
Conversations on the Exceptional Character

Henryk Iwaniec

Department of Mathematics, Rutgers University
110 Frelinghuysen road, Piscataway, NJ 08854, USA
e-mail: iwaniec@math.rutgers.edu

1 Introduction

Everything has its exceptional character, and the analytic number theory is no exception, it has one which is real and most perplexed. In this article I will tell the story how the existence or the non-existence of such a character shaped developments in arithmetic, especially for studies in the distribution of prime numbers. Many researchers are affected by this dangerous yet beautiful beast, and this author is no exception. I shall address questions and present results which I witnessed during my own studies.

Of course, the Grand Riemann Hypothesis for the Dirichlet L -functions rules out any exception! Nevertheless, after powerful researchers made serious attacks on the beast and got painfully defeated, it is now understandable that these people consider the problem to be as hard as the GRH itself. Some experts go further with prediction that the GRH will be established first for complex zeros, while the real zeros may wait long for a different treatment. In the meantime we have many ways of living with or without the exceptional character. In this article I try to show that this little dose of uncertainty is enjoyable and stimulating for many new ideas.

Acknowledgements

I would like to thank the C.I.M.E. for supporting my participation in this summer school in a wonderful scenery, and I am grateful to Alberto Perelli and Carlo Viola for their encouragement, patience and great help in preparing this article for publication.

Supported by the NSF grant DMS-0301168

2 The exceptional character and its zero

The characters $\chi \pmod{D}$ were introduced by G.L. Dirichlet for his proof of the equidistribution of primes in reduced residue classes modulo D , the essential ingredient being the non-vanishing of the series

$$L(s, \chi) = \sum_1^{\infty} \chi(n)n^{-s} \quad (2.1)$$

at $s = 1$. It is already in this connection that the case of real character is different from all the complex characters.

Throughout we assume that $\chi = \chi_D$ is the real, primitive character of conductor D , so it is given by the Kronecker symbol

$$\chi(n) = \left(\frac{D^*}{n} \right), \quad (2.2)$$

where $D^* = \chi(-1)D$. This character is associated with the field

$$K = \mathbb{Q}(\sqrt{D^*}), \quad (2.3)$$

which is real quadratic if $\chi(-1) = 1$, or imaginary quadratic if $\chi(-1) = -1$. The celebrated Class Number Formula of Dirichlet asserts that

$$L(1, \chi) = \frac{\pi h}{\sqrt{D}} \quad \text{if } D^* < -4 \quad (2.4)$$

and similar formula holds in other cases. Here $h = h(-D)$ is the class number of K . By the obvious bound $h \geq 1$ one gets

$$L(1, \chi) \geq \frac{\pi}{\sqrt{D}}. \quad (2.5)$$

Hence $L(1, \chi) \neq 0$, but one can also show this directly as follows.

Consider the convolution $\lambda = 1 * \chi$, i.e.

$$\lambda(n) = \sum_{d|n} \chi(d) = \prod_{p^\alpha || n} (1 + \chi(p) + \cdots + \chi(p^\alpha)) \geq 0.$$

For squares we have $\lambda(m^2) \geq 1$. Hence

$$T(x) = \sum_{n \leq x} \lambda(n)n^{-1/2} > \frac{1}{2} \log x.$$

On the other hand we find that

$$\begin{aligned}
 T(x) &= \sum_{dm \leq x} \chi(d)(dm)^{-1/2} \\
 &= \sum_{d \leq y} \chi(d)d^{-1/2} \sum_{m \leq x/d} m^{-1/2} + \sum_{m \leq x/y} m^{-1/2} \sum_{y < d \leq x/m} \chi(d)d^{-1/2} \\
 &= \sum_{d < y} \frac{\chi(d)}{d^{1/2}} \left\{ 2\left(\frac{x}{d}\right)^{1/2} + c + O\left(\left(\frac{d}{x}\right)^{1/2}\right) \right\} + O(Dx^{1/2}y^{-1}) \\
 &= 2L(1, \chi)x^{1/2} + O(Dx^{1/2}y^{-1} + x^{-1/2}y) \\
 &= 2L(1, \chi)\sqrt{x} + O(\sqrt{D})
 \end{aligned}$$

for $x \geq D$ by choosing $y = (xD)^{1/2}$, where the implied constant is absolute. Letting $x \rightarrow \infty$ these inequalities imply

$$L(1, \chi) \neq 0. \tag{2.6}$$

For showing (2.6) the class number formula (2.4) is dispensable, but it is a good starting place for estimating the class number $h = h(-D)$ of the imaginary quadratic fields. To this end one needs estimates of $L(1, \chi)$ (clearly (2.5) would give nothing new). By the Riemann Hypothesis for $L(s, \chi)$ it follows that

$$(\log \log D)^{-1} \ll L(1, \chi_D) \ll \log \log D, \tag{2.7}$$

hence the corresponding bounds for the class number

$$\sqrt{D}(\log \log D)^{-1} \ll h(D) \ll \sqrt{D} \log \log D. \tag{2.8}$$

Here the implied constants are absolute, effectively computable.

No chance to prove (2.8) by means available today. The best known upper bound is $L(1, \chi_D) \ll \log D$, which is easy (up to a constant). A lower bound for $L(1, \chi_D)$ is more important and the current knowledge is even less satisfactory. This problem is closely related to the zero-free region for $L(s, \chi_D)$.

At present we know that $L(s, \chi) \neq 0$ for $s = \sigma + it$ in the region

$$\sigma > 1 - \frac{c}{\log D(|t| + 1)}, \tag{2.9}$$

where c is a positive absolute constant, for any character $\chi \pmod{D}$ with at most one exception. The exceptional character is real and the exceptional zero is real and simple. This follows by classical arguments of de la Vallée-Poussin (cf. E. Landau [L1]). Hence the question:

DOES THE EXCEPTIONAL ZERO EXIST?

In this article we shall try to illuminate this matter in bright and dark colors.

Let $\chi \pmod{D}$ be the real primitive character of conductor D and $\beta = \beta_\chi$ be the largest real zero of $L(s, \chi)$. Conjecturally $\beta_\chi = 0, -1$ if $\chi(-1) = 1, -1$, respectively. We say that χ is exceptional if

$$\beta > 1 - \frac{c}{\log D} \quad (2.10)$$

for some positive constant c . One could make this concept more definite by fixing a sufficiently small value of the constant c , however we feel this would only obscure the presentation.

E. Landau [L1] said that H. Hecke knew that if χ was not exceptional, then

$$L(1, \chi) \gg (\log D)^{-1}. \quad (2.11)$$

Remark. In the exceptional case of odd character $\chi = \chi_D$ (which is associated with the imaginary quadratic field $K = \mathbb{Q}(\sqrt{-D})$) there are several, quite precise relations between β_χ , h and $L(1, \chi)$, see [GSc], [G1], [GS].

Landau made a first breakthrough in the exceptional zeros area. Let $\chi \pmod{D}$ and $\chi' \pmod{D'}$ be two distinct real primitive characters and β , β' be real zeros of $L(s, \chi)$, $L(s, \chi')$, respectively. He showed that

$$\min(\beta, \beta') \leq 1 - \frac{b}{\log DD'} \quad (2.12)$$

with some positive, absolute constant b . This shows that the exceptional zeros occur very rarely. For example, calibrating the constant c in (2.10) to $c = b/3$ one can infer from (2.12) that if $\chi \pmod{D}$ is exceptional then the next exceptional one $\chi' \pmod{D'}$ appears no sooner than for some $D' \geq D^2$.

There is a great idea in Landau's arguments which is still exercised in modern works. Generalising slightly we owe to Landau the product L -function (a quadratic lift)

$$\text{Lan}(s, f) = L(s, f) L(s, f \otimes \chi) = \sum_1^\infty a_f(n) n^{-s} \quad (2.13)$$

where

$$L(s, f) = \sum_1^\infty \lambda_f(n) n^{-s} \quad (2.14)$$

can be any natural L -function and $L(s, f \otimes \chi)$ is derived from $L(s, f)$ by twisting (= multiplying) its coefficients $\lambda_f(n)$ with $\chi(n)$. This is particularly interesting for L -functions having finite degree Euler product. Then the prime coefficients of $\text{Lan}(s, f)$ are

$$a_f(p) = \lambda(p) \lambda_f(p) \quad (2.15)$$

where

$$\lambda(p) = 1 + \chi(p). \quad (2.16)$$

The key observation is that if $L(1, \chi)$ is small then the class number h is small and $\chi(p) = 1$ is a rare event (not many primes split in the field K). Therefore $\chi(p) = -1$ and $a_f(p) = 0$ quite often. In other words $\chi(m)$ pretends to be

the Möbius function $\mu(m)$ on squarefree numbers. In this scenario $L(s, f \otimes \chi)$ pretends to be $L(s, f)^{-1}$ up to a small Euler product, and $\text{Lan}(s, f)$ behaves like a constant. This indicates that $L(s, f)$ cannot vanish at s near one, unless $L(s, f \otimes \chi)$ has a pole at $s = 1$ (natural L -functions are regular at $s \neq 1$).

Remark. Landau worked with

$$\zeta(s)L(s, \chi)L(s, \chi')L(s, \chi\chi')$$

which is the product of two $\text{Lan}(s, f)$ for $\zeta(s)$ and $L(s, \chi)$.

Paraphrasing the above observation one may say that if the exceptional zero is very close to $s = 1$ then the other zeros are further away of $s = 1$; not only the zeros of $L(s, \chi)$, but also of any other L -function. This kind of a repelling property of the exceptional zero was nicely exploited in the works of M. Deuring [D] and H. Heilbronn [H] with a remarkable result that

$$h(-D) \rightarrow \infty \text{ as } D \rightarrow \infty. \tag{2.17}$$

Shortly after that, E. Landau [L2] performed a quantitative analysis of the repelling effects and made a cute logical maneuver ending up with the lower bound

$$h(-D) \gg D^{\frac{1}{8}-\varepsilon} \tag{2.18}$$

for any $\varepsilon > 0$, the implied constant depending on ε (the original statement was a little different, but easily equivalent to (2.18)). In the same year and the same journal (the very first volume of Acta Arithmetica of 1936) C. L. Siegel [S] published the still stronger estimate

$$h(-D) \gg D^{\frac{1}{2}-\varepsilon}. \tag{2.19}$$

Note. Siegel was a much broader mathematician than Landau, however in my opinion Landau's ideas pioneered the above developments, so why did Siegel ignore Landau's contribution entirely?

3 How was the class number problem solved?

All three results (2.17), (2.18), (2.19) suffer from the serious defect of being ineffective (the implied constants in the Landau–Siegel estimates are not computable in terms of ε). For that reason one cannot use the results for the determination of all the imaginary quadratic fields with a given fixed class number h . Gauss conjectured that there are exactly nine fields with $h = 1$ (that is to say with unique factorization), the last one for $K = \mathbb{Q}(\sqrt{-163})$. Before the problem was completely solved it was known that there can be at most one more such field.

The Class Number One problem was eventually solved by arithmetical means (complex multiplication and Weber invariants) by K. Heegner [He] and

later re-done independently by H. Stark [St]. A completely different solution was given by A. Baker [B] using transcendental number theory means (linear forms in three logarithms of algebraic numbers). Next it was recognized that the linear forms in two logarithms could do the job, so the 1948 work of A. O. Gelfond and Yu. V. Linnik [GL] was sufficient to resolve the non-existence of the tenth discriminant. H. Stark also settled the class number two problem.

