Additive Combinatorics and Theoretical Computer Science*

Luca Trevisan†

May 18, 2009

Abstract

Additive combinatorics is the branch of combinatorics where the objects of study are subsets of the integers or of other abelian groups, and one is interested in properties and patterns that can be expressed in terms of linear equations. More generally, arithmetic combinatorics deals with properties and patterns that can be expressed via additions and multiplications.

In the past ten years, additive and arithmetic combinatorics have been extremely successful areas of mathematics, featuring a convergence of techniques from graph theory, analysis and ergodic theory. They have helped prove long-standing open questions in additive number theory, and they offer much promise of future progress.

Techniques from additive and arithmetic combinatorics have found several applications in computer science too, to property testing, pseudorandomness, PCP constructions, lower bounds, and extractor constructions. Typically, whenever a technique from additive or arithmetic combinatorics becomes understood by computer scientists, it finds some application.

Considering that there is still a lot of additive and arithmetic combinatorics that computer scientists do not understand (and, the field being very active, even more will be developed in the near future), there seems to be much potential for future connections and applications.

1 Introduction

Additive combinatorics differs from more classical work in extremal combinatorics for its study of subsets of the integers, or of more general abelian groups (rather than of graphs, hypergraphs, and set systems), and for asking questions that can be phrased in terms of linear equations (rather than being about cuts, intersections, subgraphs, and so on). So whereas a typical question in extremal combinatorics is:

> *At most how many edges can there be in an n-vertex graph containing no triangle?*

a typical question in additive combinatorics is

> *At most how many elements can there be in a subset of $\{1, \ldots, N\}$ containing no length-3 arithmetic progression?*

In *arithmetic* combinatorics one is interested in properties defined in terms of both addition and multiplication (over the integers, or possibly over more general rings). A typical question in arithmetic combinatorics is:

†luca@cs.berkeley.edu. U.C. Berkeley, Computer Science Division. This material is based upon work supported by the National Science Foundation under grant No. CCF-0729137 and by the BSF under grant 2002246.

If A is a set of N integers, how large or small can the sum-set $A+A := \{x+y : x, y \in A\}$, the difference set $A-A := \{x-y : x, y \in A\}$, and the product set $A*A := \{xy : x, y \in A\}$ be, and what are the relations between the sizes of these sets?

Additive (and arithmetic) combinatorics has grown to become a very active and successful area of mathematics. It has incorporated, and fused, techniques from graph theory, analysis, and ergodic theory, and a deep understanding is emerging of the connections between these seemingly disparate approaches.

A major success of this area of research has been the proof of the Green-Tao theorem [GT08b] that the primes contain arbitrarily long arithmetic progressions. This was one of the long-standing open questions in additive number theory, and it was completely open before the work of Green and Tao.[1] Later, we will say a bit more about their proof, but for now let us consider the very high-level structure of their argument. Green and Tao prove, roughly speaking, that

1. If a set D of integers has positive density inside a *pseudorandom* set R of integers[2] then D contains arbitrarily long arithmetic progressions

2. There is a set of integers R which is pseudorandom and such that the primes have positive density inside R.

The appeal of this approach is that the number-theoretic part of the argument is entirely contained in (2), and the bulk of the proof is in (1), which is a purely additive combinatorial statement whose proof requires no number theory.

It seems possible that additive combinatorics could help resolve other questions about additive patterns in the primes by a similar two-part argument, in which one proves that sets of a certain type must contain a given pattern, and then shows that the primes are a set of that type. This is very exciting because many of the classical conjectures in number theory are about additive patterns in the primes: the twin primes conjecture is the question of whether the equation $y - x = 2$ has infinitely many solutions in primes x, y, and the Goldbach conjecture is whether the equation $x + y = 2n$ has always solutions in primes x, y for every integer n. (Currently, however, the known techniques inherently fail when applied to problems such as the twin primes conjecture or the Goldbach conjecture.)

The techniques developed in additive and arithmetic combinatorics have had, over time, a number of applications to theoretical computer science, in such diverse areas as the design of property testing algorithms, the design of probabilistically checkable proofs, the construction of pseudorandom generators, the proof of communication complexity lower bounds, and the study of the notion of pseudoentropy. Typically, whenever a technique from additive combinatorics becomes known to computer scientists, it finds applications. Given that there is still a lot about additive combinatorics that computer scientists do not understand well, it is safe to predict that more applications will be found in the future, and perhaps that the reason for these connections will become more clear.

In the next section we are going to briefly review the story of *Szemerédi's Theorem* and of the Green-Tao theorem, probably the most remarkable success story in additive combinatorics.

[1] It was known that there are infinitely many length-3 progressions of primes, but it was open even to show that there are infinitely many length-4 progressions.

[2] D having positive density inside R means that there is a constant $\delta > 0$ such that, for infinitely many N, if we call $D_N := D \cap \{1, \ldots, N\}$ and $R_N := R \cap \{1, \ldots, N\}$, we have $|D_N \cap R_N| \geq \delta \cdot |R_N|$. The notion of pseudorandomness used here has quite a technical definition.

Then we will review the several computer science applications that have arisen from the various techniques that were developed to prove these two theorems. We shall conclude with a review of techniques, applications, and open questions from *arithmetic* combinatorics, and its connections to *randomness extractors* in computer science.

There are several excellent references for the reader who is interested in learning more about this area. A survey paper by Green [Gre04] and Tao's [Tao07] notes for his FOCS 2007 tutorial are excellent starting points for a computer scientist. Next, a reader might enjoy a more technical survey by Green [Gre06] and Tao's lecture at ICM 2006 [Tao06a]. The book by Tao and Vu [TV06] is an extensive and very well written review of the area. Notes and videos of a 2007 Princeton short course on additive combinatorics for computer scientists are freely available online.[3] Terence Tao's blog *What's New*[4] contains several wonderful expository notes on the subjects we discuss in this paper. Additional expository notes can be found at *In Theory*.[5]

2 Szemerédi's Theorem and The Green-Tao Theorem

2.1 Szemerédi's Theorem

Szemerédi's theorem is one of the key results of additive combinatorics. It has at least four, rather different, proofs.

