GAFA, Geom. funct. anal.
Special Volume – GAFA2000, 1 – 0
1016-443X/00/010001-0 $ 1.50+0.20/0

GAFA Geometric And Functional Analysis

ROUGH STRUCTURE AND CLASSIFICATION

W.T. Gowers

1 Introduction

When I was first asked to speak at the "Visions in Mathematics" conference, I had what I believe was a typical reaction. I wanted to try to emulate Hilbert a century ago, but since I knew that I could not possibly match his breadth of vision, I was forced to make some sort of compromise. In this paper I shall discuss several open problems, not always in areas I know much about, but they are not intended as a list of the most important questions in mathematics, or even the most important questions in the areas of mathematics that I have worked in. Rather, they are a personal selection of problems that, for one reason or another, have captured my attention over the years.

The title of this paper refers to a general theme that links several unsolved problems in combinatorics. I believe that what makes them difficult is, in many cases, the lack of adequate "rough classification" theorems. I shall explain what I mean by this in §3 and give examples of problems for which classification might very well be useful. In §4 I give a more miscellaneous list of questions, with brief accounts of why I find them interesting. Most, but not all, of the questions on this list are very well known.

As the conference drew nearer, I decided that I could not resist at least some attempt at "the vision thing" [Bu]. I am not in favour of predicting which areas of mathematics are likely to be fashionable in fifteen (let alone fifty) years' time, but in §2 I have made a more general prediction about mathematics as a whole. Briefly, my contention is that during the next century computers will become sufficiently good at proving theorems that the practice of pure mathematical research will be completely revolutionized, quite possibly to the point where we would be reluctant to recognize

The research for this article was partially supported by the Clay Mathematics Institute.

it as mathematical research at all. My reasons for this opinion are outlined in the next section.

I do not claim any originality for my views on this topic. Indeed, by the time of my lecture at the conference, some of the ideas had been touched on by other speakers and during discussion sessions. Moreover, interest in automatic theorem proving has existed for many years, and since the conference I have been referred to similar views expressed by people who work in the area. (One reference is given at the end of the next section.)

2 Will Mathematics Exist in 2099?

For at least a century, mathematicians have thought about the possibility of automating mathematics. Hilbert famously asked in his tenth problem whether there was an algorithm for solving Diophantine equations, and later extended this to the question of whether there was an algorithm which would find a proof of any mathematical statement that had one. In 1936, Turing, equally famously, formalized the notion of algorithm and soon afterwards demonstrated the insolubility of the halting problem, thus showing that no such algorithm existed. While this result, and later demonstrations that several natural and well known problems in mathematics were also impossible to solve systematically, may have initially seemed somewhat negative, they also had a positive side. The idea that all our creativity and insight might be reduced to something mechanical was, after all, not very appealing. Turing's result therefore came as a relief, since it left mathematicians with something to do.

More recently, with the rise of the theory of computational complexity and NP-completeness, it has been widely understood that this relief was somewhat misplaced, because, as Gödel in fact appreciated, an algorithm for finding proofs is not much use if it takes an unfeasibly long time. Far more threatening to mathematicians would be an algorithm which could determine, for any given mathematical statement, whether or not it had a proof of length at most n, and do this in polynomial time (ideally with a polynomial of small degree). The existence of an algorithm of this kind is one of the major open problems of mathematics, since it is equivalent to whether P=NP, but many mathematicians take comfort from the fact that nobody has found a short algorithm for solving any NP-complete problem and most experts believe that no such algorithm exists.

However, this comfort, even if you are convinced that P$\neq$NP, is still misplaced, because searching for shortish proofs of arbitrary mathematical statements is a very bad model of what mathematicians actually do. Indeed, there is absolutely no evidence that where mathematicians score over computers is in their ability to solve NP-complete questions quickly. If somebody were to unearth an *idiot savant* who could determine in his or her head whether collections of Boolean formulae were satisfiable, or even factorize large integers, then one would be forced to ask some profound questions, but that will not happen (a prediction I make with great confidence). Therefore, the only possible explanation for the great success of mathematicians over the centuries is that our search for proofs of theorems is very restricted. Indeed, this is obviously true – we tend to prefer questions that are interesting, comprehensible, and seemingly not completely out of reach, and we like our proofs to provide *explanations* rather than just formal guarantees of truth.

This line of thought leads to the following question: can the notion of a "good proof" be usefully formalized? If so, then the impact on mathematics could well be comparable to the impact of the formalizing of the notions of proof and algorithm. It is perhaps unlikely that a definition will emerge which is as indisputably correct as that of a recursive function or Turing machine, but it might be realistic to try to isolate certain properties of what we commonly regard as good proofs and formalize them. Here is an incomplete list of features often associated with good proofs.

(i) Proofs are usually clearer if they have a hierarchical structure. For example, if the main theorem follows from three lemmas that all have clear statements, then one can isolate parts of the proof, think about them separately and understand the argument without having to hold too much in one's head at once.

(ii) Many good proofs are variants of existing better known arguments. This makes them easier to understand, because all one has to do is concentrate on the parts that are new.

(iii) It is easier to understand a proof that can be reduced to a few basic ideas. (Needless to say, it is not easy to formalize what one means by a "basic idea".)

(iv) Sometimes the form of a proof is dictated to some extent by the existence of certain examples and counterexamples. These clarify the proof by suggesting approaches that might work and demonstrating that other approaches cannot work.

(v) A good definition can greatly improve a proof. In extreme cases, choosing the right definition can reduce a difficult problem to a simple exercise in checking.

(vi) Many of the best proofs, or at least the best-written proofs, come with some indication of how they were discovered. To put that the other way round, proofs that depend on magic tricks, while they may be impressive and elegant, are not completely satisfactory as explanations.

The above features are interrelated, and they all have in common that a proof with one of the features is, other things being equal, easier to understand than one without. I would like to concentrate on (vi), and adopt it as a working definition of "explanation" as opposed to "guarantee of truth". That is, I shall think of an explanation as a proof together with an account of how the proof was discovered. This account should not be something like "I was sitting in the bath and suddenly had the following brilliant idea", but rather a demystification of the proof: that is, a convincing demonstration of how a mere (mathematically trained) human being could search for a proof and find this one, without relying on excessive amounts of luck.

I am aware that this definition (which will be made more precise in a moment) does not capture all our associations with the word "explanation". In particular, it might be easier to show how one came up with a brute force argument than to show how one thought of a neat definition which, after one had seen it, made the result obvious. Nevertheless, if we wish to teach computers to find proofs, it is likely to be a good idea to reflect on how we do so ourselves.

One way of explaining a proof, in the above sense, is to show that it can be generated using what one might think of as standard mathematical reflexes. Some of these might be very general methods of approach, and others might be trying to apply well known techniques from the area of mathematics in question. Here are a few examples of what I mean by general methods of approach. For my lecture I drew flow charts, but my TeX skills are very basic, so I shall represent them here as pseudo-algorithms. A much more detailed and sophisticated discussion of mathematical reflexes can be found in Polya's classic books, such as [P, vols. I and II].

A. Generalization.

1. Is p a special case of q?
   If it seems not to be, then goto 2, else goto 3.
2. Can q be appropriately modified?
   If it can then goto 1, else goto 5.

3. Is q true?
   If it might be, then goto 4, else goto 2.
4. Try to prove q instead of p.
5. Try a different approach.

B. Modification of Known Argument.

1. Find a known statement q that resembles p.
2. Can the proof of q be easily modified to give p?
   If yes then END, else goto 3.
3. Identify a statement r, forming part of the proof of q, such that the analogous statement s, which one would like as part of the proof of p, is not easy to prove.
4. If s is obviously false then goto 6, else goto 5.
5. Try to prove s.
6. Try a different approach.

C. Spotting Patterns in Special Cases.

1. Find a special case r of p.
2. If r is trivial, then goto 3, else goto 4.
3. Can I identify the next simplest special case?
   If so, call it r and goto 2, else goto 8.
4. Can I see how to prove r?
   If so, then goto 5, else goto 7.
5. Does the proof of r generalize to a proof of p?
   If so, then END, else goto 6.
6. Can I find a different, more instructive proof of r?
   If so, then goto 5, else goto 3.
7. Try to prove r instead of p.
8. Try a different approach.

D. Habituation, or Getting Bored.

1. Have I managed to think of any interesting ideas connected with p in the last couple of hours?
   If so, then goto 2, else goto 3.
2. Think about one of the interesting ideas.
3. Make a nice cup of coffee and think about something else for a while.

As for reflexes of a more specific kind, I mean things like trying to apply well known theorems or techniques. For this, one needs to have some ability to classify problems, so that one has an idea of which theorems or techniques may be helpful. Rather than giving several examples of the use of standard methods to solve problems, let me return to the question of automating mathematics and present an imagined dialogue between a mathematician and a computer in two or three decades' time. The idea of the dialogue is that the computer is very helpful to the mathematician, while not doing anything particularly clever. This represents an unthreatening intermediate stage between what we have now, computers that act as slaves doing unbelievably boring calculations for us, and full automation of mathematics. I have written the dialogue in English, but this is supposed to be a translation of a more formal language which has not yet been invented. (I shall discuss this point a little more later.)

Mathematician. Is the following true? Let $\delta > 0$. Then for N sufficiently large, every set $A \subset \{1, 2, \ldots, N\}$ of size at least δN contains a subset of the form $\{a, a+d, a+2d\}$?

Computer. Yes. If A is non-empty, choose $a \in A$ and set $d = 0$.

M. All right all right, but what if d is not allowed to be zero?

C. Have you tried induction on N, with some $\delta = \delta(N)$ tending to zero?

M. That idea is no help at all. Give me some examples please.

C. The obvious greedy algorithm gives the set

$$\{1, 2, 4, 5, 10, 11, 13, 14, 28, 29, 31, 32, 37, 38, 40, 41, \ldots\}.$$

I notice that large parts of the set are translations of other parts. In fact, this set is very like the Cantor set, so this gives a bound of $\delta \geqslant N^{(\log 2/\log 3)-1}$.

As for random methods, let us choose each element of $\{1, 2, \ldots, N\}$ to be in A with probability δ, the choices being independent. The expected number of subsets of A of the form $\{a, a+d, a+2d\}$ is $\delta^3 N^2$, to within an absolute constant. Hence, if $\delta < N^{-2/3}$ it is possible for there to be none. The expected size of A is δN, and standard estimates tell us that A is very likely indeed to be of about this size.

Applying one further standard idea, we note that if the number of sets in A of the form $\{a, a+d, a+2d\}$ is at most half the number of points in A, then we can delete one element from each of them and still be left with a substantial proportion of A. This tells us that we can choose any δ that satisfies $C\delta^3 N^2 \leqslant \delta N$, so in particular we can find some δ of the form $cN^{-1/2}$.

M. Well, random methods often give the best answer for problems like this. Let's try to prove that $cN^{-1/2}$ is best possible.

