NOTES 0: OVERVIEW OF ANALYTIC PRIME NUMBER THEORY

TERENCE TAO

In the winter quarter (starting January 5) I will be teaching a graduate topics course entitled "An introduction to analytic prime number theory". As the name suggests, this is a course covering many of the analytic number theory techniques used to study the distribution of the prime numbers $\mathcal{P} = \{2, 3, 5, 7, 11, \dots\}$. I will list the topics I intend to cover in this course below the fold. As with my previous courses, I will place lecture notes online on my blog in advance of the physical lectures.

The type of results about primes that one aspires to prove here is well captured by *Landau's classical list of problems*:

1. *Even Goldbach conjecture*: every even number N greater than two is expressible as the sum of two primes.
2. *Twin prime conjecture*: there are infinitely many pairs $n, n+2$ which are simultaneously prime.
3. *Legendre's conjecture*: for every natural number N, there is a prime between N^2 and $(N+1)^2$.
4. There are infinitely many primes of the form $n^2 + 1$.

All four of Landau's problems remain open, but we have convincing heuristic evidence that they are all true, and in each of the four cases we have some highly non-trivial partial results, some of which will be covered in this course. We also now have some understanding of the barriers we are facing to fully resolving each of these problems, such as the parity problem; this will also be discussed in the course.

One of the main reasons that the prime numbers $\mathcal{P}$ are so difficult to deal with rigorously is that they have very little usable algebraic or geometric structure that we know how to exploit; for instance, we do not have any useful *prime generating functions*. One of course can create *non-useful* functions of this form, such as the ordered parameterisation $n \mapsto p_n$ that maps each natural number n to the n^{th} prime p_n, or one could invoke *Matiyasevich's theorem* to produce a polynomial of many variables whose only positive values are prime, but these sorts of functions have no usable structure to exploit (for instance, they give no insight into any of the Landau problems listed above; see also Remark 2 below). The various *primality tests* in the literature, while useful for practical applications (e.g. cryptography) involving primes, have also proven to be of little utility for these sorts of problems; again, see Remark 2. In fact, in order to make plausible heuristic predictions about the primes, it is best to take almost the opposite point of view to the structured viewpoint, using as a starting point the belief that the primes exhibit strong *pseudorandomness*

properties that are largely incompatible with the presence of rigid algebraic or geometric structure. We will discuss such heuristics later in this course.

It may be in the future that some usable structure to the primes (or related objects) will eventually be located (this is for instance one of the motivations in developing a rigorous theory of the *"field with one element"*, although this theory is far from being fully realised at present). For now, though, analytic and combinatorial methods have proven to be the most effective way forward, as they can often be used even in the near-complete absence of structure.

In this course, we will not discuss combinatorial approaches (such as the deployment of tools from *additive combinatorics*) in depth, but instead focus on the analytic methods. The basic principles of this approach can be summarised as follows:

1. Rather than try to isolate individual primes p in $\mathcal{P}$, one works with the set of primes $\mathcal{P}$ in *aggregate*, focusing in particular on *asymptotic statistics* of this set. For instance, rather than try to find a single pair $n, n+2$ of twin primes, one can focus instead on the *count* $|\{n \leq x : n, n+2 \in \mathcal{P}\}|$ of twin primes up to some threshold x. Similarly, one can focus on counts such as $|\{n \leq N : n, N-n \in \mathcal{P}\}|$, $|\{p \in \mathcal{P} : N^2 < p < (N+1)^2\}|$, or $|\{n \leq x : n^2+1 \in \mathcal{P}\}|$, which are the natural counts associated to the other three Landau problems. In all four of Landau's problems, the basic task is now to obtain a non-trivial lower bounds on these counts.
2. If one wishes to proceed analytically rather than combinatorially, one should convert all these counts into sums, using the fundamental identity
$$|A| = \sum_n 1_A(n),$$
(or variants thereof) for the cardinality $|A|$ of subsets A of the natural numbers $\mathbf{N}$, where 1_A is the indicator function of A (and n ranges over $\mathbf{N}$). Thus we are now interested in estimating (and particularly in lower bounding) sums such as
$$\sum_{n \leq N} 1_{\mathcal{P}}(n) 1_{\mathcal{P}}(N-n),$$
$$\sum_{n \leq x} 1_{\mathcal{P}}(n) 1_{\mathcal{P}}(n+2),$$
$$\sum_{N^2 < n < (N+1)^2} 1_{\mathcal{P}}(n),$$
or
$$\sum_{n \leq x} 1_{\mathcal{P}}(n^2+1).$$
3. Once one expresses number-theoretic problems in this fashion, we are naturally led to the more general question of how to accurately estimate (or, less ambitiously, to lower bound or upper bound) sums such as
$$\sum_n f(n)$$