Van der Waerden's Theorem and the Erdős-Turán Conjecture. A starting point for this story is van der Waerden's 1927 theorem that, for every coloring of the integers with a finite number of colors, there are arbitrarily long monochromatic arithmetic progressions. A finitary version of the theorem says that for every constants c, k, there is an $N(c,k)$ such that for every $N > N(c,k)$, no matter how we color the integers $\{1,\ldots,N\}$ with c colors, there will be a monochromatic arithmetic progression of length k. (The finitary and infinitary statements are equivalent, via a non-trivial compactness argument.)

In 1936, Erdős and Turán conjectured that the coloring in Van der Waerden's theorem was not an essential assumption, and that the "true reason" why the theorem is true is that every dense enough set of integers (in particular, the largest color class in Van der Waerden's theorem) must contain long arithmetic progressions. Specifically, Erdős and Turán conjectured that for every density $0 < \delta < 1$ and every integer k there is an $N(\delta,k)$ such that every subset $A \subseteq \{1,\ldots,N\}$ of cardinality at least δN contains a length-k arithmetic progression, provided $N > N(\delta,k)$.

Roth's Proof. This conjecture became one of the great unsolved problems in Ramsey theory. In 1953, Roth [Rot53] proved the $k = 3$ case of the Erdős-Turán conjecture. In Roth's proof, $N(\delta,3)$ is doubly exponential in $1/\delta$. In other words, Roth proves that a subset of $\{1,\ldots,N\}$ containing at least about $N/\log\log N$ elements must contain a length-3 progression. The proof is analytical, and uses Fourier analysis.

There is a simple reduction that shows that, to prove Szemerédi's theorem over the integers, it is enough to prove it for the additive group $\mathbb{Z}/N\mathbb{Z}$, with N prime. (That is, it is enough to look for arithmetic progressions mod N in dense subsets of $\{1,\ldots,N\}$.) Indeed, Roth's proof is framed in the modular setting. As noted by Meshulam [Mes95], Roth's proof becomes particularly clean if one carries it out in (the additive group of) a finite vector space such as $\mathbb{F}_3^n$ instead of $\mathbb{Z}/N\mathbb{Z}$. In the

[3] `http://www.cs.princeton.edu/theory/index.php/Main/AdditiveCombinatoricsMinicourse`

[4] `http://terrytao.wordpress.com`

[5] `http://lucatrevisan.wordpress.com`

finite vector space setting the proof also becomes quantitatively better, and it shows that in every subset $A \subseteq \mathbb{F}_3^n$ of size at least about $3^n/n$ there are three points in arithmetic progressions (that is, three points of the form $a, a+b, a+b+b$). The bound is of the form $N/\log N$, where N is the size of the group, which is much better than the $N/\log\log N$ bound in the case of $\mathbb{Z}/N\mathbb{Z}$. The proof in $\mathbb{F}_3^n$ seemingly makes very strong use of the vector space structure, but Bourgain [Bou99] has been able to "simulate" this proof in $\mathbb{Z}/N\mathbb{Z}$ by using the notion of a "Bohr set" in $\mathbb{Z}/N\mathbb{Z}$ to play a role analogous to that of "subspace" in $\mathbb{F}_3^n$. Bourgain obtains a bound that is about $N/\sqrt{\log N}$ in his 1999 paper [Bou99] and a bound around $N/(\log N)^{2/3}$ in more recent unpublished work. This improves over previous work by Heath-Brown [HB87] and Szemerédi [Sze90].

Szemerédi's Proof and the Regularity Lemma. After nearly 40 years, Szemerédi [Sze69, Sze75] proved the Erdős-Turán conjecture in the general case in 1975, and this is the result that is now known as Szemerédi's Theorem. Szemerédi's proof is combinatorial, and works via a reduction to van der Waerden's Theorem. An important ingredient in the proof is the *Szemerédi Regularity Lemma,* which, roughly speaking, asserts that every graph can be "approximated" by a very simple object whose "complexity" depends only on the quality of the approximation, not on the number of vertices of the graph. More specifically, the Regularity Lemma states that if $G = (V, E)$ is a graph, then for every ϵ and ℓ there is a $k = k(\epsilon, \ell) \geq \ell$ such that V can be partitioned into disjoint blocks of vertices $B_1, \ldots, B_k$ so that for at least a $(1-\epsilon)$ fraction of the pairs of blocks B_i, B_j, the edges between B_i and B_j form an "ϵ-regular" bipartite graph. In turn, this means that for every subset $S \subseteq B_i$ and subset $T \subseteq B_j$, the number of edges between S and T is approximately the same as we would expect if the edges between B_i and B_j were laid out at random, that is, it is a $|S| \cdot |T|/(|B_i| \cdot |B_j|) \pm \epsilon$ fraction of the total number $E(B_i, B_j)$ of edges between B_i and B_j. The point is that if B_i, B_j is an ϵ-regular pair (meaning, the above property holds), then, up to some approximation, we may "forget" what are the actual edges between B_i and B_j, and just imagine that they are the edges of a random bipartite graph with edge probability $E(B_i, B_j)/(|B_i| \cdot |B_j|)$. We may also "forget" about the non-regular pairs, so we are left with a collection of a constant number of bipartite random graphs. (Which is the "constant complexity approximation" we claimed.) Unfortunately the Szemerédi Regularity Lemma has very bad quantitative bounds: k is a tower of exponentials in ϵ and ℓ. Unfortunately, Gowers [Gow97] has shown that this tower-of-exponentials dependency is necessary. This affects Szemerédi's proof of the Erdős-Turán conjecture: in his proof, $N(\delta, 4)$ is a double tower of exponentials; one only gets that a subset of $\{1, \ldots, N\}$ of size about $N/\log^*\log^* N$ must contain a length-4 progression.[6]

Furstenberg Proof and Ergodic Theory. In 1977 Furstenberg [Fur77] found a completely different proof. He proved a "transfer theorem" showing that results about arithmetic progressions (and other related structures) can be derived from statements about certain continuous objects, and then he proved the continuous statements by "standard methods" in Ergodic Theory. The transfer theorem uses a result that requires the Axiom of Choice and so, while the proof establishes that finite values of $N(\delta, k)$ exist, it is impossible, even in principle, to extract any quantitative bound. Furstenberg's proof is very robust to changes in the problem statement, and it has been used as a starting point to prove the existence of other patterns in dense sets of integers. (These other results have been extremely hard to replicate by purely combinatorial techniques.)

