Contents

Preface

This monograph presents the results of a computation of over 175 million consecutive zeros of the Riemann zeta function near zero number 10^{20}, as well as of several other large sets of high zeros, including some near zero number 2×10^{20}. These zeros lie about 10^8 times higher than previously calculated large sets of zeros, and their computation was made possible by a fast algorithm invented by A. Schönhage and the author. Although the present implementation of this algorithm is not entirely rigorous due to incomplete control of roundoff errors, it appears to be highly accurate as well as fast, and the results indicate that all the computed zeros satisfy the Riemann Hypothesis. Various statistical studies of these zeros are presented. Some of them provide numerical evidence about conjectures that go even beyond the Riemann Hypothesis, and relate the distribution of zeros of the zeta function to that of eigenvalues of random matrices studied extensively in physics. Other studies compare the observed behavior of the zeta function to known asymptotic estimates. The computations described in this book were carried out on a Cray X-MP supercomputer.

Chapter 1

Introduction

The 10^{20}-th zero of the Riemann zeta function equals

$$1/2 + i\ 1520244011592074 7268.6290299\ldots\ .$$

It and a few of its nearest neighbors are shown in Table 1.1. All told, almost 176 million zeros near the 10^{20}-th zero were computed, as wee as over 101 million zeros near zero number 2×10^{20}. These zeros lie almost 10^8 times higher than any other large sets of zeros that had been computed before. This monograph reports statistics of these and some other high zeros and describes the algorithms that made these calculations possible

The Riemann Hypothesis (RH) has been subjected to a series of numerical investigations, starting with unpublished ones by Riemann. (See [Ed, Od3] for a history of these computations.) The latest result is that the RH is true for the first 1.5×10^9 zeros (i.e., all zeros up to height $\preceq 5 \times 10^8$) [LRW2]. This computation required about 1500 hours on modern supercomputers (primarily the Cyber 205). It only separated the zeros, and did not produce accurate values for them. The reason for not obtaining values of zeros was that the investigations in this case were very concerned with establishing the validity of the RH, and for that purpose, as was explained earlier, it is only necessary to separate the zeros of $Z(t)$. Several other large sets of zeros (four sets of roughly 10^5 zeros in each case, starting with zeros number 10^{10}, 10^{11}, 2×10^{11}, and 10^{12}) have also been computed accurately [Od2]. Those

computations took several tens of hours on Cray-1 and Cray X-MP supercomputers, and produced values of the zeros that are accurate to within about 10^{-8}. (The 10^{12}-th zero equals $1/2 + i\gamma$, with $\gamma \approx 2.7 \times 10^{11}$.) The purpose of those calculations of zeros was partially to check the validity of the RH, but the primary goal (and the reason for obtaining accurate values of the zeros) was to obtain data about the distribution of spacings between zeros of the zeta function so as to compare them to some recent conjectures. These conjectures, which are described briefly in Section 2 and which go substantially beyond the RH, originate in the Montgomery pair correlation conjecture, and relate the behavior of the zeta function zeros to eigenvalues of random hermitian matrices that are used to model energy levels in many-particle systems in physics, and to the quantum chaos theories. Agreement between these conjectures and computed values of zeros might be taken as providing some support for the Hilbert and Pólya conjecture that the RH is true because the zeros of the zeta function correspond to eigenvalues of some positive operator. While all these conjectures are highly speculative, it seemed worthwhile to test them numerically. As it turned out, the agreement between conjectures and empirical data was excellent in most cases. A few features of the data that were initially unexpected were in the end explained by relating the behavior of the zeros to that of the primes. However, there were some features of the data that were slightly counterintuitive (such as a slight excess of small spacings between consecutive zeros over that predicted by the random matrix theories), and so it seemed desirable to obtain data about even higher zeros of the zeta function.

All large-scale computations of zeros of the zeta function in the last fifty-odd years (as well as Riemann's own unpublished computations [Ed, Od3, Sie1]) relied on the Riemann-Siegel formula [Ed, Sie1, Gab], which requires roughly $t^{1/2}$ operations to compute $\zeta(1/2 + it)$ for t a large positive real number. Recently a much more efficient algorithm for computing the zeta function was invented by A. Schönhage and the author [Od1, OS]. It enables one to compute all the approximately $T^{1/2}$

zeros of $\zeta(1/2+it)$ in an interval $T \le t \le T + T^{1/2}$ in about $T^{1/2}$ steps. (This algorithm is described in detail in Section 4. At this point we only note that the above description assumes that the RH is satisfied by the zeros between heights T and $T + T^{1/2}$, and that in addition these zeros are simple and well separated. All of these conditions are satisfied in all the ranges that have been investigated.) The new algorithm has now been implemented, and used to compute the zeros described in this paper. It turns out to be fast in practice as well as in theory, and for computing large sets of zeros around the 10^{20}-th zero was at least 10^5 times faster than the straightforward application of the Riemann-Siegel formula. The computations described in this book took about 2000 hours of (otherwise idle) time on a Cray X-MP supercomputer, so they were substantial. However, without the new algorithm they would have been totally infeasible.

The computations that have been carried out with the new algorithm of [OS], and that form the basis for this paper, had several goals. The first was to test the RH numerically. If the RH is false, then counterexamples are probably more likely to be found at large heights than closer to the origin, since the behavior of the zeta function is very constrained at low heights. As it turned out, no counterexample was found. Another, more important goal, was to extend the numerical studies of [Od2] by computing accurate values of large sets of high zeros to provide additional numerical checks on various conjectures about the zeros, especially about the frequency of occurrence of small gaps. If the slightly excessive frequency of small gaps that was observed in [Od2] were to occur again at greater heights, that would cast doubt on many of the conjectures that have been made. The latest computations show excellent agreement with these conjectures in almost all the measured statistics. The excess of small spacings found in [Od2] is still somewhat ambiguous, though, as will be described in Section 2.7.

Another reason for the computations of this monograph was to produce various statistics about the zeta function at large heights. One advantage of the algorithm

of [OS] is that once the main step of the computation is done, it is easy to compute individual values of the zeta function in the covered ranges, and collect many statistics. Such statistics can be used to test various conjectures (about mean values of the zeta function, for example) and to judge how fast the zeta function approaches its asymptotic behavior. Some of the statistics presented in Section 2 show amazingly fast convergence to the asymptotic behavior, while others are far from it. It is remarkable that the most noticeable difference between observed and expected behavior occurs in the study of the distribution of values of $\log|\zeta(1/2+it)|$, which in the ranges that have been investigated is rather far from the normal distribution that has been rigorously proved to hold asymptotically (see Section 2.12 and Figure 2.11.1). On the other hand, many of the unproved conjectures are supported by the numerical data to a surprising degree (see, for example, Figures 2.4.5 and 2.12.1).

The main conclusion that can be drawn from the data in this paper is that in many respects the zeta function reaches its asymptotic behavior slowly, so that even the neighborhood of the 10^{20}-th zero does not represent what happens much higher. That this slow convergence is observed is not too surprising. For example, one important question (see Section 2.10) concerns the maximal size of the zeta function on the critical line. It is known that

$$|\zeta(1/2+it)| = O(t^{\alpha})$$

for constants α that are a bit less than $1/6$, while on the RH one would have

$$|\zeta(1/2+it)| \le t^{o(1)} \quad \text{as} \quad t \to \infty \ .$$

(This is the Lindelöf conjecture.) It would be desirable to produce convincing numerical evidence about the precise maximal size of $\zeta(1/2+it)$. However, this is hard to do. The main difficulty is that near the 10^{20}-th zero, one has $t \approx 1.5 \times 10^{19}$, so that $t^{1/6} \approx 1570$, while $(\log t)^2 \approx 1950$, so it is even hard to distinguish between

these two functions that have entirely different rates of growth. (Throughout the paper, $\log x$ denotes the natural logarithm of x.)

Some of the data from the present computations might also be useful in other number theoretic investigations. For example, the Stark method [St] for obtaining lower bounds for imaginary quadratic number fields with small class numbers depends on knowledge of pairs of zeros of the zeta function that are very close together. (Another method for bounding class numbers, that of Montgomery and Weinberger [MW], depends on zeros of Dirichlet L-functions.)

The final reason for the computations of this paper was to prove that the new algorithm of [OS] is of practical use, and not just a theoretical curiosity. Since this algorithm is complicated, this was not obvious to start with, and a large section of this paper is devoted to a description of its implementation, including various modifications that were made to the basic algorithm described in [OS]. As it turns out, the algorithm is fast, over 10^5 times faster than the older algorithms would have been near the 10^{20}-th zero. Moreover, work on this implementation has suggested many additional modifications, described in Section **??**, which can probably speed up the algorithm by another order of magnitude.

The main sets of zeros that were computed are listed in Table 1.2. The entry for $N = 10^{20}$, for example, means that 175,587,726 zeros were computed, starting with zero number $10^{20} - 30,769,710$, and ending with zero number $10^{20} + 144,818,015$, and that all these zeros are of the form $1/2 + i\gamma$ with $\gamma \approx 1.5 \times 10^{19}$. Throughout the paper, references to the $N = 10^{20}$ data set will denote these 175,587,726 zeros or some subset of them and similarly for the $N = 10^{19}, \ldots,$ data sets.

The starting points for the large data sets listed in Table 1.2 were chosen to be near zeros of round order (such as 10^{20}), to be easy to refer to. It was thought that as far as the distribution of zeros is concerned, these intervals would behave like random ones. One can also concentrate on investigating the behavior of $\zeta(1/2 + it)$ near those t where the zeta function might be expected to behave in an unusual

fashion (e.g., where it is large). Some such special values of t were found, and the computations that were carried out there are listed in Tables 3.1.1 and 3.1.2. (A full explanation of the entries in these tables is given in Section 3.) These computations produced many values of the zeta function and of gaps between zeros that are current the largest ones known.

While the computations that are described in this book did yield values of zeta function zeros at much greater heights than would be feasible with older methods, they do have one serious defect, namely that they are not rigorous. The validity of the values for the zeros that have been computed (and also of the assertion that all these zeros satisfy the RH) depends on the assumption that substantial cancellation among the roundoff errors takes place. This is due largely to the extremely large sizes of the numbers being handled, and not so much to the new algorithm, and is explained in detail in Section 4. At this point we only mention that the values of zeros that have been obtained are believed to be accurate to within $\pm 10^{-6}$ or even better for $N = 10^{20}$. This belief is based partially on the expected cancellation of errors in the computation. The strongest argument for the validity of the computations, however, comes from several large sets of zeros which were computed twice, in entirely different ways. That the numbers being computed were the same follows only from deep mathematical analysis, and is not obvious from the numbers being processed. The resulting duplicate values for the zeros agreed to the expected degree, and this is a strong argument in favor of the validity of the computations. These issues are discussed in greater detail in Section 4.7.

The remainder of this monograph is organized into three sections. Section 2 recalls the basic definitions and conjectures, and then presents the statistics of the large sets of zeros given in Table 1.2. Section 2 is organized into subsections on a variety of topics, such as large values of the zeta function, large and small gaps between consecutive zeros, and many others.

Section 3 is devoted to the zeros listed in Table 3.1.1. First the statistics of

these zeros and of various properties of the zeta function in those ranges are presented. Then some simultaneous Diophantine approximation algorithms (based on the Lovász lattice basis reduction algorithm [LLL]) are described, as well as the ways in which they have been used to produce the points of Table 3.1.1 where the zeta function was expected to behave pathologically, and where it does indeed exhibit unusual behavior.

Section 4 describes the algorithms and computations on which the results of this monograph are based. First the basic algorithm of [OS] is briefly surveyed, and then various modifications to it are described. (Some are minor, while others, such as the use of band-limited function interpolation, are much more substantial.) A discussion of various additional modifications that can be utilized in the future is included (such as the replacement of the crucial rational function evaluation algorithm of [OS] by somewhat similar algorithms that have been proposed in the context of astrophysical and fluid dynamics simulations [GR1], or ways to obtain more rigorous results). There is also a large subsection on the accuracy and validity of the computations of this paper.

Chapter 2

Large sets of zeros: conjectures and statistics

2.1. Notation and definitions

The trivial zeros of the zeta function are $-2, -4, -6, \ldots$. We will consider only the *nontrivial zeros*, which lie in the critical strip $0 < \mathrm{Re}\,(s) < 1$, and are customarily denoted by ρ. Since for every nontrivial zero ρ, $\overline{\rho}$ is also a zero, we will consider only zeros ρ with $\mathrm{Im}\,(\rho) > 0$. (There are no nontrivial zeros ρ with $\mathrm{Im}\,(\rho) = 0$.) We number these zeros $\rho_1, \rho_2, \ldots$ (counting each according to its multiplicity) so that $0 < \mathrm{Im}\,(\rho_1) \leq \mathrm{Im}\,(\rho_2) \leq \ldots$. All the zeros that have been computed so far are simple and lie on the critical line, and so can be written as $\rho_n = \frac{1}{2} + i\gamma_n$, $\gamma_n \in R^+$, with $\gamma_1 = 14.134725\ldots$, $\gamma_2 = 21.022039\ldots$, $\gamma_3 = 25.010857\ldots$, etc. In many definitions throughout the paper we will be tacitly assuming that the RH holds, as otherwise those definitions might not make sense.

Let $N(t)$ denote the number of zeros ρ with $0 < \mathrm{Im}\,(\rho) \leq t$ (counted according to their multiplicity). Then it is known unconditionally [Tit2, Chapter 9.3] that

$$N(t) = \frac{t}{2\pi} \log \frac{t}{2\pi e} + O(\log t) \quad \text{as} \quad t \to \infty\ . \tag{2.1.1}$$

Therefore $\gamma_n \sim 2\pi n/(\log n)$ as $n \to \infty$. Since the zeros become denser as the height increases, and the average vertical spacing between zeros at height t is asymptotic to $2\pi/(\log(t/(2\pi)))$, we define the normalized spacing between consecutive zeros

$1/2 + i\gamma_n$ and $1/2 + i\gamma_{n+1}$ to be

$$\delta_n = (\gamma_{n+1} - \gamma_n)\frac{\log(\gamma_n/(2\pi))}{2\pi} \ . \tag{2.1.2}$$

(Here we are assuming that both zeros satisfy the RH.) It then follows from (2.1.1) that the δ_n have mean value 1 in the sense that for any positive integers N and M,

$$\sum_{n=N+1}^{N+M} \delta_n = M + O(\log(NM)) \ . \tag{2.1.3}$$

For t real and positive (as will be the case throughout the paper) we define

$$\theta(t) = \arg[\pi^{-it/2}\Gamma(1/4 + it/2)] \ , \tag{2.1.4}$$

where the argument is defined by continuous variation of s in $\pi^{-s/2}\Gamma(s/2)$, starting at $s = 1/2$ and going up vertically. We also let

$$Z(t) = \exp(i\theta(t))\zeta(1/2 + it) \ , \tag{2.1.5}$$

so that $|Z(t)| = |\zeta(1/2 + it)|$. Then it follows from the functional equation of the zeta function that $Z(t)$ is real, and sign changes of $Z(t)$ correspond to zeros of $\zeta(s)$ on the critical line. Almost all calculations of the zeta function on the critical line compute calculate $Z(t)$ and not $\zeta(1/2 + it)$ (cf. Section 4). However, it is easy to derive one from the other.

The function $\theta(t)$ is monotonic increasing for $t \geq 7$. For $n \geq -1$, we define the n-th *Gram point* g_n to be the unique solution > 7 to

$$\theta(g_n) = n\pi \ . \tag{2.1.6}$$

We have $g_{-1} = 9.666\ldots$, $g_0 = 17.845\ldots$, etc. Gram points are about as dense as the zeros of $\zeta(s)$ (see Section 2.13 for a detailed discussion), but are much more regularly distributed. In graphs, by a *Gram point scale* we will refer to labeling Gram point g_n by n (or $n - M$ for some fixed M as n varies). For example, Fig. 2.1.1 shows $Z(t)$ near zero number 10^{20}. Figure 2.1.3 shows $Z(t)$ over a somewhat wider range.

We let

$$S(t) = \pi^{-1} \arg \zeta(1/2 + it) \ , \tag{2.1.7}$$

where the argument is defined by continuous variation of s in $\zeta(s)$, starting at $s = 2$, going up vertically to $s = 2 + it$, and then horizontally to $s = 1/2 + it$. (This definition assumes that there are no zeros ρ with $\mathrm{Im}\,(\rho) = t$.) The function $S(t)$ has jump discontinuities at heights equal to zeros. We have

$$N(t) = 1 + \pi^{-1}\theta(t) + S(t) \ , \tag{2.1.8}$$

so that (2.1.1) is a consequence of the asymptotic expansion of $\theta(t)$ (which follows from Stirling's formula [HMF])

$$\theta(t) = \frac{1}{2}t \log(t/(2\pi e)) - \pi/8 + O(t^{-1}) \quad \text{as} \quad t \to \infty \tag{2.1.9}$$

and the bound [Tit2, Theorem 9.4]

$$|S(t)| = O(\log t) \quad \text{as} \quad t \to \infty \ . \tag{2.1.10}$$

Since $N(t)$ is an integer, and $\theta(t)$ is smooth, (2.1.8) shows that $S(t)$ jumps at zeros and decreases at a very steady rate between zeros. Figure 2.1.2 shows $S(t)$ over the same range of values of t as in Fig. 2.1.1, near zero number 10^{20}. This range represents typical behavior of $S(t)$ at that height. (For rare behavior of $S(t)$, see Fig. 3.2.3.) The function $S(t)$ is of crucial importance in understanding the distribution of zeros, and Sections 2.5, 2.13, and 2.14 are devoted largely to its properties.

In comparing empirical distributions of various functions, such as $S(t)$ and δ_n, to their conjectured distributions, we will rely extensively on comparing the moments of their distributions. The method of moments has fallen into some disrepute in statistics because of its many faults, such as lack of robustness. (For example, a single outlier in the data can have a large effect, something we will see in our data.) However, there are some good reasons for using it. One is that it is easy to apply.

A more substantial one is that for many of the statistics of the zeta function, such as those of $S(t)$, or of $Z(t)$, computation of moments is currently essentially the only known tool that can be used to obtain rigorous results. In such cases moments provide the most direct way of comparing empirical distributions to theoretical results.

If a sequence of probability measures with distribution functions $F_n(x)$ is such that for every $k \geq 0$, the k-th moment

$$\mu_n(k) = \int x^k dF_n(x)$$

converges to $\mu(k)$ as $n \to \infty$, then there is a limiting measure with distribution $F(x)$ whose k-th moment is $\mu(k)$. Furthermore, if the $\mu(k)$ determine their measure uniquely, and this measure has distribution function $F(x)$, then the $F_n(x)$ converge to $F(x)$ (in the weak star sense) [Bil, pp. 342–353]. The $\mu(k)$ determine $F(x)$ uniquely if they do not grow too fast [Bil], [Fel, pp. 227–228], so that the normal distribution, for example, is characterized by its moments. On the other hand, the log-normal distribution (distribution of $\exp(\eta)$, where η is normal) is not determined uniquely by its moments [Bil],[Fel].

The standard normal distribution has the density function

$$f(x) = (2\pi)^{-1/2}\ e^{-x^2/2}\ , \tag{2.1.11}$$

with mean 0 and variance 1. Often we will be dealing with quantities (such as $S(t)$) whose known asymptotic distributions are normal, but which have variances on the order of $\log\log N$ (for zeros near zero number N). Since $\log\log N$ grows very slowly, it is to be expected that the observed data will have somewhat different variances, as second order terms are likely to be substantial. (For $N = 10^{20}$, $\log\log N = 3.82976\ldots$, so even an additive constant of 1 in the estimate of the variance makes a considerable difference.) On the other hand, it is not too unreasonable to hope that the shape of the distribution should be close to the expected one. To carry out such a comparison, we will often use a *scaled and translated empirical distributions.*

If $x_1, \ldots, x_n$ are samples (of δ_m, say, or other quantities) with mean a and variance $v = \sigma^2$ (so that σ is the *standard deviation*, or *rms value*),

$$a = \frac{1}{n} \sum_{j=1}^{n} x_j \ , \tag{2.1.12}$$

$$v = \frac{1}{n} \sum_{j=1}^{n} (x_j - a)^2 \ , \tag{2.1.13}$$

then the scaled and translated values will be

$$x_j^* = (x_j - a)/\sigma \ . \tag{2.1.14}$$

The x_j^* have mean 0 and variance 1. The tables will usually list the k-th moment of x_j^* in the k-th entry, but there will be entries giving the ordinary mean a and ordinary variance v that will be marked $k = 1^*$ and $k = 2^*$, respectively. In a few cases where the mean a is extremely small, we will use $x_j^* = x_j/\sigma$. (These cases will be easy to distinguish because the scaled 1-st moment will not be 0.)

Throughout this paper, numbers that have "..." at the end are truncated to the form that is shown, while those without "..." are rounded, but the rounding is sometimes up and sometimes down. Thus, for example, π could be represented as 3.14159..., as 3.14159, or as 3.14160. The log function will always refer to the natural logarithm. References to maximal values of a function $f(x)$ will usually mean the values of $f(x)$ for which $|f(x)|$ is maximal.

Constants such as $n_0, n_1, \ldots$, will generally be different in different sections, but will be the same within a section.

2.2. Validity of the RH and correctness of the computational results

The main question about the validity of the computations described in this paper has to do with size and cancellation of roundoff errors. This issue is discussed in detail in Section 4. Even if we assume that roundoff errors are small (as they seem to be), there remains some further lack of rigor. The set of zeros corresponding to $N = 10^{12}$, for example, is claimed to consist of exactly the zeros numbered

$10^{12} - 6,032$ to $10^{12} + 1,586,163$. Those 1,592,196 values are indeed zeros of the zeta function strictly between Gram points of orders $10^{12} - 6,034$ and $10^{12} + 1,586,162$, provided all the computational steps were correct. Given the degree of regularity in the locations of those zeros, a theorem of Turing (see [Br5, Theorem 3.2] for a modified and corrected version) allows us to conclude, for example, that the 21-st through the 1,592,176-th zeros in our set are indeed zeros numbered $10^{12} - 6,012$ through $10^{12} + 1,586,143$. However, this theorem does not exclude the possibility that, for example, the interval between Gram points $10^{12} - 6,034$ and $10^{12} - 6,014$ might contain some additional zeros. Since such additional zeros would violate either Rosser's rule (see Section 2.14) or even the RH, they seem unlikely to exist, and in any event would not affect most of the statistics to a noticeable extent, and so were assumed not to exist.

There are some further cases of nonrigorous computations in this paper. For example, the conjectured distribution of the δ_n (see Section 2.3) is complicated, and (as was done in [Od2]) was computed using Van Buren's program [VB], with some modifications by S. P. Lloyd and this author. This program uses an involved combination of variational procedures and special function expansions, and no rigorous error analysis for it is known, although it appears to be very accurate (cf. [Od2]).

Other examples of nonrigorous computation are presented by various piecewise linear approximations and other interpolation schemes used in the following sections. They are all thought to produce accurate results, but no proofs are available.

2.3. Eigenvalues of random matrices and zeros

Over the last few decades, an extensive collection of results about eigenvalues of certain types of random matrices has been obtained by mathematical physicists. The aim of these investigations was to obtain insight into the distribution of energy levels in heavy nuclei, and recently their results have been applied to studies of energy levels in other kinds of many-particle systems. Some of the references for

this field are [Be1, Be2, Be3, BG, BGS1, BGS2, BFFMPW, Meh, Por]. Not only are there many beautiful and mathematically rigorous results in this area, but there is also experimental evidence that these results do describe the behavior of physical systems [HPB]. (Because of the difficulty of the experiments, the physical data, which was obtained through a major effort over the span of several decades, is sparse and of poor quality compared to the data that can be obtained for the zeta function.)

The eigenvalue results that will be of greatest interest to us are those of the Gaussian unitary ensemble (GUE), which together with the Gaussian orthogonal ensemble (GOE) and the Gaussian symplectic ensemble (GSE) has been studied extensively. The GUE consists of $n \times n$ complex Hermitian matrices of the form $A = (a_{jk})$, where $a_{jj} = 2^{1/2}\sigma_{jj}$, $a_{jk} = \sigma_{jk} + i\eta_{jk}$ for $j < k$, and $a_{jk} = \overline{a}_{kj} = \sigma_{kj} - i\eta_{kj}$ for $j > k$, where the σ_{jk} and η_{jk} are independent standard normal variables. (The GOE consists of real symmetric matrices defined similarly.) The eigenvalues of these matrices are real, and it is their asymptotic distribution, as $n \to \infty$, that is of interest. If we denote the eigenvalues by $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, then we have the *Wigner semi-circle law*: if $M(x)$ denotes the expected number of eigenvalues $\le x$, then for all fixed real x,

$$\lim_{n\to\infty} n^{-1} M(x\sqrt{n}) = \begin{cases} \dfrac{1}{2\pi}\displaystyle\int_{-2}^{x} (4-u^2)^{1/2} du\ , & |x| < 2\ , \\ 0\ , & x \le -2\ , \\ 1\ , & x \ge 2\ . \end{cases} \tag{2.3.1}$$

This distribution law applies to much more general classes of matrices than those of the GUE and related ensembles. For the GUE (and also for the GOE and GSE) a further step is possible in that one can obtain precise information about the distribution of spacings between consecutive eigenvalues. The complete distribution of eigenvalues is known, and one can derive many limit laws. To do that one normalizes the eigenvalues (basically by stretching the distance between consecutive eigenvalues $\lambda < \lambda'$ by a factor of $(4n - \lambda^2)^{1/2}/(2\pi)$) to make the average nearest neighbor

spacing equal to 1. With this normalization, the distribution of eigenvalues looks the same everywhere (in the limit as $n \to \infty$) and one can in principle determine any desired statistic of the zeros. (Doing so in practice means evaluating a definite multidimensional integral, which is often hard, and gives rise to interesting problems.) For example, if we use w to denote a normalized eigenvalue in the GUE, then one finds that for any fixed $0 \le \alpha < \beta < \infty$,

$$\mathcal{E}(|\{w' : \ w < w', \quad w' - w \in [\alpha, \beta]\}|) \sim \int_\alpha^\beta \left(1 - \left(\frac{\sin \pi u}{\pi u}\right)^2\right) du \tag{2.3.2}$$

as $n \to \infty$, where $\mathcal{E}(z)$ is the expectation of z. We say that $1 - ((\sin \pi u)/(\pi u))^2$ is the *pair correlation function* of the GUE. (The pair correlation functions of the GOE and the GSE are different.) Equation (2.3.2) shows, for example, that it is rare for GUE eigenvalues to be close together. If the w's were obtained by choosing n points independently and uniformly from the interval $[0, n]$ and letting $n \to \infty$, the pair correlation function would be identically 1. The GUE pair correlation function in the range $0 \le u \le 3$ is drawn as the solid curve in Fig. 2.4.1, and is far from being a constant.

If w is a normalized eigenvalue of the GUE, we let $w^{(k)}$ denote the k-th smallest normalized eigenvalue of those that are $> w$. Then it is known that the k-th nearest spacings $w^{(k)} - w$ satisfy a distribution law; for all $0 \le \alpha < \beta < \infty$,

$$Prob(w^{(k)} - w \in [\alpha, \beta]) \sim \int_\alpha^\beta p(k-1, u) du \tag{2.3.3}$$

as $n \to \infty$. The probability densities $p(k, u)$ (referred to as $p_2(k; u)$ in many publications, such as [CM2, ?, MdC], where the subscript 2 denotes the GUE) are complicated functions defined in terms of linear prolate spheroidal functions. For methods of computing them, see [MdC, Od2]. Graphs of $p(0, u)$ and $p(1, u)$ are given by the solid lines in Figs. 2.4.4 and 2.4.6, respectively. Those graphs show the "rigidity" of the GUE; the eigenvalues repel each other and most of the time are close to the expected distance from their neighbors. For all $u \ge 0$,

$$1 - \left(\frac{\sin \pi u}{\pi u}\right)^2 = \sum_{k=0}^{\infty} p(k, u) \ . \tag{2.3.4}$$

We note for future reference that the $p(k,u)$ have the following Taylor series expansions around 0 [Meh, MdC]:

$$p(0,u) = \frac{\pi^2}{3}u^2 - \frac{2\pi^4}{45}u^4 + \frac{\pi^6}{315}u^6 + \cdots , \tag{2.3.5}$$

$$p(1\ ,u) = \frac{\pi^6}{4050}u^7 + \cdots, \tag{2.3.6}$$

$$p(2,u) = \frac{\pi^{12}}{5358150000}u^{14} + \ldots .$$

The normalized eigenvalues in the GUE have (in the limit as $n \to \infty$) a stationary distribution. This means that clusters of eigenvalues have the same distribution no matter where in the spectrum they are located. However, this distribution is not Markovian, so that the distribution of an eigenvalue depends not just on the preceding eigenvalue, but on all previous ones as well.

The basic results about distribution of GUE eigenvalues are completely rigorous. However, they do have many gaps. One of them is that the results are obtained by averaging over the full ensemble of GUE matrices. It is conjectured that if one considers a large random GUE matrix, the distribution of its eigenvalues will be close to that of the entire ensemble with high probability. Although numerical calculations confirm this conjecture, there is no proof of it. Also, it is thought that entries of the matrix do not have to be of exactly the form specified above for the GUE result to hold.

The main goal of this monograph (and of the preceding paper [Od2]) is to test the conjecture, which will be referred to as the *GUE hypothesis*, the *GUE theory*, or simply the *GUE*, that the zeros of the zeta function behave like eigenvalues of the GUE. More precisely, it is conjectured that the δ_n behave asymptotically like $w^{(1)} - w$ in the GUE, so that for any $0 \le \alpha < \beta < \infty$,

$$M^{-1}|\{n:\ N+1 \le n \le N+M,\quad \delta_n \in [\alpha,\beta]\}| \sim \int_\alpha^\beta p(0,u)du \tag{2.3.7}$$

as $M, N \to \infty$ with M not too small compared with N. Similarly, it is conjectured

that

$$M^{-1}|\{n:\ N+1 \le n \le N+M,\quad \delta_n + \delta_{n+1} \in [\alpha,\beta]\}| \sim \int_\alpha^\beta p(1,u)du\ . \qquad (2.3.8)$$

More generally, the same reasoning leads one to expect that for any fixed k, the empirical distribution function of $\delta_n, \delta_{n+1}, \ldots, \delta_{n+k}$ for $N+1 \le n \le N+M$ approaches the stationary process that holds for the GUE.

If the GUE hypothesis is true, it might be interpreted as providing some support for the Hilbert and Pólya conjectures [Be2, Be3, Mon1, Od3] which predict that the RH is true because the zeros of the zeta function correspond to eigenvalues of a positive linear operator. The argument is that if such an operator exists, its eigenvalues might be similar to those of a random operator (especially if, as is conjectured for the GUE, most random operators have similar eigenvalue distributions), and a random linear operator ought to be the limit of a sequence of random matrices.

If the GUE hypothesis were true, that would also be of interest in physics, as the zeta function could then be used as a model of quantum chaos [Be2, Be3].

The main theoretical support and inspiration for the GUE hypothesis comes from H. Montgomery's work on the pair correlation function of the zeros of the zeta function. Under the assumption of the RH, Montgomery showed [Mon1, Mon2] that if we define

$$F(\alpha,T) = 2\pi(T\log T)^{-1} \sum_{\substack{0<\gamma\le T\\ 0<\gamma'\le T}} T^{i\alpha(\gamma-\gamma')} \frac{4}{4+(\gamma-\gamma')^2} \qquad (2.3.9)$$

for α and T real, $T \ge 2$, then

$$F(\alpha,T) = (1+o(1))T^{-2\alpha}\log T + \alpha + o(1) \quad \text{as} \quad T\to\infty\ , \qquad (2.3.10)$$

uniformly for $0 \le \alpha \le 1$. Montgomery also observed that if the primes are distributed sufficiently uniformly in arithmetic progressions, then

$$F(\alpha,T) = 1 + o(1) \quad \text{as} \quad T\to\infty \qquad (2.3.11)$$

uniformly for $\alpha \in [a,b]$, where $1 \le a < b < \infty$ are any constants. If the conjecture (2.3.11) were true, then one would find that for any $0 < \alpha < \beta < \infty$,

$$N^{-1}|\{(n,k):\ 1 \le n \le N,\ k \ge 0,\ \ \delta_n + \delta_{n+1} + \cdots + \delta_{n+k} \in [\alpha,\beta]\}| \sim \int_\alpha^\beta \left(1 - \left(\frac{\sin \pi u}{\pi u}\right)^2\right) du \tag{2.3.12}$$

as $N \to \infty$. The relation (2.3.12) is known as the Montgomery pair correlation conjecture. It says that the pair correlation of the zeros of the zeta function is the same as that of the GUE. Since the pair correlations of the GOE and GSE are different, (indeed, they are even inconsistent with (2.3.10)), this leads one to expect that the zeros might behave like eigenvalues of the GUE rather than GOE or GSE. Therefore the discussion above was concentrated on the GUE distribution. (One possible implication of this observation is that the hypothetical Hilbert-Pólya operator is likely to be complex.)

Montgomery's hypothetical result (2.3.10) and the conjectures (2.3.11) and (2.3.12) are the main theoretical evidence we have in favor of the GUE hypothesis, and the two conjectures depend on far-reaching assumptions about pseudorandom behavior of primes. Some further evidence in favor of the GUE hypothesis was provided by Ozluk [Oz1], who showed that if one considers a function similar to $F(\alpha,T)$, but where one sums over zeros of many Dirichlet L-functions, then under the assumption of the Generalized Riemann Hypothesis for these L-functions, the analog of Montgomery's conjecture (2.3.11) is true for $1 \le \alpha \le 2$. Some further slight support for the GUE hypothesis is provided by new results of Ozluk [Oz2] on zeros of Dirichlet L-functions close to the real axis.

Extensive numerical evidence in favor of the GUE hypothesis was presented in [Od2]. It was based largely on computed values of γ_n, with $1 \le n \le 10^5$ and $10^{12}+1 \le n \le 10^{12}+10^5$. With some slight exceptions (such as the slight excess of small δ_n that was mentioned in the Introduction) this evidence was in excellent agreement with the GUE hypothesis, and the degree of agreement improved

dramatically as one went from the first 10^5 zeros to those near zero number 10^{12}. Some numerical evidence for the pair correlation conjecture for Dirichlet L-functions has been obtained since then by Hejhal [Hej5]. It involved computations of a few quadratic character L-functions for several moduli at large heights. Much more extensive data have been obtained by Rumely [**?**], who computed all zeros of all L-functions to small moduli up to height 2500. His evidence also supports the GUE hypothesis.

Various theoretical results and conjectures related to the GUE theories and the pair correlation conjecture have been obtained in recent years. Some of the references are [Be2, Be3, Be4, Fu8, Gal2, Gal3, Gal4, GM, Go1, Go2, Go3, GG, GHB, GM, HB1, Mue2].

2.4. General distribution of gaps between zeros

Figure 2.4.1 shows how well the pair correlation conjecture is satisfied. The solid line is the GUE prediction $y = 1 - ((\sin \pi x)/(\pi x))^2$. The scatterplot is based on about 8×10^6 zeros near zero number 10^{20}. Let

$$\begin{aligned} n_1 &= 10^{20} - 15,409,240, \\ n_2 &= 10^{20} - 13,366,460, \\ n_3 &= 10^{20} - 10,302,282, \\ n_4 &= 10^{20} - 6,216,711, \\ n_5 &= 10^{20} - 42,778, \\ n_6 &= 10^{20} + 15,316,087, \\ n_7 &= 10^{20} + 46,073,204, \\ n_8 &= 10^{20} + 47,098,588, \end{aligned}$$

and

$$V = \{n : \ n_i \le n < n_i + 10^6 \quad \text{for some} \ \ i, \quad 1 \le i \le 8\} \ .$$

Then for each interval $I = [\alpha, \beta)$ with $\alpha = k/20$, $\beta = \alpha + 1/20$, $0 \le k < 60$, a star is placed at the point $x = (\alpha + \beta)/2$, $y = a_{\alpha,\beta}$, where

$$a_{\alpha,\beta} = \frac{20}{8 \times 10^6} \left| \{(n,k) : \ n \in V, \ \ k \ge 0, \ \ \delta_n + \cdots + \delta_{n+k} \in [\alpha, \beta)\} \right| \ . \qquad (2.4.1)$$

The solid line is the GUE prediction $y = 1 - ((\sin \pi x)/(\pi x))^2$. As can be seen, the agreement between the conjectured and observed values is excellent.

