \mathbf{Z}$ be a finite set, let $G = (A,E)$ be a graph, and put $S = \{a+b : (a,b) \in E\}$. The following result was proved by A. Balog and E. Szemerédi in [1].

Theorem 6. *If $|A| = n$, $|E| \ge c_1 n^2$ and $|S| \le c_2 n$, then there is a subset $A' \subset A$ such that $|A'| \ge c_3 n$ and $|A' + A'| \le c_4 n$, where c_3 and c_4 only depend on c_1 and c_2.*

As Balog and Szemerédi remark, these two theorems can be combined to obtain the following result: if $|A| = n$, $|E| \ge c_1 n^2$ and $|S| \le c_2 n$, then there is a d-dimensional arithmetical progression R of size not exceeding $c_5 n$ such that $|R \cap A| \ge c_6 n$, where d, c_5, c_6 depend only on c_1 and c_2. In the next lemma we prove a slight improvement of this result.

Lemma 3. *If $|A| = n$, $|E| \geq c_1 n^2$ and $|S| \leq c_2 n$, then there is a d-dimensional arithmetical progression R of size not exceeding $c_{11} n$ such that the subgraph of G induced by the set $R \cap A$ contains at least $c_{12} n^2$ edges. Here the constants d, c_{11} and c_{12} depend only on c_1 and c_2.*

Proof. In the sequel $c_7, c_8, \dots$ will denote constants depending only on c_1 and c_2. Since $|E| \geq c_1 n^2$, there is a subset $A_1 \subset A$ such that in the subgraph of G induced by A_1, the degree of every point is greater than $c_1 n/2$. (Indeed, delete one by one each point of degree at most $c_1 n/2$. In this way we cannot remove all points of A, and the set of remaining points will have the required property.) Then $|A_1| \geq c_1 n/2$. Applying Theorem 6 to the graph induced by A_1, we obtain a subset $A' \subset A_1$ such that $|A'| \geq c_7 n$ and $|A' + A'| \leq c_8 n$. Then we apply Theorem 5 to obtain a d-dimensional arithmetical progression $R = R(z, q_1, \dots, q_d; X_1, \dots, X_d)$ such that $A' \subset R$ and the size of R does not exceed $c_9 n$.

Let $c_{10} = c_1 c_7 / 8$ and let B be the set of those points of A that are connected to at least $c_{10} n$ points of A'. Since there are at least $(c_7 n \cdot c_1 n/2)/2 = 2 c_{10} n^2$ edges starting from the points of A', we have $|B| \geq c_{10} n$. Let $\mathcal{R} = \{R + b : b \in B\}$. Since $A' \subset R$, each set $R + b$ ($b \in B$) contains at least $c_{10} n$ elements of the form $a + b$, where $a \in A'$ and $(a, b) \in E$. These elements belong to S and, by assumption, $|S| \leq c_2 n$. This implies that any pairwise disjoint subsystem of $\mathcal{R}$ contains at most c_2 / c_{10} sets. Let $\{R + b_i : i = 1, \dots, k\}$ be a maximal disjoint subsystem of $\mathcal{R}$. Then $k \leq c_2 / c_{10}$ and for every $b \in B$ there is an $i \leq k$ such that $(R + b) \cap (R + b_i) \neq \emptyset$. This implies $b \in (R - R) + b_i$, and hence $B \subset \cup_{i=1}^{k} [(R - R) + b_i]$. Let $b_{k+1} = z$ and $H = \cup_{i=1}^{k+1} [(R - R) + b_i]$. Then $A' \cup B \subset H$ as $A' \subset R \subset (R - R) + b_{k+1}$. Thus the subgraph of G induced by $H \cap A$ contains at least $c_{12} n^2$ edges ($c_{12} = c_{10}^2 / 2$), since every point of B is connected to at least $c_{10} n$ points of A'. Let $q_{d+j} = b_j$ for $j = 1, \dots, k+1$, and put

$$R' = \left\{ \sum_{i=1}^{d+k+1} x_i q_i : -X_i \leq x_i < X_i (1 \leq i \leq d) \text{ and } 0 \leq x_i < 2 (d < i \leq d+k+1) \right\}.$$

Then $H \subset R'$, and in order to complete the proof of the lemma it is enough to note that R' is a $d+k+1$-dimensional arithmetical progression of size not exceeding $2^{k+1} \prod_{i=1}^{d} (2X_i) \leq 2^{d+k+1} c_9 n = c_{11} n$. □

Lemma 4. *Let $n, k \in \mathbf{N}, c > 0$, and suppose that $S(n, k) > cn^2$. Then there are positive constants d, c', c'' depending only on c, and there exists a d-dimensional arithmetical progression $R = R(z, q_1, \dots, q_d; X_1, \dots, X_d)$ such that the* $\gcd(q_1, \dots, q_d) = 1$*, the size of R is at most $c' n$, and for a suitable integer u the set*

$$\{(a, b) \in R \times R : 2a + k(b - a) = u\} \tag{11}$$

contains at least $c'' n$ distinct pairs.

Proof. Since $S(n,k) > cn^2$, there is a set $A \subset \mathbf{C}$ such that $|A| = n$ and A contains more than cn^2 subsets similar to $\{0,1,2,k\}$. Repeating the argument of the proof of Lemma 2 with $\mathbf{Q}$ instead of $\mathbf{Q}(\alpha)$, we can see that A can be chosen to be a subset of $\mathbf{Q}$. Multiplying by a suitable integer, we may assume that $A \subset \mathbf{Z}$.

Let E denote the set of all pairs (a,b) such that each number a, b, $(a+b)/2$ and $a + \frac{k}{2}(b-a)$ belongs to A; then $|E| > cn^2$. Let $S = \{a+b : (a,b) \in E\}$, then $S \subset 2A$, and hence $|S| \le n$. Therefore, by Lemma 3, there are positive constants d, c', c'' depending only on c and there exists a d-dimensional arithmetical progression $R = R(z, q_1, \ldots, q_d;\ X_1, \ldots, X_d)$ of size not exceeding $c'n$ such that $A \cap R$ contains at least $c''n^2$ edges from E. Let $F = \{(a,b) \in E : a, b \in R\}$. Then $|F| > c''n^2$ and $|\{2a + k(b-a) : (a,b) \in F\}| \le |2 \cdot A| = n$. Consequently, there exists an element $u \in A$ such that the set defined in (11) contains at least $c''n$ distinct pairs. If $(q_1, \ldots, q_d) = 1$, then the proof is complete. Otherwise let $(q_1, \ldots, q_d) = m$, and replace R by $R(z, q_1/m, \ldots, q_d/m; X_1, \ldots, X_d)$. It is clear that this modified R satisfies the requirements. □

Lemma 5. *Let $R = R(z, q_1, \ldots, q_d; X_1, \ldots, X_d)$ be a d-dimensional arithmetical progression with $(q_1, \ldots, q_d) = 1$, and denote $s = \prod_{i=1}^d X_i$ and $M = \min_{1 \le i \le d} X_i$. Then for every positive integer k, the number of elements of R in any residue class $\mod k$ is at most $(s/M) + (s/k)$.*

Proof. For $i = 1, \ldots, d$ define

$$k_i = \frac{(k, q_1, \ldots, q_{i-1})}{(k, q_1, \ldots, q_i)}, \quad k_1 = \frac{k}{k, q_1}.$$

We have $k_l \ldots k_d = k$. Now consider the k numbers

$$y_1 q_1 + \ldots + y_d q_d, 0 \le y_i \le k_i - 1.$$

We show that they are all incongruent modulo k. Indeed, suppose that

$$y_1 q_1 + \ldots + y_d q_d \equiv y_1' q_1 + \ldots + y_d' q_d \pmod{k}.$$

Let j be the largest subscript for which $y_j \ne y_j$. Since the first $j-1$ summands are divisible by $(q_1, \ldots, q_{j-1})$, we have

$$(k, q_1, \ldots, q_{j-1}) \big| q_j (y_j - y_j'),$$

hence

$$k_j = \left. \frac{(k, q_1, \ldots, q_{j-1})}{(k, q_1, \ldots, q_j)} \right| y_j - y_j',$$

which contradicts the assumption $0 \le y_j, y_j' \le k_j - 1, y_j \ne y_j'$.

Let R^* be the set of those elements of R that are $\equiv a \pmod{k}$. The sets

$$R^* + y_1 q_1 + \ldots + y_d q_d, \quad 0 \le y_i \le k_i - 1$$

lie in different residue classes modulo k, hence they are disjoint, and they are contained in $R(z, q_1, \ldots, q_d, X_1 + k_1 - 1, \ldots, X_d + k_d - 1)$, consequently

$$k|R^*| \leq \prod (X_i + k_i - 1)$$

(we recall that $k = \prod k_i$), or

$$|R^*| \leq \frac{s}{k} \prod \left(1 + \frac{k_i - 1}{X_i}\right) \leq \frac{s}{k} \prod \left(1 + \frac{k_i - 1}{M}\right).$$

To complete the proof we show that the inequality

$$\prod_{i=1}^{d} \left(1 + \frac{k_i - 1}{M}\right) \leq 1 + \frac{(\prod k_i) - 1}{M}$$

holds for arbitrary real numbers $k_i \geq 1, M \geq 1$. For $d = 2$ this inequality asserts that

$$\left(1 + \frac{k_1 - 1}{M}\right) \left(1 + \frac{k_2 - 1}{M}\right) \leq 1 + \frac{k_1 k_2 - 1}{M}.$$

After multiplying by M and rearranging this becomes

$$\frac{(k_1 - 1)(k_2 - 1)}{M} \leq (k_1 - 1)(k_2 - 1),$$

which is true if all the variables are ≥ 1. The case for $d \geq 3$ now easily follows by an induction on d. □

Now we turn to the proof of Theorem 4. Suppose that $\limsup_{n,k \to \infty} S(n, k)/n^2 > 0$. Then there is a constant $c > 0$ such that for every K there are integers $n, k > K$ with $S(n, k) > cn^2$. Thus, by Lemma 4, there are positive constants d, c', c'' such that the following statement holds:

($*$) for every K there is a d-dimensional arithmetical progression

$$R_K = R(z^K, q_1^K, \ldots, q_d^K; X_1^K, \ldots, X_d^K)$$

and there are integers $n, k > K$ such that $(q_1^K, \ldots, q_d^K) = 1$, the size of R_K is at most $c'n$, and for a suitable integer u the set

$$\{(a, b) \in R_K \times R_K : 2a + k(b - a) = u\}$$

contains at least $c''n$ distinct pairs.

We prove that this is impossible. Let d be the smallest positive integer such that ($*$) holds for suitable positive constants c' and c''. If $2a + k(b-a) = u$ then $2a \equiv u \pmod{k}$ and hence, if R_K, n, k are as in ($*$) then $2 \cdot R_K$ contains at least $c''n$ elements in one of the residue classes mod k. This implies that R_K contains at least $c''n/2$ elements in one of the residue classes mod k. Let $M_K = \min_{1 \leq i \leq d} X_i^K$. Then, by Lemma 5, we have

$$c''/2 \leq (c'/M_K) + (c'/k). \tag{12}$$

Let $C = 4c'/c''$. If $K > C$, then $k > K$ implies $c'/k < c''/4$ and thus (12) gives $M_K < C$. By rearranging the indices we may assume that $X_d^K < C$ holds for every $K > C$. This implies $d > 1$. Indeed, if $d = 1$ then we have $|R_K| = |X_1^K| < C$ for $K > C$. This gives $|R_K \times R_K| < C^2$, which contradicts $(*)$ for $K > C^2/c''$.

We complete the proof by showing that $(*)$ also holds for $d-1$ instead of d, if we replace c'' by c''/C^2. Since d was minimal, this will provide the contradiction we were looking for.

For every K we put $R'_k = R(z^K, q_1^K, \dots, q_{d-1}^K; X_1^K, \dots, X_{d-1}^K)$. Also, for every $a \in R_K$ we choose a representation $a = z^K + \sum_{i=1}^d x_i q_i^K$ with $0 \le x_i < X_i (i = 1, \dots, d)$ and define $a' = z^K + \sum_{i=1}^{d-1} x_i q_i^K$. In this way we have defined a map $a \mapsto a'$ from R_K into R'_K.

