Currently there are no reasons to doubt the Riemann Hypothesis

The zeta function beyond the realm of computation

David W. Farmer

Abstract

We examine published arguments which suggest that the Riemann Hypothesis may not be true. In each case we provide evidence to explain why the claimed argument does not provide a good reason to doubt the Riemann Hypothesis. The evidence we cite involves a mixture of theorems in analytic number theory, theorems in random matrix theory, and illustrative examples involving the characteristic polynomials of random unitary matrices. Similar evidence is provided for four mistaken notions which appear repeatedly in the literature concerning computations of the zeta-function. A fundamental question which underlies some of the arguments is: what does the graph of the Riemann zeta-function look like in a neighborhood of its largest values? We explore that question in detail and provide a survey of results on the relationship between L-functions and the characteristic polynomials of random matrices. We highlight the key role played by the emergent phenomenon of carrier waves, which arise from fluctuations in the density of zeros. The main point of this paper is that it is possible to understand some aspects of the zeta function at large heights, but the computation evidence is misleading. [1]

1 Introduction

Should one believe the Riemann Hypothesis (RH)? Since it is a conjecture with no proposed roadmap to prove it, one point of view is that it should neither be believed nor disbelieved. Yet many mathematicians have an opinion, presumably backed up by logical reasoning.

Here we consider all published arguments for doubting RH. A paper of Ivić [65, 66] lists 4 reasons, and a paper of Blanc [14] provides a 5th reason. Three of those reasons involve speculation about the distribution of zeros and their relationship to the value distribution of the ζ-function. The fundamental question there is: what does the ζ-function look like in a neighborhood of its largest values? The majority of this paper is a survey of prior results, and speculations and heuristics based on those results, which lead to an answer to that question.

Our primary goal is to provide intuition and to be persuasive. The arguments against RH generally take the form "It would be surprising if **X**." So, the burden we bear in refuting that argument is to give good reasons why **X** is not surprising. To crystallize our main points we present 33 *Principles* which we hope are also useful for future reference.

Is the purpose of this paper to persuade that RH is true? Certainly not. But perhaps those who continue to doubt RH will realize that their belief is

---

[1] MSC2020: 11M26, 11M50

not based on good evidence. As for those who believe RH: perhaps someone will write a companion paper: *Currently there are no good reasons to believe the Riemann Hypothesis.* Note the subtle difference from the opposite of the title to this paper. Also useful would be a paper explaining why every currently known equivalence to RH is unlikely to be helpful for proving RH (as in [41]).

The ζ-function is the simplest example of an *L-function.* All L-functions have properties similar to the ζ-function, and all L-functions have an analogue of the Riemann Hypothesis. As much as possible we try to discuss the ζ-function in isolation, but in a few places it is necessary to expand our perspective. We attempt to keep this paper self-contained, providing definitions and background as needed.

The themes. Many of the surprising properties of the ζ-function arise because there are different facets to the same object. A classic example is the function usually denoted $S(t)$: it is the error term in the counting function of the zeros of the ζ-function, and it also is the imaginary part of the logarithm of $\zeta(\frac{1}{2} + it)$. That equivalence is basic complex analysis. But when pondering a specific question about $S(t)$, sometimes one perspective gives good intuition, and sometimes another. Combining perspectives can lead to surprising relations, such as Principle 8.7: in regions where $|\zeta(\frac{1}{2} + it)|$ is large, $S(t)$ tends to be decreasing.

A second theme is multiple levels of randomness. We will make extensive use of the fact that the ζ-function and characteristic polynomials of random unitary matrices have a similar sort of randomness (and also some differences, which we will describe). In the random matrix world, it is a theorem that everything just follows from the quadratic repulsion between the eigenvalues. Everything. But (thank you to an anonymous referee for helping me see it this way), that perspective is reductive and it is beneficial to have other points of view, particularly on larger scales where the individual zeros are not visible. In the ζ-function world this means that sometimes the primes will enter the discussion, and other times they will be ignored.

The sections. In Section 2 we introduce the main theme by pulling together various ideas in the paper to answer the question: What does the zeta function look like beyond the realm of computation? The remainder of the paper is primarily devoted to providing intuition which will dispel misconceptions about the answer to that question — misconceptions largely due to poor extrapolation from available computations. In Section 3 we provide basic definitions and background. In Section 4 we describe four Mistaken Notions which appear repeatedly in discussions of computations of the ζ-function, some of which play an important role in the claimed reasons to doubt RH. In Section 5 we describe the connection between the distribution of zeros and the size of the ζ-function, introducing *carrier waves* as a way to separate local from long-range behavior. In Section 6 we briefly describe the connection between the ζ-function and unitary polynomials, and in Section 7 we provide an historical account of the connections to Random Matrix Theory. This leads to Section 8, where we use large unitary matrices to illustrate phenomena which occur far outside the range in which we can compute the ζ-function. By the end of Section 8 we have a good understanding of the "typical" large values of the ζ-function, and the relationship between the carrier wave, the density wave, and $S(t)$, but it is not until Section 9 that we address the most extreme values. The primes are only briefly mentioned up to this point, a shortcoming we address in Section 10. The zeros have dominated most of our discussion, but in Section 11 we discuss randomness of the ζ-function or characteristic polynomials without reference to

2

zeros or eigenvalues. After all that preparation, in Section 12 we use information from the prior sections to refute the three arguments against RH based on the distribution of zeros and values of the ζ-function, and for completeness in Section 13 we cite recent results to refute the other two arguments against RH. Finally, in Section 14 we use the Principles to explain why the Mistaken Notions of Section 4 are, in fact, mistaken.

2 The ζ-function as a random object

We present some of the main points in this paper in a thought experiment to give intuition about the behavior of the ζ-function far beyond the range where it can be computed. It is hoped that the reader will view the claims in this overview with some skepticism, leading to closer scrutiny of the ideas and perspective in the remainder of the paper.