Figure 2.4.2 presents similar data, but this time based on just 10^6 values of n; $n_9 \le n < n_9 + 10^6$, $n_9 = 10^{12} - 6,032$. A comparison of these two graphs with Figures 1 and 2 of [Od2] is instructive. Those figures show similar graphs, but based in each case on 10^5 zeros starting with zeros number 1 and $10^{12} + 1$. The scatterplot of Fig. 2.4.2 is much smoother than that of Fig. 2 of [Od2], because the former is based on 10^6 instead of 10^5 samples, and so the sampling error is smaller. That same reason explains why the scatterplot of Fig. 2.4.1 looks smoother than that of Fig. 2.4.2. Even if we make allowances for the different sample sizes, though, it is clear that the agreement between empirical and predicted values improves dramatically from $N = 1$ to $N = 10^{12}$, and improves even more between $N = 10^{12}$ and $N = 10^{20}$. In all cases, the empirical data has more pronounced peaks and troughs than expected, but this effect decreases as the height increases.

Some of the pair correlation function oscillations can be seen even for normalized spacings that exceed 3. Figure 2.4.3 shows a graph based, just like Fig. 2.4.1, on 8×10^6 zeros near zero number 10^{20}. Here, though, the scatterplot was smoothed slightly by applying the lowess function of [BC] (an implementation of Cleveland's robust locally weighted regression [Clev]). The reason for this smoothing is that even with 8×10^6 zeros, each of the $a_{\alpha,\beta}$ defined in (2.4.1) corresponds to about 4×10^5 counts (n, k). Therefore we can expect random sampling errors about $(4 \times 10^5)^{1/2}$, which gives a variation of about 1.6×10^{-3} in the value of $a_{\alpha,\beta}$. Given the small variation in the GUE prediction $y = 1 - ((\sin \pi x)/(\pi x))^2$ over the range $3 \le x \le 5$, this random sampling error produces a confusing picture if the data is not smoothed. (Another, but slightly less effective way to produce a better picture is to use sampling intervals larger than 1/20. The resulting picture is similar to that of Fig. 2.4.3.)

Figure 2.4.3 shows that the empirical pair correlation function, even for $N = 10^{20}$, has peaks and troughs that are more pronounced than those of the conjectured

distribution, at least in the range $3 < x < 5$. This is also true in the range $5 < x < 10$.

Figures 2.4.4 and 2.4.5 show the distribution of the normalized spacings δ_n for $N = 10^{12}$ and $N = 10^{20}$, based on the 1,592,196 and 78,893,234 zeros, respectively, that have been computed. Thus, for example, in Fig. 2.4.4 a star is plotted at $x = (\alpha + \beta)/2$, $y = b_{\alpha,\beta}$ for $\alpha = k/20$, $\beta = \alpha + 1/20$, $0 \le k \le 59$, where

$$b_{\alpha,\beta} = \frac{20}{1592195} \left| \{ n : \ 10^{12} - 6032 \le n \le 10^{12} + 1586162, \ \ \delta_n \in [\alpha, \beta) \} \right| \ . \quad (2.4.2)$$

The solid lines are the GUE predictions, $y = p(0, x)$. Similarly, Figures 2.4.6 and 2.4.7 show the distribution of $\delta_n + \delta_{n+1}$. (Similar graphs based on the first 10^5 zeros are contained in [Od2].)

The graphs show excellent agreement between conjecture and numerical data, and, as was to be expected, the degree of agreement increases substantially as one goes from $N = 1$ to $N = 10^{12}$, and then improves a bit more as one goes to $N = 10^{20}$. That the disagreement is greater for $\delta_n + \delta_{n+1}$ than for δ_n is to be expected, given that $S(t)$ is small. (See Section 2.5 for a discussion of this.)

A quantitative measure of the agreement between observed and conjectured distributions is shown in Tables 2.4.1 through 2.4.3, which display moments of distributions. For each set of M zeros, $K < n \le K + M$ ($M = 1,592,196$ for $N = 10^{12}$, 78,893,234 for $N = 10^{20}$, etc., where K is close to N.) Table 2.4.1 displays

$$(M-1)^{-1} \sum_{n=K+1}^{K+M-1} (\delta_n - 1)^k \ , \quad (2.4.3)$$

while Table 2.4.2 shows

$$(M-2)^{-1} \sum_{n=K+1}^{K+M-2} (\delta_n + \delta_{n+1} - 2)^k \ , \quad (2.4.4)$$

in each case for $2 \le k \le 10$. (The values for $N = 1$ are taken from [Od2].) Table 2.4.3 shows moments of $\log \delta_n$, δ_n^{-1}, and δ_n^{-2}. The values predicted by the GUE are also shown.

Tables 2.4.1 to 2.4.3 show satisfactory agreement between observed values and conjectured ones, with the degree of agreement increasing as the height of the zeros increases. (The slightly anomalous value for the moment of δ_n^{-2} for $N = 10^{18}$ is due to one extremely small δ_n that is very unusual and will be discussed in Sections 2.6 and 2.8.)

The Kolmogorov test [KS, Section 30.49] yields a method for measuring the agreement between the observed distribution of the δ_n and the GUE predictions. If samples $x_1, \dots, x_n$ are drawn from a distribution with a continuous cumulative distribution function $F(z)$, let $F_e(z)$ denote the sample distribution function:

$$F_e(z) = n^{-1} |\{k : \ 1 \le k \le n, \quad x_k \le z\}| \ .$$

The Kolmogorov statistic is then

$$D = \sup_z |F_e(z) - F(z)| \ . \tag{2.4.5}$$

If the x_i are drawn independently from the distribution characterized by $F(z)$, then [KS, Eq. 30.132]

$$\lim_{n \to \infty} Prob(D > u n^{-1/2}) = g(u) \ , \tag{2.4.6}$$

where

$$g(u) = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp(-2r^2 u^2) \ . \tag{2.4.7}$$

Table 2.4.4 gives the Kolmogorov statistic D for δ_n and $\delta_n + \delta_{n+1}$ for several blocks of 10^6 consecutive values of n. The set denoted by $N = 10^{12}$ corresponds to $n_9 \le n < n_9 + 10^6$; the ones denoted by $N = 10^{20}(a)$, $N = 10^{20}(b)$, and $N = 10^{20}(c)$ start at $n = n_6$, $n = n_8$ and $n = n_5$, respectively, where the n_i were defined at the beginning of this section. The "$N = 10^{12}$ vs. GUE" entry, for example, gives the Kolmogorov statistic of the $N = 10^{12}$ set when it is compared to the GUE distribution. For each value of D, the "prob." column gives an estimate that this statistic would arise if the δ_n ($\delta_n + \delta_{n+1}$, respectively) were drawn independently for each n from the GUE distribution. This estimate is obtained by evaluating $g(D \times 1000)$. The

"$N = 10^{20}(a)$ vs. $N = 10^{20}(b)$" row of the table was obtained by constructing a continuous distribution from the $N = 10^{20}(b)$ data and computing the Kolmogorov statistic for the discrete $N = 10^{20}(a)$ data against this continuous distribution.

What is apparent from Table 2.4.4 is that as the height increases, the empirical distributions of δ_n and $\delta_n + \delta_{n+1}$ do approach that of the GUE. When one computes the D statistic for the δ_n in the 10 blocks of 10^5 consecutive zeros that are contained in the $N = 10^{20}(b)$ set, one obtains values ranging between 0.002 and 0.0031, which correspond to probabilities between 0.83 to 0.3 of occurring if the δ_n were drawn from the GUE distribution. Thus for sets of 10^5 zeros around zero number 10^{20}, it is essentially impossible to distinguish the empirical distribution of the δ_n from the expected one. (For $\delta_n + \delta_{n+1}$, the corresponding D values are 0.0035 and 0.00555, which gives probabilities of 0.17 and 0.004, so the fit here is slightly worse.)

The comparison of the three different sets of 10^6 zeros near zero number 10^{20} to each other is revealing. The Kolmogorov statistics D are small (especially for δ_n), and indicate that all three sets come from essentially the same distribution. What seems to be happening is that at each height, when we examine large sets of zeros, the δ_n and $\delta_n + \delta_{n+1}$ behave as if they were drawn independently from some distributions that depend on t, change slowly as t changes, and tend to the GUE distributions as $t \to \infty$.

2.5. Values of $S(t)$

The upper bound (2.1.10) for $S(t)$ is the best that is known unconditionally. The Lindelöf Hypothesis (see Section 2.9) implies that $|S(t)| = o(\log t)$ as $t \to \infty$, while the RH implies [Tit2] that

$$|S(t)| = O\left(\frac{\log t}{\log\log t}\right) \quad \text{as} \quad t \to \infty \ . \tag{2.5.1}$$

The true rate of growth is thought to be much smaller. The best lower bound that has been proved under the RH is that of Montgomery [Mon3], who showed

$$S(t) = \Omega_{\pm}\left(\left(\frac{\log t}{\log\log t}\right)^{1/2}\right) \quad \text{as} \quad t \to \infty \ . \tag{2.5.2}$$

(The best unconditional bound, due to Tsang [Ts1, Ts2], replaces the square root in (2.5.2) by a cube root.) Montgomery [Mon3] has conjectured that the quantity on the right side of (2.5.2) represents the correct rate of growth of $S(t)$, and Joyner [Joy2] has presented a heuristic argument supporting this conjecture. As we will see in Section 2.6, the GUE suggests that $|S(t)|$ might occasionally get as large as $(\log t)^{1/2}$, which would contradict the Montgomery conjecture. In any case, it is thought likely that

$$|S(t)| \le (\log t)^{1/2+o(1)} \quad \text{as} \quad t \to \infty \ . \tag{2.5.3}$$

Some lower bounds for $S(t+h) - S(t)$ are also known, see [Ts1, Ts2], for example.

Not only is $S(t)$ small, but its oscillations tend to cancel out. If we define

$$S_1(t) = \int_{t_0}^{t} S(u)du \ , \tag{2.5.4}$$

then $|S_1(t)| = O(\log t)$ unconditionally, and $|S_1(t)| = O((\log t)(\log\log t)^{-2})$ on the RH [Tit2]. The true maximal magnitude of $|S_1(t)|$ is probably again around $(\log t)^{1/2}$. (See [Ts1, Ts2] for lower bounds. The estimate $|S_1(t)| = o(\log t)$ is equivalent to the Lindelöf Hypothesis, see Notes to Chapter 13 of [Tit2].) Furthermore, if one chooses t_0 appropriately, then one obtains

$$\int_0^t S_1(u)du = O(\log t) \quad \text{as} \quad t \to \infty \ . \tag{2.5.5}$$

(The same property applies to further iterations of this process.) In addition,

$$\lim_{T\to\infty} T^{-1} \int_0^T S_1(t)^2 dt = c$$

exists for a constant $c > 0$ (Theorem 14.19 of [Tit2]).

Selberg [Sel2] proved, under the assumption of the RH, that for every fixed positive integer k,

$$\int_0^T S(t)^{2k}dt = \frac{(2k)!}{k!(2\pi)^{2k}}T(\log\log T)^k(1+O((\log\log T)^{-1})) \tag{2.5.6}$$

as $T\to\infty$. Later [Sel3] he proved similar estimates unconditionally, with $(\log\log T)^{-1}$ in the remainder term replaced by $(\log\log T)^{-1/2}$. Although it was apparently not noticed right away, these results imply (unconditionally) that $S(t)$ is asymptotically normally distributed with mean 0 and variance $2\pi^2\log\log t$, so that for $\alpha<\beta$,

$$\lim_{T\to\infty} T^{-1}\left|\left\{t: 0\le t\le T, \frac{S(t)}{(2\pi^2\log\log T)^{1/2}}\in(\alpha,\beta)\right\}\right| = (2\pi)^{-1/2}\int_\alpha^\beta e^{-x^2/2}dx\ . \tag{2.5.7}$$

For further results on moments and distributions of $S_1(t)$, $S(t+h)-S(t)$, and related functions, see [Fu1, Fu2, Fu3, Fu4, Fu8, GM, Gh1, Gh2, Go2, Joy1, Ts1, Ts2]. Goldston [Go2] has improved the estimate (2.5.6) for $k=1$ by showing, under the assumption of the RH, that

$$\int_0^T S(t)^2dt = \frac{T}{2\pi^2}\log\log T + \frac{T}{2\pi^2}\left(c_1+\int_1^\infty F(\alpha,T)\alpha^{-2}d\alpha\right)+o(T) \tag{2.5.8}$$

as $T\to\infty$, where $F(\alpha,T)$ is defined by (2.3.9), and c_1 is a constant,

$$c_1 = c_0+\sum_{m=2}^\infty\sum_p\left(-\frac{1}{m}+\frac{1}{m^2}\right)\frac{1}{p^m}\ , \tag{2.5.9}$$

where $c_0=0.577\ldots$ is Euler's constant. (The sign of the m^{-1} term is wrong in [Go2].) If Montgomery's pair correlation conjecture (2.3.11) holds, then $\int_1^\infty F(\alpha,T)\alpha^{-2}d\alpha$ is asymptotic to the constant 1, but if his conjecture were to fail, it is conceivable that the second order term in the asymptotic expansion of $\int_0^T S(t)^2dt$ might oscillate.

Table 2.5.1 presents data on the moments of $S(t)$. Statistics were collected on two intervals of the form $(\gamma_n,\gamma_{n+10^6})$, where $n=n_1=10^{12}-6,032$ for the $N=10^{12}$ data, and $n=n_2=10^{20}-48,778$ for the $N=10^{20}$ data. The average values of $S(t)$ and $S(t)^2$ for these sets are given in the $k=1^*$ and $k=2^*$ zeros. To obtain a

good comparison with the asymptotic normal distribution, the other moments were scaled, so that if we let σ^2 be the mean value of $S(t)^2$, then the $k = 1, 2, \ldots, 8$ entries denote the average values of $(\sigma^{-1}S(t))^k$, and the $k = |1|, |3|$, and $|5|$ entries the average values of $|\sigma^{-1}S(t)|^k$. Finally, the last column gives the corresponding values for the standard normal distribution. As we can see, the agreement between empirical values and asymptotic ones is reasonably good, and is somewhat better for $N = 10^{20}$ than for $N = 10^{12}$.

$S(t)$ has jumps discontinuities by 1 at zeros and decreases monotonically between zeros with derivative very close to -1 (on Gram point scale). Since there is asymptotically one zero per Gram point, the smallest mean values of $S(t)^{2k}$ for any $k \in Z^+$ that is at all conceivable would be obtained by having a zero exactly halfway between every two neighboring Gram points. This would yield a mean value of $S(t)^2$ of $1/6$. The values that are observed, 0.23 for $N = 10^{12}$ and 0.26 for $N = 10^{20}$, are not much larger than that.

That the distribution of $S(t)$ is close to the normal one can be seen visually in Fig. 2.3.1. This figure is based on determining for what fraction of values of $t \in (\gamma_{n_2}, \gamma_{n_2+10^8})$ we have $S(t) \in [k/100, (k+1)/100)$, and then scaling the resulting histogram by σ to produce a graph that can be compared to that of the standard normal distribution. It is curious that the observed distribution of $S(t)$ is less peaked than the normal one, whereas in most of the other comparisons the empirical distributions have sharper peaks than expected. It is especially interesting to compare Fig. 2.5.1 to Fig. 2.11.1, which compares the distribution of $\log|Z(t)|$ (up to a constant the harmonic conjugate of $S(t)$) to the normal distribution. In both cases the limiting distributions are known to be normal (even without assuming the RH), but the observed deviations from normal behavior are different for $S(t)$ and $\log|Z(t)|$, and are much more pronounced in the latter case.

The area between the two curves in Fig. 2.5.1 is 0.023. For the corresponding figure using the $N = 10^{12}$ data, the area is 0.029.

Since both $S(t)$ and its integral $S_1(t)$ are small, we can expect that $S(t)$ will have many sign changes, and several results in this direction have been proved, the strongest ones being due to Ghosh [**?**] and Mueller [Mue1], but they are all weak. For example, Mueller proves that gaps between consecutive zeros of $S(t)$ are $O(\log\log\log t)$. On the other hand, we know that $S(t)$ has a limiting normal distribution with variance on the order of $(\log\log t)^{1/2}$ and mean close to 0, and that it cannot vary too widely (in particular, except for jump discontinuities at zeros it is monotone decreasing with derivative close to $-\log t$). Therefore we might expect that the ratio of the number of zero crossings of $S(t)$ for $t \in (\gamma_N, \gamma_{N+M})$ to M might be roughly the fraction of t in (γ_N, γ_{N+M}) for which $|S(t)| \leq 1$. This suggests that on average there ought to be on the order of $(\log\log t)^{-1/2}$ zeros of $S(t)$ per Gram interval.

The number of sign changes of $S(t)$ in the intervals that have been investigated can be determined easily from the statistics of Gram blocks and of exceptions to Rosser's rule that have been collected. When g_n is a good Gram point (see Section 2.13 for a definition) that is not close to an exception to Rosser's rule, and is not a zero of the zeta function, then $S(g_n) = 0$, and $S(t)$ changes sign at g_n. We will count this sign change as occurring in the Gram interval $[g_n, g_{n+1})$. If $B(n,k)$ is a Gram block that has exactly k zeros, then an easy accounting shows that $S(t)$ has exactly 2 sign changes in $B(n,k)$. On the other hand, when $B(n,k)$ is an exception to Rosser's rule (see Section 2.14 for definitions), and $[g_m, g_{m+r})$ is the smallest union of Gram blocks that contains both the exception and the excess zeros, then a similar accounting shows that $[g_m, g_{m+r})$ contains exactly 2 sign changes. Thus if Gram's law (see Section 2.13) held universally, we would have an average of 2 sign changes of $S(t)$ for every zero of $zeta(s)$. Departures from Gram's law lower this average. Table 2.5.2 shows the computed averages for the different data sets. There is a steady decrease in the average, but it is slow. Since the argument in the preceding paragraph suggests a rate of decrease of $(\log\log t)^{-1/2}$, this is not surprising.

For every exception $B(n,k)$ to Rosser's rule there is a t nearby with $|S(t)| \geq 2$ (and even $|S(t)| > 2$, if zeros do not coincide with Gram points, as seems likely). Statistics about these large values of $S(t)$ were collected during investigation of exceptions to Rosser's rule. Large values of $S(t)$ are of special interest because it is only when $S(t)$ is large that unusual behavior of the zeta function can take place. Locally extreme values of $S(t)$ occur at zeros. (Each zero has associated to it two values of $S(t)$, the limits of $S(t)$ as t approaches the zero from the right or the left.) Table 2.5.3 shows the values of $S(t)$ for which $|S(t)|$ is largest in absolute value, as well as the number of zeros at which $|S(t)| > 2.3$ divided by the number of exceptions to Rosser's rule. The largest value of $|S(t)|$ that was found in these computations is 2.7916, while among the first 1.5×10^9 zeros the largest such value is 2.3137 [LRW2]. (A point t at which $|S(t)| = 2.8747$ was found later in the computations described in Section 3.) Earlier computations established that $|S(t)| < 1$ for $7 < t \leq 280$, and $|S(t)| < 2$ for $7 < t \leq 6.8 \times 10^6$.

The values of $S_1(t)$ were investigated in the two intervals $(\gamma_n, \gamma_{n+10^6})$, where $n = 10^{12} - 6,032$ (for the $N = 10^{12}$ set) and $n = 10^{20} - 48,778$ (for $N = 10^{20}$). The values of $S_1(\gamma_n)$ were chosen to make

$$\int_{\gamma_n}^{\gamma_m} S_1(t)dt = 0 \tag{2.5.10}$$

for $m = n + 10^6$. The data that were obtained are summarized in Table 2.5.4; the mean of $S_1(t)^4$, for example, refers to

$$\frac{1}{\gamma_m - \gamma_n} \int_{\gamma_n}^{\gamma_m} S_1(t)^4 dt \ .$$

In addition to the uncertain choice of $S_1(\gamma_n)$, there were additional problems in these computations due to the accumulating errors from the uncertainties in the values of zeros and $S(t)$. Values computed over shorter intervals suggest that the mean values in Table 2.5.4 are accurate. The entry for sign changes of $S_1(t)$ refers to the number of sign changes per Gram interval. This figure appears to be moderately accurate.

Changing the initial value of $S_1(\gamma_n)$ by $\pm 10^{-4}$ varied the number of computed sign changes of $S_1(t)$ for the $N = 10^{20}$ interval only between 73799 and 74089.

2.6. Extreme gaps between zeros

In its weakest form, the GUE hypothesis predicts only that (2.3.7) holds for all $0 \le \alpha < \beta < \infty$, and so it says nothing about the existence of a small number ($o(M)$, say) of large or small δ_n. A double zero of the zeta function, giving $\delta_n = 0$, would not by itself contradict this weak hypothesis. On the other hand, it is known (cf. (2.3.3) and (2.3.5)) that in the GUE,

$$Prob(\delta_n \le x) = \frac{\pi^2}{9}x^3 - \frac{2\pi^4}{225}x^5 + \frac{\pi^6}{2205}x^7 + \cdots , \tag{2.6.1}$$

so very small δ_n (roughly $o(M^{-1}/3)$ among M samples) are rare in the GUE, and a similar result holds for large δ_n. A strong form of the GUE hypothesis would predict that even extreme values of δ_n (and $\delta_n + \delta_{n+1}$) for the zeta function would behave as in the GUE model.

Given the constraints on $S(t)$ described in Section 2.5, we can expect that even if the strong form of the GUE hypothesis holds, it only applies to the zeta function at large heights, and that the lower the region under investigation, the fewer extreme values of δ_n or $\delta_n + \delta_{n+1}$ we will find. This is clear for large values of δ_n and $\delta_n + \delta_{n+1}$, as these clearly correspond to large values of $|S(t)|$. It is also true for small values of δ_n and $\delta_n + \delta_{n+1}$, though, since several zeros clustered close together again force $|S(t)|$ to be large.

What was observed in [Od2] in a comparison of the first 10^5 zeros to 10^5 zeros starting with zero number 10^{12} is that the above predictions were largely satisfied by the data. In general, there was a deficiency of extreme values of δ_n and $\delta_n + \delta_{n+1}$ (compared to the GUE prediction), but this deficiency declined as one considered the higher zeros. There was, however, one observation that went counter to expectations. The number of small δ_n that were observed at large heights was larger than predicted by the GUE theory. This excess was not large, but it was also observed in the data

for 10^5 zeros starting with zero number 2×10^{11}, as well as by some data based on the first 1.5×10^9 zeros. This excess of small spacings was very counterintuitive, and so gave rise to some suspicions about the validity of the GUE hypothesis.

Table 2.6.1 shows the extremal values of δ_n and $\delta_n + \delta_{n+1}$ that were found in each data set. (The number of zeros in each data set is given in Table 1.2.) The last column in Table 2.6.1 gives the probability that the minimal δ_n would not exceed the values in the second column if all the δ_n in the data set were drawn independently from the GUE distribution. From (2.6.1), we see that the probability that the smallest δ_n out of M that are drawn from the GUE satisfies $\delta_n \le x$ is about

$$1 - \left(1 - \frac{\pi^2}{9} x^3\right)^M \sim 1 - \exp(-\pi^2 x^3 M/9) \ . \tag{2.6.2}$$

This approximation was used to compute the last column of Table 2.6.1. We can see that most of the entries in that column are high (although not too high, which would indicate a severe deficiency of small spacings), while those for $N = 10^{18}$ (where $\delta_n = 0.001124$ for $n = 10^{18} + 12,376,780$, a case that will be discussed in sections 2.7 and 4.5) and for $N = 10^{19}$ (where $\delta_n = 0.000897$ for $n = 10^{19} + 15,987,196$ is the smallest δ_n that was found in our computations) are extremely low. The smallest value of δ_n that is known is $\delta_n = 0.000310$ for $n = 1,048,449,114$ (found by van de Lune et al. [LRW2]), and the probability of such a small spacing occurring among 1.5×10^9 samples drawn from the GUE is only 0.048. Thus the extremely small values of the δ_n do appear to be somewhat more frequent than expected. (Some more evidence pointing to this conclusion is presented in Section 3.)

When we next consider small, but slightly larger spacings, we find no evidence of an excess of small spacings. Table 2.6.2 shows the number of $\delta_n \le 1/20$ and $\le 1/10$ observed in each set (given in terms of cases per million zeros to make comparisons easier). If we consider the $N = 10^{19}$ entry for $\delta_n \le 1/20$, for example, we see that we are dealing with 2353 cases altogether, so a normal sampling error might be around 50, which is about 2%. Thus the 140.5 figure in the table is consistent with the 136.8 expected for the GUE.

Still another way to judge whether there is any anomaly in the distribution of the δ_n or the $\delta_n + \delta_{n+1}$ is to use the quantile-quantile $(q-q)$ plots to compare the observed distributions to those of the GUE. Given a sample $x_1, \ldots, x_n$, and a continuous cumulative distribution function $F(z)$ for some distribution, the $q-q$ plot is obtained by plotting $x_{(j)}$ against q_j, where $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ are the x_i sorted in increasing order, and the q_j are the theoretical quantiles defined by $F(q_j) = (j-1/2)/n$ [CCKT]. The $q-q$ plot is a sensitive method of detecting differences among distributions. In particular, while it does show the outliers that are far away from the expected position, it makes it possible to disregard them and concentrate on the main part of the distribution curve. If the x_j are drawn from the distribution corresponding to $F(z)$, and the sample size n is large, the $q-q$ plot will be close to the straight line $y = x$. In all our $q-q$ plots, straight lines $y = x$ are drawn to facilitate comparisons. (By the standards of typical statistical investigations, the sample sizes we deal with are very large, and the degree of agreement between conjecture and numerical evidence is very good, so one has to look at minute deviations.)

The $q-q$ plots of [Od2] that showed the distribution of small δ_n indicated a deficiency of small δ_n for $N = 1$, and a slight excess for $N = 2 \times 10^{11}$ and $N = 10^{12}$. Those plots were each based on 10^5 values of δ_n. When the new, more extensive data for $N = 10^{12}$ were obtained, the resulting $q-q$ plot was similar to that of Fig. 2.6.1, and did not behave like the plot in Fig. 8 of [Od2] (which was based on only 10^5 zeros). Figures 2.6.1 and 2.6.2 show $q-q$ plots of δ_n drawn from two disjoint sets of 10^6 zeros near zero number 10^{20}. While the plot of Fig. 2.6.1 might indicate a slight excess of small spacings (those in (0.02,0.04), roughly), and a slight deficiency of slightly larger spacings (where the scatterplot lies above the straight line), Fig. 2.6.2 indicates almost perfect agreement between theory and experiment. Figure 2.6.2 is not completely representative of zeros in the $N = 10^{20}$ sets, since it was the one of several $q-q$ plots based on disjoint sets of 10^6 zeros that gave the best agreement. Figure 2.6.1 is more typical in this respect.

Figures 2.6.1 and 2.6.2 provide only a little, if any, support to the theory that there is an excess of small spacings among the zeros. Some further support can be found, however, if we combine all the data from the $N = 10^8$, 10^{19}, and 10^{20} data sets, which contain 112, 314, 006 zeros, and yield 112,314,003 values of δ_n. The resulting $q - q$ plot, shown in Fig. 2.6.3, does indicate a slight excess of small δ_n (the two outliers close to the bottom of the graph are the unusually small δ_n that are minimal in the $N = 10^{18}$ and 10^{19} data sets), but the evidence is not conclusive.

When we consider the other extremal values of δ_n and $\delta_n + \delta_{n+1}$, the evidence is in much better agreement with expectation. The counts in Table 2.6.2 show that the numbers of small $\delta_n + \delta_{n+1}$, large δ_n, and large $\delta_n + \delta_{n+1}$ are all smaller than predicted by the GUE theory, but increasing towards that prediction. The $q - q$ plots of Figures 2.6.4 through 2.6.7 also support this impression; there are too few extreme values in general, but the deficiency is smaller for $N = 10^{20}$ than for $N = 10^{12}$.

Because of (2.3.6), one can expect that among the values of $\delta_n + \delta_{n+1}$ drawn from the GUE, the probability of the minimal value being $\le x$ is about

$$1 - \exp(-\pi^6 x^8 M/32400) \ .$$

The minimal value of $\delta_n + \delta_{n+1}$ of 0.1124 in the $N = 10^{20}$ data set would then occur with a probability of 0.06 in the GUE, while the corresponding probabilities for the $N = 10^{12}$, 10^{14}, 10^{16}, 10^{18}, and 10^{19} data sets are 0.93, 0.78, 0.25, 0.27, and 0.60, respectively. Thus the only one of these figures that might seem unusually small is that for the minimal $\delta + \delta_{n+1}$ for $N = 10^{20}$.

The maximal values of δ_n and $\delta_n + \delta_{n+1}$ recorded in Table 2.6.1 are all somewhat smaller than what the GUE predicts, which is not too surprising given the bounds known to hold for $S(t)$ and $S_1(t)$. For very large spacings in the GUE, des Cloizeaux and Mehta [CM2] have proved that

$$\log p(0,t) \sim -\pi^2 t^2/8 \quad \text{as} \quad t \to \infty \ , \tag{2.6.3}$$

which suggests that

$$\max_{N+1\le n\le N+M} \delta_n \sim \pi^{-1}(8\log M)^{1/2} \tag{2.6.4}$$

as $N, M \to \infty$ with M reasonably large compared to N. This is larger by about a $(\log\log M)^{1/2}$ factor than the conjecture (2.5.2) of Montgomery allows. Our data are too limited to decide whether that conjecture is right.

Values of δ_n and $\delta_n + \delta_{n+1}$ larger than those of Table 2.6.1 have been found in other computations, and are described in Section 3.2. In particular, the largest known values of δ_n and of $\delta_n + \delta_{n+1}$ are 5.1454 and 6.0165, respectively.

Even on the assumption of the RH, it is only known that $\delta_n \le 0.5172$ and $\delta_n \ge 2.337$ each occurs infinitely often [CGG1], and $\delta_n \ge 2.68$ occurs infinitely often on the assumption of the Generalized Riemann Hypothesis for Dirichlet L-functions (or at least of a Generalized Lindelöf Hypothesis) [CGG2]. On the assumption of the RH, it is also known that $\delta_n < 0.77$ and $\delta_n > 1.33$ each holds for a positive proportion of n [CGGGH]. The GUE predicts that $\delta_n < \epsilon$ and $\delta_n > \epsilon^{-1}$ should each hold for a positive proportion of n for every fixed $\epsilon > 0$. If we could prove that $\delta_n < 1/4$ holds for infinitely many n, we could obtain effective bounds for class numbers of imaginary quadratic number fields [MW]. The GUE hypothesis predicts that $\delta_n < 1/4$ for 1.6% of n's, and this is very close to what is observed in numerical data. (For $\delta_n < 1/2$ the corresponding figure is 11.3%.)

2.7. Long and short range correlations between zeros

The distribution of the eigenvalue spacings in the GUE is stationary. What this means is that for any k, the frequency with which $(\delta_n, \delta_{n+1}, \dots, \delta_{n+k}) \in Q$ for any measurable subset $Q \subseteq \mathbb{R}^{k+1}$ does not depend on the range of n, so that the distribution eigenvalue spacings looks every place the same. On the other hand, this distribution is not Markovian, so that the distribution of δ_{n+1} does not depend just on that of δ_n. Instead, δ_{n+1} is correlated to all the neighboring δ_n, $\delta_{n-1}, \dots$, as well as δ_{n+2}, $\delta_{n+3}, \dots$. In the limit, that also should be true for the zeros of the

zeta function. However, given the slow growth rate of $S(t)$, one cannot expect GUE behavior from joint distributions of $\delta_n, \delta_{n+1}, \ldots, \delta_{n+k}$ if k is large. Already the data of sections 2.3 and 2.5 show that the behavior of δ_n is much closer to the GUE prediction than that of $\delta_n + \delta_{n+1}$. That was the main reason for not investigating $\delta_n + \delta_{n+1} + \delta_{n+2}$ and higher order spacings.

When we investigate long range correlations among the zeros of the zeta function, we find phenomena connected not to the GUE, but to the distribution of primes. For example, if we let the autocovariances of a set of δ_n be defined by

$$c_k = c_k(H, M) = \frac{1}{M} \sum_{n=H+1}^{H+M} (\delta_n - 1)(\delta_{n+k} - 1) \ , \tag{2.7.1}$$

then it has been conjectured by F. J. Dyson (unpublished) that in the GUE,

$$c_k \approx -\frac{1}{2\pi^2 k^2} \tag{2.7.2}$$

for $k > 0$, with the *approx* indicating some degree of approximation, not necessarily asymptotic equality as $N, M \to \infty$. This result has not been proved for the GUE, but it is intuitively appealing for both the GUE and the zeros of the zeta function, since it says in effect that a large spacing would lead to smaller spacings nearby (and vice versa), and that this effect would diminish as one considered spacings further and further apart.

What was observed in [Od2] for the δ_n was quite different from the conjecture (2.7.2). Additional data based on the new computations are presented in Table 2.7.1. The $N = 1$ entries come from the [Od2] computations, and have $H = 0$, $M = 10^5$. The $N = 10^{12}$ and 10^{20} entries come from the new computations, and both have $M = 10^6$, with $H = 10^{12} - 6,032$ for the $N = 10^{12}$ column and $H = 10^{20} - 48,776$ for the $N = 10^{20}$ column. (A comparison of the $N = 10^{12}$ entries here with those in Table 6 of [Od2], which are based on 1/10 as many zeros indicates the size of the sampling errors.) For small k, the data in this table support Dyson's conjecture (2.7.2). For higher sets of zeros, the agreement with (2.7.2) extends to slightly higher values of k. However, for large k, we see totally different behavior. If δ_n and

δ_{n+k} were independent, then, since their mean value is 1 and variance is about 1/6, we would expect a sum of 10^6 terms of the form $(\delta_n - 1)(\delta_{n+k} - 1)$ (for $k > 0$) to be about $10^{6/2}/6 \approx 170$, and this would correspond to a value of c_k of 1.7×10^{-4}. The values in Table 2.7.1 for $9,980 \le k \le 10,000$ are usually much larger than that, which shows that there are strong long range correlation between the δ_n. The pattern of signs of the c_k also shows the nonrandom nature of the c_k. The c_k are occasionally positive, and occasionally negative, indicating that for some k, a large δ_n tends to be associated with large δ_{n+k}, while for other k it tends to be associated with small δ_{n+k}.

An explanation for the long range dependencies among the δ_n was proposed in [Od2]. It implies that the observed correlations come from primes through formulas such as that of Landau [Lan1], which says that for any fixed $y > 0$, as $N \to \infty$ we have

$$\sum_{n=1}^{N} e^{i\gamma_n y} = \begin{cases} -\dfrac{\gamma_N}{2\pi} e^{-y/2} \log p + O(e^{y/2} \log N) & \text{if } \ y = \log p^m \ , \\ O(e^{-y/2} \log N) & \text{if } \ y \ne \log p^m \ , \end{cases} \tag{2.7.3}$$

where p denotes a prime and $m \in Z^+$. The above statement assumes the RH, but Landau proved a similar unconditional result. Improvements on Landau's result (with better error terms and more explicit dependencies of the error terms on y) have been obtained by Fujii [Fu5, Fu7] and Gonek [Gon2]. (There are many formulas relating primes and zeros, and the "explicit formulas" of Guinand [Gu1, Gu2] and Weil [We1] are among the most general.)

The paper [Od2] presents the detailed explanation of how Landau's formula (2.7.3) forces the spectrum of the δ_n to consist largely of point masses at frequencies corresponding to prime powers, which then forces the initially unexpected behavior of the c_k that is seen in the tables. This explanation will not be repeated here. We will mention only that while it is not rigorous, it is supported by heuristics and numerical evidence. What we will do now is to check how well Landau's formula (2.7.3) fits with the numerical data. The main interest here is to see just how many

zeros γ_n are needed at various heights to observe the phenomenon of large values occurring at logarithms of prime powers. Some proposals have even been made to use sums like that in (2.7.3) for primality testing and factoring integers. While it seems unlikely that efficient methods could be developed by this approach, it is of some interest to see what happens when one considers a relatively short sum over high zeros.

Let

$$h(y) = \sum_{n=10^{20}+1}^{10^{20}+4\times 10^4} e^{i\gamma_n y} \ . \tag{2.7.4}$$

Figure 2.7.1 shows a graph of $2\log|h(y)|$ for $0 \le y \le 3$. It is instructive to compare this graph with that of Fig. 15 of [Od2], which is drawn on the same scale, but is based on an exponential sum of 4×10^4 zeros starting at zero number $10^{12}+1$. Both graphs show sharp peaks precisely at logarithms of prime powers, and the peaks are visibly higher at primes than at proper prime powers, as predicted by Landau's formula. (The heights of the peaks are not represented too accurately on the graph because of limited sampling.) All the prime powers $< e^3 = 20.09$ are visible. The main difference between the two graphs is that in Fig. 2.7.1 the peaks are slightly lower, and the "noise" region between the peaks has somewhat higher values. Furthermore, the regular patterns seen in the "main" regions of Fig. 15 of [Od2] (which come from sampling at regular intervals a rapidly oscillating function whose frequency and amplitude are changing slowly) is not visible in Fig. 2.7.1. These differences are probably due partly to the errors in the computed values of the γ_n near the 10^{20}-th zero and partly to taking a very short sum. 4×10^4 zeros out of the first 10^{20} is a very small proportion, so it is remarkable that the pattern of Fig. 2.7.1 is as clear as it is, since this is much better than the proved results of [Fu5, Fu7, Gon2, Lan1] might lead one to expect.

