Research Statement

Ernie Croot

September 23, 2008

I work mainly in additive combinatorics and in classical analytic number theory. Here I will describe a few of my past, current, and future research projects in rough chronological order.

1 Early work

I have always been intrigued by simple questions that require complex arguments to solve, and while a graduate student I was fortunate enough to solve a nice problem of this sort in combinatorial number theory, posed by the late P. Erdős. This problem asked if when you partition the integers 2 and higher into a finite number of classes, whether some class contains a finite collection of numbers $x_1, ..., x_k$ such that

$$1 \ = \ \frac{1}{x_1} + \cdots + \frac{1}{x_k}.$$

I showed that the answer to the question is 'yes', and my short, though rather technical proof was published in the Annals of Mathematics.

After having solved this question and several more like it in the theory of unit fractions (a unit fraction is a number $1/n$, where n is a non-zero integer), while on a postdoc at U. C. Berkeley, I started looking into a number of problems pertaining to smooth numbers, which are integers having no 'large' prime factors. In particular, I had wanted to improve an old result of A. Balog, which says that the interval $[x, x + x^{1/2+\varepsilon}]$ always contains an integer having no prime divisors exceeding x^{ε}. There are various applications in number theory and cryptology that require that the shorter interval $[x, x + c\sqrt{x}]$ contains such a number, and I discovered a way to make use of the theory of bilinear forms of Kloosterman sums, as developed by W. Duke, J. Friedlander, and H. Iwaniec, to slightly improve on the best result known, due to G. Harman, which says that such intervals contain an integer free of prime divisors exceeding $x^{1/4\sqrt{e}}$. Even though I had the idea for how to do this many years ago, I only wrote it up within the past three years, and it appeared in International Journal of Number Theory.

From the time I first had the idea of how to solve this problem on smooth numbers in $[x, x + c\sqrt{x}]$, I realized that the central obstruction to greatly improving my method, as well as several other methods for attacking the problem, concerns the local distribution of values of certain multiplicative functions (or 'almost multiplicative' functions). I came to realize that the following well-known problem had some of the same types of obstrctions that come up with smooth numbers, yet because it is simpler to state, I thought it was worth thinking about too: Suppose that $f(n)$ is a completely multiplicative function taking on the values $+1$ and -1 with equal frequency; that is,

$$\text{for all } a, b \in \mathbb{N},\ f(ab) = f(a)f(b);\ \text{and,} \sum_{n \leq x} f(n) \ = \ o(x).$$

Then, must $f(n)$ change sign about $x/2$ times over all $n \leq x$? That is, must there exist a set of about $x/2$ values $n \leq x$ such that

$$f(n) \ = \ -f(n+1) \ ?$$

This problem is related to something called the "parity problem" that comes up in the study of sieve methods, and in retrospect I now see that perhaps I chose a quite difficult problem to work on!

A. Hildebrand had the best results on this problem: He showed that for infinitely many x, the number of $n \leq x$ such that $f(n) = -f(n+1)$ exceeds

$$cx/(\log\log x)^4 \text{ (for some } c > 0).$$

Amazing as this lower bound is, it has the unfortunate limitation that it only holds for infinitely many x, instead of for all sufficiently large x. I discovered a slightly intricate method involving elementary diophantine approximation, which is different from Hildebrand's approach, and which gives a lower bound that holds for *all* x; unfortunately, it is somewhat weaker than Hildebrand's bound for those x where Hildebrand gives any non-trivial bound at all. My paper on this appeared in Journal of Number Theory a few years ago.

2 Work on arithmetic combinatorics: arithmetic progressions

Just as I came to Georgia Tech, I began working on a new project in additive combinatorics, namely the Erdős-Turan conjecture. This conjecture is an old, central problem in the subject, which asks simply whether every subset S of the naturals such that

$$\sum_{s\in S}\frac{1}{s} \ = \ \infty,$$

contains arbitrarily long arithmetic progressions. We are at present quite far from proving this even for three-term progressions; however, there are several impressive results due to K. F. Roth, W. T. Gowers, J. Bourgain, E. Szemeredi, T. Tao, B. Green, and others, which show that subsets of the integers in $\{1,...,N\}$ with certain density or structural restrictions, must contain long arithmetic progressions, where here "long" means larger than $\log\log\log\log\log N$.