C. [*Pauses for 0.001 seconds*] Actually it isn't. Behrend found a much better bound in 1946. [*Downloads paper*]
M. Oh dear, I'm out of ideas then. Could you give me a suggestion by any chance?
C. We have a set A. We want to prove that a subset of a certain form exists. The best way of proving existence is often to count.
M. [*Intrigued*] Yes, but what would that mean for a problem like this?
C. Here we wish to count the number of solutions (x, y, z) of the single linear equation $2y = x + z$. A standard tool in such situations is Fourier analysis, exponential sums, the circle method, whatever you like to call it.
M. Where is the group structure?
C. Identify $\{1, 2, \ldots, N\}$ with $\mathbb{Z}/N\mathbb{Z}$. It isn't pretty but it sometimes works. I write $\mathbb{Z}_N$ for $\mathbb{Z}/N\mathbb{Z}$ and $\hat{A}(r)$ for $\sum_{s\in A} \exp(2\pi i rs/N)$. Then the number of triples $(a, a+d, a+2d) \in A^3$ is $N^{-1}\sum_{r\in\mathbb{Z}_N} \hat{A}(r)^2\overline{\hat{A}(-2r)}$.
M. We are trying to show that that expression cannot be zero unless δ is very small.
C. [*Trained to humour mathematicians*] That is indeed almost equivalent to our original problem. Thank you. Notice that $N^{-1}\hat{A}(0)^3 = \delta^3 N^2$, exactly the number we had before when A was chosen randomly. This number is large and positive – a good sign.
M. So we would like the rest of the sum to be small enough not to cancel out the zero term.
C. Yes. I am trying a few obvious ideas such as the Cauchy-Schwarz inequality. Unfortunately, there does not seem to be any reason for the rest of the sum to be small.
M. Please show me your calculations.
C. Here they are. [*Displays them.*] They do enable me, or rather us, to prove a partial result. It states that if $\max_{r\neq 0} |\hat{A}(r)| \leqslant c\delta^2 N$, then the image of A in $\mathbb{Z}_N$ contains many sets of the desired form.

This dialogue could be continued to show what to do when $\hat{A}(r)$ is large for some r. However, the computer has already guided the mathematician to an important insight which is the basis for Roth's proof of his theorem on arithmetic progressions [Ro1]. Finishing the proof is not all that difficult, though perhaps it might be for the mathematician above, who seems a bit too reliant on his computer.

Notice that the computer is at every stage trying standard ideas: induction, a greedy algorithm, random methods, counting, taking the Fourier transform and applying Cauchy-Schwarz to a sum of products. What makes it think of *these* standard ideas, rather than some other completely inappropriate ones? Part of the answer lies in how the problem is initially put to the computer. I would envisage not the formal statement given at the beginning of the dialogue, but something more interactive. For example, this stage of the process could be menu-driven. Initially one would choose Combinatorics, or perhaps Combinatorial Number Theory. This would bring up another menu from which one would choose Extremal Problems. For the next menu one would choose the word Maximize. The next choice would be Over Sets A of Integers. Then one would choose Cardinality of A. Then a menu would come up with a choice of constraints, or perhaps one would have to type them in: $A \subset \{1, 2, \ldots, N\}$ and A contains no arithmetic progression of length three.

At the end of a process like this, the computer would have many ideas about how the problem was conventionally classified. For example, it would be natural for the computer to think of the circle method because many problems in additive number theory would involve concepts such as the number of arithmetic progressions of length three in a set A. The computer could then look for a list of standard tricks, and find the slogan in the above dialogue to do with counting solutions of a linear equation.

Notice that it is important for the computer that it has access to a mathematical database which is much more sophisticated than anything we have at present. One might wonder how the computer could have discovered Behrend's paper so quickly, since understanding ordinary text is notoriously difficult for computers. I was imagining that a great deal of work had gone into classifying papers so that Behrend's paper appeared on the database not as Proc. Nat. Acad. Sci. 23 (1946), 331–332, but rather as something like Maximize – Size of A – Subject To – (1) $A \subset \{1, 2, \ldots, N\}$ & (2) A contains no 3-AP – Lower Bound. As well as the paper, there would be a summary of the main result – Lower Bound At Least $n \exp(-c\sqrt{\log n})$.

I do not see any obstacle to creating a useful database of the above kind, other than the time and effort that would be involved. The structure of the database would be very well suited to the internet, which would also make it easier to distribute the work amongst many people, so perhaps even this obstacle is not all that great. There are of course parts of mathematics that do not lend themselves particularly easily to classification of the above kind, but this is no reason to dismiss the whole idea. Even a database covering only certain areas of mathematics would be useful.

Encoding and classifying general methods, tricks, rules of thumb and so on would obviously be harder. One of the main reasons for this is that they are often not precisely formulated, and dealing with vague statements is a well known hard problem in computer science. However, hard does not mean impossible and there has already been important progress on it. Moreover, at least some of the vague principles of mathematics do not appear to present serious difficulties. Consider for example the following simple but extremely useful consequence of the Cauchy-Schwarz inequality:

$$\left(\sum_{i=1}^{n}|a_i|\right)^2 \leqslant n\sum_{i=1}^{n}|a_i|^2 .$$

This is a useful inequality because the ℓ_2-norm is often easier to handle than the ℓ_1-norm. However, in many contexts the inequality is far too weak, so one tends to use it only when there is some reason to believe that the a_i are of roughly equal size. This sounds like a vague idea, and many mathematicians, when thinking about a problem, will not need to clarify it further, but further clarification is perfectly possible. For example, another very simple but surprisingly useful inequality is

$$\sum_{i=1}^{n}|a_i|^2 \leqslant \max_{1\leqslant i\leqslant n}|a_i|\sum_{i=1}^{n}|a_i| ,$$

which tells us that the first inequality will be sharp to within a constant C if $\max_{1\leqslant i\leqslant n}|a_i| \leqslant Cn^{-1}\sum_{i=1}^{n}|a_i|$. This is not a necessary condition, but it could form one of a number of tests that a computer could apply (or rather, attempt to apply) whenever it used the first inequality, to see whether it was likely to be expensive.

If there were to be a serious attempt to codify the many principles and working methods of mathematicians, then it would be an immensely useful exercise to return to the mathematics we already know and try to rewrite the proofs as explanations (in the restricted sense that I defined earlier). Many textbooks begin with definitions that justify themselves later by leading to interesting results: this would be forbidden in the kind of textbook I have in mind. Instead, every definition should be presented as arising naturally in the search for the solution of some problem. Of course, the problem itself should also be justified; or alternatively one could take as one's starting point that the problem is worth studying, leaving the justification to somebody else. (The question of justifying open problems was discussed by Gromov at the conference. See his contribution to this volume, especially the ends of §1 and §2. For the sake of brevity I have ignored many aspects of mathematical practice, such as making conjectures and proposing definitions. These must also be taken seriously if mathematics is to be automated.)

Even if the automation of mathematics turns out to be much more difficult than I have supposed, this exercise of rewriting could have enormous benefits, because, unsurprisingly, the question of how we might teach mathematics to computers is closely related to the question of how we should teach it to human beings. I have no time for those who believe that talent for mathematical research is a mysterious unteachable quality which either you have or you don't. Of course there is such a thing as natural ability, but most mathematicians develop their ability as they mature. At the moment, they tend to do this with very little external guidance and very little help from the literature. A series of books aimed at making transparent the thought-processes that led to certain results could show to young mathematicians why research is not as impossible as it looks and convey much more efficiently the tricks used by those with more experience.

Two examples of what I mean about the justification of definitions are provided by simplicial homology and the Riemann zeta function. When I was taught simplicial homology, I was presented with a complicated (as it seemed at the time) definition which then, miraculously, led to interesting and far more concrete results such as fixed-point theorems. Similarly, a typical introduction to the zeta function begins with Euler's product formula, which leads rapidly to interesting theorems about the primes, but is not presented as the inevitable result of thinking about the primes. For these two concepts, I can give plausible (but completely ahistorical and by no means unique) accounts of how they might have been invented. I intend to put them on my personal web page (http://www.dpmms.cam.ac.uk/~wtg10/), but do not expect to get round to this before about January 2001. I also hope to give some examples of how apparent bright ideas can be explained as the natural result of familiar methods of searching for proofs.

I am fully aware that what I have described so far is a long way from the full automation of mathematics. In particular, I have ignored some notorious problems in artificial intelligence. I have briefly discussed vagueness, but the example I gave was a simple one, and does not obviously generalize. I have not mentioned the difficulty of pattern-recognition, although recognising similarity is a very important part of mathematical research. What I have tried to show is that these difficulties are not a good reason for doing nothing. Reflecting on one's thought processes does not require any expertise in computing and is potentially very beneficial to others; and a computer with a great deal of information but only very modest intelligence could still be extremely useful.

I have also not discussed the many efforts that have already been made in the general area of automatic theorem proving, an area which (as is probably obvious) I know little about. My impression is that most of the progress so far has been from the bottom up – trying to get computers to discover simple but interesting theorems starting almost from scratch. What I have talked about is more of a top-down project to which any mathematician could contribute. It might be that eventually such a top-down approach could reduce the general problem of automating mathematics to a large number of smaller, more manageable problems, so that perhaps the mathematical community could meet the automatic theorem provers half way. (One way to learn about the state of the art in automatic theorem proving is to visit the website http://www.math.temple.edu/~zeilberg/OPINIONS.html and follow links, especially from the rather extreme opinion 36.)

Finally, then, here is my guess about the impact of computers on mathematics over the next century. Although some of the difficulties involved in automatic theorem proving are formidable, a century is a very long time in computing, so I expect computers to be better than humans at proving theorems in 2099. I do not feel particularly happy about this, but I expect the impact on my own mathematical life to be entirely positive, because a semi-intelligent database of the kind I have described would be a wonderful resource and would take a great deal of the drudgery out of research. There might even be a golden age when computers were good at exercises but not yet good at having deep insights. Then, instead of wasting a week not noticing that a hoped-for lemma had a simple counterexample, one could get the computer to check it. In other words, computers would still do the boring bits for us, but these would not be quite as boring as they are now. However, such a golden age, if it occurs, is unlikely to last for long. The next stage might be one where only a very few outstanding mathematicians could discover proofs that were inaccessible to computers. Perhaps others would concentrate on teaching, but computers would probably become better than us at this as well, at least if they themselves had been taught in anything like the way I have outlined. In the end, the work of the mathematician would be simply to learn how to use theorem-proving machines effectively and to find interesting applications for them. This would be a valuable skill, but it would hardly be pure mathematics as we know it today.

3 Rough Structure and Classification, and Related Problems

In his talk, Gromov drew attention to the fact that not many ways were known of defining manifolds. (See the subsection "Randomization" in §5 of his contribution to this volume.) For a different reason, something similar is true for graphs. This might at first seem a ridiculous assertion – all you have to do to define a graph is take the set of all pairs of elements of the set $\{1,2,\dots,n\}$ and pass to an arbitrary subset. However, while it is true that a mathematical object does not have to satisfy delicate constraints in order to be a graph, one still faces the purely logical problem that the number of comprehensible descriptions is limited. Moreover, while graph theorists often talk as though they are considering individual graphs, most problems in graph theory are, strictly speaking, about *families* $(G_n)_{n=1}^\infty$ of graphs, where the number of vertices of G_n is usually n, or perhaps some simple function of n that tends to infinity.