Figure 2.7.2 shows a graph of $2\log|h(y)|$, where $h(y)$ is again defined by (2.7.4), but this time over the region $8 \le y \le 8.05$. (This graph is based on 10^4 equally spaced values of y.) The interval from $e^8 = 2980.96$ to $e^{8.05} = 3133.79$ contains

the primes 2999, 3001, 3011, 3019, 3023, 3037, 3041, 3049, 3061, 3067, 3079, 3083, 3089, 3109, 3119, and 3121, and the prime power $5^5 = 3125$. Figure 2.7.2 fails to distinguish between several close pairs of primes. When one graphs a similar sum, but with 10 times as many zeros, as is done in Fig. 2.7.3, all the primes can be distinguished, and even 3125 can be easily discerned.

An elegant measure of long range correlations between zeros was found by Berry [Be4]. If we consider an interval of length $2\pi L(\log(T/(2\pi))^{-1}$ at height T, the expected number of zeros in it equals L. We define the number variance of the zeros by

$$V_T(L) = V_{T,H}(L) = H^{-1} \int_T^{T+H} \left\{ N\left(t + \frac{2\pi L}{\log(t/(2\pi))}\right) - N(t) - L \right\}^2 dt \ . \quad (2.7.5)$$

In the GUE, one has $V_T(L) = G(L)$, with

$$G(L) = \pi^{-2}\{\log(2\pi L) - Ci(2\pi L) - 2\pi L Si(2\pi L) + \pi^2 L - \cos(2\pi L) + 1 + c_0\} \ , \quad (2.7.6)$$

where Ci and Si are the cosine and sine integrals [HMF] and $c_0 = 0.577\ldots$ is Euler's constant. Asymptotically,

$$G(L) \sim \pi^{-2} \log(2\pi L) \quad \text{as} \quad L \to \infty \ , \quad (2.7.7)$$

while

$$G(L) \sim L \quad \text{as} \quad L- > 0 \ . \quad (2.7.8)$$

Gallagher and Mueller [GM] showed that Montgomery's pair correlation conjecture implies $V_T(L) = L - L^2 + o(L^2)$ as $L \to 0$, which is consistent with (2.7.8). (See also [Fu8].) On the other hand, the numerical evidence of [Od2] showed that $V_T(L)$ was small even for moderately large L, and so a relation like (2.7.7) appeared impossible. Motivated by this discovery, by the relations between primes and long range correlation between zeros discussed above, and by his earlier work on eigenvalues of Hamiltonians of chaotic dynamical systems [Be1, Be2, Be3], Berry [Be4] found

heuristic arguments which suggested that for any $\tau \in (0,1)$, and any $L > 0$,

$$V_T(L) \approx G(L) + B_T(L) \ , \tag{2.7.9}$$

where for $U = T(2\pi)^{-1}$,

$$B_T(L) = \pi^{-2} \left\{ 2 \sum_{\substack{p \\ p^r < U^\tau}} \sum_{r=1}^{\infty} \frac{\sin^2(\pi L r (\log p)/(\log U))}{r^2 p^r} + Ci(2\pi L\tau) - \log(2\pi L\tau) - c_0 \right\} , \tag{2.7.10}$$

and p denotes primes. Computations using 10^5 zeros near zero number 10^{12}, using values of L up to 1000, showed excellent agreement between Berry's conjecture (2.7.9) and empirical data, and those results are shown in the graphs in [Be4]. Note that the $\log(L)$ terms in $G(L)$ and $B_T(L)$ cancel out, and so for every fixed L, one can show that there is a positive function $g(L)$ such that

$$G(L) + B_T(L) = g(L) + o(1) \quad \text{as} \quad T \to \infty \ . \tag{2.7.11}$$

Moreover, if τ is held fixed, then it is easy to see that

$$G(L) + B_T(L) \sim \pi^{-2} \log\log T \quad \text{as} \quad T \ , L \to \infty, \tag{2.7.12}$$

(with L growing much more slowly than T), since the arguments of the sine in the definition (2.7.10) of $B_T(L)$ will be asymptotically equidistributed modulo 2π.

The new zeros were used to obtain further data. For $N = 10^{12}$, the number variance $V_T(L) = V_{T,H}(L)$ defined by (2.7.5) was computed with

$$T = \gamma_{n_0}, \qquad n_0 = 10^{12} - 6,032 \ ,$$
$$T + H = \gamma_{m_0}, \qquad m_0 = n_0 + 5 \times 10^5 \ .$$

For $N = 10^{20}$, the values that were chosen were

$$T = \gamma_{n_1}, \qquad n_1 = 10^{20} - 48,778 \ ,$$
$$T + H = \gamma_{m_1}, \qquad m_1 = n_1 + 5 \times 10^5 \ .$$

Berry's function (2.7.10) was computed in each case with $\tau = 1/4$. (Varying τ between 0.2 and 0.3 did not appreciably change the results, as was to be expected.)

The results of some of these computations for $N = 10^{20}$ are presented in Figs. 2.7.4 through 2.7.6. In Fig. 2.7.4, the dashed line is the graph of the GUE prediction $G(L)$, the solid line is the graph of Berry's prediction $G(L) + B_T(L)$, and the scatterplot is that of computed values of $V_T(L)$. In Figs. 2.7.5 and 2.7.6 the graphs of the computed values of $V_T(L)$ and of Berry's prediction $G(L) + B_T(L)$ were both drawn as solid lines, one superposed on the other. The slight differences between the two curves show up as slight blotches on the graph. (The empirical data is slightly more wiggly than $G(L) + B_T(L)$.) We see that even for $L = 5 \times 10^5$, the agreement between computed and predicted values is almost perfect.

A comparison of the graphs of [Be4] (and of similar graphs drawn with the more extensive data that has been obtained for $N = 10^{12}$ in the present computation) with Figs. 2.7.4 through 2.7.6 shows that for $N = 10^{20}$, the number variance oscillates less than for $N = 10^{12}$. The agreement of data with Berry's prediction is better for $N = 10^{20}$.

While Berry's prediction (2.7.9) for $V_T(L)$was based on heuristic arguments, one can prove that a version of the conjecture follows from the RH and the pair correlation conjecture (2.3.11). This will be shown in a separate paper [Od4].

2.8. Lehmer phenomenon

For the RH to be true, $|Z(t)|$ cannot have any relative minima between two consecutive zeros of $Z(t)$. Cases where

$$v_n = \max_{\gamma_n < t < \gamma_{n+1}} |Z(t)| \tag{2.8.1}$$

is very small (so that in a sense the RH is "almost violated") are referred to as Lehmer's phenomenon [Lr2], and provide some of the more interesting heuristics both for and against the RH (cf. [Od3]). In this section we present statistics on the frequency of this phenomenon (which does not have a precise definition).

The zero-locating program printed the largest value of $|Z(t)|$ that had been computed in each stretch of 10^4 zeros. To provide further information, the program was

modified for the $N = 10^{19}$, $N = 10^{20}$, and $N = 2 \times 10^{20}$ data sets to obtain statistical information about the behavior of v_n. Since getting a very good approximation to v_n would have required substantial computing time, what the program computed was the midpoint value

$$w_n = |Z((\gamma_n + \gamma_{n+1})/2)| \ . \tag{2.8.2}$$

When a value of $w_n > 250$ or $w_n < 5 \times 10^{-}4$ was encountered, it was printed together with n, γ_n, and δ_n. (However, w_n was not computed for a total of roughly 100 zeros at the ends of data sets.)

To see how good an approximation w_n was to v_n, the values

$$v_n^* = \max_{1 \le k \le 39} \left| Z\left(\gamma_n + \frac{k}{40}(\gamma_{n+1} - \gamma_n)\right)\right| \tag{2.8.3}$$

as well as of w_n were computed for $n_0 \le n \le n_0 + 8 \times 10^5 - 1$, where $n_0 = 10^{20} + 15,316,087$. Let

$$r_n = v_n^*/w_n \ . \tag{2.8.4}$$

Then the maximal value of r_n that was found was 1.43. Only 755 out of the 8×10^5 values of r_n were > 1.2, while the rms value of $r_n - 1$ was 0.029. Among the 873 values of n for which $\delta_n < 0.1$, the maximal value of r_n was 1.008, and the rms value of $r_n - 1$ was $5.1 \times 10^{-}4$. For the 898 values of n for which $\delta_n > 2.5$, the corresponding numbers were 1.29 and 0.036. For the 244 values of n for which $w_n > 100$, these numbers were 1.137 and 0.032, while for the 1426 values of n for which $w_n < 0.01$, they were 1.072 and 0.0066. Thus in general the values of w_n do provide good approximations to v_n^*, and therefore surely also to v_n. This was to be expected on the basis of the GUE predictions (in particular that the approximation would be exceptionally good when δ_n is small). The size of v_n is determined largely by the few zeros nearest to δ_n (cf. [Hej5, Hej6]), and so under the assumption of the GUE one can make quantitative predictions about the behavior of r_n.

Table 2.8.1 shows the frequency of occurrence of values of $w_n < 5 \times 10^{-}4$ among the approximately 3×10^8 values of n that were checked in the $N = 10^{19}$, $N = 10^{20}$,

and $N = 2 \times 10^{20}$ data sets. The smallest value of w_n that was found there was 1.82×10^{-6}, for $n = 10^{20} + 52,127,155$ and $\delta_n = 0.00263$, while the second smallest was 1.84×10^{-6}.

One might expect, and one does observe empirically, that the Lehmer phenomenon is associated to small values of δ_n. If δ_n is small, then one might expect that w_n is almost proportional to δ_n^2, since zeros other than γ_n and γ_{n+1} ought to contribute multiplicative factors that behave like a power of $\log \gamma_n$ on the average, and are at most $\gamma_n^{o(1)}$ as $n \to \infty$ (assuming the Lindelöf conjecture). Since the probability that $\delta_n \le x$ is about $\pi^2 x^3/9$ for x small (see Sections 2.3 and 2.6), one might conjecture that the probability of $w_n < y$ might be proportional to $y^{3/2}$. This would suggest that among the first n zeros, the smallest w_n might be $n^{-2/3} + o(1)$ as $n \to \infty$. If true, this relation would settle an old question [Ed] about the number of terms in the asymptotic part of the Riemann-Siegel formula that have to be used to separate the zeros; even the old estimate of Titchmarsh [Tit1] with an error term of $O(t^{-3/4})$ would suffice at large heights.

The above heuristic about the behavior of small w_n is supported well by empirical data. Let W denote the set formed by the $N = 10^{19}$ data set and the first 78,893,234 zeros in set $N = 10^{20}$. (In the notation of Table 4.7.1, it is the union of sets i, k, l, m, and n.) In set W, 1976 values of n have $w_n < 5 \times 10^{-}4$. Among these 1976 values, the ratio w_n/δ_n^2 varies between 0.0136 and 8.56, with a mean of 0.608 and a variance of 0.427. Thus the correlation between δ_n^2 and w_n is only modest. On the other hand, these w_n follow almost perfectly the rule conjectured above that the fraction of them that are $< y$ ought to be proportional to $y^{3/2}$. This can be seen by looking at the ratio of the k-th smallest w_n to $5 \times 10^{-4} \times (k/1976)^{2/3}$, which varies between 0.715 and 1.267, with a mean of 1.01 and a variance of 9×10^{-4}, and from looking at a $q - q$ plot of the sorted w_n against $5 \times 10^{-4} \times (k/1976)^{2/3}$. We also find good agreement between this prediction of the behavior of the small w_n and the counts of Table 2.8.1, which are based on all of the zeros in the $N = 10^{19}$, 10^{20},

and 2×10^{20} data sets. Thus on average the influence of the neighboring γ_k cancels out.

The most extreme example of the Lehmer phenomenon that was found during the computations described in this paper occurs for $n = 10^{18} + 12,376,780$, where $w_n = 5.28 \times 10^{-7}$ and $\delta_n = 0.001124$. A graph of $Z(t)$ near this point is given in Fig. 2.8.1. (Figure 2.8.1 shows also what looks like another case of the Lehmer phenomenon near Gram point $n-5$, but in that case the minimum of $Z(t)$ reaches -0.0094, and so it does not qualify under our definition.) A much more detailed view of $Z(t)$ in a small neighborhood of this Lehmer phenomenon is given in Fig. 4.7.1. (That picture plays an important role in the discussion of the validity of the present computations that is presented in Section 4.6.)

The most extreme example of the Lehmer phenomenon that is known was found by van de Lune et al. [LRW2]. For $n = 1,048,449,114$, they discovered that $\delta_n = 0.000310$, while $v_n = 2.2 \times 10^{-7}$ ($\geq w_n$). Since the height of this example is only about the square root of that for $n = 10^{18} + 12,376,780$, it could be argued that the higher example of this paper is even more extreme. However, the δ_n found by van de Lune et al. is by far the smallest of any that are known.

2.9. Large values of $\zeta(1/2+it)$

The largest value of $|Z(t)| = |\zeta(1/2+it)|$ that was encountered by van de Lune et al. [LRW2] in their investigation of the first 1.5×10^9 zeros was 117. Table 2.9.1 lists the largest values of $|Z(t)|$ that were encountered in each of the data sets computed in this paper. The main zero locating program kept track of the largest value of $|Z(t)|$ that had been computed, but did not attempt to do a systematic search for large values. However, since large values are usually associated with large δ_n, the standard zero locating procedure seemed to be quite good at finding the high peaks in $|Z(t)|$. For the $N = 10^{19}$, $N = 10^{20}$, and $N = 2 \times 10^{20}$ data sets, the more careful procedure described in Section 2.8 was employed, which provided

even more reliable statistics. The number of values of n in those two data sets for which w_n (defined by Eq. (2.8.2)) exceeded various thresholds is given in Table 2.9.2. (Section 3 lists some values of t for which $|Z(t)|$ is much larger and which were found by a different procedure.)

The rate of growth of $|Z(t)|$ is one of the most intensively studied problems in the theory of the zeta function, since bounds on it provide estimates on the density of possible zeros away from the critical line. It is easy to show that

$$|Z(t)| \le t^{\alpha+o(1)} \quad \text{as} \quad t \to \infty \tag{2.9.1}$$

with $\alpha = 1/4$. Exponential sum methods were used in the first few decades of this century to show that (2.9.1) holds with $\alpha = 1/6$, and then to successively lower this value of α. (See the Notes to Chapter 5 of [Tit2] for a list of the improvements.) Until recently, the smallest value of α for which (2.9.1) was known to hold was $\alpha = 139/858 = 0.162004\ldots$, proved by Kolesnik [Ko], and there were indications that this result was close to the limit of what the two-dimensional "exponent pair" method that was being used could yield [GK]. However, Bombieri and Iwaniec [BI] have obtained a new method that gave $\alpha = 9/56 = 0.16071\ldots$. This method was then developed by Huxley, Kolesnik, and Watt in a series of papers, and the latest result, proved by Huxley [Hux] is that (2.9.1) holds with $\alpha = 89/570 = 0.15614\ldots$.

The Lindelöf hypothesis is the statement that (2.9.1) holds with $\alpha = 0$. The RH yields a slightly stronger bound [Tit2]

$$|Z(t)| \le \exp(c(\log t)(\log\log t)^{-1}) \tag{2.9.2}$$

for some $c > 0$. On the other side, Balasubramanian and Ramachandra [Bala, BR] have shown that

$$\max_{0\le t\le T} |Z(t)| \ge \exp\left(\frac{3(\log T)^{1/2}}{4(\log\log T)^{1/2}}\right) \tag{2.9.3}$$

if T is large enough and more generally, that if $\eta > 0$, then for $T \ge T(\eta)$ and

$(\log T)^\eta \le H \le T$, we have

$$\max_{T\le t\le T+H} |Z(t)| \ge \exp\left(\frac{3(\log H)^{1/2}}{4(\log\log H)^{1/2}}\right) . \tag{2.9.4}$$

Montgomery [Mon3] has conjectured that (2.9.3) is close to the real rate of growth of $|Z(t)|$.

While the data that was collected about large values of $|Z(t)|$ probably does reflect accurately the behavior of the zeta function in these ranges, it does not help in assessing what the true rate of growth of $|Z(t)|$ is. There are two main problems. One is the relatively small number of zeros that were investigated. Since large values of $|Z(t)|$ are rare, we probably do not even have a good representation of the large values of $|Z(t)|$ for $t < \gamma_n$, $n = 10^{20}$. (This is supported by the results of Section 3, where much higher values were found by special methods.) Another problem in using our data to assess the true growth rate of $|Z(t)|$ arises from the slow approach to its true asymptotic behavior. As is noted in Section 2.12 (see especially Fig. 2.10.1), even $\log|Z(t)|$ in the ranges that have been investigated can be far from its eventual distribution. Furthermore, as was noted in the Introduction, even when one investigates at heights $t \approx 1.5\times 10^{19}$, it is hard to tell the differences in growth rates between various functions. (The situation is not as bleak as might seem from the argument used in the Introduction, since one can use sensitive tools such as ratios of values of a function at different points to estimate its growth rate, but that only helps to a limited extent.) Note that for $t = \gamma_n$, $n = 10^{20}$, the bound (2.9.3) is only 12.9.

Before concluding this section, we present some more statistics on the large values of $|Z(t)|$ that were found in the W data set, which was defined in Section 2.8, and which is a subset of the $N = 10^{19}$ and 10^{20} data sets. In set W, 565 values of n were recorded for which $w_n > 250$. The largest is $w_n = 631.7$ for $n = 10^{20} + 13,704,916$, for which $\delta_n = 3.1428$. (The maximum of $|Z(t)|$ between γ_n and γ_{n+1} is at least 641, and there is no violation of Rosser's rule near γ_n.) Of the 565 values, 94 are associated with violations of Rosser's rule. (Of the 28 values of n for which $w_n > 400$,

7 are associated with violations of Rosser's rule.) The smallest value of δ_n that was found for these 565 values of n was 2.07, and the largest was 4.03.

There was a substantial correlation between δ_n^2 and w_n among these 565 samples in set W. The ratio w_n/δ_n^2 was in the range (19.47, 64.74), with a mean of 35.47 and variance 61.32. However, at large heights one would expect this correlation to diminish, in contrast to the situation for Lehmer's phenomenon (Section 2.8). In the latter case the GUE theories predict that δ_n^2 will occasionally get as small as $n^{-2}/3$, so that the influence of the other zeros (likely to be $n^{o(1)}$ because of the Lindelöf hypothesis and the separation of the other zeros that is predicted by the GUE) will not affect the size of $Z(t)$ very much. On the other hand, the GUE theories predict that $\delta_n^2 = O(\log n)$, and since $Z(t)$ is known to get much larger (cf. (2.9.3)), this must be due to some long-range imbalances in the locations of the zeros. One model for the distribution of $Z(t)$ (first proposed informally by Montgomery, and worked out in detail by Bombieri and Hejhal [BH, Hej5, Hej6]) predicts that at large heights, the size of $Z(t)$ is determined primarily by long "amplitude" waves, which are then modulated by local distributions of zeros. This model predicts that there should be clusters of large values of $|Z(t)|$, and that over wide ranges, w_n ought to depend mostly on the "amplitude" waves, and not on δ_n. That there is a very strong correlation between the large w_n and δ_n^2 in our data might therefore indicate that we are not yet seeing the true asymptotic behavior.

2.10. Moments of $\zeta(1/2 + it)$

It is conjectured that for every $\lambda \ge 0$,

$$\lim_{T\to\infty} T^{-1}(\log T)^{-\lambda^2} \int_0^T |Z(t)|^{2\lambda} dt = c(\lambda) \tag{2.10.1}$$

exists, with $c(\lambda) > 0$ for all λ. A proof of this conjecture, or even of some much weaker bound, would be very important, since it would prove the Lindelöf conjecture. However, this conjecture is only known to be true for $\lambda = 0$ with $c(0) = 1$ (trivial), $\lambda = 1$ with $c(1) = 1$, and $\lambda = 2$ with $c(2) = (2\pi^2)^{-1}$ (see the Notes for Chapter 7

in [Tit2] for detailed information and references). No specific values have been conjectured for $c(\lambda)$ in general, but under the assumption of the RH, Conrey and Ghosh [CG1] have shown that $c(\lambda) \geq c_1(\lambda)$, where

$$c_1(\lambda) = \Gamma(1+\lambda^2)^{-1} \prod_p \left\{ \left(1-\frac{1}{p}\right)^{\lambda^2} \sum_{m=0}^{\infty} \left(\frac{\Gamma(m+\lambda)}{m!\Gamma(\lambda)}\right)^2 p^{-}m \right\} , \qquad (2.10.2)$$

and since $c(\lambda) = c_1(\lambda)$ for $\lambda = 0$ and 1, they suggested that perhaps $c(\lambda) = c_1(\lambda)$ for all $\lambda \in [0,1]$. Since $c_1(2) = (4\pi^2)^{-1} = c(2)/2$, equality of $c(\lambda)$ and $c_1(\lambda)$ is unlikely outside the range $0 \leq \lambda \leq 1$. (There is a mistake on this point in the Notes to Chapter 7 of [Tit2].) Conrey and Ghosh [**?**] have shown that the derivatives of $c_1(\lambda)$ and $c(\lambda)$ with respect to λ agree at $\lambda = 0$ and 1. Also, for $0 \leq \lambda < 2$, Heath-Brown [HB2] has shown under the assumption of the RH that if $c(\lambda)$ exists, it is not much larger than predicted by the Conrey-Ghosh conjectures.

One purpose of this section is to provide some numerical evidence about possible values of $c(\lambda)$. One might expect that if

$$r(\lambda, T, H) = H^{-1}(\log T)^{-\lambda^2} \int_T^{T+H} |Z(t)|^{2\lambda} dt , \qquad (2.10.3)$$

then $r(\lambda, T, H) \sim c(\lambda)$ as $T \to \infty$, if H grows sufficiently fast with T while λ is held fixed. Table 2.10.1 presents some values of $r(\lambda, T, H)$ computed for $T = \gamma_{n_0}$ with $n_0 = 10^{20} + 47,098,588$ and $T + H = \gamma_{n_1}, n_1 = n_0 + 10^6$. Each of the 10^6 gaps between consecutive zeros was divided into 40 intervals, $Z(t)$ was evaluated at the endpoints of these subintervals, and Simpson's rule was applied to estimate the integral. Variations on this procedure showed that it produced estimates that were accurate to at least three decimal places (and more for high moments, as Simpson's rule is least accurate for small λ, where the function has singularities at zeros that are hard to deal with). However, the values in the tables, especially for large λ, have to be used with caution because even an interval of 10^6 zeros around zero number 10^{20} is too small to be truly representative. For example, similar data was obtained for $T = \gamma_{n_2}$ with $n_2 = 10^{20} + 15,316,087$ and $T + H = \gamma_{n_3}$, $n_3 = n_2 + 8 \times 10^5$, and also for $T = \gamma_{n_4}$, $n_4 = 10^{20} - 15,409,244$, $T + H = \gamma_{n_5}$, $n_5 = n_4 + 10^6$. For $\lambda = 1$,

the values found there differed by less than 0.5% from those in Table 2.10.1, but for $\lambda = 2.5$ these values were 1.20 and 0.752 times those in Table 2.10.1, respectively. The problem is that high moments are determined largely by the few exceptionally large values of $Z(t)$, and those are rare. (See the next section for some further evidence of this.) To get a good sample, for large λ, one would need to integrate $|Z(t)|^{2\lambda}$ over much longer intervals.

The data in Table 2.10.1 are consistent with the Conrey-Ghosh conjectures that $c(\lambda) = c_1(\lambda)$ for $0 \le \lambda \le 1$. However, given the differences between the empirical data for $\lambda = 1$ and $\lambda = 2$ and the known asymptotic values, it is hard to draw any definitive conclusions. For $\lambda = 1$, estimates of the second moment of $Z(t)$ are known that are better than (2.10.1). They are of the form

$$\int_0^T Z(t)^2 dt = T(\log T - 1 - \log(2\pi) + 2c_0) + E(T) \ , \tag{2.10.4}$$

where c_0 denotes Euler's constant ($= 0.577215\ldots$), and $|E(T)| = O(T^\alpha)$ for various $\alpha < 1/3$. (The best current value of α is $139/429 + o(1)$ as $T \to \infty$, due to Kolesnik [Ko] and in a slightly sharper form to Hafner and Ivić [HI]. Note that $139/492 = 0.3240\ldots$) If we let $r^*(\lambda, T, H)$ be defined similarly to $r(\lambda, T, H)$, but with $\log T$ in (2.10.3) replaced by $\log T - \log(2\pi) + 2c_0$, we find that for the values of T and H that were used to compute Table 2.10.1, $r^*(1, T, H) = 1.004$, which is closer to the asymptotic value $c(1) = 1$ than the value of $r(1, T, H) = 0.989$. (The other two sets of values that were considered give $r^*(1, T, H) = 1.0003$ and 0.9995, respectively.) Thus a major problem in using empirical data is that we do not have good conjectures about asymptotics of moments of $Z(t)$, and that second order terms in those asymptotics are likely to be only slightly smaller than the main terms. (See also Section 2.11 on deviations between observed and expected behavior of $Z(t)$.)

Some data were obtained also about the negative moments of $|Z(t)|$. Table 2.10.2 shows some values of

$$\frac{1}{H}\int_T^{T+H} |Z(t)|^{-2\lambda} dt$$

for T and H as in Table 2.10.1. (The values for $T = \gamma_{n_2}$, $T+H = \gamma_{n_3}$, were essentially identical.) They were obtained by applying Simpson's rule to the inner 38 subintervals in every gap between consecutive zeros, and approximating $|Z(t)|$ by a linear function on the two outer subintervals.

Conrey and Ghosh [CG2] have shown (assuming the RH) that

$$\frac{1}{M}\sum_{m=1}^{M} \max_{\gamma_m < t < \gamma_m+1} Z(t)^2 \sim \frac{1}{2}(e^2-5)\log(\gamma_M/(2\pi)) \tag{2.10.5}$$

as $M \to \infty$. Since $c(1) = 1$, this means that on average $Z(t)^2$ at its maxima is $1+\frac{1}{2}(e^2-7) = 1.1945\ldots$ times the average of $Z(t)^2$ over the entire range $0 < t \le \gamma_M$. (This surprisingly small factor of 1.1945... occurs because the values of $Z(t)^2$ at the critical points where they achieve their maxima are not weighted by the lengths of the intervals on which the maxima are computed. Large values of $Z(t)$ are usually associated to large gaps between consecutive zeros.) Computation over the range from $T = \gamma_{n_2}$ to $T+H = \gamma_{n_3}$ yielded a value of 1.224... instead of the asymptotic value of $1.1945\ldots$. (The value 1.224... is probably a slight underestimate of the correct ratio, since the actual maxima were not determined, but the largest of the values at the 40 evenly spaced points was used.)

Gonek [Gon1] has shown, again assuming the RH, that

$$\frac{1}{M}\sum_{m=1}^{M} Z(\gamma_m + i\alpha\Delta)^2 \sim \left(1 - \left(\frac{\sin\pi\alpha}{\pi\alpha}\right)^2\right)\log(\gamma_M/(2\pi)) \tag{2.10.6}$$

as $M \to \infty$, where $\Delta = 2\pi(\log(\gamma_M/(2\pi)))^{-1}$. Computations for $\alpha = 0.1, 0.2, \ldots, 0.9$ and over the zeros numbered $n_4, n_4+1, \ldots, n_5-1$ showed reasonably good argument, but with the ratio of empirical data to Gonek's asymptotic estimate declining by 4% as α goes from 0.1 to 0.9.

2.11. Distribution of values of $\zeta(1/2+it)$

Since

$$\log\zeta(1/2+it) = \log|Z(t)| + \pi i S(t)\ ,$$

it is not surprising that methods that yield the distribution of $S(t)$ should give corresponding results for $\log|Z(t)|$. Selberg in unpublished manuscripts studied mean values of $(\log\zeta(1/2+it))^h(\log\zeta(1/2-it))^k$ for nonnegative integers h and k, and his results imply, for example, that for rectangles E in R^2

$$\lim_{T\to\infty}\frac{1}{T}\left|\left\{t:\ T\le t\le 2T, \frac{\log\zeta(1/2+it)}{(2^{-1}\log\log T)^{1/2}}\in E\right\}\right| = (2\pi)^{-1}\iint_E e^{-(x^2+y^2)/2}dxdy\ , \tag{2.11.1}$$

so that in particular, for any $\alpha<\beta$,

$$\lim_{T\to\infty}\frac{1}{T}\left|\left\{t:\ T\le t\le 2T,\ \alpha<\frac{\log|Z(t)|}{(2^{-1}\log\log T)^{1/2}}<\beta\right\}\right| = (2\pi)^{-1/2}\ \int_\alpha^\beta e^{-x^2/2}dx\ . \tag{2.11.2}$$

Thus the real and imaginary parts of $\log\zeta(1/2+it)$ behave like independent normal variables with means 0 and variances $(\log\log t)/2$. While Selberg's results have not been published, they were known to some mathematicians (see [Hej6, Joy1, Jut, Mon6], and some extensions of Selberg's results have been obtained by Joyner [Joy1] and Tsang [Ts2]. The weaker result (2.11.2) has been reproved by Laurinchikas [Lau1, Lau2, Lau3, Lau4, Lau5].

The critical issue here is whether the approximation (2.11.2) is accurate even for T fixed and α and β varying over wide ranges. If that is the case, then we are led to expect that something like (2.10.1) holds. Furthermore, if the approximation is accurate even for α and β relatively large (compared to T), one would expect that the maximal size of $|Z(t)|$, for $0\le t\le T$, would be $\exp((\log T)^{1/2+o(1)})$, which is conjectured by some to be the true rate of growth of $Z(t)$ (cf. Section 2.9). Thus it is of substantial interest to find out more about the tails of the distribution of $\log|Z(t)|$.

For $n_0\le n\le n_1-1$, $n_0=10^{12}-6032$, $n_1=n_0+10^6$, each interval (γ_n,γ_{n+1}) was partitioned into 40 equal subintervals, $Z(t)$ was evaluated at the endpoints of these subintervals, and a linear approximation to $Z(t)$ between consecutive evaluation

points was used to estimate

$$b_{\alpha,\beta} = \frac{1}{\gamma_{n_1} - \gamma_{n_0}} |\{t : \gamma_{n_0} \le t \le \gamma_{n_1}, \alpha \le \log|Z(t)| \le \beta\}| \tag{2.11.3}$$

for $\beta = \alpha + 1/100$, $\alpha = k/100$, $-1000 \le k < 1000$. The mean of this distribution (as derived from the $b_{\alpha,\beta}$ data) was $5.29 \times 10^{-}4$ and the variance was 2.2930. (In Fig. 2.11.1 it is labelled as the $10 = 10^{12}$ distribution.) Similar data was obtained for $n_2 \le n \le n_3 - 1$, $n_2 = 10^{20} + 15,316,087$, $n_3 = n_2 + 10^6$, and there the mean was $5.20 \times 10^{-}4$ and the variance was 2.5657. (This is the $N = 10^{20}$ distribution.) Based on (2.11.2), one would expect mean values of 0, which is very close to the calculated values, given the errors in the computation and sampling errors. The values for the variances would be expected to be $(\log\log T)/2$, where T is the height of the data set, which equals 1.635 and 1.894 for the two data sets, respectively. Since $(\log\log T)/2$ is only the asymptotic value and increases very slowly, lower order terms can be expected to be significant, and so the agreement between observed data and theory is reasonably good on this point as well. However, the shapes of the observed distributions of $\log|Z(t)|$ appear to be different from the asymptotic normal distribution. To obtain a good comparison, the two distributions for $N = 11^{12}$ and 10^{20} were each scaled to have variance 1, and were plotted in Fig. 2.10.1 together with the standard normal distribution. We see that while the fit of the $N = 10^{20}$ data is slightly better than that for $N = 10^{12}$, it is not much better. This is in great contrast to the fit of the data for $S(t)$ (which, apart from a factor of $1/pi$, is the imaginary part of $\log\zeta(1/2+it)$, while $\log|Z(t)|$ is the real part of it) which, as we see in Section 2.5 and Fig. 2.5.1, is much better. It might be of some interest to compute second order terms in the expansion of moments of $\log|Z(t)|$ to see what is responsible for the deviations from the asymptotic behavior that are visible in the data. In view of Goldston's results [Go2] (mentioned in Sections 2.5 and 2.7), it seems likely that such higher order terms depend on the pair correlation of zeros, and even on higher order correlations.

The area between the empirical distribution curve for $N = 10^{12}$ in Fig. 2.11.1

and the normal curve is 0.132, while for $N = 10^{20}$ the corresponding area is 0.114. In both cases these areas are much larger than those for the distribution curves for $S(t)$ discussed in Section 2.5, which confirms the impression one obtains by comparing Fig. 2.5.1 to Fig. 2.11.1.

Table 2.11.1 presents extensive data on the moments of $\log|Z(t)|$. The six sets of data summarized in this table were all obtained by choosing 10^6 random points in an interval of length 1.5×10^5. For $N = 10^{12}$, this interval started near zero number $10^{12} - 6,032$. For $N = 10^{18}(a)$ and $N = 10^{18}(b)$, the intervals were the same, starting near zero number $10^{18} - 8,839$ but the random sequences were different, since different seeds were chosen for the pseudorandom number generator. This was done to estimate the size of the sampling error. For $N = 10^{20}(c)$, the starting point was near zero number $10^{20} - 48,776$, while for $N = 10^{20}(d)$, it was near zero number $10^{20} + 15,316,087$. The mean and second moment for each data set are shown in the $k = 1^*$ and $k = 2^*$ entries, respectively. These were then applied to translate and scale the data sets to obtain mean equal to 0 and variance equal to 1, for ease of comparison with the standard normal distribution. The k-th entry in the table, $1 \le k \le 10$, given the k-th moment of each scaled data set, and the last column gives the corresponding value for the normal distribution (0 for k odd, $(k-1)\cdot(k-3)\cdot\ldots\cdot 3\cdot 1$ for k even).

Given that the distribution of $\log|Z(t)|$ differs so much from the expected normal one, we have to treat the data about moments of $|Z(t)|$, for example, with extreme caution, as they may not be representative of the true asymptotic behavior. Furthermore, the general distribution of $Z(t)$ may be even further from what happens higher up.

Figure 2.11.2 presents some empirical data on values of $Z(t)$. This figure is based on the values of $Z(t)$ in the three intervals covering 2.8×10^6 zeros that were described in the preceding section. For each interval between consecutive zeros, the function $|Z(t)|$ was approximated on 40 equal-sized subintervals by a linear function,

and the length of the interval on which this linear approximation was in each range $[k-1,k)$ for $k \geq 0$ was computed. If A_k denotes the length of all the intervals on which the linear approximations were in $[k-1,k)$, and

$$q_k = \frac{A_k}{\sum_{j=1}^{\infty} A_j}$$

the fraction of time spent there, then the plot in Fig. 2.11.2 shows $\log q_k$. From this graph and other graphs based on the data from each of the three main intervals separately, it appears that for $k <\sim 150$, the empirical data obtained so far represents well the long-run statistics of $|Z(t)|$ at the heights that were investigated. In particular, the curves of $\log q_k$ for the three main intervals are almost identical in that range. On the other hand, for $k >\sim 250$, the behavior of $\log q_k$ is dominated by a few large peaks of $|Z(t)|$ (which also account for a large part of the values of high moments of $Z(t)$ dealt with in the previous section). In particular, the segments of the graph in Fig. 2.11.2 that shoot up are caused by high peaks. The final region ($k \geq 353$) is due to two peaks where $|Z(t)|$ reaches about 460, and the preceding region of increase in $\log q_k$ is due to a point where $|Z(t)|$ is around 351.

2.12. Values of $\zeta'(1/2+i\gamma)$

Under the assumption of the RH and of a weak consequence of the pair correlation conjecture, namely that for some $\tau > 0$, there is a constant B such that

$$\limsup_{N\to\infty} \frac{1}{N} \Big| \{n : N \leq n2N,\ \delta_n < c\} \Big| \leq Bc^{\tau} \tag{2.12.1}$$

holds uniformly for all $c \in (0,1)$, Hejhal [Hej6] has shown that for all $\alpha < \beta$,

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ n : N \leq n \leq 2N,\ \frac{\log \left| \frac{2\pi Z'(\gamma_n)}{\log(\gamma_n(2\pi)^{-1})} \right|}{(2^{-1}\log\log N)^{1/2}} \in (\alpha,\beta) \right\} \right| = (2\pi)^{-1/2} \int_{\alpha}^{\beta} e^{-x^2/2} dx\ . \tag{2.12.2}$$

(Note that under the RH, which we assume throughout this section, $|\zeta'(\rho)| = |Z'(\gamma)|$ if $\rho = 1/2 + i\gamma$.)

