Prime numbers and L-functions

Henryk Iwaniec*

Abstract. The classical memoir by Riemann on the zeta function was motivated by questions about the distribution of prime numbers. But there are important problems concerning prime numbers which cannot be addressed along these lines, for example the representation of primes by polynomials. In this talk I will show a panorama of techniques, which modern analytic number theorists use in the study of prime numbers. Among these are sieve methods. I will explain how the primes are captured by adopting new axioms for sieve theory. I shall also discuss recent progress in traditional questions about primes, such as small gaps, and fundamental ones such as equidistribution in arithmetic progressions. However, my primary objective is to indicate the current directions in Prime Number Theory.

Mathematics Subject Classification (2000). Primary L20; Secondary N05.

Keywords. Prime numbers, L-functions.

1. Introduction

Prime numbers fascinate every mathematician, regardless of her or his field of main interest. They also capture the attention of people in other professions. I recall my popular talk in May 2005 which I delivered to engineers in my native city Elblag in Poland; never before have I heard questions about primes being asked with greater passion. Since our modern daily life is driven by computers, the prime numbers are used to combat hackers. There are offers of huge monetary awards for finding large prime numbers (which are apparently useful in cryptography). Regardless of industrial applications the prime numbers will always play a fundamental role in number theory, because they are to arithmetic as the elementary particles are to matter in physics. Primes form the heart of analytic number theory. Therefore this is a serious subject in which I have been happily working most of my life (and fortunately being paid to do so). When presenting results in this talk I shall often express my views on methods and perspectives concerning prime numbers. The tools for studying primes (like the L-functions, character sums, bilinear forms, sieve methods, combinatorial identities) are as fascinating as the results themselves; thus I will spend considerable time analyzing the strength of these tools and their potential.

This is not a survey of all that is known about prime numbers. There are truly great results concerning prime numbers, which nevertheless do not seem to give insight into

*Supported by NSF grant DMS-03-01168.

Proceedings of the International Congress
of Mathematicians, Madrid, Spain, 2006

the nature of primes. One of these, in my opinion, is the recent spectacular result of B. Green and T. Tao concerning long arithmetic progressions (you will find a full in-depth account of their result in these Congress Proceedings). My goal here is to cover various areas of analytic number theory which are oriented towards the Theory of Prime Numbers in general. Among them are the very promising developments by D. A. Goldston, J. Pintz and C. Y. Yildirim [32] concerning small gaps between primes. For a recreational style article (nevertheless deep) I refer to E. Bombieri [5], which is also very valuable for many historical details.

2. Primes versus zeros

Traditionally primes are denoted by the letter p. The set of all primes $\mathcal{P} = \{p = 2, 3, 5, 7, 11, 13, 17, 19, \dots\}$ is infinite, and in fact relatively dense. Precisely, the Prime Number Theorem asserts that $\pi(x)$, the number of primes $p \le x$, satisfies the asymptotic formula

$$\pi(x) \sim x(\log x)^{-1}, \quad \text{as } x \to \infty.$$

Hence a novice may argue that basic questions concerning the distribution of primes in various regions or in various sequences of arithmetical interest could be answered with confidence by statistical considerations. Definitely the abundance of primes is useful to support many heuristic arguments. It is often quite easy to predict where the primes are, but rigorous proofs require advanced technology. The point is, we have not yet found any structural mechanism which controls the behavior of prime numbers.

Today, for example, we cannot even determine whether there are infinitely many twin primes, although we expect there are plenty; in particular Hardy and Littlewood conjectured that

$$\pi_2(x) = |\{p \le x : p + 2 \text{ prime}\}| \sim 2cx(\log x)^{-2}$$

where $c = .6601\dots$ is a constant given by a certain product over odd primes. Somewhat related to the twin prime problem is the old question of Goldbach that every even number $N > 2$ is the sum of two primes. We have reason to believe that the number of solutions to the equation $p_1 + p_2 = N$ is quite large (it should be asymptotically $c(N)N(\log N)^{-2}$, where $c(N)$ is a positive number depending on N mildly). Nevertheless we cannot rule out the possibility that sums of two primes may miss a few even numbers. J. Pintz [60] showed that the set of even numbers $N \le X$ which are not represented by sums of two primes is extremely small, its cardinality is bounded by $O(X^{2/3})$. Note that this estimate for the missing Goldbach numbers yields the classical result of I. M. Vinogradov, that every large odd number is a sum of three primes. The meaning of a large number is, of course, subjective. But in the case of sums of primes it has provoked serious investigations. If we are not allowed to use the Grand Riemann Hypothesis then it is still not possible by powerful contemporary

computers to check that the Vinogradov theorem holds for all odd numbers > 5. One needs pure mathematics to cover the middle range (I would call it a theory of midsize numbers). J.-M. Deshouillers and his collaborators [11], [12], [14], have undertaken the task with such goals in mind (also for the Waring problem), so today we are sure that every number > 5 is a sum of at most six primes (due to O. Ramaré [62]).

One may fairly ask the questions, "Why is the Goldbach problem important, or why is it so difficult?" For the first part the answer is; "It is not important per se, it simply arises from our curiosity". I am sure many people would be happy to crack the problem, although this would make no great impact on the foundation of mathematics. For the second part the answer is; "Because it appeals to the multiplicative properties within the additive structure of integers".

Incidentally, a close analog of the twin prime conjecture for Gaussian primes appears in some problems on elliptic curves with complex multiplication (see the short communication in this Congress by Jorge Jiménez Urroz [46]).

The additive group aspects of the integers are quite well understood by means of harmonic analysis. For example, consider the Poisson summation formula

$$\sum_{m\in\mathbb{Z}^r} f(m) = \sum_{n\in\mathbb{Z}^r} \hat{f}(n)$$

where the summations on both sides are over integer vectors of the same dimension. In a slightly more general version of Poisson's formula a sum over a lattice goes to another sum over a dual lattice while the test function changes by the Fourier transform. This can be interpreted as a trace formula for a torus. The very general case of the trace formula for homogeneous spaces when the relevant group action is not commutative may look differently, however it creates similar effects. The group elements no longer correspond to other group elements, but rather they are associated with eigenvalues of a differential operator. Each side of the trace formula serves as a tool to improve our knowledge about the other side. After having established along these lines basic properties of both spectra many applications follow. This is the scheme we practice in analytic number theory (the spectral theory of automorphic forms versus sums of Kloosterman sums being a great example in the modern theory, where non-commutative harmonic analysis rules the game).

The prime numbers resist obeying this treatment unconditionally. Their dual companions are the complex zeros of the zeta function. Historically speaking the zeta function was introduced by Euler as the Dirichlet series

$$\zeta(s) = \sum_n n^{-s} = \prod_p (1 - p^{-s})^{-1}.$$

so today we call it the Riemann zeta function. Rightly so, because Riemann realized better than anybody previously that the secret of primes is revealed by the zeta function in the whole complex domain $s = \sigma + it$. Besides the above Euler product over primes,

we have the Weierstrass type product over the complex zeros

$$s(1-s)\pi^{-\frac{s}{2}}\Gamma\Big(\frac{s}{2}\Big)\zeta(s) = e^{-bs}\prod_{\rho}\Big(1-\frac{s}{\rho}\Big)e^{s/\rho}.$$

Combining the Euler product and the Weierstrass product one derives by complex variable integration the so called explicit formula

$$\sum_n \Lambda(n)f(n) = \int_1^\infty \Big(1-\frac{1}{(x-1)x(x+1)}\Big)f(x)dx - \sum_\rho F(\rho).$$

Here $\Lambda(n)$ denotes the von Mangoldt function; it is equal to $\log p$ if n is a power of p and zero elsewhere (actually it was P. Tchebyshev who first realized that counting primes p with the weight $\log p$ is more natural than with the weight one). On the right side F denotes the Mellin transform of f. This formula holds for a large class of test functions, for example, for any f which is smooth, compactly supported on $(1,\infty)$. There are also other variants of the explicit formula and for quite general L-functions.

The explicit formula (which is an involution) would be a natural analog of the trace formula for primes if only we could relate the zeros to eigenvalues of some self-adjoint operator. Hence the celebrated hypothesis of Riemann would follow (which says that all the complex zeros lie on the line $\mathrm{Re}(s)=\frac{1}{2}$, the critical line). However, in spite of many intelligent speculations (especially those inspired by Random Matrix Theory, cf. B. Conrey [9]), this vision remains a dream (Polya–Hilbert). I hope that my talk will show how much has been accomplished concerning prime numbers by roundabout methods. Yes, the RH would do a lot for primes, but, as a plain statement in the absence of intrinsic meaning of the zeros, the hypothesis does not reach far enough. Let me say with satisfaction that researchers in analytic number theory have developed tools which outperform the RH. I shall return to substantiate this claim on several occasions.