Although these approaches are brilliant and deep, what one would like to have are structural results that would allow one to instantly see why dense subsets of $\{1,...,N\}$ have long arithmetic progressions, or at the very least a classification of all subsets having fewer than the expected number of k term progressions (a random subset of $\{1,...,N\}$ of density θ has $\sim c_k\theta^k N^2$ arithmetic progressions of length k, for some constant c_k that does not depend on N or θ). This describes a substantial part of my research in this area that I began while at Georgia Tech.

To describe this work, first I need a definition. Given a finite abelian group G, and a density $\theta \in (0,1]$, we say that a subset $S \subseteq G$ is a *critical set* of density at least θ if $|S| \geq \theta|G|$, and if S has the minimal number of thee-term arithmetic progressions among all subsets of G having at least $\theta|G|$ elements. In this context a three-term arithmetic progression is a triple of points $x, y, z \in S$ satisfying $x + y = 2z$. If one knew precisely how many three-term arithmetic progressions that critical sets in $\mathbb{F}_p$ (although $\mathbb{F}_p$ is a field, we only consider the additive structure, which is an abelian group) had, and if they had as many as is widely believed, then one could solve the Erdős-Turan problem for $k = 3$. Rough descriptions of some of my recent results are given as follows:

Theorem 1. Fix a $\theta \in (0,1]$, and consider a critical set in $\mathbb{F}_p$ having at least θp elements. Compute the number of three-term progressions in this set, and normalize by dividing by p^2. Let this number be δ_p. Then, it turns out that as p runs through the prime numbers, δ_p tends to some limit $\delta \in (0,1]$. Thus, the result is saying that critical sets of the same density lying in two different fields have about the same number of three-term arithmetic progressions, upon properly normalizing. This result appeared in Canadian Math Bulletin.

Theorem 2. Given $\theta \in (0,1]$, if S is a critical set of $\mathbb{F}_p$ with at least θp elements, then S contains an arithmetic progression of length $(\log p)^{1/4-o(1)}$. Compared to the best that is known for arithmetic progressions in arbitrary sets of density θ, this is rather long: Gowers, who has the best results, showed that every subset of $\mathbb{F}_p$ of density θ has an arithmetic progression of length at least $\log\log\log\log\log p + c(\theta)$. My result appeared some years ago in Journal of Combinatorial Theory Series A.

Theorem 3. Fix a density $\theta \in (0,1]$. Given a prime p, let S be a critical set in $\mathbb{F}_p$ having at least θp elements. Then, apart from $o(p)$ elements, S is a sumset $A+B$, where $|A| = p^{1-o(1)}$ and $|B| = p^{o(1)}$. In fact, A is what is called a Bohr neighborhood; and, one can deduce from this other structural results, such as that S is approximately the union of several long arithmetic progressions (or is a sumset $A+B$, where A is an arithmetic progression of length p^δ for some $\delta \in (0,1]$). It is worth remarking that this result can be deduced from an old theorem of B. Green called the "arithmetic regularity lemma"; however, my result holds for much lower densities than can be achieved via Green's theorem. This paper has been submitted.

Theorem 4. In joint work with Olof Sisask, I developed a new proof of Roth's theorem that a positive density subset of $\{1,...,N\}$ contains a three-term arithmetic progression. This paper is to appear in Proceedings of the AMS.

I feel that it should be possible to prove much more exact structure theorems on three-term progressions than the ones I have listed above; however, I will hold off describing this here, as they are rather technical.

3 Additive combinatorics: more recent work

3.1 Sum-product inequalities and incidence theorems

In addition to these results on arithmetic progressions, I have also been working recently on some questions related to "sum-product inequalities", and on properties of sumsets.

The work on sum-product inequalities is related to an old problem considered by P. Erdős and E. Szemerédi: Suppose that A is a set of real numbers of size n. Give a lower bound for $\max(|A+A|, |A.A|)$, where here $A+A$ is the set of sums $a+b$, with $a, b \in A$, while $A.A$ is the set of products ab. The point of the question is that it is easy to construct sets A for which $A+A$ is "small" (e.g. just take A to be an arithmetic progression), but for all such constructions one will see that the product set $A.A$ is "large"; similarly, the constructions with $A.A$ small must have $A+A$ large. Erdős and Szemerédi proved that, in fact, there is a positive constant $\varepsilon > 0$ so that for n large enough, one has

$$\max(|A+A|, |A.A|) \ \gg \ |A|^{1+\varepsilon}.$$

This basic result has been sharpened and extended by numerous people, and the most recent result is due to J. Solymosi, who showed in a short and brilliant paper (all Solymosi's papers are short and brilliant!) that one may take $\varepsilon = 1/3$.