or more generally bilinear or multilinear sums such as

$$\sum_n \sum_m f(n,m)$$

or

$$\sum_{n_1,\dots,n_k} f(n_1,\dots,n_k)$$

for various functions f of arithmetic interest. (Importantly, one should also generalise to include integrals as well as sums, particularly contour integrals or integrals over the unit circle or real line, but we postpone discussion of these generalisations to later in the course.) Indeed, a huge portion of modern analytic number theory is devoted to precisely this sort of question. In many cases, we can predict an *expected main term* for such sums, and then the task is to control the *error term* between the true sum and its expected main term. It is often convenient to normalise the expected main term to be zero or negligible (e.g. by subtracting a suitable constant from f), so that one is now trying to show that a sum of signed real numbers (or perhaps complex numbers) is small. In other words, the question becomes one of rigorously establishing a significant amount of *cancellation* in one's sums (also referred to as a gain or savings over a benchmark "trivial bound"). Or to phrase it negatively, the task is to rigorously prevent a *conspiracy* of non-cancellation, caused for instance by two factors in the summand $f(n)$ exhibiting an unexpectedly large correlation with each other.

4. It is often difficult to discern cancellation (or to prevent conspiracy) directly for a given sum (such as $\sum_n f(n)$) of interest. However, analytic number theory has developed a large number of techniques to relate one sum to another, and then the strategy is to keep transforming the sum into more and more analytically tractable expressions, until one arrives at a sum for which cancellation can be directly exhibited. (Note though that there is often a short-term tradeoff between analytic tractability and algebraic simplicity; in a typical analytic number theory argument, the sums will get expanded and decomposed into many quite messy-looking sub-sums, until at some point one applies some crude estimation to replace these messy subsums by tractable ones again.) There are many transformations available, ranging such basic tools as the triangle inequality, pointwise domination, or the Cauchy-Schwarz inequality to key identities such as multiplicative number theory identities (such as the *Vaughan* identity and the Heath-Brown identity), Fourier-analytic identities (e.g. *Fourier inversion*, *Poisson summation*, or more advanced trace formulae), or complex analytic identities (e.g. the *residue theorem*, *Perron's formula*, or *Jensen's formula*). The sheer range of transformations available can be intimidating at first; there is no shortage of transformations and identities in this subject, and if one applies them randomly then one will typically just transform a difficult sum into an even more difficult and intractable expression. However, one can make progress if one is guided by the strategy of isolating and enhancing

a desired cancellation (or conspiracy) to the point where it can be easily established (or dispelled), or alternatively to reach the point where no deep cancellation is needed for the application at hand (or equivalently, that no deep conspiracy can disrupt the application).

5. One particularly powerful technique (albeit one which, ironically, can be highly "ineffective" in a certain technical sense to be discussed later) is to use one potential conspiracy to defeat another, a technique I refer to as the "dueling conspiracies" method. This technique may be unable to prevent a single strong conspiracy, but it can sometimes be used to prevent *two* or *more* such conspiracies from occurring, which is particularly useful if conspiracies come in pairs (e.g. through complex conjugation symmetry, or a functional equation). A related (but more "effective") strategy is to try to "disperse" a single conspiracy into several distinct conspiracies, which can then be used to defeat each other.

As stated before, the above strategy has not been able to establish any of the four Landau problems as stated. However, they can come close to such problems (and we now have some understanding as to why these problems remain out of reach of current methods). For instance, by using these techniques (and a lot of additional effort) one can obtain the following sample partial results in the Landau problems:

1. *Chen's theorem*: every sufficiently large even number N is expressible as the sum of a prime and an *almost prime* (the product of at most two primes). The proof proceeds by finding a nontrivial lower bound on
$$\sum_{n \leq N} 1_{\mathcal{P}}(n) 1_{\mathcal{E}_2}(N-n),$$
where $\mathcal{E}_2$ is the set of almost primes.
2. *Zhang's theorem*: There exist infinitely many pairs p_n, p_{n+1} of consecutive primes with $p_{n+1} - p_n \leq 7 \times 10^7$. The proof proceeds by giving a non-negative lower bound on the quantity
$$\sum_{x \leq n \leq 2x} (\sum_{i=1}^{k} 1_{\mathcal{P}}(n+h_i) - 1)$$
for large x and certain distinct integers $h_1, \ldots, h_k$ between 0 and 7×10^7. (The bound 7×10^7 has since been lowered to 246.)
3. *The Baker-Harman-Pintz theorem*: for sufficiently large x, there is a prime between x and $x + x^{0.525}$. Proven by finding a nontrivial lower bound on
$$\sum_{x \leq n \leq x + x^{0.525}} 1_{\mathcal{P}}(n).$$
4. *The Friedlander-Iwaniec theorem*: There are infinitely many primes of the form $n^2 + m^4$. Proven by finding a nontrivial lower bound on
$$\sum_{n,m: n^2+m^4 \leq x} 1_{\mathcal{P}}(n^2+m^4).$$

We will discuss (simpler versions of) several of these results in this course.

Of course, for the above general strategy to have any chance of succeeding, one must at some point use *some* information about the set $\mathcal{P}$ of primes. As stated previously, usefully structured parametric descriptions of $\mathcal{P}$ do not appear to be available. However, we do have two other fundamental and useful ways to describe $\mathcal{P}$:

1. (Sieve theory description) The primes $\mathcal{P}$ consist of those numbers greater than one, that are not divisible by any smaller prime.
2. (Multiplicative number theory description) The primes $\mathcal{P}$ are the multiplicative generators of the natural numbers $\mathbf{N}$: every natural number is uniquely factorisable (up to permutation) into the product of primes (the *fundamental theorem of arithmetic*).

The sieve-theoretic description and its variants lead one to a good understanding of the *almost primes*, which turn out to be excellent tools for controlling the primes themselves, although there are known limitations as to how much information on the primes one can extract from sieve-theoretic methods alone, which we will discuss later in this course. The multiplicative number theory methods lead one (after some complex or Fourier analysis) to the *Riemann zeta function* (and other L-functions, particularly the *Dirichlet L-functions*), with the distribution of zeroes (and poles) of these functions playing a particularly decisive role in the multiplicative methods.

Many of our strongest results in analytic prime number theory are ultimately obtained by incorporating some combination of the above two fundamental descriptions of $\mathcal{P}$ (or variants thereof) into the general strategy described above. In contrast, more advanced descriptions of $\mathcal{P}$, such as those coming from the various *primality tests* available, have (until now, at least) been surprisingly ineffective in practice for attacking problems such as Landau's problems. One reason for this is that such tests generally involve operations such as exponentiation $a \mapsto a^n$ or the factorial function $n \mapsto n!$, which grow too quickly to be amenable to the analytic techniques discussed above.

To give a simple illustration of these two basic approaches to the primes, let us first give two variants of the usual proof of *Euclid's theorem*:

Theorem 1 (Euclid's Theorem)**.** *There are infinitely many primes.*

Proof. (Multiplicative number theory proof) Suppose for contradiction that there were only finitely many primes $p_1, \ldots, p_n$. Then, by the fundamental theorem of arithmetic, every natural number is expressible as the product of the primes $p_1, \ldots, p_n$. But the natural number $p_1 \ldots p_n + 1$ is larger than one, but not divisible by any of the primes $p_1, \ldots, p_n$, a contradiction. □

Proof. (Sieve-theoretic proof) Suppose for contradiction that there were only finitely many primes $p_1, \ldots, p_n$. Then, by the Chinese remainder theorem, the set of natural numbers A that is not divisible by any of the $p_1, \ldots, p_n$ has density $\prod_{i=1}^n (1 - \frac{1}{p_i})$, that is to say

$$\lim_{N \to \infty} \frac{1}{N} |A \cap \{1, \ldots, N\}| = \prod_{i=1}^n (1 - \frac{1}{p_i}).$$

In particular, A has positive density and thus contains an element larger than 1. But the least such element is one further prime in addition to $p_1, \ldots, p_n$, a contradiction. □

Remark 2. *One can also phrase the proof of Euclid's theorem in a fashion that largely avoids the use of contradiction; see this previous blog post for more discussion.*

Both proofs in fact extend to give a stronger result:

Theorem 3 (Euler's Theorem)**.** *The sum $\sum_{p\in\mathcal{P}} \frac{1}{p}$ is divergent.*