Let $K > C$, and let R_K, n, k and u be as in $(*)$. Obviously, $a - a' \in \{i \cdot q_d^K : 0 \le i \le x_d^K - 1\}$ for every $a \in R_K$ and hence the set

$$P = \{2a' + k(b' - a') : (a, b) \in R_K \times R_K, 2a + k(b-a) = u\}$$

contains at most $(X_d^K)^2$ distinct elements. Since $K > C$, we have $X_d^K < C$ and thus $|P| \le C^2$. This implies that for a suitable u' the set

$$\{(a', b') \in R'_K \times R'_K : 2a' + k(b' - a') = u'\}$$

contains at least $(c''/C^2)n$ distinct pairs. Therefore, replacing c'' by c''/C^2, the $d-1$-dimensional arithmetical progression R'_K and the integers n, k, u' will satisfy the statement of $(*)$ apart from the condition that $(q_1^K, \dots, q_{d-1}^K) = 1$. But this condition can also be fulfilled if we replace R'_K by

$$R(z^K, q_1^K/m, \dots, q_{d-1}^K/m; X_1^K, \dots, X_{d-1}^K),$$

where $m = (q_1^K, \dots, q_{d-1}^k)$. □

References

1. A. Balog and E. Szemerédi, A statistical theorem of set addition, Combinatorica 14 (1994), 263–268.
2. G. Elekes and P. Erdös, Similar configurations and pseudogrids, preprint.
3. G. A. Freiman, Foundations of a structural theory of set addition (in Russian), Kazan Gos. Ped. Inst., Kazan, 1966.
4. G. A. Freiman, Foundations of a structural theory of set addition, Translation of Mathematical Monographs vol. 37, Amer. Math. Soc., Providence, R. I., USA, 1973.
5. G. A. Freiman, What is the structure of K if $K + K$ is small?, Lecture Notes in Mathematics 1240, Springer, Berlin-New York 1987, pp. 109–134.
6. I. Z. Ruzsa, Generalized arithmetical progressions and sumsets, Acta Math. Sci. Hung. 65 (1994), 379–388.

On Lipschitz Mappings Onto a Square*

Jiří Matoušek**

J. Matoušek (✉)
Department of Applied Mathematics, Institute of Theoretical
Computer Science (ITI), Charles University, Malostranské nám. 25,
118 00 Praha 1, Prague, Czech Republic
e-mail: matousek@kam.mff.cuni.cz

1. Introduction

The following problem was posed by Laczkovich [5]: Let $E \subseteq \mathbb{R}\mathbb{R}^d$ $(d \geq 2)$ be a set with positive Lebesgue measure $\lambda^d(E) > 0$. Does there exist a Lipschitz mapping $f : \mathbb{R}^d \to Q = [0,1]^d$, such that $f(E) = Q$? Preiss [6] answered this question affirmatively for $d = 2$:

Theorem 1. *Let $E \subseteq \mathbb{R}^2$ be a set with $\lambda^2(E) > 0$. There exists a Lipschitz mapping f of the plane onto the square $Q = [0,1]^2$, such that $f(E) = Q$.*

Later it was observed by Jones that this theorem is also an easy consequence of a much earlier result of Uy [8].

In this note we give a somewhat different proof of this theorem based on a well-known combinatorial lemma due to Erdös and Szekerés. By an additional trick, we also prove the following "absolute constant" version:

Theorem 2. *There exists a constant $c > 0$ such that for any E in the plane with Lebesgue measure $\lambda^2(E) = 1$ there exists a 1-Lipschitz mapping f of the plane onto the square $Q = [0,c]^2$, such that $f(E) = Q$.*

The value of c obtained from our proof is quite small. Clearly some improvement is possible, but it seems that our method is not suitable for obtaining the best possible value of c. It is easy to see that one cannot hope to push c arbitrarily close to 1 (e.g., consider E being a disk of unit area; its diameter $2/\sqrt{\pi}$ is smaller than the diagonal of the unit square, and so there is no 1-Lipschitz mapping onto the unit square).

*This paper was originally published in 1997. The present version is a minor revision from 2012, which includes some references to newer developments.

** It is my pleasure to thank University College of London for enabling my visit, during which this research was conducted. The visit was also supported by the Humboldt Foundation.

[1] Preiss in fact proves a slightly stronger statement, namely that f can be taken such that $f(\mathbb{R}^2 \setminus E)$ is countably rectifiable (i.e. it can be covered by a countable set of Lipschitz curves). In order to keep this note technically simple, we will not prove this strengthening here, although our method also provides it with some extra care.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_33,

Laczkovich's question for $d > 2$ remains open; a combinatorial problem related to an attempt to generalize our proof is stated at the end of Sect. 2.

2. Proof of Theorem 1

The metric in $\mathbb{R}^2$, implicitly considered in Theorem 1, is the usual Euclidean metric. It appears more convenient to work with the maximum metric d_∞ defined by

$$d_\infty((x_1, y_1), (x_2, y_2)) = \max(|x_1 - x_2|, |y_1 - y_2|)\,.$$

This metric will be used throughout the rest of this paper. Clearly this modification does not affect the validity of Theorem 1.

We begin the proof by a simple lemma, which is essentially contained in [6]. Let $a > 0, w > 0$ be real numbers and $\varphi : [0, a] \to [0, a]$ be a 1-Lipschitz real function. Let Q denote the square $[0, a]^2$. We partition Q into three regions defined as follows (see Fig. 1): L (resp. S, resp. U) is the set of points $(x, y) \in Q$ with $y < \varphi(x) - w$ (resp. $\varphi(x) - w \leq y \leq \varphi(x) + w$, resp. $y > \varphi(x) + w$). We define a mapping $f = f_{\varphi,w} : Q \to [0, 1] \times [0, 1 - 2w]$ (called the *contraction of* Q *along* φ *by* w) as follows:

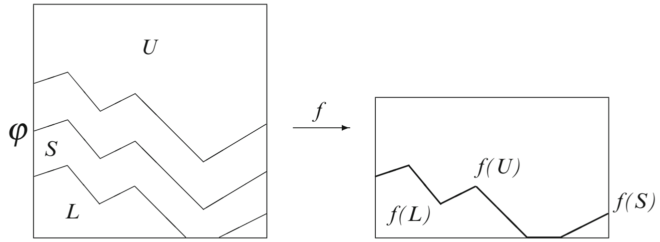


Fig. 1 Contraction along φ

$$f(x, y) = \begin{cases} (x, \min(y, 1 - 2w)) & \text{for } (x, y) \in L \\ (x, \max(y - 2w, 0)) & \text{for } (x, y) \in U \\ (x, \max(\min(\varphi(x) - w, 1 - 2w), 0)) & \text{for } (x, y) \in S. \end{cases}$$

Lemma 3. *With a, w, φ as above, $f = f_{w,\varphi}$ is a 1-Lipschitz mapping.*

Proof: Let $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$ be two points in Q and $q_1 = (u_1, v_1)$ and $q_2 = (u_2, v_2)$ their f-images, resp., and suppose that $v_1 \geq v_2$. We must show $d_\infty(p_1, p_2) \geq d_\infty(q_1, q_2)$. First, if $|u_1 - u_2| \geq |v_1 - v_2|$, then $d_\infty(q_1, q_2) = |u_1 - u_2| = |x_1 - x_2| \leq d_\infty(p_1, p_2)$, as f does not alter the x-coordinate. On the other hand, suppose that $|u_1 - u_2| < |v_1 - v_2| = v_1 - v_2$; then $q_1 \in f(S) \cup f(L)$ implies $q_2 \in f(L)$, as $f(S)$ is a graph of the 1-Lipschitz function $x \mapsto \max(0, \min(1 - 2w, \varphi(x) - w))$. Therefore f decreases the

y-coordinate of p_1 by at least as much as the y-coordinate of p_2, and we have $d_\infty(q_1,q_2) = v_1 - v_2 \le y_1 - y_2 \le d_\infty(p_1,p_2)$. □

We also note that the mapping f in the above lemma does not increase the measure of sets.

Let $E \subseteq \mathbb{R}^2$ be as in Theorem 1. By the Density Theorem for the Lebesgue measure, one can choose a square Q_0 with side a_0 and a compact subset $K \subseteq E \cap Q_0$ with $\lambda^2(K) \ge 0.99\lambda^2(Q_0)$. For notational convenience, let us assume $Q_0 = [0,a_0]^2$. Let $f_0 : \mathbb{R}^2 \to Q_0$ be a 1-Lipschitz retraction of the plane onto Q_0.

We derive Theorem 1 from the following lemma.

Lemma 4. *Let $Q = [0,a]^2$ be an axis-parallel square, let $K \subseteq Q$ be compact with $\lambda^2(K) = \lambda^2(Q)(1-\varepsilon)$, $0 < \varepsilon \le 0.01$. Then one can find a 1-Lipschitz mapping g of Q onto $Q' = [0,a']^2$ such that*

(i) $a' \ge a(1-\sqrt{\varepsilon})$,
(ii) $d_\infty(p, g(p)) \le a\sqrt{\varepsilon}$ *for any* $p \in Q$,
(iii) $\lambda^2(K') \ge \lambda^2(Q')(1-0.9\varepsilon)$, *where* $K' = g(K)$.

Assuming this lemma, the proof of Theorem 1 is finished by an inductive construction as follows. Suppose that we have already constructed a square $Q_i = [0,a_i]^2$ and a 1-Lipschitz mapping f_i of $\mathbb{R}^2$ onto Q_i in such a way that $\lambda^2(K_i) \ge \lambda^2(Q_i)(1-\varepsilon_i)$, where $K_i = f_i(K)$. We apply Lemma 4 with $Q = Q_i$, $a = a_i$, $K = K_i$, and we get a 1-Lipschitz mapping $g_i : Q_i \to Q_{i+1} = [0,a_{i+1}]^2$ with properties as in the lemma. We set $f_{i+1} = g_i \circ f_i$ and continue by the next step of the construction.

From the Lemma we get $\varepsilon_{i+1} \le 0.9\varepsilon_i$ and $a_{i+1} \ge a_i(1-\sqrt{\varepsilon}_i)$. We thus have $\varepsilon_i \le 0.01(0.9)^i$, $a_i \ge a_0 \prod_{j=0}^{i-1}(1-0.95^j/10)$, and straightforward estimates give $a = \lim_{i\to\infty} a_i \ge 0.1a_0$.

For every point $p \in \mathbb{R}^2$, we set $f(p) = \lim_{i\to\infty} f_i(p)$. Condition (ii) of Lemma 4 guarantees that the limit exists. The mapping f is clearly 1-Lipschitz, and its image is contained in the square $Q = [0,a]^2$. The set $f(K)$ is compact and it is easily seen that it is also dense in Q, and thus $f(K) = Q$. The proof of Theorem 1 is finished by rescaling f so that it maps K onto $[0,1]^2$.

Proof of Lemma 4: Let G be the set of squares with side a/n from an $n \times n$ grid covering Q. Let $B_0 \subseteq G$ be the set of squares $s \in G$ with $\lambda^2(s \setminus K) \ge \frac{15}{16}\lambda^2(s)$. Choose n so large that the squares of B_0 contain at least 1/2 of the measure of $Q \setminus K$; this is possible by elementary measure theoretic considerations. Then $|B_0| \ge \varepsilon n^2/2$.

The required mapping g will again be constructed inductively, this time in a finite number t of steps. Let us explain the first step of the construction. Let $\varphi_0 : [0,a] \to [0,a]$ be a suitable 1-Lipschitz function; its choice is the heart of the proof. The requirement on φ_0 is that either the graph of φ_0 (i.e. the set $\{(x,\varphi_0(x));\ x \in [0,a]\}$) or the graph of φ_0 rotated by $\pi/2$ (i.e. the set $\{(\varphi_0(y), y);\ y \in [0, a]\})$ contain the centers of possibly many squares of B_0, namely at least $\sqrt{|B_0|}$ such centers. We will explain the construction assuming that the first case occurs (the graph of φ_0 contains at least $\sqrt{|B_0|}$ centers). The second case is handled symmetrically, by exchanging the role of the coordinate axes.

Let $D_0 \subseteq B_0$ be the squares whose centers are contained in the graph of φ_0, $|D_0| \geq \sqrt{|B_0|}$. We let $\bar{g}_0 : Q \to Q$ be the contraction of Q along φ_0 by a/n (see the definition above Lemma 3). We define an auxiliary mapping $h_0 : [0, a] \times [0, a(n-2)/n] \to [0, a(n-2)/n]^2$, acting as the identity on the first $n-2$ columns of the grid and contracting the last two columns to a vertical segment, and we set $g_0 = h_0 \circ \bar{g}_0$. Thus, the range of g_0 is the square $Q_1 = [0, a(n-2)/n]^2$.