A lot is known about the complex zeros $\rho=\beta+i\gamma$ of $\zeta(s)$. There are plenty of them (all are in the critical strip $0<\mathrm{Re}(s)<1$, none on the line $\mathrm{Re}(s)=1$, which fact is equivalent with the PNT), namely

$$N(T) = |\{\rho=\beta+i\gamma;\ |\gamma|\le T\}| = \frac{T}{\pi}\log T + \mathrm{O}(T).$$

Today we know that over 40% of the zeros lay on the critical line (due to B. Conrey [8]), while relatively few are away from the critical line $\mathrm{Re}(s)=\frac{1}{2}$. Quantitative statements of such results are treasures of the zeta-function theory. For example, we have the following Density Theorem

$$\begin{aligned} N(\alpha,T) &= |\{\rho=\beta+i\gamma : \beta>\alpha, |\gamma|\le T\}| \\ &\ll T^{c(1-\alpha)}\log T \end{aligned} \tag{1}$$

for $\frac{1}{2}\le\alpha\le 1$ and $T\ge 2$, where c is an absolute constant.

What is special about this inequality? First it tells us that the chance to find zeros a fixed positive distance from the critical line diminishes rapidly with that distance. The Density Conjecture asserts that the density theorem holds with the exponent $c = 2$. Hence almost all the zeros are close to the critical line (this fact is known unconditionally). The DC is a lovely substitute for the RH in applications for estimating gaps between consecutive primes; it implies (among other things)

$$d_n = p_{n+1} - p_n \ll p_n^{\frac{1}{2}} (\log p_n)^2 \tag{2.1}$$

a result, which can be improved by the Riemann hypothesis only slightly. Therefore it is not surprising that the density theorems received a great attention (in various forms, slightly different than the above). Major developments were carried out from the late sixties to the late eighties. One of many original ideas that emerged from these investigations consists of reducing the counting of zeros to counting large values of special Dirichlet polynomials (naturally called the zero detectors). These values are larger than expected only at the hypothetical zeros off the critical line (they are not good for the zeros on the critical line!), so the phenomenon is rather superficial. In this context H. L. Montgomery [55] laid a foundation for the theory of Dirichlet polynomials. There are deep conjectures in his theory, which are interesting in their own right. Great progress was made by subsequent researchers, especially by M. Huxley [41], M. Jutila [49] and D. R. Heath-Brown [34]. In particular, Huxley succeeded in proving the density theorem with the exponent $c = 12/5$,which produces (after refinements by extra ingredients from sieve methods, see [36]), the asymptotic formula

$$\sum_{x-y<n\le x} \Lambda(n) \sim y, \quad y = x^{\frac{7}{12}}. \tag{2.2}$$

The density conjecture seems to be within reach of current technology, so it is extremely attractive, because it would fully eliminate the need for the Riemann hypothesis for important applications to the distribution of prime numbers (sorry, no prize of one million dollars for a proof of the density conjecture, unless you show that the sets whose cardinalities are being estimated in the density conjecture are all empty).

3. Gaps between primes

If we ask for slightly less than the asymptotic formula (2.2), say we are satisfied instead with a lower bound of the right order of magnitude, then the sieve method becomes a handy addition to the density theorems. Briefly speaking, the sieve offers a decomposition for sums over primes into terms, some of which are non-negative so they can be discarded at will. Of course, after dropping these inconvenient terms the asymptotic formula is lost, but one gains a greater flexibility when dealing with the remaining terms. To these one can apply the theory of Dirichlet polynomials more efficiently due to factorization properties, which are under control to some extent.

Consequently, one gets better bounds for gaps between consecutive primes. The best known unconditional result is

$$d_n = p_{n+1} - p_n \ll p_n^{0.525}$$

due to R. C. Baker, G. Harman and J. Pintz [2]. This is not yet what the RH yields, but it is very close. My point here is that the elementary arguments of combinatorial nature, like the exclusion-inclusion arguments of sieve methods, can be very powerful in conjunction with analytic tools (by exploring features of positivity before applying complex variable analysis).

Suppose the RH is true. Can one get a better bound for d_n if the zeros are regularly distributed on the critical line? Yes, but not very much better. The Pair Correlation Conjecture of Montgomery [56] offers some insight as to how the differences between zeros are distributed, but only with a limited precision in asymptotic formulas for the density function (up to a few main terms). Goldston, Heath-Brown and Julia Mueller explored these conjectures many times ending up with the following result:

$$d_n = \mathrm{o}(\sqrt{p_n}\log p_n).$$

Note that this estimate is just a bit too short to solve the old problem that prime numbers exist between every two consecutive squares. To this end one needs $d_n < 2\sqrt{p_n} + 5$.

Regardless of the zeta-function theory limitation, it is expected that d_n is much smaller. Some heuristic considerations of a probabilistic nature let Cramer [30] conjecture that $d_n \ll (\log n)^2$. While we believe this estimate could be true, one has to be cautious about Cramer's probabilistic model (it is too simplistic, it suffers from having no arithmetical elements). Indeed, Cramer's model suggests that the asymptotic formula (2.2) may hold for extremely short intervals, like $y = (\log x)^A$ with any constant $A > 2$. On the other hand H. Maier [54] showed that the asymptotic formula (2.2) fails even for some larger $y = y(x)$. His idea is quite simple, yet the consequences are very surprising (see more observations in the article by J. Friedlander [21]).

For probabilistic modeling of arithmetic quantities I would suggest to look for inspirations in the Random Matrix Theory. This wonderfully elaborated theory is capable of revealing hidden characteristics, which are impossible to find by naive straightforward thinking. Although the Random Matrix Theory is primarily analytic in essence, mysteriously enough every asymptotic formula predicted by the RMT so far seems to be correct, including the arithmetical factors. I do not completely comprehend why the two worlds of numbers, analytic and arithmetic in nature, manifest their co-existence here? Many interesting relations have been discovered and explained in this framework by B. Conrey, D. Farmer and others; see how some of these are articulated by B. Conrey [9].

Next question is; "How large can the gaps be between primes?" By the PNT it follows that $p_n \sim n\log n$, so d_n is about $\log n$ on average. More precisely we know that $d_n/\log n$ behaves like a random variable with Poisson distribution, this means

$$|\{n \le x;\, d_n > t\log n\}| \sim e^{-t}x \quad \text{for } t > 0, \text{ as } x \to \infty.$$

However, gaps larger than the average size do occur occasionally. Erdös and Rankin showed that infinitely often d_n can be as large as

$$c(\log n)(\log\log n)(\log\log\log\log n)(\log\log\log n)^{-2},$$

where c is a positive constant. (Once in a while D. Goldston asks, "What are the last words of a drowning analytic number theorist?" and he is still saying, "logloglog".)

Of course, for small gaps we expect to have $d_n = 2$ infinitely often (the twin prime conjecture). The problem of finding small gaps between primes sparked a great deal of interest (see Bombieri–Davenport [7], Huxley [42] and Maier [54]). Just a year ago the world was stunned by the following result:

$$\liminf_n \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

This is a magnificent achievement of D. A. Goldston, J. Pintz and C. Y. Yildirim [32] after over a decade of working on the problem by Goldston and Yildirim. They also showed that

$$\liminf_n \frac{p_{n+1} - p_n}{\sqrt{\log n}(\log\log n)^2} < \infty.$$

Their work represents a significant contribution to sieve methods. We shall return to this subject in Section 6.

4. Primes in arithmetic progressions

Primes in arithmetic progressions are building blocks for basic constructions in analytic number theory. Let $q > 1$ and $(a, q) = 1$. After Dirichlet we know that there are infinitely many primes $p \equiv a \pmod q$. His introduction of multiplicative characters $\chi \pmod q$ and L-functions

$$L(s, \chi) = \sum_n \chi(n) n^{-s} \tag{2}$$

$$= \prod_p (1 - \chi(p) p^{-s})^{-1} \tag{3}$$

are commonly considered as the beginning of analytic number theory (should Euler be the father?). The historical memoir of Riemann on the zeta-function has been naturally extended to the family of Dirichlet L-functions including the Riemann hypothesis. The so-called Grand Riemann Hypothesis asserts that all the zeros of $L(s, \chi)$ in the critical strip $0 < \mathrm{Re}(s) < 1$ are on the critical line $\mathrm{Re}(s) = \frac{1}{2}$; it is equivalent to the

asymptotic formula

$$\psi(x; q, a) = \sum_{\substack{n \le x \\ n \equiv a \pmod q}} \Lambda(n)$$
$$= \frac{x}{\phi(q)} + \mathrm{O}(\sqrt{x}(\log x)^2)$$

where the implied constant is absolute. Note that the above formula is meaningful (i.e. its main term exceeds the error term) for relatively large modulus q in terms of x, namely, it holds uniformly in $q \ll \sqrt{x}(\log x)^{-3}$.