Here is a list of a few common ways of defining families of graphs.

(i) *Repeating patterns.* Some examples of this are

  (a) K_n, the complete graph of order n;
  (b) P_n, the path with n vertices;
  (c) C_n, the cycle with n vertices;
  (d) $K_{1,n}$, the "star" with one vertex joined to n others, and no further edges;
  (e) $\{0,1\}^n$, the discrete n-dimensional cube;
  (f) a dyadic tree with a root and n further layers.

(ii) *Simple constructions from known graphs.* These include products, quotients, disjoint unions, removal of a few edges and so on.

(iii) *Use of algebra and number theory.* Two notable examples are the Paley graph and the Ramanujan graphs of Lubotzky, Phillips and Sarnak [LPS]. In general, Cayley graphs often have interesting properties.

(iv) *Random graphs.* It is not quite clear in what sense this phrase defines a family of graphs: see the discussion below.

This list could undoubtedly be continued, but I would be surprised if somebody could give me a further ten genuinely distinct and interesting classes of definitions.

Let us consider for a moment what it means to say that a graph is "random". We would like to make statements such as "A random graph on

n vertices with edge-probability $\frac{1}{3}$ has approximately $\frac{1}{3}\binom{n}{2}$ edges." Sometimes, one can make sense of such a statement by inserting the words "with high probability" in the appropriate place. However, there are contexts in which this is not adequate. For example, there is a class of graphs, known as expanders, which have direct practical applications in the design of certain algorithms. It is known that graphs chosen randomly are, with high probability, good expanders. On the other hand, faced with a particular randomly chosen graph, how can one be sure that it is not one of the exceptions? This is perhaps not a genuine practical problem, since a graph that works with probability $1-10^{-8}$ is just as useful as one that works with complete certainty (though I have heard that people in industry who buy computer systems are very suspicious of random elements in the design). Even so, a very important development in graph theory was the realization that several properties shared by almost all graphs are equivalent, and serve as a useful definition of "randomness".

These properties apply to a graph G with n vertices such that almost every vertex has degree approximately $n/2$. They show that G resembles a graph where each edge is chosen independently with probability $1/2$. (Similar results hold for other probabilities as well.) Here are three of them.

(i) For any two large sets A and B of vertices, the number $e_G(A,B)$ of pairs $(x,y)\in A\times B$ such that xy is an edge of G is $|A||B|/2+o(n^2)$ (and thus roughly the number one would expect if G was a random graph).

(ii) For any fixed (small) graph H, the number of (induced) subgraphs of G isomorphic to H is roughly what the expected number would have been if G had been chosen randomly.

(iii) Define the adjacency matrix A of G by setting $A(x,y)$ to be 1 if xy is an edge and 0 otherwise. Then the second largest eigenvalue (in modulus) of A is $o(n)$.

In connection with property (i) above, define the *density* $\delta_G(A,B)$ to be $e_G(A,B)/|A||B|$. Another way of stating the property is to say that $\delta_G(A,B)$ is approximately $1/2$ for all large sets A and B. Interestingly, one can replace (ii) by the weak-seeming assumption that G contains roughly the expected number of cycles of length four.

One of the first papers in which it was shown that the smallness of the second largest eigenvalue implied randomness properties was by Alon and Milman [AM]. The first paper dealing explicitly with the idea of pseudo-random graphs was by Thomason [T1]. The paper of Chung, Graham and Wilson [CGW] made the whole area much more systematic, giving a long list of equivalent randomness properties. They called a graph satisfying any one (and hence all) of the properties *quasirandom*, perhaps to make clear the distinction between this idea and the rather different concept of pseudorandomness in computer science.

There are two morals I would like to draw from the above results. The first is that in many contexts they allow one to give a clear meaning to the concept of a random graph, even when the graph has been explicitly chosen. (For example, the Paley graph is known to be quasirandom in this sense.) The second is that they support our intuition that two random graphs are basically the same. Of course, they will not be isomorphic, but they share all the properties that interest us, so the difference between them is not important.

I can now explain what I mean by "rough classification". It is clear that there are too many graphs for it to be sensible to try to classify them all up to isomorphism. However, as just commented, we often like to regard non-isomorphic graphs as basically the same. Implicit in that discussion was a metric which can be defined as follows. Given two graphs G and H, both with n vertices, and a bijection ϕ between $V(G)$ and $V(H)$, define

$$d_\phi(G,H) = n^{-2} \max_{A,B\subset V(G)} \left|e_G(A,B) - e_H(\phi A, \phi B)\right|$$

and let $d(G,H)$ be the minimum of all the $d_\phi(G,H)$. If G and H are close in this metric, it means that there is a one-to-one correspondence between their vertices such that the densities $\delta_G(A,B)$ in G are approximately the same as the corresponding densities in H, at least when A and B are large. A graph is quasirandom if and only if it is close to a typical random graph in this metric.

By a rough classification of graphs, I mean a choice of a small number of representative graphs, such that every graph is close in the above metric to one of the representatives. Such a classification will be useful if the representative graphs can be easily described and analysed. It is not at all obvious that a useful classification should exist, but an important result, Szemerédi's regularity lemma [S1], shows that it does.

To state the result, let us introduce a further definition. If G is a graph and A and B are sets of vertices in G, let us say that the pair (A,B) is *ϵ-regular* if, whenever A' and B' are subsets of A and B with $|A'| \geqslant \epsilon|A|$ and $|B'| \geqslant \epsilon|B|$, we have $|d_G(A',B') - d_G(A,B)| \leqslant \epsilon$. If A and B are disjoint, this condition, which is very similar to the first property of quasirandom graphs listed above, states that the edges between A and B behave very like those of a random bipartite graph on A and B with edge probability $d_G(A,B)$. Szemerédi's regularity lemma is as follows.

LEMMA. *Let $\epsilon > 0$. Then there exists a positive integer $k = k(\epsilon)$ such that, given any finite graph G, the vertices of G can be partitioned into k classes $V_1, \dots, V_k$, all of size $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$, such that all but at most $\epsilon\binom{k}{2}$ of the pairs (V_i, V_j) (with $i < j$) are ϵ-regular.*

A common way of defining graphs is to choose a few sets $V_1, \dots, V_k$ and a symmetric $[0,1]$-valued matrix $(p_{ij})_{i,j=1}^k$, and to join $x \in V_i$ to $y \in V_j$ with probability p_{ij}, all choices being independent. Szemerédi's regularity lemma states that *every* graph is like that, at least up to closeness in the metric defined earlier. Since random graphs are quite well understood, this gives us a useful classification.

One consequence of this result is that if G is a graph, and H is an induced subgraph of G formed by choosing half the vertices of G at random, then H looks very similar to G. The reason is that, with high probability, the vertex set of H will be made up of roughly half of each of the sets V_i in the partition given by Szemerédi's lemma, and the regularity of a pair (V_i, V_j) is inherited by large subsets (V_i', V_j'), at the cost of a slight worsening of the parameter ϵ. This general principle about random subgraphs is hard to demonstrate without the use of some result along the lines of the regularity lemma (though there are weaker versions that suffice).

A major drawback with the regularity lemma is that the dependence of k on ϵ is very bad – a tower of twos of height proportional to ϵ^{-5}. Moreover, it has been shown that this unfortunate situation is necessary. In [G1] a graph is constructed which gives a lower bound for k of a tower of twos of height proportional to $\epsilon^{-1/16}$. This limits the use of the lemma to situations where the dependence on ϵ is not important.

Before I move on to talk about open problems, let me briefly discuss the other half of my title: rough structure. Mathematicians vary considerably in the amount of structure they like, and this allows them to be roughly classified in a useful way. At one end of the spectrum are those who start with so many assumptions that even finding a new object that satisfies the assumptions is a great achievement. Those who manage to do so are often rewarded with the discovery of beautiful and detailed patterns. At the other end are mathematicians who like more elbow room. They operate under fewer constraints, so the objects they study exist in great profusion and the problems that arise are of a very different character.

As my discussion of classification already indicates, this distinction, though it undoubtedly exists, is not quite as great as it might seem. I would like to blur it a little further by showing that patterns can emerge unexpectedly even in very combinatorial problems.

There is a theorem of Freiman [Fr1,2] that illustrates this extremely well. Let A be a set of integers, and write $A+A$ for the set $\{x+y : x, y \in A\}$. It is an easy exercise to show that if $|A| = n$ then $2n-1 \leqslant |A+A| \leqslant \frac{1}{2}n(n+1)$ and that both bounds are tight. Freiman's theorem gives a complete description of A when $A+A$ is small, or to be precise, when $|A+A| \leqslant Cn$ for some fixed constant C. An obvious way to construct such a set is to let P be an arithmetic progression of size at most $Cn/2$ and to let A be an arbitrary subset of P of size n. However, this is not the only source of examples: one can take a "two-dimensional" arithmetic progression, which is a set of the form

$$\{a + r_1 d_1 + r_2 d_2 : 0 \leqslant r_1 < m_1, 0 \leqslant r_2 < m_2\}\,.$$

It can be checked that if A is a two-dimensional progression, then $|A+A| \leqslant 4|A|$. More generally, $A+A$ will be small if A is a large subset of a d-dimensional arithmetic progression for some small d. (The definition should be obvious.)

A further easy exercise is to show that if $|A+A| = 2n-1$, then A must be an arithmetic progression of length n. Harder than this, but still not too hard, is to show that if $|A+A|$ is just a little larger than $2n$, at most $2.1n$ say, then there is an arithmetic progression of length slightly greater than n that contains A. The strongest result of this kind is due to Freiman.

Theorem. *Let A be a set of integers of size n. If $0 \leqslant b \leqslant n-3$ and $|A+A| \leqslant 2n-1+b$, then A is contained in an arithmetic progression of length $n+b$.*

This result shows that A is forced to be one-dimensional if $|A+A| \leqslant 3n-4$. This bound is tight, because if m is any integer larger than $2n-3$ and A is the set $\{1, 2, \dots, n-1, m\}$, then $|A+A| = 3n-3$.

Freiman's main theorem is much deeper, and tells us what happens when we gradually relax the condition on the size of the sumset. It states that the *only* sets with small sumsets are large subsets of low-dimensional arithmetic progressions.

Theorem. *For every C there exist constants $K = K(C)$ and $d = d(C)$ such that if A is any set of n integers and $|A+A| \leqslant Cn$, then A is a* *subset of a generalized arithmetic progression of dimension at most d and cardinality at most Kn.*

Notice that the assumption made in Freiman's theorem is combinatorial, but the conclusion is very definitely structural, in an appropriate rough sense. This is not particularly surprising when C is close to the extremal value of 2, when A is, just as one might expect, a small perturbation of some easily described extremal set, but for larger values of C the result is much less to be expected, and is not at all easy to prove. A beautiful and completely different proof of Freiman's theorem was obtained by Ruzsa [Ru], and a considerably cleaned up version of Freiman's original argument has recently been given by Bilu [Bi].