2.1 Snapshots of the ζ-function

As described in Section 3, we will consider the function $Z(t)$, which contains the same information as the Riemann ζ-function, but it is a real-valued function of a real variable, so we can graph it. What might the graph of $Z(t)$ look like for $t \approx 100^{100^{100}}$? Current methods fail well before 10^{40}, so such a computation is not remotely plausible. But suppose in some distant future a computer algebra package allowed one to graph $Z(t)$ for such large t, on an interval of, for example, width $40/(2\pi \log(t/2\pi))$ — an interval which should contain around 40 zeros. By Principle 7.3 and Principle 7.4, combined with Principle 5.1 or Principle 10.2, it is plausible that the graph might look like this:
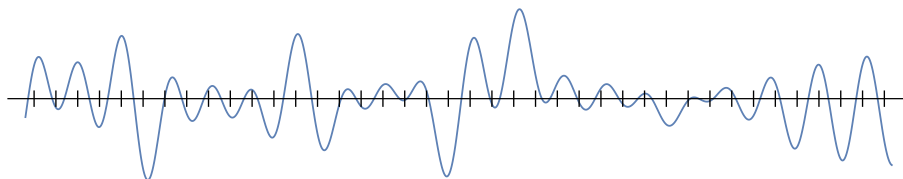

Figure 2.1 What $Z(t)$ looks like, on some interval which is expected to contain around 40 zeros, near $t \approx 100^{100^{100}}$. The tick marks are separated by the average gap between zeros at that height.

Neither axis in Figure 2.1 has a scale, although we are told enough information to determine the width covered by the horizontal axis. The scale on the vertical axis is the mystery.

By Principle 5.1, the Z-function is close to being a polynomial. In particular, locally the behavior of the Z-function is determined by its zeros, up to a local constant factor. So, if the relative spacing of the zeros in Figure 2.1 is plausible for t of that size, then the given graph is accurate: as is typical for computer algebra systems, the vertical scale is automatically adjusted to show the features of the graph.

By Principle 7.3, the distribution of zeros in Figure 2.1 is indeed plausible, and furthermore we expect to find many intervals of that width at that height, on which the graph of the Z-function looks similar to Figure 2.1. The only thing we don't know is: what is the vertical scale of that graph?

By Selberg's Theorem (3.15), $\log Z(t)$ is normally distributed with mean zero and variance $\sqrt{\frac{1}{2} \log \log t}$. So, at height $t \approx 100^{100^{100}}$, half the time $\log Z(t)$ is likely to be of size ≈ 15, and the other half of the time it is likely to be ≈ -15. Exponentiating, we find that the vertical scale in Figure 2.1, or in any other interval at that height where the graph looks like Figure 2.1, is likely to be around 10^6, or it is likely to be around 10^{-6}, and both possibilities are equally probable. In particular, the local distribution of zeros is not the main factor in the size of the ζ-function.

This fundamental fact, that the size of the ζ-function is not strongly dependent on the local distribution of zeros, is hard to believe when your intuition only comes from computations in the modest range currently accessible by computers.

Snapshots do not show the big picture. By **snapshot** we refer to an image showing the graph of a function, having an aspect ratio between $1/10$ and 2, over an interval where the function has a moderate number of interesting features, with the vertical scale adjusted so that those features are visible. Figure 2.1 is an example, as are the numerous other graphs in this paper. A snapshot of the ζ-function typically covers an interval containing between 10 and 100 zeros. As mentioned above, a snapshot of the Z-function depends only on the location of the nearby zeros. More precisely, it depends only on the relative sizes of the gaps between zeros. Near the middle of the snapshot all that matters is the zero gaps within the snapshot; near the edges, the gaps immediately outside the snapshot are relevant.

Any specific (finite) sequence of normalized zero gaps will never happen (except once, if it is specifically constructed in that way), but precision is not relevant, because a snapshot is a continuous function of the zero gaps. Given an open neighborhood of any specific (finite) gap sequence, such an approximate gap sequence will occur infinitely many times, and in fact will occur a positive proportion of the time among all normalized gap sequences of that length — the precise proportion depending on the given sequence and the size of the neighborhood. See Principle 7.3 or Principle 7.4. For example, Figure 4.1 will occur a positive proportion of the time (to within the resolution of the human eye) as a snapshot of the Z-function.

To summarize:

Principle 2.2 Snapshots of the ζ-function are well understood. *Any given finite sequence of normalized gaps between zeros will (approximately) occur a positive proportion of the time, as predicted by the Random Matrix Model for zeros of the zeta function. The appearance of the graph of the Z-function over that region, ignoring the vertical scale, is determined by the relative spacing of those zeros.*

In particular, one does not expect to see anything surprising when looking at a new snapshot of the Z-function.

What counts as "surprising" changes over time. In the early computations of zeros of the ζ-function, it was surprising that occasionally two zeros would be very close together (known as **Lehmer pairs**). Today such pairs of close zeros are expected, and those occur at the frequency predicted by Principle 7.3.

3 Background on the ζ- function

The Riemann zeta-function (which we will call the ζ-function) is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \qquad (3.1)$$

for $\sigma > 1$, where $s = \sigma + it$ is a complex variable. The ζ-function has a meromorphic continuation to the complex plane, with a simple pole at $s = 1$ with residue 1.