Recently A. Granville and H. Stark [GS] showed a new inequality between the class number $h(-D)$ of $K = \mathbb{Q}(\sqrt{-D})$ and reduced quadratic forms (a, b, c) of discriminant $-D$, that is solutions of the equation

$$-D = b^2 - 4ac \quad (3.1)$$

in integers a, b, c with

$$-a < b \leq a < c, \text{ or } 0 \leq b \leq a = c. \quad (3.2)$$

Note that the reduction condition (3.2) means that the root

$$z = \frac{-b + i\sqrt{D}}{2a} \quad (3.3)$$

is in the standard fundamental domain of the modular group $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. They showed that

$$h(-D) \geq \left(\frac{\pi}{3} + o(1)\right) \frac{\sqrt{D}}{\log D} \sum_{(a,b,c)} a^{-1}. \quad (3.4)$$

Since the principal form with $a = 1$ is always there we get $h(-D) \gg \sqrt{D}/\log D$, and with some extra work one can deduce from (3.4) that $\chi = \chi_D$ is not exceptional. Fine, but the formula (3.4) of Granville–Stark is conditional, they need a uniform abc -conjecture for number fields, specifically for the Hilbert class field which is an extension of K of degree $h(-D)$! In spite of this criticism I strongly recommend this paper for learning a number of beautiful arguments.

A new excitement arose with the work of D. Goldfeld [G2] who succeeded in giving an effective lower bound

$$h(-D) \gg \prod_{p|D} \left(1 - \frac{2\sqrt{p}}{p+1}\right) \log D. \quad (3.5)$$

We shall give a brief sketch how this remarkable bound is derived. But first we point out some historical facts. In principle there is no reason to abandon the repelling property of an exceptional zero; one can still produce an effective result provided such an exceptional zero has a numerical value. OK, but believing in the Grand Riemann Hypothesis one cannot expect to find a real

zero of any natural L -function which would be qualified to play a role of the repellent. A close analysis of Siegel's arguments reveals that any zero $\beta > \frac{1}{2}$ has some power of repelling; although not as strong as the zero near the point $s = 1$, yet sufficient for showing effectively that

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

The only hope along such ideas is to use an L -function which vanishes at the central point $\beta = \frac{1}{2}$, at least this assumption does not contradict the GRH. Hence the first question is: does the central zero have an effect on the class number? In the remarkable paper by J. Friedlander [F] we find the answer: yes it does, and the impact depends on the order of the central zero! The second question is: how to find L -functions which do vanish at the central point? If $L(s, f)$ is self-dual and has the root number -1 , that is the complete function $\Lambda(s, f)$ which includes the local factors at infinite places satisfies the functional equation

$$\Lambda(s, f) = -\Lambda(1 - s, f), \tag{3.6}$$

then, of course, $L(\frac{1}{2}, f) = 0$. Alas, no such function was known until J. V. Armitage [A] gave an example of an L -function of a number field (the Dirichlet L -functions $L(s, \chi)$ cannot vanish at $s = \frac{1}{2}$ by a folk conjecture). After this example Friedlander was able to apply his ideas giving an effective estimate for the class number of relative quadratic extensions. His work anticipated further research by Goldfeld.

A lot more possibilities were offered by elliptic curves. According to the Birch–Swinnerton-Dyer conjecture, the Hasse–Weil L -function of an elliptic curve E/\mathbb{Q} vanishes at the central point to the order equal to the rank of the group of rational points. Goldfeld needed an L -function with central zero of order at least three. It is easy to point out a candidate as it is easy to construct an elliptic curve of rank $g = 3$, but proving that it is modular with the corresponding L -function vanishing to that order is a much harder problem. Ten years after Goldfeld's publication such an L -function was provided by B. Gross and D. Zagier [GZ], making the estimate (3.5) unconditional. Still, to make (3.5) practical (for example for the determination of all the imaginary quadratic fields $K = \mathbb{Q}(\sqrt{-D})$ with the class number $h = 3, 4, 5$, etc.) one needs a numerical value of the implied constant; so J. Oesterlé [O] refined Goldfeld's work and obtained a pretty neat estimate (3.5) with the implied constant $1/55$.

The best one can hope for to obtain along Goldfeld's arguments is $h(-D) \gg (\log D)^{g-2}$ when an L -function with the central zero of multiplicity g is employed. However there are popular problems which require a better effective lower bound for $h(-D)$, such as the

EULER IDONEAL NUMBER PROBLEM. Find all discriminants $-D$ for which the class group of $K = \mathbb{Q}(\sqrt{-D})$ has exactly one class in each genus.

By the genus theory, if $-D$ is an idoneal discriminant then $h(-D) = 2^{\omega(D)-1}$, where $\omega(D)$ is the number of distinct prime divisors of D . Because

$\omega(D)$ can be as large as $\log D / \log \log D$, the problem of Euler calls for an effective lower bound

$$h(-D) \gg D^{c/\log \log D}, \quad \text{with } c > \log 2. \quad (3.7)$$

Of course, Landau's estimate (2.18) tells us that the number of idoneal discriminants is finite, yet we cannot determine all of them.

4 How and why do the central zeros work?

Very briefly we mention the main ideas behind the bound (3.5). There is no particular reason to restrict ourselves to the Hasse–Weil L -functions of elliptic curves, except that they are natural and available sources for multiple central zeros.

Let $f \in S_k(N)$ be a primitive cusp form of weight $k \geq 2$, k -even, and level N , that is a Hecke form on $\Gamma_0(N)$. This has the Fourier expansion

$$f(z) = \sum_1^{\infty} \lambda_f(n) n^{(k-1)/2} e(nz) \quad (4.1)$$

with coefficients $\lambda_f(n)$ which are eigenvalues of Hecke operators T_n for all n . With our normalization the associated L -function

$$L(s, f) = \sum_1^{\infty} \lambda_f(n) n^{-s} \quad (4.2)$$

converges absolutely in $\operatorname{Re} s > 1$ (because of the Ramanujan conjecture $|\lambda_f(n)| \leq \tau(n)$ proved by P. Deligne), it has the Euler product

$$L(s, f) = \prod_p (1 - \lambda_f(p) p^{-s} + \chi_0(p) p^{-2s})^{-1} \quad (4.3)$$

where $\chi_0 \pmod{N}$ is the principal character, and the complete product

$$\Lambda(s, f) = \left(\frac{\sqrt{N}}{2\pi} \right)^s \Gamma\left(s + \frac{k-1}{2}\right) L(s, f) \quad (4.4)$$

(which is entire) satisfies the self-dual functional equation

$$\Lambda(s, f) = w(f) \Lambda(1-s, f). \quad (4.5)$$

Here $w(f) = \pm 1$ is called the root number, or the sign of the functional equation.

Let $\chi = \chi_D$ be the real character (the Kronecker symbol) associated with the imaginary quadratic field $K = \mathbb{Q}(\sqrt{-D})$. For simplicity assume that $(D, N) = 1$. The twisted form

$$f_\chi(z) = \sum_1^\infty \chi(n)\lambda_f(n)n^{(k-1)/2}e(nz) \tag{4.6}$$

is also a primitive form of weight k and level $N_\chi = ND^2$, the L -function

$$L(s, f_\chi) = \sum_1^\infty \chi(n)\lambda_f(n)n^{-s} \tag{4.7}$$

has appropriate Euler product, while the complete product

$$\Lambda(s, f_\chi) = \left(\frac{D\sqrt{N}}{2\pi}\right)^s \Gamma\left(s + \frac{k-1}{2}\right)L(s, f_\chi) \tag{4.8}$$

is entire and satisfies the functional equation

$$\Lambda(s, f_\chi) = w(f_\chi)\Lambda(1-s, f_\chi) \tag{4.9}$$

with the root number

$$w(f_\chi) = \chi(-N)w(f). \tag{4.10}$$

Given f and χ we consider the Landau product

$$L(s) = L(s, f)L(s, f_\chi) = \sum_1^\infty a(n)n^{-s}. \tag{4.11}$$

This is an L -function with Euler product of degree four. The complete product

$$\Lambda(s) = Q^s \Gamma^2\left(s + \frac{k-1}{2}\right)L(s, f)L(s, f_\chi) \tag{4.12}$$

with

$$Q = \frac{DN}{4\pi^2} \tag{4.13}$$

satisfies the functional equation

$$\Lambda(s) = w\Lambda(1-s), \quad w = \chi(-N). \tag{4.14}$$

From here we compute the derivative of order $g \geq 0$ of $\Lambda(s)$ at $s = \frac{1}{2}$ by way of moving the integration in

$$\frac{g!}{2\pi i} \int_{(1)} \Lambda\left(s + \frac{1}{2}\right)s^{-g-1} ds.$$

We obtain

$$Q^{-1/2} \Lambda^{(g)}(1/2) = (1 + (-1)^g w)S \tag{4.15}$$

where

$$S = \sum_1^\infty \frac{a(n)}{\sqrt{n}} V\left(\frac{n}{Q}\right) \tag{4.16}$$

and $V(y)$ is the Mellin transform of $g!\Gamma^2(s + k/2)s^{-g-1}$.

Assuming that $L(s)$ vanishes at $s = \frac{1}{2}$ of order larger than g and that

$$w = (-1)^g \tag{4.17}$$

we get

$$S = 0. \tag{4.18}$$

This is not possible if the class number $h = h(-D)$ is very small. The key point is that many coefficients $a(n)$ vanish so S is well approximated by a product over small primes. Waving hands a bit we can pull out from S a positive factor which takes squares, then we reduce S to a sum which looks like

$$S^{\flat} = \sum_{m < Q}^{\flat} \frac{a(m)}{\sqrt{m}} \left(\log \frac{Q}{m}\right)^g. \tag{4.19}$$

Here the superscript \flat means that the summation is restricted to squarefree numbers. For m squarefree we have

$$a(m) = \lambda(m)\lambda_f(m) \ll \lambda(m)\tau(m).$$

One also shows that

$$\sum_{y < m \leq x}^{\flat} \lambda(m)\tau(m)m^{-1/2} \ll h \left(\frac{1}{y} + \frac{x}{D}\right)^{1/2}. \tag{4.20}$$

Now, another crucial point is that the sum S^{\flat} runs over $m < Q$ with $Q \ll D$ (for N fixed), so (4.20) is extremely sharp in this range, giving

$$S^{\flat} = \sum_{m < y} \frac{a(m)}{\sqrt{m}} \left(\log \frac{Q}{m}\right)^g + O(h) \tag{4.21}$$

with $y = (\log D)^{2g}$. Now assuming that $h \ll (\log D)^{g-2}$ one can approximate the short sum (4.21) essentially by the product

$$(\log D)^g \prod_{p < y} \left(1 + \frac{a(p)}{\sqrt{p}}\right) \tag{4.22}$$

and eventually one draws a contradiction. Of course, in details the arguments are more complicated (cf. [IK]), but their key points look like above. As we mentioned in the transition from S to S^{\flat} a factor taking squares is pulled out, this factor is essentially $L(1, \text{sym}^2 f)$ which is positive. We would also like to point out that the L -functions $L(s, f)$ for automorphic forms f on GL_n with $n > 2$ would not do the job, because the corresponding partial sum S^{\flat} is

longer than D . On the other hand if we could employ $L(s, f)$ with f on GL_1 then S^b is of length \sqrt{D} and arguments similar to the above yield

$$h(-D) \gg D^{1/4} \log D. \tag{4.23}$$

Well, this is a wishful thinking; the GL_1 automorphic L -functions are just the Dirichlet L -functions for real characters, and none of these is expected to vanish at the central point $s = \frac{1}{2}$!