It is easily checked that the mapping g_0 maps each square of the grid G into some square of G in Q_1. All squares in D_0 are contracted to pieces of Lipschitz curves.

We define $B_1 = \{s \in G;\ \exists s' \in B_0 : g_0(s') \subseteq s \text{ and } \lambda^2(g_0(s')) > 0\}$. Thus, we have $|B_1| \leq |B_0| - |D_0|$, so at least $|D_0|$ squares of B_0 are “killed” by g_0.

Similarly we define mappings $g_1, g_2, \ldots$. Having defined the mapping g_{i-1}, we put $B_i = \{s \in G;\ \exists s' \in B_{i-1} : g_{i-1}(s') \subseteq s \text{ and } \lambda^2(g_{i-1}(s')) > 0\}$ and we construct an appropriate 1-Lipschitz mapping $g_i : Q_i \to Q_{i+1}$, where $Q_i = [0, a(n-2i)/n]^2$. The mapping g_i kills at least $\sqrt{|B_i|}$ squares of B_i, in the sense that $|B_i| - |B_{i+1}| \geq \sqrt{|B_i|}$.

Let t be the first index such that $|B_0| - |B_t| \geq \varepsilon n^2/5$. For $i < t$, we thus have $|B_0| - |B_i| < \varepsilon n^2/5$ (and therefore $|B_i| > \varepsilon n^2/4$). In particular, we get $\varepsilon n^2/5 > |B_0| - |B_t| \geq \sum_{i=0}^{t-2} \sqrt{|B_i|} \geq (t-1)n\sqrt{\varepsilon}/2$, which implies that $t < \frac{2}{5}n\sqrt{\varepsilon} + 1 \leq n\sqrt{\varepsilon}/2$ (since we may assume that $n\sqrt{\varepsilon}$ is large).

By composing the mappings $g_0, \ldots, g_{t-1}$, we obtain a 1-Lipschitz mapping g of Q onto $Q' = [0, a']^2$, with $a' = a(1 - 2t/n) \geq a(1 - \sqrt{\varepsilon})$. Under this mapping, at least $\varepsilon n^2/5$ of the squares of B_0 have images of measure 0, and these squares contain at least $\frac{15}{16}(a^2/n^2)(\varepsilon n^2/5) = \frac{3}{16}\varepsilon a^2$ of the measure of $Q \setminus K$. Hence $\lambda^2(g(Q \setminus K)) \leq \frac{13}{16}\varepsilon a^2$ and $\lambda^2(g(K)) \geq \lambda^2(Q' \setminus g(Q \setminus K)) \geq a'^2 - \frac{13}{16}\varepsilon a^2$. Using the assumption $\varepsilon \leq 0.01$, we have $a' \geq a(1 - \sqrt{\varepsilon}) \geq 0.9a$, and thus $\lambda^2(g(K)) \geq a'^2(1 - \frac{10}{9} \cdot \frac{13}{16}\varepsilon) \geq a'^2(1 - 0.9\varepsilon)$ as required.

It remains to show how to choose the mappings φ_i, whose graphs or rotated graphs pass through sufficiently many center points of the squares of B_i. Call a set $D \subseteq \mathbb{R}^2$ 1-*Lipschitz in the y-coordinate* (resp. *in the x-coordinate*) if $|y_1 - y_2| \leq |x_1 - x_2|$ (resp. $|x_1 - x_2| \leq |y_1 - y_2|$) for any two points $(x_1, y_1), (x_2, y_2) \in D$. It is easy to see that if D is a finite set which is 1-Lipschitz in the y-coordinate, then there is a 1-Lipschitz function whose graph contains D. Hence it suffices to establish the following

Lemma 5. *Let P be an m point set in the plane, $k = \lceil \sqrt{m}\, \rceil$. Then there exists either a k point subset of P 1-Lipschitz in the x-coordinate or a k point subset of P 1-Lipschitz in the y-coordinate.*

Proof: Call a set C in the plane a *nondecreasing chain* if for any two points $(x_1, y_1), (x_2, y_2) \in C$ with $x_1 \leq x_2$ we have $y_1 \leq y_2$; a *nonincreasing chain* is defined analogously with an opposite inequality for the y-coordinates. A lemma due to Erdös and Szekerés [4] can be stated as follows: Any m point set in the plane contains a k point nondecreasing chain or a k point nonincreasing chain. To see the relevance to the above lemma, rotate the coordinate system by $\pi/4$; then nondecreasing chains become precisely 1-Lipschitz subsets in the x-coordinate and nonincreasing chains become 1-Lipschitz subsets in the x-coordinate. □

For a direct application of the above method for Laczkovich's problem in dimension d, one would need the following: Given an m-point set P in $\mathbb{R}$, find a subset S of size $cm^{1-1/d}$ ($c > 0$ a constant) which is 1-Lipschitz in the x_i-coordinate for some $i = 1, 2, \ldots, d$. Here a subset $S \subseteq P$ is C-Lipschitz in the x_i-coordinate if for any two points $a = (a_1, \ldots, a_d), b = (b_1, \ldots, b_d) \in S$ one has $|a_i - b_i| \leq C \max_{i \neq j}(|a_j - b_j|)$. As noted by Gábor Tardos, already for $d = 3$ such a large 1-Lipschitz subset need not exist; see [7]. Indeed, let $P \subset \mathbb{R}^3$ be the set $\{(i, j, i + i);\ i, j = 1, 2, \ldots \sqrt{m}\}$; it is not difficult to check that there is no subset of size exceeding $O(\sqrt{m})$ which is 1-Lipschitz in one of the coordinates. Still, it is possible that there always exists a subset of size $cm^{1-1/d}$ which is C-Lipschitz in one of the coordinates, C a constant. This problem looks interesting in its own right, although it is not completely clear whether a positive answer would solve Laczkovich's problem.

Update from the year 2012. Alberti, Csörnyei, and Preiss gave a negative solution to the last stated problem (apparently still unpublished; see [1]). However, the following variant of the problem, from [2], still remains open, and a positive answer would also have interesting analytic consequences: Is there a constant C such that for every n^d-point set in $\mathbb{R}^d$ there exists a coordinate system and a subset of at least n^{d-1} points that is C-Lipschitz in one of the coordinates? Further significant progress regarding Laczkovich's problem and related questions was announced by Csörnyei and Jones (see [2]).

3. Proof of Theorem 2

In the above proof of Theorem 1, the only part where an arbitrarily large part of the measure of E is wasted is the initial application of the Density Theorem for the choice of the square Q_0 with $\lambda^2(Q_0 \cap E) > 0.99\lambda^2(Q_0)$. In the rest of the proof, $Q_0 \cap E$ is then mapped onto a square with side at most 10 times smaller than the side of Q_0. Thus, it suffices to prove the following lemma:

Lemma 6. *For any $\varepsilon_0 \in (0, 1)$ and any $E \subset \mathbb{R}^2$ with $\lambda^2(E) = 1$, there exists a 1-Lipschitz mapping $f : \mathbb{R}^2 \to Q = [0, c]^2$ with $\lambda^2(f(E)) \geq (1 - \varepsilon_0)\lambda^2(Q)$, where $c = c(\varepsilon_0) > 0$ only depends on ε_0.*

Proof: The proof is probabilistic, using the so-called Second Moment Method; see, e.g., [3].

Fix $c = \sqrt{\varepsilon_0}/7$. For simplicity, we will construct a 2-Lipschitz mapping onto $[0,c]^2$; the Lemma is then obtained by taking $c/2$ instead of c and rescaling. Choose an integer N large enough so that some axis-parallel square S with side $3cN$ contains at least $9/10$ of the measure of E. Let G be an $N \times N$ grid consisting of $3c \times 3c$ axis-parallel squares; the squares of G are denoted by s_{ij}, $i = 1, 2, \ldots N$ denoting the row index and $j = 1, 2, \ldots, N$ the column index. Let M_{ij} denote the middle $c \times c$ square in s_{ij}, and set $M = \bigcup_{i,j=1}^{N} M_{ij}$. Since 9 translational copies of the set M cover the square S, it is possible to place the grid G in such a way that $\lambda^2(M \cap E) \geq 1/10$; we assume that such a placement has been fixed.

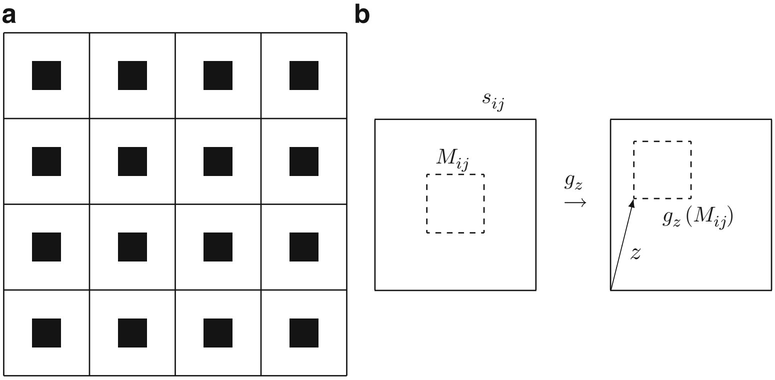


Fig. 2 (**a**) The grid G and the set M; (**b**) the mapping g_z

Consider a grid square s_{ij}. Let $z = z_{ij} \in [0, 2c]^2$ be a vector. We define a mapping $g_z : s_{ij} \to s_{ij}$ as follows. We require that g_z is the identity map on the boundary ∂s_{ij} of s_{ij}, and on M_{ij}, g_z acts as the translation by the vector $z - (c, c)$ (see Fig. 2). With respect to the d_∞ metric, g_z is a 2-Lipschitz mapping on $\partial s_{ij} \cup M_{ij}$. We extend it to a 2-Lipschitz mapping $s_{ij} \to s_{ij}$ (this is possible since for any metric space Y and any $X \subseteq Y$, a C-Lipschitz mapping from X to the plane with the d_∞ metric can be extended to a C-Lipschitz mapping defined on Y; see, e.g., [9]).

For later use, we observe that if z is randomly chosen from the uniform probability distribution on $[0, 2c]^2$, then for any fixed $x \in M_{ij}$, we have

$$\mathrm{Prob}(x \in g_z(E \cap M_{ij})) \geq \frac{\lambda^2(E \cap M_{ij})}{4c^2} \tag{1}$$

(since for each $y \in E \cap M_{ij}$, $g_{x-y+(c,c)}$ sends y to x).

Consider $Z = \{(z_{ij};\ i, j = 1, 2, \ldots, N);\ z_{ij} \in [0, 2c]^2\}$ with the product probability measure, and let $\mathbf{z} \in Z$ be a random element. Let $\bar{g}_{\mathbf{z}} : G \to G$

be the 2-Lipschitz mapping whose restriction on each s_{ij} is $g_{z_{ij}}$. Further we define a 1-Lipschitz mapping $f_1 : G \to [0, 3c]^2$ corresponding to a "folding" of the grid G (as if G were a piece of paper and we wanted to fold it to one little square). Formally, if $p \in s_{ij}$ is a point in G with displacement (x, y) from the lower left corner of s_{ij}, we set

$$f_1(p) = \begin{cases} (x, y) & \text{for } i, j \text{ odd,} \\ (x, 3c - y) & \text{for } i \text{ odd, } j \text{ even,} \\ (3c - x, y) & \text{for } i \text{ even, } j \text{ odd,} \\ (3c - x, 3c - y) & \text{for } i, j \text{ even.} \end{cases}$$

Finally, we let f_0 be the retraction of the plane onto the grid G, and $f_2 : [0, 3c]^2 \to Q$ be the mapping translating the middle $c \times c$ square of $[0, 3c]^2$ onto $Q = [0, c]^2$ and contracting the rest of $[0, 3c]^2$ onto the boundary of Q. Both f_0, f_2 are 1-Lipschitz. We put $f_{\mathbf{z}} = f_2 \circ f_1 \circ \bar{g}_{\mathbf{z}} \circ f_0$. This is a 2-Lipschitz map.