The ζ-function has a symmetry, known as the **functional equation**, which can be expressed in several ways. With

$$X(s) := \pi^{s-\frac{1}{2}} \frac{\Gamma(\frac{1}{2} - \frac{1}{2}s)}{\Gamma(\frac{1}{2}s)} \qquad (3.2)$$

where Γ is the Euler Gamma-function, we have

$$\zeta(s) = X(s)\zeta(1-s), \qquad (3.3)$$

or equivalently

$$\xi(s) := \tfrac{1}{2}s(s-1)\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s)\zeta(s) \qquad (3.4)$$
$$= \xi(1-s), \qquad (3.5)$$

or equivalently

$$Z(t) := X(\tfrac{1}{2} + it)^{-\frac{1}{2}}\zeta(\tfrac{1}{2} + it) \quad \text{ is real if } \quad t \in \mathbb{R}. \qquad (3.6)$$

The factor $\frac{1}{2}s(s-1)$ in (3.4) is irrelevant to the invariance under $s \leftrightarrow 1-s$, but it is traditionally included so that $\xi(s)$ is an entire function. In (3.6) the square root is chosen so that $Z(0) = \zeta(\frac{1}{2}) \approx -1.46$ and $Z(t)$ is analytic for $|\Im(t)| < \frac{1}{2}$. It is common to refer to $|t|$, the magnitude of the imaginary part of $s = \sigma + it$, as the **height** when referring to the behavior of $\zeta(s)$ or $Z(t)$ in a particular region.

The **Hardy Z-function** (3.6) is useful because it can be graphed, and it tells us essentially everything we might want to know about the ζ-function because $|Z(t)| = |\zeta(\frac{1}{2} + it)|$ if $t \in \mathbb{R}$. Figure 3.1 shows $Z(t)$ and $\log|Z(t)|$ for $5429.29 < t < 5466.44$, along with the function $S(t)$ which we will introduce shortly.
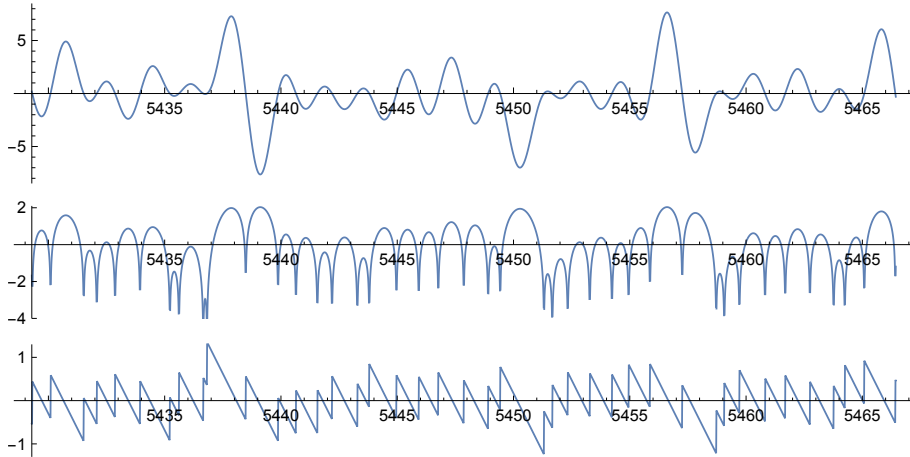

Figure 3.1 $Z(t)$, $\log|Z(t)|$, and $S(t)$ near the 5000th zero $\gamma_{5000} \approx 5447.86$.

The critical facts (pun intended) about $Z(t)$ are that it is smooth, and if t is real then $Z(t)$ is real and $|Z(t)| = |\zeta(\frac{1}{2} + it)|$. Those facts do not uniquely determine $Z(t)$: there is still a global choice of ± 1. One might ask whether the choice matters, and if it does, is the standard choice for the square-root in (3.6) the correct option?

Principle 3.2 $Z(t)$ *is statistically indistinguishable from* $-Z(t)$.

In other words, any measure of distance between probability distributions cannot distinguish between the distributions of $Z(t)$ and $-Z(t)$ for $t \in [T, 2T]$, as $T \to \infty$.

Thus, the global choice of sign for $Z(t)$ does not matter. Indeed, $\xi(\frac{1}{2} + it)$ is also real for $t \in \mathbb{R}$, and has the same critical zeros as $Z(t)$, but it has the opposite sign of $Z(t)$. So, the standard normalizations for those functions disagree with each other. A consequence of Principle 3.2 is that if k is a non-negative integer then

$$\int_T^{2T} Z(t)^{2k+1} \, dt = o(T), \tag{3.7}$$

and as $T \to \infty$ that integral changes sign infinitely many times. Equation (3.7) is a theorem for $k = 0$, use the Riemann-Siegel formula (4.1) and integrate term-by-term, and a conjecture for larger k.

Principle 3.2 is a stronger statement than asserting that to leading order the value distribution of $Z(t)$ is symmetric. In contrast, the average value of $\zeta(\frac{1}{2} + it)$ is 1 in a very strong sense: if k is a non-negative integer then

$$\int_T^{2T} \zeta(\tfrac{1}{2} + it)^k \, dt \sim T, \tag{3.8}$$

which can be shown by moving the integral to the region of absolute convergence of the Dirichlet series and then integrating term-by-term. Such small biases (meaning: effects of size 1, when the main term is growing very slowly) are the underlying cause of many misconceptions arising from numerical computations, see (4.4) and Principle 4.2.

3.1 We care about zeros because we care about primes

Riemann's great insight was that the zeros of the ζ-function encode information about the primes. Based at least partially on numerical computation, he formulated the conjecture which is now known as the **Riemann Hypothesis**:

Conjecture 3.3 The Riemann Hypothesis (RH). *The zeros of* $\zeta(s)$ *with* $0 < \sigma < 1$ *lie on the line* $\sigma = \frac{1}{2}$*; equivalently, all zeros of* $\xi(s)$ *lie on the line* $\sigma = \frac{1}{2}$*; equivalently, all zeros of* $Z(t)$ *with* $|\Im(t)| < \frac{1}{2}$ *are real.*

Riemann's original formulation involved yet another function: all the zeros of $\Xi(t) := \xi(\frac{1}{2} + it)$ are real. RH has been rigorously verified for $t < 3 \times 10^{12}$, involving more than 12.3×10^{12} zeros [82].