However one can derive the effective bound (4.23) from the more plausible hypothesis that

$$L(1/2, \chi_D) \geq 0. \tag{4.24}$$

Indeed we have (cf. (22.60) of [IK])

$$\zeta_K(1/2) = \zeta(1/2)L(1/2, \chi_D) = \frac{1}{2} \sum_{(a,b,c)} a^{-1/2} \log \left(\frac{\sqrt{D}}{2a} \right) + O(hD^{-1/4}),$$

where (a, b, c) runs over reduced forms, so $1 \leq a \leq \sqrt{D/3}$. Since $\zeta(\frac{1}{2})L(\frac{1}{2}, \chi_D) \leq 0$ this yields

$$h \gg \sum \left(\frac{\sqrt{D}}{a} \right)^{1/4} \log \frac{\sqrt{D}}{a} \tag{4.25}$$

giving (4.23) from just one term $a = 1$ (the principal form).

Because of the spectacular consequence (4.23) of the assumption (4.24), it seems that the latter is out of reach by the current technology. Of course, the GRH implies (4.24), but it also implies (2.8), so taking this road is pointless.

Closing this section we state an effective lower bound for $h(-D)$ which can be rigorously established by following the above guidelines.

Theorem 4.1 *Suppose that $L(s)$ given by (4.11) vanishes at $s = \frac{1}{2}$ to order $m \geq 3$. Then*

$$h(-D) \gg \theta(D) (\log D)^{g-1} \tag{4.26}$$

where $g = m - 1$ or $g = m - 2$ according to the parity condition $(-1)^g = w$. Here $\theta(D)$ is a mild factor, precisely

$$\theta(D) = \prod_{p|D} \left(1 + \frac{1}{p} \right)^{-3} \left(1 + \frac{2\sqrt{p}}{p+1} \right)^{-1}$$

while the implied constant depends only on the cusp form $f \in S_k(N)$ and is effectively computable.

Remark. For the purpose of proving Goldfeld's lower bound (3.5) Gross-Zagier delivered the following elliptic curve

$$E : -139y^2 = x^3 + 10x^2 - 20x + 8, \tag{4.27}$$

which is modular of conductor $N = 37 \cdot 139^2$ and rank $r = 3$.

5 What if the GRH holds except for real zeros?

If you are not afraid of confrontation with complex zeros of the $L(s)$, then be more productive working with the logarithmic derivative $L'(s)/L(s)$ rather than with the partial sums of $L^{(g)}(s)$. P. Sarnak and A. Zaharescu [SZ] have taken this route to improve Goldfeld's bound (3.5) significantly.

Theorem 5.1 *Let $L(s, f)$ vanish at $s = \frac{1}{2}$ of order ≥ 3 . Let $-D$ be a fundamental discriminant with $\chi_D(N) = 1$. Suppose $L(s) = L(s, f)L(s, f \otimes \chi_D)$ has all its zeros either on the critical line $\operatorname{Re} s = \frac{1}{2}$ or on the real line $\operatorname{Im} s = 0$. Then*

$$h(-D) \gg D^{\frac{1}{6}-\delta} \quad (5.1)$$

for any $\delta > 0$, the implied constant depending effectively on δ and f .

Theorem 5.1 is our variation on the work of Sarnak–Zaharescu. Their arguments are somewhat different and their bound (5.1) has the exponent $1/10$ in place of $1/6$. Moreover they worked only with the L -functions associated with the elliptic curve (4.27). But they also established a few other interesting results, some of which are ineffective.

To explain what is behind the proof of Theorem 5.1 we appeal to the so called “explicit formula”

$$\begin{aligned} \sum_{L(\rho)=0} \phi\left(\frac{\gamma}{2\pi} \log R\right) &= 2\widehat{\phi}(0) \frac{\log D}{\log R} + \phi(0) \\ &\quad - 2 \sum_{p \mid D} \lambda_f(p) \frac{\lambda(p)}{\sqrt{p}} \widehat{\phi}\left(\frac{\log p}{\log R}\right) \frac{\log p}{\log R} + O\left(\frac{\log \log D}{\log R}\right). \end{aligned} \quad (5.2)$$

This is derived by integrating $L'(s)/L(s)$ against a test function ϕ , using the functional equation (4.14) and Cauchy's residue theorem. Here $\phi(x)$ is an even function whose Fourier transform $\widehat{\phi}(y)$ is continuous and compactly supported, so $\phi(x)$ is entire, $R \geq 2$ is a parameter to be chosen later, and the implied constant depends only on the cusp form $f \in S_k(N)$ and the test function ϕ . To be fair we must admit that the exact explicit formula contains terms over prime powers which we put into the error term; this involves an estimate for the logarithmic derivative of $L(s, \operatorname{sym}^2 f)$ which follows by using the standard zero-free region near $s = 1$.

Suppose $\widehat{\phi}(y)$ is supported in $[-1, 1]$. Thinking of $h = h(-D)$ being small, say $h \leq D^{\frac{1}{6}-\delta}$, we can estimate the sum over primes in (5.2) by

$$\sum_{p < R} \frac{\lambda(p)}{\sqrt{p}} \frac{\log p}{\log R} \ll (\log D)^{-\delta} \quad (5.3)$$

for any R with $hD^{1/2} \leq R \leq h^{-2}D^{1-3\delta}$. Later we shall choose $R = D^{\frac{2}{3}-\delta}$. This is not an easy bound; it shows that $\lambda(p) = 1 + \chi(p) = 0$ very often. Hence the explicit formula (5.2) reduces to

$$\sum_{L(\rho)=0} \phi\left(\frac{\gamma}{2\pi} \log R\right) = 2\widehat{\phi}(0) \frac{\log D}{\log R} + \phi(0) + O((\log D)^{-\delta}). \quad (5.4)$$

Now we are ready to play with (5.4), that is to say we want to pick up a test function $\phi(x)$ for which (5.4) is false. Already at first glance (5.4) is an improbable expression for most reasonable $\phi(x)$, because for what reason the zeros $\rho = \beta + i\gamma$ of $L(s)$ can be so regularly distributed to generate the functional $\phi \rightarrow 3\widehat{\phi}(0) + \phi(0)$? As we do not know much about the spacing of zeros, our chance for contradiction goes by estimations. More chance if we can make every term $\phi((\gamma/2\pi) \log R)$ non-negative, so we can pick up the largest one and drop the others. For this reason we assume that all the zeros lay on two lines, $\beta = \frac{1}{2}$ or $\gamma = 0$. Specifically we choose the Fourier pair (as in Sarnak–Zaharescu)

$$\phi(x) = \left(\frac{\sin \pi x}{\pi x}\right)^2, \quad \widehat{\phi}(y) = \max(1 - |y|, 0) \quad (5.5)$$

giving

$$m \leq 2 \frac{\log D}{\log R} + 1 + O((\log D)^{-\delta}), \quad (5.6)$$

where m is the multiplicity of the zero of $L(s)$ at $s = \frac{1}{2}$. For $R = D^{\frac{2}{3}-\delta}$ this implies $m < 4$, that is $m \leq 3$. However we assumed that $L(s, f)$ has zero at $s = \frac{1}{2}$ of order ≥ 3 , and we also know that $L(s) = L(s, f)L(s, f \otimes \chi)$ has the root number $w = \chi(-N) = \chi(-1) = -1$, so m is even, $m \geq 4$. This contradiction completes the proof of Theorem 5.1.

Remarks. The final blow in the proof of Theorem 5.1 is powered by the positivity arguments. This is an excellent example of the strength of the real-variable harmonic analysis when coupled with the positivity ideas. The positivity arguments are hard to implement to complex domains, so the hypothesis that all zeros are on specific lines is critical.

6 Subnormal gaps between critical zeros

A simple central zero of an L -function yields no effect on the class number, still if it has large order then it does. But what about the complex zeros on the critical line, so to speak the critical zeros, which appear in abundance? More hopefully one should ask if some clustering of the critical zeros can be as effective as the high order central zero. This possibility was contemplated in the literature long before the central zero effects. Indeed the fundamental work of H. L. Montgomery [M] on the pair correlation of zeros was motivated by the class number problems. In a joint paper Montgomery–Weinberger [MW] used zeros of a fixed real Dirichlet L -function which are close to the central point $s = \frac{1}{2}$, by means of which they were able to perform quite strong computations for the imaginary quadratic fields $K = \mathbb{Q}(\sqrt{-D})$ with the

class number $h = h(-D) = 1, 2$ (see [MW] for precise results and for some other relevant claims).

Recently B. Conrey and H. Iwaniec [CI] considered the Hecke L -function

$$L(s; \psi) = \sum_{\mathfrak{a}} \psi(\mathfrak{a})(N\mathfrak{a})^{-s} \quad (6.1)$$

associated with the imaginary quadratic field $K = \mathbb{Q}(\sqrt{-D})$. Here \mathfrak{a} runs over the non-zero integral ideals of K and $\psi \in \widehat{\mathcal{C}}(K)$ is a character of the class group. Although $L(s; \psi)$ does not factor as the Landau product (2.13) (unless ψ is a genus character), it possesses the same crucial feature, namely the lacunarity of the coefficients

$$\lambda_\psi(n) = \sum_{N\mathfrak{a}=n} \psi(\mathfrak{a}) \quad (6.2)$$

if the class number $h(-D)$ is ridiculously small.

Have in mind that the corresponding theta series

$$\theta(z; \psi) = \sum_0^\infty \lambda_\psi(n) e(nz),$$

with $\lambda_\psi(0) = h/2$ for the trivial character and $\lambda_\psi(0) = 0$ otherwise, is a modular form of weight $k = 1$, level D and Nebentypus χ_D (it is a cusp form if ψ is a complex character). This yields the functional equation (self-dual)

$$\Lambda(s; \psi) = \left(\frac{\sqrt{D}}{2\pi}\right)^s \Gamma(s) L(s; \psi) = \Lambda(1-s; \psi). \quad (6.3)$$

By contour integration one can show that the number of zeros of $L(s; \psi)$ in the rectangle $s = \sigma + it$ with $0 \leq \sigma \leq 1$, $0 < t \leq T$ satisfies

$$N(T; \psi) = \frac{T}{\pi} \log \frac{T\sqrt{D}}{2\pi e} + O(\log DT). \quad (6.4)$$

Hence one can say (assuming GRH) that the average gap between consecutive zeros $\rho = \frac{1}{2} + i\gamma$ and $\rho' = \frac{1}{2} + i\gamma'$ is about $\pi/\log \gamma$.

We have shown in [CI] that if the gap is a little smaller than the average for sufficiently many pairs of zeros on the critical line (no GRH is required) then $h(-D) \gg \sqrt{D}(\log D)^{-A}$ for some constant $A > 0$. This result may not appeal to everybody, because our L -function $L(s; \psi)$ is intimately related with the field $K = \mathbb{Q}(\sqrt{-D})$, so are its zeros. Well, we can draw a more impressive statement from the zeta function of K (the case of the trivial class group character)

$$\zeta_K(s) = \zeta(s)L(s, \chi_D). \quad (6.5)$$

Since we do not need all the zeros, we choose only those of $\zeta(s)$ which apparently have nothing in common with the character χ_D .

Theorem 6.1 *Let $\rho = \frac{1}{2} + i\gamma$ denote the zeros of $\zeta(s)$ on the critical line and $\rho' = \frac{1}{2} + i\gamma'$ denote the nearest zero to ρ on the critical line ($\rho' = \rho$ if ρ is a multiple zero). Suppose*

$$\#\left\{\rho; 0 < \gamma < T, |\gamma - \gamma'| \leq \frac{\pi}{\log \gamma} \left(1 - \frac{1}{\sqrt{\log \gamma}}\right)\right\} \gg T(\log T)^{4/5} \quad (6.6)$$

for any $T \geq 2005$. Then

$$h(-D) \gg \sqrt{D}(\log D)^{-90} \quad (6.7)$$

where the implied constant is effectively computable.