As was the case with the values of $Z(t)$, we would like to obtain more information about the tails of the distribution of $Z'(\gamma_n)$, and in particular about the moments. Let us define

$$J_\lambda(T) = \sum_{\gamma_n \le T} |Z'(\gamma_n)|^{2\lambda} . \tag{2.12.3}$$

Then $J_\lambda(T)$ exists for all $\lambda \ge 0$, and if the zeros of the zeta function are all simple (as they are conjectured to be, and as is the case with all of the zeros that have been dealt with numerically) then $J_\lambda(T)$ also exists for $\lambda < 0$. The only nontrivial asymptotic result was proved by Gonek [Gon1] under the assumption of the RH;

$$J_1(T) \sim \frac{T}{24\pi}(\log T)^4 \quad \text{as} \quad T \to \infty .$$

It is clear that $J_0(T) = N(T) \sim \frac{T}{2\pi}\log T$ as $T \to \infty$, and it is known (cf. [Tit2, Section 14.27]) that $J_{-1/2}(T)/T \to \infty$ as $T \to \infty$. Gonek [Gon3] has also shown that (under the RH) $J_{-1}(T) \ge cT$ for some $c > 0$. If the limit law (2.12.2) holds even for small N, and the tails of the distribution of $Z'(\gamma_n)$ are not too large, then we might expect (as was suggested by Hejhal [Hej6] and stated explicitly by Gonek [Gon3]) that $J_\lambda(T)$ is on the order of

$$T(\log T)^{(\lambda+1)^2} \quad \text{as} \quad T \to \infty . \tag{2.12.4}$$

Furthermore, Gonek [Gon3] has conjectured that

$$J_{-1}(T) \sim 3\pi^{-2}T \quad \text{as} \quad T \to \infty . \tag{2.12.5}$$

Approximate values were obtained for $|Z'(\gamma_n)|$, $n_0 \le n \le n_0 + 10^6 - 1$, where $n_0 = 10^{20} + 15,316,107$. Since the behavior of $Z(t)$ is determined primarily by zeros close to t (cf. [BH, Hej6]), it was assumed that for t near γ_n, $Z(t)$ is approximated well by

$$a \prod_{j=-20}^{20} (t - \gamma_n + j) , \tag{2.12.6}$$

where a, representing the influence of zeros further from γ_n, is almost a constant, and this led to approximating $Z'(\gamma_n)$ by

$$\epsilon^{-1} Z(\gamma_n+\epsilon) \prod_{\substack{j=-20\\ j\neq 0}}^{20} \frac{\gamma_n+j}{\gamma_n+\epsilon-\gamma_n+j}\ , \tag{2.12.7}$$

where $\epsilon = (\gamma_{n+1}-\gamma_n)/40$. Varying the number of terms in the heuristic approximation (2.12.6) as well as varying ϵ showed that (2.12.7) does produce good approximation to $Z'(\gamma_n)$.

The smallest value of $|Z'(\gamma_n)|$ that was found was 0.13, while the largest was 2.47×10^3. The values of $\log|Z'(\gamma_n)|$ had a mean of 3.35 and a variance of 1.14, in contrast to 1.91 and 1.9, respectively, which are predicted by Hejhal's result (2.12.2). Given the slow rate of growth of these quantities, second order terms in the asymptotic results are likely to be large, so this difference between expected and observed values is probably not significant. If we let

$$v_n = (\log|Z'(\gamma_n)| - m)/\sigma\ , \tag{2.12.8}$$

where m is the mean and σ the standard deviation of our set of $\log|Z'(\gamma_n)|$, then Fig. 2.12.1 shows a comparison of the distribution of v_n with the standard normal distribution. The line is the standard normal density, while the scatterplot represents a histogram of v_n; for each interval $[\alpha,\beta)$, $\alpha = k/50$, $\beta=\alpha+1/50$, a star is placed at (x,y), $x=((\alpha+\beta)/2-m)/\sigma$, $y=\sigma b_{\alpha,\beta}$, where

$$b_{\alpha,\beta} = 5010^6\,|\{n:\ v_n\in[\alpha,\beta)\}|\,. \tag{2.12.9}$$

It is worth noting that the distributions of $\log|Z(t)|$ and $\log|Z'(\gamma_n)|$ are both supposed to be asymptotically normal, but the convergence appears much faster for $\log|Z'(\gamma_n)|$, as is revealed by a comparison of Fig. 2.12.1 to Fig. 2.11.1. This is true even though the asymptotic normality of $\log|Z(t)|$ is an unconditional theorem, while that of $\log|Z'(\gamma_n)|$ depends on unproved assumptions.

Table 2.12.1 lists the moments of the v_n and of the asymptotic normal distribution. The entry for $k=1,\dots,10$ denotes the k-th moment of v_n and the normal

distribution, while the $k = 1^*$ and $k = 2^*$ entries give the first two moments of $\log|Z'(\gamma_n)|$, respectively. Comparison with Table 2.11.1 again shows much better agreement between empirical and expected values for $\log|Z'(\gamma_n)|$ than for $\log|Z(t)|$.

Moments of $Z'(\gamma_n)$ for the 10^6 values that were computed are shown in Table 2.12.2. Since (2.12.4) suggests that for M relatively large,

$$J_\lambda^*(M,N) = \frac{1}{M}\sum_{n=N+1}^{N+M}|Z'(\gamma_n)|^{2\lambda} \tag{2.12.10}$$

ought to be of the order of magnitude of

$$(\log\gamma_N)^{(\lambda+1)^2-1}\ ,$$

while for $\lambda = 1$ and $\lambda = -1$ we ought to have the more precise relations

$$J_1^*(M,N) \sim \frac{1}{12}(\log\gamma_N)^3\ , \tag{2.12.11}$$

$$J_1^*(M,N) \sim 6\pi^{-2}(\log\gamma_N)^{-1} \tag{2.12.12}$$

(the asymptotic relations holding as $M, N \to \infty$ with M relatively large). Table 2.12.2 shows the ratio of empirical to expected values, namely

$$r_\lambda = J_\lambda^*(10^6, n_0-1)(\log\gamma_{n_0})^{1-(\lambda+1)^2}\ . \tag{2.12.13}$$

The value for $\lambda = 1$ is in excellent agreement with (2.12.11) (which is a theorem under the assumption of the RH), while the value for $\lambda = -1$ is reasonably consistent with that of (2.12.12). Since (2.12.12) is derived from Gonek's conjecture (2.12.5), this supports the conjecture.

A theorem announced by Fujii [Fu4] (which assumes the RH) states that

$$\begin{aligned}\sum_{0<\gamma_n\le T}\zeta'(1/2+i\gamma_n) = \frac{T}{4\pi}\log^2\frac{T}{2\pi} + c_0\frac{T}{2\pi}\log\frac{T}{2\pi}\\ +c_1\frac{T}{2\pi} + O(T^{9/10+o(1)})\end{aligned} \tag{2.12.14}$$

as $T\to\infty$, where c_0 and c_1 are explicit constants. This turns out to be in excellent agreement with the empirical result

$$\sum_{n=n_0}^{n_0+10^6-1}\zeta'(1/2+i\gamma_n) = 2\ .181\times 10^6 + i8.7\times 10^3. \tag{2.12.15}$$

The approximate procedure (2.12.7) that was used to estimate $Z'(\gamma_n)$ can be replaced by a much more rigorous and accurate method. The algorithm of citeOS that was used to compute $Z(t)$ precomputes a set of values from which $Z(t)$ is obtained by interpolation. However, the main interpolation formula (4.4.15) can be differentiated with respect to t, which enables one to compute $Z'(t)$ (and therefore also $\zeta'(1/2+it)$) from the basic data. If such a program were written, it could be used to check the speed of convergence of the distribution of $\log|\zeta'(1/2+it)|$ to the Gaussian limit that has been shown to hold under the assumption of the RH by Hejhal [Hej6].

2.13. Gram points and blocks

Gram's law is the empirical observation that $Z(t)$ usually changes sign in each *Gram interval* $G_n = [g_n, g_{n+1})$, $n \geq -1$. (The Gram points g_n are defined in Section **??**.) Gram [**?**] observed that it held in the range of values he investigated, but he conjectured that it would fail eventually. The first counterexample occurs for $G_1 25$, and was discovered by Hutchinson [**?**]. If Gram's law held universally, the RH would obviously be true. However, it is known that this "law" fails infinitely often. On the other hand, it does hold for a large fraction of cases. For $n \leq 1.5 \times 10^9$, Gram's law holds 72.79% of the time [LRW2], among 10^6 Gram intervals near zero number 10^{12}, it holds 70.82% of the time, and among 10^6 Gram intervals near zero number 10^{20}, it holds 68.9% of the time. (Under the GUE and some further assumptions to be discussed later, one might expect that asymptotically, Gram's law should hold about 66.3% of the time.)

One barely plausible reason Gram's law might hold (and why the RH might hold) is that in the Riemann Siegel formula for $Z(t)$ (see Eq. (4.2.2)) the leading term equals $2(-1)^n$ at $t = g_n$. If this term, which is the largest, were always dominant, then Gram's law and the RH would follow. We now know this to be false, but there is still interest in the behavior of $Z(t)$ at Gram points, since sign changes of $Z(t)$

correspond to zeros of the zeta function on the critical line.

A Gram point g_n is called *good* if $(-1)^n Z(g_n) > 0$, and *bad* otherwise. A *Gram block* is an interval $B_n = [g_n, g_{n+k})$ such that g_n and g_{n+k} are good Gram points, while $g_{n+1}, \ldots, g_{n+k-1}$ are bad Gram points. The *length* of a Gram block $B_n = [g_n, g_{n+k})$ is k. The *pattern of zeros* in a Gram block $B_n = [g_n, g_{n+k})$ is the string $a_1 \cdots a_k$, where a_i denotes the number of zeros of $Z(t)$ in $[g_{n+i-1}, g_{n+i})$. Since no Gram interval with more than 4 zeros has even been found, writing $a_1 \cdots a_k$ without comma separators is unambiguous. (Gram intervals with arbitrarily many zeros almost surely exist, but given the GUE predictions about zeros repelling each other, they are likely to be rare.)

The statistics that have been collected on Gram intervals and blocks (as well as on exceptions to Rosser's rule, which are discussed in Section **??**) are subject to errors, not only because of the roundoff problems that have been mentioned before and are discussed extensively in Section 4, but also because even if the computations of $Z(t)$ were exact, Gram points were determined only approximately, so that the determinations of the signs of $Z(g_n)$ were not always certain. No special precautions were taken to deal with this problem (such as checking on the size of the computed value of $Z(g_n)$) as it was felt that this was unlikely to affect general statistics.

The computations of [LRW2] of the first 1.5×10^9 zeros found only 6 Gram blocks of length 9, and none of lengths ≥ 10. In contrast, the maximal lengths of Gram blocks found during the present computations were 9 for $N = 10^{12}$, 9 for $N = 10^{14}$, 11 for $N = 10^{16}$, 13 for $N = 10^{18}$ (1 case), 12 for $N = 10^{19}$, 14 for $N = 10^{20}$ (2 cases, with zero patterns 01111111113110 and 01111111111130), and 13 for $N = 2 \times 10^{20}$.

Table 2.13.1 gives the fraction of Gram blocks in given data sets with given lengths. The $N = 1$ and $N = 1.4 \times 10^9$ data is derived from Table 1 of [LRW2], and comes from two sets of 10^8 Gram intervals each, the first one starting at g_0, the second at g_n for $n = 1.4 \times 10^8$. The $N = 10^{12}$ data is based on only 1,590,000 Gram intervals.

The main program did not keep track of Gram blocks according to their pattern of zeros. However, a special study was made of two blocks of 10^6 Gram intervals each, one starting at g_{n_1}, $n_1 = 10^{12} - 6,034$, the other at g_{n_2}, $n_2 = 10^{20} - 42,780$, which for the remainder of this section will be referred to as the $N = 10^{12}$ and $N = 10^{20}$ data sets, respectively.

If a Gram block $B(n,k)$ contains exactly k zeros (so is not associated with a violation of Rosser's rule, see Section 2.14) then its zero pattern must be either 211...110, or 011...112, or 011...131...110 (where the data ... refer to any strings of consecutive 1's, but these strings might be even shorter than indicated). pattern might be missing). Van de Lune et al. [LRW2] noted in their computations that for a fixed k, the first two zero patterns seemed to be much more frequent than the third one, and that the frequencies seemed stable as the height of zeros increased. The new computations, however, show a steady decrease in the frequency with which the third pattern appears as the height increases. Table 2.13.2 shows the observed frequencies. The $N = 1$ entry is drawn from Table 2 of [LRW2], which is based on statistics of 3 sets of 10^8 Gram intervals each, starting at g_m with $m = 0$, 7×10^8, and 1.4×10^9, respectively. Only Gram blocks of length k with exactly k zeros are considered, and the entry in the table gives the fraction of all such Gram blocks that have a zero pattern containing a 3. The decrease in the frequency of the third zero pattern is puzzling. The GUE theories suggest that this pattern ought to occur a positive proportion of the time.

Table 2.13.3 presents data on the fraction of Gram intervals that contain a given number of zeros. The $N = 1$ and $N = 1.4 \times 10^9$ data sets are the same as in Table 2.13.1, and these entries come from Table 5 of [LRW2]. Note that there were no Gram intervals with ≥ 4 zeros in the $N = 10^{12}$ and $N = 10^{20}$ sets (although such intervals did turn up in other data sets around the 10^{20}-th zero, for example).

The GUE entry in Table 2.13.3 was derived by assuming that a Gram interval does not differ from any other interval of that length, and so the entry in the table

for a given m in the GUE row is the probability that an interval of length 1 contains exactly m zeros. Since the averages of $S(t)$ do increase as t increases, it seems reasonable to expect that at large heights the local distribution of the zeros will be independent of Gram points, which leads to the above assumption (cf. [Fu4]). In other words, the expectation is that at large heights, any grid of points spaced like the Gram points would exhibit similar behavior with respect to location of zeros.

If the zeros at large heights are distributed independently of the Gram points, in the sense above, namely that shifting all the Gram points in a large interval by the same amount would not affect the statistics of Gram intervals and blocks, then we can expect that if we define

$$z_n = \frac{\gamma_n - g_m}{g_m + 1 - g_m} \quad \text{if} \quad \gamma_n \in [g_m\ , g_{m+1}),$$

then the z_n will be distributed uniformly in the unit interval [Fu4]. (So far, the equidistribution of the γ_n has been shown only modulo much coarser grids, see [Hl, Fu3].) Figure 2.13.1 shows the distribution of z_n for the two data sets $N = 10^{12}$ and $N = 10^{20}$ as well as for a third set, labelled $N = 10^6$, derived from the 10^6 zeros γ_n with $10^6 + 1 \le n \le 2 \times 10^6$. In each case a histogram was prepared giving the number of $z_n \in [j/1000,\ (j+1)/1000)$, $0 \le j < 10^3$, and these data were used to derive the smooth curve in the picture using the lowess function of [BC]. A perfectly uniform distribution would correspond to a straight horizontal segment at height 1, while the most nonuniform distribution (which also would minimize the moments of $|S(t)|$), corresponds to a point mass at $1/2$ and 0 elsewhere. The $N = 10^{20}$ curve is much closer to this conjectured uniform behavior than the $N = 10^{12}$ curve, and neither is far away from it. Even the $N = 10^6$ curve is not very far from uniform behavior. The area between the curve in Fig. 2.13.1 and the straight horizontal segment at height 1 is 0.105 for $N = 10^6$, 0.051 for $N = 10^{12}$, and 0.028 for $N = 10^{20}$. Thus is appears to be of the order of $(\log N)^{-1}$.

A quantitative study of the extent to which the sign of $Z(g_n)$ might coincide with $(-1)^n$ was started by Titchmarsh [Tit0], who showed (as might be expected

from the Riemann-Siegel formula (4.2.2)) that as $M \to \infty$,

$$M^{-1} \sum_{n=1}^{M} Z(g_n) = o(1) , \tag{2.13.1}$$

$$M^{-1} \sum_{n=1}^{M} (-1)^n Z(g_n) = 2 + o(1) , \tag{2.13.2}$$

as well as

$$M^{-1} \sum_{n=1}^{M} Z(g_n) Z(g_{n+1}) = -2(1 + c_0) + o(1) \tag{2.13.3}$$

where $c_0 = 0.577\ldots$ is Euler's constant. These results have been strengthened and extended considerably by Moser [Mos1, Mos2, Mos3, Mos4, Mos5, Mos6, Mos7, Mos8, Mos9, Mos11, Mos12, Mos14]. Table 2.13.4 presents some averages involving the $Z(g_n)$ that were computed, using the 2 sets of 10^6 values each that were specified above. For example, the $|Z^3(g_n)|$ entry gives the value of

$$10^{-6} \sum_{n=R}^{R+10^6-1} |Z^3(g_n)|$$

for the appropriate R. We see that the computational results are in excellent agreement with Titchmarsh's results (2.13.1) to (2.13.3).

2.14. Violations of Rosser's rule

Rosser's rule, formulated on the basis of empirical evidence, states that a Gram block $B(n,k)$ contains at least k zeros. It thus requires less regularity than Gram's law, yet if Rosser's rule were to hold universally, it would imply the RH (just as the validity of Gram's law would), and would also imply that every Gram block $B(n,k)$ contains exactly k zeros. However, it is easy to see that Rosser's rule holding up to height T is equivalent to the bound $|S(t)| < 2$ holding for $t < T$, which contradicts the unboundedness of $S(t)$. Thus Rosser's rule has to fail infinitely often.

$S(t)$ grows very slowly, and Rosser's rule holds for most Gram blocks that have been checked. The first exception to Rosser's rule (defined as a Gram block $B(n,k)$ which has fewer than k zeros) is $B(n,2)$ with $n = 13,999,525$ [Br5]. There are

only 15 exceptions to Rosser's rule for $n \le 7.5 \times 10^7$ [Br5], and 3055 exceptions for $n \le 1.5 \times 10^9$ [LRW2]. Among the values of n with $1.4 \times 10^9 \le n \le 1.5 \times 10^9$, there were 0.287 exceptions per 10^6 zeros.

The new computations found 62528 exceptions to Rosser's rule. Table 2.14.1 shows how many occurred in each data set and their density. Not only are the exceptions in the new data sets more frequent, but they also are much more varied than those found among the first 1.5×10^9 zeros. If $B(n,k)$ is an exception to Rosser's rule, then k will be referred to as the *length* of the exception. The pattern of zeros inside this block has to be $011\cdots110$. (For notation, see Section 2.13.) To describe the exception, we have to specify where the two "missing zeros" are located. We will use the notation

$$kXa_1a_2\cdots a_m, \qquad X = L \ \text{ or } \ R\,, \tag{2.14.1}$$

to denote an exception $B(n,k)$ where the missing zeros are to the left of $B(n,k)$ (if $X = L$) or to the right of it (if $X = R$), and where $a_1a_2\cdots a_m$ denotes the pattern of zeros in the smallest union of Gram blocks that is adjacent to $B(n,k)$ and contains the missing zeros. Thus, for example, $3L0312$ denotes an exception of length 3, where the pattern of zeros in $[g_n - 4, g_n + 3)$ is 0312010. This is not a completely unambiguous description, but it suffices for all the cases that have been encountered, as no case of 3 exceptions to Rosser's rule that are close together has been found. We will refer to (2.14.1) as the *type* of the exception, and m will be called the *length of the excess block*. With this notation the 3055 exceptions among the first 1.5×10^9 zeros fall into just 13 types:

$$2R3,\ \ 2L3,\ \ 2R40,\ \ 2L04,\ \ 2R22,\ \ 2L22,\ \ 2R230\,,$$

$$2L032,\ \ 2R410,\ \ 3R3,\ \ 3L3,\ \ 3R40,\ \ 3L04\,,$$

with 2715 of them being of types $2R3$ and $2L3$, and only 82 being of length 3. In particular, all lengths of exceptions and lengths of excess blocks are ≤ 3.

The exceptions found during the new computations fall into 206 distinct types.

The relative frequencies of the most popular types in the new data sets and also among the first 1.5×10^9 zeros are shown in Table 2.14.2.

The maximal length of an exception that was found is 9, and it occurs in two exceptions, one of type $9R3$ and one of type $9L3$. There are 15 exceptions of length 8, 84 of length 7, and 416 of length 6. The maximal length of an excess block is 9, and occurs in 2 exceptions of types $2R211113110$ and $3L011311112$. There are 8 cases of excess blocks of length 8, and 46 of length 7.

Some of the 62528 exceptions to Rosser's rule that were found in the main computations occur very close to each other. There are several cases where 2 exceptions are separated by a single Gram interval. The smallest such case that was found is that of $B(n,3)$ and $B(n+4,5)$, where $n = 10^{16} + 3,916,331$, and the pattern of zeros in $[g_n, g_{n+1}0)$ is 0103011103. No case was found where two exceptions are adjacent. (However, Section 3 presents results of other computations that found several examples of this phenomenon.) Finally, no example of 3 exceptions close to each other has been found.

Chapter 3

Special points for the zeta function

3.1. Introduction

The main computations described in Section 2 were carried out at heights that were thought likely to be random as far as the behavior of the zeta function is concerned. Thus the data that were collected were likely to represent long-run statistics of the zeta function at those heights. However, what is most interesting is not to study the typical behavior but to look at extreme values. It would be desirable, for example, to determine where the smallest spacing of consecutive zeros up to some height is without finding all the zeros up to that height. No way to do this is known. It is not even known how to find places where consecutive zeros are very close to each other. The problem is that we would need a way to determine places where both the zeta function and its derivative are small, and this is not feasible currently. On the other hand, there are ways to determine values of t where $\zeta(1/2 + it)$ is likely to be very large. Such methods have been used before [KW, vdL, Od2], and the method described later in this section is a development of the method that was mentioned briefly in [Od2]. These methods determine values of t for which the large initial terms in formulas for $\zeta(1/2+it)$ have the same argument, and therefore add up to a large quantity that often is not cancelled by the remaining terms.

One reason for the interest in large values of $\zeta(1/2+it)$ is that one can think of a large peak as "pushing aside" the zeros that would normally lie in that area, and if these zeros were pushed off the critical line, one would find a counterexample to the RH. No such counterexamples were found in the computations described here, but many interesting phenomena were observed.

Section 3.2 presents the results of the computations near the special points. Section 3.3 describes the diophantine approximation algorithms that were used to construct these special points. Finally, Section 3.4 discusses how these algorithms could be improved, and what other computations could be attempted in the future.

3.2. Computational results

The computations of this section, which are summarized in Tables 3.2.1 and 3.2.2, found 5,168,540 zeros. As is the case with the main sets of zeros described in Section 2, even if the programs are correct and roundoff errors do not matter, it is not absolutely certain that the few dozen zeros at the ends of the data sets are indeed all the zeros in those ranges (cf. Section 2.2). However, for the purpose of exposition, it will be assumed that they are.

There were 22 separate computations, and the sets of zeros and special points associated to them will be denoted by the letters A through V. Table 3.2.1 shows the first zero of each data set, the number of zeros in that set, and the value of $t_0 = (\gamma_n + \gamma_{n+1})/2$ for that n for which $|Z((\gamma_n + \gamma_{n+1})/2)|$ is largest among all n in the data set. Table 3.2.2 then shows the value of $Z(t)$ at $t = t_0$, the largest value of $S(t)$ in a neighborhood of t_0 (which is always the largest $S(t)$ in a given data set, but which does not always occur at γ_n or γ_{n+1}), the value of δ_n (which in all cases is the largest δ_m in a given data set), and the pattern of zeros in a union of Gram blocks that include t_0 (see Section 2.13 for notation).

The entries in Table 3.2.2 show that the attempt to produce unusual behavior of the zeta function was successful. The value of $|Z(t)| \approx 1580$ found in set U is

far higher than 641, the largest value that was found in the main computations. Similarly, the value of $\delta_n = 5.1454$ from set C is the largest δ_n that has been found so far, and the value of $S(t) = 2.8747$ from set T is a record for this function. Figures 2.2.1 and 3.3.2 show graphs of $Z(t)$ near the special value of $t = t_0$ from set T, and Fig. 3.2.3 shows a graph of $S(t)$ in that same range.

Figures 3.2.1 and 3.2.2 are typical of those for the other sets in that they display a single high peak of $|Z(t)|$, with other nearby values of $Z(t)$ much smaller. For example, in looking at stretches of about 30 Gram intervals centered at the special points, one finds only 3 peaks among all 22 data sets where the sign of $Z(t)$ was opposite to that at the main peak, and $|Z(t)| > 30$ was satisfied. The largest value of $|Z(t)|$ in such secondary peaks was 36. Thus we are probably still not seeing the expected behavior of large values of $Z(t)$ that is discussed at the end of Section 2.9.

The general distribution of zeros as well as other properties of the zeta function in the ranges covered here were not too remarkable, aside from the behavior near the peaks of $|Z(t)|$. Exactly 3 midpoint values w_n (see (2.8.2) for a definition) that were > 250 were found away from the special values of t, but they were all < 304. Exactly 100 values of $w_n < 5 \times 10^{-4}$ were found, the smallest of them 2.47×10^{-5}. The smallest value of δ_n that was found was 3.29×10^{-3}, with the second smallest 7.62×10^{-3}. (Since the probability the minimal δ_n of 5,168,540 being drawn from the GUE ensemble is $\leq 3.29 \times 10^{-3}$ is about 0.18, this is consistent with the tendency that was observed before of the minimal δ_n being somewhat smaller than expected.) There were 5459 values of n with $\delta_n < 0.1$, and 844 values of n with $\delta_n > 2.8$. The largest 22 δ_n that were found are the ones given in Table 3.2.2. The 23-rd largest δ_n was 3.50. There were 1861 values of $\delta_n + \delta_n < 0.6$, the smallest of them 0.2512, and 525 values of $\delta_n + \delta_{n+1} > 4$, the largest of them 6.0165. (If $n = 35,200,636,070,992,305,894$, so that $\delta_n = 4.3214$ is the largest δ_m in set V, then for this n we have $\delta_{n-1} + \delta_n = 6.0165$.)

An initial concern about these computations was that they might give a distorted

view of various properties of the zeta function, such as the distribution of δ_n, for example, at the heights being investigated. This was thought possible because the special points t_0 were chosen so that the initial terms in the Riemann-Siegel formula for $Z(t_0)$ behave as if t_0 were close to 0. Thus it seemed possible that aside from the vicinity of the special point t_0, where $Z(t)$ is large, $Z(t)$ might behave as if t were small, and so would be very constrained. However, that appears not to be the case. The agreement between the distributions of the δ_n in our small sets and the GUE prediction is good when one compares graphs prepared like those of Figures 2.4.4 and 2.4.6, and also when one prepares $q-q$ plots. In those comparisons, the presence of one huge outlier does not make much of a difference. On the other hand, when comparing moments of $\delta_n - 1$, one sees substantial differences, especially for high moments. These are easy to explain. When computing the mean value of $(\delta_n - 1)^{10}$ over 2.5×10^5 zeros, for example, a single value of $\delta_n = 5$ will contribute $4^{10} \cdot 4 \cdot 10^{-6} = 4.1943$ to the mean, whereas the GUE prediction for that mean is only 0.488.

The maximal value of $|S(t)|$ does not always occur at one of the two zeros adjacent to the highest peak of $|Z(t)|$. For example, in set V, if we let $n = 35,200,636,070,992,171,653$, then $w_n = 1329.5$, $\delta_n = 4.3214$, but $\delta_{n-1} = 1.6951$, and $S(\gamma_{n-1}+) = 2.8314$, $S(\gamma_n-) = 1.1363$, $S(\gamma_n+) = 2.1363$, $S(\gamma_{n+1}-) = 2.1851$, $S(\gamma_{n+1}+) = -1.1851$.

Exactly 614 exceptions to Rosser's rule were found. They fall into 45 types, each of which had occurred in the main computations. The longest exception had length 6, and the longest excess block also had length 6. On the other hand, a new phenomenon was observed in 21 of the 22 data sets, namely that of 2 exceptions to Rosser's rule being adjacent to each other. Thus for example, the zero pattern 22000022 near the special point t_0 for set A corresponds to an exception of type $2L22$ followed immediately by an exception of type $2R22$. This phenomenon has been observed only in the 21 cases exhibited in Table 3.2.2.

The basic conclusion to be drawn from the computations of this section is that the idea of looking for special points where the zeta function behaves in unusual ways is sound, and does produce interesting results. It also shows that investigating only a random selection of about 10^8 out of the first 10^{20} zeros misses some of the most intriguing places.

3.3. Diophantine approximation algorithms and special points

The Riemann-Siegel formula (Eq. (4.2.2)), as well as other "approximate functional equations" show that the size of $\zeta(1/2+it)$ is determined by the size of the sum of an initial segment of the divergent Dirichlet series

$$\sum_{n=1}^{\infty} n^{-1/2-it} \ . \tag{3.3.1}$$

One can also hope that the size of this sum is determined largely by the size of a partial Euler product,

$$P_X(t) = \prod_{p \le X} (1 - p^{-1/2-it})^{-1} \ . \tag{3.3.2}$$

The basic strategy for finding large values of $\zeta(1/2+it)$ is to find t such that $|P_X(t)|$ is large, and if it is, compute $\zeta(1/2+it)$. (In practice, it has turned out to be helpful to first check that $|P_Y(t)|$ is large for some $Y > X$. This eliminated many candidate values of t.) There is no guarantee that this approach will succeed, but it appears to work well.

To find values of t that make $|P_X(t)|$ large, we search for values of t such that each of the p^{it} is close to 1, as that makes each term in the product maximal. Thus we need to find a t for which there exist integers $m_1, \dots, m_n$ such that each of $t \log p_k - 2\pi m_k$ is small, $1 \le k \le n$, where $n = \pi(X)$ and $p_1, p_2, \dots, p_n$ are the primes $\le X$. This is an instance of a homogeneous simultaneous diophantine approximation problem. We solve it using the Lovász lattice basis reduction algorithm [LLL], which has now become the basic tool in solving a variety of diophantine approximation

problems in high dimensions. Given a basis for a lattice in which the vectors have integer coordinates, this algorithm produces another basis of short vectors. While the new *reduced* basis is not guaranteed to contain the shortest vector in the lattice, the algorithm has polynomial running time and variants of it are efficient in practice. The papers [LO2, OtR] contain some examples of the applications of this algorithm.

The lattices to which the Lovász algorithm was applied have as their basis the rows of the following $(n+1)\times(n+1)$ matrix:

$$\begin{pmatrix} [\alpha_1 2^{m-r}\log p_1] & [\alpha_2 2^{m-r}\log p_2] & \cdots & [\alpha_n 2^{m-r}\log p_n] & 1 \\ [2\pi\alpha_1 2^m] & 0 & \cdots & 0 & 0 \\ 0 & [2\pi\alpha_2 2^m] & \cdots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & [2\pi\alpha_n 2^m] & 0 \end{pmatrix} , \tag{3.3.3}$$

where $\alpha_k = p_k^{-1/4}$, and $m > r > 0$ are integers. A typical vector in the reduced basis is then of the form

$$(M[\alpha_1 2^{m-r}\log p_1] - m_1[2\pi\alpha_1 2^m], \ldots, M[\alpha_n 2^{m-r}\log p_n] - m_n[2\pi\alpha_n 2^m],\ M) , \tag{3.3.4}$$

where $M, m_1, \ldots, m_n$ are integers. For this vector to be short, M and each of

$$M[\alpha_k 2^{m-r}\log p_k] - m_k[2\pi\alpha_k 2^m] \tag{3.3.5}$$

have to be relatively small. For the difference in (3.3.5) not to be large,

$$M2^{-r}\log p_k - 2\pi m_k \tag{3.3.6}$$

must be small, so that $t = M2^{-r}, m_1, \ldots, m_n$ gives a solution to our basic problem.

The function of the α_k in the definition (3.3.3) of the lattice basis is to take advantage of the fact that in trying to make $P_X(t)$ large, it is more important that the $p = 2$ term be large than that the $p = 79$ term be large, say. If

$$t\log p_k - 2\pi m_k = \epsilon_k , \tag{3.3.7}$$

and the ϵ_k are small, then

$$\log|P_X(t)| - \log P_X(0) \approx -\frac{1}{2}\sum_{p\le X} \epsilon_k^2 p_k^{-1/2} \ , \tag{3.3.8}$$

and so we really wish to minimize $\sum \epsilon_k^2 p_k^{-1/2}$. Since the Lovász algorithm attempts to minimize the Euclidean norm of vectors, the definition of the α_k induces it to produce the desired result.

The implementation of the Lovász algorithm that was used in the computations of this section was essentially the same as that of [LO2, OtR], and will not be described here. Just as those implementations, it computed the Gram-Schmidt factors in floating point approximations, and not in exact rational arithmetic, to make the computations practical. For each initial basis, several iterations were performed; after reducing a given basis, the rows of the reduced basis were permuted, and the Lovász algorithm was applied to that basis. This had roughly the same effect as the procedure followed in [LO2], in which several permutations of the initial basis were reduced separately, in that additional reductions gave sometimes better and sometimes worse results.

As in [LO2, OtR], the lattice basis reduction algorithm was implemented using Brent's MP multiple precision package [Br4]. The lattice basis of the form (3.3.3) to which it was applied usually had $40 \le n \le 85$, $70 \le m \le 75$, and $11 \le r \le 16$, and usually about 6 successive reductions were performed. All the values of t from all the reductions (several thousand values in total) were collected and used to compute $|P_Y(t)|$ with Y about $p_{95} = 499$. Those t for which $|P_Y(t)|$ was largest (in a given range of values of t) were then used for the computations described in Section 3.2.

3.4. Possible extensions

One possible way to obtain even better values of t is to speed up the implementation of the Lovász algorithm. The Brent MP package [Br4] was written to be portable and is not very efficient, and on a machine like the Cray X-MP is about 10 times

slower than a program customized for this machine could be. Also, there are some nice methods for speeding up the Lovász algorithm itself that have been developed by Radziszowski and Kreher [RK]. All these improvements could be used to reduce lattices of larger dimensions or reduce more permutations of a given basis. Another approach might be to develop better lattice basis reduction algorithms. Several approaches are available, such as those of Schnorr [Sch1, Sch2], but apparently none of them have been implemented yet. Any one of those approaches could also be combined with simpler tricks, such as that of trying to maximize a product like that of (3.3.2), but where some of the large primes are replaced by slightly larger primes.

All the above approaches have major limitations. Logarithms of primes are rationally independent, and ought to behave like independent random variables as far as multidimensional diophantine approximation properties are concerned. This means that given any fixed subset S of them, values of t for which all the $t \log p$ for $p \in S$ are small modulo 2π are likely to be far apart, and if S is large, the smallest value of t of this kind is likely to be large. Therefore to find values of t for which $\zeta(1/2+it)$ is large, we probably need algorithms that can find vectors in extremely high dimensional lattices that are only slightly shorter than usual, as opposed to the method that has been used, which finds very short vectors in low dimensions. It is doubtful that any of the approaches suggested above could yield such algorithms.

Computations with some of the values of t that were found during the main computations of Section 2 and for which $\zeta(1/2+it)$ is large confirm the suggestion above that such large values arise typically from unpredictable interactions of many large primes and not from an almost perfect lining up of a small set of initial primes. Therefore further searches for values of t with $\zeta(1/2+it)$ large using algorithms known or foreseeable today might produce additional interesting phenomena, but is not likely to find all the large values.

Simultaneous diophantine approximation algorithms could also be applied to find other values of t for which $\zeta(1/2+it)$ is unusual. For example, the values of

t in Table 3.2.1 all lie close to values of u for which $S(u)$ is large, but that is a by-product of having a large gap between zeros in that region. One could also try to search directly for values of t for which $S(t)$ is large. There are various formulas for $S(t)$, such as those of Selberg (Theorem 14.21 of [Tit2]) or Goldston [Go2] (see Section 2.7). The main term in Selberg's formula suggests that to make $S(t)$ large, we ought to search for t such that

$$\sum_{p \le X} \operatorname{Arg}(1 - p^{-1/2-it}) \tag{3.4.1}$$

is large in absolute value. This task can be formulated easily as a diophantine approximation problem, but to obtain large values, it appears that we need to deal with large X, which tends to produce impracticably large values of t. There are two culprits here. One is that the contribution of the sum in (3.4.1) to $S(t)$ is divided by π. The other one is that the error term in Selberg's formula is large compared to the main term in ranges of t that are of interest. (This is to be expected, since $|S(t)| < 2.9$ for all values that have been computed, while the remainder terms in Selberg's formula have to produce the jumps by 1 of $S(t)$ at zeros, since the main term is continuous.)

The final conclusion to be drawn from the above discussion is that searches for special values of $\zeta(1/2 + it)$ do produce interesting results and can be improved somewhat, but there is no method in sight that is likely to produce all the points of interest.