The zeros of $\zeta(s)$ with $0 < \sigma < 1$ are traditionally denoted by $\rho = \beta + i\gamma$. By the functional equation and the fact that $\zeta(s)$ is real if s is real, if $\rho = \beta + i\gamma$ is a zero of $\zeta(s)$ then so is $\rho = 1 - \beta + i\gamma$. So either $\beta = \frac{1}{2}$, as predicted by RH, or there is a pair of zeros of $\zeta(s)$ (or $\xi(s)$) located symmetrically around the **critical line** $\sigma = \frac{1}{2}$. Equivalently, a failure of RH corresponds to a pair of complex conjugate zeros of $Z(t)$. Assuming RH, the zeros of $\zeta(s)$ with positive imaginary part are denoted $\frac{1}{2} + i\gamma_n$ with $0 < \gamma_1 < \gamma_2 < \cdots$. That notation would break down if there were repeated zeros, but no plausible reason has been

given to expect a multiple zero. Various discussions in this paper may implicitly assume all zeros are simple, an assumption which generally is irrelevant to the points being made.

There are 41 zeros of the ζ-function visible in Figure 3.1. RH implies that (for $|t| > 3$) all local maxima of $Z(t)$ (or $\xi(\frac{1}{2} + it)$) are positive and all local minima are negative, a condition which is visible in the first two graphs in Figure 3.1. The converse is not necessarily true — a failure of RH need not cause a negative maximum or a positive minimum, although that is mistakenly asserted in the official statement of the Riemann Hypothesis Millennium Problem [19]. The issue is the distance of a hypothetical non-critical zero from the critical line, see [42]. So, one cannot trivially deduce from the graph of $Z(t)$ or $\log(|Z(t)|)$ in Figure 3.1 that RH holds for $5430 < t < 5448$.

It is possible to verify RH on an interval by a computer calculation, based on two properties of the ζ-function. One property we have already seen: $Z(t)$ is real when t is real, so one can count critical zeros by looking for sign changes. The other is that the zeros of the ζ-function have a nice counting function with a small and computable error term. Let $N(T)$ be the number of zeros of $\zeta(s)$ with $0 < \gamma_n \leq T$. We have [101], assuming T is not the imaginary part of a zero of $\zeta(s)$,

$$N(T) = \frac{1}{2\pi} T \log T - \frac{\log 2\pi e}{2\pi} T + \tfrac{7}{8} + S(T) + O(T^{-1}) \qquad (3.9)$$

where

$$S(T) = \frac{1}{\pi} \arg \zeta(\tfrac{1}{2} + iT) = \frac{1}{\pi} \Im \log \zeta(\tfrac{1}{2} + iT). \qquad (3.10)$$

The argument in (3.10) is determined by continuous variation along the line $2 + it$ for $0 \leq t \leq T$, and then along the line $\sigma + iT$ for $2 \geq \sigma \geq \frac{1}{2}$. One can take either (3.9) or (3.10) as the definition, with the other as a theorem.

We will see that the function $S(t)$ grows very slowly. The third plot in Figure 3.1 illustrates the basic properties of $S(t)$: it has a jump discontinuity (of height 1) at a simple critical zero, and where it is continuous it is approximately linear with slope $-\frac{1}{2\pi} \log t$.

By (3.9) and (3.10) one can rigorously prove that RH holds on an interval: use sign changes to count real zeros of $Z(t)$, then compute the change in $S(t)$ to determine the change in $N(t)$ and thus find the total number of zeros. Check if those two quantities are equal. A sophisticated version of this idea is known as **Turing's method**; see [21] for a modern treatment. In Figure 3.1 a failure of RH would correspond to a jump discontinuity in $S(t)$ of height 2, at a point where $Z(t)$ does not have a zero. Such a jump does not occur in the graph of $S(t)$, so by also considering either $Z(t)$ or $\log |Z(t)|$, one can "see" a proof in Figure 3.1 that RH is true for $5430 < t < 5448$.

3.2 Unfolding the zeros

By (3.9) the zeros at larger height are on average closer together. Specifically, at height T the average gap between zeros is $2\pi / \log T$, and $\gamma_n \approx 2\pi n / \log n$. When discussing the statistics of the zeros, in particular the gaps between zeros, it is helpful to use the **normalized** or **unfolded** zeros $\tilde{\gamma}_n = \frac{1}{2\pi} \gamma_n \log \gamma_n$. Note that $\tilde{\gamma}_n \sim n$ and $\tilde{\gamma}_{n+1} - \tilde{\gamma}_n$ equals 1 on average.

When it is necessary to be more precise than the asymptotic behavior, we set $\tilde{\gamma}_n = \widetilde{N}(\gamma_n)$, where

$$\widetilde{N}(T) = \frac{1}{2\pi} T \log T - \frac{\log 2\pi e}{2\pi} T + \tfrac{7}{8}. \qquad (3.11)$$

We leave it as an exercise to determine the average value of $\tilde{\gamma}_n - n$ as $n \to \infty$. The answer is in Subsection 8.6. That answer also explains the usual definition of $S(t)$ in the case t is the imaginary part of a zero of the ζ-function.

3.3 The size of $Z(t)$ and $S(t)$

The function $S(t)$ is both the error term in the zero counting function $N(t)$ and the imaginary part of $\log \zeta(\frac{1}{2} + it)$. Viewed as an error term for a counting function, it is perhaps surprising that $S(t)$ grows very slowly: much more slowly that the error term in the prime number theorem, or the Dirichlet divisor problem, or other counting problems, for example. A consequence is that the zeros cannot stray too far from their expected location. We will see that this has profound implications for the behavior of the ζ-function.