Have in mind that each of $\zeta(s)$, $L(s, \chi_D)$ has asymptotically half the number of zeros of $\zeta_K(s)$, so that relative to $\zeta(s)$ in Theorem 6.1 we are counting the gaps which are a little smaller than the half of the average gap. Our condition (6.6) is quite realistic, because the Pair Correlation Conjecture of Montgomery asserts that the zeros of $\zeta(s)$ are not equidistributed. In fact the PCC implies that

$$|\gamma - \gamma'| < \frac{2\pi\vartheta}{\log \gamma} \quad (6.8)$$

with any $\vartheta > 0$, for a positive proportion of zeros. The best unconditional estimate (6.8) is known with $\vartheta = 0.68$ by Montgomery–Odlyzko [MO], $\vartheta = 0.5171$ by Conrey–Ghosh–Gonek [CGG] and $\vartheta = 0.5169$ by Conrey–Iwaniec (unpublished). For the effective bound (6.7) we need (6.8) with some $\vartheta < \frac{1}{2}$.

Remark. At the meeting in Seattle of August 1996 D. R. Heath-Brown gave a talk “Small Class Number and the Pair Correlation of Zeros” in which he showed how the assumption of the class number being small distorts the Pair Correlation Conjecture of Montgomery. His and our arguments have similar roots.

The main principles of the proof of Theorem 6.1 can be seen quickly (but of course, the details are formidable) from the “approximate functional equation”

$$L(s; \psi) = \sum_{n \leq t\sqrt{D}} \lambda_\psi(n) n^{-s} + X(s) \sum_{n \leq t\sqrt{D}} \lambda_\psi(n) n^{s-1} + \dots$$

on the line $s = \frac{1}{2} + it$. Because $\lambda_\psi(n)$ are lacunary (assuming the class number is relatively small) the two partial sums can be shortened substantially, so the variation of $L(s; \psi)$ in t is mostly controlled by the gamma factor

$$X\left(\frac{1}{2} + it\right) = \left(\frac{2\pi e}{t\sqrt{D}}\right)^{2it} \{1 + O(1/t)\}$$

(a “root number” in the t -aspect). In other words the “infinite place” leads the spin while the “finite places” are too weak and too few to disturb. Therefore in this illusory scenario the zeros of $L(s; \psi)$ should follow the equidistribution law, but we postulated otherwise, hence the contradiction.

From the above discussion one may also get an idea why the PCC predicts a density function for differences between zeros to be other than constant; the reason might be that the “finite places” generate the periodicities n^{it} with distinct frequencies as n varies around t .

Another interesting lesson one can draw from the above situation is that the very popular perception that the zeros of very different L -functions operate in their own independent ways, that they do not see each other so cannot conspire, is not wise. This idealistic view may appeal to math philosophers, but when the tools of analytic number theory break the sky we find a more fascinating and complex structure.

7 Fifty percent is not enough!

... for winning in a democracy, neither for ruling out the exceptional character. In recent investigations we (see Iwaniec–Sarnak [IS]) took an opposite direction for attacking the problem of the exceptional character. Rather than using the central zeros of L -functions as repellents, we need families of L -functions whose central values are positive, not very small.

For this presentation we take the set $\mathcal{H}_k(N)$ of cusp forms f of weight $k \geq 2$, k -even which are primitive on the group $\Gamma_0(N)$ (i.e. which are eigenfunctions of all the Hecke operators T_n , $n = 1, 2, 3, \dots$). The basic properties of the associated L -functions are (4.1)–(4.14). The Hilbert space structure of the linear space $S_k(N)$ plays a role in our arguments (the Petersson formula brings Kloosterman sums which are our tools), and the transition from spectral to arithmetical normalizations is achieved by the factors

$$\omega_f = \zeta_N(2) L(1, \text{sym}^2 f)^{-1}, \quad (7.1)$$

where $\zeta_N(s)$ denotes the zeta function with the local factors at primes $p \mid N$ omitted, and $L(s, \text{sym}^2 f)$ is the L -function associated with the symmetric square representation of f . These are mild factors since

$$(\log kN)^{-2} \ll L(1, \text{sym}^2 f) \ll (\log kN)^2. \quad (7.2)$$

The upper bound is an easy consequence of the Ramanujan conjecture (proved by P. Deligne), while the lower bound is essentially saying that $L(s, \text{sym}^2 f)$ has no exceptional zero which is now known as fact due to Hoffstein–Lockhart [HL]. Actually we do not make use of (7.2), because the factors ω_f are kept present in our averagings over the family $\mathcal{H}_k(N)$. We have

$$\sum_{f \in \mathcal{H}_k(N)} \omega_f X_f \sim N \quad (7.3)$$

for each of the vectors $X_f = 1$, $X_f = L(\frac{1}{2}, f)$, $X_f = L(\frac{1}{2}, f_\chi)$, and

$$\sum_{f \in \mathcal{H}_k(N)} \omega_f L(\tfrac{1}{2}, f) L(\tfrac{1}{2}, f_\chi) \sim NL(1, \chi) \tag{7.4}$$

as $N \rightarrow \infty$ over squarefree numbers, uniformly for $D \leq N^\delta$ with $\delta > 0$ a small fixed constant (recall that D is the conductor of $\chi = \chi_D$).

The great attraction of the formula (7.4) is the fact that the central values $L(\tfrac{1}{2}, f)$, $L(\tfrac{1}{2}, f_\chi)$ are known unconditionally to be non-negative. Of course, one can deduce this from the GRH, yet we can do it without (see Waldspurger [Wa], Kohnen–Zagier [KZ], Katok–Sarnak [KS], Guo [Gu]). The non-negativity of $L(\tfrac{1}{2}, f)$ has much to do with f being a GL_2 form. Recall that this property for self-dual GL_1 forms (the Dirichlet real characters) would have immediate consequences for the class number (see (4.23)), unfortunately it is not provable without recourse to the GRH. It is not easy to show that $L(\tfrac{1}{2}, f) \geq 0$, $L(\tfrac{1}{2}, f_\chi) \geq 0$ for any cusp form $f \in \mathcal{H}_k(N)$, but these estimates are not actually deep. One may get an idea why the central values are non-negative by considering a simple example of f whose coefficients are

$$a(n) = \sum_{ad=n} \chi(a)\bar{\chi}(d)(a/d)^{ir}.$$

In this case

$$L(\tfrac{1}{2}, f) = |L(\tfrac{1}{2} + ir, \chi)|^2 \geq 0.$$

One may express the central values of automorphic L -functions by sums of squares in a more profound fashion. For the CM forms F. Rodríguez Villegas [R-V] takes squares of a theta-series. A cute proof of $L(\tfrac{1}{2}, f) \geq 0$ follows as a by-product in the recent investigations of W. Luo–P. Sarnak [LS] in quantum chaos.

Another important feature of the asymptotic formula (7.4) is its “purity”, that is to say the absence of lower order terms involving the derivative $L'(1, \chi)$. Therefore if $L(1, \chi)$ is very small then almost all the products $L(\tfrac{1}{2}, f)L(\tfrac{1}{2}, f_\chi)$ are very small. Before speculating further let us restrict the summation (7.4) to forms for which the root number of $L(s, f)L(s, f_\chi)$ is one,

$$w = w_f w_{f_\chi} = w_f w_f \chi(-N) = \chi(-N) = 1$$

(because the L -functions with root number -1 vanish at the central point trivially by the functional equation).

One can establish that a lot of $L(\tfrac{1}{2}, f)$ and $L(\tfrac{1}{2}, f_\chi)$ are not very small, say

$$L(\tfrac{1}{2}, f) \geq (\log N)^{-2} \tag{7.5}$$

$$L(\tfrac{1}{2}, f_\chi) \geq (\log N)^{-2}, \tag{7.6}$$

using the classical idea of averaging of mollified values. If the two sets of f 's for which both (7.5) and (7.6) hold had a large intersection (positive percentage) we could conclude from (7.3), (7.4) by the non-negativity that

$$L(1, \chi) \gg (\log D)^{-4}. \quad (7.7)$$

We did succeed to show that (7.5) holds for at least 50% of forms $f \in \mathcal{H}_k(N)$ with $\varepsilon_f = 1$, and that (7.6) holds for at least 50% of forms $f \in \mathcal{H}_k(N)$ with $\varepsilon_{f_\chi} = 1$. These results are just too short to ensure a significantly large intersection.

It is hard to believe that a character $\chi \pmod{D}$ can be so vicious to divide (by twisting) any respectful family of L -functions into two equal size classes (almost), giving all the power to one class and nothing for the other class. Yet, we cannot destroy such feature by present tools. Having (7.6) for 50% forms it suffices to get (7.5) for slightly more than 50%. The latter task seems to be quite promising, because the character issue is irrelevant! Not really! Actually we undertook the task with stronger tools offered by averaging over the level N . Consequently we were able to attach to $L(\frac{1}{2}, f)$ a mollifying factor longer than N (which puts us beyond diagonal) leading to (7.5) for more than 50% of the forms f with $w_f = 1$. But (7.6) is not useful for every N , here we need the root number condition $\chi(-N) = 1$. Ironically, if one installs this condition to averaging over the level, then the off-diagonal terms are badly affected, and the excess over 50% disappears! We are convinced there is a magic conspiracy out there which prevents us from cracking the existence of the exceptional character along our lines.

Perhaps one should build a comprehensive theory which explains all the peculiar loops in which we are often trapped when venturing beyond the diagonal path.

8 Exceptional primes

An easy way of handling problems is to avoid them. Better yet one may find that the obstacle which is hard to eliminate can be exploited to reach the goal in other ways. The case of the exceptional character is a spectacular example in this regard. We shall present a few applications of the exceptional character for producing primes in tide areas where even the GRH fails to work. Having tasted the results one may only wish that the exceptional character is a real thing, not an illusion which researchers of several generations tried to kill.

The good reason for liking the real exceptional characters $\chi(m)$ is that they pretend to be the Möbius function $\mu(m)$ at almost all squarefree integers m . In the same time the characters are periodic functions, so one can apply a Fourier analysis in place of zeros of L -functions. One needs a quantitative measure of how closely $\chi(m)$ approximates to $\mu(m)$. To this end consider

$$\Delta(z, x) = \sum_{z < n \leq x} \lambda(n) n^{-1}. \quad (8.1)$$

Recall that $\lambda = 1 * \chi$, and χ is the real character of conductor D , not necessarily exceptional. We have

$$\Delta(z, x) = L(1, \chi) [\log x + O(\log z)] \tag{8.2}$$

if $x > z \geq D^2$. Hence $\lambda(n)$ vanishes very often if $L(1, \chi)$ is very small, and $\chi(p) = -1$ very often. We can see this phenomenon better from estimates for

$$\delta(z, x) = \sum_{z \leq p < x} \lambda(p)p^{-1}. \tag{8.3}$$

By the inequality $\delta(z, x)\Delta(1, z) \leq \Delta(z, xz)$ we get by (8.2)

$$\delta(z, x)\Delta(1, z) \leq L(1, \chi) [\log x + O(\log z)]. \tag{8.4}$$

Applying the trivial bound $\Delta(1, z) \geq 1$ we get

$$\delta(z, x) \leq L(1, \chi) [\log x + O(\log z)] \tag{8.5}$$

if $x > z \geq D^2$. One can also estimate $\delta(z, x)$ in terms of any real zero, say β , of $L(s, \chi)$. Indeed we have

$$\begin{aligned} \Delta(1, z) &> z^{\beta-1} \sum_{1 \leq n < z} \lambda(n)n^{-\beta} \left(1 - \frac{n}{z}\right) \\ &= \frac{1}{2\pi i} \int_{(1)} \zeta(s + \beta)L(s + \beta, \chi)z^{s+\beta-1} \frac{ds}{s(s+1)} \\ &= L(1, \chi)(1 - \beta)^{-1}(2 - \beta)^{-1} + O(q^{1/4}z^{-1/2}) \\ &> L(1, \chi)(1 - \beta)^{-1} \end{aligned}$$

by moving the integration to the line $\text{Re } s = \frac{1}{2} - \beta$, provided $x > z \geq D^2$. Inserting this bound to (8.4) we obtain

$$\delta(z, x) < (1 - \beta) [\log x + O(\log z)]. \tag{8.6}$$

The implied constants in (8.5), (8.6) are absolute. These inequalities show that $\delta(z, x)$ is very small so $\chi(p) = -1$ for almost all p in the range $D^2 \leq p \leq D^A$, A a constant, provided χ is exceptional.