[6] Indeed, even achieving the bound $N/\log^*\log^* N$ requires Shelah's proof [She90] of van der Waerden's theorem, which came much later; the bound coming from Szemerédi's proof and the proof of van der Waerden's theorem available in 1975 was around $N/\alpha(N)$, where α is the inverse of the Ackermann function.

Gowers's Analytical Proof and Gowers Uniformity. A great quantitative breakthrough came with the work of Gowers [Gow98, Gow01], who returned to the analytic approach of Roth, and was able to devise new analytic techniques to deal with the case of arithmetic progressions of length 4 or more. As in Roth's proof, Gowers is able to show the existence of arithmetic progressions in subsets of $\{1, \ldots, N\}$ of size at least $N/(\log \log N)^{c_k}$, where $c_k > 0$ is a positive constant for each fixed k.

Roth's proof was based on the following reasoning: let $A \subseteq \mathbb{Z}/N\mathbb{Z}$ be a set of δN elements in which we want to show the existence of length-3 arithmetic progressions. Then we look at the Fourier coefficients of the indicator function $1_A : \mathbb{Z}/N\mathbb{Z} \to \{0, 1\}$ of A: if all the coefficients (except the zero-th) of 1_A are small (smaller than about $\delta^2/2$) then A has about as many length-3 progressions as a random set of density δ; in particular at least one length-3 progression. Otherwise, A is correlated with a character, and this can be exploited to reduce the problem of finding a length-3 arithmetic progression in A to the problem of finding a length-3 arithmetic progression in A', a subset of $\mathbb{Z}/N'\mathbb{Z}$ where N' is about $\sqrt{N}$ and, crucially, A' has now density $\delta + \Omega(\delta^2)$ in N'. Then we repeat the same reasoning, and each time we reduce to a case in which we are looking for arithmetic progressions in a denser set. Once the density exceeds $2/3$, then we know that there are length-3 arithmetic progressions by a simple probabilistic argument.

The second part of Roth's proof holds for arithmetic progressions of any length. If 1_A has a large Fourier coefficient, then the denser subset $A' \subseteq \mathbb{Z}/N'\mathbb{Z}$ that we construct is such that the existence of a length-k arithmetic progression in A' implies the existence of such a progression in A, for every k. Unfortunately, the first part does not hold any more: all the Fourier coefficients of 1_A being small is not a sufficient condition for A to have approximately the same number of length-4 progression as a random set of the same size.

To get around this difficulty, Gowers introduces a family of norms for functions $f : \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$, which are well defined for any function $f : G \to [-1, 1]$ where G is an abelian group (indeed the norms are also well defined if $f : G \to \mathbb{C}$ is a complex-valued bounded function); these norms are now called *Gowers uniformity norms*, or just *Gowers norms*. Gowers shows that the if $A \subseteq \mathbb{Z}/N\mathbb{Z}$ has size δN, and the k-th Gowers norm of the function $1_A - \delta$ is small (at most $\delta^{O_k(1)}$) then A contains approximately as many length $(k + 1)$ arithmetic progressions as a random set of density δ; in particular A contains arithmetic progressions. If the k-th Gowers norm of $1_A - \delta$ is not small, then it can be argued that 1_A has a sort of "low degree local approximation" by a polynomial-like object of degree $(k - 1)$, and this can be used to construct a set A' of density $\delta + \delta^{O_k(1)}$ in $\mathbb{Z}/N'\mathbb{Z}$, where $N' = N^{\Omega_{k,\delta}(1)}$ such that finding arithmetic progressions of length $k + 1$ in A' gives us progressions of length $k + 1$ in A.

Gowers's work on the uniformity norm introduced a sort of "higher degree Fourier analysis" which has been greatly developed by Green and Tao, and which appears to be a very powerful tool.

The Proof via Hypergraph Regularity. Ruzsa and Szemerédi [RS76] showed in 1976 how to prove Roth's Theorem (the case of Szemerédi's Theorem for progressions of length 3) as a simple (but rather clever) application of the Regularity Lemma. They show that: (1) from the Regularity Lemma one can derive the *Triangle Removal Lemma*, stating that if an n-vertex graph has $\Omega(n^2)$ edge-disjoint triangles, then it has $\Omega(n^3)$ triangles in total, and (2) Roth's Theorem follows from the Triangle Removal Lemma because from a subset $A \subseteq \mathbb{Z}/N\mathbb{Z}$ with no length-3 arithmetic progression we can construct a graph that has $3N$ vertices, $N|A|$ edge-disjoint triangles, and no other triangle. (So we have a contradiction to the Triangle Removal Lemma if $|A| = \Omega(|N|)$.) Part (2) of the proof can be extended to length-k progressions provided one has a corresponding "k-clique removal lemma in $(k - 1)$-uniform hypergraphs." By analogy with the Ruzsa-Szemerédi proof one would

expect the latter statement to follow from a generalization to hypergraphs of the Regularity Lemma. Unfortunately, when one tries and formulate a Hypergraph Regularity Lemma, one tends to get either a statement that is too weak to imply the removal lemma, or a statement that is so strong that it's actually false. Recently, however, the long-standing open question of devising a Ruzsa-Szemerédi-like proof of the full Szemerédi's Theorem has been resolved, in independent work by Nagle, Rödl, Schacht, and Skokan [RS04, NRS06], Gowers [Gow07], and Tao [Tao06c].

2.2 Patterns in the Primes

The Green-Tao Theorem. A long standing conjecture in additive number theory was the existence of arbitrarily long (and infinitely many) arithmetic progressions in the primes. While it was known that there are infinitely many length-3 arithmetic progressions in the primes, even the case of length-4 progressions was open until recently.

There were two approaches to this question. One approach is to view the question as a special case of the Hardy-Littlewood conjecture. Hardy and Littlewood conjectured that, roughly speaking, for every system of linear equations, the number of solutions to the equations in primes is approximately the same as the number of solutions given by a probabilistic model, in which instead of the primes we consider a random set of integers selected so that they do not have small factors and so that they have the same density as the primes. This is a very strong statement, which, for example, immediately implies the twin primes conjecture because, in this model, the equation $y - x = 2$ has $\Theta(N/(\log N)^2)$ solutions in the interval $\{1, \ldots, N\}$. Hence the Hardy-Littlewood conjectures not only predicts the existence of infinitely many twin primes, but even the existence of $\Theta(N/(\log N)^2)$ twin primes pairs in $\{1, \ldots, N\}$, and, indeed, even the constant implicit in the $\Theta()$ notation. (Experiments are in agreement with the prediction.)