In analytic number theory, the uniformity of asymptotic formulas, or inequalities, with respect to the involved parameters, is the key issue, because these parameters constitute structural components for connecting distinct sets of numbers. We shall see this machine in action when producing primes in special sparse sequences by sieve methods. In this regard the Grand Riemann Hypothesis would be most useful. The great virtue of GRH for all natural L-functions (like L-functions on ideals in number fields, L-functions of Galois representations, L-functions of elliptic curves, and ultimately the automorphic L-functions of any degree) is its ability to yield strong and neat estimates of great uniformity with respect to the relevant invariants. For industrial applications we should use GRH without hesitation. But for a critical researcher the prospect of obtaining extra strong results might deter him from attacking the GRH. My point is that current ideas (in analytic number theory) are not capable, in fact not even aimed to penetrate the subject so deeply. By design many methods are successful in breaking only through the surface of the problem, which is critical for its solution. For example one does not need the full strength of the Lindelöf hypothesis for L-functions in terms of the conductor (another consequence of the GRH), but a small improvement in the convexity bound is just sufficient for proving major results. H. Weyl and D. Burgess get credit for establishing the first subconvexity bounds for the Riemann zeta function in the s-aspect, and for the Dirichlet L-functions in the conductor aspect, respectively. Do not expect that a small improvement of a convexity bound is possible by squeezing the functional equation arguments; in every case it is the state-of-the-art technique which crashes the barrier. I refer to the full in-depth presentation by P. Michel in the Number Theory Section of this Congress for other examples and for many original ideas with surprising applications.

We now return to primes in arithmetic progressions. It is known that $L(s, \chi)$, for any character χ of conductor q, does not vanish in the region

$$\sigma > 1 - \frac{c}{\log q(|t| + 1)}, \quad s = \sigma + it \tag{4.1}$$

where c is a positive absolute constant, with at most one exception. The exceptional character is real and the exceptional zero is also real and simple (due to E. Landau [52]). The problems of the exceptional character are fascinating, so we shall speak more about these issues in a separate section. By the above zero-free region one

derives the unconditional formula

$$\psi(x;q,a) = \frac{x}{\phi(q)} - \frac{\chi(a)}{\phi(q)}\frac{x^{\beta}}{\beta} + \mathrm{O}(x\,\exp(-b\sqrt{\log x})) \tag{4.2}$$

where χ is the exceptional character, β is the exceptional zero, so

$$1 - \frac{c}{\log q} < \beta < 1, \tag{4.3}$$

and b is a positive absolute constant. Here on the right-hand side, the second term disappears if the exceptional character does not exist. However, if the exceptional character does exist with the exceptional zero very close to one, then it distorts the asymptotic for $\psi(x;q,a)$ dramatically. Depending on the value $\chi(a) = -1, 1$, we find that the asymptotic number of prime $p \equiv a \pmod q$, either doubles or reduces to nothing, respectively. Of course, we do not believe in this phenomenon, yet we cannot rule it out. In fact there are many situations where the existence of the exceptional character would be welcome.

Even if the GRH is true the primes are not very uniformly distributed among various residue classes. In 1853 P. Tchebyshev noticed that the class 3 (mod 4) has a tendency of representing more primes than the class 1 (mod 4). Of course, the bias must not be large, because of the uniformity guaranteed by the GRH. A hundred years later S. Knapowski and P. Turan [51] succeeded in justifying this observation rigorously creating the so called Comparative Prime Number Theory, which is based on another great invention, the Turan Power Sum Method. Their results stimulated further investigations by several authors, who discovered other phenomena. M. Rubinstein and P. Sarnak [63] gave a quite precise characterization of those moduli and the residue classes for which the Tchebyshev bias is present. They also provided the measures which describe quite precisely the distribution of the error terms

$$E(x;q,a) = \psi(x;q,a) - \frac{x}{\phi(q)}$$

simultaneously with respect to specific collections of the classes $a \pmod q$. Better yet, their work illuminates the interface where primes and zeros communicate with each other.

While we may have to wait a long time for a proof of the GRH, its main consequence is already established in practical terms, that is to say, the error term $E(x;q,a)$ is showed to be relatively small on average with respect to the modulus q over the same range which the GRH covers. Indeed we know that

$$\sum_{\substack{q\le Q\\(q,a)=1}} \lambda(q,a)E(x;q,q) \ll x(\log x)^{-A} \tag{4.4}$$

where $\lambda(q,a)$ are arbitrary real numbers absolutely bounded, A is any positive number and $Q = \sqrt{x}(\log x)^{-B}$, with B depending only on A. Here the quality of the

bound is not impressive, we save only a factor $(\log x)^A$ relative to the trivial bound, which nevertheless is sufficient for most applications. More important is that we have here useful estimates for quite large moduli. Because the coefficients $\lambda(q,a)$ are arbitrary we can sum the error terms $E(x;q,a)$ with absolute values, so no cancellation occurs. In this format the above estimate was established in 1965 independently by E. Bombieri [3] and A. I. Vinogradov [67] (actually the original statement of Bombieri was slightly more refined). This was a great triumph of the then new technology, the large sieve. The Bombieri–Vinogradov theorem turned out to be particularly useful in applications of combinatorial sieve methods, having the effect of replacing the GRH.

Yet, neither the GRH nor the Bombieri–Vinogradov theorems are the last words about primes in arithmetic progressions. We expect that (4.4) holds with $Q = x^{1-\varepsilon}$, or even better that each error term satisfy

$$E(x;q,a) \ll x^{\varepsilon}(x/q)^{\frac{1}{2}}, \quad \text{if } 1 \le q \le x.$$

If one attempts to prove the latter estimate by an appeal to the explicit formula, then the task boils down to having extremely regular distribution of zeros of L-functions on the critical line (mostly the zeros near the central point $s = \frac{1}{2}$ play a role). However, nothing like that is known, and nobody so far has had the courage to formulate the required distribution. Let me say that the analysis based on the n-level correlation theory (cf. Z. Rudnick and P. Sarnak [64]) does seem to hint for the problem in question, however not enough. Nevertheless, an important progress was made in the eighties by working directly with primes. We know today (see Bombieri–Friedlander–Iwaniec [6] and Fouvry [16]) that (4.4) holds with $Q = x^{\frac{4}{7}-\varepsilon}$, provided a is fixed and the coefficients $\lambda(q) = \lambda(q,a)$ are well factorable (this means that for any factorization $Q_1Q_2 = Q$ with $Q_1 > 1$, $Q_2 > 1$ one can represent the coefficients $\lambda(q)$ as a convolution of two new coefficients supported on numbers less than Q_1 and Q_2, respectively). These coefficients are almost as good as any other in applications, especially in conjunction with sieve methods.

When the modulus q is close to x there are not sufficiently many primes $p \le x$, to warrant the equidistribution among the residue classes $a(\mathrm{mod}q)$. J. Friedlander and A. Granville [22], building upon an idea of H. Maier, have shown that for every B the asymptotic formula $\psi(x;q,a) \sim x/\phi(q)$ cannot hold uniformly in the range $q \le x(\log x)^{-B}$. Even if we consider averaging over the moduli the situation does not improve, for they have shown that the Bombieri–Vinogradov estimate (4.4) cannot hold for $Q = x(\log x)^{-B}$, where A, B are arbitrary positive constants.

Primes in arithmetic progressions appear in various contexts, and not just as tools for shaping other things. For example the arithmetic in a number field (a finite algebraic extension of the field of rational numbers) requires a good knowledge of prime ideals in the ring of algebraic integers of that field. One can find all of them by factoring the rational primes p. If the extension is abelian the kind of factorization depends almost exclusively on the residue class of p modulo the conductor of the field. Clearly, a good question is; "What is the first prime ideal in a number field

(precisely the non-rational prime ideal of the lowest norm), that is to say the smallest "elementary particle" of the field?"

In particular we have the question; "What is the least prime number in an arithmetic progression?", say $p_{\min}(q, a) \equiv a \pmod q$. The best known asymptotic formula (the Siegel–Walfisz theorem)

$$\psi(x; q, a) = \frac{x}{\phi(q)} + \mathrm{O}(x(\log x)^{-A}) \tag{4.5}$$

provides a poor estimate, while the GRH would tell us that $p_{\min}(q, a) \ll q^{2+\varepsilon}$. One of the deepest results in the Prime Number Theory is the estimation of Yu. V. Linnik

$$p_{\min}(q, a) \ll q^L,$$

where L and the implied constant are absolute, effectively computable. The original arguments of Linnik, and later refinements by several authors (cf. [39]), are gems of the theory. Among many strong ingredients one finds the repulsion property of the exceptional real zero of a real character L-function. It is a fascinating subject to which we give more attention in Section 6. Recently J. Friedlander and H. Iwaniec (see [45]) developed a different approach to the Linnik theorem avoiding many of these ingredients; our method does not essentially appeal to the zeros of L-functions but instead it applies a lot of sieve arguments. The best known Linnik constant $L = 5.5$ is due to D. R. Heath-Brown [37], while it is expected that $L = 1 + \varepsilon$ should be fine. Note that the statement $p_{\min}(q, a) = \mathrm{o}(q \log q)$ would be false! For more delicate results and fine speculations we refer to A. Granville [29].