Chapter 4

Algorithms and their implementation

4.1. Introduction

The main result of [OS], namely Theorem 5.1, can be reformulated for the case of computations of $\zeta(1/2+it)$ as follows:

> *For any $a \in [0,1/2]$ and any positive constants δ and c_1, there is an effectively computable constant $c_2 = c_2(\delta, c_1, a)$ and an algorithm that for every $T > 0$ will perform $\leq c_2 T^{1/2+\delta}$ operations on numbers of $\leq c_2 \log T$ bits using $\leq c_2 T^{a+\delta}$ bits of storage and will then be capable of computing any value $Z(t)$ for $T \leq t \leq T + T^a$ to within $\pm T^{-c_1}$ in $\leq c_2 T^{\delta}$ operations using the precomputed and stored values.*

This result is completely rigorous, but implementing it as it is described in [OS] presents difficulties because of the need for high precision and large storage. This section shows a modified version of the algorithm that is practical, but which does sacrifice some of the rigor of the basic result to achieve speed. Many of the choices that were made in the implementation were forced or at least suggested by the hardware and software that was used, and would have been made differently on another machine.

All the main computations were carried out on a Cray X-MP supercomputer with 2 processors and 4 million words of main memory. Although occasionally both processors were used, there was no true parallel processing involved, as the programs did not interact with each other. The Cray computers have 64-bit words, with 48-bit mantissas (including the sign bit), which give slightly over 14 decimal digits of precision in the standard single precision (*sp*) floating point numbers, and slightly over 28 decimal digits in double precision (*dp*). (See [Od2] for a more extended discussion of this issue.) A crucial part of the algorithm, as will be shown later, involves computing $\exp(it\log n)$for n ranging up to about $t^{1/2}$. Since $t \approx 1.5 \times 10^{19}$ near the 10^{20}-th zero, $t\log n$ is on the order of 10^{20}, and so if we do the computations in *dp*, then after reducing modulo 2π we are left with only about 8 decimal digits of accuracy, and this is also true for values of $\exp(it\log n)$ that we obtain after exponentiating. This is only barely acceptable, and accounts for most of the lack of rigor in the computations. Attempting this computation in *sp* would produce a totally meaningless answer. On the other hand, the Cray is designed for *sp* computations that vectorize. All *dp* computations are done in software, and although some of them are vectorized by the latest Cray compilers, some *dp* arithmetic operations are about 100 times slower than *sp* ones. Therefore even though *dp* computations were by themselves only barely accurate enough, it was necessary to do as much computing as possible in *sp* to obtain high speed. To achieve this, some hybrid methods described in Sections 4.2 and 4.3 were used.

The problems outlined above of getting sufficient accuracy were due not to the nature of the new algorithm but to the large height at which the computations were undertaken. Implementation of any of the older algorithms (such as that of the Riemann-Siegel formula discussed in Section 4.2) would have had to cope with the same difficulties. (No matter which algorithm was used, supercomputers like the Cray would be essential in practice, since less powerful machines typically have only 32-bit words, which would require using multiple precision packages, which are

prohibitively slow.) The new algorithm does introduce some additional sources of errors, however, which would make rigorous error analysis harder than it would be for the older methods even if higher precision computations were employed.

The present implementation applies only to computations of the zeta function on the critical line. The algorithm of [OS] can also be used to compute the zeta function on other lines, and this has applications to problems such as that of computing $\pi(x)$ [LO4], but no attempt was made to write programs to carry out such applications. The method of [OS] also applies to the computation of Dirichlet L-function and related functions. Only minor modifications to the present implementation would be needed to compute Dirichlet L-functions, and this may be done in the future.

The main computations were all carried out on a Cray X-MP supercomputer running the Unicos 2.0 operating system, with some of the final statistical computations done under Unicos 3.0. The language of the main computations was Fortran, with several different compilers being used. Various UNIX$^{\text{TM}}$ tools, such as the Awk programming language [AKW], were utilized. Many of the statistical studies of zeros were carried out on a DEC VAX 8550 computer using Fortran, Awk, or (especially) the S statistical programming language [BC]. S was also used to produce all the graphs in this paper.

4.2. Zero-locating program

The program for locating zeros is based on the Riemann-Siegel formula [Ed, Gab, Iv, Sie1, Tit2], which has been the basic tool for all zeta function computations at large height during the last 60 years. This formula says that if

$$\tau = t/(2\pi),\ k_1 = \left\lfloor \tau^{1/2} \right\rfloor,\ z = 2(\tau^{1/2} - k_1) - 1\ , \tag{4.2.1}$$

then for any $m \geq 0$,

$$Z(t) = 2\sum_{k=1}^{k_1} k^{-1/2}\cos(t\log k - \theta(t)) \\ + \ (-1)^{k_1+1}\tau^{-1/4}\sum_{j=0}^{m}\Phi_j(z)(-1)^j\tau^{-j/2} + R_m(\tau) \ , \tag{4.2.2}$$

where the $\Phi_j(z)$ are certain entire functions that can be expressed in terms of derivatives of

$$\Phi_0(z) = \frac{\cos\{\pi(4z^2+3)/8\}}{\cos(\pi z)} \ ,$$

and

$$R_m(\tau) = O(\tau^{-(2m+3)/4}) \quad \text{as} \quad \tau \to \infty \ . \tag{4.2.3}$$

Gabcke [Gab] has obtained essentially optimal bounds for the remainder terms $R_m(\tau)$, and the one used in the new computations was

$$|R_1(\tau)| \leq 0.053t^{-5/4} \quad \text{for} \quad t \geq 200 \ . \tag{4.2.4}$$

The asymptotic expansion terms $\Phi_0(z)$ and $\Phi_1(z)$ were computed using their Taylor series expansions [CR, Gab].

The main difficulty in computing $Z(t)$ using the Riemann-Siegel formula is in the evaluation of the cosine sum in (4.2.2). (For t near the 10^{29}-th zero, $k_1 \approx 1.5\times 10^9$.) In the new implementation it was computed as the sum of two terms,

$$Z_1(t) = 2\sum_{k=1}^{k_0-1} k^{-1/2}\cos(t\log k - \theta(t)) \ , \tag{4.2.5}$$

and

$$\text{Re } e^{-i\theta(t)}F(t) \ , \tag{4.2.6}$$

where

$$F(t) = F(k_0, k_1; t) = \sum_{k=k_0}^{k_1} 2k^{-1/2}\exp(it\log k) \ . \tag{4.2.7}$$

The advantage of the new algorithm over the straightforward term-by-term evaluation of the Riemann-Siegel formula is in the method of evaluating $F(t)$, which is

an adaptation of the method presented in [OS], and is described in detail in Sections 4.3 and 4.4. We will now describe the computations of $\theta(t)$, of $Z_1(t)$, and of the zero-locating procedure.

One could take $k_0 = 1$, in which case $Z_1(t) = 0$ identically, but for technical reasons having to do with the speed of the algorithm for computing $F(t)$ it was advantageous not to do this, and in practice one had $100 \le k_0 \le 500$. (See Table 4.5.1 for some values.) The method used to compute $Z_1(t)$ was essentially the same as that used in [Od2] for computing the entire cosine sum in the Riemann-Siegel formula. The argument t was always maintained as a *dp* variable. Another *dp* variable, t_0, was also maintained, which normally had the property that $|t - t_0| \le 10$. Three arrays, d_n, q_n, u_n, $1 \le n \le k_0 - 1$, were also used; d_n was the *dp* value of $\log n$, q_n was the value of $2n^{-1/2}$, computed in *dp* but stored in *sp*, and u_n was the value of $t_0 \log n$ reduced modulo 2π, where the computation was again done in *dp* but the stored value was in *sp*. To compute $Z_1(t)$ for a new value of t, t was compared to t_0. If $|t - t_0| > 10$, t_0 was set to t, and the u_n were recomputed. At that point (and also if $|t - t_0| \le 10$ was satisfied initially) δ was defined as the *sp* value of $t - t_0$, t_1 as the *dp* value of $t_0 + \delta$, $\theta(t_1)$ was computed in *dp*, reduced mod 2π, and converted to an *sp* variable v. Finally, $Z_1(t)$ was computed as the sum (in *sp*) of

$$w_n = q_n \cos(\delta e_n + u_n - v), \quad 1 \le n \le k_0 - 1 \ ,$$

where e_n is the *sp* value of d_n (obtained by truncation).

For the computation of $\theta(t)$, another *dp* variable $\tilde{t}_0$ was maintained together with the *dp* value of $\theta(\tilde{t}_0)$ and with *dp* or *sp* (depending on order) values of derivatives of $\theta(t)$ at $\tilde{t}_0$. When $|t - \tilde{t}_0| \le 50$ was satisfied, $\theta(t)$ was computed from the stored values using its Taylor series expansion around $\tilde{t}_0$, using partially *dp* and partially *sp* arithmetic. When $|t - \tilde{t}_0| > 50$, $\tilde{t}_0$ was set to t and $\theta(t)$ and its derivatives were computed in *dp* (or *sp* for higher derivatives) using Stirling's formula. The reason for this involved procedure was to avoid using the Cray *dp* logarithm function, which was extremely slow when the program was being written. Later, a new version

of the *dp* logarithm routine was installed in the system libraries that is about 4 times faster than the old one, so that this procedure does not gain much. However, the old procedure was retained, both because it was still faster, and because of the considerations of accuracy and reliability of the computational results that are described in Section 4.6.

The procedure for locating zeros was the standard one of finding Gram blocks and searching for the expected number of sign changes of $Z(t)$ in them. When a violation of Rosser's rule was encountered, the program searched neighboring Gram blocks. Once all the zeros were separated, they were located to a nominal accuracy (i.e, disregarding any inaccuracy in the computation) of $\pm 2 \times 10^{-8}$ by the Brent combination [Br1] of linear and quadratic interpolation. The sophisticated zero-locating strategies of [LRW1, LRW2] were not employed, and about 8.5 evaluations of $Z(t)$ were used on average to compute each zero. (An additional 1 evaluation of $Z(t)$ per zero was performed to determine the value of $Z(t)$ halfway between zeros.)

4.3. Odlyzko-Schönhage algorithm

The function $F(t) = F(k_0, k_1; t)$ is computed in two stages. In the first, precomputation stage, which accounts for most of the computing time, $F(t)$ is computed at a uniform grid of points.

$$t = T,\ T + \delta, \ldots,\ T + (R-1)\delta\ . \tag{4.3.1}$$

The second stage, described in Section 4.4, is fast, and computes the values of $F(t)$ for $T + A \le t \le T + (R-1)\delta - A$ for a certain constant A from the stored values of $F(T)$, $F(T+\delta), \ldots, F(T+(R-1)\delta)$. This section describes the precomputation phase. It is based on [OS] with only minor modifications, and although it is essentially complete, it is technical. The description in [OS] does not cover the details of the implementation, but is more conceptual and easier to read, and is therefore

likely to be preferable for those interested only in the basic ideas of the algorithm and not in the details.

Let $r \in Z^+$, and define

$$R = 2^r, \quad \omega = \exp(2\pi i/R) \ . \tag{4.3.2}$$

In principle any R for which the Fast Fourier Transform (FFT) can be applied efficiently could be used, but it was convenient to work with powers of 2. The values of r that were used in the main computations were $r = 17, 19, 23,$ and 24.

For $-R/2 \le h < R/2$, define

$$u_h = \sum_{j=0}^{R-1} F(T + j\delta)\omega^{-hj} \ . \tag{4.3.3}$$

Once the u_h are computed, the $F(T + j\delta)$ can be obtained from them fast through the FFT:

$$F(T + j\delta) = R^{-1} \sum_{h=-R/2}^{R/2-1} u_h \omega^{jh} \ . \tag{4.3.4}$$

This computation takes a negligible amount of time.

Using the definition (4.2.7) of $F(t)$ in (4.3.3), exchanging the orders of summation, and summing the geometric series that arises, one obtains

$$u_h = \omega^h \sum_{k=k_0}^{k_1} \frac{a_k}{\omega^h - b_k} \ , \tag{4.3.5}$$

where the β_k are defined so that $-R/2 \le \beta_k < R/2$ and

$$b_k = \exp(2\pi i \beta_k/R) = \exp(i\delta \log k) \ , \tag{4.3.6}$$

$$a_k = 2k^{-1/2} e^{iT \log k}(1 - e^{iR\delta \log k}) \ . \tag{4.3.7}$$

Write

$$f(z) = \sum_{k=k_0}^{k_1} \frac{a_k}{z - b_k} \ , \tag{4.3.8}$$

Then we need to evaluate $f(\omega^h)$ for $-R/2 \le h < R/2$. Term-by-term evaluation of the sum in (4.2.8) would require on the order of $k_1 R$ operations, which of the same

complexity as evaluating the Riemann-Siegel formula in the standard way at each point $T + j\delta$. However, the new algorithm of [OS] leads to much faster evaluation of the $f(\omega^h)$ by means of Taylor series expansions. Let $\langle x \rangle$ denote the nearest integer to x, let $\|x\|_R$ denote the "cyclic distance" modulo R:

$$\|x\|_R = \min_m |x - mR| \ ,$$

and for integers p, q with $q \geq 0$, $3^q \leq R/2 + 1$, $-R/2 \leq p < R/2$, $|p3^q| \leq R/2 - 1 + (3^q - 1)/2$, define

$$I_{p,q} = \left\{ k : k_0 \leq k \leq k_1, \|\beta_k - p3^q\|_R \geq 3^q - 1, \|\beta_k - \langle p/3 \rangle 3^{q+1\infty}\|_R < 3^{q+1\infty} - 1 \right\} . \tag{4.3.9}$$

Then it can be shown easily (*cf.* [OS]) that each k belongs to at most 6 different $I_{p,q}$ for a fixed q.

Let $Q = \lfloor \log_3(R/2 + 1) \rfloor$. Then for any h, $-R/2 \leq h < R/2$, it is easy to see (*cf.* [OS]) that $\{k_0, k_0 + 1, \ldots, k_1\}$ is the disjoint union of the sets $I_{\langle h3^{-q} \rangle, q}$ for $0 \leq q \leq Q$. Hence if

$$f_{p,q}(z) = \sum_{k \in I_{p,q}} \frac{a_k}{z - b_k} \ , \tag{4.3.10}$$

then for $-R/2 \leq h < R/2$, we have

$$f(\omega^h) = \sum_{q=0}^{Q} f_{\langle h3^{-q} \rangle, q}(\omega^h) \ . \tag{4.3.11}$$

The new algorithm evaluates the functions $f_{p,q}(z)$ at points $z = \omega^h$ with $\langle h3^{-q} \rangle = p$. For $q \leq Q_1$, ordinary evaluation of the sum in (4.3.10) is used. For $Q_1 < q \leq Q$, the function $f_{p,q}(z)$ is expanded in its Taylor series around the point

$$z_{p,q} = \exp(2\pi i p 3^q / R) \ . \tag{4.3.12}$$

It is easy to show (*cf.* [OS]) that these Taylor series converge fast, so not too many terms in them have to be kept. Finally, these Taylor series are used to evaluate the $f_{p,q}(\omega^h)$.

The threshold Q_1 was taken to be 3 in all the computations after some experiments showed that it was reasonably close to the optimal choice. The Taylor series method is inefficient when $|I_{p,q}|$ is small, since its overhead is large. A slight improvement in the program could be obtained by selecting which method to use based on $|I_{p,q}|$ and not on q alone.

The main computation proceeds in stages indexed by integers m, $-R/2 \le m \le R/2 - 1$. In stage m, only $k \in S_m$ are considered, where

$$S_m = \{k : k_0 \le k \le k_1,\ \beta_k \in [m, m+1)\} \ . \tag{4.3.13}$$

For any p and q, if $k \in I_{p,q}$ for some $k \in S_m$, then $S_m \subseteq I_{p,q}$, which makes bookkeeping for the various computations easy. The distribution of the β_k is nonuniform, with almost all the time being spent in the small fraction of stages m for which $|S_m|$ is large.

Each stage m is further subdivided into substages corresponding to a partition of S_m into blocks $S_{m,j}$, $1 \le j \le s$, of consecutive k's with $|S_{m,j}| \le 2560$ for all j, and $|S_{m,j}| < 2560$ being possible only for $j = s$. (For almost all stages $|S_m| \le 2560$, and so $s = 1$.) This was done to keep the sizes of the auxiliary arrays small, and also to have their lengths be multiples of 64, the length of Cray vector registers.

Suppose that

$$S_{m,j} = \{k :\ k_2 \le k \le k_3\} \ .$$

Several auxiliary arrays are defined. The most important and most time consuming to compute is the d_k array, $k_2 \le k \le k_3$, with d_k being an approximation to the dp value of $\log k$. The set $S_{m,i}$ is partitioned into blocks of 64 consecutive values of k (with the last block possibly being smaller), and if a block consists of k's with $k_4 \le k \le k_5 \le k_4 + 63$, then d_{k_5} is computed using the Cray dp logarithm routine, and the d_k, $k_4 \le k < k_5$ are then computed from d_{k_5} using Taylor series expansions. When the program was first written, this involved procedure was about 6 times faster (for computations near zero number 10^{20}) than using the Cray dp logarithm routine,

which served to cut the running time of the entire rational evaluation program by over 30%. (As a result, the computation of the d_k now takes about 10% of the total running time instead of the roughly half that was required by the earliest version of the program which involved the Cray *dp* logarithm function.) The latest Cray mathematical subroutine libraries have a *dp* logarithm routine that is about 4 times faster than the old one, and so the procedure described above is only about 1.5 times as fast as using the standard Cray *dp* logarithm all the time would be. (Much faster variants of this method are possible, as is shown in Section 4.7.1.)

Once the d_k are computed, they are used to calculate $T \log k$ modulo 2π in *dp*, which is then converted to *sp* and used to compute $\exp(iT \log k)$ utilizing the Cray cosine and sine routines. The $2k^{-1/2}$ factor is also computed in *sp* arithmetic. Finally the difference $1 - \exp(iR\delta \log k)$ is computed in the form

$$-2i \exp\left(i\frac{1}{2}R\delta \log k\right) \sin\left(\frac{1}{2}R\delta \log k\right) \ , \tag{4.3.14}$$

where *sp* arithmetic is used for the trigonometric functions, but $2^{-1}R\delta \log k$ is computed in *dp* and reduced modulo 2π in *dp*, for reasons that will be explained later. All these factors are then combined using *sp* arithmetic to obtain a_k. The b_k are also computed in *sp*.

For $q = 2$ and 3, ordinary complex *sp* arithmetic is used to evaluate the $a_k/(\omega^h - b_k)$ for $k \in S_{m,j}$ and these are added to stored variables corresponding to $f(\omega^h)$. For $q \geq 4$, complex *sp* arithmetic is used to compute the coefficients $a_k(z_{p,q} - b_k)^{-n-1}$ for $0 \leq n \leq V$ in the Taylor series expansion

$$\frac{a_k}{z - b_k} = \sum_{n=0}^{\infty} a_k(z_{p,q} - b_k)^{-n-1}(z_{p,q} - z)^n \tag{4.3.15}$$

around $z_{p,q}$, and these are added to the arrays holding the Taylor series coefficients of $f_{p,q}(z)$. The number of terms V depends on m, p, q, and is chosen so as to make the V-th computed coefficient about 10^{-15} times the size of the 0-th one. Except for q close to Q, V is almost always < 50. After all the S_m have been processed, the Taylor series of the $f_{p,q}(z)$ are used to compute the $f_{\langle h3^{-q},q\rangle}(\omega^h)$ in *sp* arithmetic

for $q \geq 4$, and these numbers are then added to the variables corresponding to $f(\omega^h)$. Since the a_k are only accurate to 9 or 10 decimal digits in the computations near zero number 10^{20}, one could take V much smaller for computations at such large heights, say about 2/3 of the present value, without significantly affecting the accuracy of the final results. This would speed up the main program by about 15%. This modification was not made in the programs to keep them the same for all heights.

For $q = 0$ and 1, a special procedure is used, since here b_k and $z_{p,q} = \omega^h$ (for $h = p3^q$) are close to each other, and so computing $z_{p,q} - b_k$ in *sp* would lead to large errors. Instead, we use the expansion

$$\omega^h - b_k = -2i \exp(\pi i(h + \beta_k)/R) \sin(\pi(h - \beta_k)/R). \tag{4.3.16}$$

The $\pi(h - \beta_k)/R$ factor is evaluated in *dp*, reduced modulo 2π, and converted to *sp* before being used to evaluate the sine. If $(h - \beta_k)/R$ is small, the definition (4.3.6) of β_k shows that $2^{-1}R\delta \log k$ reduced modulo 2π cannot be too large, and the ratio of the two sines in (4.3.14) and (4.3.16) is bounded by R in absolute values. The Cray $\sin(x)$ routine is accurate for small x, since it computes $x(\sin(x)/x)$, and so the quotient of the quantities in (4.3.14) and (4.3.16) is evaluated accurately. (The computation of $2^{-1}R\delta \log k$ modulo 2π, which was mentioned above, is done in *dp* to make sure that the arguments of sine in these computations are accurate.)

Aside from the *dp* operations, which are often not vectorized by the Cray compilers, most of the computations were written so they would be vectorized automatically by the compiler. (No assembly language routines were used.) This is even true of the Taylor series expansions, since those are almost always performed on large sets of k's simultaneously, so the inner loops are written to run on k, and not on the index of the Taylor series term being evaluated. (This does require the use of some auxiliary arrays, but since at most 2560 k's are considered in each stage, storage in not a problem.) As will be described in Section 4.7.1, some of the crucial loops in the program are executed at the rate of over 100 million floating point operations per

second, which is fast for Fortran programs, since the cycle time on the Cray X-MP is 9.5 nanoseconds.

The above sketch of the implementation of the rational function evaluation algorithm applies directly only for the runs with $r = 17$ and 19. For $r = 23$ and 24, a modified version of the algorithm had to be used because of space restrictions that are discussed at greater length in Section 4.5. In the implementation discussed above, a complex array of length R is kept for the values $f(\omega^h)$, $-R/2 \le h < R/2$, as well as arrays for the Taylor series coefficients of the $f_{p,q}(z)$ with $q > Q_1 = 3$. In the versions used for $r \ge 23$, the program works on 2^{r-17} segments of values of h, each of length $2^{17} = 131072$. If we denote one such segment by

$$H = \{h : \ h_0 \le h < h_0 + 2^17\}$$

($h_0 = -R/2$, $-R/2 + 2^{17}$, etc.), then the main program computes the contribution to $f(\omega^h)$ for $h \in H$ of all k such that $k \in I_{p,q}$ with some $0 \le q \le 8$, $p = \langle h'3^{-q} \rangle$ for some $h' \in H$, where these contributions are computed as before, namely directly for $q \le Q_1$, and through Taylor series expansions for $q > Q_1$. These values are stored in a file. Another file is also created, which contains the contributions to the Taylor series coefficients for $q \ge 9$ of all the

$$k \in \bigcup_{m \in H} S_m \ .$$

As different H's are processed, the Taylor series contributions for $q \ge 9$ are added, and at the end they are combined with the previously computed contributions of $q \le 8$ to obtain the values of $f(\omega^h)$.

The algorithm is involved, and its running time depends on a complex combination of various factors. A rough indication of where most of the time is spent is provided by Table 4.3.1. It is based on experiments with the algorithm for $r = 17$, when it is applied to evaluate the $f(\omega^h)$ for $k_0 \approx 1.5 \times 10^9$, $k_1 = k_0 + 10^6$, $T \approx 1.5 \times 10^{19}$, $\delta = 0.15$. The total running time was 132 seconds. The figures in Table 4.3.1

should be treated with caution as only a rough indication of where most of the computational effort was spent.

The basic FFT routines that were used were those of Bailey [Bail]. They were written especially for the Cray-2, where they are both faster and more accurate than the standard Cray routines. On the Cray X-MP, Bailey's routines are slightly slower than the standard Cray ones. They were selected because of their greater accuracy, although in comparison with the errors in the rational function evaluation algorithm, the additional errors introduced by the FFT program are negligible. The time needed for the FFT itself was completely negligible, with complex transform on 2^{19} points taking under 1 second, much less time than it took to read in the data.

Because of space limitations on the Cray X-MP, Bailey's routines could be used directly only for $r = 17$ and 19. For $r = 23$ and 24 it was necessary to perform extensive reformatting operations. Suppose that we wish to take the FFT of $v_0, \dots, v_{M-1}$, where $M = 2^g K$, say, and we can only carry out FFT of length K in core. If $w_0, \dots, w_{M-1}$ is the Fourier transform of the v_j, then

$$\begin{aligned} w_h &= \sum_{j=0}^{M-1} v_j \exp(2\pi i h j/M) \\ &= \sum_{s=0}^{2^g-1} \exp(2\pi i h s/M) \sum_{m=0}^{K-1} v_{2^g m+s} \exp(2\pi i m h/K) \ . \end{aligned} \tag{4.3.17}$$

The inner sum above is just the Fourier transform of a sequence of length K, and can be handled by the FFT directly. To implement this, one needs to create new data sets consisting of the subsequences $v_{2^g m+s}$, $0 \le m \le K-1$, carry out the FFT on them, and then combine them to obtain the w_h. For computations with $r = 23$, for example, Bailey's algorithm is used with $K = 2^{19}$, so that for each of the $16 = 2^4$ FFT's, all 2^{23} values v_j have to be read, and after all the FFT's are done, 16 passes through the data are performed to compute and store the decimal linear combination given by (4.3.17). This takes about an hour of elapsed time (the exact length depending on the load on the system), although very little computing time. For $r = 24$, the total time is about 4 times longer. For large computations, it

would be worthwhile to use more efficient procedures, some of which are discussed in Section 4.7. Such procedures would have been advantageous even for computations on the scale described here, and the only reason they were not carried out was the additional programming effort that would have been required, and the limited facilities for data storage that were available.

4.4. Band-limited function interpolation

Section 4.3 shows how $F(t)$, $F(T+\delta), \ldots, F(T+(R-1)\delta)$ are computed. In general, though, we need to compute $F(t)$ for various $t \in (T, T+(R-1)\delta)$ that are not predictable a priori. The approach that was presented in [OS] was to compute several of the derivatives $F^{(h)}(t)$ at the grid points $t = T, T+\delta, \ldots, T+(R-1)\delta$, and then to compute desired values of $F(t)$ by expanding in a Taylor series around the nearest grid point. Since the derivatives $F^{(h)}(t)$ are representable as sums similar to that for $F(t)$, they can be computed by a variant of the algorithm described in Section 4.3. However, the need to use a dense grid and to store the derivatives $F^{(h)}(t)$ at the grid points make this approach inefficient. Another possible approach is that of interpolating values of $Z(t)$ from the values computed on the grid $T, T+\delta, \ldots, T+(R-1)\delta$, as is done in [Hej5], for example, where $Z(t)$ is approximated as if it were a polynomial through the Lagrange interpolation formula. This method also appears inefficient, and furthermore it is not rigorous.

The method that is used to compute $F(t)$ for t not a grid point is based on band-limited function interpolation techniques. If

$$G(t) = \int_{-\tau}^{\tau} g(x) e^{ixt} dx \ , \tag{4.4.1}$$

then it's been known for a long time that $G(t)$ is determined by its samples at the points $n\pi/\tau$, $n \in Z$, provided only that $G(t)$ satisfies some mild conditions, and further that $G(t)$ is then representable by the "cardinal series"

$$G(t) = \sum_{n=-\infty}^{\infty} G\left(\frac{n\pi}{\tau}\right) \frac{\sin(\tau t - n\pi)}{\tau t - n\pi} \ . \tag{4.4.2}$$

Results of this type have a long history, going back to E. Borel, Hadamard, de la Vallée Poussin, E. T. and J. M. Whittaker, and Ferrar in the mathematical literature, and to Nyquist, Kotelnikov, Shannon, and Someya in engineering (see [Hig] for a history), and are the basis for digital sound transmission and storage, for example. Two comprehensive surveys of the literature in this area are those of Butzer et al. [BSS] and Jerri [Jer].

The cardinal series in (4.4.2) is not suitable for the interpolation of $F(t)$ because, aside from the question of whether the expansion (4.4.2) is valid for $F(t)$, the sum in (4.4.2) converges slowly. We use instead a formula for $G(t)$ that involves a sum of $G(n\pi/\beta)$ for some $\beta > \tau$, which thus involves more frequent (and thus less efficient) sampling of $G(t)$, but in which the coefficients of $G(n\pi/\beta)$ decrease rapidly. The basic approach appears to be well-known to many analysts and communications engineers, but no published reference for the result we use was found, so a proof is sketched below. (See [BSS, Jer] for other possible approaches.)

Suppose that $G(t)$ satisfies (4.4.1), where $g(x)$ will be assumed for the moment to be in $L^2(-\tau, \tau)$. Take $\beta > \tau$ and define $g(x) = 0$ for $\tau < |x| < \beta$, and then extend $g(x)$ to the entire real line by making it periodic with period 2β. Then we have

$$g(x) = \sum_n a_n \exp(2\pi i n x/(2\beta)) \ , \tag{4.4.3}$$

where

$$a_n = (2\beta)^{-1} \int_{-\beta}^{\beta} g(x) \exp(-2\pi i n x/(2\beta)) dx \ . \tag{4.4.4}$$

Eq. (4.4.1) then shows that

$$a_n = (2\beta)^{-1} G(-n\pi/\beta) \ . \tag{4.4.5}$$

Next, choose $\lambda, \epsilon > 0$ so that

$$\tau \le \lambda - \epsilon < \lambda + \epsilon \le \beta \ , \tag{4.4.6}$$

and let $H(x)$ be some continuous function with $H(x) = 0$ for $|x| > \epsilon$, and

$$\int_{-\infty}^{\infty} H(x) dx = 1 \ . \tag{4.4.7}$$

Further, let $\chi(x)$ be the characteristic function of the interval $[-\lambda, \lambda]$, and let $ustarv$ denote the convolution of the functions u and v:

$$(u * v)(x) = \int_{-\infty}^{\infty} u(y)v(x-y)dy \ .$$

Then

$$(\chi * H)(x) = \int_{x-\lambda}^{x+\lambda} H(y)dy = \begin{cases} 1, & |x| \le \lambda - \epsilon \ , \\ 0, & |x| \ge \lambda + \epsilon \ . \end{cases} \tag{4.4.8}$$

Therefore

$$G(t) = \int_{-\tau}^{\tau} g(x)e^{ixt}dx = \int_{-\infty}^{\infty} g(x)e^{ixt}(\chi * H)(x)dx \ . \tag{4.4.9}$$

Substituting the Fourier series (4.4.3) into the last expression above and using (4.4.5) yields

$$G(t) = (2\beta)^{-1} \sum_n G(-n\pi/\beta) \int_{-\infty}^{\infty} e^{ixn\pi/\beta + ixt}(\chi * H)(x)dx \ . \tag{4.4.10}$$

If we change n to $-n$ in the above formula, then the integral above is just the Fourier transform of $\chi * H$ evaluated at $n\pi/\beta - t$, which is the product of the Fourier transforms of χ and H. If $h(t)$ is the Fourier transform of $H(x)$,

$$h(t) = \int_{-\infty}^{\infty} H(x)e^{ixt}dx \ , \tag{4.4.11}$$

then we obtain

$$G(t) = \frac{\lambda}{\beta} \sum_n G(n\pi/\beta) \frac{\sin \lambda(n\pi/\beta - t)}{\lambda(n\pi/\beta - t)} \ h(n\pi/\beta - t) \ . \tag{4.4.12}$$

The interpolation formula (4.4.12) was derived under the assumption that $g(x) \in L^2(-\tau, \tau)$, but by taking limits, it is easy to see that this formula holds when $g(x)$ is a finite linear combination of delta functions, as well as in more general settings.

The formula (4.4.12) can be applied directly with $G(t) = F(t)$ for $\tau = \log k_1$, but since the spectrum of $F(t)$ is limited to $[\log k_0, \log k_1]$, it is more efficient to apply it with

$$G(t) = F(t)e^{-i\alpha t} \ , \tag{4.4.13}$$

where

$$\alpha = \frac{1}{2}(\log k_1 + \log k_0) \ . \tag{4.4.14}$$

Then Eq. (4.4.12) yields

$$F(t) = \frac{\lambda}{\beta} \sum_n F\left(\frac{n\pi}{\beta}\right) e^{-i\alpha(n\pi/\beta - t)} \frac{\sin \lambda(n\pi/\beta - t)}{\lambda(n\pi/\beta - t)} h(n\pi/\beta - t) \ , \tag{4.4.15}$$

valid for any β and λ that satisfy (4.4.6), where we now take

$$\tau = \frac{1}{2}(\log k_1 - \log k_2) \ . \tag{4.4.16}$$

We choose

$$\beta = \pi/\delta \ , \tag{4.4.17}$$

$$\lambda = (\beta + \tau)/2 \ , \tag{4.4.18}$$

$$\epsilon = (\beta - \tau)/2 \ , \tag{4.4.19}$$

and take

$$h(u) = \frac{c}{\sinh(c)} \frac{\sinh(c^2 - \epsilon^2 u^2)^1/2}{(c^2 - \epsilon^2 u^2)^1/2} \ , \tag{4.4.20}$$

where c is a constant that was equal to 30 in most of the computations. A typical set of values used for the computation of one of the large sets of zeros near zero number 10^{20} is

$$\begin{aligned} k_0 &= 450 \ , \\ k_1 &= 1,555,488,184 \ , \\ \alpha &= 13.637\ldots \ , \\ \tau &= 7.5279\ldots \ , \\ \delta &= 0.29 \ , \\ \beta &= 10.833\ldots \ , \\ \lambda &= 9.1804\ldots \ , \\ \epsilon &= 1.65258\ldots \ , \\ c &= 30 \ . \end{aligned} \tag{4.4.21}$$

Note that the distances between consecutive Gram points are 0.148433..., so there is only about one grid point at which $F(t)$ is evaluated for every two Gram intervals.

Many different kernels $h(u)$ could have been used for the interpolation. The specific function $h(u)$ of (4.4.20) was suggested by B. F. Logan. He had discovered a long time ago [Kai] that $h(u)$ is a remarkably good approximation to the principal eigenfunction of the finite Fourier transform, which led to its widespread use in some signal processing applications, as well as in some problems in number theory [MO]. More important for our application are some further optimality properties of $h(u)$ that have been proved by Logan [Log1, Log2]. The formula (4.4.15) is evaluated by summing the terms in the series corresponding to n with $n\pi/\beta$ close to t, and neglecting the remainder of the sum. If we do not use any special knowledge of the behavior of $F(n\pi/\beta)$ or of $\sin(\lambda(n\pi/\beta - t))$, and we sum the series in (4.4.15) over n with $|n\pi/\beta - t| < c/\epsilon$, then we need to minimize,

$$\int_{|u|>c\epsilon^{-1}} |h(u)u^{-1}|du \ ,$$

and Logan's results show that this minimum is achieved by the function defined in eqneq4320, and equals

$$2\log \frac{1+e^{-c}}{1-e^{-c}} \ .$$

(For $c = 30$, this quantity is $\approx 2e^{-30} \approx 1.9 \times 10^{-13}$.) Interpolation using the formula (4.4.15) is performed over approximately the interval $[T + c/\epsilon, T + (R-1)\delta - c/\epsilon]$.

For the set of parameters listed in (4.4.21), the interpolating sum in (4.4.15) was estimated by explicitly evaluating and adding up about 120 terms of the sum. Increasing δ (without changing k_0) increases the length of the interval over which $F(t)$ can be computed, and therefore increases the number of zeros that can be calculated. This has practically no effect on the running time of the rational function evaluation program (assuming the number of grid points stays the same), but increases the time needed by the zero-locating program, both because there are more zeros to be processed, and because more terms in the interpolation formula (4.4.15) have to be evaluated. Increasing k_0 allows one to increase δ (and so the number of zeros that can be computed) without changing ϵ (and thus the number of terms that

have to be computed in (4.4.15)). Such a change, however, increases the running time of the zero-locating program by increasing the number of terms in the sum $Z_1(t)$. The choice of parameters listed in (4.4.21) was not optimized carefully, and could undoubtedly be modified to obtain a more efficient algorithm.

4.5. Space and time requirements

Table 4.5.1 shows the running times of the rational function evaluation program in some of the computations that were carried out. The first column denotes the zero set. Upper case letters refer to the computations near the special points described in Section 3 and listed in Tables 3.2.1 and 3.2.2. Lowercase letters refer to computations listed in Table 4.6.1. These were primarily the large sums that are described in Section 2, together with some smaller computations designed to check the accuracy of the larger ones. (See Section 4.6 for a discussion of the reasons for such computations.) The FFT computations were fast, by comparisons, especially for $R \le 2^{19}$. (For $R = 2^{23}$ and 2^{24}, they took several hours of elapsed time, most of it spent reading and writing disk files to rearrange the data, but only seconds of computing times.) The zero-locating program took slightly under 90 minutes per million zeros when $\delta \approx 0.3$ (and less for smaller δ), so that the computation of the roughly 3.3×10^7 zeros in set n took about 46 hours (2800 minutes) in addition to the 102 hours for the rational function evaluation program.