I have presented Szemerédi's regularity lemma and Freiman's theorem in a very positive way, and they both have important applications. However, they do not solve everything. For the rest of this section I shall discuss problems which appear to be hard to solve without some sort of rough classification theorem. In each case, the classification theorem would have to be better than anything we have at present. I begin with three problems in elementary Ramsey theory.

1. Bounds for diagonal Ramsey numbers. Ramsey's theorem asserts that for every positive integer k there exists a positive integer N such that if the edges of the complete graph on N vertices are coloured red and blue, then there must be k of the vertices with all edges joining them of the same colour. (The graph spanned by these vertices is called *monochromatic*.) The smallest N with this property is denoted by $R(k,k)$. It is known that $R(3,3)=6$ and $R(4,4)=18$, but $R(k,k)$ is not known for any $k \geqslant 5$. It is unlikely that a reasonable formula for $R(k,k)$ exists, so a much more interesting question is how it behaves asymptotically. Simple arguments show that $2^{k/2} \leqslant R(k,k) \leqslant 2^{2k}$, but, despite strenuous efforts on the part of many mathematicians, no significant improvement on either bound has been obtained. (By significant I mean a lower bound of $2^{\alpha k}$ with $\alpha > 1/2$ or an upper bound of $2^{\beta k}$ with $\beta < 2$.)

There is an equivalent definition of $R(k,k)$, which I shall sometimes use without further comment. It is the smallest N such that any graph with N vertices contains a clique or independent set with k vertices (that is, k vertices either completely joined or not joined at all).

2. Covering by triangle-free graphs. The Ramsey number $R(3,3,\dots,3)$ (where 3 occurs r times) is the smallest integer N such that if the complete graph on N vertices is coloured with r colours, then there must be a triangle with all its edges of the same colour. Relatively straightforward arguments show that this number lies between e^r and r^r. It is an open question whether there is an upper bound of the form C^r for some absolute constant C.

An equivalent formulation of this problem is as follows: if the complete graph on n vertices is to be written as a union of k triangle-free graphs, how large must k be? One can do it if $k = \lceil \log_2 n \rceil$. Is it necessary to use at least $c \log n$ graphs?

There is a further equivalent formulation of the problem in terms of Shannon capacity – see the contribution of Alon in this volume for a discussion of this concept.

3. Ramsey numbers under additional constraints. The following is a beautiful problem of Erdős and Hajnal. Let H be a fixed finite graph. If G is assumed not to contain any induced copy of H, then by how much does the bound for $R(k,k)$ improve? In particular, does there exist a constant C depending on H only such that every graph with at least k^C vertices containing no induced H has a complete or empty subgraph with k vertices?

What is it about these three problems that makes them hard? The most important reason seems to be that it is not obvious what the extremal graphs should be. For the first problem, the lower bound $R(k,k) \geqslant 2^{k/2}$ was proved by Erdős [E] in 1947. His proof can be summarized as follows: colour each edge red or blue randomly and independently with probability one half, and check that the expected number of monochromatic complete subgraphs with k vertices is less than one. To put that even more loosely: a random colouring works.

Once one has seen this proof, it comes to seem so natural that it is tempting to conjecture that the best possible bound, or something close to it, is attained by a random graph, and indeed many who have thought about the problem believe this to be so. However, an unsettling result of Thomason [T2] suggests that the proof of such a fact will not be easy. A natural approach to estimating $R(k,k)$ is to consider the following more general problem. Let k and n be positive integers, and suppose that the edges of a complete graph on n vertices are coloured red and blue. What is the smallest possible number of monochromatic complete subgraphs of order k? Erdős conjectured that a random colouring should be best possible, which would suggest a bound of approximately $2.2^{-\binom{k}{2}}\binom{n}{k}$. This result is true when $k = 3$, but Thomason showed that it is false when $k = 4$.

It is important to emphasize the sense in which random graphs are not best. Some graphs, such as the remarkable Ramanujan graphs, improve on random constructions by being "more random than random". For example, Ramanujan graphs are d-regular and have a second largest eigenvalue of $2\sqrt{d-1}$, which is asymptotically as small as possible. Although random graphs also tend to have a second largest eigenvalue of size $C\sqrt{d-1}$, the constant C is larger than 2. This is not the nature of the non-randomness of Thomason's example, and indeed cannot be, because any graph which is *better* than random will be quasirandom in the sense defined earlier, and therefore it and its complement will contain about the same number of copies of K_4 as a random graph.

Thomason's example is constructed as follows. He first chooses a fixed graph H with k vertices. (One example that works is a certain product of seven triangles.) He then partitions the vertices of a complete graph into equal-sized sets $V_1, \dots, V_k$ and colours an edge between $x \in V_i$ and $y \in V_j$ blue if ij is an edge of H and red otherwise. Details can be found in [T2,3]. Notice that Szemerédi's regularity lemma and the facts about quasirandom graphs imply that any example is *necessarily* similar to Thomason's in that it will involve partitioning the vertices into a few sets and changing the edge probabilities between the sets from $1/2$ to something else, although it is quite surprising that Thomason manages with probabilities of zero and one.

As for the second problem, one obvious way to cover n edges with $\lceil \log_2 n \rceil$ triangle-free graphs is to use bipartite ones. One can let the vertices of G be all 01-sequences $(\epsilon_1, \dots, \epsilon_k)$ of length k, and for each i define a bipartite graph H_i by joining $(\epsilon_1, \dots, \epsilon_k)$ to $(\eta_1, \dots, \eta_k)$ if and only if $\epsilon_i \neq \eta_i$. Once again, we have a very natural example which looks as though it might be best possible. However, it is not.

In order to demonstrate this, we make a simple observation. Suppose that G is a complete graph on m vertices which can be covered by r triangle-free graphs $H_1, \dots, H_r$. Now let $k > 1$ and consider the complete graph with vertex set G^k. For $1 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant k$ define a graph H_{ij} by joining $(x_1, \dots, x_k)$ to $(y_1, \dots, y_k)$ if and only if $x_j y_j$ is an edge of H_i. It is easy to check that the graphs H_{ij} are triangle-free and cover all edges of G^k. The number of graphs H_{ij} is rk and G^k has m^k vertices. If $n = m^k$ then $rk = \log n/((\log m)/r)$, so to beat the first bound, all we need to do is find an example where $(\log m)/r$ is greater than $\log 2$. One such example exists when $m = 5$. The complete graph on five vertices can be covered with two five-cycles, and $(\log 5)/2 > \log 2$. A better example still is obtained as follows. Let G be the complete graph K_{41}, and identify the vertices of G with the integers modulo 41. The quartic residues form a subgroup Q of the multiplicative group of $\mathbb{F}_{41}$. This subgroup has index four, and it can be checked that no two of its elements differ by one. We now colour the edges xy of G with four colours according to the coset of $x-y$. (Since -1 is a quartic residue, the coset of $x-y$ equals the coset of $y-x$.) It is easy to see that a monochromatic triangle would imply the existence of four non-zero elements a,u,v,w of $\mathbb{F}_{41}$ such that $au^4+av^4=aw^4$. This in turn would imply the existence of non-zero p and q such that $1+p^4=q^4$, which were checked not to exist. Therefore, each colour class represents a triangle free graph, and in this case $(\log m)/r$ is $(\log 41)/4$, which is larger than $(\log 5)/2$.

Even this is not the best known example, which (as I learned from Noga Alon at the conference) is related to a certain Cayley graph with over three hundred vertices and was discovered by a computer.

For the third problem, if H is a moderately complicated graph with seven vertices, say, then there is no hope of guessing what the extremal graph G will be. The result is plausible, because the condition that G contains no induced copy of H rules out almost any use of randomness, which seems to be very important if a graph is not to contain large cliques or independent sets.

It is interesting to think about why for some extremal problems it is easy to identify the best possible examples, while for others it is much harder. The three problems above are all hard because there are two completely different techniques for producing examples, and one can produce a whole spectrum of further examples by mixing them. For example, the following graph G shows that $R(k,k)\geqslant k^2$. Let the vertices of G be $V_1\cup\cdots\cup V_k$, where the V_i are disjoint sets of size k. Join $x\in V_i$ to $y\in V_j$ if and only if $i\neq j$. At one time Turán believed that the bound of k^2 was sharp, and although we now know that he was completely wrong, Thomason's example shows that it is difficult to dismiss examples that mix this sort of explicit approach with a probabilistic one. This also accounts for the difficulty with the third problem. A similar phenomenon occurs with the second, in that there is a spectrum of methods of constructing triangle-free graphs. At one end is a complete bipartite graph and at the other is a random graph with appropriate edge-probabilities (and a few deletions). In between, one can construct a triangle-free graph H on r vertices using random methods, and then blow it up to a graph G on n vertices by partitioning the vertices of

G into sets $V_1, \ldots, V_r$ and joining $x \in V_i$ to $y \in V_j$ if and only if ij is an edge of H.

It is because the extremal graphs for these problems are not obvious that there seems to be a need for some sort of classification. Returning to the problem of estimating $R(k,k)$, although the random graph may well give the right answer, it appears to be necessary to take account of graphs such as the one constructed by Thomason, even if only to rule them out. One might imagine that Szemerédi's regularity lemma was ideally suited to the purpose, but the poor bounds in the lemma make it far too crude. In fact, even the definition of quasirandom graphs is too crude, because it is perfectly possible for a graph G on n vertices to be $n/2$-regular and have a second largest eigenvalue proportional to $\sqrt{n}$, but to contain a complete subgraph with $\sqrt{n}$ vertices. In other words, the distinction between two quasirandom graphs as previously defined can be very important in the context of Ramsey's theorem. This suggests the need for a stronger definition: as far as I know, nothing suitable has been proposed.

I have concentrated so far on graph theory, but similar difficulties arise in many other contexts. Here are two problems in combinatorial number theory.

4. Roth's theorem on arithmetic progressions. In 1974, Szemerédi [S2] proved the following theorem, which had been conjectured by Erdős and Turán in 1936 [ET].

Theorem. *Let $k \geqslant 3$ be an integer and let $\delta > 0$. Then if N is sufficiently large, every subset A of $\{1, 2, \ldots, N\}$ of cardinality at least δN contains an arithmetic progression of length k.*

The correct asymptotic dependence of N on k and δ is still a major open problem, but the problem I wish to concentrate on here is more specific. It is to determine the dependence of N on δ when $k = 3$.

The first progress towards the Erdős-Turán conjecture was due to Roth [Ro1], who showed that the conjecture was true when $k = 3$, and that N could be taken to be $\geqslant \exp\exp(C/\delta)$. This was later improved to $\exp(1/\delta^C)$ by Szemerédi [S3] and Heath-Brown [He]. Very recently, it has been shown by Bourgain [Bo3] that C can be $2 + \epsilon$. More precisely, if $\delta \geqslant c \log\log N (\log N)^{-1/2}$, then A must contain an arithmetic progression of length three.