Needless to say, proving this kind of pseudorandomness for the primes is out of reach of current techniques (but see below for substantial progress made by Green and Tao in the past two years), but strong pseudorandomness results have been established for the sets of "almost primes," of integers with few divisors. We know, for example, that $y - x = 2$ has infinitely many solutions in which at least one of x, y is prime, and the other integer has at most two divisors.

The other approach was via stronger versions of Szemerédi's theorem. We know from Gowers's proof that any subset of $\{1, \ldots, N\}$ with at least about $N/(\log\log N)^{c_k}$ elements contains a length-k arithmetic progression: If we could show it for sets of size at least $N/\log N$, then we would be done, because the theorem would directly apply to the primes. Unfortunately the current techniques seem very far from such a result; even for progressions of length 3 the most recent results of Bourgain fail to reach the required density.

The great achievement of Green and Tao [GT08b] is to *combine* these two approaches, and to take advantage of the known partial results both on the pseudorandomness approach (which apply to the almost primes but not yet to the primes) and on the density approach (which apply to set of positive density, or even of density $1/(\log\log n)^c$, but not to the density of the primes).

As we mentioned in the introduction, the high level structure of the proof is to show: (1) Every set D having positive density inside a pseudorandom set R of integers must have arbitrarily long arithmetic progressions; and (2) The primes have positive density inside the set of almost primes, and the set of almost primes is pseudorandom.[7]

[7]This is a rather oversimplified account; instead of dealing with actual sets, for example, the proof refers to "measures," that is mappings of the integers into non-negative reals – sets correspond to measures taking value in $\{0, 1\}$. Instead of the primes one looks at a continuous approximation, and instead of the almost primes one looks at a rather complicated measure. The pseudorandomness condition is quantitative, and D and R are restricted to an interval $\{1, \ldots, N\}$, so that the entire proof is finitary.

Most of the work is done to establish part (1), and to do so while requiring a notion of pseudorandomness for R that be weak enough to be provable for the almost primes. The main step in proving part (1) is in showing that if D is a dense subset of a pseudorandom set R, then there is a set M, of positive density in all the integers, which is a "model" for D and, in a certain technical sense, is "indistinguishable" from D. (Technically, the Gowers norm of the difference between the indicator function of D, appropriately scaled, and the indicator function of M, is small.) Then by Szemerédi's theorem, for every k, M contains length k progressions; indeed a random k-element progression in $\{1, \ldots, N\}$ is entirely contained in M with probability $\Omega_k(1)$. (A fact that follows from the standard Szemerédi's theorem.) The proof is completed by arguing that D too must contain length-k arithmetic progressions, otherwise the indistinguishability condition between D and M would be violated.

Other Additive Patterns. Green and Tao have developed an approach to treat several special cases of the Hardy-Littlewood conjecture, rather than just arithmetic progressions, and also to obtain the exact asymptotics. They lay out a program in [GT06] that works as follows: given a system of linear equations (such that we would like to count the number of solutions in primes), it is possible to associate a "complexity" k to the system, and the number of solutions in primes to the system is as predicted by the Hardy-Littlewood conjecture provided that, roughly speaking, the indicator function of the primes has small $(k+1)$-th Gowers norm. Then they show that verifying the latter condition can be broken up into two pieces: showing that (1) the Mobius function (an approximation of the indicator function of the primes) has low correlation with "k-step nilsequences" (a generalization of degree-k polynomials) and (2) that a function that has low correlation with k-step nilsequences has low $(k+1)$-th Gowers norm.

Unfortunately this program does not cover the twin primes conjecture, because the equation $y - x = 2$ has *infinite* complexity according to the Green-Tao definition. The equations defining length k arithmetic progressions, however, have complexity $k-2$ and there are several additional examples of low-complexity equations.

Green and Tao have recently completed the proof of part (1) of their program [GT09, GT08a], for all k. Regarding part (2), the conjecture that low correlation with nilsequences implies low Gowers norm is the *Gowers inverse conjecture*, and so far Green and Tao have proved the $k = 2$ case [GT05].

It has also been of interest, as a "model" for the case of functions $f : \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$, to consider a version of the Gowers inverse conjecture in finite vector spaces, that is for bounded functions $f : \mathbb{F}_p^n \to [-1, 1]$. In such a case, the finite field Gowers inverse conjecture, as formulated by Green and Tao, is that if $f : \mathbb{F}_p^n \to [-1, 1]$ has $(k+1)$-th Gowers norm at least δ, then there is a polynomial q of degree at most k such that $f()$ and $\omega^{q()}$ have correlation $\Omega_\delta(1)$, where ω is a p-th root of unity. This is proved for odd p in [GT05]. (These simple polynomial phase functions replace the much more complex notion of nilsequence when we work in finite fields.)

Gowers and Wolf [GW07] show that there are systems of linear equations of Green-Tao complexity larger than k whose number of solutions is still controlled by the $(k+1)$-th Gowers norm. They provide an alternative, tighter, notion of complexity which can be substituted to the Green-Tao notion to prove more general results.

3 Applications of the Szemerédi Regularity Lemma

The Regularity Lemma gives a way to approximate an arbitrary graph by a "low complexity model," and so it is an appealing starting point for algorithmic applications: if we have to solve a

computational problem on a given graph, we may first construct a Szemerédi approximation, then solve the problem on the approximating graph (which might be easier given the simpler structure of the latter) and then argue that an exact solution in the approximating graph is an approximate solution in the original graph.

Szemerédi's proof of the Regularity Lemma is non-constructive, that is, it shows that a partition satisfying the conditions of the lemma exists, but it does not show how to construct it. Alon et al. [ADL$^+$94] provide an algorithmic proof of the Regularity Lemma, and present various applications. Frieze and Kannan [FK96] showed that one can find polynomial time approximation schemes for graph optimization problems in dense graphs via the Regularity Lemma, and that, for such applications, a weaker version of the Regularity Lemma is sufficient. The Frieze-Kannan *Weak Regularity Lemma* provides a weaker notion of approximation (or regularity) for the partition, but the number of blocks in the partition is "just" singly exponential in the approximation parameter ϵ, instead of being a tower of exponentials. Frieze and Kannan also give an algorithmic proof of the Regularity Lemma via eigenvalue calculations [FK99a] and formulate a general Regularity Lemma for matrices [FK99b] (the graph Regularity Lemma being the special case for matrices which are adjacency matrices of graphs).