Comparison of entries g and i, and also of k and n, shows that increasing the number of grid points (and therefore the number of zeros that can be computed) has relatively little effect on the running time of the rational function evaluation program; around the 10^{20}-th zero, going from 1.6×10^7 zeros to 3.2×10^7 zeros increases the running time by less than 17%. The reason for not using even larger grids was lack of memory.

Lack of memory, both core and disk, was the main constraint in planning the program from the beginning. Computing time was not a major limitation. Around

2000 hours were used for all the computations reported here, which is substantial. At the time these computations were carried out, however, the Cray was lightly utilized, and so although only time that would have been idle otherwise was used, a lot of it was available. As a result, minimizing the running time of the program was not of high priority. (Various possible improvements are discussed in Section 4.7.)

The Cray X-MP that was used had 2 processors and 4 million words of memory (32Mb, or megabytes). In practice a maximum of 25Mb was available for a single process, and when such a process ran, one processor stood stands idle. The first version of the rational function evaluation program to be implemented had $R = 2^{19}$ and maintained all the auxiliary arrays in memory all the time, and as a result required over 15Mb. This program was used to compute the zeros in sets b, c and e of Table 4.6.1, as well as the $N = 10^14$ set of Table 1.2, but it would not run if there was any other process of over 10Mb that was running. For $R = 2^{17}$, the corresponding program (which was used to compute all the small sets of zeros of Section 3) requires only about 5Mb, and so was able to utilize much more of the spare time that was available, since sometimes it would even run when there was one process of $\leq$ 20Mb running, and all other waiting processes were too large to fit into the remaining memory. (This did not happen in all such cases because of the way the scheduler was working.) Most of the large computations were carried out with the segmented version of the program that is described at the end of Section 4.3. For $R = 2^{23}$, it requires 8Mb. (This can be lowered to below 5Mb with some simple rewriting of the program.) The main zero locating program also uses about 8Mb of space. In this program the space requirement can be lowered to below 1Mb very easily, since only small segments of the values of $F(t)$ at grid points are needed at any time. The reduction of process size to 8Mb seemed sufficient, however, to take advantage of available time.

The limitation on core memory was overcome by segmenting the rational function evaluation program. A limitation harder to overcome was the lack of disk storage

space. Most of the large computations had $R = 2^{23}$, which meant that 2^{23} complex values of $F(t)$ were being computed and stored, which comes to 128Mb. Moreover, this data had to be reformatted for the FFT application, so that at least for short periods, 256Mb had to be stored. (An in-place FFT program would have eliminated the need for the extra storage, and thus would have led to the computation of twice as many values, but this option was not used since it would have required larger storage of final files. This is discussed further in Section 4.7.) Disk space, even for temporary storage, was extremely scarce during these computations, and so this seemed to be close to the limit of what could be easily computed at that time. For $R = 2^{24}$ (set n in Tables 4.5.1 and 4.6.1), the peak storage requirement is 512Mb, and was satisfied only because W. M. Coughran kindly made available some of his dedicated disk space.

Some of the ways of overcoming the memory limitations are discussed in Section 4.7.1.

Over 2000Mb of data from these computations (mostly values of $F(t)$ at grid points, but also some listings of zeros, as well as various other data) have been stored on an optical disk, and are available for further studies. While optical disk storage technology appears to be very reliable, some of the data may have been corrupted in moving it over a local area network, and so may not be usable. (When the possibility of such errors was realized, a system of parity checks was instituted for later data sets, to prevent such problems from arising.)

4.6. Correctness of computational results

The main defect in the computations reported here is that they lack rigorous error bounds. This is owing to the combination of the height at which the zeta function was computed and the computer hardware that was available. Even if one were to use the standard term-by-term evaluation of the Riemann-Siegel formula, this problem would be severe. The main difficulty there would be in evaluating the

first sum in (4.2.2), which is (neglecting the $\theta(t)$ term) of the form

$$2\sum_{k=1}^{k_1} k^{-1/2}\cos(t\log k)\ . \tag{4.6.1}$$

Near the 10^{20}-th zero, $k_1 \approx 1.5\times 10^9$, $t \approx 1.5\times 10^{19}$, and so for almost all values of k in the sum, $t\log k \approx 3\times 10^{20}$. Therefore, since dp arithmetic on the Cray is performed with about 28 decimal digits of precision, the values of $t\log k$ that are computed are accurate only to within about $\pm 10^{-8}$, and hence the values of $t\log k$ after reduction modulo 2π are only likely to be accurate to within $\pm 10^{-8}$, and the sum in (4.6.1) is likely to be evaluated with an error of

$$E = 2\times 10^{-8}\sum_{k=1}^{1.5\times 10^9} \epsilon_k k^{-1/2}, \qquad -1 \le \epsilon_k \le 1\ , \tag{4.6.2}$$

even if the computations of the $k^{-1/2}$ and of the sum in (4.6.1) are performed in infinite precision. Given no special knowledge of the ϵ_k, all one can say is that

$$|E| \le 2\times 10^{-8}\sum_{k=1}^{1.5\times 10^9} k^{-1/2} \approx 1.5\times 10^{-3}\ . \tag{4.6.3}$$

Many examples of Lehmer's phenomenon that have been found where the maximum of $|Z(t)|$ between zeros is substantially less than the bound (4.6.3), and so in these cases one not be even certain that all the zeros are on the critical line, much less be able to locate them accurately.

The use of multiprecision arithmetic packages would solve the above roundoff problem, but at a high price in computing time. In the straightforward evaluation of the Riemann-Siegel formula, one can gain a few extra digits of guaranteed accuracy by a method described at the end of Section 4.7.1. In the new method that was used for this paper, there are several additional difficulties. The rational function evaluation method computes the $f(\omega^h)$ as sums of a large number of terms, and some of these terms are Taylor series expansions whose coefficients are obtained by adding up numerous other expansions. It would be a difficult task to obtain good error estimates for all these operations. Further, even if one succeeded, it would

be necessary to also bound the errors in the FFT routines and in band-limited function interpolation. The FFT, for example, is known for its good properties in controlling errors, but this applies to only a limited extent when one considers worst-case behavior and has to worry even about roundoff errors in addition.

The above roundoff problem would not arise if one used machines with larger word sizes than the Cray's 64 bit ones, but such computers are unlikely to become available in the near future. Another solution would be to restrict computations to lower heights. However, it seemed desirable to obtain information from as high up as possible, since the zeta function approaches its asymptotic behavior slowly. Therefore it was necessary to abandon rigor in the computations.

While no rigorous error bounds have been obtained, the computational results are thought to be accurate. One reason for thinking this is based on heuristics. The bound (4.6.3) is very conservative in that it is sharp only when almost all the ϵ_k in (4.6.2) are close to $+1$ or almost all are close to -1. In practice, one expects that the ϵ_k will be practically independent of each other, and if that is so, then even under the assumption that the ϵ_k take only the extreme value ± 1, we find that the rms value of E is only

$$2 \times 10^{-8} \left(\sum_{k=1}^{1.5\times 10^9} k^{-1} \right)^{1/2} \approx 9 \times 10^{-8} \ .$$

This is the typical error we expect as a result of cancellation among various roundoff errors. Similarly, one expects substantial cancellation in the rational function evaluation, in the FFT, and in band-limited function interpolation.

The need to rely on cancellation of roundoff errors introduces another level of uncertainty to the computation. Although it is common and accepted in numerical analysis, statistics, and the physical sciences, it is seldom encountered in pure mathematics. This uncertainly is added to the usual uncertainties about reliability of hardware, the design of hardware floating point units (*cf.* [Br, Od2, Schr1, Schr2]), the correctness of manufacturers' mathematical subroutines (*cf.* [SF]), the reliability

of compilers, and finally the correctness of the main programs. All these problems occur with reasonably high frequency. To add to the long list of problems that have been found, we mention that with some Cray Fortran compilers, the test program for the Brent MP package [Br4], which computes π, $\exp(\pi(163/9)^{1/2})$, and $\exp(\pi(163)^{1/2})$ to about 100 decimal places, produced all the digits of π correctly, but gave it a negative sign, and produced totally unrecognizable numbers for the remaining two problems. Mathematicians usually insist on a higher standard of rigor than this.

In some mathematical computations it is not important to have absolute assurance of correctness, since the results are used only to obtain insight into behavior of various functions or systems, and eventually conventional proofs that make no appeal to any computations are constructed (*cf.* [Od0]). In many other cases, though, such as that of the Four Color Theorem [AH], or of some proofs in dynamical systems (*cf.* [Lan]), computational results are an integral part of the proof. There is a school of thought that questions the validity of all such proofs.

The computations of this paper are in one sense even more questionable than those mentioned above, since they depend not only on the correctness of the hardware and software, but also on quasi-random cancellation of roundoff errors. This is to some extent worse than relying on usual probabilistic algorithms, since in these at least the coin tosses are really independent, so one can talk of rigorous probabilistic results. (This assumes, of course, that one can obtain really random bits, but that is another topic we will not deal with it here.) In our case there is no true randomness, as the roundoff process is deterministic. Moreover, the zeta function is certainly nonrandom, and so it is certainly conceivable that the errors in the evaluations of $Z(t)$ might arrange themselves to conceal a violation of the RH. It is for this reason that the previous numerical verifications of the RH, such as those of Brent [Br5] and of van de Lune, te Riele, and Winter [LRW2], were done very carefully. For example, those investigators did not even rely on their machines'

cosine routines, and were careful in the analysis of their error terms. As a result, the validity of the verification of the RH for the first 1.5×10^9 zeros by van de Lune et al. relies only on the assumptions that the hardware and compilers were reliable, their program was correct (it is available for inspection in [LRW1] and some further modifications are in [WR]), and that their machines' *dp* cosine routines (used to provide data to the linear interpolation routines that compute cosines) were at least moderately accurate.

The new programs do not have the same assurance of correctness that those of [LRW2] do. However, in a sense they can be argued to be even more trustworthy. The reason for this is that large parts of the computations were done twice. In general, redoing the same computation on the same machine with the same program provides a check only against certain intermittent errors. In our case, though, the computations were quite different. The grids T, $T+\delta$, $T+2\delta, \ldots$, at which $F(t)$ was being evaluated were always different. As a result, the rational functions $f(z)$ that were being evaluated at the R-th roots of unity (where R was sometimes the same and sometimes different in different computations) were different for the two grids. Therefore the numbers that resulted from the application of the FFT, and were used for band-limited function interpolation, were different. That what was being computed in the two calculations was in both cases $Z(t)$ was thus not apparent at all from the numbers being processed, and is a result of the involved analysis of Sections 4.2 to 4.4. That the two values that were obtained were the same to within the expected error serves as evidence that they are indeed values of $Z(t)$, since it would require a very unusual combination of errors for the two computations to yield the same answers otherwise. This method thus serves to check not only the roundoff errors, but also the hardware, compilers, operating systems, and the program themselves.

Care was taken to minimize the parts of the computations of $Z(t)$ that were common to the overlapping sets. The values of δ were always different. The com-

putations of $Z_1(t)$ and of the asymptotic expansion in the Riemann-Siegel formula were harder to make distinct. However, the procedure for computing $\theta(t)$ outlined in Section 4.2 served to make the values of $\theta(t)$ in different computations slightly different, so that in locating zeros, different computations dealt with different values of t.

The above method of computing $Z(t)$ in two different ways that ought to yield the same result only because of deep mathematical results is not novel. In early computations of π, such as that of Shanks and Wrench [SW] (see [BB] for history of this subject and much more efficient modern methods), π was calculated through two different Machin-like formulas. In the case of [SW], they were

$$\pi = 24 \arctan\left(\frac{1}{8}\right) + 8 \arctan\left(\frac{1}{57}\right) + 4 \arctan\left(\frac{1}{239}\right)$$

and

$$\pi = 48\arctan\left(\frac{1}{18}\right) + 32 \arctan\left(\frac{1}{57}\right) - 20 \arctan\left(\frac{1}{239}\right) .$$

Since again only a very unusual combination of errors could give the same answer by both methods, obtaining the same result provides a convincing (although non-rigorous) argument in favor of correctness.

With new algorithm, the comparison of results of overlapping computations did lead to the uncovering of one error in the program that had escaped detection in several earlier runs. After one computation of over 1.6×10^7 zeros near zero number 10^{20} (set m of Table 4.6.1, described below), the set below it (set l of Table 4.6.1) was computed. However, a computation of the zeros in the segment overlapping set m showed apparent violations of the RH when the values of $F(t)$ from set l were being used, although no such violations were found using set m. Interpolation of values of $F(t)$ from set m to give the values of $F(t)$ on the grid of set l revealed that the computed values of $F(t)$ at the grid points $f_j = T + j\delta$ of set l differed from the (presumably correct) ones derived from set m by $c(-1)^j$, where c was a certain constant. This immediately suggested that in set l, $f(-1)$ was being

evaluated incorrectly. An inspection of the code revealed a simple mistake having to do with indexing of the roots of unity ω^h in the segmented program, and it was easy to correct the data. This bug had not revealed itself before because the unusual combination of having a pole of the rational function $f(z)$ in a certain range close to -1, which was required for the code to produce incorrect output, had not occurred in the earlier runs.

How close to each other are the values of $Z(t)$ computed in different runs? Let us consider the neighborhood of the extreme example of Lehmer's phenomenon near γ_n, $n = 10^{18} + 12,376,780$, where the minimum of $Z(t)$ between γ_n and γ_{n+1} is only -5.3×10^{-7}. This is the example that comes closest to violating the RH among all those found in our computations, and the obvious question is whether one can be sure that the RH is indeed satisfied by rho_n and rho_{n+1}. This example was found in a computation of 1.7×10^7 zeros, and to confirm the accuracy of the computed values of $Z(t)$, two additional computations were carried out, each of about 1.5×10^5 zeros, and each centered close to γ_n. The starting points of the three computations and the grid spacings δ were distinct in all three computations to assure maximal independence in the computed values. When the results of the these runs are plotted on the scale of Fig. 2.7.1, they are indistinguishable. When one plots $Z(t)$ only in the immediate vicinity of γ_n and γ_{n+1}, as in Fig. 4.6.1, the three graphs are still indistinguishable. It is only when one goes over to the scale of Fig. 4.6.2, which shows $Z(t)$ near its minimal value in (γ_n, γ_{n+1}), that differences are apparent. This graph was prepared by computing $Z(t)$ from each run at intervals of 10^{-6} times the length of a Gram interval (so that Fig. 4.6.2 corresponds to about 150 evenly spaced values of t) and connecting the points obtained that way by lines. (The function of the lines was to enable the reader to tell which values come from the same data set.) The jagged appearance of the lines is the result of the quantization and roundoff errors. (Note that changing a value of $t \approx 1.7 \times 10^{17}$ by 10^{-6} times the length of a Gram interval affects only the last 15 or so bits in the *dp* representation of t.)

Given the scale of Fig. 4.6.2, the three curves are close together, and thus provide convincing evidence that the claimed values of $Z(t)$, γ_n, and γ_{n+1} are indeed highly accurate, and that the RH is not violated in this region.

All indications from preliminary runs were that the new algorithm was highly accurate, and storage of two complete data sets needed to perform a detailed comparison would have been hard to arrange. Therefore it was decided not to recompute all the values of zeros near zero number 10^{20} using different grids, but to have different computations cover consecutive ranges with some overlap. Table 4.6.1 shows all the different sets of zeros that were computed and that overlapped other ranges. The three sets of zeros that were referred to above in the discussion of Lehmer's phenomenon, for example, are listed under f, g, and h in Table 4.6.1. (The set a consists of the zeros computed in [Od2] by the standard Riemann-Siegel formula method, and so its values of zeros are very trustworthy.) The four main computations (near $N = 10^{20}$) are those in sets k, l, m, and n, and each one overlaps each of its neighbors in about 10^6 zeros. The small set o was computed as an additional check, since the smaller grid spacing and fewer grid points were expected to produce more accurate values, and the somewhat different program used was an extra check on programming mistakes. (This was also the motivation behind some of the other computations of small sets of zeros, such as that of j. The medium size sets, b, c and e, were computed by the earliest of all versions of the new program.)

A few large scale statistical comparisons were made of the values of $Z(t)$ produced in different computations. For example, to compare sets m and o of Table 4.6.1, the values of $Z(t)$ were computed using data from each set at 5×10^6 points spaced $1/300$ apart (about 45 per Gram interval) starting at $t = 1.52024401159207401 \times 10^{19}$. The largest difference (in absolute value) was 1.5×10^{-6}, and the rms difference was 5×10^{-8}.

While the errors made in computing $Z(t)$ are of some interest, the main question is that of accuracy in computing the γ_n, which depends not only on accuracy of

values of $Z(t)$, but on the size of $Z(t)$ and $Z'(t)$ near zeros. Therefore extensive and careful comparisons were made of the differences in values of γ_n computed in different sets. Table 4.6.2 summarizes the results of these comparisons. The "a vs. b" entry, for example, shows that the values for the 101,053 zeros common to sets a and b different by no more than 2.5×10^{-9}, and the rms difference was $3.7 \times 10^{-1}1$. (These are differences in the values of the γ_n. Should the values of two adjacent zeros in set l, for example, each be off by ϵ, with one value too small by ϵ and the other too large by ϵ, the resulting value of δ_n would be off by $\pm 14\epsilon$.) The maximal differences increase as one looks down the table, as was to be expected. They all stay small, though, and are the main justification for the claimed validity of the data.

The rms difference entries in Table 4.6.2 should be treated with great caution. One reason is that the zero-locating program was only asked to compute the zeros to a nominal accuracy of $\pm 2 \times 10^{-8}$ (for zeros near zero number 10^{20}; somewhat higher accuracy was specified for lower zeros). Because of the mixture of linear and quadratic interpolation that was used, usually the convergence of the algorithm at the end of a particular search was quadratic, and so accuracy much greater than the specified one was reached in almost all cases. Thus the fact that the rms figures in Table 4.6.2 are substantially below the specified accuracy of 2×10^{-8} is the result of many happy accidents, and not a matter of design. Another reason not to rely on the rms figures is that often they were inflated by the programs that were used for the comparisons. Since there was no reason to expect accuracy better than $\pm 10^{-8}$, *sp* programs were used for most of the computations of Table 4.6.2, which led to the loss of the last few bits of precision. (The anomalously large rms value for the "l vs. m" entry, as compared to the "k vs. l" and "m vs. n" entries, which cover roughly the same number of zeros at about the same height, is almost certainly due to the use of a *sp* array in a data conversion routine, for example.) Thus in general the rms figures in Table 4.6.2 are upper bounds for the rms errors achievable with

the new algorithm, but should not be regarded as accurate estimates.

One source of errors in the computation of the γ_n lies in the method of calculating $\theta(t)$. Some of these errors come from the roundoff difficulties associated with handling large numbers within the limited precision that was available. Other errors came from the Taylor series expansion procedure, described in Section 4.2, that was used to compute $\theta(t)$. Some indication of the errors introduced this way can be obtained from the data produced during the main runs. Because of the inefficient search procedure near exception to Rosser's rule (described in Section 4.2), usually about ten zeros near each exception were computed twice. The two values were hardly ever identical, since the zero locating program was usually invoked with different arguments. Differences caused by this factor were usually extremely small. Much larger were differences caused by the fact that often the two computations calculated $\theta(t)$ for nearby values of t by expanding around different values of t_0. The largest difference in the computed values of the γ_n that was found that is due to this phenomenon is 2.7×10^{-7} for $n = 10^{20} + 31,141,844$. (The second largest was only slightly more than half as large.) This zero is located near two exceptions to Rosser's rule that are close to each other, with the peak value of $Z(t)$ in that region equal to 257.6. The pattern of zeros (starting at Gram point g_{n-19}) is 2111110110030101311, and $Z(t)$ is small in a large neighborhood of γ_n ($|Z(t)| < 0.015$ over approximately the whole Gram interval that contains γ_n). $Z'(\gamma_n) = 0.7$ is small, and so the computed location of γ_n is sensitive to errors in the computation of $\theta(t)$.

Other tests to determine the sensitivity of the computed values of the zeros to errors in the computation of $\theta(t)$ were also performed. For example, the zeros in set o were computed several times, always using the same rational function evaluation output for the interpolation of band-limited functions, but modifying the strategy of evaluating $\theta(t)$ by forcing more frequent recomputations of t_0, or simply the use of different sets of t_0's. The resulting values for the zeros had differences (when compared to the basic computation of the zeros in that set) that were $\leq 10^{-7}$ in

absolute value, and $\leq 2 \times 10^{-9}$ in rms value. Thus the basic conclusion from these tests is that errors in the computed $\theta(t)$ were not a significant factor.

Another reason for trusting the computational results of this paper is that the results of the most time-consuming part, the rational function evaluation, are transformed by the FFT before being used for the computation of zeros. This means that any error in this part of the computation affects the computation of all zeros, and so if it is substantial, is likely to lead to an apparent counterexample to the RH. This is in contrast to the standard methods, such as that of [LRW2], in which a single mistake affects the computation of only one value of $Z(t)$.

The final, and in many ways most convincing, although unrigorous argument in favor of the correctness of the computations reported here is that they did not find any counterexamples to the RH. This might seem a strange argument. The point of it is that if the RH is true, it is only barely true, in the sense that even tiny changes in the formulas used to compute the zeta function yield functions that no longer satisfy the RH. Many deliberate as well as accidental experiments were performed in which some of the parameters in the programs were modified slightly, and they almost invariably ended up giving apparent counterexamples to the RH. For example, in set V of Section 3, the minimal w_n (see Section 2.8 for definitions) that was found was 1.43×10^{-4}, so perturbing $Z(t)$ by smaller quantities could not produce apparent counterexamples to the RH. In particular, dropping the asymptotic expansion part of the Riemann-Siegel formula does not produce visible problems in this set, although it does in other ones that have more extreme cases of Lehmer's phenomenon, and in all cases it perturbs the computed values of the zeros. On the other hand, only slightly larger perturbations do produce apparent counterexamples. One also finds counterexamples when one computes

$$Z(t) - 2k^{-1/2} \cos(t \log k - \theta(t))$$

for $k = 10^6$. Also, when one computes

$$Z_1(t) + \text{Re } e^{-\theta(t)} \; F(t - \alpha)$$

with $\alpha = 10^{-4}$ instead of $\alpha = 0$, apparent counterexamples to the RH appear. (In all these cases, the apparent counterexamples refer to cases where the function being computed has a positive relative minimum or a negative relative maximum.)

4.7. Possible improvements

At large heights, the new algorithm is much faster than previous methods. The computation of 10^5 zeros near zero number 10^{12} in [Od2] took about 15 hours on a Cray X-MP using direct evaluation of the Riemann-Siegel formula. Set n of Table 4.6.1 contains almost 3.3×10^7 zeros near zero number 10^{20}, and it was computed in about 150 hours on the same machine. Since the Riemann-Siegel formula involves about 7.5×10^3 times more terms near zero number 10^{20} than near zero number 10^{12}, computing all the zeros in set n by the method of [Od2] would have required about

$$15 \times 300 \times 7500 \approx 3.7 \times 10^7$$

hours, or more than 2×10^5 times longer than the new algorithm required.

While the current implementation of the new algorithm is much more efficient than previous algorithms, it is far from optimal. The author's main interest was in demonstrating that the new algorithm was indeed faster than old ones, and in obtaining data about zeros of the zeta function. Since spare computer time was available, saving programming effort was often chosen over efficiency of the program. The following subsections present some of the ways in which the program could be modified to run faster or to produce more accurate results. They might be useful in future computations. It seems likely that the ideas in Sections 4.7.1 and 4.7.2 could be used to increase the speed of the algorithm by another order of magnitude on the Cray X-MP. This might make it possible to compute large sets of zeros near zero number 10^{22}, for example.

All the main programs can be parallelized, and one can achieve high performance this way. (For the rational function evaluation program, there are some examples of

similar algorithms, discussed in Section 4.7.2, that have been implemented effectively on parallel computers by Greengard and Gropp [GG], Zhao [Zh], and Zhao and Johnsson [ZJ].) One difficulty in using existing multiprocessors would likely be their relatively low precision. Our discussion will be oriented towards more standard vector processors, however.

4.7.1. Faster and more accurate computations

Many parts of the rational function evaluation program can be speeded up. Table 4.3.1 shows that the present strategy of computing *dp* values of $\log k$ takes about 10% of the running time. While this is 1.5 times faster than using the current standard Cray *dp* routines would be (and 6 times faster than using the old Cray routines), it can be improved substantially by modifying the program slightly. For example, the following method is about 3.5 faster (for $k \approx 10^9$) than the one currently used. Group the $k \in S_{m,j}$ into consecutive blocks, say $k_4 \le k \le k_5$, with $k_5 - k_4 \precsim 2 \times k_4 \times 10^{-6}$, and let $k_6 = \lfloor (k_4 + k_5)/2 \rfloor$. Compute the *dp* $\log k_6$ using the Cray routine, and compute the $h_j = 1/(j \cdot k_6^j)$ in *dp* for $1 \le j \le 4$. Initially, for $k_4 \le k \le k_5$, assign to d_k the value of $\log k_6$, and then, for $1 \le j \le 4$, modify each d_k, $k_4 \le k \le k_5$, by subtracting from it $h_j(k - k_6)^j$, where the $(k - k_6)^j$ are taken from a precomputed integer array. For $j = 4$, the multiplication of h_j and $(k - k_6)^j$ can be carried out in *sp*. Further improvements can probably be obtained with further experimentation. (The speeds that can be achieved in this part of the program depend strongly on how the compiler treats *dp* computations.)

Substantial savings can be obtained by modifying the procedures used to evaluate the $f_{p,q}(\omega^h)$. Instead of choosing a uniform threshold Q_1, the decision whether to use Taylor series expansions or direct evaluation can be made dependent on the size of $I_{p,q}$. Reducing the number of terms in the Taylor series as mentioned in Section 4.3 can reduce the total running time by at least 15%, and even greater savings are probably possible by more careful choices.

The zero-locating program can be improved in several ways. The computation of $\theta(t)$ takes about 7% of the time for the choice of parameters in (4.4.21). This time can be reduced practically to zero, since around the 10^{20}-th zero, $\theta(t)$ is almost linear and the distances between consecutive Gram points are almost constant.

The strategy for locating zeros can be improved, especially in the neighborhoods of exceptions to Rosser's rule, where it is grossly inefficient. Currently about 8.5 evaluations of $Z(t)$ are used to compute each zero (with one more evaluation to obtain the value of $Z(t)$ at the midpoint between adjacent zeros). It should be possible to devise strategies that take better advantage of the previously computed values and of the expected behavior of $Z(t)$. This might involve computing $Z'(t)$ (which can be obtained from the interpolation formula (4.4.15)) and modifying Brent's algorithm [Br1]. A good model for this approach is the algorithm of van de Lune et al. [LRW1, LRW2], which uses fewer than 1.2 evaluations of $Z(t)$ per zero to separate the zeros.

About 45% of the time of the zero-locating program (for the parameters listed in (4.4.21)) is spent in band-limited function interpolation. (The computation of $Z_1(t)$ takes approximately 25%.) Logan's kernel (4.4.20) has desirable optimality properties in terms of rate of convergence of the interpolating sum (4.4.15), but it is somewhat hard to compute. It is possible that other kernels could be constructed that would give slower convergence, and therefore require evaluating explicitly more terms in the sum in (4.4.15), but which would be much more efficient to compute. Another approach would be to initially use the formula (4.4.15) only for values of t that belong to a grid somewhat finer than that of T, $T+\delta$, $T+2\delta$, ..., say at points $T+k\delta/1000$, $k \geq 0$. For such values of t, it would only be necessary to precompute the values of

$$e^{i\alpha m\delta/1000}\frac{\sin(\lambda m\delta/1000)}{(\lambda m\delta/10000)}h(m\delta/1000) \tag{4.7.1.1}$$

for $|m| \precsim 1000c/(\delta\epsilon)$, and later evaluations of the interpolation formula would be reduced to inner products of vectors. With this approach, even without the use

of assembly language, one could compute interpolation sums of the form (4.4.15) at the rate of about 1.2×10^7 terms in the sum per second on the Cray X-MP. (If one selects δ to be a simple rational multiple of the average gap between Gram points, say $\delta = 2(g_{n+1} - g_n)$, another factor of 2 improvement in speed is possible by taking advantage of the fact that only Re $e^{-i\theta(t)}F(t)$ is needed, and $\theta(t)$ is close to linear.) In the existing program, they are computed at the rate of only about 5×10^5 terms per second. Once the zero was proved to lie between two points of the subgrid $T + k\delta/1000$, one could locate it more accurately by evaluating $Z(t)$ at several neighboring points of the subgrid, using some numerical interpolation method to obtain an approximation to the zero that is likely to be accurate. Finally this location of the zero could be confirmed by using the standard interpolation method to evaluate $Z(t)$ at points on either side of the zero.

All the main programs are written in Fortran, and many parts of the computation are handled by subroutines. Some slight improvements can be expected from replacing subroutine calls, which are slow on the Cray machines, by in-line code. Much greater speedups are likely to be achieved by using assembly language. Currently none of the *dp* operations vectorize. However, given the structure in the *dp* operations employed in the new algorithms, it should be possible to write assembly language code that would vectorize these operations. Another case where assembly language ought to produce much faster programs is in the computation of Taylor series coefficients and the evaluation of Taylor series, which, as is shown by Table 4.3.1, account for more than half of the running time of the rational function evaluation program. Let c_n be defined by

$$\sum_{k \in S_m, j} \frac{a_k}{z - b_k} = \sum_{n=0}^{\infty} c_n (z_{p,q} - z)^n \ ,$$

so that

$$c_n = \sum_{k \in S_m, j} a_k (z_{p,q} - b_k)^{-n-1} \ ,$$

and the c_n corresponding to different sets $S_{m,j}$ are added together to obtain the n-th

Taylor series coefficient of $f_{p,q}(z)$. The c_n are computed in the present program by using two complex *sp* arrays, u_k and v_k, $k \in S_{m,j}$, with

$$\begin{aligned} u_k &= (z_{p,q} - b_k)^{-1} , \\ v_k &= a_k (z_{p,q} - b_k)^{-n} . \end{aligned}$$

To compute c_n, v_k is assigned the value $u_k v_k$ for $k \in S_{m,j}$, and the Cray library function *csum* is invoked to sum the new u_k. (*Csum* sums a complex *sp* array in a vectorized way.) Each c_n thus requires $|S_{m,j}|$ complex *sp* multiplications and $|S_{m,j}| - 1$ complex *sp* additions. These combinations of complex multiplications and additions are carried out at the rate of 1.29×10^7 per seconds. Since each complex multiplication involves 4 real multiplications and 2 real additions, and each complex addition requires 2 real additions, the Cray is performing 1.03×10^8 floating point operations per seconds, which is good when one recalls that the cycle time of the Cray X-MP is 9.5 nanoseconds. However, it should be possible to take advantage of the fact that each of the basic combinations of a complex multiplication and complex addition involves 4 real multiplications and 4 real additions. Since the Cray can do an addition and a multiplication at the same time, it ought to be feasible to write the code so that additions coming from computation of some particular n would be done at the same time as the multiplications for $n + 1$, say. This would give a speedup factor 2, if the data transfers could be arranged appropriately. (It might even be possible to obtain some savings in Fortran, without resorting to assembly language.) Similar improvements can probably be obtained in other parts of the program.

Table 4.3.1 shows that many different parts of the rational function evaluation program consume noticeable fractions of computing time, and so for maximal efficiency one would have to work on all of them. This is also true of the zero-locating program. No matter to what extent these programs are optimized, however, one can obtain some savings by optimizing the choice of the parameters δ, k_0, etc., which were not chosen very carefully in the computations that are reported here.

Memory constraints can be overcome by using larger mass storage devices. These could be larger magnetic disks (the maximal total storage requirement for computing the 3.3×10^7 zeros in set n was 512Mb, and much larger disks are commercially available), or even magnetic tapes (especially the high capacity digital tapes that are becoming available) or optical disks. Substantial savings can be realized by using in-place FFT programs, which perform the FFT on a set of data with little additional space being required. Most such algorithms require about $\log R$ passes through the data for an R-point FFT. D. H. Bailey has pointed out that there is an algorithm of Gentleman [Gen] which requires only three passes through the data, and so is particularly attractive for our application.

One factor that might facilitate very large scale computation with the new algorithm is that most of computing time of the rational function evaluation program is spent in only a few segments, because of the nonuniform distribution of the β_k. This means that one can first perform most of the computation with very little storage, and only a couple of hours would be needed to deal with the overwhelming majority of segments and the FFT program, and it's only during that time that substantial storage would be needed. Afterwards, the data can be stored away even on slow mass storage devices, since the zero-locating program requires only small segments of data at a time. Furthermore, storage space can sometimes be used more fully by choosing R not to be a power of 2.

At the time the program was implemented, limitations on disk storage, capacity of local area networks, and availability of long-term storage on optical disks were such that utilizing the methods suggested above seemed very cumbersome. Right now, however, with the availability of an automatic optical disk changer and larger disks, it would be much easier to carry out some of these improvements.

It is possible to gain some additional speed by taking advantage of the nonuniform distribution of the β_k. Instead of computing $F(k_0, k_1, t)$ at a grid of point

$t = T, T+\delta, \ldots, T+(R-1)\delta$, one can compute $F(m_j, m_{m+1}; t)$ for $0 \le j < s$, where

$$m_0 = k_0 < m_1 < \cdots < m_s = k_1 \ ,$$

where the m_{j+1}/m_j are roughly equal, and where t now would run over a much sparser grid, approximately $t = T$, $T+s\delta$, $T+2s\delta$, ..., since the range of frequencies in each $F(m_j, m_{j+1}; t)$ would be much narrower. In order not to use too much space, one could then compute $F(m_j, m_{j+1}; t)$ at only $R/2^{s-j}$ points at a time. This would mean that the $F(m_j, m_{j+1}; t)$ for $j \le s-2$ would be computed on several adjacent grids of points, but since in the ranges that are being considered now, the running time depends mostly on k_1, and to a much smaller extent on R, it appears that one could obtain substantial savings.

It should be possible to obtain slightly more accurate results at very small additional cost in computing time. This might be useful if one were to do some computing near zero number 10^{22}, for example. The main source of inaccuracy in the present program is in computing $T \log k \bmod 2\pi$, $2 \le k \le k_1$. While multiprecision programs such as Brent's MP package [Br4] are likely to be too slow to be used on each such term separately, one can apply them for some values of k spaced far apart, and then use Taylor series expansions in terms of k to obtain the other values. One has

$$T \log(k+h) = T \log k + Th/k - Th^2/(2k^2) + \cdots \ ,$$

and so if $|h/k| \le 10^{-3}$, say, dp arithmetic would give about 4 more decimal digits of accuracy than the method that is now used.

4.7.2. Greengard-Rokhlin algorithm

The possible modifications to the present implementation that are discussed in Section 4.7.1 are small programming improvements. It is also possible to change the basic rational function evaluation algorithm, by modifying the functions $f_{p,q}(z)$ (see Section 4.3). What is needed is a collection of functions $\tilde{f}_j(z)$ such that for every h,

$$f(\omega^h) = \sum_{j \in J(h)} \tilde{f}_j(\omega^h)$$

for some subset $J(h)$, and such that all the $\tilde{f}_j(\omega^h)$ can be computed efficiently by direct evaluation or by using their Taylor series expansions. The functions $f_{p,q}(z)$ of [OS] and of Section 4.3 are only one of many choices.

As was already noted in [OS] in remarks about the Trummer problem, the rational function evaluation algorithm of that paper can be extended to the evaluation of much more general functions. Another algorithm for the evaluation of Coulomb and gravitational potentials was invented by Greengard and Rokhlin [GR1], and was subsequently improved, extended,and applied to several additional problems by several investigators [AGR, CGR, GG, GR2, GR3, Kat, Zh, ZJ]. It seems to offer the possibility of a substantial improvement in the speed of the zeta function program. Its underlying principle is the same as that of the algorithm of [OS], namely of aggregating the contributions of those poles of the function that are close together. However, it works differently. To avoid unnecessary notation, we explain briefly how it would be applied to the zeta function problem, which is the evaluation of $f(\omega^h)$, although it is more general than that. In the algorithm of [OS] and this book, the functions $\tilde{f}_j(z)$ are usually evaluated by obtaining their Taylor series expansions around points outside the regions containing their poles. For each m, and each $k \in S_m$, the coefficients $a_k(z_{p,q} - b_k)^{-n-1}$ in (4.3.15) are evaluated for $0 \le n \le V$ (V is usually around 40) and for each pair p, q such that $S_m \subseteq I_{p,q}$, $Q_1 \le q \le Q$. Since there are about $Q \approx \log R$ such pairs p, q, the total effort involves on the order of

$$V k_1 \log R \tag{4.7.2.1}$$

basic arithmetic operations.

In the Greengard-Rokhlin algorithm, one would compute instead the coefficients $a_k(b_k - z_m)^n$ in the expansions

$$\frac{a_k}{z - b_k} = \sum_{n=0}^{\infty} a_k (b - z_m)^n (z - z_m)^{-n-1} \ , \tag{4.7.2.2}$$

for example. Here z_m would be some point located among the b_k with $k \in S_m$, say

$$z_m = \exp(2\pi i(m + 1/2)/R) \ .$$

Expansions of the type (4.7.2.2) converge for z away from z_m and b_k, say for $|z - z_m| \geq 3/2$. The coefficients $a_k(b_k - z_m)^n$ would again be computed for $0 \leq n \leq V$, with V of about the same size as in the present algorithm. Addition of the coefficients $a_k(b_k - z_m)^n$ for $k \in S_m$ gives an expansion

$$\sum_{k \in S_m} \frac{a_k}{z - b_k} = \sum_{n=0}^{V} A_n^{(m)} (z - z_m)^{-n-1} + \cdots \tag{4.7.2.3}$$

that can be used to compute the contribution of the $k \in S_m$ at $z = \omega^h$ for ω^h close to z_m. For ω^h that are further away, one would combine the contributions of several S_l's, say S_m, S_{m+1}, S_{m+2}, and obtain an expansion around z_{m+1}. The crucial point about the Greengard-Rokhlin algorithm is that unlike in the algorithm of [OS] and this book, this expansion around z_{m+1} would not be done by recomputing the contribution of each $k \in S_m cup S_{m+1} cup S_{m+2}$, but by translating the previously computed expansions; e.g.,

$$\sum_{n=0}^{V} A_n^{(m)} (z - z_m)^{-n-1} = \sum_{n=0}^{V} B_n^{(m)} (z - z_{m+1})^{-n-1} + \cdots , \tag{4.7.2.4}$$

where the $B_n^{(m)}$ are derived from the $A_n^{(m)}$ by linear transformations coming from the binomial expansion, without reference to the a_k and b_k for $k \in S_m$. The straightforward formulas for the $B_n^{(m)}$ take on the order of V^2 operations to compute them from the $A_n^{(m)}$. As a result, since there would again be on the order of $\log R$ levels in the hierarchy of expansions, with each level having only $1/3$ or $1/2$ of the expansions in the level below, obtaining all the expansion coefficients would take on the order of

$$k_1 V + V^2 R \tag{4.7.2.5}$$

operations. For $k_1 \approx R$ and $V \approx \log R$, as would be true for the zeta function algorithm in the absence of memory constraints, and also for the Coulomb or gravitational potential calculations of [GR1] and related papers, this is about the same operation count as for the present algorithm (see (4.7.2.1)). However, for present and foreseeable computations of the zeta function at large heights, R is much smaller

than k_1 ($R \approx 1.6 \times 10^7$ as compared to $k_1 \approx 1.5 \times 10^9$ in the largest computation of this paper), and so the Greengard-Rokhlin algorithm is likely to give much faster rational function evaluation. An order of magnitude improvement seems likely in the time needed to evaluate the expansion coefficients. At present, Taylor series expansions consume about half of the time, so even eliminating them entirely would only double the speed of the program. However, with faster coefficient expansion techniques, one could also use these methods to evaluate contributions to $f(\omega^h)$ of b_k that are closer to ω^h than is done at present, and this would give much greater gains in efficiency. Moreover, once these parts of the program were improved, the improvements to other parts (such as that of evaluating dp values of $\log k$) that have been suggested would become much more significant, and all of them together could increase the speed of the entire algorithm by an order of magnitude, especially if assembly language was used as suggested in Section 4.7.1.

The basic idea of translating an expansion that is at the heart of the Greengard-Rokhlin algorithm for pole expansions can also be used for Taylor expansions of the kind that are used in the present algorithm, but it is not as efficient in this setting. Also, since the $B_n^{(m)}$ are derived from the $A_n^{(m)}$ by a convolution, one can do this computation in fewer than the V^2 steps of the straightforward algorithm, by using FFT-based methods. Greengard and Rokhlin [GR2] report some improvements obtained this way, but they are only about 2 or 3 for 2-dimensional problems, and about 8 for 3-dimensional ones. Since the zeta function problem is essentially a 1-dimensional one (with all the poles and points of evaluation on the unit circle), we might expect small improvements from this source. This might be counteracted to some extent by the fact that the order of expansion V that is used with the zeta function is higher than in [GR2], so the overhead might be smaller, and noticeable savings might still be obtained.

One aspect of this multipole expansions of Greengard and Rokhlin that would have to be investigated carefully before their algorithm could be used for the zeta

function computation is its accuracy. However, based on the results reported so far [GR1, GR2], that is not likely to be a problem.

4.7.3. Computations of low zeros

The new algorithm is much more efficient than the implementation of the standard Riemann-Siegel formula evaluation in [Od2] even around zero number 10^{12}. However, this advantage might not hold or be as noticeable around zero 1.5×10^9, especially if one were only interested in separating zeros, and not computing them accurately (so that only about 1.2 evaluations of $Z(t)$ would be needed per zero, instead of the 10 or so of the current implementation of the algorithm of [OS]). Thus if one were interested in extending the numerical verification of the RH beyond the 1.5×10^9 zeros of [LRW2], the present implementation might not help much. This is due to a large extent to the design of the program, which was aimed at computing around the 10^{20}-th zero, and so various parameters were chosen with that goal in mind. It is likely that the program could be rewritten to be much faster at lower heights, and with more extensive use of *dp* arithmetic rigorous error analysis could be performed for it, but this would represent a substantial programming effort.

What we present here is a combination of several techniques that ought to give a simple algorithm for computing $Z(t)$ that ought to be about an order of magnitude faster than the algorithm of [LRW2], and for which rigorous error analysis could be performed. The basic idea is to again compute $Z_1(t)$ in the standard way, and to compute $F(t)$ on a uniformly spaced grid of points T, $T + \delta, \ldots$, and to use band-limited function interpolation to then obtain $F(t)$ at intermediate points, as is explained in Section 4.4. The band-limited function interpolation method errors can be bounded rigorously. If $k_0 \approx t^{1/4}$, it would suffice to compute $F(t)$ once every 4 Gram intervals, but to shorten the interpolation computations and to control the errors better (through having to sum fewer terms in the series in (4.4.15)), it could be preferable to sample somewhat more frequently.

The evaluation of $F(t)$ in the suggested method would be performed not by the method of [OS], but by forming arrays a_k and b_k, $k_0 \le k \le k_1$, with

$$a_k = 2k^{-1/2}e^{iT\log k}, \quad b_k = e^{i\delta \log k} \ . \tag{4.7.3.1}$$

$F(T)$ would then be the sum of the a_k. Next, we assign to a_k the value of $a_k \cdot b_k$, and sum the new a_k, to obtain $F(T+\delta)$. Repeating this operation yields all the $F(T+j\delta)$. Since complex *sp* multiplications and additions are vectorized by the Cray Fortran compiler, this method would be fast, as was already noted in [Od2]. (To avoid loss of accuracy in repeated multiplications, it would be advisable to use this method only on short stretches of the grid $T, T+\delta, \ldots$, so that at $t = T+100j\delta$, for example, one would recompute $a_k = 2k^{-1/2}\exp(it\log k)$ from scratch.)

A further improvement can be obtained by using the Euler product, as was suggested by A. Schönhage in a slightly different context. To compute $F(t)$, we do not need to compute all the $2k^{-1/2}\exp(it\log k)$ for $k_0 \le k \le k_1$ explicitly. Instead, we can compute them for all k, $2 \le k \le k_1$, such that $(k,P) = 1$, where $P = 2\times 3\times 5\times\cdots\times p_h$ is the product of the first h primes (with h small, say $h = 4$ or 5). (This computation would be done by a modification of the method presented above.) Then, to obtain $F(t)$, we can compute, for each k, $1 \le k \le k_1$,

$$2k^{-1/2}e^{it\log k} \sum_{\substack{g\in Q \\ k_0 \le kg \le k_1}} g^{-1/2}e^{it\log g} \ , \tag{4.7.3.2}$$

where Q is the set of integers all of whose prime factors are $\le p_h$. Since the sum in (4.7.3.2) would be the same for many k, this operation would be vectorizable.

The methods presented in this section could also be useful for very accurate computations of high zeros. If one were to find an extreme example of Lehmer's phenomenon at large heights, or even a suspected counterexample to the RH, where it would be necessary to obtain more accurate values of $Z(t)$ than are given by the present implementation of the algorithm of [OS], writing an improved version of this algorithm with a guaranteed error bound would be laborious, and might require a

prohibitive amount of time to run. On the other hand, since values of $Z(t)$ in only a short interval would likely be needed, the method of this section (combined with the suggestion at the end of Section 4.7.1 about increased accuracy) might be adequate to resolve any uncertainties.

Acknowledgements

Since this paper is essentially a continuation of [Od2], the author would like to thank P. Barrucand, R. A. Becker, O. Bohigas, R. P. Brent, W. S. Cleveland, F. J. Dyson, P. X. Gallagher, E. Grosswald, D. A. Hejhal, B. Kleiner, H. J. Landau, S. P. Lloyd, C. L. Mallows, H. L. Montgomery, A. Pandey, H.J. J. te Riele, L. Schoenfeld, N. L. Schryer, D. S. Slepian, and D. T. Winter, all of whom helped in the preparation of that earlier paper, and many of whom helped with this one. Additional help in the preparation of this paper was provided by D. H. Bailey, M. V. Berry, J. B. Conrey, L. Greengard, A. Ivić and B. F. Logan. D. A. Hejhal and J. van de Lune provided exceptionally extensive and useful comments. A. Schönhage is due special thanks for the joint work that resulted in the algorithm that made these computations possible.

The author is grateful to W. M. Coughran, P. Glick, R. H. Knag, and P. J. Weinberger for help in obtaining time and storage space on computers and to D. H. Bailey for providing fast Fourier transform programs.

Bibliography

[AKW] A. V. Aho, B. W. Kernighan, and P. J. Weinberger, *The AWK Programmming Language,* Addison-Wesley, 1988.

[AGR] J. Ambrosiano, L. Greengard, and V. Rokhlin, The fast multipole method for gridless particle simulations, *Computer Physics Comm. 48* (1988), 117–125.

[AH] K. Appel and W. Haken, The four color proof suffices, *Math. Intelligencer 8* (no. 1) (1986), 10–20, 58.

[Bail] D. H. Bailey, A high-performance FFT algorithm for vector supercomputers, *Intern. J. Supercomputer Appl. 2, no. 1,* (1988), 82–87.

[Bala] R. Balasubramanian , On the frequency of Titchmarsh's phenomenon for $\zeta(s)$. IV, *Hardy-Ramanujan J. 9* (1986), 1–10.

[BR] R. Balasubramanian and K. Ramachandra, On the frequency of Titchmarsh's phenomenon for $\zeta(s)$. III, *Proc. Indian Acad. Sci., Sect. A 86* (1977), 341–351.

[BC] R. A. Becker and J. M. Chambers, *S: An Interactive Environment for Data Analysis and Graphics*, Wadsworth, 1984.

[Be1] M. V. Berry, Semiclassical theory of spectral rigidity, *Proc. Royal Soc. London A 400* (1985), 229–251.

[Be2] M. V. Berry, Riemann's zeta function: A model for quantum chaos?, pp. 1–17 in *Quantum Chaos and Statistical Nuclear Physics*, T. Seligman and H. Nishioka, eds., Lecture Notes in Physics #263, Springer, 1986.

[Be3] M. V. Berry, Quantum chaology, *Proc. Royal Soc. London A 413* (1987), 183–198.

[Be4] M. V. Berry, Semiclassical formula for the number variance of the Riemann zeros, *Nonlinearity 1* (1988), 399–407.

[Bil] P. Billingsley, *Probability and Measure*, Wiley, 1979.

[BG] O. Bohigas and M.-J. Giannoni, Chaotic motion and random matrix theories, pp. 1–99 in *Mathematical and Computational Methods in Nuclear Physics*, J. S. Dehesa, J. M. G. Gomez, and A. Polls, eds., Lecture Notes in Physics #209, Springer 1984.

[BGS1] O. Bohigas, M. J. Giannoni, and C. Schmit, Characterization of chaotic quantum spectra and universality of level fluctuation laws, *Physical Rev. Letters 52* (1984), 1–4.

[BGS2] O. Bohigas, M. J. Giannoni, and C. Schmit, Spectral properties of the Laplacian and random matrix theories, *J. Physique-Lettres 45* (1984), L1015–22.

[BHP] O. Bohigas, R. U. Haq, and A. Pandey, Higher-order correlations in spectra of complex systems, *Phys. Rev. Letters 54* (1985), 1645–1648.

[BH] E. Bombieri and D. A. Hejhal, Sur les zéros des fonctions zêta d'Epstein, *Comp. Rend. Acad. Sci. Paris, Sér. I, 304* (1987), 213–217.

[BI] E. Bombieri and H. Iwaniec, On the order of $\zeta(1/2+it)$, *Ann. Scuola Norm. Sup. Pisa, Ser. IV, 13* (1986), 449–472.

[BB] J. M. Borwein and P. B. Borwein, *Pi and the AGM*, Wiley-Interscience, 1987.

[Br1] R. P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, 1973.

[Br4] R. P. Brent, A Fortran multiple precision arithmetic package, *ACM Trans. Math. Software 4* (1978), 57–70.

[Br5] R. P. Brent, On the zeros of the Riemann zeta function in the critical strip, *Math. Comp., 33* (1979), 1361–1372.

[BFFMPW] T. A. Brody, J. Flores, J. P. French, P. A. Mello, A. Pandey, and S. S. M. Wong, Random-matrix physics: spectrum and strength fluctuations, *Rev. Modern Physics 53* (1981), 385–479.

[Br] W. S. Brown, A simple but realistic model of floating-point computations, *ACM Trans. Math. Software 7* (1981), 445–480.

[BSS] P. L. Butzer, W. Splettstösser, and R. L. Stens, The sampling theorem and linear prediction in signal analysis, *Jber. d. Deutschen Math.-Verein. 90* (1988), 1–70.

[CGR] J. Carrier, L. Greengard, and V. Rokhlin, A fast adaptive algorithm for particle simulations, *SIAM J. Stat. Sci. Comp. 9* (1988), 669–686.

[CCKT] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey, *Graphical Methods for Data Analysis*, Wadsworth, 1983.

[Clev] W. S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Amer. Statist. Assoc. 74* (1979), 829–836.

[CM1] J. des Cloizeaux and M. L. Mehta, Some asymptotic expressions for prolate spheroidal functions and for the eigenvalues of differential and integral equations of which they are solutions, *J. Math. Phys. 13* (1972), 1745–1754.

[CM2] J. des Cloizeaux and M. L. Mehta, Asymptotic behavior of spacing distributions for the eigenvalues of random matrices, *J. Math. Phys. 14* (1973), 1648–1650.

[CG1] J. B. Conrey and A. Ghosh, On mean values of the zeta-function, *Mathematika 31* (1984), 159–161.

[CG2] J. B. Conrey and A. Ghosh, A mean value theorem for the Riemann zeta-function at its relative extrema on the critical line, *J. London Math. Soc. (2) 32* (1985), 193–202.

[CGGGH] J. B. Conrey, A. Ghosh, D. A. Goldston, S. M. Gonek and D. R. Heath-Brown, On the distribution of gaps between zeros of the zeta function, *Quarterly J. of Math. Oxford* (2), *36* (1985), 43–51.

[CGG1] 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.

[CGG2] J. B. Conrey, A. Ghosh, and S. M. Gonek, Large gaps between zeros of the zeta-function, *Mathematika 33* (1986), 212–238.

[CR] F. D. Crary and J. B. Rosser, High precision coefficients related to the zeta function, MRC Technical Summary Report #1344, Univ. of Wisconsin, Madison, May 1975, 171 pp.; reviewed by R. P. Brent in *Math. Comp. 31* (1977), 803–804.

[Dav] D. Davies, An approximate functional equation for Dirichlet L-functions, *Proc. Royal Soc. Ser. A 284* (1965), 224–236.

[Deu] M. Deuring, Asymptotische Entwicklungen der Dirichletschen L-Reihen, *Math. Ann. 168* (1967), 1–30.

[Dy1] F. J. Dyson, Statistical theory of the energy levels of complex systems. II, *J. Math. Phys. 3* (1962), 157–165.

[Ed] H. M. Edwards, *Riemann's Zeta Function*, Academic Press, 1974.

[Fel] W. Feller, *An Introduction to Probability Theory and its Applications,* vol. 2, 2nd ed., Wiley, 1971.

[Fu1] A. Fujii, On the zeros of Dirichlet L-functions. I. *Trans. Amer. Math. Soc. 196* (1974), 225–235.

[Fu2] A. Fujii, On the uniformity of the distribution of zeros of the Riemann zeta function, *J. reine angew. Math. 302* (1978), 167–205.

[Fu3] A. Fujii, A prime number theorem in the theory of the Riemann zeta function, *J. reine angew. Math. 307/308* (1979), 113–129.

[Fu4] A. Fujii, On the zeros of Dirichlet L-functions. II (With corrections to "On the zeros of Dirichlet L-function. I" and the subsequent papers), *Trans. Amer. Math. Soc. 267* (1981), 33–40.

[Fu5] A. Fujii, On the uniformity of the distribution of zeros of the Riemann zeta function (II), *Comm. Math. Univ. Sancti Pauli 31* (1982), 99–113.

[Fu6] A. Fujii, Zeros, primes and rationals, *Proc. Japan Acad. Ser. A, 58* (1982), 373–376.

[Fu7] A. Fujii, Uniform distribution of the zeros of the Riemann zeta function and the mean value theorems of Dirichlet L-functions, *Proc. Japan Acad. Ser. A, 63* (1987), 370–373.

[Fu8] A. Fujii, Gram's law for the zeta zeros and the eigenvalues of gaussian unitary ensembles, *Proc. Japan Acad. Ser. A, 63* (1987), 392–395.

[Fu9] A. Fujii, Zeta zeros and Dirichlet L-functions, *Proc. Japan Acad. Ser. A, 64* (1988), 215–218.

[Gab] W. Gabcke, *Neue Herleitung und explicite Restabschätzung der Riemann-Siegel-Formel*, Ph.D. Dissertation, Göttingen, 1979.

[Gal2] P. X. Gallagher, Pair correlation of zeros of the zeta function, *J. reine angew. Math. 362* (1985), 72–86.

[Gal3] P. X. Gallagher, Applications of Guinand's formula, pp. 135–157 in *Analytic Number Theory and Diophantine Problems*, A. C. Adolphson, J. B. Conrey, A. Ghosh, and R. I. Yager, eds., Birkhäuser, 1987.

[Gal4] P. X. Gallagher, A double sum over primes and zeros of the zeta function, in *Number Theory, Trace Formulas, and Discrete Groups,* K. E. Aubert, E. Bombieri, and D. M. Goldfeld, eds., Proc. 1987 Selberg Symposium, Academic Press, 1989, pp. 229–240.

[GM] P. X. Gallagher and J. H. Mueller, Primes and zeros in short intervals, *J. reine angew. Math. 303/304* (1978), 205–220.

[Gen] W. M. Gentleman, Fast Fourier transforms — for fun and profit, *AFIPS Proc. 29* (1966), 563–578.

[Gh1] A. Ghosh, On Riemann's zeta-function–sign changes of $S(T)$, pp. 25–46 in *Recent Progress in Analytic Number Theory*, vol. 1, H. Halberstam and C. Hooley, eds., Academic Press, 1981.

[Gh2] A. Ghosh, On the Riemann zeta-function–mean value theorems and the distribution of $|S(t)|$, *J. Number Theory 17* (1983), 93–102.

[Go1] D. A. Goldston, Prime numbers and the pair correlation of zeros of the zeta function, pp. 82–91 in *Topics in Analytic Number Theory*, S. W. Graham and J. D. Vaaler, eds., Univ. Texas Press, 1985.

[Go2] D. A. Goldston, On the function $S(T)$ in the theory of the Riemann zeta function, *J. Number Theory 27* (1987), 149–177.

[Go3] D. A. Goldston, On the pair correlation conjecture for zeros of the Riemann zeta function, *J. reine angew. Math. 385* (1988), 24–40.

[GG] D. A. Goldston and S. M. Gonek, A note on the number of primes in short intervals, to be published.

[GHB] D. A. Goldston and D. R. Heath-Brown, A note on the differences between consecutive primes, *Math. Ann. 266* (1984), 317–320.

[GM] D. A. Goldston and H. L. Montgomery, Pair correlation of zeros and primes in short intervals, pp. 183–203 in *Analytic Number Theory and Diophantine Problems*, A. C. Adolphson, J. B. Conrey, A. Ghosh, and R. I. Yager, eds., Birkhäuser, 1987.

[Gon0] S. M. Gonek, *Analytic Properties of Zeta and L-Functions*, Ph.D. Dissertation, Univ. Michigan, 1979.

[Gon1] S. M. Gonek, Mean values of the Riemann zeta-function and its derivatives, *Invent. math. 75* (1984), 123–141.

[Gon2] S. M. Gonek, A formula of Landau and mean values of $\zeta(s)$, pp. 92–97 in *Topics in Analytic Number Theory*, S. W. Graham and J. D. Vaaler, eds., Univ. Texas Press, 1985.

[Gon3] S. M. Gonek, On negative moments of the Riemann zeta function, to be published.

[GK] S. W. Graham and G. Kolesnik, One and two dimensional exponential sums, pp. 205–222 in *Analytic Number Theory and Diophantine Problems*, A. C. Adolphson, J. B. Conrey, A. Ghosh, and R. I. Yager, eds., Birkhäuser, 1987.

[GG] L. Greengard and W. D. Gropp, A parallel version of the fast multipole algorithm, to be published.

[GR1] L. Greengard and V. Rokhlin, A fast algorithm for particle simulations, *J. Computational Phys. 73* (1987), 325–348.

[GR2] L. Greengard and V. Rokhlin, Rapid evaluation of potential fields in three dimensions, to be published.

[GR3] L. Greengard and V. Rokhlin, On the efficient implementation of the fast multipole algorithm, to be published.

[Gu1] A. P. Guinand, A summation formula in the theory of prime numbers, *Proc. London Math. Soc.* (2) *50* (1948), 107–119.

[Gu2] A. P. Guinand, Fourier reciprocities and the Riemann zeta-function, *Proc. London Math. Soc. (2) 51* (1949), 401–414.

[Gut] M. C. Gutzwiller, Stochastic behavior in quantum scattering, *Physica 7D* (1983), 341–355.

[HI] J. L. Hafner and A. Ivić, On the mean-square of the Riemann zeta-function on the critical line, J. Number Theory *32* (1989), 151–191.

[HMF] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, eds., National Bureau of Standards, 9th printing, 1970.

[HPB] R. U. Haq, A. Pandey, and O. Bohigas, Fluctuation properties of nuclear energy levels: Do theory and experiment agree? *Physical Rev. Letters 48* (1982), 1086–1089.

[HB1] D. R. Heath-Brown, Gaps between primes and the pair correlation of zeros of the zeta-function, *Acta Arith. 41* (1982), 85–99.

[HB2] D. R. Heath-Brown, Fractional moments of the Riemann zeta-function. II, to be published.

[Hej1] D. A. Hejhal, The Selberg trace formula and Riemann zeta function, *Duke Math. J. 43* (1976), 441–482.

[Hej5] D. A. Hejhal, Zeros of Epstein zeta functions and supercomputers, pp. 1362–1384 in *Proc. Intern. Congress Math. 1986*, Amer. Math. Soc., 1987.

[Hej6] D. A. Hejhal, On the distribution of $\log|\zeta'(1/2+it)|$, in *Number Theory, Trace Formulas, and Discrete Groups,* K. E. Aubert, E. Bombieri, and D. M. Goldfeld, eds., Proc. 1987 Selberg Symposium, Academic Press, 1989, pp. 343–370.

[Hig] J. R. Higgins, Five short stories about the cardinal series, *Bull. Amer. Math. Soc. 12* (1985), 45–89.

[Hl] E. Hlawka, Über die Gleichverteilung gewisser Folgen, welche mit den Nullstellen der Zetafunktion zusammenhängen, *Sitzungsber. Öst. Akad. Wiss. Math.-Naturw. Kl. II 184* (1975), 459–471.

[Hux] M. N. Huxley, Exponential sums and the Riemann zeta function. IV, in preparation.

[Iv] A. Ivić, *The Riemann Zeta-function*, Wiley, 1985.

[Jer] A. J. Jerri, The Shannon sampling theorem — its various extensions and aplications: a tutorial review, *Proc. IEEE 65* (1977), 1565–1596.

[Joy1] D. Joyner, *Distribution theorems for L-functions*, Longman, 1986.

[Joy2] D. Joyner, On the Dyson-Montgomery hypothesis, preprint.

[Jut] M. Jutila, On the value distribution of the zeta-function on the critical line, *Bull. London Math. Soc. 15* (1983), 513–518.

[Kai] J. F. Kaiser, Design methods for sampled data filters, pp. 221–236 in *Proc. First Allerton Conf. Circuit and System Theory,* Monticello, Ilinois, 1963.

[KW] E. Karkoschka and P. Werner, Einige Ausnahmen zur Rosserschen Regel in der Theorie der Riemannschen Zetafunktion, *Computing 27* (1981), 57–69.

[Kat] J. Katzenelson, Computational structure of the N-body problem, *SIAM J. Stat. Sci. Comp. 10* (1989), 787–815.

[KS] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Griffin, 1981.

[Ko] G. Kolesnik, On the method of exponent pairs, *Acta Arith. 45* (1985), 115–143.

[LO2] J. C. Lagarias and A. M. Odlyzko, Solving low-density subset sum problems, *J. ACM 32* (1985), 229–246. (Preliminary version in *Proc. 24–th IEEE Foundations Computer Science Symp.,* pp. 1–10, 1983.)

[LO4] J. C. Lagarias and A. M. Odlyzko, Computing $\pi(x)$: An analytic method, *J. Algorithms 8* (1987), 173–191.

[Lan1] E. Landau, Über die Nullstellen der Zetafunktion, *Math. Ann. 71* (1911), 548–564.

[Lan] O. E. Lanford III, Computer-assisted proofs in analysis, pp. 1385–1394 in *Proc. Intern. Congress Math. 1986*, Amer. Math. Soc. 1987.

[Lau1] A. Laurinchikas, Riemann zeta function on the critical line, *Litovsk. Mat. Sb. 25, no. 2,* (1985), 114–118. (In Russian.) English translation in *Lithuanian Math. J. 25* (1985), 145–148.

[Lau2] A. Laurinchikas, Moments of the Riemann zeta-function on the critical line, *Mat. Zametki 39* (1986), 483–493. (In Russian.) English translation in *Math. Notes Akad. Sci. USSR 39* (1986), 267–272.

[Lau3] A. Laurinchikas, Limit theorem for the Riemann zeta-function on the critical line. I, *Litovsk. Mat. Sb. 27, no. 1,* (1987), 113–132. (In Russian.) English translation in *Lithuanian Math. J. 27* (1987), 63–75.

[Lau4] A. Laurinchikas, Limit theorem for the Riemann zeta-function on the critical line. II, *Litovsk. Mat. Sb. 27, no. 3,* (1987), 489–500. (In Russian.) English translation in *Lithuanian Math. J. 27* (1987), 236–243.

[Lau5] A. Laurinchikas, A limit theorem for Dirichlet L-functions on the critical line, *Litovsk. Mat. Sb. 27, no. 1,* (1987), 699–710. (In Russian.)

[LLL] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, Factoring polynomials with rational coefficients, *Math. Ann. 261* (1982), 515–534.

[Li] J. E. Littlewood, On the zeros of the Riemann zeta-function, *Proc. Cambridge Philos. Soc. 22* (1924), 295–318.

[Log1] B. F. Logan, Optimal truncation of the Hilbert transform kernel for bounded high-pass functions, pp. 10–12 in *Proc. 5–th Annual Princeton Conf. Information Sci. Systems*, Princeton Univ., 1971.

[Log2] B. F. Logan, Bounds for the tails of sharp-cutoff filter kernels, *SIAM J. Math. Anal. 19* (1988), 372–376.

[Lr2] D. H. Lehmer, On the roots of the Riemann zeta-function, *Acta Math. 95* (1956), 291–298.

[LRW1] J. van de Lune, H. J. J. te Riele, and D. T. Winter, Rigorous high speed separation of zeros of Riemann's zeta function, Report NW 113/81, Mathematical Center, Amsterdam, 1981.

[LRW2] J. van de Lune, H. J. J. te Riele, and D. T. Winter, On the zeros of the Riemann zeta function in the critical strip. IV., *Math. Comp. 46* (1986), 667–681.

[Meh] M. L. Mehta, *Random Matrices*, 2nd revised and enlarged ed., Academic Press, 1991.

[MdC] M. L. Mehta and J. des Cloizeaux, The probabilities for several consecutive eigenvalues of a random matrix, *Indian J. Pure Appl. Math. 3* (1972), 329–351.

[Mon1] H. L. Montgomery, The pair correlation of zeros of the zeta function, pp. 181–193 in *Analytic Number Theory,* H. G. Diamond, ed., Proc. Symp. Pure Math. *24*, Amer. Math. Soc., Providence 1973.

[Mon2] H. L. Montgomery, Distribution of zeros of the Riemann zeta function, *Proc. Int. Congress Math. Vancouver* (1974), 379–381.

[Mon3] H. L. Montgomery, Extreme values of the Riemann zeta function, *Comm. Math. Helv. 52* (1977), 511–518.

[Mon6] H. L. Montgomery, Selberg's work on the zeta function, in *Number Theory, Trace Formulas, and Discrete Groups,* K. E. Aubert, E. Bombieri, and D. M. Goldfeld, eds., Proc. 1987 Selberg Symposium, Academic Press, 1989, pp. 157–168.

[MO] H. L. Montgomery and A. M. Odlyzko, Gaps between zeros of the zeta function, pp. 1079–1106 in *Topics in Classical Number Theory: Coll. Math. Soc. Janos Bolyai 34,* G. Halász, ed., North-Holland, 1984.

[MW] H. L. Montgomery and P. J. Weinberger, Notes on small class numbers, *Acta Arith. 24* (1973/74), 529–542.

[Mos1] J. Moser, On a certain sum in the theory of the Riemann zeta function, (in Russian), *Acta Arith. 31* (1976), 31–43.

[Mos2] J. Moser, On a Hardy-Littlewood theorem in the theory of the Riemann zeta function, (in Russian), *Acta Arith. 31* (1976), 45–51.

[Mos3] J. Moser, On Gram's law in the theory of the Riemann zeta function, (in Russian), *Acta Arith. 32* (1977), 107–113.

[Mos4] J. Moser, Proof of a hypothesis of E. C. Titchmarsh in the theory of the Riemann zeta function, (in Russian), *Acta Arith. 36* (1980), 147–156.

[Mos5] J. Moser, On the roots of the equations $Z'(t) = 0$, (in Russian), *Acta Arith. 40* (1981), 79–89.

[Mos6] J. Moser, Corrections to the papers: *Acta Arith. 31* (1976), pp. 31–43; *31* (1976), pp. 45–51; *35* (1979), pp. 403–404, (in Russian), *Acta Arith. 40* (1981), 97–107.

[Mos7] J. Moser, New consequences of the Riemann-Siegel formula, (in Russian), *Acta Arith. 42* (1982), 1–10.

[Mos8] J. Moser, On a certain biquadratic sum in the theory of the Riemann zeta function, (in Russian), *Acta Math. Univ. Comen. 42–43* (1983), 35–39.

[Mos9] J. Moser, Properties of the sequence $Z[t_v(\tau)]$ in the theory of the Riemann zeta function, (in Russian), *Acta Math. Univ. Comen. 42–43* (1983), 55–63.

[Mos10] J. Moser, On some lower bounds for the distance of consecutive zeros of the function $\zeta(1/2+it)$, (in Russian), *Acta Math. Univ. Comen. 44–45* (1984), 75–80.

[Mos11] J. Moser, On a cubic formula in the theory of the Riemann zeta function, (in Russian), *Acta Math. Univ. Comen. 44–45* (1984), 81–89.

[Mos12] J. Moser, New mean value theorems for the function $|\zeta(1/2+it)|^2$, (in Russian), *Acta Math. Univ. Comen. 46–47* (1985), 21–40.

[Mos13] J. Moser, On the behavior of positive and negative values of the function $Z(t)$ in the theory of the Riemann zeta function, (in Russian), *Acta Math. Univ. Comen. 46–47* (1985), 41–48.

[Mos14] J. Moser, On a cubic sum in the theory of the Riemann zeta function, (in Russian), *Acta Math. Univ. Comen. 46–47* (1985), 63–74.

[Mue1] J. H. Mueller, On the Riemann zeta function $\zeta(s)$ — gaps between sign changes of $S(t)$, *Mathematika 29* (1982), 264–269.

[Mue2] J. H. Mueller, Arithmetic equivalent of essential simplicity of zeta zeros, *Trans. Amer. Math. Soc. 275* (1983), 175–183.

[Od0] A. M. Odlyzko, Applications of symbolic mathematics to mathematics, in *Applications of Computer Algebra*, R. Pavelle, ed., Kluwer-Nijhoff, 1985, pp. 95–111.

[Od1] A. M. Odlyzko, New analytic algorithms in number theory, pp. 466–475 in *Proc. Intern. Congress Math. 1986*, Amer. Math. Soc. 1987.

[Od2] A. M. Odlyzko, On the distribution of spacings between zeros of the zeta function, *Math. Comp. 48* (1987), 273–308.

[Od3] A. M. Odlyzko, Zeros of the Riemann zeta function: Conjectures and computations, manuscript in preparation.

[Od4] A. M. Odlyzko, The number variance of zeros of the Riemann zeta function, manuscript in preparation.

[OtR] A. M. Odlyzko and H. J. J. te Riele, Disproof of the Mertens conjecture, *J. reine angew. Math. 357* (1985), 138–160.

[OS] A. M. Odlyzko and A. Schönhage, Fast algorithms for multiple evaluations of the Riemann zeta function, *Trans. Amer. Math. Soc. 309* (1988), 797–809.

[Oz1] A. E. Ozluk, *Pair correlation of zeros of Dirichlet L-functions*, Ph.D. dissertation, Univ. of Michigan, Ann Arbor, Mich., 1982.

[Oz2] A. E. Ozluk, On the pair correlation of zeros of Dirichlet L-functions, *Proc. First Conf. Canadian Number Theory Assoc. (Banff, 1988)*, R. A. Mollin, ed., W. de Gruyter, 1989, to appear.

[Por] C. E. Porter, ed., *Statistical Theories of Spectra: Fluctuations*, Academic Press, 1965.

[RK] S. P. Radziszowski and D. L. Kreher, Solving subset sum problems with the L^3 algorithm, *J. Combin. Math. Combin. Comput. 3*, (1988), 49–63.

[Sch1] C. P. Schnorr, A more efficient algorithm for a lattice basis reduction, *J. Algorithms 9* (1988), 47–62.

[Sch2] C. P. Schnorr, A hierarchy of polynomial time lattice basis reduction algorithms, *Theoretical Computer Science 53* (1987), 201–224.

[Schr1] N. L. Schryer, A test of a computer's floating-point arithmetic unit, AT&T Bell Laboratories Computing Science Technical Report #89, 1981.

[Schr2] N. L. Schryer, A case study in testing: floating-point arithmetic, to be published.

[Sel2] A. Selberg, On the remainder in the formula for $N(T)$, the number of zeros of $\zeta(s)$ in the strip $0 < t < T$, *Avh. Norske Vid. Akad. Oslo I. Mat.-Naturvid. Kl., no. 1* (1944), 1–17.

[Sel3] A. Selberg, Contributions to the theory of the Riemann zeta-function, *Arch. for Math. og Naturv. B, 48* (1946), 89–155.

[SW] D. Shanks and J. W. Wrench, Jr., Calculation of π to 100,000 decimals, *Math. Comp. 16* (1962), 76–99.

[Sie1] C. L. Siegel, Über Riemanns Nachlass zur analytischen Zahlentheorie, *Quellen und Studien zur Geschichte der Math. Astr. Phys. 2* (1932), 45–80. Reprinted in C. L. Siegel, *Gesammelte Abhandlungen*, Springer, 1966, Vol. 1, pp. 275–310.

[Sie2] C. L. Siegel, Contributions to the theory of the Dirichlet L-series and the Epstein zeta functions, *Ann. Math. 44* (1943), 143–172. Reprinted in C. L. Siegel, *Gesammelte Abhandlungen*, Springer, 1966, Vol. 2, pp. 360–389.

[SF] E. H. Spafford and J. C. Flaspohler, A report on the accuracy of some floating point math functions on selected computers, Georgia Inst. Tech., School of Inform. Comp. Sci., Report GIT-ICS 85/06, revised Jan. 1986.

[St] H. M. Stark, On complex quadratic fields with class number two, *Math. Comp. 29* (1975), 289–302.

[Tit0] E. C. Titchmarsh, On van der Corput's method and the zeta-function of Riemann. IV, *Quart. J. Math. 5* (1934), 98–105.