Assuming RH, Littlewood [101] showed that $S(T) = O(\log T / \log \log T)$. It is conjectured [43] that

$$|S(T)| \leq (1 + o(1)) \frac{1}{\pi} \sqrt{\frac{1}{2} \log T \log \log T}, \qquad (3.12)$$

and that bound is sharp. The results for $Z(t)$ are analogous: on RH we have [101] $\log |Z(t)| = O(\log t / \log \log t)$, and the conjecture (with sharp constant) is

$$\log |Z(T)| \leq (1 + o(1)) \sqrt{\frac{1}{2} \log T \log \log T}. \qquad (3.13)$$

In the other direction [18], the current best result is that there exists $C > 0$ such that there exist arbitrarily large T with:

$$\log |Z(T)| > C \sqrt{\frac{\log T \log \log \log T}{\log \log T}}. \qquad (3.14)$$

The typical size of $Z(t)$ is much smaller. Selberg [94] proved that if t is chosen uniformly at random from $[T, 2T]$, then

$$\frac{\log \zeta(\frac{1}{2} + it)}{\sqrt{\frac{1}{2} \log \log T}} \to N(0, 1) \qquad \text{as} \qquad T \to \infty, \qquad (3.15)$$

where $N(0, 1)$ is the standard (complex) Gaussian, and (3.15) indicates convergence in distribution. In particular, $\log |Z(t)|$ and $S(t)$ each have a (real) Gaussian distribution, which are different only because the definition of $S(t)$ contains a factor of $1/\pi$, and those distributions are independent.

In numerical computations of the ζ-function the scale factor in Selberg's theorem is practically irrelevant: at $T_{BH} = 10^{33}$, which is approximately the largest height where the ζ-function has been calculated [17], Selberg's theorem (assuming it is valid at that low height) says that the typical size of $\log |Z(t)|$ is $\sqrt{\frac{1}{2} \log \log T_{BH}} \approx 1.5$. The conjectured extreme values are larger, $\sqrt{\frac{1}{2} \log T_{BH} \log \log T_{BH}} \approx 12.8$, so within the realm of current computation one might expect to see $|Z(t)|$ larger than $300,000$. Unfortunately, the extreme values are rare, and possibly (due to lower order terms) the predicted largest values do not occur until significantly greater heights. The largest value of $Z(t)$ found in [17] is approximately 16244 near 3.92×10^{31}, which is 3.7% less than the largest computed value [100]. The largest calculated value of $S(t)$ is 3.345 near $t = 7.7573 \times 10^{27}$.

Figure 3.4 shows $Z(t)$, $\log|Z(t)|$, and $S(t)$ in a neighborhood of the largest value of $Z(t)$ found in [17]. (Note that there is no good linear scale on which to plot $Z(t)$ in that region.) The author thanks Jonathan Bober and Ghaith Hiary for providing open access to their extensive data [59].
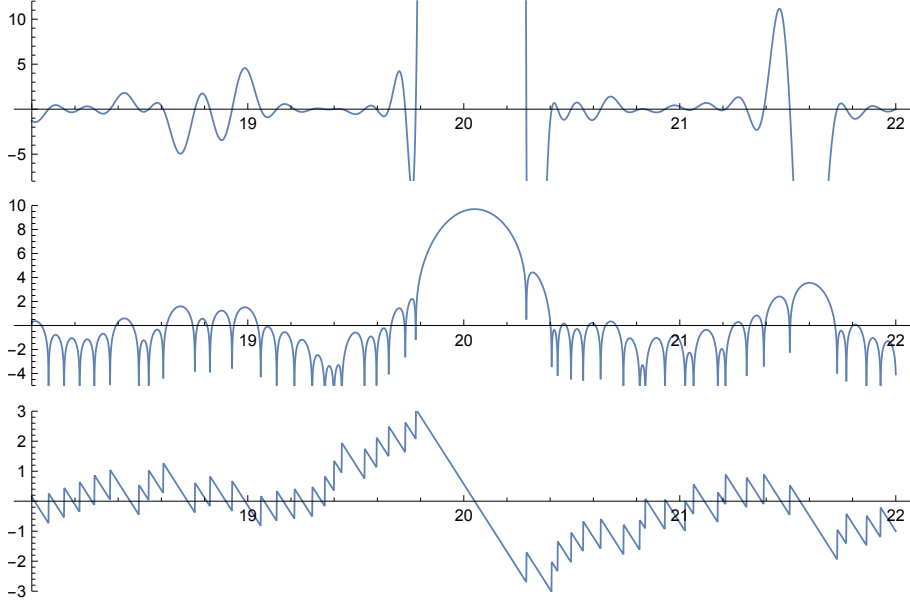

Figure 3.4 Plots of $Z(T_{big}+t)$, $\log|Z(T_{big}+t)|$, and $S(T_{big}+t)$ where $T_{big} = 3924676458989430915525116928 4084 \approx 3.9 \times 10^{31}$.

In Figure 3.4 we see that this particular large value of $Z(t)$ arises from a large gap between neighboring zeros. The large zero gap also contributes to the large value of $S(t)$: a zero gap of K times the local average spacing must be accompanied by a value $|S(t)| \geq K/2$. The large zero gap in Figure 3.4 is 5.93 times the local average. The relationship between the size of an isolated large zero gap and the local maximum of $Z(t)$ is subtle: see Principle 9.2. The relationship between a large gap and $S(t)$ is almost trivial, but we record it for later use. We write $S^+(t)$ and $S^-(t)$ for the right and left limiting values, respectively.

Principle 3.5 *If $\tilde{\gamma}_{j+1} - \tilde{\gamma}_j = K$, then $S^+(\gamma_{j+1}) - S^-(\gamma_j) = K$. So in particular either $|S^+(\gamma_{j+1})| \geq K/2$ or $|S^-(\gamma_j)| \geq K/2$. Thus, an upper bound on $|S(t)|$ implies a comparable upper bound on the size of the normalized zero gaps.*

Note that Principle 3.5 does not say that a large value of $S(t)$ must be accompanied by a large gap between zeros.