The best known bound in the other direction is given by an ingenious example discovered by Behrend in 1946 [B]. It shows that the inequality $N \geqslant \exp(C \log(\delta^{-1})^2)$, or equivalently $\delta \geqslant \exp(-c\sqrt{\log N})$, is not enough to guarantee an arithmetic progression of length three.

This situation is interesting for a number of reasons. First, the bounds leave open the question of whether a density $\delta = (\log N)^{-1}$ is enough. This is an important question because $(\log N)^{-1}$ is the density of the primes in the interval $\{1, 2, \ldots, N\}$. A result of van der Corput, based on techniques of Vinogradov, shows that the primes contain infinitely many arithmetic progressions of length three. It is still possible, however, that there is a proof of this fact which depends only on the density of the primes and not on more detailed facts about their distribution.

A second reason is that although Behrend's example gives a very weak bound compared with what is known in the other direction, it does at least show that pure random methods do not give an extremal set. (This point was discussed by the mathematician and the computer in the dialogue in the previous section.)

5. Bounds for Freiman's theorem. In the statement of Freiman's theorem given earlier, no mention was made of the dependence of K and d on C. This is still far from being completely resolved. See the article of Bilu [Bi] for details about what is known. Improving the known bounds would have important applications. To give some idea of what is *not* known, here are two specific questions. If A is a set of n integers and $|A + A| \leqslant C|A|$, can anything be said about the structure of A if $C = n^\alpha$ for some $\alpha > 0$, or even if $C = \log n$? The existing proofs give no information at all in this situation. For the second question, let A satisfy the same conditions. Does A have a subset B of size cn such that B is contained in a generalized arithmetic progression P of cardinality K and dimension d, where c and K have some reasonable dependence on C and $d = A \log C$ for some absolute constant A? For applications, it would still be very interesting if d could be bounded above by $(\log C)^A$.

These two problems raise difficulties that are similar in spirit to the difficulties in the graph-theoretic problems. Behrend's example shows that there is not an obvious extremal set for Problem 4, and once again there is a conflict between random and explicit techniques which partially accounts for this situation. Similarly, the sets that arise in any consideration of Freiman's theorem are very varied. So for both problems a classification would almost certainly be useful.

More specifically, for both problems one finds oneself wanting to understand subsets $A \subset \mathbb{Z}_N$ of size αN, where α is a constant, or perhaps a number that tends to zero very slowly as N tends to infinity. (Here $\mathbb{Z}_N$

is the group of integers modulo N.) A good way to do this is to look at the Fourier transform of the characteristic function of A, defined by $\hat{A}(r) = \sum_{s\in A} e(rs/N)$, where $e(x)$ is notation for $\exp(2\pi ix)$. The coefficient $\hat{A}(0)$ is just the cardinality of A. If $\hat{A}(r)$ is significantly smaller than $\alpha^2 N$ for every non-zero r, then A behaves in many ways like a random subset of $\mathbb{Z}_N$, where every element is chosen with probability α. We therefore have a good definition of quasirandomness for subsets of $\mathbb{Z}_N$. See [CG] for more details about this.

Notice that the information conveyed by Fourier coefficients about subsets of $\mathbb{Z}_N$ is very similar to the information conveyed by eigenvalues about graphs. This is not a coincidence: if A is a subset of $\mathbb{Z}_N$, one can define a directed graph G_A with vertex set $\mathbb{Z}_N$ by having an edge from x to y if and only if $y - x \in A$, and it is easy to check that the eigenvalues of G_A are the Fourier coefficients of A. The analogies go further. Just as one can try to classify graphs, one can also try to classify subsets of $\mathbb{Z}_N$. For this one needs a notion of "not interestingly different": it turns out that the difference between two sets A and B is for many purposes unimportant if $\hat{A}(r) - \hat{B}(r)$ is small for every r.

Fourier analysis has been an essential part of the best arguments known for analysing subsets A of $\mathbb{Z}_N$ in connection with Problems 4 and 5. The reason is that, because perturbing the Fourier coefficients of a set A does not make an important difference, we only have to worry about the large coefficients. There are very few of these, so we have the following efficient way of encoding the essential information about A: choose a constant β, let $K = \{r : |\hat{A}(r)| \geqslant \beta N\}$ and for $r \in K$ let $f_A(r) = \hat{A}(r)$. It might seem as though mapping sets A to their functions f_A is a ready-made classification scheme, but I would regard "semi-classification" as a better description, because it is not easy to use all the information contained in f_A. The known arguments tend to throw away all information about the set K except for its cardinality, but the structure of K is very important as well. In particular, the behaviour of A depends very much on whether there are simple arithmetical relationships between the elements of K, such as two of them adding to a third. Unfortunately, it is not easy to see how to incorporate this information into arguments, and this seems to be a barrier to further progress on the two problems. Once again, what appears to be missing is a better classification scheme. (Similar remarks apply to eigenvalues of graphs. Useful though they are, for many problems it is important to know how the corresponding eigenvectors fit together, and this information is hard to use.)

The next two problems are designed to show that difficulties similar to the ones discussed so far can arise in many contexts.

6. A problem about subsets of the n-sphere. Denote by S_{n-1} the Euclidean unit sphere of $\mathbb{R}_n$. Let $m < n$, let $\epsilon > 0$ and let A be a subset of S_{n-1} that intersects every m-dimensional subspace of $\mathbb{R}_n$. Write A_ϵ for the set $\{x : d(x, A) \leqslant \epsilon\}$. There are interesting questions one can ask about the set A_ϵ. For example, how small can its measure be? For given values of m and ϵ, what is the largest r such that A_ϵ is forced to contain the unit sphere of an r-dimensional subspace?

These questions are hard, because there are two natural ways of constructing a set A with the given property. The first is to take the unit sphere of an $(n - m)$-dimensional subspace, and the second is to take a union of randomly placed spherical caps of some given radius. Again we are in a situation where many other constructions can be devised that mix these two ideas (for example, instead of caps, which are expansions of zero-dimensional sets, one could take expansions of k-dimensional sets for some small k). It seems likely that if A is required to be centrally symmetric, then the measure of A_ϵ is minimized when A is the sphere of an $(n - m)$-dimensional subspace. On the other hand, it does not seem necessary for A_ϵ to contain the sphere of such a large subspace. Perhaps one can do better by allowing A_ϵ to be larger, but less regular.

For some values of m and ϵ, a rough classification of subsets of S_{n-1} might help to determine a good bound for r. One would like to exploit the fact that sets A for which A_ϵ is small have some structure, and deduce from this that r is larger than is implied by measure-theoretic considerations alone.

One possible method of providing *upper* bounds for r is to let B be the unit sphere of an $(n - m)$-dimensional subspace, let ϕ be a continuous antipodal map from S_{n-1} to itself and let $A = \phi(B)$. This makes A a sort of topological sphere of dimension $n - m$. Unfortunately, it is hard to think of a choice for ϕ which will provide a good bound and be easy to analyse.

7. A problem about layers of the discrete cube. Let $r < s < t < n$ be positive integers, write $[n]^{(s)}$ for the set of all subsets of $[n] = \{1, 2, \ldots, n\}$ of size s and let $\mathcal{A}$ be a subset of $[n]^{(s)}$. That is, $\mathcal{A}$ is a collection of sets of size s. Write $\partial_r\mathcal{A}$ for the collection of all sets $B \in [n]^{(r)}$ such that $B \subset A$ for some $A \in \mathcal{A}$, and $\partial^t\mathcal{A}$ for the collection of all sets $C \in [n]^{(t)}$ such that $A \subset C$ for some $A \in \mathcal{A}$. Suppose that every set in

$[n]^{(t)}$ contains some set in $\mathcal{A}$, so $\partial^t \mathcal{A} = [n]^{(t)}$. How small can $\partial_r \mathcal{A}$ be?

The nature of this problem depends a great deal on the values of r, s, t and n. Notice that, once again, there are two completely different ways of constructing a set $\mathcal{A}$ that satisfies the conditions of the problem. If one wishes to minimize the cardinality of $\mathcal{A}$, then it is best to choose it randomly. However, amongst all sets $\mathcal{A}$ of a given cardinality, random ones are those for which $\partial_r \mathcal{A}$ is largest. An alternative construction exploits the pigeonhole principle in a simple way. For convenience let us suppose that s divides t and that $k = t/s$ divides n. Partition the set $[n]$ into k equal-sized subsets $V_1 \cup \cdots \cup V_k$ and let $\mathcal{A}$ be the collection of all sets A of size r such that $A \subset V_i$ for some i. Every set of size t intersects at least one of the V_i in at least $n/k = s$ points, so $\mathcal{A}$ satisfies the required condition.

Let us suppose for a moment that s is much smaller than t. If r is only very slightly smaller than s, then it looks as though a random choice of $\mathcal{A}$ is best, since each set of $\mathcal{A}$ will contain only a few sets in $[n]^{(r)}$. If on the other hand r is much smaller than s, then the explicit construction turns out to be better. Between these two extremes, there is a range of values of r where it is not obvious which construction is better, or whether something "in between" is more suitable. In fact, I do not even know how to prove what appears to be the case in the extreme situations – that the examples I have sketched are best possible.

This problem is perhaps slightly artificial, but it is a simple example of a type of difficulty I have met several times when thinking about other questions, and I doubt whether I am alone in this. Whatever one thinks about the interest or otherwise of the problem, it is hard to imagine a solution being dull. Once again, it looks as though a rough classification would be very useful, this time of subsets of $[n]^{(s)}$. (There do exist versions of Szemerédi's regularity lemma when s is fixed, but they would not help if one took $s = \sqrt{n}$, for example.)

8. Finding rough classifications. Arising from the previous problems is the general problem of finding new and useful rough classifications. Amongst the objects that one might like to classify are the following.

(i) Graphs. The problem here is to find a useful replacement of Szemerédi's regularity lemma. Because the bad bound in the lemma is (in a very weak sense!) tight, replacing the lemma would mean replacing the description it gives of a general graph by something more complicated, in return for much better bounds. Another interesting project is to find alternative proofs for results that use the lemma (just as it is interesting to replace a proof that relies on the classification of finite simple groups by one that does not). See [KoS], [GrRR] and [G2, Prop. 12] for some examples of how it can be avoided. For more general information on the regularity lemma and its variants, see [KoS].

(ii) Triangle-free graphs. Two ways of constructing these are to use random methods, or to take a given triangle-free graph H with vertices $x_1, \dots, x_k$ and to define a graph G with vertex set partitioned into sets $V_1, \dots, V_k$ by joining $x \in V_i$ to $y \in V_j$ if and only if $x_i x_j$ is an edge of H. Combining these methods, and other simple operations such as subgraphs and disjoint unions, yields a large class of triangle-free graphs. Do techniques such as these get anywhere near to generating *all* triangle-free graphs?

(iii) Sparse graphs. It has long been recognised that a major drawback with Szemerédi's regularity lemma is that the only graphs for which it gives any information are those with a number of edges proportional to $\binom{n}{2}$, while many problems of great interest concern graphs with, say, n^α edges for some $\alpha \in (1,2)$. Some variants of the regularity lemma do exist for such sparse graphs, but they fall well short of anything that one might wish to call a classification.