Because the uniformity in estimates for $\pi(x; q, a)$ with respect to q is vital in practice, there are many interesting results which are not perfect, but non-trivial for very large moduli. In this regard the sieve methods have an advantage over the analytic methods. First, using the elementary Brun sieve, Titchmarsh showed that $\pi(x; q, a) < cx/\phi(q) \log(x/q)$ for all $q < x$, where the constant c is absolute. This is a problem of the one-dimensional sieve (often called the linear sieve). The very elegant method of Selberg, which is optimal in this case, leads to the Brun–Titchmarsh estimate with $c = 2$. The same neat estimate can be also achieved by a large sieve type argument (a Hilbert inequality due to H. L. Montgomery and R. C. Vaughan [57], [58]). It is intrinsic to the linear sieve that we miss the correct estimate by factor of two (the parity problem, see E. Bombieri [4]). Therefore it was surprising when Y. Motohashi [59] improved this estimate in some ranges by incorporating analytic arguments with the sieve theory. His work inspired further developments of the linear sieve theory (see [43]). The key new feature is the bilinear form structure of the remainder term which can be better estimated by methods of exponential sums over a finite field. Recently J. Friedlander and H. Iwaniec [25] employed estimates for extremely short exponential sums of Kloosterman type (based on the original ideas of Karatsuba, and reminiscent of the Vinogradov exponential sums method) getting an improvement essentially in the whole range,

$$\pi(x; q, a) \leq (2 - \delta)x/\phi(q) \log(x/q)$$

for $x^\alpha < q < x^\beta$ with any fixed $0 < \alpha < \beta < 1$ and some $\delta > 0$ depending only on α, β (we assume that x is large in terms of α, β). This estimate does not break the parity barrier of sieve theory; it would do so if we had $\log(x/q)$ replaced by $\log x$. However we are skeptical that one can go that far with the sieve arguments, because the consequences would be fantastic, namely the non-existence of the exceptional zero (apply the formula (4.2)).

5. Problems of exceptional character

Perhaps there is nothing more exciting in analytic number theory than debates about the exceptional character. I have written a long survey on the subject [45], and I now repeat some of my observations here. For brevity let me restrict the story to the odd real characters; although many remarks are valid for the even characters as well. Such a character $\chi_D(n) = \left(\frac{-D}{n}\right)$ is the Kronecker symbol, whose values $+1, -1, 0$ characterize factorization of rational primes into ideals in the imaginary quadratic field $K = Q(\sqrt{-D})$. Here $-D$ is the discriminant of the field K and $D > 0$ is the conductor of the character χ_D.

Let me begin with an intriguing observation by L. Euler, that the polynomial $x^2 - x + 41$ takes prime values for all $0 \le x \le 40$. No, do not hope for many such amusing examples! We know today that the reason for seeing so many first prime values of the Euler polynomial is that its splitting field $K = Q(\sqrt{-163})$ has the class number one, that is to say, every integral ideal of K is principal. G. Rabinowitsch [61] made it clear with his criterion for all the discriminants $-D$ with $K = Q(\sqrt{-D})$ having class number one. Long ago C. F. Gauss conjectured that there are exactly nine such fields, and hence our saga began. After the early 1930s (Deuring, Heilbronn, Linfoot), we knew that the Gauss list is complete except possibly for one missing discriminant, so the problem was to show that the tenth discriminant did not exist! Numerical computations were useless until we got an effective bound for the class number $h(-D)$ in terms of D.

By the Dirichlet class number formula

$$L(1, \chi_D) = \frac{\pi h(-D)}{\sqrt{D}}, \quad \text{if } D > 4 \tag{5.1}$$

and by the estimates $(\log\log D)^{-1} \ll L(1, \chi_D) \ll \log\log D$, which follow from the Riemann Hypothesis for $L(s, \chi_D)$, we infer a pretty good location for the class number

$$\sqrt{D}(\log\log D)^{-1} \ll h(-D) \ll \sqrt{D}\log\log D.$$

Of course, this would end the saga for someone who takes the Riemann hypothesis for granted, but we are not willing to do so. Therefore, we are looking for an unconditional lower bound for $L(1, \chi_D)$. If there is no exceptional zero of $L(s, \chi_D)$ in the region (4.1), then $L(1, \chi_D) \gg 1/\log D$, and consequently $h(-D) \gg \sqrt{D}/\log D$, where

the implied constant is effectively computable. This lower bound would be more than sufficient to determine all the imaginary quadratic fields $K = Q(\sqrt{-D})$ with any fixed class number $h(-D) = 1, 2, 3, \ldots$. Ironically the original idea of Dirichlet for estimating $L(1, \chi_D)$ (which was needed for the existence of primes in arithmetic progressions) uses the trivial bound $h(-D) \geq 1$ in the formula (5.1), offering nothing useful for the class number problem itself. E. Landau [53] gave the first non-trivial bound $h(-D) \gg D^{1/8}$ (shortly after C. L. Siegel [66] improved it), which is quite impressive, but still useless for solving the Gauss problem. Landau's estimate is defective (so is Siegel's), because the implied constant is ineffective. This translates into saying that the exceptional zero of $L(s, \chi_D)$ is not too close to one. Specifically Siegel proved that for every $\varepsilon > 0$ there exists a constant $c(\varepsilon) > 0$ such that

$$\beta \leq 1 - c(\varepsilon) q^{-\varepsilon}. \tag{5.2}$$

The constant $c(\varepsilon)$ cannot be computed. This deficiency doesn't matter for many applications, but one must be aware that certain statements with ineffective constants have no content. For example, consider the following grotesque theorem: There is a constant $T > 0$, such that if every complex zero of ζ(s) with height $< T$ is on the critical line then the RH is true. However Siegel's result is serious, it is indispensable for the Bombieri–Vinogradov theorem.

One of many intriguing characteristics of the exceptional character is its repelling property. Roughly speaking if the exceptional zero of $L(s, \chi_D)$ is closer to the point $s = 1$, then the other zeros are farther away from $s = 1$, not only the real zeros of $L(s, \chi_D)$, but also all the zeros of any other natural L-function. This effect seems to be pretty universal. Here is how the mystery can be explained in a few steps:

- Suppose a real zero of $L(s, \chi_D)$ is close to $s = 1$.
- Then $L(1, \chi_D)$ is very small.
- Consequently, by the class number formula $h(-D)$ is small.
- Therefore, the prime numbers which split in $K = Q(\sqrt{-D})$ are rare.
- Hence the character χ_D takes value -1 at almost all primes.
- This says that χ_D pretends to be the Möbius function on squarefree numbers, because both are multiplicative.
- While also being periodic the character χ_D works nicely with any natural L-function by twisting.
- The natural L-function after twisting is still entire and in the same time it pretends to be the inverse.
- In conclusion the natural L-function, whatever it is, cannot vanish in vast regions.

This colorful scenario is a dream which we wish were true. For one reason the exceptional character could help prove beautiful theorems about primes without recourse to the Grand Riemann Hypothesis. In fact, we shall see that the existence of the exceptional character can permit us to do better than what we can do with the GRH.

Before indulging ourselves in this illusory situation, let me come back to reality with some historical points. To derive effective results, in principle, there is no reason to abandon the repelling property of a real zero; provided this special zero is really real, that is it has a numerical value. Fine, but how can one produce this repellant if we believe in the GRH? The only hope along such ideas is to find an L-function which vanishes at the central point $\beta = \frac{1}{2}$. A quick examination of Siegel's arguments reveals that any zero $\beta > \frac{1}{2}$ has some power of repelling, which is not as strong as that of the zero near the point $s = 1$, yet sufficiently strong for showing effectively that

$$h(-d) \gg D^{\beta-\frac{1}{2}}(\log D)^{-1}.$$

In view of this property the first question that arises is; "Does the central zero have an effect on the class number?" In the remarkable paper by J. Friedlander [20] we find the answer; "Yes it does and the impact depends on the order of the central zero!" The second question is; "How does one find L-functions which do vanish at the central point with sufficient multiplicity?" Definitely the Dirichlet L-functions do not qualify (by a folk conjecture $L\left(\frac{1}{2}, \chi\right) > 0$ for any real character χ). Obviously, if $L(s, f)$ is self-dual and has the root number -1 in its functional equation, then $L\left(\frac{1}{2}, f\right) = 0$. Alas, not a single such case was known until J. V. Armitage [1] gave an example of an L-function of a number field.

A lot more possibilities are offered by elliptic curves. Indeed, according to the Birch and Swinnerton-Dyer conjecture the Hasse–Weil L-function of an elliptic curve E/Q vanishes at the central point to the order equal to the rank of the group of rational points. D. Goldfeld [31] first took this route successfully assuming he was given an L-function which vanishes at the central point to order three (a double zero at the central point is not repelling). It is easy to point out the candidate as it is easy to construct an elliptic curve of rank $g = 3$ (by forcing three points to lay on a curve), but proving that it is modular with the corresponding L-function vanishing to the correct order at the central point is much harder a problem. Ten years after Goldfeld's work such an L-function was provided by B. Gross and D. Zagier [33], concluding with the lower bound

$$h(-D) \gg \prod_{p|D}(1 - \frac{2\sqrt{p}}{p+1}) \log D.$$

This bound is effective, so today one can determine (time permitting) all the imaginary quadratic fields $K = Q(\sqrt{-D})$ which have a given class number.

Some renowned researchers contemplated that the GRH could hold for any natural L-function except possibly for some real zeros very close to the point $s = 1$. But

recently P. Sarnak and A. Zaharescu [65] proved that such a zero would ruin the GRH very badly. Briefly speaking, the L-functions for certain cusp forms would have complex zeros off the critical line.