Proof. (Multiplicative number theory proof) By the fundamental theorem of arithmetic, every natural number is expressible uniquely as the product $p_1^{a_1} \ldots p_n^{a_n}$ of primes in increasing order. In particular, we have the identity

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p\in\mathcal{P}} (1 + \frac{1}{p} + \frac{1}{p^2} + \ldots)$$

(both sides make sense in $[0, +\infty]$ as everything is unsigned). Since the left-hand side is divergent, the right-hand side is as well. But

$$(1 + \frac{1}{p} + \frac{1}{p^2} + \ldots) = \exp(\frac{1}{p} + O(\frac{1}{p^2}))$$

and $\sum_{p\in\mathcal{P}} \frac{1}{p^2} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$, so $\sum_{p\in\mathcal{P}} \frac{1}{p}$ must be divergent. □

Proof. (Sieve-theoretic proof) Suppose for contradiction that the sum $\sum_{p\in\mathcal{P}} \frac{1}{p}$ is convergent. For each natural number k, let A_k be the set of natural numbers not divisible by the first k primes $p_1, \ldots, p_k$, and let A be the set of numbers not divisible by any prime in $\mathcal{P}$. As in the previous proof, each A_k has density $\prod_{i=1}^{k}(1 - \frac{1}{p_i})$. Also, since $\{1, \ldots, N\}$ contains at most $\frac{N}{p}$ multiples of p, we have from the union bound that

$$|A \cap \{1, \ldots, N\}| = |A_k \cap \{1, \ldots, N\}| - O(N \sum_{i>k} \frac{1}{p_i}).$$

Since $\sum_{i=1}^{\infty} \frac{1}{p_i}$ is assumed to be convergent, we conclude that the density of A_k converges to the density of A; thus A has density $\prod_{i=1}^{\infty}(1 - \frac{1}{p_i})$, which is non-zero by the hypothesis that $\sum_{i=1}^{\infty} \frac{1}{p_i}$ converges. On the other hand, since the primes are the only numbers greater than one not divisible by smaller primes, A is just $\{1\}$, which has density zero, giving the desired contradiction. □

Remark 4. *We have seen how easy it is to prove Euler's theorem by analytic methods. In contrast, there does not seem to be any known proof of this theorem that proceeds by using any sort of prime-generating formula or a primality test, which is further evidence that such tools are not the most effective way to make progress on problems such as Landau's problems. (But the weaker theorem of Euclid, Theorem 1, can sometimes be proven by such devices.)*

The two proofs of Theorem 2 given above are essentially the same proof, as is hinted at by the geometric series identity

$$1 + \frac{1}{p} + \frac{1}{p^2} + \cdots = \left(1 - \frac{1}{p}\right)^{-1}.$$

One can also see the *Riemann zeta function* begin to make an appearance in both proofs. Once one goes beyond Euler's theorem, though, the sieve-theoretic and multiplicative methods begin to diverge significantly. On one hand, sieve theory can still handle to some extent sets such as twin primes, despite the lack of multiplicative structure (one simply has to sieve out two residue classes per prime, rather than one); on the other, multiplicative number theory can attain results such as the *prime number theorem* for which purely sieve theoretic techniques have not been able to establish. The deepest results in analytic number theory will typically require a combination of both sieve-theoretic methods and multiplicative methods in conjunction with the many transforms discussed earlier (and, in many cases, additional inputs from other fields of mathematics such as arithmetic geometry, ergodic theory, or additive combinatorics).

1. Topics Covered

Analytic prime number theory is a vast subject (the 615-page *text of Iwaniec and Kowalski*, for instance, gives a good indication as to its scope). I will therefore have to be somewhat selective in deciding what subset of this field to cover. I have chosen the following "core" topics to focus on:

- Elementary multiplicative number theory.
- Heuristic random models for the primes.
- The basic theory of the Riemann zeta function and Dirichlet L-functions, and their relationship with the primes.
- Zero-free regions for the zeta function and the Dirichet L-function, including Siegel's theorem.
- The prime number theorem, the Siegel-Walfisz theorem, and the Bombieri-Vinogradov theorem.
- Sieve theory, small and large gaps between the primes, and the parity problem.
- Exponential sum estimates over the integers, and the Vinogradov-Korobov zero-free region.
- Zero density estimates, Hohiesel's theorem, and Linnik's theorem.
- Exponential sum estimates over finite fields, and improved distribution estimates for the primes.
- (If time permits) Exponential sum estimates over the primes, the circle method, and Vinogradov's three-primes theorem.