Fix any point $x \in [0, c]^2$, and let X_{ij} be the 0/1 indicator variable for the event $x \in f_{\mathbf{z}}(E \cap M_{ij})$ (this is a random variable depending on the choice of $\mathbf{z}$), and set $X = \sum_{i,j=1}^{N} X_{ij}$. From (1), we get for the expectation of each X_{ij}

$$\mathrm{E}\, X_{ij} \geq \frac{1}{4c^2} \lambda^2(E \cap M_{ij}) \,,$$

and so

$$\mathrm{E}\, X = \sum_{i,j=1}^{N} \mathrm{E}\, X_{ij} \geq \frac{1}{40c^2} > \frac{1}{\varepsilon_0} \,.$$

As the variables X_{ij} are independent, we have

$$\operatorname{Var} X = \sum_{i,j=1}^{N} \operatorname{Var} X_{ij} \leq \sum_{i,j=1}^{N} \mathrm{E}\, X_{ij}^2 = \sum_{i,j} \mathrm{E}\, X_{ij} = \mathrm{E}\, X \,.$$

From Chebyshev's Inequality we thus get

$$\operatorname{Prob}(X = 0) \leq \operatorname{Prob}(|X - \mathrm{E}\, X| \geq \mathrm{E}\, X) \leq \frac{\operatorname{Var} X}{(\mathrm{E}\, X)^2} \leq \frac{1}{\mathrm{E}\, X} < \varepsilon_0 \,.$$

The event $X \neq 0$ means $x \in f_{\mathbf{z}}(M \cap E)$, where x is our fixed point of Q. Considering all x's simultaneously and using Fubini's Theorem, we get that there exists a $\mathbf{z}_0 \in Z$ such that $\lambda^2(f_{\mathbf{z}_0}(E \cap M)) \geq (1 - \varepsilon_0)\lambda^2(Q)$. This concludes the proof of Lemma 6. □

Acknowledgements I would like to thank David Preiss for explaining me his proof, useful discussions and for his hospitality during my visit.

References

1. G. Alberti, M. Csörnyei, D. Preiss: Structure of null sets in the plane and applications, in *European Congress of Mathematics: Stockholm, June 27–July 2, 2004 (A. Laptev ed.)*, Europ. Math. Soc., Zurich 2005, pages 3–22.
2. G. Alberti, M. Csörnyei, D. Preiss: Differentiability of Lipschitz functions, structure of null sets, and other problems, in *Proc. ICM 2010*, vol. III, World Scientific, Hackensack, NJ, pages 1379–1394.
3. N. Alon, J. Spencer, P. Erdös: *The probabilistic method.* Cambridge Univ. Press 1992.
4. P. Erdös, G. Szekerés: A combinatorial problem in geometry. *Compositio Math.* 2(1935) 463–470.
5. M. Laczkovich: Paradoxical decompositions using Lipschitz functions, *Real Analysis Exchange* 17(1991–92), 439–443.
6. D. Preiss, manuscript, 1992.
7. T. Szabó, G. Tardos: A multidimensional generalization of the Erdős–Szekeres lemma on monotone subsequences, *Combinatorics, Probability and Computing* 10(2001) 557–565.
8. N. X. Uy, Removable sets of analytic functions satisfying a Lipschitz condition, *Ark. Mat.* 17(1979), 19–27.
9. J. H. Wells, L. R. Williams: *Embeddings and extensions in analysis*, Springer-Verlag 1975.

A Remark on Transversal Numbers

János Pach

J. Pach (✉)
EPFL, Lausanne, Switzerland

Rényi Institute, Budapest, Hungary
e-mail: pach@cims.nyu.edu

> "What does the Hungarian parrot say?"
> "Log. Log log log log . . . "
> (Riddle. Folklore.)

1. Introduction

In his classical monograph published in 1935, Dénes König [23] included one of Paul Erdős's first remarkable results: an infinite version of the Menger theorem. This result (as well as the König-Hall theorem for bipartite graphs, and many related results covered in the book) can be reformulated as a statement about transversals of certain hypergraphs.

Let H be a hypergraph with vertex set $V(H)$ and edge set $E(H)$. A subset $T \subseteq V(H)$ is called a *transversal* of H if it meets every edge $E \in E(H)$. The *transversal number* $\tau(H)$ is defined as the minimum cardinality of a transversal of H. Clearly, $\tau(H) \geq \nu(H)$, where $\nu(H)$ denotes the maximum number of pairwise disjoint edges of H. In the above mentioned examples, $\tau(H) = \nu(H)$ holds for the corresponding hypergraphs. However, in general it is impossible to bound τ from above by any function of ν, without putting some restriction on the structure of H.

One of Erdős's closest friends and collaborators, Tibor Gallai (who is also quoted in König's book) once said: "I don't care for bounds involving $\log n$'s and $\log\log n$'s. I like exact answers. But Paul has always been most interested in asymptotic results." In fact, this quality of Erdős has contributed a great deal to the discovery and to the development of the "probabilistic method" (see [4, 14]).

The search for "exact answers" (e.g. to the perfect graph conjecture of Berge [7]) has revealed some important connections between transversal problems and linear programming that led to the deeper understanding of the König-Hall-Menger-type theorems. It proved to be useful to introduce another parameter, the *fractional transversal number* of a hypergraph, defined by

$$\tau^{\star}(H) = \min_{t} \sum_{x \in V(H)} t(x),$$

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_34,

where the minimum is taken over all non-negative functions $t : V(H) \longrightarrow \mathbb{R}$ with the property that

$$\sum_{x\in E} t(x) \geq 1 \qquad \text{for every } E \in E(H).$$

Obviously, $\tau(H) \geq \tau^\star(H) \geq \nu(H)$, and $\tau^\star(H)$ can be easily calculated by linear programming. (See [25] and [30].)

At the same time, the probabilistic (or shall we say, asymptotic) approach has also led to many exciting discoveries about extremal problems related to transversals (e.g. Ramsey-Turán-type theorems, property B). It was pointed out by Vapnik and Chervonenkis [32] that in some important families of hypergraphs a relatively small set of randomly selected vertices will, with high probability, be a transversal. They defined the *dimension* of a hypergraph as the size of the largest subset $A \subseteq V(H)$ with the property that for every $B \subseteq A$ there exists an edge $E \in E(H)$ such that $E \cap A = B$. Adapting the original ideas of [32] and [21], it was shown in [22] (see also [27]) that

$$\tau(H) \leq (1 + o(1)) \dim(H)\tau^\star(H) \log \tau^\star(H), \tag{1}$$

as $\tau^\star \longrightarrow \infty$, and that this bound is almost tight.

Ding, Seymour and Winkler [11] have introduced another parameter of a hypergraph, closely related to its dimension. They defined $\lambda(H)$ as the size of the largest collection of edges $\{E_1, \ldots, E_k\} \subseteq E(H)$ with the property that for every pair $(E_i, E_j), 1 \leq i \neq j \leq k$, there exists a vertex x such that $x \in E_i \cap E_j$ but $x \notin E_h$ for any $h \neq i, j$. Combining (1) with Ramsey's theorem, they showed that

$$\tau(H) \leq 6\lambda^2(H)(\lambda(H) + \nu(H)) \binom{\lambda(H) + \nu(H)}{\lambda(H)}^2 \tag{2}$$

holds for every hypergraph H.

As far as we know, Haussner and Welzl [21] were the first to recognize that (1) has a wide range of interesting geometric applications, due to the fact that a large variety of hypergraphs defined by geometric means have low Vapnik-Chervonenkis dimensions.

The aim of this note is to illustrate the power of this approach by two examples. In Sect. 2 we show that (2) easily implies some far-reaching generalizations of results of Erdős and Szekeres [15, 16]. In Sect. 3, we use (2) to extend and to give alternative proofs for some old results of Gyárfás and Lehel (see [19, 20, 24]) bounding the transversal numbers of box hypergraphs.

2. Covering with Boxes

Given two points $p, q, \in \mathbb{R}^d$, let $\mathrm{Box}[p, q]$ be defined as the smallest box containing p and q, whose edges are parallel to the axes of the coordinate system. The following theorem settles a conjecture of Bárány and Lehel [6], who established the first non-trivial result of this kind.

Theorem 1. *Any finite (or compact) set $P \subseteq \mathbb{R}^d$ contains a subset with at most $2^{2^{d+2}}$ elements, $\{p_i | 1 \leq i \leq 2^{2^{d+2}}\}$, such that*

$$P \subseteq \bigcup_{i,j=1}^{2^{2^{d+2}}} \text{Box}[p_i, p_j].$$

Proof. Let H be a hypergraph on the vertex set

$$V(H) = \{\text{Box}[p,q] | p, q \in P\},$$

defined as follows. Associate with each point $r \in P$ the set

$$E_r = \{\text{Box}[p,q] | r \in \text{Box}[p,q]\},$$

and let $E(H) = \{E_r | r \in P\}$.

Clearly, $E_p \cap E_q \neq \emptyset$ for any $p, q \in P$, because $\text{Box}[p,q] \in E_p \cap E_q$. Hence, $\nu(H) = 1$.

According to a well-known lemma of Erdős and Szekeres [15], any sequence of $k^2 + 1$ real numbers contains a monotone subsequence of length $k + 1$. By repeated application of this statement, we obtain that any set of $2^{2^{d-1}} + 1$ points in $\mathbb{R}^d$ has three elements p_i, p_j, p_k with $p_k \in \text{Box}[p_i, p_j]$. This immediately implies that

$$\lambda(H) \leq 2^{2^{d-1}}.$$

Indeed, for any family of more than $2^{2^{d-1}}$ edges $E_{p_1}, \ldots E_{p_\lambda} \in E(H)$, one can choose three distinct indices i, j, k with $p_k \in \text{Box}[p_i, p_j]$, which yields that

$$E_{p_i} \cap E_{p_j} \subseteq E_{p_k}.$$

Thus, we can apply (2) to obtain

$$\tau(H) \leq 6\lambda^2(H) \left(\lambda(H) + 1\right)^3 < 2^{2^{d+2}},$$

and the result follows. □

As was shown in [6], the bound $2^{2^{d+2}}$ in Theorem 1 is nearly optimal.

In fact, the above argument yields a slightly stronger result.

Theorem 2. *Let $P \subseteq \mathbb{R}^d$ by any compact set, and let $\mathcal{B}$ be any family of boxes in parallel position with the property that for any $(\nu+1)$-element subset $P' \subseteq P$ there is a box $B \in \mathcal{B}$ which covers at least two points of $P' (d, \nu \geq 1)$. Then one can choose at most $\binom{2^{2^d}+\nu}{\nu}^5$ members of $\mathcal{B}$ such that their union will cover P.*

In [16], Erdős and Szekeres proved the following.

Lemma 1. *Every set $P \subseteq \mathbb{R}^2$ with at least 2^k elements contains three points p_1, p_2, p_3 such that $\sphericalangle p_1, p_2, p_3 \geq \pi \left(1 - \frac{1}{k}\right)$.*

Our next result, which improves a theorem of Bárány [5], can be regarded as a generalization of Lemma 1.

Theorem 3. *Let d be a positive integer, $\varepsilon > 0$. Every finite (or compact) set $P \subseteq \mathbb{R}^d$ has a subset of at most $2^{(c/\varepsilon)^{d-1}}$ elements, $P' = \{p_1, p_2, \ldots\}$, with the property that for any $p \in P \setminus P'$ there exist $p_i, p_j \in P'$ satisfying*

$$\sphericalangle p_i p p_j \geq \pi - \varepsilon.$$

(Here $c \leq 8$ is a constant.)

Proof. For $d \geq 2$, $\varepsilon > 0$ fixed, let us cover the unit hemisphere centered at $O \in \mathbb{R}^d$ with $(4/\varepsilon)^{d-1}$ spherical $(d-1)$-dimensional simplices $S_1, S_2, \ldots$ such that the diameter of each S_i is at most $\varepsilon/2$. Let $h_{t1}, \ldots, h_{td}$ denote the hyperplanes induced by O and the ($(d-2)$-dimensional) facets of S_t. A parallelotope whose facets are parallel to $h_{t1}, \ldots, h_{td}$, respectively, is called a *box of type* t. The smallest box of type t containing $p, q \in \mathbb{R}^d$ will be denoted by $\mathrm{Box}_t[p,q]$.

For any $p, q \in \mathbb{R}^d$, choose an index t such that pq is parallel to Os for some $s \in S_t$, and let $\mathrm{Box}[p,q] = \mathrm{Box}_t[p,q]$. Notice that if $r \in \mathrm{Box}\ [p,q]$ then $\sphericalangle prq \geq \pi - \varepsilon$.

Just like in the previous proof, define a hypergraph H by

$$\begin{aligned} V(H) &= \{\mathrm{Box}[p,q] | p,q \in P\}, \\ E(H) &= \{E_r | r \in P\}, \end{aligned}$$

where $E_r = \{\mathrm{Box}[p,q] | r \in \mathrm{Box}[p,q]\}$, and observe that it is sufficient to bound the transversal number of H. Clearly, $\nu(H) = 1$.