(iv) Subsets of Abelian groups, especially $\mathbb{Z}_N$ and $\mathbb{Z}$, with two sets regarded as similar if their Fourier transforms are close.

(v) Subsets of $\mathbb{Z}_N$, with two sets regarded as similar if their characteristic functions are close in the following norm:

$$\|f\| = N^{-4} \sum_{x,a,b,c \in \mathbb{Z}_N} \prod_{\epsilon \in \{0,1\}^3} f_\epsilon(x - \epsilon_1 a - \epsilon_2 b - \epsilon_3 c)\,.$$

Here f_ϵ denotes f if $\epsilon_1 + \epsilon_2 + \epsilon_3$ is even and the complex conjugate of f otherwise. It turns out that, while the large Fourier coefficients of a set A contain all the information about arithmetic progressions of length three, they do not do so for longer progressions. If A and B are close in the above norm, then they will contain approximately the same number of arithmetic progressions of length four – hence its importance. For more details see [G2]. A classification with respect to this norm would necessarily be much more sophisticated than a classification with respect to closeness of Fourier coefficients.

(vi) Subsets of metric spaces such as $[0,1]^2$ and S_{n-1} with two sets regarded as similar if their metric properties are roughly the same. There are many ways of making this question precise, some of which have been studied extensively. One possibility is to say that subsets A and B of a metric space X are similar if there is a small $\epsilon > 0$ and there are 1-Lipschitz maps $\phi : A \to X$ and $\psi : B \to X$ such that every point of B is within ϵ of some point of ϕA, every point of A is within ϵ of some point of ψB, and $d(x, \psi\phi x)$ and $d(y, \phi\psi y)$ are at most ϵ for every $x \in A$ and $y \in B$.

(vii) Subsets of combinatorial structures such as the discrete cube $\{0,1\}^n$ and $[n]^{(r)}$.

4 A Further Selection of Problems

Every problem in this section had to satisfy two criteria. First, I myself should have spent at least an hour trying to solve it, and second, I should have at least some intention of spending more time on it in the future. These were enough to rule out most of the interesting problems in mathematics, leaving me with a manageable list.

9. The Riemann hypothesis. This is so well known that I shall not even bother to state it here. I include it because I do not wish to make some kind of statement by not doing so. However, I have nothing much to say about it. Instead, I refer the reader to the vast literature on the subject, including parts of the articles of Connes, Iwaniec and Sarnak in this volume.

10. The Kakeya problem. This is closely related to the Riemann hypothesis. It has many formulations, of which here are three.

(i) Let A be a subset of $\mathbb{R}^d$ such that for every line L in $\mathbb{R}^d$, A contains a line L' parallel to L. Such a set is often called a Kakeya set, or a Besicovitch set. Besicovitch showed that the measure of A can be zero, but all known examples have Hausdorff dimension d. Bourgain [Bo4] has shown that the Hausdorff dimension of A must be at least $\frac{13d}{25} + \frac{12}{25}$, which is a significant improvement on a relatively easy bound of $\frac{d}{2}+1$. (Very recently this bound was improved further by Katz and Tao to $\frac{6d}{11} + \frac{5}{11}$ [KaT].) It is known that the dimension must be d when $d = 2$, but for $d > 2$ the problem is open. The first non-trivial lower bound when $d = 3$ was discovered by Bourgain [Bo2]. His argument showed that the dimension of a three-dimensional Kakeya set was at least $7/3$. This was improved by Wolff to $5/2$ [W]. This bound was recently improved very slightly by Katz, Laba and Tao [KaLT], an interesting result as $5/2$ was a natural limit in Wolff's proof and surpassing it was very difficult.

(ii) Let S be the square $\{(x,y,z) \in \mathbb{R}^3 : z = 0, 0 \leqslant x,y \leqslant 1\}$ and let P be the point $(1/2,1/2,1)$. Now divide S into n^2 subsquares S_{ij} of sidelength n^{-1} in the obvious way. For each of these subsquares, form a cone C_{ij} by taking the convex hull of S_{ij} and P. Translate each cone C_{ij} in an xy-direction (or in other words, slide its base to somewhere else in the xy-plane) and call the translated cone C'_{ij}. Then how small can the volume of $\bigcup_{i,j=1}^n C'_{ij}$ be? If the smallest possible volume behaves like $n^{-\alpha}$ for some $\alpha > 0$, then there is a Kakeya set in $\mathbb{R}^3$ with Hausdorff dimension less than three. The converse is also true. The obvious d-dimensional analogue of this equivalence holds as well.

(iii) Let m and r be integers, and for every d between 1 and r, let $P_d \subset \mathbb{Z}$ be an arithmetic progression of length m and common difference d. How small can the union $\bigcup_{d=1}^r P_d$ be? In particular, if $r = m^C$ for some fixed integer $C \geqslant 2$, then can one improve on the trivial bound of mr by a factor of m^α for some $\alpha > 0$? Once again, a construction that achieved this could be translated into an example of a Kakeya set with less than full dimension.

The third version of the problem gives some clue that it has a number-theoretical content. A close connection with a conjecture of Montgomery (see [Mo, Chapter 7]), and hence with the zeros of the zeta function, was uncovered by Bourgain in his modestly titled paper [Bo1]. Ever since a famous paper of Fefferman [F], it has been realized that Kakeya sets are also of great importance in harmonic analysis. A solution to the Kakeya problem would therefore be a major advance for several different reasons.

11. Does P equal NP? I have already mentioned this problem in §2. Since not all mathematicians feel comfortable stating it, let me give what I find the cleanest definition of the class of NP problems. First, it is convenient to think not about problems but about Boolean functions, that is, functions $\phi : \{0,1\}^n \to \{0,1\}$. A problem such as the travelling salesman problem can easily be regarded as such a function: encode your graph G with n vertices as a 01-sequence η_G of length $\binom{n}{2}$ in an obvious way and define $\phi(\eta_G)$ to be 1 if G contains a Hamilton cycle and 0 otherwise. A Boolean function ϕ belongs to P if there is a polynomial-time algorithm for computing ϕ. (Strictly speaking, one should talk about a sequence ϕ_n of functions, where ϕ_n has $f(n)$ variables, and the time taken to compute ϕ_n is bounded above by a polynomial function of $f(n)$.) It belongs to NP if it is a *projection* of a function in P, in the following sense: there is an integer m bounded above by a polynomial in n and a function $\psi : \{0,1\}^m \times \{0,1\}^n$ which belongs to P such that $\phi(\eta) = 1$ if and only if there exists $\xi \in \{0,1\}^m$

such that $\psi(\xi, \eta) = 1$.

For the travelling salesman problem, ξ will represent an ordering of the vertices, η a graph, and $\psi(\xi, \eta) = 1$ if and only if the ordering of the vertices given by ξ defines a Hamilton cycle in the graph given by η. This is very easy to check, so ψ belongs to P, and the function ϕ defined earlier, which is 1 if and only if $\psi(\xi, \eta) = 1$ for some ξ, is in NP.

One further definition, which I shall not give precisely, is that of an NP-complete function. An important discovery of Cook [Co] and Karp [K], which explains the interest in the P=NP problem, is that many NP problems (or corresponding Boolean functions) are *universal*, in the sense that a polynomial-time algorithm for solving one of them can be converted into a polynomial-time algorithm for solving an arbitrary problem in NP. Such a problem is called NP-complete. In fact, hundreds of such problems are now known, and many of them are of direct practical importance.

The P=NP problem may look off-putting because the set of all functions Turing-computable in polynomial time is somehow not very mathematical. I have sometimes seen it written that the problem is difficult because it is hard to rule out *all* possible polynomial-time algorithms. However, there is a very clean and completely mathematical reformulation of the problem in terms of so-called circuit complexity which shows that this is not really the difficulty at all. I shall give the definition in an equivalent form.

Let $\phi : \{0,1\}^n \to \{0,1\}$ be a Boolean function and let $A = \{\eta : \phi(\eta){=}1\}$. For each $1 \leqslant i \leqslant n$, let $E_i = \{\eta : \eta_i = 1\}$, and call the sets E_i and their complements *elementary* sets. The *circuit complexity* of ϕ, or A, is the length N of the shortest sequence $A_1, A_2, \dots, A_N$ such that $A_N = A$ and each A_i is either an elementary set, the complement of some earlier A_j or the union or intersection of two earlier A_js.

A problem which is almost, but not quite, equivalent to whether P=NP is the following. Let ϕ be a Boolean function in NP. Must the circuit complexity of ϕ be bounded above by a polynomial in the number of variables? A polynomial-time algorithm for computing ϕ can be converted into a sequence $A_1, \dots, A_N$ of polynomial length, so a superpolynomial bound for the circuit complexity would imply that P and NP are not equal. It is theoretically possible that the converse to this is false, but it is hard to envisage a construction of a sequence $A_1, \dots, A_N$ with so little regularity that it, or some adaptation of it, could not be efficiently generated algorithmically. I am therefore convinced that circuit complexity is the right approach to the problem. (This is the standard view of theoretical computer scientists, but not of all mathematicians.)

Unfortunately, it does not make the problem easy, and the best known lower bound for the circuit complexity of an NP-function is still only linear. A beautiful and important recent paper of Razborov and Rudich [RR] goes a long way towards explaining why the problem is difficult, and is required reading for anybody who would like to prove that P$\neq$NP. A natural approach to proving lower bounds for circuit complexity is to define some measure of simplicity, prove that sets A with small circuit complexity must be simple, and prove that the set A corresponding to some NP-function is not simple. Razborov and Rudich define a notion of "natural proof", which basically means a proof resulting from a "natural" measure of simplicity. They show that all the known bounds in circuit complexity have been achieved using natural proofs in their sense, and they then show that any natural proof that P$\neq$NP could be converted into a polynomial-time algorithm for factorizing large integers. This is convincing evidence that a proof that P$\neq$NP would have to be very strange indeed. For more details, see the contributions of Razborov and Wigderson in this volume, or the paper of Razborov and Rudich itself, which is well written and not hard to read.

The following problem (mentioned by Wigderson at the conference) has a similar flavour to the P=NP problem, but has nothing to do with computers and appears at first glance to be much simpler. Let M be an $n\times n$ invertible matrix over $\mathbb{F}_2$. Gaussian elimination tells us that M can be converted into the identity matrix by a sequence of at most n^2 elementary row operations. Moreover, a simple counting argument (there are fewer than $2n^2$ elementary row operations and almost 2^{n^2} non-singular matrices) shows that almost all matrices need at least $cn^2/\log n$ operations. The problem is to give an explicit example of a matrix that needs a superlinear number. ("Explicit" does not mean "unambiguously defined", or one could just enumerate all the matrices and take the first one that worked. One possible interpretation is that it should be possible to generate an explicit matrix, or rather sequence of matrices, with a polynomial-time algorithm.) This problem remains stubbornly open, and a solution might well shed light on the P=NP problem itself.