In 1996, Goldreich, Goldwasser and Ron [GGR98] introduced the general notion of *property testing*, abstracting previous work on polynomial low-degree tests motivated by program testing and PCP. The notion is to distinguish an object that has the property from an object that is ϵ-far from the property; the latter is defined depending on the problem and the representation. For dense n-vertex graphs in the adjacency matrix representation, being ϵ-far means that one has to add or remove at least ϵn^2 edges to turn one graph into the other. In this setup, the Ruzsa-Szemerédi paper gives a property testing algorithm of complexity $O_\epsilon(1)$ for the property of being *triangle-free*. (Pick $q = O_\epsilon(1)$ vertices, and check if there is a triangle among them; if the graph is triangle free, then the algorithm will never see a triangle; be if we need to remove $\geq \epsilon n^2$ edges in order to remove all triangles, then by the Triangle Removal Lemma, the graph has $\Omega_\epsilon(n^3)$ triangles and so, for a proper choice of q, the algorithm has probability at least $3/4$ of seeing a triangle.)

Alon et al. [AFKS00] show that previous results in extremal combinatorics imply testers for the k-colorability problem studied in [GGR98], although less efficient ones than the testers in [GGR98]. Alon et al. also show how to prove property testing results via the Regularity Lemma. Alon [Alo02] studies property testing problems of the following form: for a fixed graph H, given a graph G, does G contain an edge-induced copy of H? If H is a triangle, this testing problem is the Triangle Removal Lemma, and Behrend's example [Beh46] (plus some extra work) shows that the complexity of any tester, while constant in ϵ, cannot be polynomial in $1/\epsilon$. Alon shows that this is the case whenever H is not bipartite, while, if H is bipartite, the complexity is always polynomial in $1/\epsilon$.

The Regularity Lemma, and its variations, has remained an important tool in the study of graph property testing, and of more general property testing problems. In a few instances, research motivated by property testing has intersected the research in additive combinatorics motivated by the search for a "Ruzsa-Szemerédi"-style proof of Szemerédi's theorem.

For example, Alon and Shapira [AS08] prove a stronger form of the Regularity Lemma as a step in their proof that every monotone graph property is testable in the adjacency matrix model; this stronger form is analogous to a form of Regularity Lemma proved by Tao [Tao06b] as a step towards the hypergraph regularity proof of Szemerédi's theorem. More recently, Austin and Tao [AT08] have used hypergraph regularity machinery to prove general results about property testing in hypergraphs.

4 Applications of Gowers Norms

A first connection between Gowers norms and computer science comes from the problem of *low degree test*, in which given a function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ we want to distinguish the case in which f is a degree-d polynomial from the case in which f is ϵ-far (meaning, it differ in at least ϵ fraction of inputs) from all degree-d polynomial. This question has been studied because of its applications to program testing and probabilistically checkable proofs [BLR93, BFL91, GLR$^+$91], and it remains an important problem. It is a type of property testing problem, and in fact it was the inspiration for the general definition of property testing given by Goldreich, Goldwasser and Ron.

Typically, the low degree test problem is studied in settings where d is smaller than the size p of the field. Alon et al. [AKK$^+$03] considered the setting of degree larger than field size, and specifically the case of testing low-degree in $\mathbb{F}_2^n$. They proposed and analyzed a testing algorithm, and showed that if the test, for degree d, accepts with probability $1 - \epsilon$ a function f, then there is a degree d polynomial over $\mathbb{F}_2$ that differs in at most a $O(\epsilon 2^d)$ fraction of inputs from f.

Up to normalization, the acceptance probability of the test of Alon et al. is *precisely* the $(d+1)$-th Gowers norm of f, and the result of Alon et al. shows that if the $(d+1)$-th Gowers norm of a function $f : \mathbb{F}_2^n \to [-1, 1]$ is very close to 1, then f is very close to a function of the form $(-1)^q$, where q is a degree-d polynomial.

Samorodnitsky [Sam07] shows that if the test of Alon et al. for quadratic polynomials accepts a function f with probability at least $1/2 + \epsilon$, then f has correlation $\Omega_\epsilon(1)$ with $(-1)^q$, where q is a quadratic polynomial. This establishes the Gowers inverse conjecture in $\mathbb{F}_2^n$ for the third Gowers norm, a case that was not covered by the work of Green and Tao [GT05], which deals with groups of odd order. The work of Samorodnitsky was independent of the work of Green and Tao.

Samorodnitsky and Trevisan [ST06] show that, for every k, if the k-th Gowers norm of a function $f : \mathbb{F}_p^n \to [-1, 1]$ is at least ϵ, then f has a variable of *influence* at least $\Omega_{\epsilon,k}(1)$. This is incomparable with the Gowers inverse conjecture and it has applications to the design of probabilistically checkable proofs. (We refer to the paper for more details.)

Bogdanov and Viola [BV07] use Gowers norms to analyse a pseudorandom generator construction that fools low-degree polynomials in finite fields. The analysis of the generator for polynomials of degree d is conditional on the inverse conjecture for the d-th Gowers norm in finite fields, and, so their result is unconditional for cubic polynomials because of the work of Green, Tao, and Samorodnitsky.

Lovett [Lov08] and Viola [Vio08] provide unconditional constructions whose analysis does not directly use the Gowers norm, but rather the structure of inductive use of Cauchy-Schwarz which is typical of proofs that use the Gowers norm.

Viola and Wigderson [VW07] use Gowers norms to prove "XOR lemmas" for correlations with low degree polynomials and for communications complexity. For low degree polynomials, they use the low-degree test of Alon et al. [AKK$^+$03] to argue that correlation with low-degree polynomials is approximated by the Gowers norm, and then they argue that the Gowers norm is reduced exponentially by XOR-ing functions on independent inputs; the structure of the proof for communication complexity is similar, and, instead of the low-degree test of Alon et al. Viola and Wigderson use a result of Chung and Tetali [CT93] and Raz [Raz00] (see also [BNS92]) which relates the number-on-the-forehead communication complexity of a function [CFL83] to an expression which is similar to the Gowers norm.

