Combinatorial and Additive Number Theory Problem Sessions

Participants of CANT 2012

May 27, 2012

Abstract

These notes are a summary of the problem session discussions during the problem sessions of CANT 2012 (May 22nd to May 25th, 2012 at the CUNY Graduate Center). A list of participants is given in Appendix 5; whenever possible, comments are attributed to the speaker.

For more information, visit the conference homepage at

`http://www.theoryofnumbers.com/`

or email either the typist at sjm1@williams.edu or the organizer at melvyn.nathanson@lehman.cuny.edu.

Warning: These notes were LaTeX-ed in real-time by Steven J. Miller (save for Thursday's notes, which were drafted by Ryan Ronan, and Friday's by Steven Senger); all errors should be attributed solely to the typist and/or the copious libations which were generously donated by individuals.

Finally, many people here are involved with REUs and are looking for problems. If you are interested in working on one of these, either in an REU or on your own, in order to avoid multiple groups working on the same projects at the same time, please email me at sjm1@williams.edu.

1 Problem Session I: Tuesday, May 22nd (Chair Renling Jin)

- *From Renling Jin, jinr@cofc.edu:* Define a subset of the natural numbers B to be an essential component if for all $A \subset \mathbb{N}$, $\sigma(A+B) > \sigma(A)$ if $0 < \sigma(A) < 1$. B is an extraordinary component if

  $$\liminf_{\sigma(A)\to 0} \frac{\sigma(A+B)}{\sigma(A)} \ = \ \infty.$$

  Here

  $$\sigma(A) \ = \ \inf_{x\geq 1} \frac{A(x)}{x}.$$

  Ruzsa conjectured that every essential component is an extraordinary component.

  What are the essential components we know? If

  $$B \ = \ \{k^2 : k \in \mathbb{N}\}$$

  then

  $$\sigma(A+B) \ \geq \ \sigma(A)^{1-1/4}$$

  since B is a basis of order four. We get

  $$\frac{\sigma(A+B)}{\sigma(A)} \ \geq \ \frac{1}{\sqrt[4]{\sigma(A)}}.$$

  Similar for cubes or k-powers.

- *From Steven J. Miller, sjm1@williams.edu:* We say a set A is a More Sums Than Differences Set, or an MSTD set, if $|A+A| > |A-A|$, where

  $$\begin{aligned} A+A &= \{a_i + a_j : a_i, a_j \in A\} \\ A-A &= \{a_i - a_j : a_i, a_j \in A\}. \end{aligned}$$

  As addition is commutative and subtraction is not, it's expected that 'most' sets are difference dominated; however, Martin and O'Bryant proved that a positive percentage of sets are sum-dominated. There are explicit constructions of infinite families of sum-dominant sets. Initially the best result was a density of $n^c 2^{n/2}/2^n$, then $1/n^4$ (or $1/n^2$), and now the record is $1/n$ (where our sets A are chosen uniformly from subsets of $\{0, 1, \dots, n-1\}$). Can you find an 'explicit' family that is a positive percentage.

- *From Urban Larsson, urban.larsson@yahoo.se:* Let $A = \{0, 1, 3, 4, \ldots\}$ for a set that avoids arithmetic progression, thought to be best set to avoid arithmetic progression but not (comes from a greedy construction). Equivalence with a base 3 construction: $A = \{0, 1, 10, 11, 100, 101, \ldots\}$ gives $A((3^n + 1)/2) = 2^n$, where $A(n) = \#\{i \in A \mid i < n\}$. Hence, for all n, $A(n) < Cn^{\log 2/\log 3} \approx n^{2/3}$. Study impartial heap games. Is it possible to find a game such that the P and N-positions correspond to the numbers in this construction? (A position is in N if and only if the first player wins.) In some sense such that:

P	P	N	P	P	N	N
0	1	2	3	4	5	6

  We rather use three heaps of sizes in three-term arithmetic progression. A legal move is to erase the largest pile and then to announce one of the smaller piles as the new largest pile. Notation (x, y), where x is the number of tokens in the smallest heap and y in the second smallest. In the table below, the first entry is the outcome, the second is the position, the third is the Grundy value, and the fourth are the options.

P	$(0,1)$	0	
N	$(0,2)$	1	$(0,1)$
P	$(0,3)$	0	$(1,2)$
N	$(0,4)$	0	$(0,2),\ (2,3)$
N	$(1,2)$	1	$(0,1)$
N	$(1,3)$	0	$(1,2),$
P	$(1,4)$	0	$(0,2),\ (2,3)$
N	$(2,3)$	2	$(0,1),\ (1,2)$
N	$(2,4)$	3	$(1,2),\ (0,2),\ (2,3)$
N	$(3,4)$	0	$(1,2),\ (0,2),\ (2,3)$

  The P positions (Grundy value 0) have both lower heap sizes in the set A. The N positions have Grundy values > 0, defined as the *minimal exclusive* of the Grundy values of the options. What are they? Is it possible to extend the game by *adjoining moves* to obtain limsup$A(n)/n^{\log 2/\log 3} = \infty$? The game generalizes to k-term arithmetic progressions and the Sidon-condition for example.

How do we extend such games? We need a general definition for the family of games. A *ruleset* is a set of finite sets of positive integers. From a *position* consisting of a set S of non-negative integers, choose one of the numbers $s \in S$ and a set M of numbers from the given ruleset. The next position, which is a set of nonnegative numbers, is $\{s - m \mid m \in M\}$, provided max$M \leq s$. We get a recursive definition of the set A which determines the P-positions for a given M. A position S is in P if and only if $S \subset A$. That is S is in N iff $S \cap A \neq \emptyset$. In this sense we can abuse notation and regard A as the set of "P-positions". A game extension of M is $M \cup M'$, for M' a set of finite sets of nonnegative numbers. For our game the set M is $M = \{\{d, 2d\} \mid d > 0\}$. One first example of a game extension is $M = \{\{d, 2d\} \mid d > 0\} \cup \{\{1\}\}$. Question: does the set A become less dense for this game than for our original AP-avoiding game?

2 Problem Session II: Wednesday, May 23rd (Chair Steven J Miller)

- *From Steven J Miller, sjm1@williams.edu:* We investigated in `http://arxiv.org/pdf/1109.4700v2.pdf`) properties of $|A + A|$ and $A + A$ as A varies uniformly over all subsets of $\{0, 1, \dots, n-1\}$. How does the behavior change if we change the probability of choosing various A's (see for example my work with Peter Hegarty: `http://arxiv.org/pdf/0707.3417v5`).

  Another related problem is to 'clean-up' the formula we have for the variance. This involves sums of products of Fibonacci numbers – can the answer be simplified?

  What about the expected values of $2kA$ versus $kA - kA$.

- *From Ryan Ronan, ryan.p.ronan@gmail.com:* Earlier today I discussed joint work on generalized Ramanujan primes, `http://arxiv.org/pdf/1108.0475`. One natural question is whether or not for each prime p there is some constant c_p such that p is a c_p-Ramanujan prime.

  Another question is the distribution of c-Ramanujan primes among the primes, in particular the length of runs of these and non-these. It can take awhile for the limiting behavior of primes to set in; it's dangerous to make conjectures based on small sized data sets. Are the calculations here sufficiently far enough down the number line to have hit the limiting behavior? For a related question, perhaps the Cramer model is not the right model to use to build predictions, and instead we should use a modified sieve to construct 'random primes'. It would be worthwhile to do so and see what happens / what the predictions are.

- *From Steven Senger, senger@math.udel.edu:* Have a subset A of a finite field $\mathbb{F}_q$ satisfying for all ϵ and δ positive (1) $|A|\,|AA| \geq q^{3/2+\epsilon}$, (2) $|AA| \leq q^{1-\delta}$. For all generalized geometric progressions G with $|G| \approx |AA|$ we have $|(AA+1) \setminus G| \geq q^{\delta}$. Can reduce the size constraint (1)? Can we increase the size of $|(AA+1) \setminus G| \geq q^{\delta}$?

- *From Kevin O'Bryant, obryant@gmail.com:* How far out can you go $\{x_1, x_2, x_3, x_4, \cdots\}$ such that the first four are in the first four quadrant, the first nine in the first nine subdivisions (3×3), the first 16 in the first 4×4 and so on.... We know this can't go on forever, violates Schmidt.

  The discrepancy of the sequence $\{x_i\}$ is

  $$\mathrm{Disc}(\{x_i\}_{i=1}^d) \ = \ \sup_R \left| \frac{\#\{x_i \in R\}}{d} - A(R) \right|.$$

  We have $\mathrm{Disc}(\{x_i\}_{i=1}^d) \ \geq \ C\frac{\log d}{d}$. If we spread the points too well, the discrepancy gets very low.

  Let me rephrase – I strongly believe that this logarithmic factor will kill this arrangement.

3 Problem Session III: Thursday, May 24th (Chair Alex Iosevich)

- *From Jerry Hu, HuJ@uhv.edu:*

  This problem is related to Nathanson's talk "The Calkin-Wilf tree and a forest of linear fractional transformations" from Tuesday. Recall the form of the Calkin-Wilf tree, where we have:

$$\begin{array}{ccccccc} & & & \frac{a}{b} & & & \\ & & \swarrow & & \searrow & & \\ & \dfrac{a}{a+b} & & & & \dfrac{a+b}{b} & \\ \swarrow & & \searrow & & \swarrow & & \searrow \\ \dfrac{a}{2a+b} & & \dfrac{2a+b}{a+b} & & \dfrac{a+b}{a+2b} & & \dfrac{a+2b}{b}. \end{array}$$

When a and b are both initialized as 1, each positive rational number appears on the tree exactly once.

The question is: how can we generalize this? More specifically, do there exist other trees of the form

$$\begin{array}{ccc} & z & \\ \swarrow & & \searrow \\ L(z) & & R(z) \end{array}$$

in which every positive rational number appears exactly once? Can we find all, or any, nontrivial functional pairs L, R such that this condition holds?

- *From Nathan Kaplan, nathanckaplan@gmail.com:* Here is a problem about counting lines among points in $\mathbb{F}_3^n$. I will give two different kinds of motivation for why someone might be interested in this.

  The card game SET is played with 81 distinct cards, each of which has four attributes (number, color, shading, and shape), where each attribute has three possibilities. We can identify a card with a 4-tuple (x_1, x_2, x_3, x_4), where each $x_i \in \mathbb{F}_3$. The game is played by collecting sets. A set is a collection of three cards (x, y, z) such that for each of the four attributes each card is the same or all three cards are different. It is equivalent that the vectors in $\mathbb{F}_3^4$ represented by our three cards take the form $(x, y, -(x+y))$, or equivalently, $(x, x+d, x+2d)$. Therefore, we see that what we are looking for is a three term arithmetic progression in $\mathbb{F}_3^n$. In $\mathbb{F}_3^n$ a 3-term AP is equivalent to a line. A set of vectors with no 3-term AP is called a cap set. The cap set problem asks, "What is the maximum size of a cap set in $\mathbb{F}_3^n$?". This problem is very hard and has been well-studied. Exact answers are known only for $n \leq 6$. We note that for $n = 3$ the cap set problem is equivalent to asking for the maximum number of SET cards one can have so that there is no set among them. The answer to this is 20 and an argument is given in the paper *The Card Game Set* by Benjamin Davis and Diane MacLagan.

There is a related problem motivated by SET which does not seem to have appeared in the literature. The game is usually played by dealing out 12 cards. We know that it is possible to have no sets at all, but we could ask for the largest number of sets which could occur among 12 cards. I can show that this is 14, but the argument is sort of ad hoc and not so satisfying. I have not found anything written before about the following question. What is the maximum number of lines that m points in $\mathbb{F}_3^n$ can contain? Note that any two points determine a unique line, so if a set contains many lines, then it determines few lines. Equivalently we could ask for the minimum number of lines determined by m points in $\mathbb{F}_3^n$. This question is very general and includes the cap set problem as a subcase. This is because the number of lines contained in a subset of $\mathbb{F}_3^n$ determines the number of lines contained in its complement, so if we know the maximum number of lines among any collection of m points for all m, then we also know the minimum number of lines among m points.

Here is the actual problem I am asking. In the argument for the maximum number of lines among 12 points in $\mathbb{F}_3^4$ is 14, it is clear that the maximum number of lines among 12 points in $\mathbb{F}_3^n$ is 14 for any $n \geq 3$. That is, if we want lots of lines, the best thing that we can do is to put our points into the smallest possible dimensional subspace that can contain them.

Conjecture 3.1. *Fix $m \geq 0$ and let $d = \lceil \log_3(m) \rceil$. For any $n \geq d$, the maximum number of lines contained among m points in $\mathbb{F}_3^n$ is equal to the maximum number of lines contained among m points in $\mathbb{F}_3^d$.*

I think that this is probably true and that the proof for it is probably easy. One could also ask similar questions for $\mathbb{F}_q^n$ for other q.

Here is some extra motivation that the cap set problem is interesting. Tic-Tac-Toe on a $3 \times 3 \times 3$ board can never end in a draw no matter how many moves are made by each player. This is the first case of a more general phenomenon, the Hales-Jewett Theorem. Given k, there exists a d such that Tic-Tac-Toe on a $k \times \cdots k = [k]^n$ board (where it takes k in a row to win) cannot end in a draw no matter how many times each player moves. A more precise statement is that for large enough n, either a set or its complement must contain a combinatorial line. I won't define exactly what a combinatorial line is, but it is a slightly more restrictive condition than a Tic-Tac-Toe line, which is slightly more restrictive than the type of line described above in the discussion of SET.

A few years ago, the initial Polymath project organized by Tim Gowers was focused on giving a combinatorial proof of the Density Hales-Jewett Theorem. The only previous proof of this theorem involved arguments from ergodic theory. Let $c_{n,k}$ be the largest number of points of $[k]^n$ which does not contain a combinatorial line. Let $c'_{n,k}$ be the largest number of points of $[k]^n$ which does not contain a geometric line (you can think of this as a Tic-Tac-Toe line. These are called Moser numbers. Finally, let $c''_{n,k}$ be the largest number of points of $[k]^n$ without a line of the type described above. Clearly $c''_{n,k} \leq c'_{n,k} \leq c_{n,k}$.

Theorem 3.2 (Density Hales-Jewett)**.** *Fix* $k \geq 1$*. Then*

$$\lim_{n\to\infty} \frac{c_{n,k}}{n^k} = 0.$$

This result is important in understanding the growth of cap sets. The Polymath project also proved the best known lower bound for $c_{n,k}$. It is quite difficult to compute these numbers in general, even for small k. We mentioned above that $c''_{4,3} = 20$ and it is also known that $c''_{5,3} = 45$ and that $c''_{6,3} = 112$. This last statement determines the maximum number of lines among $3^6 - 112$ points in $[3]^6$, for example. The Polymath project also determined more values of $c_{n,3}$ and $c'_{n,3}$ than previously known.

Since so much work has gone into understanding large subsets of $[k]^n$ with no lines, it seems reasonable to study collections of points which contain the largest possible number of lines.

4 Problem Session IV: Friday, May 25th (Chair Kevin O'Bryant)

The following papers are relevant for the problems proposed by Steven Miller.

- `http://arxiv.org/abs/1107.2718`

- `http://arxiv.org/abs/1008.3204`
- `http://arxiv.org/abs/1008.3202` (the gap paper referenced below is in preprint stage, but available upon request).
- `http://www.emis.de/journals/INTEGERS/papers/j57/j57.pdf` (Hannah Alpert).

Proposed problems.

- *From Steven J. Miller, sjm1@williams.edu:* The following problems are related to Zeckendorf decompositions. Many of these are currently being studied by my summer REU students in the Williams 2012 SMALL program. If you are interested in working on these, please email me at sjm1@williams.edu.

  ⋄ We know every number has a unique Zeckendorf decomposition, and appropriately localized the number of summands converges to being a Gaussian. What happens if we have a decomposition where some integers have multiple representations? What if there are some integers that have no representations? Instead of counting the total number of summands, what if you just count how many of each summand one has (so in decimal 4031 wouldn't count as $4 + 0 + 3 + 1$ but $1 + 0 + 1 + 1$).

  ⋄ We have formulas for the limiting distribution of gaps between summands of Fibonacci and some generalized Fibonacci sequences. Try to find formulas for general recurrence relations as a function of the coefficients of the relations. Do this for the signed Fibonacci decomposition (see Hannah Alpert's paper; can we generalize signed distributions to other recurrence relations). What about the distribution of the largest gap (that should grow with n for numbers between H_n and H_{n+1}). If we appropriately normalize it, does it have a nice limiting distribution?