12. An effective version of Roth's theorem. By Roth's theorem I now mean not his theorem on arithmetic progressions but his much more famous theorem on rational approximation of algebraic numbers [Ro2]. This states that for every algebraic number α and every $\epsilon > 0$ there exists

$c > 0$ such that for any pair p, q of positive integers we have the inequality $|\alpha - p/q| \geqslant c/q^{2+\epsilon}$. This is a very considerable strengthening of Liouville's theorem. The problem is that the proof gives no information at all about the dependence of c on α and ϵ. The same is true of earlier, weaker results of Thue and Siegel.

The known results in this direction are very modest. For example, Baker [B] has proved an effective version of Roth's theorem in the case $\alpha = 2^{1/3}$, obtaining the inequality $|2^{1/3} - p/q| \geqslant 10^{-6}/q^{2.995}$, which should be compared with the lower bound from Liouville's theorem of c/q^3, where c can be taken to be $1/16$.

A closely related problem, mentioned by Furstenberg in his talk at this conference, concerns the continued-fraction expansion of algebraic numbers. If α is a quadratic irrational, then everything is known, and the partial quotients are eventually periodic. Roth's theorem implies that the partial quotients of an algebraic number cannot grow too quickly, but, except in the quadratic case, there is no example for which it is known whether they are bounded.

13. Littlewood's conjecture. This problem was discussed by Margulis at the conference. Littlewood conjectured that if α and β are any two real numbers and if $\|x\|$ denotes the distance from x to the nearest integer, then $\liminf_{n\in\mathbb{N}} n\,\|n\alpha\|\,\|n\beta\| = 0$.

To understand the problem, consider a quadratic irrational α. Liouville's theorem is equivalent to the statement that $\|n\alpha\|$ is always bounded below by cn^{-1}. On the other hand, usually $\|n\alpha\|$ is much bigger than this. One can therefore ask whether it is possible to find two numbers α and β, neither of which can be well approximated by rationals, with the further property that $\|n\beta\|$ is large whenever $\|n\alpha\|$ is small and vice-versa. Littlewood's conjecture is a precise formulation of this question.

The existence of such a pair of numbers α, β would have the more combinatorial consequence that for every n one could find a set X of n points in $[0,1]^3$ and an absolute constant $c > 0$ such that the inequality $\prod_{i=1}^3 |x_i - y_i| \geqslant cn^{-1}$ held for any two distinct points $x, y \in X$. Indeed, writing $\langle t\rangle$ for the fractional part of t, a suitable set would be $\{(mn^{-1}, \langle\alpha m\rangle, \langle\beta m\rangle) : 1 \leqslant m \leqslant n\}$. As far as I know, the existence of such a set X is an open question, so it is possible that Littlewood's conjecture, if it is true, has a combinatorial proof. A simple argument shows that one can find $n/\log n$ points with the given property. Any improvement would be interesting.

If Littlewood's conjecture is false, then a good candidate for a counterexample would be the pair $(\sqrt{2}, \sqrt{3})$. It seems to be very hard to analyse concrete examples such as this, and perhaps this is the main difficulty with the problem.

14. Self-avoiding random walks. In common with many mathematicians, I am fascinated by the long list of results in statistical physics that have been obtained by physicists using non-rigorous arguments. One of the most appealing source of problems is the area of self-avoiding random walks. A self-avoiding walk of length n is a path $x_0, x_1, \dots, x_n$ of vertices in the lattice $\mathbb{Z}^d$, where $x_0 = 0$, x_{i-1} and x_i are adjacent for $1 \leqslant i \leqslant n$ and there are no repeats. If one places the uniform distribution on the set of all self-avoiding walks of length n in $\mathbb{Z}^2$ or $\mathbb{Z}^3$, then one obtains an important probabilistic model about which a great deal is known to physicists and almost nothing to mathematicians. I will mention just a few problems.

Nearly all the most interesting questions are special cases of the following vague one: what does a typical self-avoiding walk of length n look like? Perhaps the most obvious question is the expected distance from the origin to x_n. In two dimensions, the arguments of physicists show that the answer ought to be $n^{3/4}$, but no upper bound of the form $n^{1-\epsilon}$ has been proved rigorously. The situation for lower bounds is even worse. One might expect a trivial argument to show a lower bound of $cn^{1/2}$, since a self-avoiding walk surely travels further than a pure random walk. However, it is not known how to prove rigorously that a typical self-avoiding walk does not tend to fold back on itself. Of course, most points of a self-avoiding walk must have distance $cn^{1/2}$ from the origin by a trivial counting argument, so this last question is specifically to do with the end point.

Another question concerns the number $f(n)$ of self-avoiding walks of length n. It is easy to see that f is submultiplicative, so a standard argument shows that $f(n) = C^n g(n)$ for some constant C, known as the connective constant, and some function g that grows subexponentially. The exact value of C is not known, but this is not a particularly interesting problem for two reasons. First, C can in principle be determined to any accuracy by an exact counting of all walks of some suitable length n_0 (although this rapidly becomes impossible in practice). More importantly, C does not have a property known as *universality*.

One can change the model of self-avoiding walks in many ways, such as changing the lattice in which it lives. It is known by physicists that certain parameters are unaffected by such changes. For example, the expected end-to-end distance is always proportional to $n^{3/4}$. These unchanging parameters are called universal, and are regarded as having physical significance. The connective constant is not universal in this sense, as it certainly does depend on the underlying lattice.

On the other hand, the function g does seem to be universal. Once again, much more is known to physicists than to mathematicians: they tell us that $g(n)$ grows like $n^{11/32}$. One might at first wonder how one could hope to study the function g when the connective constant C is not known. However, it is a simple exercise to show that g has the following interpretation: $g(n)$ grows like n^{α} if and only if the probability that two self-avoiding random walks of length n do not intersect (except at the origin) behaves like $n^{-\alpha}$. This shows that g does have some sort of independent significance. From a pure mathematical point of view, it is a very interesting open problem to prove that g is bounded above by a polynomial. The best known bound in two dimensions is due to Hammersley and Welsh [HaW], who showed by an ingenious elementary argument that $g(n) \leqslant \exp(cn^{1/2})$.

The reason the hypothesis of universality is plausible is that one would expect the macroscopic behaviour of a very long self-avoiding walk not to depend too critically on the precise behaviour at very small magnitude scales. Connected with this belief is the view that there should be some sort of appropriately scaled limit of self-avoiding random walks, playing the role that Brownian motion and Wiener measure play for ordinary random walks. This is another important open problem.

These problems are all at their most interesting in two and three dimensions. Since the Hausdorff dimension of a Brownian path is two, one does not expect the self-avoidance constraint to be all that significant in four dimensions or more. This intuition has in fact been justified rigorously when the dimension is at least five, by Hara and Slade [HarS1,2]. Even though their results are not very surprising, they are a major achievement and the proofs are not at all easy. For much more detail on all these problems and results, see the excellent book of Madras and Slade [MS].

15. Untying the unknot. There has been a great deal of excitement about knot theory in the last twenty years, including the discovery of important new knot invariants. However, the following basic problem is open: does there exist a polynomial-time algorithm for recognising a diagram of the unknot, and for producing a sequence of Reidemeister moves that unties it? (The time is a function of the number of crossings in the diagram.) If such an algorithm could be found, it would open up a completely new approach to recognising and classifying knots. I know at least one expert in knot theory who strongly believes in the existence of such an algorithm. It is by no means trivial to find *any* algorithm for recognising the unknot, but this has been done by Haken [H]. I myself have performed the following scientific experiment. I take a long piece of string, tie the two ends together with a small knot, and then entangle the resulting loop, trying to make it hard to disentangle (but without pulling anything tight). I have never managed to cause myself the slightest difficulty.

16. The Knaster hypothesis. This is the following question. Let $f : S_n \to \mathbb{R}$ be continuous and let $x_1, \dots, x_{n+1}$ be any $n+1$ points in S_n. Must there exist an orthogonal map α such that the values $f(\alpha x_i)$ are all equal? When $n = 1$ a positive answer follows trivially from the intermediate value theorem. I think the conjecture is known to be true when $n = 2$ but open for $n \geqslant 3$. It is also known for certain very special sets of points $x_1, \dots, x_{n+1}$.

One motivation for the Knaster hypothesis, aside from the intrinsic appeal of the question, is that, as observed by Milman [Mi], a positive answer would provide a completely new proof of one of the central results of Banach space theory – Dvoretzky's theorem [D]. This states that for any k and $\epsilon > 0$ there exists n such that every n-dimensional normed space has a k-dimensional subspace whose Banach-Mazur distance from a Hilbert space is at most $1+\epsilon$. (Geometrically, it says that an n-dimensional centrally symmetric convex body has an almost ellipsoidal k-dimensional cross section.) Dvoretzky's theorem follows from the Knaster hypothesis as follows. Consider S_{n-1} as the unit sphere of ℓ_2^n, let $x_1, \dots, x_n$ be a δ-net of the unit sphere of a k-dimensional subspace V, and let f be the restriction of an arbitrary norm $\|.\|$ on $\mathbb{R}^n$ to S_{n-1}. By the Knaster hypothesis, there is a rotation α such that all the values $\|\alpha x_i\|$ are equal. Standard arguments then show that all points x in the subspace αV have roughly the same ratio $\|x\| / \|x\|_2$. This argument, when made precise, gives the best possible estimate for the asymptotic dependence of n on ϵ for fixed k, which is still an open problem.

Certain natural approaches to the Knaster hypothesis turn out to fail for good reasons, and many people have the feeling that it may well be false. However, even a considerable weakening, such as proving it for $\sqrt{n}$ points in S_n, would be useful for Dvoretzky's theorem.

17. The density Hales-Jewett theorem. The Hales-Jewett theorem concerns colourings of the set $\{1, 2, \dots, k\}^N$. If x belongs to this set, A is a subset of $\{1, 2, \ldots, N\}$ and $1 \leqslant j \leqslant k$, then let us write $x \oplus jA$ for the point y obtained from x by changing x_i to j if $i \in A$ and leaving it unaltered otherwise. A *combinatorial line* is a set of the form $\{x \oplus jA : 1 \leqslant j \leqslant k\}$. For example, when $k = 3$ and $N = 5$, then the point $x = (3, 3, 1, 1, 2)$ and the set $\{1, 3, 5\}$ give rise to the combinatorial line

$$\{(1,3,1,1,1), (2,3,2,1,2), (3,3,3,1,3)\} .$$

Notice that if the set $\{1, 2, \ldots, k\}^N$ is imagined as an N-dimensional grid, then combinatorial lines are special cases of geometrical lines, that is, sets of k collinear points.

The Hales-Jewett theorem states that for any k and r there exists N such that if the set $\{1, 2, \ldots, k\}^N$ is coloured with r colours, then there must be a combinatorial line consisting of points of the same colour. This theorem is easily seen to be a strengthening of van der Waerden's theorem, and like van der Waerden's theorem it has a density version: for every k and every $\delta > 0$ there exists N such that every subset $A \subset \{1, 2, \ldots, k\}^N$ with at least δk^N elements contains a combinatorial line.