The last year has seen two surprising developments about the Gowers inverse conjecture in finite fields. Lovett, Meshulam, and Samorodnitsky [LMS08] and Green and Tao [GT07a] found a counterexample in $\mathbb{F}_2^n$ to the conjecture for the fourth norm (the first case that was still open);

in the same paper, Green and Tao [GT07a] prove a weak form of the Gowers k-th norm inverse conjecture (which hold only for functions f which are themselves polynomials of low degree) for the case in which the characteristic of the field is larger than k. Bergelson, Tao and Ziegler [BTZ09, TZ08a] have recently proved the full Gowers inverse conjecture in finite fields, again provided the characteristic of the field is large enough.

5 Ergodic-Theoretic Techniques

I think it is fair to say that the ergodic-theoretic techniques in additive combinatorics are the least understood by theoretical computer scientists, and certainly by the author. The power of these techniques to prove purely combinatorial results has been demonstrated time and again, from Furstenberg's proof of Szemerédi's theorem in the 1970s to the recently announced proof of the Gowers inverse conjecture in finite fields [BTZ09, TZ08a]. It would certainly be useful to computer scientists to understand better how these techniques work. There are at least two advantages to this family of techniques.

One is that they allow to pass from an finite quantitative statement that may involve several ϵs and δs to a cleaner qualitative infinitary statement. (See the post by Terence Tao, on "hard analysis" versus "soft analysis."[8]) The work in computer science on "graph limits" studied by Lovász and Szegedy and by Borgs et al. [LS06, BCL$^+$06] is an example of how the approach of turning a discrete quantitative problem into an infinitary qualitative problem can be advantageous in computer science too.

Once cast as an infinitary problem, usually about a specific measure space with a specific measure-preserving operator, the next step is to generalize the statement that one is trying to prove to all measure spaces, and all measure-preserving operators, essentially forgetting all the structure of the problem. This seemingly makes the task of proving the statement harder, however since many proof techniques have an inductive (usually via *transfinite* induction) structure, it can be easier to prove a more general statement, because then one can rely on a more powerful inductive hypothesis.

Let us return for a moment to the Green-Tao proof that the primes contain arbitrarily long arithmetic progressions. Recall that a key step in the proof was, given a pseudorandom set R of integers and a dense subset D of R, to construct a dense model set M of D such that M and D are indistinguishable. Green and Tao [GT08b] describe the argument used in the proof as a finitary version of arguments in ergodic theory.

A core step in the proof, which appears in more explicit and general form as [TZ08b, Theorem 7.1], is essentially the above statement, but with "pseudorandom" and "indistinguishable" defined as in complexity theory, in terms of bounds on the distinguishing probability of certain families of adversaries. (We should remark that this is neither the formalism nor the intuition used in [GT08b, TZ08b], although it would be possible to state their result in this language.) Some work is then needed to go from this notion of "indistinguishability" to a notion defined in terms of Gowers norms.

Reingold et al. [RTTV08] give a new proof of the "dense model theorem" [TZ08b, Theorem 7.1] via Nisan's proof of Impagliazzo hardcore set theorem [Imp95], using duality of linear programming, in the form of optimal strategy for two-players zero sum games. Gowers [Gow08] independently discovers the same approach in rather different language, and finds several applications. Impagliazzo

[8] `http://terrytao.wordpress.com/2007/05/23/soft-analysis-hard-analysis-and-the-finite-convergence-principle/`

[Imp08] shows that the dense model theorem can be derived directly from the statement of a strong form of his hardcore set theorem; such a strong form of the hardcore set theorem had been proved by Holenstein [Hol05].

The proof of Reingold et al. [RTTV08] involves a reduction of polynomial complexity, and it gives a new characterization of the notion of pseudoentropy. In particular, it implies that if one has partial information about the seed of a pseudorandom generator then the output of the generator still has large pseudoentropy. This latter result was discovered independently by Dziembowski and Pietrzak [DP08] who use it to construct stream ciphers that remain secure even if partial information about the key leaks.

The construction of a dense model via duality of linear programming in [RTTV08, Gow08] is non-constructive, while the model set is explicitly constructed in the "finitary ergodic theoretic" proof of [GT08b, TZ08b]. The latter proof, however, when applied to complexity theory, involves a reduction of exponential complexity. Trevisan, Tulsiani and Vadhan [TTV09] provide a third (and a fourth) proof that is constructive and, in complexity-theoretic settings, involves a reduction of polynomial complexity.

6 Arithmetic Combinatorics

In this section we shall limit ourselves to discuss *sum-product* theorems in finite fields. The motivation for such results came from the *Kakeya problem* in analysis.

A *Besicovitch set*, or *Kakeya set*, is a set that, for every direction, contains a unit-length segment parallel to that direction. The Kakeya problem is, roughly speaking, how "large" such a set must be. It is known that, for every n, there are Kakeya sets in $\mathbb{R}^n$ of Lebesgue measure 0; if, however, one looks at more refined measures of "size," the Kakeya conjecture is that the Hausdorff and Minkowski dimension of any Kakeya set is n, the largest possible. This has various applications in analysis and is the source of several fascinating connections; the interested reader will enjoy Terence Tao's excellent expository article on this problem [Tao01].

In the late 1990s, Bourgain showed that one could make progress on this problem by exploiting one of the staples of additive combinatorics: an understanding of the structures of subsets A of a group such that $A + A := \{a + b : a \in A, b \in A\}$ is small. This kind of combinatorial reasoning also proves lower bounds (and, usually, it does so in a cleaner way) to the size of a Kakeya set in $\mathbb{F}_p^n$, a problem which is rather interesting in its own.

In this more combinatorial setup, the *Finite Field Kakeya Conjecture*, formulated by Wolff in 1999 [Wol99], was that a Kakeya set in $\mathbb{F}_p^n$ must have size at least $\Omega_n(q^n)$, where the constant in the $\Omega()$ notation depends on n but not on q. Motivated by this conjecture, Wolff made the *finite field sum-product conjecture* that for every subset $A \subseteq \mathbb{F}_p$, either $|A + A|$ or $|A \cdot A|$ is at least $\min\{p, |A|^{1+\epsilon}\}$ for an absolute positive constant $\epsilon > 0$. If A is a set of integers, Erdős and Szemerédi had shown that this is true, but their proof (and later improvements) uses geometric properties that are not true for finite geometries.