By the definition of $\lambda(H)$, one can select $\lambda(H) = \lambda$ elements $p_1, \ldots, p_\lambda \in P$ with the property that any two of them is enclosed in a box of some type, which does not cover any other p_k. More precisely, for every $1 \leq i < j \leq \lambda$ there exists $t(i,j) \leq (4/\varepsilon)^{d-1}$ such that

$$\{p_1, \ldots, p_\lambda\} \cap \mathrm{Box}_{t(i,j)}[p_i, p_j] = \{p_i, p_j\}.$$

Obviously, p_i and p_j are two antipodal vertices of $\mathrm{Box}_{t(i,j)}[p_i, p_j]$, and every box has 2^{d-1} pairs of antipodal vertices. Let us color the segments $p_i p_j (1 \leq i < j \leq \lambda)$ with $(4/\varepsilon)^{d-1} 2^{d-1}$ colors according to the value of $t(i,j)$ and to the particular position of the diagonal $p_i p_j$ within $\mathrm{Box}_{t(i,j)}[p_i, p_j]$. It is easy to see that the segments of the same fixed color form a bipartite subgraph of the complete graph K_λ on the vertex set $p_1, \ldots, p_\lambda$. Hence the chromatic number of K_λ,

$$\lambda(H) \leq 2^{(4/\varepsilon)^{d-1} 2^{d-1}},$$

and the result follows from (2). □

It is not hard to see that the bound in Theorem 3 is asymptotically tight, apart from the exact value of c (see [5, 13]).

Combining these observations with an analogue of Turán's theorem for hypergraphs, we immediately obtain the following result related to a problem of Conway, Croft, Erdős and Guy [8].

Corollary 1. *There exists a constant $c > 0$ such that, for any set of n distinct points $p_1, \dots p_n \in \mathbb{R}^d$, the number of triples $i < j < k$ for which $\sphericalangle p_i p_j p_k > \pi - \varepsilon$, is at least $\lfloor n^3 / 2^{(c/\varepsilon)^{d-1}} \rfloor$. Moreover, apart from the value of c, this bound cannot be improved.*

Finally, we mention another straightforward generalization of Lemma 1.

Theorem 4. *Let P be any set of at least $k^{(c/\varepsilon)^{d-1}}$ points in $\mathbb{R}^d$, where c is a suitable constant. Then one can find $p_0, \dots, p_k \in P$ such that they are "almost collinear", i.e., $\sphericalangle p_{i-1} p_i p_{i+1} > \pi - \varepsilon$ for every $i (1 \le i < k)$.*

3. Gallai-Type Theorems

Many problems in geometric transversal theory were motivated by the following famous question of Gallai. Given a family of pairwise intersecting disks in the plane, what is the smallest number of needles required to pierce all of them? (The answer is three. See [9, 10, 12, 17].)

First we show that (2) implies the following result of Gyárfás and Lehel.

Theorem 5 ([**20**]). *For any positive integers k and ν, there exists a number $f = f(k, \nu)$ with the following property. Let H be any finite family of subsets of $\mathbb{R}$ such that each of them can be obtained as the union of at most k intervals. If H has no $\nu + 1$ pairwise disjoint members, then all of its members can be pierced by at most f points.*

Proof. In order to apply (2), we have to bound $\lambda(H)$. Let $E_1, \dots, E_\lambda$ be some members (edges) of H such that, for any $i < j$, $E_i \cap E_j$ has a point x_{ij} which does not belong to any other E_h $(h \ne i, j)$. Write each E_i $(1 \le i \le \lambda)$ as the union of k intervals,

$$E_i = I_{i1} \cup \ldots \cup I_{ik}.$$

If $x_{ij} \in I_{ip} \cap I_{jq}$ for some $i < j$, then (E_i, E_j) is called a pair of type (p, q). (A pair may have several different types.)

It is easy to check that there are no four edges E_i such that all $\binom{4}{2} = 6$ pairs determined by them are of the same type. Thus,

$$\lambda < R_{k^2}(4),$$

where $R_s(t)$ denotes the smallest number R such that any complete graph of R vertices, whose edges are colored with s colors, has a monochromatic complete subgraph of t vertices (cf. [18]). Hence, the theorem is true with

$$f(k,\nu) \le 6\binom{R_{k^2}(4)+\nu}{\nu}^5. \qquad \square$$

Theorem 5 does not generalize to subsets of the plane that can be obtained as the union of k axis-parallel rectangles. Indeed, let $H = \{E_i | 1 \le i \le n\}$, where

$$E_i = \left\{(x,y) \in \mathbb{R}^2 \mid 0 \le x, y \le n \ \text{ and } \ \min(|x-i|, |y-i|) \le \frac{1}{4}\right\}.$$

Then $\nu(H) = 1$, while $\lambda(H) = \tau(H) = n$.

However, one can easily establish the following.

Theorem 6. *Let F be a family of open domains in the plane such that each of them is bounded by a closed Jordan curve, and any two of them share at most two boundary points. Furthermore, let H be a finite set system, whose every element can be obtained by taking the union of at most k members of F. If H has no $\nu+1$ pairwise disjoint elements, then all of its elements can be pierced by at most $g(k,\nu)$ points (where g depends only on k and ν).*

Proof. Pick λ elements (edges) of H,

$$E_i = I_{i1} \cup \ldots \cup I_{ik} \qquad (I_{ip} \in F, 1 \le i \le \lambda, 1 \le p \le k),$$

and suitable points

$$x_{ij} \in (E_i \cap E_j) \setminus \cup_{h \ne i,j} E_h,$$

as in the previous proof. After defining the type of a pair $(E_i, E_j), i < j$ in exactly the same way as above, now one can argue that there are no 6 edges E_i such that all the $\binom{6}{2} = 15$ pairs determined by them have the same type (p,q). Assume, for contradiction, that e.g. $E_1, \ldots, E_6$ satisfy this condition for some $p \ne q$. Then any I_{ip} $(1 \le i \le 3)$ and any I_{jq} $(4 \le j \le 6)$ have a common interior point (x_{ij}) which is not covered by any other E_k $(k \ne i,j)$. We can conclude (by tedious case analysis) that there exist pairwise disjoint connected open subsets $I'_{ip} \subseteq I_{ip}$ $(1 \le i \le 3)$, $I'_{jq} \subseteq I_{jq}$ $(4 \le j \le 6)$ such that every I'_{ip} and I'_{jq} share a common boundary segment. This contradicts Kuratowski's theorem on planar maps. The case $p = q$ can be treated similarly.

Thus, $\lambda < R_{k^2}(6)$ and the result follows. We could also apply Theorem 1.1 of Sharir [31] to deduce $\lambda < R_{k^2}(c)$ with a much larger constant $c > 6$. $\square$

Theorem 6 can be applied to the family F_C of all homothetic copies of a convex set C in the plane. The special case when C is a convex polygon with a bounded number of sides was settled by Gyárfás [19]. (An easy compactness argument shows that C does not need to be strictly convex.)

For any hypergraph H and for any integer $t \ge 1$, let $\nu_t(H)$ denote the maximum number of (not necessarily distinct) edges of H such that every vertex is contained in at most t of them. Furthermore, let $\lambda_t(H)$ be the size of the largest collection of edges $\{E_i | i \in I\} \subseteq E(H)$ with the property that for any t-tuple $J \subseteq I$ there exists $x_J \in V(H)$ such that

$$x_J \in \left(\bigcap_{i \in J} E_i \right) \setminus \left(\bigcup_{i \notin J} E_i \right).$$

Clearly, $\nu_1(H) = \nu(H)$ and $\lambda_2(H) = \lambda(H)$.

Ding, Seymour and Winkler [11] have established an upper bound for $\tau(H)$ in terms of $\nu_t(H)$ and $\lambda_{t+1}(H)$, for any fixed $t \geq 1$. Applying their result with $t = 2$, we obtain the following generalization of Theorem 5 for the plane.

Theorem 7. *Let H be a finite family of open sets in the plane such that*

(i) Every member of H is bounded by at most k closed Jordan curves;
(ii) Any two distinct members of H have at most ℓ boundary points in common.

Assume that among any $\nu + 1$ members of H there are three with non–empty intersection. Then all members of H can be pierced by at most $f(k, \ell, \nu)$ points, where f does not depend on H.

In higher dimensions we obtain e.g. the following result.

Theorem 8. *Let H be a finite family of not necessarily connected polyhedra in $\mathbb{R}^d (d \geq 2)$. Assume that every member of H has at most k vertices, and that among any $\nu + 1$ members of H there are $d + 1$ whose intersection is non-empty. Then all members of H can be pierced by at most $g(d, k, \nu)$ points, where g does not depend on H.*

The special case of Theorem 8, when every member of H is the union of a bounded number of axis-parallel boxes, was proved by Lehel [24].

4. Concluding Remarks

Alon, Brightwell, Kierstead, Kostochka, and Winkler [1] further analyzed the hypergraph H defined in the proof of Theorem 1. Using the same corollary of the Erdős-Szekeres theorem [15] as we did, they gave an upper bound on $\tau^*(H)$. Using (1) rather than (2), they obtained a slight improvement on Theorem 1: Instead of a subset of size $2^{2^{d+2}}$, there exists a subset of size

$$2^{2^d + d + \log d + \log\log d + O(1)}$$

with the required property. Here log stands for the natural logarithm.

Pálvölgyi and Gyárfás [28] fine-tuned the above arguments and further improved the bound to

$$2^{2^{d-1}+\log d+\log\log d+O(1)}.$$

Many related transversal problems concerning hypergraphs defined by geometric means and hypergraphs formed by the neighborhoods of the vertices of a graph were studied in [1–3] and [26], respectively. In most cases, it turns out that these hypergraphs or some others derived from them have small λ or small VC-dimension. Therefore, one can apply the estimates (1) or (2).

References

1. N. Alon, G. Brightwell, H. A. Kierstead, A. V. Kostochka, and P. Winkler: Dominating sets in k-majority tournaments, Journal of Combinatorial Theory B 96 (2006), 374–387.
2. N. Alon, G. Kalai, J. Matoušek, and R. Meshulam: Transversal numbers for hypergraphs arising in geometry, Adv. in Appl. Math. 29 (2002), 79–101.
3. N. Alon and D. J. Kleitman: Piercing convex sets and the Hadwiger–Debrunner (p,q)–problem, Adv. Math. 96 (1992), 103–112.
4. N. Alon and J. Spencer: The Probabilistic Method, J. Wiley Interscience, New York, 1991.
5. I. Bárány: An extension of the Erdős–Szekeres theorem on large angles, Combinatorica 7 (1987), 161–169.
6. I. Bárány and J. Lehel: Covering with Euclidean boxes, European J. Combinatorics 8 (1987), 113–119.
7. C. Berge: Graphs and Hypergraphs, North–Holland, 1982.
8. J. H. Conway, H. T. Croft, P. Erdős and M. J. T. Guy: On the distribution of values of angles determined by coplanar points, J. London Math. Soc. (2) 19 (1979), 137–143.
9. L. Danzer: Zur Lösung des Gallaischen Problems über Kreisscheiben in der euklidischen Ebene, Studia Sci. Math. Hung. 21 (1986), 111–134.
10. L. Danzer, B. Grünbaum and V. Klee: Helly's theorem and its relatives, in: Convexity, Proc. Symp. Pure Math., Vol. 7, Amer. Math. Soc., Providence, 1963, 100–181.
11. G. Ding, P. Seymour and P. Winkler: Bounding the vertex cover number of a hypergraph, Combinatorica 14 (1994), 23–34.
12. J. Eckhoff: Helly, Radon and Carathéodory type theorems, in: Handbook of Convex Geometry (P. Gruber, J. Wills, eds.), North–Holland, Amsterdam, 1993, 389–448.
13. P. Erdős and Z. Füredi: The greatest angle among n points in the d–dimensional Euclidean space, Annals of Discrete Mathematics 17 (1983), 275–283.
14. P. Erdős and J. Spencer: Probabilistic Methods in Combinatorics, Akadémiai Kiadó, Budapest and Academic Press, New York, 1974.
15. P. Erdős and G. Szekeres: A combinatorial problem in geometry, Compositio Math. 2 (1935), 463–470.
16. P. Erdős and G. Szekeres: On some extremum problems in elementary geometry, Ann. Univ. Sci. Budapest. Eötvös, Sect. Math. III–IV (1960–61), 53–62.
17. J. Goodman, R. Pollack and R. Wenger: Geometric transversal theory, in: New Trends in Discrete and Computational Geometry (J. Pach, ed.), Springer–Verlag, Berlin, 1993, 163–198.
18. R. Graham, B. Rothschild and J. Spencer: Ramsey Theory, J. Wiley and Sons, New York, 1980.