As will be explored in detail, Figure 3.4 does not illustrate the typical behavior for the large values of $Z(t)$ and $S(t)$. One way to see this is from Selberg's theorem [94] that $\log|Z(t)|$ and $S(t)$ are independently distributed. One of those being typically large should have no effect on the other. Figure 3.4 illustrates that a large zero gap causes large values of $Z(t)$ and $S(t)$ to occur in close proximity, therefore that cannot be the typical behavior near a large value. (This discussion does not address the question of independence in the tails of the distributions of $\log|Z(t)|$ and $S(t)$.)

In Section 4 we briefly explore the history of finding large values of $Z(t)$ and explain why that work has inadvertently led to a mistaken impression of what the graph of $Z(t)$ looks like in a neighborhood of its largest values. In Section 5 we introduce *carrier waves*, which are the actual cause of the largest values of $Z(t)$.

9

4 Misleading ideas about large values

Computation has been an important tool for studying the ζ-function ever since Riemann calculated the first few zeros by hand. Computers have enabled large-scale computations: large on the the human scale but minuscule on an absolute scale. In reference to whether existing computations should be seen as evidence for RH, Andrew Odlyzko [79] has sounded a cautionary note: the true nature of the ζ-function is unlikely to be revealed until we reach regions where $S(t)$ is routinely over 100. Since (by Selberg's theorem) $S(t)$ is typically of size $\sqrt{\frac{1}{2}\log\log t}$, such regions will be inaccessible for a long time.

Despite Odlyzko's warning, there are certain aspects of the ζ-function which appear in computations and have influenced the direction of research, but which do not accurately portray the true nature of the ζ-function. These misconceptions are partially based on the way the ζ-function is computed, which we describe next.

4.1 Computations of $Z(t)$

The earliest large-scale computations of the ζ-function calculated $Z(t)$ using the **Riemann-Siegel formula**:

$$Z(t) = 2 \sum_{n < \sqrt{t/2\pi}} n^{-\frac{1}{2}} \cos(\theta(t) - t\log n) \ + \ \text{remainder} \tag{4.1}$$

where

$$\theta(t) = \arg\left(\pi^{-it/2}\Gamma\left(\frac{1}{4} + i\frac{t}{2}\right)\right)$$

$$= \frac{t}{2}\log\left(\frac{t}{2\pi}\right) - \frac{t}{2} - \frac{\pi}{8} + O(t^{-1}). \tag{4.2}$$

The function $\theta(t)$ arises in an alternate expression for $Z(t)$:

$$Z(t) = e^{i\theta(t)}\zeta(\tfrac{1}{2} + it). \tag{4.3}$$

The $n = 1$ term in (4.1) is the largest, and if t is small then that term has a strong influence on the overall sum. The points where $\cos(\theta(t)) = \pm 1$ are known as **Gram points**, numbered so that $\theta(g_m) = m\pi$. Almost equivalently, Gram points are the locations where $\zeta(\tfrac{1}{2} + it)$ is real but nonzero. (The "almost" is because 0 is not considered to be a Gram point, and it has not been proven (but certainly it is true) that the ζ-function does not vanish at any Gram point.)

Gram [54] noted that the first several zeros of $Z(t)$ interlace the Gram points. In other words, when t is small the other terms are insufficient to flip the sign of the first term at a Gram point.

In the course verifying RH for the first 75 million zeros, Brent [23] found that $Z(t)$ was unusually large, more than 79.6, at the $70\,354\,406$th Gram point. Furthermore, at that point *the first 72 terms in the Riemann-Siegel formula (4.1) were positive.* Figure 4.1 shows a graph of $Z(t)$ near that Gram point.
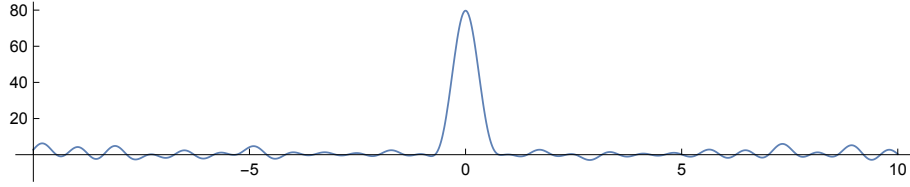
Figure 4.1 Plot of $Z(g_{70354406} + t)$ for $-10 < t < 10$, where $g_{70354406} \approx$ 30694257.761 is a Gram point.

At a Gram point the first term in the Riemann-Siegel formula is 1 or -1, and the other terms have no particular bias, so it is not surprising that [101]

$$\sum_{n \leq N,\ n \text{ even}} Z(g_n) \sim N \qquad \text{and} \qquad \sum_{n \leq N,\ n \text{ odd}} Z(g_n) \sim -N. \qquad (4.4)$$

Those averages are similar to the fact that the average value of $\zeta(\frac{1}{2} + it)$ is 1, see (3.8). Such biases, combined with Selberg's central limit theorem, are major contributors to:

Principle 4.2 *For questions about the size of the ζ-function, no numerical computation can give reliable evidence because the true nature of the ζ-function reveals itself on the scale of $\sqrt{\log \log T}$, which within the realm of computation is indistinguishable from a bias of order $O(1)$.*

Principle 4.2, and other sources for misinterpreting the data, are explored throughout this paper.

4.2 Three unfounded inferences

Brent's observations have re-appeared in subsequent numerical computations. The result is that reinforcing those observations has been (at least partially) a goal of recent computational work, particularly as it relates to understanding the largest values of the ζ-function. That is unfortunate because the largest values as they appear in computations are not representative of the largest values at greater height. Thus, the impressions one has from those data are not helping to build intuition for the true nature of the ζ-function. Quite the opposite: those numerical examples tempt one into mistaken notions.

The notions we refer to are:

Mistaken Notion 4.3 *The largest values of the ζ-function occur when there is a particularly large gap between zeros.*

Mistaken Notion 4.4 *The largest values of the ζ-function occur when a large number of initial terms in the Riemann-Siegel formula (4.1) have the same sign.*

Mistaken Notion 4.5 *Counterexamples to RH are more likely to occur near an unusually large gap between zeros.*

These Notions are part of the folklore of the ζ-function and the author has observed conversations on these Notions at many conferences, often accompanied by refutations. Thus, it seemed prudent to address these issues explicitly.