Let us assume the Grand Riemann Hypothesis. A simple or double zero at the center has no visible effect on the class number. But what about the other zeros on the critical line, which we know appear in abundance? Rather than asking for high multiplicity, more hopefully, one should ask if some clustering of the complex zeros do the job. Yes indeed, to wipe out the exceptional zeros of real Dirichlet L-functions one only needs to appeal to the zeros of the Riemann zeta function. B. Conrey and H. Iwaniec [10] showed that if the gaps between zeros of $\zeta(s)$ are smaller than half of the average value sufficiently often, then the exceptional zeros do not exist (see the original paper for more precise statements). Of course, one must question whether the required small gaps do occur in reality. We cannot yet prove it, but we have strong evidence deduced from the Pair Correlation Theory of H. L. Montgomery [56].

We used to think that the zeros of distinct L-functions do not see each other, that they are governed by independent distribution laws (read the Katz–Sarnak philosophy [50]). This is not so clear today by the results of [10] (the Riemann zeta function conspires against the Dirichlet L-functions?).

The L-functions co-exist and interact strongly in families. Analytic number theorists are very successful in exploring families associated with objects, which are in some sense orthogonal and complete in an appropriate ambient space (like the Hilbert space of modular forms). For example, let H_k be the basis of the linear space of cusp forms of weight k which are simultaneous eigenfunctions of all the Hecke operators on the modular group $\mathrm{SL}(2, \mathbb{Z})$. Let $H(K)$ be the union of H_k for $k \leq K$, $k \equiv 0(\bmod 4)$, so $H(K)$ has about K^2 forms. Let $L(s, f)$ be the Hecke L-function associated with f in $H(K)$. Note that the root number (the sign of the functional equation) is $+1$ (we normalize the Hecke operators so that the central point of any $L(s, f)$ is at $s = \frac{1}{2}$). Motivated by questions of the exceptional character H. Iwaniec and P. Sarnak [48] proved that at least 50% of the L-functions in the set $H(K)$ do not vanish at the central point. Actually we gave fairly good positive lower bounds. We believe that 100% of these central values are strictly positive (and relatively large). So what is special about the 50%? If we got just a bit more, then we could say good-bye to the exceptional zero. Keep in mind that here we took a somewhat opposite direction for attacking the exceptional zero, that is to say, we do not explore zeros of L-functions as repellants, but instead we utilize a lot of positive central values.

So far we have tried vigorously to eliminate the exceptional zero, because it is a pest in many areas of analytic number theory. However, as we have said previously, in some applications the exceptional character and its exceptional zero are very welcome. In particular the exceptional character helps to deal with prime numbers in exotic sequences where the Riemann hypothesis is not applicable. Recall that the real character χ_D is said to be exceptional if the corresponding L-function has a real zero $\beta > 1 - c/\log D$, for some fixed sufficiently small positive constant c. Assuming that this happens for arbitrarily large modulus D one can show the existence of

primes in many sparse sequences. I call these "the illusory primes", because today nobody believes that the exceptional zeros exist in reality. First D. R. Heath-Brown [35] showed under the above condition that there are infinitely many twin primes. J. Friedlander and H. Iwaniec [26], [27], [28] found illusory primes in other sets. For example we showed under similar conditions (see [28]) that the polynomial $x^2 + y^6$ represents infinitely many primes.

Having enjoyed the assistance of the exceptional zeros for the Dirichlet L-functions in the quest for prime numbers we can only dream of extending this illusory world to other kind of L-functions. But we already know that the L-functions of cusp forms are not exceptional; nor are the associated symmetric square L-functions (due to Goldfeld, Hoffstein, Lockhart, Lieman, Ramakrishnan). In view of these results and related intense investigations, the real character appears to be the hardest stubborn case!

6. Capturing primes by sieve methods

The sieve methods were created with great expectation for finding the twin primes and for proving the Goldbach conjecture. The first ideas of Viggo Brun of 1915–1924 followed the exclusion-inclusion procedure as in the ancient Eratosthenes sieve. Over fifty years many ramifications of this approach have been developed, notably the combinatorial sieve, the Selberg upper bound sieve and the Bombieri asymptotic sieve (see my article [44] in the Proceedings of the ICM in Helsinki). Although the principal ideas are elementary, it was necessary to incorporate analytic arguments for the finest estimates. In the most important cases, like the linear sieve, we know the optimal results. Unfortunately they are too weak to give prime numbers in general sequences for which the methods apply. Not because we overlooked something, but rather because of an intrinsic barrier, which is called the *parity phenomenon*. The parity phenomenon is best explained in the context of Bombieri's asymptotic sieve [4]. This says that within the classical conditions for the sieve one cannot sift out all numbers having the same parity of the number of their prime divisors. Never mind producing primes; we cannot even produce numbers having either one, three, five or seven prime divisors. However under the best circumstances we can obtain numbers having either 2006 or 2007 prime divisors. Similarly we can also obtain numbers having either one or two prime divisors, but we are not able to determine which of these numbers are there, probably both.

Therefore, in order to distinguish primes from numbers having two prime divisors it is necessary to extend the system of sieve conditions by adding a new condition. We shall explain the new idea by modifying the asymptotic sieve of Bombieri.

Suppose $\mathcal{A} = (a_n)$ is a sequence of real, non-negative numbers. We are after the sum

$$S(x) = \sum_{n \le x} a_n \Lambda(n). \tag{6.1}$$

Recall that $\Lambda(n)$ denotes the von Mangoldt function which is supported on powers of prime numbers (in practice it is easy to ignore the high powers). Assume we are given natural approximations to the sums

$$A_d(x) = \sum_{\substack{n \le x \\ n \equiv 0 (\bmod d)}} a_n,$$

specifically we have

$$A_d(x) = g(d)A(x) + r_d(x)$$

where $g(d)$ is a multiplicative functions with $0 \le g(p) < 1$ for all p and $r_d(x)$ is considered to be an error term, which is relatively small. Here $g(d)$ represents the density of the subsequence of elements a_n with n being divisible by d. Naturally we assume that $g(d)$ is multiplicative, because we believe that the divisibility by distinct primes are independent events (this is not exactly true in stronger models of the sieve). Writing $\Lambda(n)$ as a convolution of the Möbius function and the logarithm, or more conveniently writing

$$\Lambda(n) = \sum_{d|n} \mu(d) \log d,$$

we arrange $S(x) = H(x)A(x) + R(x)$, where

$$H(x) = -\sum_{d \le x} \mu(d)(\log d)g(d),$$

$$R(x) = -\sum_{d \le x} \mu(d)(\log d)r_d(x).$$

The density function $g(d)$ usually satisfies natural regularity conditions, which imply that $H(x)$ has a limit

$$H(x) \sim H = \prod_p (1 - g(p)) \left(1 - \frac{1}{p}\right)^{-1}. \tag{6.2}$$

Ignoring the remainder term $R(x)$ one is led to the following asymptotic formula

$$S(x) \sim HA(x), \quad \text{as } x \to \infty. \tag{6.3}$$

Although one cannot ignore the remainder terms arbitrarily, the above asymptotic formula is conjectured to hold for every natural sequences $\mathcal{A} = (a_n)$. This agrees with formulas derived by various heuristic arguments, for instance by the circle method.

Why do we expect that the remainder term $R(x)$ is insignificant? There are primarily two reasons. First if the moduli d are relatively small, say $d < D$ with some $D = D(x) \le x$, then the individual error terms $r_d(x)$ are small. But for d

large, closer to x, the error terms $r_d(x)$ are comparable in size with the main terms $g(d)A(x)$. Hence their contribution is insignificant not because they are relatively small, but because of a cancellation in $R(x)$, which is due to the sign change of the Möbius function $\mu(d)$. The first part with $d < D$ belongs to the classical system of sieve conditions while the remaining part is critical for breaking the parity barrier. Looking behind the scene is the randomness of the Möbius function, which we loosely articulate as the following principle:

Randomness of the Möbius function. *The Möbius function $\mu(d)$ changes sign with unbiased fashion towards any natural sequence $c(d)$, thus producing a considerable cancellation in sums of the twisted terms $\mu(d)c(d)$.*

It is hard to imagine that a natural sequence conspires with the Möbius function, so it is frequently save to accept the heuristic formula (6.3). Then why do we face the parity barrier of sieve methods which prevents us from capturing prime numbers? Because the combinatorial sieve constructions make use of a truncated Möbius function which is obviously biased to the pure Möbius function! The Selberg sieve is somewhat different, but not much in this regard.

Estimating sums of terms twisted by the Möbius function is usually as difficult as that of the von Mangoldt function, so we are not yet done. In the next step of the sieve for primes we convert the twisted sums to bilinear forms. The latter can be estimated using a variety of tools of operator theory with the most successful being the duality principle and the large sieve inequality (not a sieve method, please). Here the structure of a bilinear form plays a vital role. It works for quite general coefficients so one can escape from the vicious circle created by the Möbius function. Let me give a simple but quite general result which is derived along the above lines.