19. A. Gyárfás: A Ramsey–type theorem and its applications to relatives of Helly's theorem, Periodica Math. Hung. 3 (1973), 261–270.
20. A. Gyárfás and J. Lehel: A Helly–type problem in trees, Coll. Math. Soc. J. Bolyai 4. Combinatorial Theory and its Applications, North–Holland, Amsterdam, 1969, 571–584.
21. D. Haussler and E. Welzl: ε–nets and simplex range queries, Discrete Comput. Geometry 2 (1987), 127–151.
22. J. Komlós, J. Pach and G. Woeginger: Almost tight bounds for epsilon–nets, Discrete Comput. Geometry 7 (1992), 163–173.
23. D. König: Theory of Finite and Infinite Graphs, Birkhäuser Verlag, Basel – Boston, 1990.
24. J. Lehel: Gallai–type results for multiple boxes and forests, Europ. J. Combinatorics 9 (1988), 113–120.
25. L. Lovász: Normal hypergraphs and the perfect graph conjecture, Discrete Math. 2 (1972), 253–267.
26. T. Luczak and S. Thomassé: Coloring dense graphs via VC-dimension, arXiv:1007.1670.
27. J. Pach and P. Agarwal: Combinatorial Geometry, J. Wiley and Sons, New York, 1995.
28. D. Pálvölgyi and A. Gyárfás: Domination in transitive colorings of tournaments, manuscript, 2012.
29. F. P. Ramsey: On a problem of formal logic, Proc. London Math. Soc. 30 (1930), 264–286.
30. A. Schrijver: Linear and Integer Programming, J. Wiley and Sons, New York, 1986.
31. M. Sharir: On k–sets in arrangements of curves and surfaces, Discrete Comput. Geom. 6 (1991), 593–613.
32. V. N. Vapnik and A. Ya. Chervonenkis: On the uniform convergence of relative frequencies of events to their probabilities, Theory Probab. Appl. 16 (1971), 264–280.

In Praise of the Gram Matrix

Moshe Rosenfeld

M. Rosenfeld (✉)
Department of Computer Science, Pacific Lutheran University,
Tacoma, WA 98447, USA
e-mail: moishe@u.washington.edu

Summary. We use the Gram matrix to prove that the largest number of points in R^d such that the distance between all pairs is an odd integer (the square root of an odd integer) is $\leq d+2$ and we characterize all dimensions d for which the upper bound is attained. We also use the Gram matrix to obtain an upper bound for the smallest angle determined by sets of n lines through the origin in R^d.

1. Introduction

An $n \times n$ symmetric matrix M is positive semi-definite if $\langle Mx, x\rangle \geq 0$ for all vectors $x \in E^n$. Equivalently, M is positive semi-definite if it is symmetric and its eigenvalues are nonnegative. The Gram Matrix (Grammian) of the set of vectors $\{u_l, \dots, u_n\} \subset E^d$ is the $n \times n$ matrix M defined by $M = (\langle u_i, u_j\rangle)$. This matrix is a positive semi-definite matrix and its rank is the dimension of the subspace spanned by $\{u_1, \dots, u_n\}$. It is well known that if N is a positive semi-definite matrix of order $n \times n$ and rank d, then N is the Grammian of a set of n vectors $\{u_1, \dots, u_n\} \subset E^d$. In other words, $N = (\langle u_i, u_j\rangle)$. This bi-directional relation between vectors, their inner products and the Grammian has provided a powerful tool for solving problems in combinatorics and combinatorial geometry, it plays a central role in the extensive work of Seidel on equiangular lines, it was used in [4] to solve a problem of L. Lovász where they proved that $\sqrt[3]{2n^2} \geq \max \| \sum_{i=1}^{n} u_i \| \geq c\frac{\sqrt[3]{n^2}}{\sqrt{\ln n}}$, the max taken over all families of n almost orthogonal unit vectors in R^n, that is all n tuples of vectors such that among any three vectors there is at least one orthogonal pair, it was used in [9] to answer a question of P. Erdös showing that the maximum size of a set of almost orthogonal lines in R^d is $2d$ and the list can go on and on. One could fill a large volume exploring the many applications of the Gram matrix. The work in this note was motivated by the following simple and attractive problem that appeared in the Nov. 1993 54th W. L. Putnam Mathematical Competition (problem B-5): Can four points in the plane have pairwise odd integral distances? The answer is NO and the Gram matrix provides a very short proof of that. Surprisingly, we cannot do much better if we relax the distance requirement and permit points to have distances whose square is an odd integer. We prove that if the square of all distances among

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_35,

pairs of n points in R^d is an odd integer then $n \leq d+2$. We show that the upper bound is attained if and only if $d = 2 \mod 4$. A slight modification of this proof yields also an alternative proof to a theorem of R. L. Graham, B. L. Rothschild and E. G. Strauss [3] that $d+2$ points in R^d with odd integral distances exist if and only if $d = 14 \mod 16$. We conclude by using the Gram matrix to obtain an upper bound for the smallest angle determined by sets of n lines through the origin in R^d. Cases for which this upper bound is attained are discussed and also some related open problems. In Sect. 2 we discuss distances among points and related open problems and in Sect. 3 we discuss angles among lines.

2. Odd Distances Among Points in R^d

Theorem 1. *Let $p_1, \ldots, p_n$ be n points in R^d such that $\|p_i - p_j\|^2 = 1 \mod 2$ if $i \neq j$. Then*

$$n \leq \begin{cases} d+2 & if d = 2 \mod 4 \\ d+1 & if d \neq 2 \mod 4 \end{cases}$$

and this is best possible.

Proof. Let $u_k = p_k - p_1$ and let $M = (\langle u_k, u_j \rangle)$ be the Gram Matrix of the vectors $\{u_k\}$. M is a positive semi-definite matrix of rank $\leq d$ and order $n-1$, hence if $n - 1 > d \det(M) = 0$ and therefore $\det(2M) = 0$. From $2\langle u_i, u_j \rangle = \|u_i\|^2 + \|u_i\|^2 - \|u_i - u_j\|^2$ and $\|u_i - u_j\|^2 = \|p_i - p_j\|^2 = 1 \mod 2$ we deduce that $2\langle u_i, u_j \rangle$ is an odd integer. From the above discussion it follows that:

$$A = 2M \mod 4 = \begin{pmatrix} 2 & a_{1,2} & a_{1,3} \ldots a_{1,n+1} \\ a_{1,2} & 2 & a_{2,3} \ldots a_{2,n+1} \\ \cdot & \cdot & \cdot \ \ldots \ \cdot \\ \cdot & \cdot & \cdot \ \ldots \ \cdot \\ a_{n+1,1} & a_{n+1,2} & \cdot \ \ldots \ 2 \end{pmatrix} \quad (a_{ij} = a_{ji} = \pm 1).$$

Let $A_{ij}(x)$ be the matrix derived from the matrix A by replacing a_{ij} and a_{ji} by x. Consider the quadratic $p(x) = \det(A_{ij}(x)) = ax^2 + bx + c$. Since all entries in A are integers clearly, a, b and c are also integers. Furthermore, since $A_{ji}(x)$ is symmetric we also have $b = 2k$. Hence $p(1) - p(-1) = 4k$. This means that by replacing any pair of symmetric -1's by 1's the value of the determinant mod 4 remains unchanged or $\det(A) \mod 4 = \det(J + I) \mod 44$ (the matrix J is the square matrix with $J_{ij} = 1$). Clearly 1 is an eigenvalue of $J + I$ with multiplicity $n - 2$ and since the sum of the entries in each row is n, n is the remaining eigenvalue. Therefore $\det(A) \mod 4 = \det(J + I) \mod 4 = n \mod 4$. As noted above, $\det(A) = 0$ if $n > d + 1$ hence we must have $n = 0 \mod 4$. If $n > d + 3$ and if we had n points in R^d

with mutually odd integral squared distances then $n = 0 \mod 4$. But then removing one of the points will yield a set of $n-1$ such points in R^d while $n-1 \neq 0 \mod 4$ and $n-1 > d+1$ which is impossible. Hence $n \leq d+2$. It remains to show that for all $d = 2 \mod 4$ it is possible to construct $d+2$ points in R^d with mutually odd integral squared distances. The following construction using the Gram matrix will do the job.

Let $d = 4k+2$ and let A be a matrix of order $4k+3$ with $A_{ii} = 2k+1$ and $A_{ij} = -\frac{1}{2} i \neq j$. Since $A - (2k+\frac{3}{2})I = -\frac{1}{2}J$, $(2k+\frac{3}{2})$ is an eigenvalue of A and since J has rank 1 the multiplicity of this eigenvalue is $4k+2$. The sum of the entries in each row is zero, hence 0 is the remaining eigenvalue of A. This implies that $\text{rank}(A) = 4k+2 = d$ that A is positive semi-definite and can be written as: $A = M \cdot M^{tr}$ where M is a $(4k+3) \times (4k+2)$ matrix. The rows of M will determine $4k+3$ points $\{p_i\}$ in R^d such that: $\|p_i\|^2 = 2k+1$ and $\|p_i - p_j\|^2 = \|p_i\|^2 + \|p_j\| + 1 = 4k+3$. These points together with the origin yield $4k+4 = d+2$ points as claimed. Geometrically, the $d+2$ points consist of the $d+1$ points of a regular simplex with side $\sqrt{4k+3}$ and its center. To see this, recall that the length of the edge of the regular simplex inscribed in the unit sphere in R^d is $\sqrt{2+2/d} = \sqrt{\frac{4k+3}{2k+1}}$ if $d = 4k+2$. Hence if we inflate the regular unit simplex by $\sqrt{2k+1}$ the distance from the center to the vertices will be $\sqrt{2k+1}$ and the distance between the other vertices will be $\sqrt{4k+3}$. □

A slight modification of the above proof yields an alternative proof to the following theorem of Graham, Rothschild and Strauss [3].

Theorem 2. *For the existence of $n+2$ points in R^n so that the distance between any two of them is an odd integer it is necessary and sufficient that $(n+2) = 0 \mod 16$.*

Proof. As in the previous proof, let $\{p_0, \ldots, P_{n+1}\}$ be $n+2$ points in R^n with mutual distances $\|p_i - p_j\| = 1 \mod 2$. Let $u_i = p_i - p_0$, $i = 1, \ldots, n+1$. The matrix $M = (\langle u_i, u_j \rangle)$ has rank $\leq d$ and order $n+1$. Hence if $n \geq d$ we have:

$$\det(M) = 0 \tag{1}$$

$$\det(M) = 0 \Rightarrow \det(2M) = 0 \Rightarrow \det(2M) \mod 16 = 0. \tag{2}$$

Since $\|u_i, u_j\| = \|p_i, p_j\| = 1 \mod 2$ we have:

$$\|u_i, u_j\|^2 = 1 \mod 8 \tag{3}$$

$$2\langle u_i, u_j \rangle = \|u_i\|^2 + \|u_j\|^2 - \|u_i - u_j\|^2 = 1 \mod 8. \tag{4}$$

Since $\|u_i\| = 1 \mod 2$ we have:

$$2\langle u_i, u_i \rangle^2 = 2\|u_i\|^2 = 2 \mod 16. \tag{5}$$

From (4) and (5) we get:

$$2M \quad \bmod 16 = \begin{pmatrix} 2 & a_{1,2} & a_{1,3} \dots & a_{1,n+1} \\ a_{2,1} & 2 & a_{2,3} \dots & a_{2,n+1} \\ \cdot & \cdot & \cdot \ \dots & \cdot \\ \cdot & \cdot & \cdot \ \dots & \cdot \\ a_{n+1,1} & a_{n+1,2} & \cdot \ \dots & 2 \end{pmatrix}$$

$$= A \quad (\text{where } a_{ij} = a_{ji} = 1 \text{ or } 9).$$

Again we let $A_{ij}(X)$ be the matrix derived from the matrix A by replacing a_{ij} and a_{ji} by x and consider the quadratic $p(x) = \det(A_{ij}(x)) = ax^2+bx+c$. As in Theorem 1, a, b and c are integers and $b = 2k$. Hence $p(9) - p(1) = 80a+16k$. This means that replacing any pair of symmetric 9's by 1's the value of the determinant mod 16 is unchanged, in other words, $\det(A) \mod 16 = \det(J + I) \mod 16$.

Since the eigenvalues of $J + I$ are $(n + 2)$ and $(1)^{\{n\}}$ we have:

$$\det(J + I) \mod 16 = (n + 2) \mod 16.$$

Thus $\det(M)$ can be 0 only if $(n + 2) = 0 \mod 16$. Again, the Gram matrix can be used to describe a construction of $d + 2$ points so that the distance between any two of them is an odd integer. As in Theorem 1, it will suffice to construct a positive semi-definite matrix M of order $d + 1$, rank d, with $M_{ii} = (2k_i + 1)^2$ and $M_{ii} + M_{jj} - 2M_{ij} = (2k_{ij} + 1)^2$. For $d = 16k - 2$ let M be the matrix of order $16k - 1$ shown below.

$$M = \begin{pmatrix} (4k-1)^2 & \frac{(4k-1)^2}{2} & \cdot & \cdot\cdot & \frac{(4k-1)^2}{2} \\ \frac{(4k-1)^2}{2} & (6k-1)^2 & \frac{8k^2-8k+1}{2} & \cdot\cdot & \frac{8k^2-8k+1}{2} \\ \cdot & \frac{8k^2-8k+1}{2} & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\cdot & \cdot \\ \frac{(4k-1)^2}{2} & \frac{8k^2-8k+1}{2} & \cdot & \cdot\cdot & (6k-1)^2 \end{pmatrix}$$

A simple but tedious computation shows that $(8k - 1)$ times the top row is the sum of the remaining $16k-2$ rows. Hence 0 is an eigenvalue of M. Clearly, $((6k - 1)^2 - \frac{8k^2-28k+1}{2})$ is an eigenvalue with multiplicity $16k - 3$ and using the trace we can see that the remaining eigenvalue is also positive. Hence M is positive semi-definite of rank $16k - 2$. Since $2(6k - 1)^2 - (8k^2 - 8k + 1) = (8k - 1)^2$ and $(6k - 1)^2 + (4k - 1)^2 - (4k - 1)^2 = (6k - 1)^2 M$ is the desired matrix. The matrix M is the Gram matrix of the vectors determined by the $d + 2$ points constructed in [3]. □

Remark 1. *Unlike the $d + 2$ points constructed in Theorem 1, this set is a 3-distance set. H. Harbroth (private communication) showed that there is no 2-distance set of $d+2$ points in R^d so that the two distances are odd integers. On the other extreme it might be interesting to construct $d + 2$ points so that all distances (squared distances) are distinct odd integers. Both theorems raise* *some interesting questions already in the plane. If we define a graph whose vertices are the points in the plane and connect two vertices by an edge if their distance (distance squared) is an odd integer then this graph does not contain a K_4 (K_5). Do either of the two graphs have a finite chromatic number? (Clearly if the odd distance-squared graph has a finite chromatic number so does the odd distance graph). This problem is similar to Nelson's classical problem of coloring the plane so that points at distance* 1 *get distinct colors.*

As noted by P. Erdős, Turàn's theorem implies that the maximum number of distances among n *points in the plane that are odd integers is* $\leq n^2/3$. *On the other hand, one can easily construct* n *points on the line with* $n^2/4$ *odd distances. Erdős asked whether it is possible to construct* n *points in the plane so that more than* $n^2/4$ *of the distances will be an odd integer. This question as well as the similar question with odd squared-distances is also interesting in higher dimensions.*

3. Angles Among Lines in R^d

Let $L = \{L_1, \dots, L_n\} \subset R^d$ be a set of lines through the origin. Let $\alpha(L)$ denote the smallest angle among all angles determined by pairs of distinct lines from L. Let $\alpha(n,d) = \sup\{\alpha(L) | L = \{L_1, \dots, L_n\} \subset R^d\}$.

Theorem 3. *[1]* $\cos\alpha(n,d) \geq \sqrt{\frac{n-d}{d(n-1)}}$.

[2] For each n *and* d *there is a set of lines* $L = \{L_1, \dots, L_n\} \subset R^d$ *such that* $\cos\alpha(L) = \cos\alpha(n,d)$.

Proof. We assume that $n > d$. Let u_i be a unit vector along the line L_i and let M be the Grammian of $\{u_1, \dots, u_n\}$. Note that $\cos^2 a(L) = \max\langle u_i, u_j\rangle^2$. Since M is positive semi-definite of rank $\leq d$ it has at most d nonzero eigenvalues $\{\lambda_1, \dots, \lambda_d\}$, $\lambda_i \geq 0$. Since the u_i's are unit vectors, $M_{ii} = 1$ and therefore:

$$\operatorname{trace}(M) = n = \sum_{i=1}^{d} \lambda_i \tag{6}$$

$$\operatorname{trace}(M^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} \langle u_i, u_j\rangle^2 = \sum_{i=1}^{d} \lambda_i^2. \tag{7}$$

From (6):

$$\sum_{i=1}^{d} \lambda_i^2 \geq d\left(\frac{n}{d}\right)^2 = \frac{n^2}{d}.$$

From (7):

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \langle u_i, u_j\rangle^2 = n + 2\sum_{i=1}^{n}\sum_{j=i+1} \langle u_i, u_j\rangle^2 \geq \frac{n^2}{d}.$$

Or:

$$n(n-1)\max(\langle u_i,u_j\rangle^2) \geq \frac{n^2}{d} - n = \frac{n(n-d)}{d} \Rightarrow \max(\langle u_i,u_j\rangle^2) \geq \frac{n-d}{d(n-1)}.$$

To prove (7) note that $\max\cos^2\alpha(L) = \max\{\max(\langle u_i,u_j\rangle^2)\}$ where the inner maximum is taken over the $\binom{n}{2}$ pairs of distinct vectors and the outer maximum is taken over all n-tuples of unit vectors in E^d (which is a closed bounded set in E^{nd}). Since this is a continuous function the maximum is attained. □

Remark 2. *Note that if equality holds in* (3) *then* $\langle u_i,u_j\rangle^2 = \frac{n-d}{d(n-1)}$ *or the lines form a set of equiangular lines. It is well known that the existence of* n *equiangular lines in* E^d *with angle* θ *is equivalent to the existence of a graph with* n *vertices so that the smallest eigenvalue of its Seidel matrix is* $-\frac{1}{\arccos\theta}$ *and has multiplicity* $n-d$*. Already in* E^3 *it seems that the determination of the exact value of* $\alpha(n,3)$ *may be a very difficult problem. For instance, it is not possible to construct a set* L *of* 5 *equiangular lines in* E^3 *with* $\cos\alpha(L) = \frac{1}{\sqrt{6}}$ *even though we can construct a set of* 5 *equiangular lines. Indeed the only way to do it is by taking* 5 *of the* 6 *diagonals of the icosahedron. The angle determined by any pair of these lines is:* $\arccos\frac{1}{\sqrt{5}}$*. Raphael Robinson (private communication) asked whether* $\alpha(5,3) = \alpha(6,3)$*, or more generally if there are integers* n *for which* $\alpha(n,3) = \alpha(n+1,3)$*? For spherical caps there are a few values for which this is the case, L. Danzer [6] proved that the widest angle for* 11 *spherical caps is the same as for* 12 *(*$\arccos\frac{1}{\sqrt{5}} = 63°26'5.8''$*). He also showed that for* 10 *spherical caps the widest angle is* $\geq 66°8'48.3''$ *hence spherical caps cannot be used to prove that* $\alpha(5,3) = \alpha(6,3)$*, for more information see [7] and [8]. Since the maximum number of equiangular lines in* E^3 *is* 6*, the bound also will not be attained for* $n > 6$*. It will also be interesting to know whether there are values of* $n \geq 5$ *for which the extremal configuration is not unique.*

For $n = d+1$ *we get* $\cos(\alpha(d+1,d)) = \frac{1}{d}$ *and this is always attained by the* $d+1$ *lines connecting the center of the regular simplex to its vertices. There are many sporadic cases where the bound is attained. For instance* $\alpha(16,6) = \frac{1}{3}$ *and there are 16 equiangular lines in* R^6 *with angle* $\arccos(\frac{1}{3})$*;* $\alpha(176,22) = \alpha(276,23) = \frac{1}{5}$ *and in both these cases there are appropriate sets of equiangular lines (see [5]). A particularly attractive case is the case of* $2d$ *lines in* R^d*. In this case* $\cos\alpha(2d,d) = \sqrt{\frac{2d-d}{d(2d-1)}} = \sqrt{\frac{1}{2d-1}}$*. This bound is attained for* $d = 3,5$ *but not for* $d = 4$*. Indeed, there is no Seidel Matrix of order* 8 *with smallest eigenvalue* $-\sqrt{7}$*. On the other hand for* $d = 5$ *it is possible to construct* 10 *lines in* R^5 *so that the angle between any distinct pair is* $\arccos\frac{1}{3}$*: to see this consider the Seidel matrix of the Petersen graph. It's eigenvalues are* $(3)^{\{5\}}, (-3)^{\{5\}}$ *hence this matrix can be used to construct* 10 *equiangular lines in* R^5 *with angle* $\arccos\frac{1}{3}$*. Recall that a conference matrix of order* n *is a symmetric* $n\times n$ *matrix* A *with* 0*'s along* *the diagonal,* ± 1 *elsewhere satisfying* $A^2 = (n-1)I$ *([1]). Its eigenvalues are* $\pm\sqrt{n-1}$*. Hence if a conference matrix of order* $2d$ *exists it will yield* $2d$ *equiangular lines in* R^d *with angle* $\alpha(2d, d)$*. A necessary condition for the existence of a conference matrix of order* n *is that* $n = 2 \mod 4$ *and* $n-1$ *is a sum of two squares (J. H. van Lint and J. J. Seidel). There are exactly* 4 *distinct conference matrices of order* 26 *(Weisfeiler [10]), there are at least* 18 *conference matrices of order* 50*. In each one of these cases* $\alpha(2d, d)$ *will be attained by distinct configurations. Conference matrices of orders* 6*,* 14*,* 30*,* 38*,* 42*,* 46 *are known to exist hence in all corresponding dimensions the bound for the largest angle is attained.*

Acknowledgements The author is indebted to Branko Grünbaum and Victor Klee for helpful discussions during preparation of this note.

References

1. V. Belevitch, Conference networks and Hadamard matrices, Ann. Soc. Sci. Bruxelles, Ser. I 82 (1968), pp. 13–32.
2. L. Danzer, Finite point-sets on S^2, Discrete Mathematics, Vol. 60 (1986), pp. 3–66.
3. R. L. Graham, B. L. Rothschild and E. G. Strauss, Are there $n+2$ points in E^n with odd integral distances? Amer. Math. Monthly, 81 (1974), pp. 21–25.
4. B. S. Kashin and S. V. Konyagin, On systems of vectors in a Hilbert space, Proceedings of the Steklov Institute of Mathematics (AMS Translation) 1983, Issue 3, pp. 67–70.
5. P. W. H. Lemmens and J. J. Seidel, Equiangular lines, J. of algebra 24 (1973), pp. 494–512.
6. J. H. van Lint and J. J. Seidel, Equilateral points sets in elliptic geometry, Proc. Kon. Nederl. Akad. Wet., Ser. A, 69 (1966), pp. 335–348.
7. Raphael M. Robinson, Arrangement of 24 points on a sphere, Math. Annalen 144 (1961), pp. 17–48.
8. Raphael M. Robinson, Finite sets of points on a sphere with each nearest to five others, Math. Annalen 1979 (1969), pp. 296–318.
9. M. Rosenfeld, Almost orthogonal lines in E^d, Applied Geometry and Discrete Mathematics, The "Victor Klee Festschrift," DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 4 (1991), pp. 489–492.
10. B. Weisfeiler, On construction and identification of graphs, Lecture Notes 558, Springer-Verlag (1976).

On Mutually Avoiding Sets*

Pavel Valtr

P. Valtr (✉)
Department of Applied Mathematics and Institute for Computer Science (ITI),
Charles University, Malostranské nám. 25, 11800 Praha 1, Czech Republic

Graduiertenkolleg “Algorithmische Diskrete Mathematik”, Fachbereich
Mathematik, Freie Universität Berlin, Takustrasse 9, 14195 Berlin, Germany
e-mail: valtr@kam.mff.cuni.cz

Summary. Two finite sets of points in the plane are called mutually avoiding if any straight line passing through two points of any one of these two sets does not intersect the convex hull of the other set. For any integer n, we construct a set of n points in general position in the plane which contains no pair of mutually avoiding sets of size more than $\mathcal{O}(\sqrt{n})$ each. The given bound is tight up to a constant factor, since Aronov et al. [1] showed a polynomial-time algorithm for finding two mutually avoiding sets of size $\Omega(\sqrt{n})$ each in any set of n points in general position in the plane.