The finite field sum-product conjecture had also been formulated in the 1990s by Avi Wigderson, motivated by the problem of finding "seedless" randomness extractors for independent sources. There had been no progress on this problem since the 1985 paper by Chor and Goldreich [CG88] which had introduced it. Zuckerman [Zuc90] had considered the following question: suppose X, Y, Z are independent random variables ranging over $\mathbb{F}_p$, each of min-entropy at least k; is it true that $X \cdot Y + Z \bmod p$ must have min-entropy at least $(1+\epsilon) \cdot k$, for some absolute constant $\epsilon > 0$? A positive solution to this question would allow to make progress on the extractor problem. Zuckerman showed that a positive answer follows by assuming the *Generalized Paley Graph Conjecture*, which is still

open. Wigderson noted that Zuckerman's conjecture was a "statistical analog" of a sum-product theorem, and so formulated the sum-product conjecture in finite fields as a plausible intermediate step in proving Zuckerman's conjecture.

Bourgain, Katz and Tao [BKT04] proved in 2004 the finite field sum-product conjecture. (For sets A that were not too small and not too large; this restriction was removed by Konyagin [Kon03].) Soon afterwards, Barak, Impagliazzo and Wigderson [BIW04] proved Zuckerman's conjecture about the three independent random variables and made the first progress on seedless extractors for independent sources since 1985. Rapid progress followed, which was quite dramatic, including a 2006 paper by Barak, Rao, Shaltiel, and Wigderson [BRSW06] making the first progress since the 1981 Frankl-Wilson construction [FW81] of Ramsey graphs. There remain many open problems, especially on the task originally formulated by Chor and Goldreich of extracting near unbiased randomness bits from just two independent sources. Bourgain [Bou05] reported some progress on this problem.

It was a great recent surprise that the Kakeya problem in finite fields was proved by Zeev Dvir [Dvi08], a complexity theorist, with a three-page argument. Dvir's ideas have had additional extractor applications [DW08] and are likely to have further impact.

7 Some Major Open Questions

There probably remain additional applications to be discovered of ideas from additive combinatorics to computer science, maybe to circuit complexity, or proof complexity, or computational learning theory. Certainly, much work remains to be done on extractors from two independent sources, a problem from which the ideas from arithmetic combinatorics have had a very substantial impact.

Dvir's proof of the finite field Kakeya conjecture shows that it is also possible for computer scientists to approach some of the well-studied open questions in arithmetic combinatorics. In a spirit of probably unwarranted optimism, I shall now review my three favorite open questions in additive combinatorics. (They are all well studied and perhaps hopeless.)

Open Question 1 (The cap problem in $\mathbb{F}_3$) *What is the size of the largest subset $A \subseteq \mathbb{F}_3^n$ that does not contain length-3 arithmetic progression?*

The Roth-Meshulam proof implies that the size is at most about $3^n/n$, but the best lower bound is only about 2.2^n. Is there a better upper bound? Is there a $(3 - o(1))^n$ lower bound?

Open Question 2 (Bounds in the triangle removal lemma) *The triangle removal lemma states that if an N-vertex graph has at least δN^2 edge-disjoint triangles, then it has at least $\epsilon(\delta) \cdot N^3$ triangles. The dependency of ϵ on δ cannot be polynomial (ϵ can be at most about $\delta^{\log 1/\delta}$) and is at most a tower of exponentials. Is the dependency singly exponential?*

It would certainly be interesting to have any negative result not based on Behrend's example, and it would be remarkable to have a positive result even of tower type but, say, with a tower of logarithmic depth in $1/\delta$, because this would give a result which would not be provable via the Regularity Lemma.

Open Question 3 (The polynomial Freiman-Ruzsa Conjecture – See [Gre04]) *Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a function such that $\mathbb{P}_{x,y,z}[F(x) + F(y) + F(z) = F(x + y + z)] \geq \epsilon$. Is it true that there is a matrix A and a vector b such that $\mathbb{P}_x[F(x) = Ax + b] \geq \epsilon^{O(1)}$?*

It is known how to construct A, b so that the agreement is at least $e^{-O(poly(1/\epsilon))}$, and this was proved independently by Green [Gre04] and Samorodnitsky [Sam07]. See a survey paper by Viola [Vio07] for an exposition of the proof. Green and Tao have improved the agreement to $e^{-O(1/\epsilon)}$ [GT07b].

Acknowledgements. I am grateful to Anil Ada, Lane Hemaspaandra, Troy Lee, Shachar Lovett, Ryan O'Donnell, Terry Tao, Salil Vadhan, and Avi Wigderson for their comments, corrections, and pointers to the literature.

References

[ADL+94] Noga Alon, Richard A. Duke, Hanno Lefmann, Vojtech Rödl, and Raphael Yuster. The algorithmic aspects of the regularity lema. *Journal of Algorithms*, 16:80–109, 1994.

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[AKK+03] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing low-degree polynomials over $GF(2)$. In *Proceedings of RANDOM-APPROX*, pages 188–199, 2003.

[Alo02] Noga Alon. Testing subgraphs in large graphs. *Random Structures and Algorithms*, 21(3-4):359–370, 2002.

[AS08] Noga Alon and Asaf Shapira. Every monotone graph property is testable. *SIAM Journal on Computing*, 38(2):505–522, 2008.

[AT08] Tim Austin and Terence Tao. On the testability and repair of hereditary hypergraph properties. arXiv:0801.2179, 2008.

[BCL+06] Christian Borgs, Jennifer T. Chayes, Laszlo Lovasz, Vera T. Sos, Balazs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 261–270, 2006.

[Beh46] Felix A. Behrend. On the sets of integers which contain no three in arithmetic progression. *Proc. Nat. Acad. Sci.*, 23:331–332, 1946.

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3–40, 1991. Preliminary version in *Proc. of FOCS'90*.

[BIW04] Boaz Barak, Russell Impagliazzo, and Avi Wigderson. Extracting randomness using few independent sources. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 384–393, 2004.

[BKT04] Jean Bourgain, Nets Katz, and Terence Tao. A sum-product estimate for finite fields, and applications. *Geometric and Functional Analysis*, 14:27–57, 2004.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993. Preliminary version in *Proc. of STOC'90*.

[BNS92] László Babai, Noam Nisan, and Mario Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992.

[Bou99] Jean Bourgain. On triples in arithmetic progression. *Geometric and Functional Analysis*, 9:968–984, 1999.