Perhaps the most elegant proof of sum-product inequalities, which achieved a lower bound with $\varepsilon = 1/4$, is due to G. Elekes, who applied the Szemerédi-Trotter incidence theorem in a powerful and beautiful way. Essentially, his proof boils down to showing that if $A.A$ and $A+A$ were "small", say both were of size $|A|^{1+\varepsilon}$ for small $\varepsilon > 0$, then there would be too many "rich lines"

$$(a \cdot t, \ b+t), \ t \ \in \ \mathbb{R}, \ \text{where } a, b \in A,$$

passing through the grid

$$(A.A) \ \times \ (A+A).$$

Here, by "rich line", we mean a line that hits a grid in too many places. Note that each of these lines above hits the grid in $|A|$ places.

This led me to wonder what more one could prove much more, especially the following variant of a conjecture of Solymosi:

Conjecture. For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds for all n sufficiently large: Suppose that A and B are sets of n real numbers. Then, there can be at most n^{ε} lines in general position (no three meeting at a point, no line parallel to any others) such that each intersects $A \times B$ in more than $n^{1-\delta}$ points.

I have not yet been able to prove this conjecture; but, I feel that I have all the ingredients to do so, and just need the time to assemble them into a proof. A significant ingredient in my hypothetical proof, and surely a good first step at showing Solymosi's conjecture regardless of the approach, is the following result, which I coauthored with my student Evan Borenstein:

Theorem 5. For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds for all n sufficiently large: Suppose that A and B are sets of n real numbers. Then, if one has a family of lines containing at least n^{ε} distinct slopes, and where each line is parallel to at least n^{ε} others, at least one of them must fail to be $n^{1-\delta}$-rich – i.e. must fail to intersect the grid in at least $n^{1-\delta}$ points.

It is a fairly straightforward exercise to show that Solymosi's conjecture implies this theorem. Also note that this theorem is in spirit related to Solymosi's conjecture, as both pertain to small families of lines ($n^{O(\varepsilon)}$ lines) that are "rich" in a grid.

Besides working on the above conjecture of Solymosi, I also have an idea for how to give an alternate proof of sum-product inequalities, though the parameter ε involved will be weaker than the recent result of Solymosi ($\varepsilon = 1/3$). However, my idea can be generalized to prove theorems about rich "cuves" passing through grids $A \times B$, instead of merely rich lines; and, the results that it gives are stronger than those produced by the Szemerédi-Trotter theorem for curves. This should have many applications to other additive combinatorial problems; indeed, my original motivation for proving such a theorem was to attack certain basic questions pertaining to which polynomials $f(x,y) \in \mathbb{R}[x,y]$ are "expanders". [1]

3.2 The structure of sumsets

Another area of additive combinatorics that I have recently been working on concerns the structure of sumsets $A + B$. One recent result on this was one I coauthored with Tomasz Schoen, which shows that so long as one has a "large" spectral gap between any two consecutive Fourier coefficients of a set $A \subseteq \mathbb{F}_p$, then the sumset $A + A$ must cover $1 - \varepsilon$ proportion of $\mathbb{F}_p$. This can be thought of as a generalization of the familiar fact that if the second-largest Fourier coefficient of a set is small relative to the largest one, then $A+A$ must be "large". Furthermore, we have generalized the result to k-fold sums $A+A+\cdots+A$, for $k \geq 3$, and were able to produce nice spectral

[1] We say that $f(x,y)$ is an "expander" if there exists $\varepsilon > 0$ such that the image of $A \times A$ under the map f has size at least $|A|^{1+\varepsilon}$ once $|A|$ is large enough.

gap conditions guaranteeing that the sum equals all of $\mathbb{F}_p$, not merely $1-\varepsilon$ fraction of $\mathbb{F}_p$. Our paper on this is to appear in Acta Arithmetica.

In order to motivate another recent theorem of mine on sumsets, it is worth briefly mentioning the Balog-Szemerédi-Gowers theorem: Certain finite sets A taken from an additive abelian group have the property that when one forms the sumset $A+A$, one does not get "growth" – i.e. one gets that $|A+A|$ is "not significantly larger" than the set A itself. For example, if the additive abelian group is the integers, and if one lets $A := \{1, 2, ..., N\}$, then $|A+A| = 2N-1$, which is "only" about double the size of A. On the other hand, if one were to let $A := \{2, 4, 8, ..., 2^N\}$, then $A+A$ would have size $\sim N^2/2$, so would be significantly larger than A. Now, certain sets A, although they produce large sets $A+A$, nonethless have a lot of "additive structure", specifically there are a small number of $s \in A+A$ for which very many of the pairs $(a, b) \in A \times A$ lead to $s = a+b$. In such a case, one might guess that it should be possible to pass to a large subset $A' \subseteq A$ such that $A' + A'$ is "small"; that is, one might guess it is possible to "filter out" the contribution of those $s \in A+A$ having few representations as a sum $a+b$, $a, b \in A$. This guess is exactly what the Balog-Szemerédi-Gowers theorem tells us, and is stated as follows:

Balog-Szemerédi-Gowers Theorem. Suppose that A is a finite subset of size n of an additive abelian group, and that the number of quadruples a_1, a_2, a_3, a_4 satisfying

$$a_1 + a_2 \ = \ a_3 + a_4$$

is at least $n^{3-\varepsilon}$. Then, there exists a subset $A' \subseteq A$ satisfying

$$|A'| \ \gg_\varepsilon \ n^{1-f(\varepsilon)}, \text{ and } |A' + A'| \ \ll_\varepsilon \ n^{1+g(\varepsilon)}$$

where f and g are fixed polynomials (they do not depend on ε).

The theorem that I and Evan Borenstein proved involves multiple sums, and can be thought of as a variant of a similar sort of result due to Sudakov, Szemerédi and Vu, which appeared in their paper *On a question of Erdős and Moser*, Duke Math. Jour **129** (2005), 129-155. Our theorem is given as follows:

Theorem 6. For every $0 < \varepsilon < 1/2$ and $c > 1$, there exists $\delta > 0$, such that the following holds for all k sufficiently large, and all sufficiently large finite subsets A of an abelian group: Suppose that

$$S \ \subseteq \ A \times A \times \cdots \times A \ = \ A^k,$$

and let

$$\Sigma(S) \ := \ \{a_1 + \cdots + a_k \ : \ (a_1, ..., a_k) \in S\}.$$

If

$$|S| \ \geq \ |A|^{k-\delta}, \text{ and } |\Sigma(S)| \ < \ |A|^c,$$

then there exists

$$A' \ \subseteq \ A, \ |A'| \ \geq \ |A|^{1-\varepsilon},$$

such that

$$|\ell A'| \ = \ |A' + \cdots + A'| \ \leq \ |A'|^{c(1+\varepsilon\ell)}.$$

The way that our theorem differs from that of Sudakov, Szemerédi, and Vu is the fact that we get much better control on the growth of the size of these sums $A' + \cdots + A'$. Indeed, for ℓ small and ε near 0, the sum $A' + \cdots + A'$ is at most about $|A|^c$ in size.

There are many other projects on additive combinatorics that I am also currently considering; unfortunately, it would take me quite a lot of space to describe them all!

4 Analytic Number theory: recent work

Finally, besides these projects in additive combinatorics that form the core of my research, I have been working on many projects while at Georgia Tech with students and other colleagues. One of these concerns the square dependence problem, which has its origins in the theory of integer factorization and cryptology, and is the following basic question: Select integers $n_1, n_2, ... \leq N$ at random until some subset of them has product equal to a square. What is the expected stopping time for this process?

Schroppel and Pomerance obtained upper and lower bounds for this expected stopping time $T = T(N)$, and deduced

$$T \in [y_0, y_0^{1+o(1)}], \text{ where } y_0 = y_0(N) = \exp(\sqrt{2 \log N \log\log N});$$

however, they gave no estimate for this $o(1)$, and so, in particular, could not give an asymptotic estimate for T (nor even anything close to an asymptotic estimate). Recently, myself, A. Granville, R. Pemantle and P. Tetali showed that

$$T \in [(\pi/4)(e^{-\gamma} - o(1))y_0, \ (e^{-\gamma} + o(1))y_0].$$

Moreover, we believe that the upper bound here is the true stopping time!

If the lower bound could be brought up to equal the upper bound, then we will have established the true order of the stopping time in the square dependence problem. Such an accomplishment would bring an end to a long, interesting, and beautiful research programme initiated by Pomerance and others, to rigorously analyze the running time of a major component of several different integer factoring algorithms. Besides the purely mathematical implications, our work (suitably generalized and refined) might also lead to better-performing integer factoring algorithms.