In order to cover all this material, I will focus on more qualitative results, as opposed to the strongest quantitative results, in particular I will not attempt to optimise many of the numerical constants and exponents appearing in various estimates. This also allows me to downplay the role of some key components of the field which

are not essential for establishing the core results of this course at such a qualitative level:

- I will minimise the use of algebraic number theory tools (such as the class number formula).
- I will avoid deploying the functional equation (or related identities, such as Poisson summation) if they are unnecessary at a qualitative level (though I will note when the functional equation can be used to improve the quantitative results). As it turns out, all of the core results mentioned above can in fact be derived without ever invoking the functional equation, although one usually gets poorer numerical exponents as a consequence.
- Somewhat related to this, I will reduce the reliance on complex analytic methods as compared to more traditional presentations of the material, relying in some places instead on Fourier-analytic substitutes, or on results about harmonic functions. (But I will not go as far as deploying the primarily real-variable "pretentious" approach to analytic number theory currently in development by Granville and Soundararajan, although my approach here does align in spirit with that approach.)
- The discussion on sieve methods will be somewhat abridged, focusing primarily on the Selberg sieve, which is a good general-purpose sieve for qualitative applications at least.
- I will almost certainly avoid any discussion of automorphic forms methods.
- Similarly, I will not cover methods that rely on additive combinatorics or ergodic theory.

Of course, many of these additional topics are well covered in existing textbooks, such as the above-mentioned text of Iwaniec and Kowalski (or, for the finer points of sieve theory, the *text of Friedlander and Iwaniec*). Other good texts that can be used for supplementary reading are Davenport's *"Multiplicative number theory"* and Montgomery-Vaughan's *"Multiplicative number theory I."*. As for prerequisites: some exposure to complex analysis, Fourier analysis, and real analysis will be particularly helpful, although we will review some of this material as needed (particularly with regard to complex analysis and the theory of harmonic functions). Experience with other quantitative areas of mathematics in which lower bounds, upper bounds, and other forms of estimation are emphasised (e.g. asymptotic combinatorics or theoretical computer science) will also be useful. Knowledge of algebraic number theory or arithmetic geometry will add a valuable additional perspective to the course, but will not be necessary to follow most of the material.

2. Notation

In this course, all sums will be understood to be over the natural numbers unless otherwise specified, with the exception of sums over the variable p (or variants such as p_1, p_2, etc.), which will be understood to be over primes.

We will use asymptotic notation in two contexts, one in which there is no asymptotic parameter present, and one in which there is an asymptotic parameter (such as

x) that is going to infinity. In the non-asymptotic setting (which is the default context if no asymptotic parameter is explicitly specified), we use $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote an estimate of the form $|X| \leq CY$, where C is an absolute constant. In some cases we would like the implied constant C to depend on some additional parameters such as k, in which case we will denote this by subscripts, for instance $X = O_k(Y)$ denotes the claim that $|X| \leq C_k Y$ for some C_k depending on k.

In some cases it will instead be convenient to work in an asymptotic setting, in which there is an explicitly designated asymptotic parameter (such as x) going to infinity. In that case, all mathematical objects will be permitted to depend on this asymptotic parameter, unless they are explicitly referred to as being fixed. We then use $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote the claim that $|X| \leq CY$ for some fixed C. Note that in slight contrast to the non-asymptotic setting, the implied constant C here is allowed to depend on other parameters, so long as these parameters are also fixed. As such, the asymptotic setting can be a convenient way to manage dependencies of various implied constants on parameters. In the asymptotic setting we also use $X = o(Y)$ to denote the claim that $|X| \leq cY$, where c is a quantity which goes to zero as the asymptotic parameter goes to infinity.

Remark 5. *In later posts we will make a distinction between implied constants C that are effective (they can be computed, at least in principle, by some explicit method) and those at are ineffective (they can be proven to be finite, but there is no algorithm known to compute them in finite time).*

We use $d|n$ to denote the assertion that d divides n, and a (q) to denote the residue class of a modulo q.

We use 1_E to denote the indicator function of a set E, thus $1_E(x) = 1$ when $x \in E$ and $1_E(x) = 0$ otherwise. Similarly, for any mathematical statement S, we use 1_S to denote the value 1 when S is true and 0 when S is false. Thus for instance $1_{2|n} = 1_{n\ (2)=0}$ is the indicator function of the even numbers.

We use $|E|$ to denote the cardinality of a set E.