[Bou05] Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(1):1–32, 2005.

[BRSW06] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2-source dispersers for sub-polynomial entropy and ramsey graphs beating the Frankl-Wilson construction. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 671–680, 2006.

[BTZ09] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of f^ω. arXiv:0901.2602, 2009.

[BV07] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 41–51, 2007.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proceedings of the 15th ACM Symposium on Theory of Computing*, pages 94–99, 1983.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, April 1988.

[CT93] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM Journal on Discrete Mathematics*, 6(1):110–123, 1993.

[DP08] Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 293–302, 2008.

[Dvi08] Zeev Dvir. On the size of Kakeya sets in finite fields. *Journal of the AMS*, to appear, 2008.

[DW08] Zeev Dvir and Avi Wigderson. Kakeya sets, new mergers and old extractors. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 625–633, 2008.

[FK96] Alan Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science*, pages 12–20, 1996.

[FK99a] Alan Frieze and Ravi Kannan. A simple algorithm for constructing Szemerédi's regularity partition. *Electronic Journal of Combinatorics*, 6, 1999.

[FK99b] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[Fur77] Hillel Furstenberg. Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, 31:204–256, 1977.

[FW81] Peter Frankl and Richard M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, 1981.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

[GLR+91] Peter Gemmell, Richard Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proceedings of the 23rd ACM Symposium on Theory of Computing*, pages 32–42, 1991.

[Gow97] Timothy Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geometric and Functional Analysis*, 7(2):322–337, 1997.

[Gow98] Timothy Gowers. A new proof of Szeméredi's theorem for progressions of length four. *Geometric and Functional Analysis*, 8(3):529–551, 1998.

[Gow01] Timothy Gowers. A new proof of Szeméredi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001.

[Gow07] Timothy Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Annals of Mathematics*, 166:897–946, 2007.

[Gow08] Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. arXiv:0811.3103, 2008.

[Gre04] Ben Green. Finite field models in additive combinatorics. arXiv:math.NT/0409420, 2004.

[Gre06] Ben Green. Montreal lecture notes on quadratic fourier analysis. arXiv:math/0604089, 2006.

[GT05] Ben Green and Terence Tao. An inverse theorem for the Gowers U^3 norm. math.NT/0503014, 2005.

[GT06] Ben Green and Terence Tao. Linear equations in primes. math.NT/0606088, 2006.

[GT07a] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. arXiv:0711.3191, 2007.

[GT07b] Ben Green and Terence Tao. A note on the Freiman and Balog-Szemerédi-Gowers theorems in finite fields. arXiv:math/0701585, 2007.

[GT08a] Ben Green and Terence Tao. The Möbius function is asymptotically orthogonal to nilsequences. arXiv:0807.1736, 2008.

[GT08b] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167:481–547, 2008.

[GT09] Ben Green and Terence Tao. Quadratic uniformity of the mobius function. *Annales de l'Institut Fourier*, 58:18631935, 2009.

[GW07] Timothy Gowers and Julia Wolf. The true complexity of a system of linear equations. arXiv:0711.0185, 2007.

[HB87] David R. Heath-Brown. Integer sets containing no arithmetic progressions. *Journal of the London Mathematical Society*, 35:385394, 1987.

[Hol05] Thomas Holenstein. Key agreement from weak bit agreement. In *Proceedings of the 37th ACM Symposium on Theory of Computing*, pages 664–673, 2005.

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science*, pages 538–545, 1995.

[Imp08] Russell Impagliazzo. Personal Communication, 2008.

[Kon03] Sergei Konyagin. A sum-product estimate in fields of prime order. math.NT/0304217, 2003.

[LMS08] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 547–556, 2008.

[Lov08] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 557–562, 2008.

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. of Combinatorial Theory, Series B*, 6(96):933–957, 2006.

[Mes95] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory Series A*, 71:168–172, 1995.

[NRS06] Brendan Nagle, Vojtech Rödl, and Mathias Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures and Algorithms*, 26:1–67, 2006.

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2), 2000.

[Rot53] Klaus Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[RS76] Imre Ruzsa and Endre Szemerédi. Triple systems with no six points carrying three triangles. In *Proceedings of the Fifth Hungarian Colloquium on Combinatorics*, pages 939–945, 1976. Volume II.

[RS04] Vojtech Rödl and Jozef Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures and Algorithms*, 25(1):1–42, 2004.

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 76–85, 2008.

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 506–515, 2007.

[She90] Saharon Shelah. Primitive recursive bounds for Van der Waerden numbers. *Journal of Symbolic Logic*, 55:887–888, 1990.

[ST06] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 11–20, 2006.

[Sze69] Endre Szemerédi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hung.*, 20:89–104, 1969.

[Sze75] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245, 1975.

[Sze90] Endre Szemerédi. Integer sets containing no arithmetic progressions. *Acta Math. Hungar.*, 56(1-2):155158, 1990.

[Tao01] Terence Tao. From rotating needles to stability of waves: Emerging connections between combinatorics, analysis, and PDE. *Notices of the AMS*, 48(3):294–303, 2001.

[Tao06a] Terence Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. In *Proceedings of the International Congress of Mathematicians*, 2006.

[Tao06b] Terence Tao. Szemerédi's regularity lemma revisited. *Contributions to Discrete Mathematics*, 1:8–28, 2006.

[Tao06c] Terence Tao. A variant of the hypergraph removal lemma. *Journal of Combinatorial Theory, Series A*, 113:1257–1280, 2006.

[Tao07] Terence Tao. Structure and randomness in combinatorics. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 3–18, 2007.

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *Proceedings of the 24th IEEE Conference on Computational Complexity*, 2009.

[TV06] Terence Tao and Van Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

[TZ08a] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. arXiv:0810.5527, 2008.

[TZ08b] Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Mathematica*, 201:213305, 2008.

[Vio07] Emanuele Viola. Selected results in additive combinatorics: an exposition. Technical Report TR07-103, Electronic Colloquium on Computational Complexity, 2007.

[Vio08] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 124–127, 2008.

[VW07] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*, pages 141–154, 2007.

[Wol99] Thomas Wolff. Recent work connected with the Kakeya problem. In *Prospects in mathematics (Princeton, NJ, 1996)*, pages 129–162. AMS, 1999.

[Zuc90] David Zuckerman. General weak random sources. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 534–543, 1990.