Now how this observation can be used for applications to prime numbers? We start from the zeta-function of $K = \mathbb{Q}(\sqrt{D^*})$

$$\zeta_K(s) = \sum_1^\infty \lambda(n)n^{-s} = \zeta(s)L(s, \chi). \tag{8.7}$$

Define the multiplicative function $\nu(m)$ by

$$\frac{1}{\zeta_K(s)} = \sum_1^\infty \nu(m)m^{-s}. \tag{8.8}$$

Note that $|\nu(m)| \leq \lambda(m)$ and

$$\nu(m) = \mu(m)\lambda(m) \text{ if } m \text{ is squarefree.} \quad (8.9)$$

If $\lambda(n)$ is lacunary (i.e. $\lambda(n)$ vanishes very often), then so is $\mu(m)$. Next, writing $\zeta'(s)/\zeta(s) = L(s, \chi)\zeta'(s)/\zeta_K(s)$ we find

$$A(n) = \sum_{klm=n} \chi(k)(\log l)\nu(m). \quad (8.10)$$

We also introduce the function

$$\lambda'(d) = \sum_{kl=d} \chi(k) \log l, \quad (8.11)$$

so (8.10) becomes

$$A(n) = \sum_{dm=n} \lambda'(d)\nu(m). \quad (8.12)$$

One can easily view $\lambda'(d)$ as a divisor-like function, because \log is smooth and slowly increasing while χ is periodic with a relatively small period. Moreover, if χ is exceptional then $\nu(m)$ is lacunary, so it contributes to (8.12) very little only at small m . Therefore one can see (8.12) as an approximation to the von Mangoldt function by a divisor-like function. By means of this formula in many interesting applications one can accomplish results for primes as strong as for the divisor function.

The situation described above is a little bit oversimplified. In practice a serious difficulty occurs with handling the lacunary part of (8.12), say

$$A_*(n) = \sum_{\substack{dm=n \\ m > D^2}} \lambda'(d)\nu(m), \quad (8.13)$$

especially when $A_*(n)$ is applied against a sparse sequence $\mathcal{A} = (a_n)$. We estimate (8.13) by

$$|A_*(n)| \leq \tau(n)(\log n) \sum_{\substack{m|n \\ m > D^2}} \lambda(m).$$

We deal with $\tau(n)\log n$ crudely by special devices which allow us to ignore this factor, so we are left essentially with

$$A_\infty(n) = \sum_{\substack{m|n \\ m > D^2}} \lambda(m). \quad (8.14)$$

From the above partitions we arrive at (essentially)

$$\sum_{n \leq x} a_n A(n) = \sum_{\substack{dm \leq x \\ m \leq D^2}} a_{dm} \lambda'(d)\nu(m) + O\left((\log x)^{2005} \sum_{n \leq x} a_n A_\infty(n)\right).$$

In estimating the sum

$$\sum_{n \leq x} a_n \Lambda_\infty(n) \tag{8.15}$$

we cannot forget that in our mind $\lambda(m)$ is lacunary. If $\mathcal{A} = (a_n)$ is not sparse then one can disconnect a_n from $\Lambda_\infty(n)$ by Cauchy's inequality and estimate the resulting sums separately and quite easily. A great challenge appears for very sparse sequences. We open the convolution $\lambda = 1 * \chi$ in (8.14) and consider $\Lambda_\infty(n)$ to be like the divisor function $\tau_3(n)$ rather than like $\tau(n)$ in the main term. After having opened the $\lambda(m)$ our analysis of the error term must be asymptotically accurate, because at the end we must observe a crucial cancellation which reflects the lacunarity of $\lambda(m)$.

Having said this we conclude that the existence of the exceptional character creates a useful substitute for $\Lambda(n)$ in terms of divisor-like functions of degree three. Therefore various methods of analytic number theory which are capable of showing an asymptotic formula for

$$\sum_{n \leq x} a_n \tau_3(n) \tag{8.16}$$

are likely to be modified to yield an asymptotic formula for

$$\sum_{n \leq x} a_n \Lambda(n). \tag{8.17}$$

In a series of papers J. Friedlander and H. Iwaniec [FI2], [FI3], [FI4] realized these ideas for a few very sparse sequences. For example we got the following formula for primes in a short interval

$$\psi(x) - \psi(x - y) = y \{ 1 + O(L(1, \chi)(\log x)^{r^r}) \} \tag{8.18}$$

for $x \geq y \geq x^{39/79}$, $x \geq D^r$ where $r = 18,290$ and the implied constant is absolute. The result is unconditional, but it is useful only under special conditions, such as

$$L(1, \chi) \leq (\log D)^{-1-r^r}$$

and $D^r \leq x \leq D^{2r}$. Note that $\frac{39}{79} < \frac{1}{2}$, so the interval in (8.18) can be very short. The Riemann Hypothesis does not work for intervals shorter than $[x - \sqrt{x}, x]$.

Similar ideas (however more precise with respect to the powers of logarithms) were used earlier by D. R. Heath-Brown [H-B1] with an impressive conclusion that if there is an infinite sequence of exceptional zeros, then there are infinitely many twin primes.

9 The least prime in an arithmetic progression

9.1 Introduction

In the previous sections we have been trying either to eliminate the exceptional character from the surface of the Earth, or to employ it for producing

impressive, yet illusory results. However one can play both tunes in a complementary fashion to end up with completely unconditional results and effective ones, unlike the Landau–Siegel type. The celebrated work of Yu. V. Linnik [L] on the least prime in an arithmetic progression is a true masterpiece of this kind.

Let $p_{\min}(q, a)$ denote the first prime $p \equiv a \pmod{q}$. Linnik proved that

$$p_{\min}(q, a) \ll q^L \tag{9.1}$$

for any $q > 1$, $(a, q) = 1$, where L and the implied constant are absolute and effectively computable. The GRH gives (9.1) with any $L > 2$, while the best known result with $L = 5.5$ is due to D. R. Heath-Brown [H-B3]. The best possible (9.1) should be with any $L > 1$.

Using arguments similar to these in the proof of (8.18) we [FI2] showed that

$$\psi(x; q, a) = \frac{\psi(x)}{\varphi(q)} \{1 + O(L(1, \chi)(\log x)^{r^r})\} \tag{9.2}$$

for $(a, q) = 1$, $D \nmid q$ and any $x \geq \max\{q^{462/233}, D^r\}$ with $r = 554, 401$, where the implied constant is absolute. If $\chi \pmod{D}$ is exceptional in the sense that

$$L(1, \chi) \leq (\log D)^{-1-r^r} \tag{9.3}$$

then (9.2) implies (9.1) with $L = 2 - \frac{1}{59}$ for q in the range

$$D^r \leq q \leq \exp(L(1, \chi)^{-r/(r^r+1)}). \tag{9.4}$$

Earlier Heath-Brown [H-B2] also succeeded in bringing the Linnik constant L close to 2, but not below 2, under the assumption of the existence of exceptional characters (our condition (9.3) is a bit stronger).

As we described in the previous section at some point our arguments depend on the sum (8.16), specifically for the sequence $\mathcal{A} = (a_n)$ of the characteristic function of the progression $n \equiv a \pmod{q}$. By no means this is an easy sum if $x < q^2$; just mention we had to modify the result of [FI1], which is proved by an appeal to the Riemann Hypothesis for varieties.

Back to Linnik's bound (9.1) there are several interesting points to say about its original proof in regard to the theory of Dirichlet L -functions. All the proofs up to now use essentially the following three principles:

P1: THE ZERO-FREE REGION (2.9).

P2: THE LOG-FREE ZERO DENSITY ESTIMATE:

$$\sum_{\chi \pmod{q}} N_{\chi}(\alpha, T) \leq a(qT)^{b(1-\alpha)} \tag{9.5}$$

where $N_{\chi}(\alpha, T)$ denotes the number of zeros $\rho = \beta + i\gamma$ of $L(s, \chi)$ with $\beta \geq \alpha$, $|\gamma| < T$ for $\frac{1}{2} \leq \alpha \leq 1$, $T \geq 1$, and a, b are positive absolute constants.

P3: THE EXCEPTIONAL ZERO REPULSION:

If $\beta > 1 - c/\log q$ is a real zero of $L(s, \chi)$ for a real character $\chi \pmod{q}$, then there is no other zero of any L -functions with characters modulo q in the region

$$\sigma \geq 1 - \frac{d |\log(1 - \beta) \log q|}{\log q(|t| + 1)} \tag{9.6}$$

where c, d are positive, small, absolute constants.

The first principle is classical, the other two are due to Linnik. The principles P2, P3 set the theory of Dirichlet L -functions at the most profound level. Yes, they will be obsolete soon after the GRH is proved, but for the time being (perhaps a very long time) these principles are treasures on their own right.

Having paid tribute to P2, P3, we are going to show the Linnik bound (9.1) without using these principles. Our arguments (a joint work with J. Friedlander) reveal a new potential of sieve methods. First we treat the case when the exceptional character is available, because the arguments are quick and require almost nothing from the theory of L -functions, not even P1 nor the Prime Number Theorem. The second case, with no exceptional character existing, is somewhat longer. In this case we do use P1, however by a hard work one could dispense with it. The point is that using sieve we are not aiming at an asymptotic formula for primes $p \equiv a \pmod{q}$, so the “parity barrier” of linear sieve is not a problem, the primes can be produced along the elementary lines à la Chebyshev. Anyway, there is no reason to work hard without P1, when the derivation of the zero-free region (2.9) is by today standards very easy.

We replace P2 by a much simpler result:

Proposition 9.1 *Let $\rho = \beta + i\gamma$ run over the zeros of $L(s, \chi)$ for a character $\chi \pmod{q}$. Put*

$$A(t) = \sum_{\rho} (1 + (1 - \beta) \log q)^{-1} (1 + |t - \gamma| \log q)^{-2}. \tag{9.7}$$

For any real t we have

$$A(t) \leq \frac{3}{2} + \frac{\log(|t| + c)}{2 \log q} \tag{9.8}$$

where c is a positive absolute constant.

Remark. A bound for $A(t)$ with $|t| \leq q$ by any fixed number suffices for our applications, because we are not going to give a numerical value of the constant L in (9.1).

Proof of (9.8). For any $s = \sigma + it$ with $1 < \sigma \leq 2$ we have

$$-\operatorname{Re} \frac{L'}{L}(s, \chi) = \frac{1}{2} \log q|s| - \sum_{\rho} \operatorname{Re} \frac{1}{s - \rho} + O(1),$$

$$\left| \frac{L'}{L}(s, \chi) \right| \leq -\frac{\zeta'}{\zeta}(\sigma) = \frac{1}{\sigma - 1} + O(1),$$

$$\operatorname{Re} \frac{1}{s - \rho} = \frac{\sigma - \beta}{(\sigma - \beta)^2 + (t - \gamma)^2} \geq \frac{1}{\sigma - \beta} \left(1 + \frac{|t - \gamma|}{\sigma - 1} \right)^{-2}.$$

Hence

$$\sum_{\rho} \frac{\sigma - 1}{\sigma - \beta} \left(1 + \frac{|t - \gamma|}{\sigma - 1} \right)^{-2} \leq 1 + \frac{1}{2}(\sigma - 1)(\log q|s| + O(1)).$$

For $\sigma = 1 + 1/\log q$ the left side is equal to $A(t)$ giving the bound (9.8).