Proposition (sieve for primes). *Suppose a sequence of non-negative numbers $\mathcal{A} = (a_n)$ has the density function $g(d)$ which is multiplicative with $0 \le g(p) < 1$ and*

$$\sum_{p\le y} g(p)\log p = \log y + c_g + O(1/\log y), \tag{6.4}$$

for any $y \ge 2$, with c_g a constant. Suppose

$$\sum_{d\le D} |r_d(x)| \le A(x)(\log x)^{-2}, \tag{6.5}$$

$$\sum_{\ell} \Big| \sum_{\substack{\ell m\le x\\ z<m\le z^2}} \mu(m)a_{\ell m}\Big| \le A(x)(\log x)^{-2}, \tag{6.6}$$

where $z = x^\delta$, $D = x^{1-\delta}$ with some small $\delta > 0$. Then we have

$$\sum_{n\le x} a_n\Lambda(n) \sim HA(x), \quad \text{as } x \to \infty. \tag{6.7}$$

The first condition (6.5) is classical in sieve theory, and sometimes it can be established for $D = x^{1-\delta}$ with δ arbitrarily small. In this case the first condition alone suffices to derive an asymptotic formula (due to Bombieri)

$$\sum_{n\le x} a_n \Lambda_k(n) \sim kHA(x)(\log x)^{k-1}$$

for any $k \ge 2$, where $\Lambda_k(n)$ is the von Mangoldt function of order k which is supported on numbers having at most k distinct prime divisors. As we mentioned before this asymptotic formula must fail for $k = 1$, because of the parity barrier. Our second condition (6.6) takes care of this barrier. This bilinear form estimate is much harder to establish, yet it is in the realm of modern technology.

Example (Fouvry–Iwaniec). The sequence $\mathcal{A} = (a_n)$ with

$$a_n = \sum_{\ell^2+m^2=n} \Lambda(\ell)$$

satisfies (6.5) and (6.6). Therefore we have

$$\sum_{\ell^2+m^2\le x} \Lambda(\ell)\Lambda(\ell^2+m^2) \sim Hx$$

where H is a positive constant. Hence there are infinitely many primes of type $p = \ell^2 + m^2$ where ℓ is also prime.

Presently there are various variants of sieve axioms which are capable of producing primes (see Friedlander and Iwaniec [23]), Heath-Brown [37]). In the above proposition the axioms are realistic only if the sequence $\mathcal{A} = (a_n)$ is relatively dense, while the other versions can handle quite sparse sequences. Of course the verification of these axioms for sparse sequences is even harder, but the results are more impressive.

Example (Friedlander–Iwaniec). We have

$$\sum_{a^2+b^4\le x} \Lambda(a^2+b^4) \sim Hx^{\frac{3}{4}}.$$

Example (Heath-Brown). He has

$$\sum_{a^3+2b^3\le x} \Lambda(a^3+2b^3) \sim Hx^{\frac{2}{3}}.$$

Let me point out that in every case considered so far the prime producing sieve does not actually produce primes. At best what it does is allow the search for primes in a target sequence to be augmented by using primes in another sequence which has a simpler structure so we know it contains primes by standard analytic arguments, usually like the zeta function methods. Yes, it is a steal, but not easy. Usually these

transformations are far more advanced than a proof of the PNT in the comparative sequence.

The same can be said about the prime producing sieve in the work of Goldston–Pintz–Yildirim [32], although their approach is very different. They start with a collection of some distinct positive integers $h_1, \dots, h_r$, to which they associate the arithmetic function

$$W(m) = \sum_{1 \le i \le r} \Lambda^\flat(m - h_i) - \log m$$

where $\Lambda^\flat(n) = \log n$ if n is prime, and zero otherwise. Assume that the Bombieri–Vinogradov theorem holds with moduli $q \le Q = x^\theta$, for some $\theta > \frac{1}{2}$, and that the number of shifts r is sufficiently large in terms of θ. Summing $W(m)$ over $m \le x$ with certain non-negative weights (Selberg's sieve weights of various dimensions) they managed to show that $W(m)$ is positive for many $m \le x$. For these m's at least two of the shifted numbers $m - h_1, \dots, m - h_r$ are primes (these primes are not produced by sieve weights, they are borrowed from the Bombieri–Vinogradov theorem, which extracts them from the Siegel–Walfish theorem). Consequently the gap between these primes is bounded by $H = \max |h_i - h_j|$. Note that the result requires the level of distribution of primes in arithmetic progressions to be x^θ with $\theta > \frac{1}{2}$, which the GRH does not reach. Nevertheless, as we said in Section 4 we believe this should hold with any $\theta < 1$. Assuming this conjecture (the Elliott–Halberstam conjecture) the arguments of Goldston–Pintz–Yildirim lead to a conclusion that there are infinitely many gaps between distinct primes which do not exceed 16. Wow!

7. Bilinear forms technique for sums over primes

The core of Prime Number Theory consists of the distribution of primes in short segments, in arithmetic progressions, in homogeneous polynomials and in other similar sparse sequences. This territory expands to number fields, where the prime ideals play the role of primes. Here new challenging aspects emerge, particularly important being the uniformity of estimates with respect to the field invariants. For example, the distribution of prime ideals in the Galois conjugacy classes (the Tchebotarev theorem) is considered as a non-abelian analog of the distribution of primes in arithmetic progressions. Then one goes further into the territory of automorphic forms, where for example one may study the Hecke eigenvalues at primes. Why at primes? Because these eigenvalues admit a geometric interpretation (consider the group of points of an elliptic curve over a finite field). One may view this section of the theory of primes as a natural continuation of the celebrated memoir of Riemann on the zeta function. Here analytic arguments come to fruition through the zeros of L-functions. H. Davenport named this territory "multiplicative number theory".

Now I would like to go through a different territory of sums over primes which can hardly be treated by L-function methods. They are distinguished by the idea of

bilinear forms. Consider the sum

$$S_f(x) = \sum_{p \le x} f(p)$$

where f is an arithmetic function defined on primes, but not necessarily on all positive integers. Of course, one can always extend f to all integers arbitrarily, however a natural extension may not suggest itself easily. Very often, neither the associated Dirichlet series

$$\sum_n f(n) n^{-s},$$

nor the Fourier series

$$\sum_n f(n) e(nz)$$

has any beneficial properties (these series are logical creatures if f is multiplicative or additive respectively).

The bilinear forms technique rearranges the sum $S_f(x)$ into a number of other sums (as in sieve methods), the key one being of type

$$S_f(M, N) = \sum_{m \le M} \sum_{n \le N} \alpha_m \beta_n f(m, n)$$

with some specific coefficients α_m, β_n. Because of the complexity of these coefficients one cannot expect to treat them in any other way than as general complex numbers. For this reason the sum $S_f(M, N)$ is almost a genuine bilinear form. A non-trivial estimate is still possible; provided f does not resemble a multiplicative function, and f has a considerable sign variation. Paraphrasing, we are looking for a non-trivial estimate of the norm of the corresponding operator, which in this case is the largest eigenvalue of the corresponding matrix. While there are lots of possibilities here, it turns out that an application of Cauchy's inequality is most popular, because it is very flexible. I do not mean a straightforward application, in many situations it precedes with elaborate preparation of the bilinear form by exploiting its characteristics. For example it is wise to perform some kind of linearization before applying Cauchy's inequality to avoid the increase of the dimension of the resulting lattice point problem (see how in [17] we landed in the domain of complex integers when counting the rational primes of type $p = \ell^2 + m^2$ with ℓ prime).

Let me state one of many results about general sums over primes which exhibit the bilinear form technique.

Theorem (Duke–Friedlander–Iwaniec). *Let $f(n)$ be any sequence of complex numbers with $|f(n)| \le \tau(n)$. Suppose*

$$\sum_{d \le y} \Big| \sum_{dm \le x} f(dm) \Big| \le x (\log x)^{-2},$$

$$\sum_m \Big| \sum_{\substack{pm\le x\\ w<p<v}} p^{it} f(pm)\Big| \le x(\log x)^{-4}$$

for $x \ge e^{1/\varepsilon}$ with $0 < \varepsilon \le \frac{1}{12}$, where t is any real number and the parameters y, v, w are given by $y = x^{\frac12-\varepsilon}$, $v = x^{\frac13-\varepsilon}$, $w = x^{\varepsilon/10\log\log x}$. Then we have

$$\sum_{p\le x} f(p) \ll \varepsilon x(\log x)^{-1},$$

where the implied constant is absolute.

Remark. The inner sum coefficients in the first double sum are constant and in the second double sum they are p^{it}, so virtually they are general because t is not fixed.

This theorem was crafted for a specific sequence in mind, namely

$$f(n) = \sum_{\nu^2+1\equiv 0\,(n)} e\Big(\frac{\nu h}{n}\Big)$$

where $h \ne 0$ is a fixed integer. That this sequence satisfies the above condition is itself a great problem which we solved using the spectral theory of automorphic forms. Hence we obtain

$$\frac{1}{\pi(x)} \sum_{p\le x} \sum_{\nu^2+1\equiv 0(p)} e\Big(\frac{\nu h}{p}\Big) \to 0, \quad \text{as } x \to \infty.$$

In other words, using Weyl's criteria from equidistribution theory we proved that the roots of the polynomial $P(X) = X^2 + 1$ in the finite field of p elements are uniformly distributed when p runs over primes (for every $p \equiv 1 \pmod 4$ there are two roots, and none if $p \equiv 3 \pmod 4$). Similar results are established for other quadratic polynomials irreducible over $\mathbb{Z}$. The problem for higher degree irreducible polynomials is out of reach by current technology. Probably the uniform distribution of roots holds no matter what is the Galois group of the polynomial.

According to a theorem of Fermat every prime $p \equiv 1 \pmod 4$ is a sum of two squares, $p = a^2 + b^2$. This representation is unique if we require a, b to be positive and b odd. We call the Jacobi symbol $s_p = \big(\frac{a}{b}\big)$ the spin of the corresponding Gaussian prime $\pi = a + bi$. As a by-product of the work on primes represented by the polynomials $X^2 + Y^4$ we have shown that the spin changes quite regularly, in fact we got (see [24])

$$\sum_{\substack{p\le x\\ p\equiv 1(4)}} s_p \ll x^{\frac{76}{77}}.$$

8. Sums over primes related to modular forms

A great variety of interesting arithmetic functions appear in modular forms, for example the Fourier coefficients of cusp forms. For an obvious reason the bilinear form technique does not work for sums over primes of a multiplicative function. In particular this would fail for the Hecke eigenvalues of any classical cusp form (holomorphic of integral weight). However it works nicely for the Fourier coefficients $\rho(n)$ of any metaplectic cusp form (a modular form on a congruence group of half-integral weight with respect to the theta multiplier). Indeed W. Duke and H. Iwaniec [15] established the following estimate

$$\sum_{p\leq x}\rho(p) \ll x^{\frac{156}{157}}.$$

D. R. Heath-Brown and S. Patterson got an estimate of similar nature for cubic Gauss sums. When investigating the low lying zeros of classical cusp forms in [47] we encountered sums of type

$$S_f(x) = \sum_{p\leq x}\lambda(p)e(2\sqrt{p})$$

where $\lambda(p)$ are the eigenvalues of the Hecke operators T_p acting on a classical cusp form on the full modular group. We normalize these so that the Ramanujan conjecture (proved by P. Deligne) becomes $|\lambda(p)| \leq 2$. Due to the twist by the exponential factor the function $f(n) = \lambda(n)e(2\sqrt{n})$ is not multiplicative, so it is possible to show by bilinear form techniques that

$$S_f(x) \ll x^{\frac{5}{6}}.$$

In the theory of exponential sums and character sums there is a reasonable expectation that cancellation should be of the order of the square root of the number of terms, unless there is a natural reason to prevent this from happening, in which case one gets a larger main term. Clearly this philosophy is consistent with the Grand Riemann Hypothesis. Indeed we can show that the sum of $f(n)$ over all integers $n \leq x$ is bounded by $O(x^{\frac{1}{2}+\varepsilon})$. Should this bound be also true for the sum restricted by primes? One would think so according to the above philosophy, but our findings suggest different things! To be convinced we applied the Möbius Function Randomness Principle which led us to the surprising asymptotic formula

$$S_f(x) \sim cx^{\frac{3}{4}}(\log x)^{-1}$$

where $c \neq 0$ is a constant (c is the value of the associated symmetric square L-function at $s = 1$, up to some elementary factors). On the other hand for the sums over primes with no twists the GRH implies

$$\sum_{p\leq x}\lambda_f(p) \ll x^{\frac{1}{2}+\varepsilon}, \tag{4}$$

$$\sum_{p\le x} e(2\sqrt{p}) \ll x^{\frac{1}{2}+\varepsilon}. \tag{5}$$

Comparing the above three estimates one may conclude that the Hecke eigenvalues $\lambda(p)$ are somewhat biased towards the exponential function $e(2\sqrt{p})$. Why does it happen at primes but not at all integers? It would be interesting to understand this behavior within the structure of modular forms.

Some important arithmetic functions at primes do not obey the law of equidistribution with respect to the natural (Lebesgue) measure. For example the angles of the Hecke eigenvalues $\lambda(p) = 2\cos(\theta_p)$ for a given cusp form on the full modular group (so it is not of complex multiplication type) are conjectured to be equidistributed with respect to the Sato–Tate measure $2\pi^{-1}(\sin\theta)^2 d\theta$ on $[0, \pi]$. This would follow (as part of Langlands' program) from the conjecture that all the symmetric power L-functions associated with the given cusp form are holomorphic in the half-plane $\mathrm{Re}(s) \ge 1$ and do not vanish on the line $\mathrm{Re}(s) = 1$.

The angles of the classical Kloosterman sums

$$K(a, b; p) = \sum_{xy\equiv 1\,(p)} e\left(\frac{ax+by}{p}\right) = 2\sqrt{p}\cos\theta_p$$

with $ab \ne 0$ are also conjectured to have the Sato–Tate measure of equidistribution. Sadly enough we do not even know whether they change sign infinitely often. However E. Fouvry and P. Michel [18], [19] showed that the Kloosterman sums to moduli, which are the product of at most twenty-three distinct primes, do change sign very often. This is a very deep work. Besides creating innovations in sieve methods they gave original transformations which reduce the problem to the Riemann hypothesis for varieties over a finite field (proved by P. Deligne).

The Kloosterman sums twisted by the real character

$$S(a, b; p) = \sum_{xy\equiv 1\,(p)} \left(\frac{x}{p}\right) e\left(\frac{ax+by}{p}\right) = 2\sqrt{p}\cos\theta_p$$

(named Salie sums), behave differently. Their angles are known to be equidistributed with respect to the natural (Lebesgue) measure. This fact was proved by W. Duke, J. Friedlander and H. Iwaniec [13] using the bilinear forms technique and the spectral theory of automorphic forms.

9. Closing remarks

Analytic number theory is fortunate to have one of the most famous unsolved problems, the Riemann hypothesis. Not so fortunately, this puts us in a defensive position, because outsiders who are unfamiliar with the depth of the problem, in their pursuit for the ultimate truth, tend to judge our abilities rather harshly. In concluding this

talk I wish to emphasize my advocacy for analytic number theory by saying again that the theory flourishes with or without the Riemann hypothesis. Actually, many brilliant ideas have evolved while one was trying to avoid the Riemann hypothesis, and results were found which cannot be derived from the Riemann hypothesis. So, do not cry, there is a healthy life without the Riemann hypothesis. I can imagine a clever person who proves the Riemann hypothesis, only to be disappointed not to find new important applications. Well, an award of one million dollars should dry the tears; no applications are required!

References

[1] Armitage, J. V., Zeta functions with a zero at $s = \frac{1}{2}$. *Invent. Math.* **15** (1972), 199–205.

[2] Baker, R. C., Harman, G., Pintz, J., The difference between consecutive primes, II. *Proc. London Math. Soc.* (3) **83** (3) (2001), 532–562.

[3] Bombieri, E., On the large sieve. *Mathematika* **12** (1965), 201–225.

[4] Bombieri, E., The asymptotic sieve. *Rend. Accad. Naz. XL* (5) **1/2** (1975/76), 243–269.

[5] Bombieri, E., Prime numbers from recreational mathematics to practical applications. Preprint of IAS.

[6] Bombieri, E., Friedlander, J. B., Iwaniec, H., Primes in arithmetic progressions to large moduli. *Acta Math.* **156** (3–4) (1986), 203–251.

[7] Bombieri, E., Davenport, H., Small differences between prime numbers. *Proc. Roy. Soc. Ser. A* **293** (1966), 1–18.

[8] Conrey, J. B., More than two fifths of the zeros of the Riemann zeta function are on the critical line. *J. Reine Angew. Math.* **399** (1989), 1–26.

[9] Conrey, J. B., L-functions and random matrices. In *Mathematics Unlimited—2001 and beyond*, Springer-Verlag, Berlin 2001, 331–352.

[10] Conrey, B., Iwaniec, H., Spacing of zeros of Hecke L-functions and the class number problem. *Acta Arith.* **103** (3) (2002), 259–312.

[11] Deshouillers, J.-M., te Riele, H., On the probabilistic complexity of numerically checking the binary Goldbach conjecture in certain intervals. In *Number Theory and its Applications* (Kyoto, 1997), Dev. Math. 2, Kluwer Acadademic Publishers, Dordrecht 1999, 89–99.

[12] Deshouillers, J.-M., Effinger, G., te Riele, H., Zinoviev, D., A complete Vinogradov 3-primes theorem under the Riemann hypothesis. *Electron. Res. Announc. Amer. Math. Soc.* **3** (1997), 99–104 (electronic).

[13] Duke, W., Friedlander, J. B., Iwaniec, H., Equidistribution of roots of a quadratic congruence to prime moduli. *Ann. of Math.* (2) **141** (2) (1995), 423–441.

[14] Deshouillers, J.-M., Hennecart, F., Landreau, B., Waring's problem for sixteen biquadrates—numerical results. *Colloque International de Théorie des Nombres* (Talence, 1999); *J. Théor. Nombres Bordeaux* **12** (2) (2000), 411–422.

[15] Duke, W., Iwaniec, H., Bilinear forms in the Fourier coefficients of half-integral weight cusp forms and sums over primes. *Math. Ann.* **286** (4) (1990), 783–802.