This density theorem is due to Furstenberg and Katznelson [FuK] and makes heavy use of ergodic theory. Moreover, no proof is known that does not use these techniques. It would be very interesting to find a non-ergodic proof even in the case $k = 3$. From the theorem in this case it is a simple exercise to deduce the following result. For every $\delta > 0$ there exists N such that every subset of the two-dimensional grid $\{1, 2, \ldots, N\}^2$ of size at least δN^2 contains a triple of the form $\{(x, y), (x, y + d), (x + d, y)\}$, that is, an aligned isosceles rightangled triangle. Even this statement lacks any proof that gives a reasonable upper bound for N in terms of δ.

18. Danzer's problem. Is it possible to find a set X of points in $\mathbb{R}^2$ such that every convex body of area 1 contains at least one point of X and not more than C points, where C is an absolute constant? If K is a convex body of area 1, then there are homothetic rectangles R_1 and R_2 such that $R_1 \subset K \subset R_2$ and such that the area of R_2 is at most four times that of R_1. It follows that it is enough to consider rectangles in the statement of the problem.

There is a small connection between this problem and some of the questions in the previous section, because the obvious random and explicit methods for constructing such a set X both fail, but for very different reasons.

Danzer's actual question is weaker than the above. He asks for a set X which intersects every convex body of area 1 and also satisfies the inequality

$|X \cap RD| \leqslant CR^2$ for every R, where D is the unit disc about the origin. A beautiful paper of Bambah and Woods [BW] shows that X cannot be a finite union of translates of lattices, thus ruling out what at first appears to be a promising source of potential examples.

19. Billiards in polygons. There is a famous example, known as Kafka's study, which shows that one can design a connected (but non-convex) room such that, even if the walls are entirely lined with mirrors, no single light source can illuminate the whole of the room. The example exploits the fact that a beam of light which passes between the two foci of an ellipse will return between the two foci after it has been reflected off the boundary of the ellipse. It is not known whether there is a polygonal example.

A closely related question is the following. Suppose an infinitesimally small billiard ball is set moving on a frictionless (and pocketless) polygonal table. Will it return arbitrarily close to its starting point and direction? The answer is known to be positive if the angles between all the sides of the polygon are rational multiples of π, as then there are only finitely many directions to consider, and the problem can easily be reduced to a known and easy result about so-called interval exchange maps.

Let $[0,1)$ be partitioned into m half-open intervals in two different ways, $I_1 \cup \dots \cup I_m$ and $J_1 \cup \dots \cup J_m$ in such a way that I_r and J_r have the same length for every $r \leqslant m$. Let ϕ_r be the translation of I_r to J_r. Let ϕ be the map which takes $x \in I_r$ to $\phi_r(x)$. An *interval exchange map* from $[0,1)$ to $[0,1)$ is any map of this kind. It is quite straightforward to prove that such maps are recurrent, in the sense that for every x and every $\epsilon > 0$ there exists $N > 0$ such that the distance from $\phi^N x$ to x is at most ϵ.

I have asked around and found nobody who can tell me whether the obvious two-dimensional generalization of the above result is true or false. That is, suppose one chops a square into rectangles and rearranges them (by translation) so that they form the original square again. Is the resulting rectangle exchange map recurrent? I believe that it is, but some experts in dynamical systems to whom I have asked the problem have the opposite instinct.

References

[AM] N. Alon, V.D. Milman, Eigenvalues, expanders and superconcentrators, Proc. 25th Annual FOCS, Singer Island, FL, IEEE, New York (1984), 320–322.

[BW] R.P. BAMBAH, A.C. WOODS, On a problem of Danzer, Pacific J. Math. 37 (1971), 295–301.

[B] F.A. BEHREND, On sets of integers which contain no three in arithmetic progression, Proc. Nat. Acad. Sci. 23 (1946), 331–332.

[Bi] Y. BILU, Structure of sets with small sumset, Astérisque 258 (1999), 77–108.

[Bo1] J. BOURGAIN, Remarks on Montgomery's conjectures on Dirichlet series, Geometric Aspects of Functional Analysis (1989-1990), Springer Lecture Notes in Mathematics 1469 (1991), 153–165.

[Bo2] J. BOURGAIN, Besicovitch type maximal operators and applications to Fourier analysis, GAFA 1 (1991), 147–187.

[Bo3] J. BOURGAIN, On triples in arithmetic progression, GAFA 9:5 (1999), 968–984.

[Bo4] J. BOURGAIN, On the dimension of Kakeya sets and related maximal inequalities, GAFA 9:2 (1999), 256–282.

[Bu] George Bush, Interview, Time, 26th Jan. 1987.

[CG] F.R.K. CHUNG, R.L. GRAHAM, Quasi-random subsets of $\mathbb{Z}_n$, J. Comb. Th. A 61 (1992), 64–86.

[CGW] F.R.K. CHUNG, R.L. GRAHAM, R.M. WILSON, Quasi-random graphs, Combinatorica 9 (1989), 345–362.

[Co] S.A. COOK, The complexity of theorem proving procedures, Proc. 3rd Annual ACM Symposium on the Theory of Computing (1971), 151–158.

[D] A. DVORETZKY, Some results on convex bodies and Banach spaces, Proc. Symp. on Linear Spaces, Jerusalem (1961), 123–160.

[E] P. ERDŐS, Some remarks on the theory of graphs, Bull. Amer. Math. Soc. 53 (1947), 292–294.

[ET] P. ERDŐS, P. TURÁN, On some sequences of integers, J. London Math. Soc. 11 (1936), 261–264.

[F] C. FEFFERMAN, The multiplier problem for the ball, Annals of Math. 94 (1971), 330–336.

[FuK] H. FURSTENBERG, Y. KATZNELSON, A density version of the Hales-Jewett theorem, J. D'Analyse Math. 57 (1991), 64–119.

[Fr1] G.A. FREIMAN, Foundations of a Structural Theory of Set Addition (in Russian), Kazan Gos. Ped. Inst., Kazan 1966.

[Fr2] G.A. FREIMAN, Foundations of a Structural Theory of Set Addition, Translations of Mathematical Monographs 37, Amer. Math. Soc., Providence, R.I., USA, 1973.

[G1] W.T. GOWERS, Lower bounds of tower type for Szemerédi's uniformity lemma, GAFA 7 (1997), 322–337.

[G2] W.T. GOWERS, A new proof of Szemerédi's theorem for arithmetic progressions of length four, GAFA 8 (1998), 529–551.

[GrRR] R.L. GRAHAM, V. RÖDL, A. RUCIŃSKI, On graphs with linear Ramsey numbers, preprint.

[H] W. HAKEN, Theorie der Normalflächen, ein Isotopiekriterium für den Kreisnoten, Acta Math. 105, 245–375.

[HaW] J.M. HAMMERSLEY, D.J.A. WELSH, Further results on the rate of convergence to the connective constant of the hypercubical lattice, Quart. J. Math. Oxford Ser. (2) 13 (1962), 108–110.

[HarS1] T. HARA, G. SLADE, Self-avoiding walk in five or more dimensions, Comm. Math. Phys. 147 (1992), 101–136.

[HarS2] T. HARA, G. SLADE, The lace expansion for self-avoiding walk in five or more dimensions, Rev. Math. Phys. 4 (1992), 235–327.

[He] D.R. HEATH-BROWN, Integer sets containing no arithmetic progressions, J. London Math. Soc. (2) 35 (1987), 385–394.

[K] R.M. KARP, Reducibility among combinatorial problems, Complexity of Computer Computations, Proc. Sympos., IBM Thomas J. Watson Res. Centr, Yorktown Heights, N.Y., 1972, (R.E. Miller, J.W. Thatcher, eds.) Plenum Press, New York 1972, 85–103.

[KaLT] N.H. KATZ, I. LABA, T. TAO, An improved bound on the Minkowski dimension of Besicovitch sets in $\mathbb{R}^3$, Annals of Math., to appear.

[KaT] N.H. KATZ, T. TAO, A new bound on partial sum-sets and difference-sets, and applications to the Kakeya conjecture, submitted.

[KoS] J. KOMLÓS, M. SIMONOVITS, Szemerédi's Regularity Lemma and its applications in Graph Theory, in "Combinatorics, Paul Erdős is 80 (Vol 2)", Bolyai Society Math. Studies 2, 295–352, Kesthely (Hungary) 1993, Budapest 1996.

[LPS] A. LUBOTZKY, R. PHILLIPS, P. SARNAK, Explicit expanders and the Ramanujan conjectures, Proceedings of the 18th ACM Symposium on the Theory of Computing 1986, 240–246; also Combinatorica 8 (1988), 261–277.

[MS] N. MADRAS, G. SLADE, The Self-Avoiding Random Walk, Birkhäuser, Boston, 1992.

[Mi] V.D. MILMAN, A few observations on the connections between local theory and some other fields, in Geometric Aspects of Functional Analysis, Israel seminar (GAFA) 1986-1987 (J. Lindenstrauss, V.D. Milman, eds.), Springer LNM 1317, (1988), 283–289.

[Mo] H.L. MONTGOMERY, Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis, CBMS Regional Conference Series in Math. 84, AMS 1994.

[P] G. POLYA, Mathematics and Plausible Reasoning, Vols. I and II, Princeton University Press, 1954.

[RR] A.A. RAZBOROV, S. RUDICH, Natural proofs, in 26th Annual ACM Symposium on the Theory of Computing (STOC '94, Montreal, PQ, 1994); also J. Comput. System Sci. 55 (1997), 24–35.

[Ro1] K. ROTH, On certain sets of integers, J. London Math. Soc. 28 (1953), 245–252.

[Ro2] K. ROTH, Rational approximations to algebraic numbers, Mathematika 2 (1955), 1–20 (with corrigendum p. 168).

[Ru] I.Z. RUZSA, Generalized arithmetic progressions and sumsets, Acta Math. Hungar. 65 (1995), 379–388.

[S1] E. SZEMERÉDI, Regular partitions of graphs, Colloques Internationaux C.N.R.S. 260 – Problèmes Combinatoires et Théorie des Graphes, Orsay 1976, 399–401.

[S2] E. SZEMERÉDI, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 299–345.

[S3] E. SZEMERÉDI, Integer sets containing no arithmetic progressions, Acta Math. Hungar. 56 (1990), 155–158.

[T1] A.G. THOMASON, Pseudo-random graphs, Proceedings of Random Graphs, Poznán 1985 (M. Karonski, ed.), Annals of Discrete Mathematics 33, 307–331.

[T2] A.G. THOMASON, A disproof of a conjecture of Erdős in Ramsey theory, J. London Math. Soc. 39 (1989), 246–255.

[T3] A.G. THOMASON, Graph products and monochromatic multiplicities, Combinatorica 17 (1997), 125–134.

[W] T.H. WOLFF, An improved bound for Kakeya type maximal functions, Revista Mat. Iberoamericana 11 (1995), 651–674.

W.T. GOWERS, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK