Ninth Lecture: 21/4

Remark 9.1. Recall that an *equivalence relation* on a set X is a subset $\mathcal{R}$ of $X \times X$ satisfying the following three properties:

Reflexivity: $(x,\, x) \in \mathcal{R}$ for all $x \in X$. In words, every element of x is related to itself.

Symmetry: $(x,\, y) \in \mathcal{R} \Leftrightarrow (y,\, x) \in \mathcal{R}$. In words, x is related to y if and only if y is related to x.

Transitivity: If $(x,\, y) \in \mathcal{R}$ and $(y,\, z) \in \mathcal{R}$ then $(x,\, z) \in \mathcal{R}$. In words, if x is related to y and y to z, then x is related to z.

An equivalence relation on a set X gives rise to a partition of the set into so-called *equivalence classes*, such that two elements are related if and only if they lie in the same class.

We can now give a further interpretation of the Stirling numbers $S(n,\, k)$, namely: $S(n,\, k)$ is the number of different equivalence relations on an n-element set which give rise to exactly k equivalence classes.

Example 9.2. The number of ways to distribute n identical balls into k identical bins such that no bin is left empty is denoted $p(n,\, k)$, and is usually referred to as the number of *partitions of n into exactly k parts*. There are a number of different recurrences for partition numbers, perhaps the simplest is

$$p(n,\, k) = p(n-1,\, k-1) + p(n-k,\, k), \tag{8.1}$$

which is an exercise on Homework 1. See also Exercise 12.4.2 in Biggs, for example. We can also note some special cases:

(i) $p(n,\, 1) = 1$ since the only partition into one part is n itself.

(ii) $p(n,\, n) = 1$ since the only paritition of n into n parts is $1 + 1 + \cdots + 1$ (n times).

(iii) $p(n,\, n-1) = 1$ since the only paritition of n into $n-1$ parts is to have one part equal to 2 and all other parts equal to 1, i.e.: $2 + 1 + \cdots + 1$.

(iv) $p(n,\, 2) = \lfloor n/2 \rfloor$ since every partition of n into two parts is of the form $k+(n-k)$, where $\lceil n/2 \rceil \leq k \leq n-1$.

To summarise (i)-(iv):

$$p(n,\, 1) = p(n,\, n) = p(n,\, n-1) = 1, \quad p(n,\, 2) = \lfloor \frac{n}{2} \rfloor. \tag{8.2}$$

The function

$$p(n) = \sum_{k=1}^{n} p(n,\, k)$$

has been extensively studied in Number Theory. It is the total number of partitions of n, or just the "partition function". A famous problem, which was essentially solved in 1918 by Hardy and Ramanujan (and independently by Uspensky in 1920), was to determine a good asymptotic estimate for $p(n)$. Their result is that

$$p(n) \sim \frac{e^{\pi\sqrt{2n/3}}}{4\sqrt{3}\, n}, \tag{8.3}$$

where $f(n) \sim g(n)$ means that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$. Note that the growth rate of $p(n)$ is superpolynomial but subexponential, an indication of a highly non-trivial behaviour since, for example, it follows from previous lectures that sequences satisfying linear recurrences always exhibit either polynomial or exponential growth. To do justice to the rich theory on partitions is beyond the scope of this course. See wiki. For the record, we just list all partitions of 8:

$$\begin{gathered}
8, \quad 7+1, \quad 6+2, \quad 6+1+1, \quad 5+3, \quad 5+2+1, \\
5+1+1+1, \quad 4+4, \quad 4+3+1, \quad 4+2+2, \quad 4+2+1+1, \\
4+1+1+1+1, \quad 3+3+2, \quad 3+3+1+1, \quad 3+2+2+1, \quad 3+2+1+1+1, \\
3+1+1+1+1+1, \quad 2+2+2+2, \quad 2+2+2+1+1, \quad 2+2+1+1+1+1, \\
2+1+1+1+1+1+1, \quad 1+1+1+1+1+1+1+1.
\end{gathered}$$

Hence $p(8) = 22$ and, breaking it down,

$$\begin{gathered}
p(8,\,1) = 1, \quad p(8,\,2) = 4, \quad p(8,\,3) = 5, \quad p(8,\,4) = 5, \\
p(8,\,5) = 3, \quad p(8,\,6) = 2, \quad p(8,\,7) = 1, \quad p(8,\,8) = 1.
\end{gathered}$$

SPECIAL TOPIC 1: GENERATING FUNCTIONS IN ADDITIVE NUMBER THEORY.

Additive number theory is, as the name suggests, an area of research in Number Theory. Some of the most famous problems of number theory can be placed in this setting but, starting from work of people like Erdős and Turán in the early 20th century, it has become a field of research in its own right, in which combinatorial questions and methods play a major role. The field has exploded in popularity in the last 20 years or so. Our goal here will be to introduce one of the central notions in the area, that of *basis*, and to prove a classical result of Erdős and Turán, which involves a very clever use of a certain generating function. The interested reader can find more on additive number theory in the lecture notes for MMA300 on my homepage.

For simplicity, all sets in what follows are assumed to be subsets of $\mathbb{N}_0$.

Definition 9.3. Let $A \subseteq \mathbb{N}_0$. The *sumset* $A + A$ is defined as

$$A + A = \{a_1 + a_2 : a_1,\, a_2 \in A, \text{ where } a_1 = a_2 \text{ is allowed}\}.$$

Example 9.4. Let $A = \{0,\, 1,\, 3,\, 4,\, 7\}$. Then

$$\begin{gathered}
0 = 0+0, \;\; 1 = 0+1, \;\; 2 = 1+1, \;\; 3 = 0+3, \;\; 4 = 1+3, \;\; 5 = 1+4, \;\; 6 = 3+3, \\
7 = 0+7 = 3+4, \;\; 8 = 1+7 = 4+4, \;\; 10 = 3+7, \;\; 11 = 4+7, \;\; 14 = 7+7,
\end{gathered}$$

so $A + A = \{0,\, 1,\, 2,\, 3,\, 4,\, 5,\, 6,\, 7,\, 8,\, 10,\, 11,\, 14\}$.

Notation 9.5. It is common in this subject to write $2A$ instaed of $A + A$. Do *not*

confuse $2A$ with the set $\{2a : a \in A\}$.

We can extend Definition 9.3 to so-called "higher order sumsets":

Definition 9.6. Let $A \subseteq \mathbb{N}_0$ and $k \in \mathbb{N}$. The *k-fold sumset* kA is defined as

$$kA = \{a_1 + a_2 + \cdots + a_k : a_i \in A, \text{ where repititions are allowed}\}.$$

Remark 9.7. If A is a finite set with $|A| = n$, then $|2A| \leq \frac{n(n+1)}{2}$, since this is the total number of possible choices of a pair $\{a_1, a_2\}$ of elements of A: there are $\binom{n}{2}$ choices in which $a_1 \neq a_2$ and n choices in which $a_1 = a_2$.

An upper bound for the size of kA, for general k, is an exercise on Homework 2.

We now turn our attention to infinite sets of non-negative integers.

Definition 9.8. Let $A \subseteq \mathbb{N}_0$ and $k \in \mathbb{N}$. The set A is said to be a *basis of order k* if $kA = \mathbb{N}_0$. It is said to be an *asymptotic* basis of order k if $\mathbb{N}_0 \backslash kA$ is a finite set, in other words, if kA contains all sufficiently large positive integers.

Example 9.9. Let $A_2 = \{n^2 : n \in \mathbb{N}_0\}$. Lagrange's Theorem (1770) states that A_2 is a basis of order 4. It is not a basis of order 3, since Gauss proved that a number is a sum of three squares if and only if it is not of the form $4^k(8l + 7)$, for some $k,\ l \in \mathbb{N}_0$.

More generally, for any $k \geq 2$, let $A_k = \{n^k : n \in \mathbb{N}_0\}$. It was first proven by Hilbert (1909) that every A_k is a basis of some order. To determine the minimum order of A_k as an asymptotic basis[1] is called *Waring's Problem*. It has been solved only for $k = 2$, as stated above, and for $k = 4$: Davenport (1939) proved that A_4 is an asymptotic basis of order 16, but of no smaller order. For A_3, the answer is known to lie between 4 and 7 and, if you can solve this problem, you'll surely get a Fields Medal ! For more information on Waring's Problem, see wiki.

Example 9.10. Let $\mathcal{P}$ denote the set of primes and $A := \mathcal{P} \cup \{0, 1\}$. Vinogradov (1937) proved that every sufficiently large odd number is a sum of three primes, from which it follows that A is an asymptotic basis of order 4. In fact, Vinogradov's method works for all odd numbers geater than about 10^{400}. Hence, it was a major breakthrough in 2013 when Helfgott proved that *every* odd number greater than or equal to 7 is a sum of three primes. In consequence, A is a basis of order 4. Helfgott's work is a tour de force in technique (the paper is well over 100 pages long). Vinogradov used the so-called *Hardy-Littlewood circle method*, which has become a standard tool in analytic number theory. Helfgott managed to find improvements to the classical approach which got the method to work for numbers greater than 10^{30} or so. It being 2013, this was just about small enough to be able to check all odd numbers up to that point on a computer.

The *Goldbach conjecture* states that every even number greater than or equal to 4 is a sum of two primes. If true, this would imply that our set A is a basis of order 3. This is another problem which, if you solve it, will get you a Fields Medal ! For more info,

[1] a more interesting problem, for technical reasons, than the order as a basis.

see wiki.

For our purposes, the point of Examples 9.9 and 9.10 is that they show that the concepts of basis and asymptotic basis cover some very classical problems in Number Theory. A more "combinatorial perspective" was introduced by people like Erdős and Turán, from the 1930s onwards, who began by asking how one might find (asymptotic) bases of a given order which are as "efficient/sparse" as possible. It follows from Remark 9.7 that, if A is an asymptotic basis of order k, then A contains at least on the order of $n^{1/k}$ of the integers up to n. However, the set of primes in Example 9.10 is much, much denser. The Prime Number Theorem (1896) asserts that $\pi(n) \sim \frac{n}{\ln n}$, where $\pi(n)$ is the number of primes up to n. Hence, the set A in Example 9.10 is certainly "more than dense enough" to be an asymptotic basis of order 3. The difficulty with the Goldbach conjecture is that it is not, of course, about a randomly chosen collection of numbers, but a set of numbers with very specific properties, namely all its elements are primes. So, for example, only one of its members is an even integer.

One might expect, however, that it is possible to choose elements of a set A in a "more random manner" so that it is an asymptotic basis of order k but much sparser than the primes, in the best case containing only on the order of $n^{1/k}$ of the numbers up to n, for all large n. Erdős and Turán asked if one could achieve the "Holy Grail" and construct a set A such that every sufficiently large integer could be expressed as a sum of two elements from A in *exactly one way* ? They proved that this is, in fact, impossible, and we will reproduce their proof below. To formulate precise statements, it is convenient to introduce one further piece of terminology:

Notation 9.11. Let $A \subseteq \mathbb{N}_0$ and $n \in \mathbb{N}_0$. We denote by $r_2(A, n)$ the number of ways that n can be expressed as the sum of two elements of A, i.e.:

$$r_2(A, n) = \#\{\{a_1, a_2\} : a_1, a_2 \in A, \ a_1 + a_2 = n\}. \tag{9.4}$$

Considered as a function $r_2(A, \cdot) : \mathbb{N}_0 \to \mathbb{N}_0$, this is called the *(2-fold) representation function* of the set A.

The main result in Special Topic 1 is the following:

Theorem 9.12. (Erdős-Turán 1941) *There is no integer $t \geq 1$ and subset $A \subseteq \mathbb{N}_0$ such that $r_2(A, n) = t$ for all sufficiently large n. In other words, if A is an asymptotic basis of order 2, then its 2-fold representation function cannot be ultimately constant.*

The proof will be given next day. We finish by stating a famous open problem, which would be a major strengthening of the previous theorem. It's yet another opportunity for you to get a Fields Medal !!

Conjecture 9.13. (Erdős-Turán 1941) *Suppose $A \subseteq \mathbb{N}_0$ is an asymptotic basis of order two. Then the function $r_2(A, \cdot)$ is unbounded, i.e.: $\limsup_{n\to\infty} r_2(A, n) = \infty$.*