- *From Mizan Khan, khanm@easternct.edu:* Let

  $$\mathcal{H}_n := \{(x, y) \in \mathbb{Z} \times \mathbb{Z} : xy \equiv 1 \pmod{n}, 1 \leq x, y \leq n - 1)\}.$$

  Consider the convex closure of $\mathcal{H}_n$– what can we say about the number of vertices in this convex closure? Let $v(n)$ be the number of vertices. Easily, $v(n) \geq 2(\tau(n-1) - 1)$, where τ is the number of positive divisors.

It is easy to see that $\limsup v(n) = \infty$. Can we show that $\lim_{n\to\infty} v(n) = \infty$?

Also, consider $D(n) = v(n) - 2(\tau(n-1) - 1)$. We know that $D(n) > 0$ for a set of density 1 in the naturals and furthermore $D(n) = 0$ on a set which is $\gg \frac{x}{\log x}$. Can we improve the second estimate?

- *From Steven Senger, senger@math.udel.edu:* We will call a family of sets, $P_n \subset [0,1]^2$, s-adaptable if they satisfy the following bound:

  $$\frac{1}{\binom{n}{2}} \sum_{x\neq y; x,y\in P_n} |x-y|^{-s} \lesssim 1.$$

  The Szemerédi-Trotter incidence theorem says that for a set of n points and m “reasonable" curves in the plane, the number of incidences of points and curves is bounded above by

  $$I \lesssim (nm)^{\frac{2}{3}} + n + m.$$

  Can we get better incidence bounds for s-adaptable sets? Specifically, can we get tighter bounds in the case of n points and n circles centered at those points?

- *From Nathan Pflueger, pflueger@math.harvard.edu:* Suppose S is a numerical semigroup, $S \subset \mathbb{N}_+$, closed under addition, i.e., $S + S \subset S$. Let $G := \mathbb{N}_+ \backslash S$. Define the weight of S to be $w(s) = |\{(x,y) \in S \times G : 0 < x < y\}|$. Define the irreducible elements of S to be the minimal generators. Define the effective weight of S to be $w_{eff}(s) = |\{(x,y) \in S_{irred} \times G : 0 < x < y\}|$. Let the genus of S be $g = |G|$.

  For example, $S = \langle 3, 5\rangle$. Then $w(s) = 4$, and $w_{eff}(s) = 3$.

  Can we characterize the genus g subgroups of largest effective weight? We believe the largest is $\approx \frac{g^2}{4}$, and in the form $\langle a, a+1, \ldots, b-1, b\rangle$, where $b < 2a$.

  This comes from algebraic geometry. Pick a point p on an algebraic curve or surface. $S = \{ord_p(f) : f$ is a rational function$\}$, where $ord_p(f)$ is the order of the single pole at p of f.

5 Speaker List

- John Bryk, John Jay College (CUNY)
- Mei-Chu Chang, University of California-Riverside
- Emel Demirel, Bergen County College
- Frederic Gilbert, Ecole Polytechnique, Paris
- Christopher Hanusa, Queens College (CUNY)
- Charles Helou, Penn State Brandywine
- Jerry Hu, University of Houston - Victoria
- Alex Iosevich, University of Rochester
- Geoff Iyer, University of Michigan
- Renling Jin, College of Charleston
- Nathan Kaplan, Harvard University
- Mizan R. Khan, Eastern Connecticut State University
- Sandra Kingan, Brooklyn College (CUNY)
- Alex Kontorovich, Yale University
- Urban Larsson, Chalmers University of Technology and University of Gothenburg
- Oleg Lazarev, Princeton University
- Xian-Jin Li, Brigham Young University
- Neil Lyall, University of Georgia
- Steven J. Miller, Williams College
- Rishi Nath, York College (CUNY)
- Mel Nathanson, Lehman College (CUNY)

- Kevin O’Bryant, College of Staten Island (CUNY)
- Kerry Ojakian, St. Joseph’s College, New York
- Ryan Ronan, Cooper Union
- Steven Senger, University of Delaware
- Jonathan Sondow, New York
- Liyang Zhang, Williams College
- Wei Zhang, Columbia University