9.2 The case with an exceptional character

Let $\chi \pmod{q}$ be a real, non-principal character. We do not really assume that χ is exceptional, so we end up with unconditional results, which will be useful in the final conclusion only when χ is the exceptional character.

We shall apply sieve to the sequence $\mathcal{A} = (\lambda(n)a_n)$, where $\lambda = 1 * \chi$ and a_n is the characteristic function of the arithmetic progression $n \equiv a \pmod{q}$ with $(a, q) = 1$. Clearly we must assume that

$$\chi(a) = 1, \tag{9.9}$$

or else there is nothing but zero in \mathcal{A} . We need to evaluate the sums of type

$$A_d(x) = \sum_{\substack{n \leq x \\ n \equiv 0 \pmod{d}}} \lambda(n)a_n = \sum_{\substack{n \leq x/d \\ n \equiv a\bar{d} \pmod{q}}} \lambda(dn)$$

for $(d, q) = 1$. Think of $\lambda(n)$ as the Hecke eigenvalues of the Eisenstein series of weight one and the central character (Nebentypus) χ . Hence $\lambda(n)$ is multiplicative,

$$\lambda(dn) = \sum_{\delta | (d, n)} \mu(\delta) \chi(\delta) \lambda(d/\delta) \lambda(n/\delta).$$

This yields

$$A_d(x) = \sum_{\delta | d} \mu(\delta) \chi(\delta) \lambda(d/\delta) A(x/\delta d; q, a\bar{\delta d})$$

where

$$A(y; q, \alpha) = \sum_{\substack{m \leq y \\ m \equiv \alpha \pmod{q}}} \lambda(m).$$

Here we write

$$\lambda(m) = \sum_{kl=m} \chi(k) = (1 + \chi(m)) \sum_{\substack{kl=m \\ k < l}} \chi(k) + \chi(\sqrt{m})$$

for $(m, q) = 1$, where the last term $\chi(\sqrt{m})$ vanishes, unless \sqrt{m} is an integer (a traditional convention), that is if $m = k^2$. This gives

$$A(y; q, \alpha) = (1 + \chi(\alpha))A^*(y; q, \alpha) + O(\sqrt{y})$$

where

$$\begin{aligned} A^*(y; q, \alpha) &= \sum_{\substack{kl \leq y, k < l \\ kl \equiv \alpha \pmod{q}}} \chi(k) = \frac{1}{q} \sum_{k < \sqrt{y}} \chi(k) \left(\frac{y}{k} - k \right) + O(\sqrt{y}) \\ &= L(1, \chi) y q^{-1} + O(\sqrt{y}). \end{aligned}$$

Hence

$$A(y; q, \alpha) = (1 + \chi(\alpha))L(1, \chi) y q^{-1} + O(\sqrt{y})$$

where the implied constant is absolute. Here we have $\alpha = a\bar{\delta}d$, $\chi(\alpha) = \chi(a)\chi(d/\delta) = \chi(d/\delta) = 1$, or else $\chi(d/\delta) = 0$. Hence we obtain

$$A_d(x) = 2L(1, \chi) \frac{\nu(d)x}{dq} + O(\tau_3(d)\sqrt{x/d}) \tag{9.10}$$

where

$$\nu(d) = \sum_{\delta|d} \mu(\delta) \frac{\chi(\delta)}{\delta} \lambda\left(\frac{d}{\delta}\right).$$

This is multiplicative with

$$\nu(p) = 1 + \chi(p) \left(1 - \frac{1}{p}\right). \tag{9.11}$$

We write the approximation (9.10) in the familiar sieve format

$$A_d(x) = g(d)X + r_d(x) \tag{9.12}$$

where $g(d) = \nu(d)/d$ stands for the sifting density function,

$$X = 2L(1, \chi) x q^{-1} \tag{9.13}$$

and $r_d(x)$ is the error term, $r_d(x) \ll \tau_3(d)\sqrt{x/d}$. Hence the remainder term of level y satisfies

$$R(x, y) = \sum_{d < y} |r_d(x)| \ll \sqrt{xy} (\log y)^2,$$

where the implied constant is absolute.

We seek primes so we wish to estimate $S(\mathcal{A}, \sqrt{x})$. Under normal conditions the task would be beyond the capability of a linear sieve. However we think that $\chi(p) = -1$ very often for the exceptional character, so the density function at such primes is very small, $g(p) = p^{-2}$. In this scenario we have a sieve problem of small dimension, and the Fundamental Lemma of sieve theory does the job,

$$S(\mathcal{A}, z) = X V(z) \{1 + O(e^{-s})\} + O(\sqrt{xy} (\log y)^2) \quad (9.14)$$

where $s = \log y / \log z \geq 2$ and

$$V(z) = \prod_{\substack{p < z \\ p \nmid q}} (1 - g(p)) = \prod_{\substack{p < z \\ p \nmid q}} \left(1 - \frac{1}{p}\right) \left(1 - \frac{\chi(p)}{p}\right). \quad (9.15)$$

We do not need the full strength of e^{-s} in (9.14), a weaker term s^{-1} suffices. Choosing

$$y = \frac{x}{q^3 (\log x)^8}, \quad x \geq q^8, \quad (9.16)$$

we see that the error term in (9.14) is negligible giving

$$S(\mathcal{A}, z) = X V(z) \left\{1 + O\left(\frac{\log z}{\log x}\right)\right\}. \quad (9.17)$$

From $S(\mathcal{A}, z)$ we go to $S(\mathcal{A}, \sqrt{x})$ by Buchstab's formula

$$S(\mathcal{A}, \sqrt{x}) = S(\mathcal{A}, z) - \sum_{z \leq p < \sqrt{x}} S(\mathcal{A}_p, p).$$

For every $z \leq p < \sqrt{x}$ we estimate $S(\mathcal{A}_p, p)$ by an upper-bound sieve of level y/p getting

$$S(\mathcal{A}_p, p) \ll g(p) V(p) X + \sqrt{xy/p} (\log y)^2.$$

Adding these estimates we arrive at

$$S(\mathcal{A}, \sqrt{x}) = X V(z) \left\{1 + O\left(\frac{\log z}{\log x} + \delta(z, x)\right)\right\} \quad (9.18)$$

where $\delta(z, x)$ is defined by (8.3) and was estimated twice in (8.5) and (8.6), in the range $x > z \geq q^2$. Hence we conclude (still unconditional result)

Lemma 9.2 *Let $\chi \pmod{q}$ be a real, non-principal character and β be any real zero of $L(s, \chi)$. Suppose $\chi(a) = 1$. Then for $x \geq q^8$ we have*

$$\pi(x; q, a) = 2L(1, \chi) V(q^2) \frac{x}{q} \left\{1 + O\left(\frac{\log q}{\log x} + (1 - \beta) \log x\right)\right\} \quad (9.19)$$

where the implied constant is absolute. The factor $1 - \beta$ can be replaced by $L(1, \chi)$.

Corollary 9.3 *Let the condition of Lemma 9.2 be satisfied. Then for x in the segment*

$$q^A \leq x \leq e^{1/A(1-\beta)}, \tag{9.20}$$

where A is any large constant, $A \geq 8$, we have

$$\pi(x; q, a) > \frac{x L^*(1, \chi)}{\varphi(q) \log q} \tag{9.21}$$

where

$$L^*(1, \chi) = L(1, \chi) \prod_{p < q^2} (1 - \chi(p)). \tag{9.22}$$

Of course, the segment (9.20) is not void only if $1 - \beta \leq A^{-2}(\log q)^{-1}$, which with a large constant A means that χ is an exceptional character. Assuming that this is the case we get the Linnik bound (9.1) with $L = A$.

9.3 A parity-preserving sieve inequality

Next we are going to apply sieve to the sequence $\mathcal{A} = (a_n)$ which is the characteristic function of the arithmetic progression $n \equiv a \pmod{q}$. Our goal is to estimate

$$S(\mathcal{A}, z) = \sum_{\substack{n \leq x \\ (n, P(z))=1}} a_n$$

for $z = \sqrt{x}$. For $(d, q) = 1$ we have

$$A_d(x) = \frac{x}{dq} + O(1)$$

so we have a problem of linear sieve. For the level of distribution of \mathcal{A} we take

$$y = \frac{x}{q} (\log x)^{-4/3}. \tag{9.23}$$

Recall that the lower-bound linear sieve works only in the range $z \leq \sqrt{y}$, which is not a problem because

$$S(\mathcal{A}, \sqrt{x}) = S(\mathcal{A}, \sqrt{y}) + O\left(\frac{x \log q}{\varphi(q) (\log x)^2}\right) \tag{9.24}$$

by the Brun-Titchmarsh estimate.

Next the linear sieve gives (see [I], Rutgers notes)

$$S(\mathcal{A}, \sqrt{y}) = S^-(\mathcal{A}, \sqrt{y}) + \sum_{n \text{ even}} S_n(\mathcal{A}, \sqrt{y}).$$

Usually one discards all terms $S_n(\mathcal{A}, \sqrt{y})$ getting the lower bound

$$S(\mathcal{A}, z) \geq S^-(\mathcal{A}, z) = X V^-(y, z) + O(y),$$

where $X = xq^{-1}$ and $O(y)$ is the bound for the remainder term. The main term equals

$$V^-(y, z) = \{f(s) + O((\log y)^{-1/3})\}V(z)$$

with $s = \log y / \log z$. For $z = \sqrt{y}$ we get $f(2) = 0$, so the sum $S^-(\mathcal{A}, \sqrt{y})$ is negligible, and we must keep the terms $S_n(\mathcal{A}, \sqrt{y})$. We get

$$S(\mathcal{A}, \sqrt{y}) \geq \sum_{n \text{ even}} S_n(\mathcal{A}, \sqrt{y}) + O(y). \tag{9.25}$$

We only exploit the term for $n = 4$, which is

$$S_4(\mathcal{A}, \sqrt{y}) = \sum'_{p_4 < p_3 < p_2 < p_1 < \sqrt{y}} \dots \sum' S(\mathcal{A}_{p_1 p_2 p_3 p_4}, p_4)$$

where the summation is restricted by the conditions $p_1 p_2^3 < y$, $p_1 p_2 p_3 p_4^3 \geq y$. Dropping more terms we deduce that

$$S(\mathcal{A}, \sqrt{y}) \geq \frac{1}{24} \sum''_{\substack{p_4 p_3 p_2 p_1 p \leq x \\ p_4 p_3 p_2 p_1 p \equiv a \pmod{q}}} \dots \sum'' \sum 1 + O(y) \tag{9.26}$$

where the superscript '' indicates that the prime variables p_r run independently over the segment

$$y^{1/6} < p_r < y^{1/5}, \quad r = 1, 2, 3, 4. \tag{9.27}$$

Remarks. In (9.26) we have estimated a sum over primes (essentially) by a sum over products of five primes. The other sums $S_n(\mathcal{A}, \sqrt{y})$ with n even (which we discarded) run essentially over products of $n + 1$ primes (if y is close to x), so the parity is odd throughout all terms of (9.25). Therefore the formula (9.25) does not break the parity which is the barrier for getting primes within the traditional axioms of sieve theory.