[16] Fouvry, É., Autour du théoréme de Bombieri-Vinogradov. *Acta Math.* **152** (3–4) (1984), 219–244.

[17] Fouvry, E., Iwaniec, H., Gaussian primes. *Acta Arith.* **79** (3) (1997), 249–287.

[18] Fouvry, E., Michel, P., Crible asymptotique et sommes de Kloosterman. In *Proceedings of the Session in Analytic Number Theory and Diophantine Equations*, Bonner Math. Schriften 360, Universität Bonn, Bonn 2003.

[19] Fouvry, E., Michel, P., Sur le changement de signe des sommes de Kloosterman. *Ann. of Math.*, to appear.

[20] Friedlander, J. B. On the class numbers of certain quadratic extensions. *Acta Arith.* **28** (4) (1975/76), 391–393.

[21] Friedlander, J. B., Irregularities in the distribution of primes. In *Advances in Number Theory* (Kingston, ON, 1991), Oxford Sci. Publ., Oxford University Press, New York 1993, 17–30.

[22] Friedlander, J. B., Granville, A., Limitations to the equi-distribution of primes. I. *Ann. of Math.* (2) **129** (2) (1989), 363–382.

[23] Friedlander, J. B., Iwaniec, H., Using a parity-sensitive sieve to count prime values of a polynomial. *Proc. Nat. Acad. Sci. U.S.A.* **94** (4) (1997), 1054–1058.

[24] Friedlander, J. B., Iwaniec, H., The polynomial $X^2 + Y^4$ captures its primes. *Ann. of Math.* (2) **148** (3) (1998), 945–1040.

[25] Friedlander, J. B., Iwaniec, H., The Brun-Titchmarsh theorem. In *Analytic Number Theory* (Kyoto, 1996), London Math. Soc. Lecture Note Ser. 247, Cambridge University Press, Cambridge 1997, 85–93.

[26] Friedlander, J. B., Iwaniec, H., Exceptional characters and prime numbers in arithmetic progressions. *Internat. Math. Res. Notices* **2003** (37) (2003), 2033–2050.

[27] Friedlander, J. B., Iwaniec, H., Exceptional characters and prime numbers in short intervals. *Selecta Math.* (*N.S.*) **10** (1) (2004), 61–69.

[28] Friedlander, J. B., Iwaniec, H., The illusory sieve. *Int. J. Number Theory* **1** (4) (2005), 459–494.

[29] Granville, A., Least primes in arithmetic progressions. In *Théorie des Nombres* (Quebec, PQ, 1987), Walter de Gruyter, Berlin 1989, 306–321.

[30] Granville, A., Harald Cramér and the distribution of prime numbers. *Harald Cramér Symposium* (Stockholm, 1993); *Scand. Actuar. J.* **1995** (1) (1995), 12–28.

[31] Goldfeld, D. M., The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3** (4) (1976), 624–663.

[32] Goldston, D. A., Pintz, J., Yildirim, C. Y., Small gaps between primes. Preprint, May 26, 2005.

[33] Gross, B. H., Zagier, D. B., Heegner points and derivatives of L-series. *Invent. Math.* **84** (2) (1986), 225–320.

[34] Heath-Brown, D. R., Zero density estimates for the Riemann zeta-function and Dirichlet L-functions. *J. London Math. Soc.* (2) **19** (2) (1979), 221–232.

[35] Heath-Brown, D. R., Prime twins and Siegel zeros. *Proc. London Math. Soc.* (3) **47** (2) (1983), 193–224.

[36] Heath-Brown, D. R., The number of primes in a short interval. *J. Reine Angew. Math.* **389** (1988), 22–63.

[37] Heath-Brown, D. R., Primes represented by $x^3 + 2y^3$. *Acta Math.* **186** (1) (2001), 1–84.

[38] Heath-Brown, D. R., Prime number theory and the Riemann zeta-function. In *Recent Perspectives in Random Matrix Theory and Number Theory*, London Math. Soc. Lecture Note Ser. 322, Cambridge University Press, Cambridge 2005, 1–30.

[39] Heath-Brown, D. R., Zero-free regions for Dirichlet L-functions, and the least prime in an arithmetic progression. *Proc. London Math. Soc.* (3) **64** (2) (1992), 265–338.

[40] Hildebrand, A., Maier, H., Irregularities in the distribution of primes in short intervals. *J. Reine Angew. Math.* **397** (1989), 162–193.

[41] Huxley, M. N., On the difference between consecutive primes. *Invent. Math.* **15** (1972), 164–170.

[42] Huxley, M. N., Small differences between consecutive primes. *Mathematika* **20** (1973), 229–232.

[43] Iwaniec, H., A new form of the error term in the linear sieve. *Acta Arith.* **37** (1980), 307–320.

[44] Iwaniec, H., Sieve methods. In *Proceedings of the International Congress of Mathematicians* (Helsinki, 1978), Acad. Sci. Fennica, Helsinki 1980, 357–364.

[45] Iwaniec, H., Conversations on the exceptional character. In *Analytic Number Theory*, Lecture Notes in Math. 1891, Springer-Verlag, Berlin 2006, 97–132.

[46] Iwaniec, H., Jiménez, U. J., Almost prime orders of CM elliptic curves modulo primes. *International Congress of Mathematicians* (Madrid, 2006), Short Communication.

[47] Iwaniec, H., Luo, W., Sarnak, P., Low lying zeros of families of L-functions. *Inst. Hautes Études Sci. Publ. Math.* **91** (2000), 55–131.

[48] Iwaniec, H., Sarnak, P., The non-vanishing of central values of automorphic L-functions and Landau-Siegel zeros. *Israel J. Math.* **120** (Part A) (2000), 155–177.

[49] Jutila, M., Zero-density estimates for L-functions. *Acta Arith.* **32** (1) (1977), 55–62.

[50] Katz, N. M., Sarnak, P., Zeroes of zeta functions and symmetry. *Bull. Amer. Math. Soc.* (*N.S.*) **36** (1) (1999), 1–26.

[51] Knapowski, S., Turán, P., On prime numbers $\equiv$ 1 resp. 3 mod 4. In *Number theory and algebra*, Academic Press, New York 1977, 157–165.

[52] Landau, E., Über die Nullstellen der Dirichletschen Reihen und der Riemannschen ζ-Funktion. *Arkiv für Mat. Astr. och Fysik* **16** (1921).

[53] Landau, E., Bemerkungen zum Heilbronnschen Satz. *Acta Arithmetica* **1** (1936), 1–18.

[54] Maier, H., Small differences between prime numbers. *Michigan Math. J.* **35** (3) (1988), 323–344.

[55] Montgomery, H. L., *Topics in multiplicative number theory*. Lecture Notes in Math. 227, Springer-Verlag, Berlin 1971.

[56] Montgomery, H. L., The pair correlation of zeros of the zeta function. In *Analytic Number Theory* (St. Louis Univ., St. Louis, Mo., 1972), Proc. Sympos. Pure Math. XXIV, Amer. Math. Soc., Providence, R.I., 1973, 181–193.

[57] Montgomery, H. L., Vaughan, R.C., The large sieve. *Mathematika* **20** (1973), 119–134.

[58] Montgomery, H. L., Vaughan, R.C., Hilbert's inequality. *J. London Math. Soc.* (2) **8** (1974), 73–82.

[59] Motohashi, Y., On some improvements of the Brun-Titchmarsh theorem. *J. Math. Soc. Japan* **26** (1974), 306–323.

[60] Pintz, J., Recent results on the Goldbach conjecture. In *Elementare und Analytische Zahlentheorie* (ed. by W. Schwarz and J. Steuding), Franz Steiner Verlag, Stuttgart 2006, 220–254.

[61] Rabinowitsch, G., Eindeutigkeit der Zerlegung in Primzahlfaktoren in quadratischen Zahlkörpern. In *Proceedings of the Fifth International Congress of Mathematicians* (Cambridge, 1912), Vol. 1, Cambridge University Press, Cambridge 1913, 418–421.

[62] Ramaré, O., On Šnirelman's constant. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **22** (4) (1995), 645–706.

[63] Rubinstein, M., Sarnak, P., Chebyshev's bias. *Experiment. Math.* **3** (3) (1994), 173–197.

[64] Rudnick, Z., Sarnak, P., The n-level correlations of zeros of the zeta function. *C. R. Acad. Sci. Paris Sér. I Math* **319** (10) (1994), 1027–1032.

[65] Sarnak, P., Zaharescu, A., Some remarks on Landau-Siegel zeros. *Duke Math. J.* **111** (3) (2002), 495–507.

[66] Siegel, C. L., Über die Classenzahl quadratischer Zahlkörper. *Acta Arithmetica* **1** (1936), 83–86.

[67] Vinogradov, A. I., The density hypothesis for Dirichet L-series. *Izv. Akad. Nauk SSSR Ser. Mat.* **29** (1965), 903–934 (in Russian).

Department of Mathematics, Rutgers, The State University of New Jersey, 110 Frelinghuysen Road, Hill Center-Busch Campus, Piscataway, NJ 08854-8019, U.S.A.

E-mail: iwaniec@math.rutgers.edu