We do not attribute those Notions as conjectures due to any specific person. Indeed, in many cases where those Notions appear in the literature (implicitly or explicitly), the writer acknowledges the Notion without necessarily endorsing it. For example, concerning Mistaken Notion 4.4, Brent [23] says "This suggests that 'interesting' regions might be predicted by finding values of t such that the first few terms in the Riemann-Siegel sum reinforce each other." The "interesting" behavior of an unusually large gap does indeed occur when the initial terms reinforce, at least in the range accessible by computers. That

Notion is not Mistaken in the limited context of computing the ζ-function at moderate height.

The even more "interesting" behavior of a possible counterexample to RH requires a second step to get to Mistaken Notion 4.5. Odlyzko [81] explains it this way: "One reason for the interest in large values of $\zeta(1/2 + it)$ is that one could think of a large peak as 'pushing aside' the zeros that would normally lie in that area, and if these zeros were pushed off of the critical line, one would find a counterexample to the RH." Odlyzko is not endorsing Mistaken Notion 4.5, merely noting the reasoning which might have led to its formulation.

Brent's and Odlyzko's use of "scare quotes" is a further indication that they are not endorsing those Notions as fundamental principles. Indeed, it may be that individually the vast majority of experts understand the limitations of what can be learned from existing computations. But the frequency with which those Notions have been repeated, not always with qualifiers such as "it has been said that...", indicates the need for clarification. In Section 14 we revisit these Notions in the context of the Principles discussed in this paper, justifying the claim that the Notions are indeed "Mistaken".

4.3 Gram's law

If (g_n) are the Gram points for the Z-function, then we say **Gram's law holds** for the **Gram interval** (g_m, g_{m+1}) if that interval contains exactly one zero of the Z-function. Gram noted that this law holds for the first 15 zeros, and Hutchinson [63] found that it holds for the first 126 zeros, and he also coined the term "Gram's law". Subsequent work found that it holds for more than 91% of the first 1000 zeros, and most papers on zero calculations report statistics on Gram's law. Such numerical evidence, and repeated attention, leads to the archetypal mistaken notion concerning zeros:

Mistaken Notion 4.6 *Gram points are special, and Gram's law is helpful for developing intuition about the location of zeros of the ζ-function.*

It has been known for a long time that Gram's law fails infinitely often [101] and in fact fails a positive proportion of the time (although there are good reasons to believe it is true more than 66% of the time, see [102, 55]). Gram's law is not mistaken merely because it is not always true: it is mistaken because it paints a picture of the zeros which inhibits one from gaining a proper intuition about the behavior of the ζ-function at large height. Also, as we will explain in Subsection 14.4, if Gram's law is stated properly, it is true 0% of the time.

5 Separating out the local zero spacing

Figure 3.1 and Figure 3.4 show that the local spacing of the zeros has an influence on the size of the ζ-function: when there is a large gap the function is larger, and when there is a small gap the function stays small. Unfortunately, those observations require one to *ignore the vertical scale*. The specific spacing of nearby zeros strongly influences the *relative* sizes of the *nearby* maxima and minima. But the local spacings have very little to do with whether or not the actual function values are particularly large or small, compared to what one would expect for the ζ-function in that region.

Indeed, by the end of this section we will see that the local spacing of zeros is not the leading order contribution to the size of the ζ-function.

In general, the size of the ζ-function is controlled by the ζ-function's **carrier wave**. The terminology is due to Hejhal [58], with detailed discussion by

Bombieri and Hejhal [20]. The idea of carrier waves is based on some unpublished speculations of H.L. Montgomery. The main idea is that, on a logarithmic scale, the ζ-function changes its size slowly. If it is large, then outside a set with small measure, it usually stays large for a while. (The "set with small measure" is small neighborhoods of the zeros.) If it is small, it usually stays small for a while. Here "a while" can be interpreted as "on an interval containing many zeros". In particular, the typical large values occur as clusters of large local maxima and minima, not as a single isolated large maximum. That is contrary to what we see in graphs of $Z(t)$, which is why we have Mistaken Notion 4.3. That is why for many people, this idea is in the "I find that hard to believe" category. Indeed, the phenomenon of carrier waves is not visible in any numeric computation of the ζ-function, because the scale of the carrier waves, just like the scale of $S(t)$, grows so slowly that it appears bounded within the range we can compute. However, we can take a first-principles approach to defining what we mean by "carrier wave", and then build intuition by looking at illustrative examples.

5.1 The wave as a local constant factor

Suppose $f(t)$ is a high degree polynomial, with zeros $\gamma_1 < \ldots < \gamma_M$, and suppose we want to understand the graph of f near t_0. Further suppose that t_0 is roughly near the middle of the γ_j. We can write

$$f(t) = a_0 \prod_{\gamma_j \text{ near } t_0} \left(1 - \frac{t}{\gamma_j} \right) \prod_{\gamma_j \text{ far from } t_0} \left(1 - \frac{t}{\gamma_j} \right) \tag{5.1}$$

$$\approx A_0 \prod_{\gamma_j \text{ near } t_0} \left(1 - \frac{t}{\gamma_j} \right) \qquad \text{for } t \text{ very close to } t_0. \tag{5.2}$$

The approximation in the second line above will be valid for t in a neighborhood of t_0 if the zeros far from t_0 are balanced on either side of t_0, meaning that the product over "γ_j far from t_0" is approximately constant near t_0.

   L-functions are much like high degree polynomials, and the global spacing of their zeros is very regular (as can be seen from the small error term in the zero counting function $N(T)$, see (3.9)). Thus we have:

Principle 5.1 *The behavior of $Z(t)$ near t_0 depends on three things: a global factor independent of t_0, the arrangement of the zeros near t_0, and a scale factor which depends on the zeros far from t_0 and which does not change too quickly as a function of t_0.*

   Figure 5.3 illustrates some of the ideas in Principle 5.1.