We have chosen to work with products of five primes $p_4 p_3 p_2 p_1 p$ rather than three for technical advantage (products of larger fixed odd number of primes would also be fine).

Before applying characters to detect the congruence $p_4 p_3 p_2 p_1 p \equiv a \pmod{q}$ we exploit the positivity, and partition the sum $\sum'' \dots \sum'' \sum$ in (9.26) into suitable blocks so the separation of variables will not be an issue later. It is essential that we can do it at this point without much loss, because the forthcoming arguments will be so delicate that anything like partial summation will inflict unreparable damage (certainly losing a logarithmic factor will kill

the arguments). To this end we fix a smooth function $f(u)$ supported on $[\frac{1}{2}, 1]$ with $0 \leq f(u) \leq 1$, and put

$$\psi_{\mathcal{X}}(x; q, a) = \sum_{\substack{p p_1 p_2 p_3 p_4 \equiv a \pmod{q} \\ x_r < p_r < 2x_r}} \dots \sum f(p p_1 p_2 x_3 x_4 / x) \log p$$

where $\mathcal{X} = [x_1, x_2, x_3, x_4]$ runs over the vectors of dyadic partition points of the segment $[y^{1/6}, y^{1/5}]$. Then (9.26) yields

$$S(\mathcal{A}, \sqrt{y}) \geq \frac{1}{24 \log x} \sum_{\mathcal{X}} \psi_{\mathcal{X}}(x; q, a) + O(y). \tag{9.28}$$

Notice that we did not partition p and we included the variables p_1, p_2 together with p in the argument of the smoothing function f , while p_3, p_4 are excluded from f . These seemingly technical devices will play nicely in relevant character sums.

Let $\psi_{\mathcal{X}}(x)$ denote the corresponding sum over $p p_1 p_2 p_3 p_4$ with the congruence condition dropped, i.e. $\psi_{\mathcal{X}}(x) = \psi_{\mathcal{X}}(x; 1, 1)$, so we have

$$\psi_{\mathcal{X}}(x) \asymp x / (\log x_1)(\log x_2)(\log x_3)(\log x_4), \tag{9.29}$$

where the implied constant depends only on f . Our goal is to show that

$$\psi_{\mathcal{X}}(x; q, a) \asymp \frac{\psi_{\mathcal{X}}(x)}{\varphi(q)} \tag{9.30}$$

for all relevant \mathcal{X} subject to some conditions on x, q, a to be specified later. Hence we derive by (9.24), (9.28) that

$$\pi(x; q, a) \gg \frac{\pi(x)}{\varphi(q)} \tag{9.31}$$

subject to the same conditions on x, q, a .

9.4 Estimation of $\psi_{\mathcal{X}}(x; q, a)$

Applying the orthogonality of characters we write

$$\psi_{\mathcal{X}}(x; q, a) = \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \psi_{\mathcal{X}}(x, \chi)$$

where

$$\psi_{\mathcal{X}}(x, \chi) = \sum_{\substack{p p_1 p_2 p_3 p_4 \\ x_r < p_r < 2x_r}} \dots \sum \chi(p p_1 p_2 p_3 p_4) f(p p_1 p_2 x_3 x_4 / x) \log p. \tag{9.32}$$

Denote $\mathcal{W} = [x_1, x_2]$, $w = x / x_3 x_4$ and

$$\psi_{\mathcal{W}}(w, \chi) = \sum_{\substack{p_1 p_1 p_2 \\ x_r < p_r < 2x_r}} \chi(p_1 p_1 p_2) f(p_1 p_1 p_2 / w) \log p, \quad (9.33)$$

so (9.32) becomes

$$\psi_{\mathcal{X}}(x, \chi) = \psi_{\mathcal{W}}(w, \chi) \left(\sum_{x_3 < p_3 < 2x_3} \chi(p_3) \right) \left(\sum_{x_4 < p_4 < 2x_4} \chi(p_4) \right). \quad (9.34)$$

The principal character $\chi_0 \pmod{q}$ gives the main term. We also put aside the contribution of the real character, say $\chi_1 \pmod{q}$, because it will require a special treatment when χ_1 is exceptional. We get

$$\psi_{\mathcal{X}}(x; q, a) = \frac{1}{\varphi(q)} \{ \psi_{\mathcal{X}}(x) + \chi_1(a) \psi_{\mathcal{X}}(x, \chi_1) + \Delta_{\mathcal{X}}(x; q, a) \} \quad (9.35)$$

where $\Delta_{\mathcal{X}}(x; q, a)$ denotes the contribution of all the characters $\chi \neq \chi_0, \chi_1$.

We estimate $\Delta_{\mathcal{X}}(x; q, a)$ in the following fashion which resembles the circle method for ternary additive problems (we have here a multiplicative analog):

$$|\Delta_{\mathcal{X}}(x; q, a)| \leq \max_{\chi \neq \chi_0, \chi_1} |\psi_{\mathcal{W}}(w, \chi)| \left(\sum_{\chi} \left| \sum_{p_3} \chi(p_3) \right|^2 \right)^{1/2} \left(\sum_{\chi} \left| \sum_{p_4} \chi(p_4) \right|^2 \right)^{1/2}.$$

It is easy to see that for any $X \geq q^2$

$$\sum_{\chi \pmod{q}} \left| \sum_{X < p < 2X} \chi(p) \right|^2 \ll \left(\frac{X}{\log X} \right)^2$$

where the implied constant is absolute. To this end square out and estimate the resulting sum over primes $p_1 \equiv p_2 \pmod{q}$ by the Brun-Titchmarsh theorem.

Now we need a modest, but non-trivial estimate of $\psi_{\mathcal{W}}(w, \chi)$ for every $\chi \neq \chi_0, \chi_1$ (it is like asking for a non-trivial estimate of the corresponding exponential sum in the circle method at every point of the minor arc). In Section 9.6 we prove that for $\chi \neq \chi_0, \chi_1$

$$\psi_{\mathcal{W}}(w, \chi) \ll \left(\frac{\log q}{\log w} + \frac{1}{\log q} \right) \frac{w}{(\log x_1)(\log x_2)} \quad (9.36)$$

provided $x_1, x_2 \geq q$ and $x_1 x_2 q^2 \leq w^{3/4}$. Hence we get

$$\Delta_{\mathcal{X}}(x; q, a) \ll \left(\frac{\log q}{\log x} + \frac{1}{\log q} \right) \psi_{\mathcal{X}}(x) \quad (9.37)$$

where the implied constant is absolute (we assume that $x \geq q^8$).

For estimating $\psi_{\mathcal{X}}(x, \chi_1)$ with the real character χ_1 we have two options. First in Section 9.6 we prove that

$$\psi_{\mathcal{W}}(w, \chi_1) \ll \left(\frac{1}{(1 - \beta_1) \log w} + \frac{\log q}{\log w} + \frac{1}{\log q} \right) \frac{w}{(\log x_1)(\log x_2)}, \quad (9.38)$$

where β_1 is the largest real zero of $L(s, \chi_1)$. Hence by trivial estimations of sums over p_3, p_4 in (9.34) we get

$$\psi_{\mathcal{X}}(x, \chi_1) \ll \left(\frac{1}{(1 - \beta_1) \log x} + \frac{\log q}{\log x} + \frac{1}{\log q} \right) \psi_{\mathcal{X}}(x). \quad (9.39)$$

The second option is to replace every $\chi(p), \chi(p_r), r = 1, 2, 3, 4$ in (9.32) by -1 getting

$$\psi_{\mathcal{X}}(x, \chi_1) = -\{1 + O(\delta(z, x))\} \psi_{\mathcal{X}}(x)$$

where $\delta(z, x)$ is defined by (8.3). Using (8.6) we get

$$\psi_{\mathcal{X}}(x, \chi_1) = -\{1 + O((1 - \beta_1) \log x)\} \psi_{\mathcal{X}}(x). \quad (9.40)$$

9.5 Conclusion

We are now ready to derive Linnik's bound (9.1) from the assorted results in Sections 9.2, 9.3, 9.4.

Suppose $\chi_1 \pmod{q}$ is a non-principal real character such that $L(s, \chi_1)$ has a real zero β_1 with

$$(1 - \beta_1) \log q \leq A^{-2}, \quad (9.41)$$

where A is a large constant, $A \geq 8$. If $\chi_1(a) = 1$ then (9.21) yields (9.1) with $L = A$. If $\chi_1(a) = -1$ then (9.35), (9.40), (9.37) yield

$$\psi_{\mathcal{X}}(x; q, a) = \frac{2}{\varphi(q)} \psi_{\mathcal{X}}(x) \{1 + O(1/A)\}$$

for $q^A \leq x \leq e^{1/A(1-\beta_1)}$. Hence we get (9.31) which yields (9.1) with $L = A$.

Now we can assume that the largest real zero β_1 of $L(s, \chi_1)$ does not satisfy (9.41). Then we get by (9.35), (9.39), (9.37) that

$$\begin{aligned} \psi_{\mathcal{X}}(x; q, a) &= \frac{1}{\varphi(q)} \psi_{\mathcal{X}}(x) \left\{ 1 + O\left(A^2 \frac{\log q}{\log x} + \frac{1}{\log q} \right) \right\} \\ &\gg \frac{\psi_{\mathcal{X}}(x)}{\varphi(q)} \end{aligned}$$

if $x \geq q^{A^2 B}$, where B is a large constant. Hence we get (9.31) for every $(a, q) = 1$, which yields (9.1) with $L = A^2 B$.

9.6 Appendix. Character sums over triple-primes

In this section we give a non-trivial estimate for the character sum $\psi_{\mathcal{W}}(w, \chi)$ defined by (9.33).

Proposition 9.4 *Let $\chi \pmod{q}$ be a non-trivial character. Put*

$$\delta_{\chi} = \min_{\rho} \{1 - \beta; |\gamma| \leq \log q\} \quad (9.42)$$

where $\rho = \beta + i\gamma$ denote zeros of $L(s, \chi)$. For $x_1, x_2 \geq q$ and $x_1 x_2 q^2 \leq w^{3/4}$ we have

$$\psi_{\mathcal{W}}(w, \chi) \ll \left\{ \frac{1}{\delta_{\chi} \log w} + \frac{1}{\log q} \right\} \frac{w}{(\log x_1)(\log x_2)} \quad (9.43)$$

where the implied constant is absolute.

Clearly Proposition 9.4 and the classical zero-free region (2.9) imply (9.36) and (9.38).

Estimating trivially one gets $\psi_{\mathcal{W}}(w, \chi) \ll w/(\log x_1)(\log x_2)$, so $\{\dots\}$ is the saving factor (if w, q are large).