1. Introduction

Let A and B be two disjoint finite sets of points in the plane such that their union contains no three points on a line. We say that A *avoids* B if no straight line determined by a pair of points of A intersects the convex hull of B. A and B are called *mutually avoiding* if A avoids B and B avoids A. In this note we investigate the maximum size of a pair of mutually avoiding sets in a given set of n points in the plane.

Aronov et al. [1] showed that any set of n points in general position in the plane (i.e., no three points lie on a line) contains a pair of mutually avoiding sets, both of size at least $\sqrt{n/12}$. Moreover, they gave an algorithm for finding such a pair of sets in time $\mathcal{O}(n \log n)$. In Sect. 2 we construct, for any integer n, a set of n points in general position in the plane which contains no pair of mutually avoiding sets of size more than $11\sqrt{n}$ each.

Mutually avoiding sets in a d-dimensional space are defined similarly. Any set of n points in general position in $\mathbb{R}^d$ contains a pair of mutually avoiding sets, both of size at least $\Omega(n^{\frac{1}{d^2-d+1}})$ (see [1]). On the other hand, our method described for the planar case in Sect. 2 yields a construction of sets of n points in R^d with no pair of mutually avoiding sets of size more than $\mathcal{O}(n^{1-1/d})$.

Now we recall some definitions from [1]. A set of line segments, each joining a pair of the given points, is called a *crossing family* if any two line segments intersect in the interior. Two line segments are called *parallel* if

*This research was supported by “Deutsche Forschungsgemeinschaft”, grant We 1265/2–1 and by Charles University grant No. 351.

R.L. Graham et al. (eds.), *The Mathematics of Paul Erdős I*,
DOI 10.1007/978-1-4614-7258-2_36,

they are two opposite sides of a convex quadrilateral. In other words, two line segments are parallel if their endpoints form two mutually avoiding sets of size 2. It is an easy observation that any pair of avoiding sets of size s can be rebuilt into s pairwise parallel line segments or into a crossing family of size s. Aronov et al. [1] used this observation and the above result on mutually avoiding sets for finding a crossing family of size $\Omega(\sqrt{n})$ and a set of $\Omega(\sqrt{n})$ pairwise parallel line segments.

The result on pairwise parallel line segments was strengthened and extended to a higher dimension by Pach. Pach [8] showed that any set of n points in general position in $\mathbb{R}^d$ contains at least $\Omega(n^{1/d})$ d-dimensional simplices (i.e., $(d+1)$-point subsets) which are pairwise mutually avoiding.

In Sect. 3 we show a relation between mutually avoiding sets and Erdős' well-known empty-hexagon-problem.

2. Sets with Small Mutually Avoiding Subsets

For a finite set P of points in the plane, let $q(P)$ denote the ratio of the maximum distance of any pair of points of P to the minimum distance of any pair of points of P. For example, if P is a square grid $\sqrt{n} \times \sqrt{n}$ then $q(P) = \sqrt{2}(\sqrt{n}-1)$. In this section we show:

Theorem 1. *Let $c > 0$ be a positive constant. Then any set P of n points in the plane satisfying $q(P) \leq c\sqrt{n}$ contains no pair of mutually avoiding sets of size more than $\lceil 2(\sqrt{17}+1)c\sqrt{n}\rceil$ each.*

One of the basic results about covering says that for any integer $n \geq 2$ there is a set P of n points in the plane with $q(P) < c_0\sqrt{n}$, where $c_0 = \sqrt{2\sqrt{3}/\pi} \approx 1.05$ (see [5]). Such a set P can be found as the triangle grid inside a disk of appropriate size. If we slightly perturb the points of P, we obtain a set in general position still satisfying $q(P) < c_0\sqrt{n}$. According to Theorem 1 this set contains no pair of mutually avoiding sets of size more than $11\sqrt{n}$. (It is obvious for $n \leq 100$. For $n > 100$ we use the estimation $\lceil 2(\sqrt{17}+1)c_0\sqrt{n}\rceil < 11\sqrt{n}$.)

Proof of Theorem 1. Let P be a set of n points in the plane satisfying $q(P) \leq c\sqrt{n}$. Without loss of generality, we may and shall assume that the minimum distance in P is 1. Let A and B be two mutually avoiding subsets of the set P. Define Cartesian coordinates so that for some positive constant $d \in (0, \frac{1}{2}c\sqrt{n}\rangle$ all points of $A \cup B$ lie in the closed strip between the two vertical lines $p : x = -d$ and $q : x = d$, and one of the sets A and B, say A, has a point on the line p and a point on the line q. Moreover, let the topmost point b_0 of the set B lie on the x-axis and let the set A lie "above" the set B (i.e., the set A lies above any straight line connecting two points of B). Since b_0 lies on the x-axis, all points of $A \cup B$ lie between the two horizontal lines $r : y = -c\sqrt{n}$

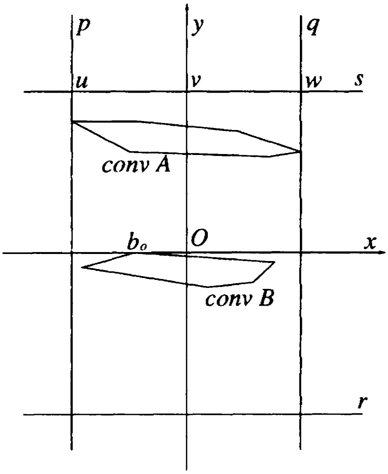


Fig. 1 Auxiliary lines and points

and $s : y = c\sqrt{n}$. Define three points u, v, w as those points in which the line s intersects the line p, the y-axis, and the line q, respectively (see Fig. 1).

For each point $b \in B$, let $f(b)$ be that point in which the line segment b_v intersects the x-axis. Now we show that, for any two points $b, b' \in B$, the distance $|f(b)f(b')|$ between $f(b)$ and $f(b')$ is greater than $\frac{d}{(\sqrt{17}+1)c\sqrt{n}}$.

If the line bb' is horizontal then $|f(b)f(b')| \geq \frac{1}{2}|bb'| \geq \frac{1}{2} > \frac{d}{(\sqrt{17}+1)c\sqrt{n}}$. If the line bb' is not horizontal then it intersects the line s in some point g outside the segment uw (see Fig. 2). Thus $|gv| > d$. Without loss of generality, assume that b is closer to the line s than b'. Let z be that point on $b'v$, for which bz is horizontal.

Now estimate

$$|bz| = \frac{|bb'|}{|gb'|} \cdot |gv| > |bb'| \cdot \frac{|gv|}{|gv| + |vb'|} > 1 \cdot \frac{|gv|}{|gv| + \sqrt{(2c\sqrt{n})^2 + d^2}} >$$

$$\frac{d}{d + \sqrt{(2c\sqrt{n})^2 + d^2}} \geq \frac{d}{d + \frac{1}{2}c\sqrt{n} + \sqrt{(2c\sqrt{n})^2 + (\frac{1}{2}c\sqrt{n})^2}} = \frac{d}{\left(\frac{1}{2} + \sqrt{\frac{17}{4}c\sqrt{n}}\right)}$$

and

$$|f(b)f(b')| \geq \frac{1}{2}|bz| > \frac{d}{(\sqrt{17}+1)c\sqrt{n}}.$$

Since the points $f(b)$, $b \in B$ are placed on a line segment of length $2d$, the size of B is at most $\lceil 2d / \frac{d}{(\sqrt{17}+1)c\sqrt{n}} \rceil = \lceil 2(\sqrt{17}+1)c\sqrt{n} \rceil$. □

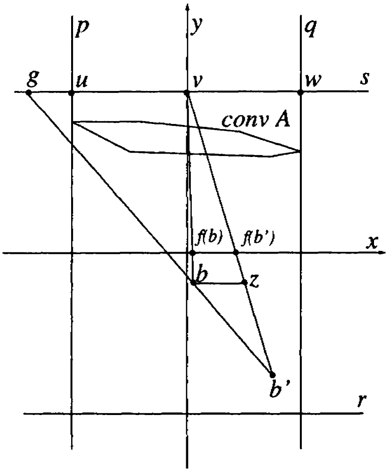


Fig. 2 Auxiliary points and triangles

3. Relation Between Mutually Avoiding Sets and the Empty-Hexagon Problem

Let A be a finite set of points in general position in the plane. A subset S of A of size k is called *convex* if its elements are vertices of a convex k-gon. If S is convex and the interior of the corresponding convex k-gon contains no point of A, then S is called a *k-hole* (or *an empty k-gon*). The classical Erdős-Szekeres Theorem [4] (1935) says that if the size of A is at least $\binom{2k-4}{k-2}+1$ then A contains a convex subset of size k.

Erdős [3] asked whether the following sharpening of the Erdős-Szekeres theorem is true. Is there a least integer $n(k)$ such that any set of $n(k)$ points in general position in the plane contains a k-hole? He pointed out that $n(4)=5$ and Harborth [6] proved $n(5)=10$. However, as Horton [7] shows, $n(k)$ does not exist for $k \geq 7$. The question about the existence of $n(6)$ (the empty-hexagon-problem) is still open. After a definition we formulate a conjecture which, if true, would imply that the number $n(6)$ exists.

Definition 3. *Let A be a finite set of points in general position in the plane. Let $k \geq 2, l \geq 2$. A subset S of A of size $k+l$ is called a (k,l)-set if S is a union of two disjoint sets K and L so that the following three conditions hold:*

(i) $|K|=k, \quad |L|=l,$
(ii) K and L are mutually avoiding,
(iii) the convex hull of S contains no points of $A-S$.

Conjecture 1 (**Bárány, Valtr**). *For any two integers $k \geq 2$ and $l \geq 2$, there is an integer $p(k, l)$ such that any set of at least $p(k, l)$ points in general position in the plane contains a (k, l)-set.*

If Conjecture 1 is true for $k = l = 6$ then the number $n(6)$ exists. It follows from the fact that any $(6, 6)$-set contains a 6-hole (it can be shown that either one of the corresponding sets K and L is a 6-hole or there is a 6-hole containing three points of K and three points of L).

Note that all known constructions of large sets with no 6-hole (see [7, 9]) satisfy Conjecture 1 already for rather small integers $p(k, l)$.

We cannot even prove that the numbers $p(k, 2)$, $k \geq 5$ exist. (Note that $p(2, 2) = 5$ and $p(3, 2) = 7$ are the minimum values of $p(k, 2)$, $k = 2, 3$, for which Conjecture 1 holds.) The existence of numbers $p(k, 2)$, $k \geq 2$ would imply the following conjecture:

Conjecture 2 (**Bárány, Valtr**). *For any integer $k > 0$, there is an integer $R(k)$ such that any set of at least $R(k)$ points in general position in the plane contains $k + 2$ points $x, y, z_1, z_2, \ldots, z_k$ such that the k sets $\{x, y, z_i\}$, $i = 1, \ldots, k$ are 3-holes (i.e., they form empty triangles).*

Bárány [2] proved that Conjecture 2 holds for $k \leq 10$.

References

1. B. Aronov, P. Erdős, W. Goddard, D. J. Kleitman, M. Klugerman, J. Pach, and L. J. Schulman, Crossing Families, *Combinatorica* 14 (1994), 127–134; also Proc. Seventh Annual Sympos. on CompoGeom., ACM Press, New York (1991), 351–356.
2. I. Bárány, personal communication.
3. P. Erdős, On some problems of elementary and combinatorial geometry, Ann. Mat. Pura. Appl. (4) 103 (1975),99–108.
4. P. Erdős and G. Szekeres, A combinatorial problem in geometry, Compositio Math. 2 (1935) ,463–470.
5. L. Fejes Tóth, Regular figures, Pergamon Press, Oxford 1964.
6. H. Harborth, Konvexe Fünfecke in ebenen Punktmengen, Elem. Math. 33 (1978), 116–118.
7. J. D. Horton, Sets with no empty convex 7-gons, Canadian Math. Bull. 26 (1983), 482–484.
8. J. Pach, Notes on Geometric Graph Theory, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 6 (1991), 273–285.
9. P. Valtr, Convex independent sets and 7-holes in restricted planar point sets, Discrete Comput. Geom. 7 (1992), 135–152.