[Tit1] E. C. Titchmarsh, The zeros of the Riemann zeta-function, *Proc. Royal Soc. London 151* (1935), 234–255 and *157* (1936), 261–263.

[Tit2] E. C. Titchmarsh, *The Theory of the Riemann Zeta-function*, 2nd ed. (revised by D. R. Heath-Brown), Oxford Univ. Press, 1986.

[Ts1] K.-M. Tsang, *The Distribution of the Values of the Riemann Zeta-function*, Ph.D. Dissertation, Princeton, 1984.

[Ts2] K.-M. Tsang, Some Ω-theorems for the Riemann zeta-function, *Acta Arith. 46* (1986), 369–395.

[VB] A. L. Van Buren, A Fortran computer program for calculating the linear prolate functions, Report 7994, Naval Research Laboratory, Washington, May 1976.

[vdL] J. van de Lune, Some observations concerning the zero-curves of the real and imaginary parts of Riemann's zeta function, Report ZW 201/83, Mathematical Center Amsterdam, December 1983.

[Wa] N. Watt, Exponential sums and the Riemann zeta function. II, *J. London Math. Soc.*, to appear.

[We1] A. Weil, Sur les "formules explicites" de la theorie des nombres premiers, *Comm. Sem. Math. Univ. Lund*, tome supplementaire (1952), 252–265.

[Whit] J. M. Whittaker, *Interpolatory Function Theory,* Cambridge Univ. Press, 1935.

[WR] D. Winter and H. te Riele, Optimization of a program for the verification of the Riemann hypothesis, *Supercomputer 5* (1985), 29–32.

[Zh] Feng Zhao, An $O(N)$ algorithm for three-dimensional N-body simulations, MIT AI Lab. report #995, October 1987.

[ZJ] Feng Zhao and L. Johnsson, The parallel multipole method on the Connection Machine, paper in preparation.

Table 1.1. Several zeros of the Riemann zeta function near zero number 10^{20}. All zeros are of the form $1/2 + i\gamma_n$.

n	$\gamma_n - 15,202,440,115,920,740,000$
$10^{20} - 6$	7267.894628
$10^{20} - 5$	7267.988948
$10^{20} - 4$	7268.077538
$10^{20} - 3$	7268.258252
$10^{20} - 2$	7268.337163
$10^{20} - 1$	7268.563308
10^{20}	7268.629029
$10^{20} + 1$	7268.828625
$10^{20} + 2$	7268.972156
$10^{20} + 3$	7269.122460
$10^{20} + 4$	7269.241484
$10^{20} + 5$	7269.313890

Table 1.2. Large computed sets of zeros of the Riemann zeta function.

N	number of zeros	index of first zero in set	approximate height of zero no. N
10^6	$1,000,1052$	$N + 1$	6.003×10^5
10^{12}	$1,592,196$	$N - 6,032$	2.677×10^{11}
10^{14}	$1,685,452$	$N - 736$	2.251×10^{13}
10^{16}	$16,480,973$	$N - 5,946$	1.941×10^{15}
10^{18}	$16,671,047$	$N - 8,839$	1.706×10^{17}
10^{19}	$16,749,725$	$N - 13,607$	1.608×10^{18}
10^{20}	$175,587,726$	$N - 30,769,710$	1.520×10^{19}
2×10^{20}	$101,305,325$	$N - 633,984$	2.991×10^{19}

Table 2.4.1. Moments of $\delta_n - 1$.

k	$N = 1$	$N = 10^6$	$N = 10^{12}$	$N = 10^{16}$	$N = 10^{18}$	$N = 10^{20}$	$N = 2 \times 10^{20}$	GUE
2	0.161	0.167	0.176	0.177	0.178	0.178	0.178	0.180
3	0.031	0.032	0.035	0.036	0.036	0.037	0.037	0.038
4	0.081	0.088	0.096	0.098	0.098	0.099	0.099	0.101
5	0.046	0.052	0.059	0.061	0.062	0.062	0.062	0.066
6	0.075	0.087	0.100	0.103	0.105	0.106	0.106	0.111
7	0.072	0.089	0.109	0.113	0.115	0.116	0.116	0.124
8	0.103	0.136	0.171	0.178	0.180	0.183	0.183	0.197
9	0.126	0.182	0.246	0.258	0.261	0.266	0.266	0.290
10	0.181	0.283	0.408	0.431	0.434	0.444	0.444	0.488

Table 2.4.2. Moments of $\delta_n + \delta_{n+1} - 2$.

k	$N = 1$	$N = 10^6$	$N = 10^{12}$	$N = 10^{16}$	$N = 10^{18}$	$N = 10^{20}$	$N = 2 \times 10^{20}$	GUE
2	0.207	0.218	0.236	0.241	0.242	0.243	0.243	0.249
3	0.028	0.029	0.027	0.027	0.027	0.028	0.028	0.030
4	0.123	0.143	0.167	0.173	0.175	0.176	0.177	0.185
5	0.047	0.057	0.062	0.064	0.065	0.066	0.066	0.073
6	0.119	0.158	0.204	0.214	0.218	0.220	0.220	0.237
7	0.078	0.113	0.151	0.159	0.162	0.164	0.164	0.185
8	0.155	0.246	0.370	0.393	0.401	0.406	0.407	0.451
9	0.142	0.250	0.423	0.453	0.465	0.470	0.471	0.544
10	0.252	0.482	0.909	0.985	1.016	1.025	1.029	1.178

Table 2.4.3. Moments of $\log \delta_n$, $\delta_n^{-}1$, and δ_n^{-2}.

moments of	$N = 1$	$N = 10^6$	$N = 10^{12}$	$N = 10^{16}$	$N = 10^{18}$	$N = 10^{20}$	$N = 2 \times 10^{20}$	GUE
$\log \delta_n$	-0.0912	-0.0960	-0.1013	-0.1022	-0.1025	-0.1027	-0.1027	-0.1035
δ_n^{-1}	1.2363	1.2534	1.2700	1.2725	1.2733	1.2737	1.2738	1.2758
δ_n^{-2}	2.2235	2.4153	2.5277	2.5309	2.5855	2.5475	2.5545	2.5633

Table 2.4.4. Kolmogorov statistic for δ_n and $\delta_n + \delta_{n+1}$, for blocks of 10^6 zeros.

	δ_n		$\delta_n + \delta_{n+1}$	
	D	prob.	D	prob.
$N = 10^{12}$ vs. GUE	0.00419	10^{-15}	0.00819	10^{-58}
$N = 10^{20}(a)$ vs. GUE	0.00180	3×10^{-3}	0.00318	3×10^{-9}
$N = 10^{20}(b)$ vs. GUE	0.00152	2×10^{-2}	0.00399	3×10^{-14}
$N = 10^{20}(a)$ vs. $N = 10^{20}(b)$	0.00108	0.19	0.00119	0.12
$N = 10^{20}(a)$ vs. $N = 10^{20}(c)$	0.00082	0.51	0.00123	0.10
$N = 10^{20}(b)$ vs. $N - 10^{20}(c)$	0.00089	0.41	0.00096	0.32

Table 2.5.1. Moments of scaled values of $S(t)$ computed from two intervals of 10^6 zeros each near $N = 10^{12}$ and 10^{20}.

k	$N = 10^{12}$	$N = 10^{20}$	normal
1	1.2×10^{-5}	-6.3×10^{-6}	0
2	1.0	1.0	1
3	3.9×10^{-4}	-4.7×10^{-4}	0
4	2.792	2.831	3
5	4.8×10^{-3}	-9.1×10^{-3}	0
6	12.22	12.71	15
7	0.050	-0.140	0
8	70.98	76.57	105
$\|1\|$	0.8058	0.8042	$0.79788\ldots$
$\|3\|$	1.3130	1.3458	$1.5957\ldots$
$\|5\|$	5.597	5.742	$6.3830\ldots$
$1*$	5.9×10^{-6}	-3.2×10^{-6}	
$2*$	0.2330808	0.2606901	

Table 2.5.2. Average number of sign changes of $S(t)$ per Gram interval.

N	$S(t)$ sign changes
10^6	1.719
10^{12}	1.600
10^{14}	1.575
10^{16}	1.556
10^{18}	1.538
10^{19}	1.531
10^{20}	1.524
2×10^{20}	1.522

Table 2.5.3. Largest values of $|S(t)|$ in various data sets and fraction of exceptions to Rosser's rule that had $|S(t)| > 2.3$.

N	largest $S(t)$	fraction of cases with $\lvert S(t)\rvert > 2.3$
10^{12}	2.1918	–
10^{14}	−2.2784	–
10^{16}	−2.4639	0.0123
10^{18}	2.6121	0.0175
10^{19}	−2.5698	0.0162
10^{20}	2.7916	0.0240
2×10^{20}	2.6271	0.0224

Table 2.5.4. Statistics of $S_1(t)$.

	$N = 10^{12}$	$N = 10^{20}$
mean of $S_1(t)^2$	0.0793	0.0793
mean of $S_1(t)^3$	0.0058	0.0058
mean of $S_1(t)^4$	0.0148	0.0148
$\max S_1(t)$	0.966	0.996
$\min S_1(t)$	−0.786	−0.768
no. sign changes	0.120	0.074

Table 2.6.1. Extremal values of δ_n and $\delta_n + \delta_{n+1}$, and the probability that the minimum value of δ_n in the GUE in a sample of the same size would not exceed the minimal value that was found.

N	$\min \delta_n$	$\max \delta_n$	$\min(\delta_n + \delta_{n+1})$	$\max(\delta_n + \delta_{n+1})$	prob. $\min \delta_n$
10^6	0.00545	3.3035	0.2914	4.0683	0.16
10^{12}	0.00649	3.5098	0.2952	4.5833	0.38
10^{14}	0.00935	3.4716	0.2723	4.6564	0.78
10^{16}	0.00454	4.1637	0.1664	4.9921	0.82
10^{18}	0.00112	3.9869	0.1680	5.0401	0.025
10^{19}	0.00090	3.8089	0.1918	5.0588	0.013
10^{20}	0.00197	4.0258	0.1124	5.2125	0.77
2×10^{20}	0.00121	4.0215	0.1377	5.0859	0.18

Table 2.6.2. Frequencies of large and small δ_n and $\delta_n + \delta_n$ (number of cases per million zeros) and the GUE predictions.

N	$\delta_n \le 0.05$	$\delta_n \le 0.1$	$\delta_n \ge 2.8$	$\delta_n + \delta_{n+1} \le 0.6$	$\delta_n + \delta_{n+1} \ge 4$
10^6	121.9	945	87.0	126.9	8.0
10^{12}	126.2	1055	157.6	331.0	94.8
10^{14}	118.7	1103	156.6	329.9	97.9
10^{16}	130.9	1070	164.4	341.1	107.7
10^{18}	135.3	1088	169.9	356.1	108.5
10^{19}	140.5	1084	170.2	362.5	114.0
10^{20}	134.7	1077	174.3	363.3	111.9
2×10^{20}	135.8	1074	172.8	366.8	112.4
GUE	136.8	1088	196.8	386.3	135.7

Table 2.7.1 Autocovariances of the δ_n.

k	$N = 1$	$N = 10^{12}$	$N = 10^{20}$
0	.1607429	0.1754737	0.1781405
1	−.0574023	−0.0576441	−0.0566976
2	−.0126083	−0.0143034	−0.0143122
3	−.0065874	−0.0055030	−0.0065465
4	−.0045317	−0.0026406	−0.0028474
5	−.0031454	−0.0016681	−0.0019375
6	−.0011362	−0.0013422	−0.0014018
7	−.0007084	−0.0009186	−0.0006824
8	−.0013904	−0.0010702	−0.0006266
9	.0013483	−0.0007598	−0.0005397
10	.0034456	−0.0006851	−0.0004818
11	.0018714	−0.0006116	−0.0002820
12	−.0002503	−0.0004058	−0.0004115
13	−.0005412	−0.0006459	−0.0003212
14	.0025227	−0.0005569	−0.0003363
15	.0046388	−0.0007091	−0.0003671
16	.0025451	−0.0001529	−0.0001061
17	.0010829	−0.0000236	−0.0004597
18	−.0001093	0.0004387	−0.0000046
19	−.0057139	0.0001141	−0.0003378
20	−.0133596	−0.0000075	0.0000028
9980		0.0020484	0.0018166
9981		−0.0037100	0.0012394
9982		−0.0030168	0.0003898
9983		0.0029465	−0.0015079
9984		0.0043783	−0.0019355
9985		−0.0010326	−0.0012999
9986		−0.0034815	0.0001715
9987		0.0000487	0.0014113
9988		−0.0012679	0.0021382
9989		−0.0037964	0.0004500
9990		0.0003175	−0.0005050
9991		0.0048778	−0.0014679
9992		0.0062130	−0.0018540
9993		0.0053806	−0.0002132
9994		0.0011459	0.0014712
9995		−0.0048852	0.0013364
9996		−0.0057967	0.0010678
9997		−0.0056723	0.0001780
9998		−0.0034737	−0.0014741
9999		0.0031196	−0.0020779
10000		0.0074084	−0.0014374

Table 2.8.1. Frequency of the Lehmer phenomenon in the $N = 10^{19}$, $N = 10^{20}$, and $10 = 2 \times 10^{20}$ data sets.

x	no. values $< x$
0.0005	6190
0.0004	4422
0.0003	2877
0.0002	1534
0.0001	573
0.00005	208
0.00002	61
0.00001	24

Table 2.9.1. Largest values of $|\zeta(1/2+it)|$ that were found.

N	$\max\|Z(t)\|$
10^{12}	176
10^{14}	246
10^{16}	460
10^{18}	376
10^{19}	448
10^{20}	641
2×10^{20}	628

Table 2.9.2. Frequency of large values of $|\zeta(1/2+it)|$, $N=10^{19}$, $N=10^{20}$, and $N=2\times 10^{20}$.

x	no. values $> x$
250	1851
300	671
350	288
400	111
450	46
500	17

Table 2.10.1. Mean values of $|\zeta(1/2+it)|$.

λ	$r(\lambda, H)$	$c_1(\lambda)$	$c_2(\lambda)$	$c(\lambda)$
.1	1.004	1.0042		
.2	1.034	1.0172		
.3	1.067	1.0381		
.4	1.098	1.0640		
.5	1.123	1.0904		
.6	1.135	1.1113		
.7	1.132	1.1195		
.8	1.107	1.1076		
.9	1.060	1.0690		
1.0	.989	1.0	1.0	1.0
1.1	.896	.901	0.906	
1.2	.787	.776	0.795	
1.3	.667	.637	0.672	
1.4	.554	.494	0.544	
1.5	.426	.360	0.421	
1.6	.319	.246	0.309	
1.7	.229	.157	0.215	
1.8	.156	.092	0.142	
1.9	.101	.050	0.086	
2.0	.0624	.025	0.051	.051
2.1	.0364	.012		
2.2	.0201	.0049		
2.3	.0105	.0019		
2.4	.00522	.00066		
2.5	.00239	.00021		

Table 2.10.2. Negative moments of $|\zeta(1/2+it)|$.

λ	mean values of $\lvert Z(t)\rvert^{-2\lambda}$
0.1	1.06
0.2	1.27
0.3	1.83
0.4	3.77

Table 2.11.1. Moments of the scaled distribution of $\log|\zeta(1/2+it)|$ obtained from 10^6 random samples near zero number N and the moments of the normal distribution.

k	$N=10^{12}$	$N=10^{18}(a)$	$N=10^{18}(b)$	$N=10^{20}(c)$	$N=10^{20}(d)$	normal
1	0.0	0.0	0.0	0.0	0.0	0
2	1.0	1.0	1.0	1.0	1.0	1
3	−0.61867	−0.54505	−0.54199	−0.53625	−0.55069	0
4	4.1319	3.9441	3.9491	3.9233	3.9647	3
5	−9.0528	−7.8024	−7.8610	−7.6238	−7.8839	0
6	44.065	39.717	40.360	38.434	39.393	15
7	−175.39	−159.45	−162.86	−144.78	−148.77	0
8	900.06	930.19	930.70	758.57	765.54	105
9	−4700.06	−6065.28	−5692.4	−4002.5	−3934.7	0
10	27016.2	48430.0	40818.3	24060.5	22722.9	945
1*	−0.0003725	−0.0009607	0.00101075	−0.00159534	0.00054934	
2*	2.29679	2.52283	2.51805	2.57360	2.51778	

Table 2.12.1. Moments of scaled values of $\log|\zeta'(1/2+it)|$ computed from 10^6 zeros for $N = 10^{20}$.

k	scaled moments of $\log\|Z'(\gamma)\|$	moments of normal distribution
1	0.0	0
2	1.0	1
3	− 0.03377	0
4	3.0182	3
5	− 0.59687	0
6	15.4522	15
7	− 9.0568	0
8	115.378	105
9	− 144.031	0
10	1180.33	945
1∗	3.34571	
2∗	12.3312	

Table 2.12.2. Moments of $|\zeta'(1/2+i\gamma_n)|$ divided by conjectured main term, two sets of zeros near zero number 10^{20}.

λ	first set of 5×10^5 zeros	second set of 5×10^5 zeros
−5.	6.08×10^{-22}	4.68×10^{-20}
−4.5	$1.29\times10^{-}16$	5.40×10^{-15}
−4.	$4.21\times10^{-}12$	9.45×10^{-11}
−3.5	2.15×10^{-8}	2.54×10^{-7}
−3.	1.74×10^{-5}	1.07×10^{-4}
−2.5	2.29×10^{-3}	7.41×10^{-3}
−2.	5.29×10^{-2}	9.51×10^{-2}
−1.5	0.275	0.322
−1.	0.640	0.644
−0.5	1.079	1.078
0.	1.0	1.0
0.5	0.436	0.436
1.	8.17×10^{-2}	8.17×10^{-2}
1.5	5.31×10^{-3}	5.32×10^{-3}
2.	9.26×10^{-5}	9.48×10^{-5}
2.5	3.53×10^{-7}	3.83×10^{-7}
3.	2.59×10^{-10}	3.10×10^{-10}
3.5	3.38×10^{-14}	4.65×10^{-14}
4.	$7.50\times10^{-}19$	1.22×10^{-18}
4.5	$2.73\times10^{-}24$	$5.34\times10^{-}24$
5.	1.59×10^{-30}	3.79×10^{-30}

Table 2.13.1. Fractions of Gram blocks of various lengths.

N	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	≥ 8
1	0.8449	0.1249	0.0258	0.0041	3.1×10^{-4}	1.7×10^{-5}	6.4×10^{-7}	0
1.4×10^{8}	0.8325	0.1289	0.0305	0.0069	1.03×10^{-3}	8.2×10^{-5}	6.0×10^{-6}	2.8×10^{-7}
10^{12}	0.8178	0.1326	0.0356	0.0106	2.8×10^{-3}	5.4×10^{-4}	6.7×10^{-5}	9.4×10^{-6}
10^{14}	0.8099	0.1347	0.0380	0.0122	3.9×10^{-3}	1.1×10^{-3}	2.1×10^{-4}	4.3×10^{-5}
10^{16}	0.8045	0.1357	0.0393	0.0135	4.8×10^{-3}	1.6×10^{-3}	4.5×10^{-4}	1.1×10^{-4}
10^{18}	0.7998	0.1364	0.0407	0.0147	5.5×10^{-3}	2.1×10^{-3}	6.9×10^{-4}	2.5×10^{-4}
10^{19}	0.7977	0.1368	0.0412	0.0150	5.9×10^{-3}	2.3×10^{-3}	8.2×10^{-4}	3.3×10^{-4}
10^{20}	0.7957	0.1371	0.0417	0.0155	6.2×10^{-3}	2.5×10^{-3}	9.3×10^{-4}	4.3×10^{-4}
2×10^{20}	0.7952	0.1372	0.0418	0.0156	6.2×10^{-3}	2.5×10^{-3}	9.7×10^{-4}	4.6×10^{-4}

Table 2.13.2. Fraction of Gram blocks of given length k that have exactly k zeros and contain a Gram interval with 3 zeros.

N	$k=3$	$k=4$	$k=5$	$k=6$
1	0.0511	0.0799	0.1737	0.5448
10^{12}	0.0449	0.0541	0.0776	0.1032
10^{20}	0.0356	0.0413	0.0447	0.0392

Table 2.13.3. Fractions of Gram intervals that contain m zeros, and the GUE prediction.

N	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
1	0.13197	0.73772	0.12864	0.00167	10^{-8}
1.4×10^9	0.13965	0.72254	0.13598	0.00183	3×10^{-8}
10^{12}	0.14787	0.70625	0.14388	0.00200	–
10^{20}	0.15748	0.68709	0.15339	0.00204	–
GUE	0.17022	0.66143	0.16649	0.00186	4×10^{-7}

Table 2.13.4. Averages of $Z(g_n)$ and related functions.

average of	$N=10^{12}$	$N=10^{20}$
$Z(g_n)$	1.12×10^{-2}	-8.801×10^{-4}
$\lvert Z(g_n)\rvert$	2.6213	2.952707053204
$(-1)^n Z(g_n)$	2.0000	2.0000
$Z(g_n)^2$	27.65	45.47
$(-1)^n Z(g_n)^2$	0.1415	−0.1945
$Z(g_n)^3$	5.539	104.98
$\lvert Z(g_n)^3\rvert$	749.8	2240.4
$(-1)^n Z(g_n)^3$	692.7	1919.1
$Z(g_n)^4$	37645	238921
$(-1)^n Z(g_n)^4$	110.3	31305
$Z(g_n)^6$	2.821×10^8	1.062×10^{10}
$(-1)^6 Z(g_n)^6$	1.175×10^6	5.803×10^9
$Z(g_n)Z(g_{n+1})$	−3.1387	−3.1606
$\lvert Z(g_n)Z(g_{n+1})\rvert$	13.028	22.122
$(-1)^n Z(g_n)Z(g_{n+1})$	-7.73×10^{-3}	0.282
$Z(g_n)^2Z(g_{n+1})^2$	6068	46070

Table 2.14.1. Number of exceptions to Rosser's rule.

N	no. exceptions	exceptions per million zeros
10^{12}	38	23.9
10^{14}	87	51.6
10^{16}	1539	93.4
10^{18}	2453	147.1
10^{19}	2780	166.0
10^{20}	34570	196.9
2×10^{20}	21061	207.9

Table 2.14.2. Relative frequencies of the most frequent types of exceptions to Rosser's rule.

type	first 1.5×10^9 zeros	$N = 10^{12}, \ldots, 10^{20}, 2 \times 10^{20}$
$2L22$	0.0363	0.1040
$2R22$	0.0373	0.1022
$2L3$	0.4501	0.0973
$2R3$	0.4386	0.0965
$3R22$	–	0.0553
$3L22$	–	0.0553
$3R3$	0.0101	0.0529
$3L3$	0.0151	0.0517
$2L212$	–	0.0502
$2R212$	–	0.0497
$3L212$	–	0.0260
$3R212$	–	0.0248
$4R3$	–	0.0217
$4L3$	–	0.0214
$4L22$	–	0.0207
$4R22$	–	0.0201
$2R2112$	–	0.0197
$2L2112$	–	0.0187
$4L212$	–	0.0081
$3L2112$	–	0.0079
total	0.9892	0.8961

Table 3.2.1. Special points for the zeta function.

set	index of first zero	approx. index of first zero	no. zeros	special point
A	1789820229889768	1.8×10^{15}	213298	366350755915100.830671
B	3225901860089967	3.2×10^{15}	202337	648244850785931.253497
C	4817290207847018	4.8×10^{15}	224580	956149582979864.127715
D	5097943069948350	5.1×10^{15}	204441	1010102804832220.857487
E	6901069159073074	6.9×10^{15}	206276	1354828108521396.144683
F	18950008168234690	1.9×10^{16}	220040	3609764047662162.288453
G	22460777057881112	2.2×10^{16}	221960	4257232978148261.305478
H	42024941452698132	4.2×10^{16}	230978	7821904288693735.919567
I	51214985107007070	5.1×10^{16}	238512	9478467782100661.935759
J	71764726511399980	7.2×10^{16}	221752	13154657441819662.863688
K	76038726777613110	7.6×10^{16}	242968	13915273262098117.070642
L	76935378855702384	7.7×10^{16}	238556	14074693071712087.957658
M	153808369585296620	1.5×10^{17}	228170	27596944669957270.886813
N	233803646149078564	2.3×10^{17}	242576	41467826318647943.357194
O	253172315703241351	2.5×10^{17}	234879	44805187485720884.423354
P	473670769727688896	4.7×10^{17}	254092	82413269794748757.568756
Q	1250710180558723404	1.3×10^{18}	246054	212059301707021086.999247
R	4710265558902545324	4.7×10^{18}	254632	771729629469964785.437895
S	4795416924536726612	4.8×10^{18}	250812	785323253967853754.707393
T	17623088585596705508	1.8×10^{19}	262932	2793650241983592679.318477
U	32220179491036385680	3.2×10^{19}	263299	5032868769288289111.005891
V	35200636070992171652	3.5×10^{19}	265396	5486648117377526447.759269

Table 3.2.2. Zeta function at special points.

set	$Z(t)$	large $S(t)$	$\max \delta_n$	zero pattern
A	-396.2	-2.4235	4.5200	22000022
B	459.7	-2.3202	3.7948	301000122
C	-663.5	-2.6410	5.1454	220000212
D	598.6	2.7575	4.9347	21120000212
E	571.0	2.3145	3.8490	22010002112
F	-523.7	-2.4394	3.9353	301000212
G	843.9	2.7022	5.0612	2210000212
H	-555.7	2.3022	4.3147	2120000212
I	581.6	-2.1748	4.2731	2120000212
J	-720.0	2.3142	4.4956	2120000212
K	-767.4	2.6654	4.8923	2120000122
L	-780.0	2.6238	4.8025	2120000212
M	-831.3	-2.4475	4.4574	220000122
N	724.0	-2.7654	4.5515	2111000122
O	-874.6	-2.6160	4.7116	2120000122
P	-918.8	2.2410	4.4529	3110000212
Q	-971.3	-2.6178	4.6669	221000012112
R	754.7	2.1360	3.8989	22100010212
S	-1065.2	-2.5178	4.4694	2120000122
T	1036.7	2.8747	4.8433	3110001022
U	1580.6	-2.4862	4.5683	220200001212
V	-1329.5	2.8314	4.3214	221100002112

Table 4.3.1. Profile of rational function evaluation algorithm: computations for $N = 10^{20}$, $R = 2^{17}$.

step	time
evaluate Taylor series coefficients	49.1%
Taylor series expansions, $q \geq 4$	6.5%
$q = 3$ terms	12.2%
$q = 2$ terms	4.9%
$q = 1$ terms	7.4%
$q = 0$ terms	3.8%
evaluate $dp \log(k)$	9.6%
compute a_k, b_k, etc.	6.5%

Table 4.5.1. Running times (in minutes) of the main rational function evaluation program.

set	R	k_0	k_1	approx. T	δ	time
A	2^{17}	500	$7,635,871$	3.7×10^{14}	0.323	24
d	2^{23}	100	$17,577,894$	1.9×10^{15}	0.37	438
H	2^{17}	500	$35,283,065$	7.8×10^{15}	0.319	86
P	2^{17}	500	$114,527,198$	8.2×10^{16}	0.3287	261
g	2^{17}	100	$164,755,715$	1.7×10^{17}	0.2	380
f	2^{23}	200	$164,755,715$	1.7×10^{17}	0.33	1115
i	2^{23}	450	$505,829,004$	1.6×10^{18}	0.313	2133
U	2^{17}	500	$894,989,353$	5.0×10^{18}	0.3067	1975
k	2^{23}	450	$1,555,488,184$	1.5×10^{19}	0.28961	5250
n	2^{24}	450	$1,555,488,184$	1.5×10^{19}	0.2903	6099
p	2^{25}	450	$1,555,488,184$	1.5×10^{19}	0.2901	7537
r	2^{24}	450	$2,181,996,752$	3.0×10^{19}	0.29025	7998

Table 4.6.1. Large sets of zeros, showing duplication of computed values.

set	index of first zero in set	number of zeros
a	$10^{12}+1$	$101,053$
b	$10^{12}-6,032$	$1,592,196$
c	$10^{16}-4,930$	$1,584,442$
d	$10^{16}-5,946$	$16,480,973$
e	$10^{18}-8,394$	$1,419,501$
f	$10^{18}-8,839$	$16,671,047$
g	$10^{18}+12,333,574$	$157,608$
h	$10^{18}+12,345,608$	$140,684$
i	$10^{19}-13,607$	$16,749,725$
j	$10^{19}-45,597$	$135,161$
k	$10^{20}-30,769,710$	$16,366,702$
l	$10^{20}-15,409,240$	$16,341,831$
m	$10^{20}-48,778$	$16,388,741$
n	$10^{20}+15,311,688$	$32,811,834$
o	$10^{20}-48,867$	$132,188$
p	$10^{20}+47,110,546$	$65,578,910$
q	$10^{20}+111,678,401$	$33,139,615$
r	$2\times10^{20}-633,984$	$33,330,777$
s	$2\times10^{20}+31,673,368$	$33,199,868$
t	$2\times10^{20}+63,843,862$	$36,827,479$

Table 4.6.2. Comparison of values for zeros obtained in different computations.

sets of zeros	max. difference	rms difference
a vs. b	2.5×10^{-9}	3.7×10^{-11}
c vs. d	5.7×10^{-8}	3.4×10^{-10}
e vs. f	4.2×10^{-8}	3.3×10^{-10}
f vs. g	2.2×10^{-8}	2.3×10^{-10}
g vs. h	3.5×10^{-8}	1.6×10^{-10}
i vs. j	2.6×10^{-8}	7.5×10^{-10}
k vs. l	7.3×10^{-7}	3.8×10^{-9}
l vs. m	7.6×10^{-7}	5.9×10^{-9}
m vs. n	5.8×10^{-7}	3.9×10^{-9}
m vs. o	2.7×10^{-7}	4.2×10^{-9}
n vs. p	5.7×10^{-7}	4.4×10^{-9}
p vs. q	5.2×10^{-7}	4.1×10^{-9}
r vs. s	1.1×10^{-6}	9.5×10^{-9}
s vs. t	4.7×10^{-7}	7.5×10^{-9}

Figure Captions

Figure 2.1.1. $Z(t)$ near zero number 10^{20}. The horizontal axis extends from Gram point number $10^{20} - 8$ to Gram point number $10^{20} + 4$.

Figure 2.1.2. $S(t)$ near zero number 10^{20}. The range of t is the same as in Fig. 2.1.1, and the jumps by 1 occur at zeros of the zeta function numbered $10^{20} - 6$ to $10^{20} + 5$.

Figure 2.1.3. $Z(t)$ near zero number 10^{20}. The horizontal axis extends from Gram point number $10^{20} - 50$ to Gram point number $10^{20} + 50$.

Figure 2.4.1. Pair correlation of zeros of the zeta function. Solid line: GUE prediction. Scatterplot: empirical data based on 8×10^6 zeros near zero number 10^{20}.

Figure 2.4.2. Pair correlation of zeros of the zeta function. Solid line: GUE prediction. Scatterplot: empirical data based on 10^6 zeros near zero number 10^{12}.

Figure 2.4.3. Pair correlation of zeros of the zeta function. Solid line: GUE prediction. Scatterplot: empirical data based on 8×10^6 zeros near zero number 10^{20}. Scatterplot smoothed.

Figure 2.4.4. Probability density of the normalized spacings δ_n. Solid line: GUE prediction. Scatterplot: empirical data based on 1,592,196 zeros near zero number 10^{12}.

Figure 2.4.5. Probability density of the normalized spacings δ_n. Solid line: GUE prediction. Scatterplot: empirical data based on 78,893,234 zeros near zero number 10^{20}.

Figure 2.4.6. Probability density of the normalized spacings $\delta_n + \delta_{n+1}$. Solid line: GUE prediction. Scatterplot: empirical data based on 1,592,196 zeros near zero number 10^{12}.

Figure 2.4.7. Probability density of the normalized spacings $\delta_n + \delta_{n+1}$. Solid line: GUE prediction. Scatterplot: empirical data based on 78,893,234 zeros near zero number 10^{20}.

Figure 2.5.1. Comparison of the scaled distribution of $S(t)$ for $N = 10^{20}$ to the asymptotic normal distribution.

Figure 2.6.1. Initial segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on 10^6 consecutive values of n, starting with $n = 10^{20} - 42,778$. Straight line $y = x$ drawn to facilitate comparisons.

Figure 2.6.2. Initial segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on 10^6 consecutive values of n, starting with $n = 10^{20} + 15,316,087$.

Figure 2.6.3. Initial segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on 112,314,003 values of n from $N = 10^18$, 10^{19}, and 10^{20} data sets.

Figure 2.6.4. Initial segment of the quantile-quantile plot of the normalized spacings $\delta_n + \delta_n + 1$ against the GUE prediction. Data based on 10^6 consecutive values of n, starting with $n = 10^{12} - 6,032$.

Figure 2.6.5. Initial segment of the quantile-quantile plot of the normalized spacings $\delta_n + \delta_n + 1$ against the GUE prediction. Data based on 10^6 consecutive values of n, starting with $n = 10^{20} - 42,778$.

Figure 2.6.6. Find segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on 10^6 consecutive values of n, starting with $n = 10^{12} - 6,032$.

Figure 2.6.7. Final segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on 10^6 consecutive values of n, starting with $n = 10^{20} - 42,778$.

Figure 2.7.1. Graph of $2\log|\sum \exp(i\gamma_n y)|$, where n runs over $10^{20} + 1 \le n \le 10^{20} + 40,000$, and values < 0 and > 16 are deleted.

Figure 2.7.2. Graph of $2\log|\sum \exp(i\gamma_n y)|$, where n runs over $10^{20} + 1 \le n \le 10^{20} + 40,000$, and values < 0 are deleted.

Figure 2.7.3. Graph of $2\log|\sum \exp(i\gamma_n y)|$, where n now runs over $10^{20} + 1 \le n \le 10^{20} + 400,000$, and values < 0 are deleted.

Figure 2.7.4. Variance of the number of zeros in an interval of length L for the GUE (dashed line), for 5×10^5 zeros near zero number 10^{20} (scatterplot), and Berry's prediction (solid line).

Figure 2.7.5. Variance of the number of zeros in an interval of length L based on 5×10^5 zeros near zero number 10^{20}.

Figure 2.7.6. Variance of the number of zeros in an interval of length L based on 5×10^5 zeros near zero number 10^{20}.

Figure 2.8.1. Neighborhood of an example of Lehmer's phenomenon. Graph of $Z(t)$ between Gram points $n - 6$ and $n + 6$, where $n = 10^{18} + 12,376,778$. The point between Gram points $n + 1$ and $n + 2$ where $Z(t)$ is seemingly tangent to the zero line represents 2 zeros with $\delta_{n+2} = 0.0011$ and the minimum of $Z(t)$ between those zeros equal to -5×10^{-7}. For a smaller scale picture of this phenomenon, see Fig. 4.6.1. The other point of near tangency, near Gram point $n - 3$, has minimum of $Z(t)$ of -0.0094.

Figure 2.11.1. Comparison of the distribution of $\log|\zeta(1/2 + it)|$ over two ranges of 10^6 zeros each near zeros number 10^{12} and 10^{20} to that of the normal distribution.

Figure 2.11.2. For each k, plots the logarithm of the fraction of time that $|\zeta(1/2 + it)| \in [k - 1, k)$. Data obtained from 3 intervals covering 2.8×10^6 zeros near zero number 10^{20}.

Figure 2.12.1. Scaled distribution of 10^6 values of $log|\zeta'(1/2 + i\gamma)|$ for $N = 10^{20}$ vs. the conjectured standard normal distribution.

Figure 2.13.1. Distribution modulo 1 of γ_n on Gram point scale, for two sets of 10^6 zeros each. Curves derived by smoothing a histogram.

Figure 3.2.1. $Z(t)$ near the point where the largest known value of $S(t)$ occurs. The horizontal axis extends from Gram point number $n = 17,623,088,585,596,834,905$ to Gram point number $n + 30$.

Figure 3.2.2. $Z(t)$ near the point where the largest known value of $S(t)$ occurs. The horizontal axis extends from Gram point number $n + 9$ to Gram point number $n + 21$, where $n = 17,623,088,585,596,834,905$. The high peak of $Z(t)$ has been cut off. This is a smaller scale view of the central part of Fig. 3.2.1.

Figure 3.2.3. $S(t)$ near the point where its largest known value occurs. The range of t is the same as in Fig. 3.2.2.

Figure 4.6.1. Small neighborhood of an example of Lehmer's phenomenon. Graph of $Z(t)$ on a segment of the interval between Gram points n and $n + 1$ (corresponding to 0 and 1 on the scale of the figure), where $n = 10^{18} + 12,376,799$. Enlargement of a section of Fig. 2.8.1.

Figure 4.6.2. Small scale view of Lehmer's phenomenon. Enlargement of a section of Fig. 4.6.1. The three curves represent three different computations of $Z(t)$ on this segment.