The proof of Proposition 9.4 does not require the zero-free region, although at some point we use the Prime Number Theorem in the form

$$\psi(y) = y + O(y(\log y)^{-4}) \quad (9.44)$$

which helps to simplify the arguments. We start by the “explicit formula”

$$\sum_n \chi(n) \Lambda(n) f(n/w) = - \sum_{\beta \geq 1/2} \widehat{f}(\rho) w^{\rho} + O(\sqrt{w} (\log q)^2),$$

where $\widehat{f}(s)$ is the Mellin transform of $f(u)$, $\rho = \beta + i\gamma$ run over the zeros of $L(s, \chi)$ and the implied constant is absolute. This gives

$$\begin{aligned} \psi_{\mathcal{W}}(w, \chi) = - \sum_{\beta \geq 1/2} \widehat{f}(\rho) w^{\rho} & \left(\sum_{p_1} \chi(p_1) p_1^{-\rho} \right) \left(\sum_{p_2} \chi(p_2) p_2^{-\rho} \right) \\ & + O(\sqrt{w} (\log q)^2). \end{aligned}$$

Note that the error term absorbs the contribution of prime powers p^2, p^3, \dots which are missing in $\psi_{\mathcal{W}}(w, \chi)$. We have

$$\widehat{f}(\rho) \ll |\rho|^{-3}. \quad (9.45)$$

Hence

$$\sum_{|\gamma| > T} |\widehat{f}(\rho)| \ll T^{-2} \log q. \quad (9.46)$$

We choose $T = \log q$ and estimate the term $\psi_{\mathcal{W}}(w, \chi)$ with $|\gamma| > T$ trivially, getting

$$\psi_{\mathcal{W}}(w, \chi) = - \sum^{\rho} \widehat{f}(\rho) \left(\frac{w}{x_1 x_2} \right)^{\rho} \left(\sum_{p_1} \chi(p_1) (x_1/p_1)^{\rho} \right) \left(\sum_{p_2} \chi(p_2) (x_2/p_2)^{\rho} \right) + O(w/(\log x_1)(\log x_2) \log q).$$

Here \sum^{ρ} denotes summation of the zeros $\rho = \beta + i\gamma$ of $L(s, \chi)$ restricted by $\beta \geq \frac{1}{2}$, $|\gamma| \leq \log q$. We have

$$\left| \left(\frac{w}{x_1 x_2} \right)^{\rho} \right| \leq \frac{w}{x_1 x_2} \left(\frac{x_1 x_2 q^2}{w} \right)^{\delta_x} q^{2(\beta-1)},$$

$$\left(\frac{x_1 x_2 q^2}{w} \right)^{\delta_x} \leq w^{-\delta_x/4} \leq 4/\delta_x \log w.$$

Hence

$$\psi_{\mathcal{W}}(w, \chi) \ll T_{\chi}(x_1, x_2) w/x_1 x_2 \delta_x \log w + w/(\log x_1)(\log x_2) \log q,$$

where

$$T_{\chi}(x_1, x_2) = \sum^{\rho} q^{2(\beta-1)} \left| \sum_{x_1 < p_1 < 2x_1} \chi(p_1) (x_1/p_1)^{\rho} \right| \left| \sum_{x_2 < p_2 < 2x_2} \chi(p_2) (x_2/p_2)^{\rho} \right|. \quad (9.47)$$

Note that we ignored the factor (9.45) because it does not help, the problem occurs with bounded zeros.

To complete the proof of (9.43) it remains to show that

$$T_{\chi}(x_1, x_2) \ll x_1 x_2 / (\log x_1)(\log x_2) \quad (9.48)$$

where the implied constant is absolute. If we knew that

$$\sum^{\rho} q^{2(\beta-1)} \ll 1, \quad (9.49)$$

then (9.48) would quickly follow by trivial estimation of the sums over primes p_1, p_2 . The bound (9.49) is true, it is a kind of log-free density bound for the zeros of $L(s, \chi)$ of height $|\gamma| \leq \log q$. However we avoid (9.49) (whose proof would be quite long) by gaining a bit from cancellation in the sums over p_1, p_2 . When the variation of $\rho = \beta + i\gamma$ with respect to γ exceeds $(\log q)^{-1}$ we do have a change in the argument of $(x_r/p_r)^{\rho}$ as p_r varies in $x_r < p_r < 2x_r$, $\log x_r \gg \log q$. This observation should explain why we did not want to separate p from p_1, p_2 in the smoothing function $f(p p_1 p_2/w)$. Moreover it is worth mentioning that for this purpose we use two prime variables p_1, p_2 rather than one, because we can apply the duality principle.

Lemma 9.5 *For $x \geq q$ and any complex numbers c_p we have*

$$\sum_p^\rho q^{2(\beta-1)} \left| \sum_{x < p < 2x} c_p(x/p)^\rho \right|^2 \ll \frac{x}{\log x} \sum_p |c_p|^2 \quad (9.50)$$

where the implied constant is absolute.

Clearly (9.48) follows from (9.50) by Cauchy's inequality.

For the proof of Lemma 9.5 it suffices to show that for any complex numbers a_ρ

$$\sum_{x < p < 2x} \frac{\log p}{p} \left| \sum_p^\rho a_\rho(x/p)^\rho q^{\beta-1} \right|^2 \ll \sum_p^\rho |a_\rho|^2. \quad (9.51)$$

First we smooth the outer summation by introducing a factor $h(p/x)$, then we square out and execute the summation over p getting

$$\sum_p \frac{\log p}{p} h(p/x)(x/p)^{\rho_1 + \bar{\rho}_2} \ll (1 + |\gamma_1 - \gamma_2| \log x)^{-2}.$$

This follows by partial summation using (9.44) and that the Fourier transform of $h(u)$ satisfies $\widehat{h}(v) \ll (1 + |v|)^{-2}$. Hence the left side of (9.51) is estimated by

$$\sum_p^{\rho_1} \sum_p^{\rho_2} |a_{\rho_1} a_{\rho_2}| (1 + |\gamma_1 - \gamma_2| \log q)^{-2} q^{\beta_1 + \beta_2 - 2} \leq \sum_p^\rho A(\gamma) |a_\rho|^2$$

where $A(t)$ is defined and estimated in Proposition 9.1. This proves (9.51), hence (9.50) by the duality, and finally (9.43).

References

- [A] J. V. Armitage, *Zeta functions with a zero at $s = \frac{1}{2}$* , Invent. Math. 15 (1972), 199-205.
- [B] A. Baker, *Linear forms in the logarithms of algebraic numbers*, Mathematika 13 (1966), 204-216.
- [CGG] J. B. Conrey, A. Ghosh and S. M. Gonek, *A note on gaps between zeros of the zeta function*, Bull. London Math. Soc. 16 (1984), 421-424.
- [CI] B. Conrey and H. Iwaniec, *Spacing of zeros of Hecke L-functions and the class number problem*, Acta Arith. 103 (2002), 259-312.
- [D] M. Deuring, *Imaginär-quadratische Zahlkörper mit der Klassenzahl (1)*, Math. Z. 37 (1933), 405-415.
- [F] J. B. Friedlander, *On the class numbers of certain quadratic extensions*, Acta Arith. 28 (1976), 391-393.
- [FI1] J. B. Friedlander and H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem*, Ann. Math. (2) 121 (1985), 319-350.
- [FI2] J. B. Friedlander and H. Iwaniec, *Exceptional characters and prime numbers in arithmetic progressions*, Int. Math. Res. Notices 37 (2003), 2033-2050.
- [FI3] J. B. Friedlander and H. Iwaniec, *Exceptional characters and prime numbers in short intervals*, Selecta Math. 10 (2004), 61-69.

- [FI4] J. B. Friedlander and H. Iwaniec, *The illusory sieve*, preprint.
- [GL] A. O. Gelfond and Yu. V. Linnik, *On Thue's method and the effectiveness problem in quadratic fields*, Dokl. Akad. Nauk SSSR 61 (1948), 773-776 (in Russian).
- [G1] D. Goldfeld, *An asymptotic formula relating the Siegel zero and the class number of quadratic fields*, Ann. Scuola Norm. Sup. Pisa (4) 2 (1975), 611-615.
- [G2] D. Goldfeld, *The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer*, Ann. Scuola Norm. Sup. Pisa (4) 3 (1976), 623-663.
- [GSc] D. Goldfeld and A. Schinzel, *On Siegel's zero*, Ann. Scuola Norm. Sup. Pisa (4) 2 (1975), 571-583.
- [GS] A. Granville and H. M. Stark, *ABC implies no "Siegel zeros" for L-functions of characters with negative discriminant*, Invent. Math. 139 (2000), 509-523.
- [GZ] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. 84 (1986), 225-320.
- [Gu] J. Guo, *On the positivity of the central critical values of automorphic L-functions for $GL(2)$* , Duke Math. J. 83 (1996), 157-190.
- [H-B1] D. R. Heath-Brown, *Prime twins and Siegel zeros*, Proc. London Math. Soc. (3) 47 (1983), 193-224.
- [H-B2] D. R. Heath-Brown, *Siegel zeros and the least prime in an arithmetic progression*, Quart. J. Math. Oxford (2) 41 (1990), no. 164, 405-418.
- [H-B3] D. R. Heath-Brown, *Zero-free regions for Dirichlet L-functions, and the least prime in an arithmetic progression*, Proc. London Math. Soc. (3) 64 (1992), 265-338.
- [He] K. Heegner, *Diophantische Analysis und Modulfunktionen*, Math. Z. 56 (1952), 227-253.
- [H] H. Heilbronn, *On the class-number in imaginary quadratic fields*, Quart. J. Math. Oxford 5 (1934), 150-160.
- [HL] J. Hoffstein and P. Lockhart, *Coefficients of Maass forms and the Siegel zero*, Ann. Math. (2) 140 (1994), 161-181.
- [I] H. Iwaniec, *Sieve Methods*, Graduate Course Notes, Rutgers, 1996.
- [IK] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Amer. Math. Soc. Colloquium Publications, vol. 53, 2004.
- [IS] H. Iwaniec and P. Sarnak, *The non-vanishing of central values of automorphic L-functions and Landau-Siegel zeros*, Israel J. Math. 120 (2000), 155-177.
- [KS] S. Katok and P. Sarnak, *Heegner points, cycles and Maass forms*, Israel J. Math. 84 (1993), 193-227.
- [KZ] W. Kohnen and D. Zagier, *Values of L-series of modular forms at the center of the critical strip*, Invent. Math. 64 (1981), 175-198.
- [L1] E. Landau, *Über die Klassenzahl imaginär-quadratischer Zahlkörper*, Gött. Nachr. (1918), 285-295.
- [L2] E. Landau, *Bemerkungen zum Heilbronnschen Satz*, Acta Arith. 1 (1936), 1-18.
- [L] Yu. V. Linnik, *On the least prime in an arithmetic progression, I. The basic theorem; II. The Deuring-Heilbronn's phenomenon*, Math. Sb. 15 (1944), 139-178 and 347-368.
- [LS] W. Luo and P. Sarnak, *Quantum variance for Hecke eigenforms*, Ann. Sci. Ecole Norm. Sup. (4) 37 (2004), 769-799.

- [M] H. L. Montgomery, *The pair correlation of zeros of the zeta function*, Proc. Sympos. Pure Math. 24, Amer. Math. Soc. 1973, 181-193.
- [MO] H. L. Montgomery and A. M. Odlyzko, *Gaps between zeros of the zeta function*, Topics in classical number theory (Budapest 1981), Colloq. Math. Soc. János Bolyai 34, North-Holland, Amsterdam 1984, 1079-1106.
- [MW] H. L. Montgomery and P. J. Weinberger, *Notes on small class numbers*, Acta Arith. 24 (1974), 529-542.
- [O] J. Oesterlé, *Nombres de classes des corps quadratiques imaginaires*, Sémin. N. Bourbaki (1983-84), exp. 631, Astérisque no. 121-122 (1985), 309-323.
- [R-V] F. Rodríguez Villegas, *Square root formulas for central values of Hecke L -series, II*, Duke Math. J. 72 (1993), 431-440.
- [SZ] P. Sarnak and A. Zaharescu, *Some remarks on Landau-Siegel zeros*, Duke Math. J. 111 (2002), 495-507.
- [S] C. L. Siegel, *Über die Classenzahl quadratischer Zahlkörper*, Acta Arith. 1 (1936), 83-86.
- [St] H. M. Stark, *A complete determination of the complex quadratic fields of class-number one*, Michigan Math. J. 14 (1967), 1-27.
- [Wa] J.-L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entier*, J. Math. Pures Appl. (9) 60 (1981), 375-484.