   Principle 5.1 does not specify which of the two factors that depend on t_0 are most relevant to the size of $Z(t)$. Any graph of $Z(t)$ in the range accessible by current computers, such as in Figure 3.4, makes it appear that the arrangement of the zeros is more important. We will see that those examples are misleading.

5.2 Carrier waves

The scale factor $A_0 = A_0(t_0)$ has been termed the **carrier wave** by Hejhal [58] and Bombieri and Hejhal [20], making rigorous a speculation of Montgomery. Montgomery's preliminary calculations suggested that the carrier wave for the Riemann zeta-function should not vary too much over a window of width $\exp(\delta_T \log \log T)/\log T$, for some function $\delta_T \to 0$. Bombieri and Hejhal [20]

show that, for most T, the carrier wave does not vary significantly over a window of width $M/\log T$, for any fixed $M > 0$, as $T \to \infty$. That is: across a span of M consecutive zeros, for any fixed M, the size of $\log|\zeta(\frac{1}{2}+it)|$ usually varies very little. (The proof in [20] might actually show that one can take $M = \log\log(T)^\kappa$ for any $\kappa < \frac{1}{4}$, with (6.21) in that paper providing the limiting constraint, but those details would need to be checked.) See Subsection 11.3 for a discussion of how wide is the carrier wave.

Bombieri and Hejhal used carrier waves as the key idea toward their proof of the following surprising theorem: if $L_1(s)$ and $L_2(s)$ are L-functions which individually satisfy RH and have the same functional equation, and $\alpha \in \mathbb{R}$, then $L_1(s) + \alpha L_2(s)$ has 100% of its zeros on the critical line (there are technical conditions which we have omitted). The proof is: the $\log(L_j(\frac{1}{2}+it))$ are independently and normally distributed with a large variance, so most of the time one of the L_j is much larger than the other, and furthermore (because of the carrier wave) it stays larger across an arbitrarily many zeros. Therefore in that region the zeros of the linear combination are very close to the zeros of the larger L-function, and that accounts for most of the zeros. (The large variance is an important ingredient: given two Gaussians with variance σ^2, the probability that they differ by less than $\sqrt{\sigma}$ is $O(\sigma^{-\frac{1}{2}})$. The above argument requires $\sigma \to \infty$.)

Note that Selberg's theorem on the normal distribution of $L(\frac{1}{2}+it)$ is not sufficient: one needs the additional fact that the carrier wave causes the larger L-function to stay large over a significant range.

Thus we have:

Principle 5.2 *The carrier wave is responsible for the bulk of the value distribution of an L-function, with the variation due to the local zero distribution playing a secondary role. In particular, it is the carrier wave which obeys Selberg's central limit theorem.*

Principle 5.2 explains why, in the range accessible by current computers, the observed value distribution of $\log|Z(t)|$ departs significantly from normal. The local zero spacing contributes a lesser (typically, bounded) amount. But a bounded amount is significant in the range where the carrier waves are very small.

5.3 Measuring the wave

We now describe a way to measure and observe the carrier wave. The goal is to isolate the contribution of the nearby zeros, as suggested in (5.2). The idea is to think of the zeros as parameters which can change: we can slide the zeros side-to-side, and this will cause a change in the graph of $Z(t)$. (This is easier to picture if all the zeros are real, which we will assume for the purposes of this thought experiment.) If we slide the zeros apart, making a large zero gap, then the function will acquire a large local maximum. If we slide the zeros so they become more equally spaced, then the maxima of the function will be approximately the same size. In the extreme case of moving the zeros to be equally spaced, the result will be the (scaled and shifted) cosine function.

It is not possible to achieve a *globally* equal spacing for the zeros, because the local average spacing of the zeros is not constant. Instead we focus on "nearby" zeros, where the average spacing is close to constant. Suppose $t_0 \in [\gamma_M, \gamma_{M+1}]$. If K is large enough and we slide the zeros $\gamma_{M-K}, \ldots, \gamma_{M+K}$ to be equally spaced, then near t_0 the resulting function will look like $a\cos(b(x+c))$ for some $a, b, c \in \mathbb{R}$.

We will run the above process in reverse. That is, start with $a\cos(b(x+c))$, pick a region of interest in the critical strip containing zeros $\gamma_{M-K}, \ldots, \gamma_{M+K}$,

and then "move" the cosine zeros to the ζ-zeros. This suggests that

$$Z(t) \approx A_Z \, a \cos(b(t+c)) \prod_{j=M-K}^{M+K} \frac{t-\gamma_j}{t-g_j}, \qquad (5.3)$$

where the g_j are zeros of $\cos(b(t+c))$, indexed so that g_j is close to γ_j, the parameters a, b, c depend on t_0 and K, and A_Z is a normalization factor. The approximation should be good near t_0 if K is large enough and t_0 is near the middle of the interval $[\gamma_{M-K}, \gamma_{M+K+1}]$. This approximation works because locally the Z-function is determined by the nearby zeros and a scale factor.

A similar approximation was considered by Hiary and Odlyzko ([60], Section 6), who refer to it as "HP", because essentially it uses a partial Hadamard Product for the approximation. They find that the quality of the approximation improves linearly with the number of zeros in the product. A difference in the approach here is to recognize that the zeros far from t_0 are basically contributing a multiplicative constant near t_0.

We apply (5.3) to $Z(t)$ from Figure 3.1, with

$$\begin{aligned}
t_0 &= 5448 \\
M &= 5000 \\
\gamma_M &\approx 5447.8619 \\
K &= 10 \\
A_Z &= 2 \\
a &= 1.0991 \\
b &= 3.3825 \\
c &= 0.17008.
\end{aligned} \qquad (5.4)$$

The parameter $A_Z = 2$ is explained after Note 5.4, as is the fact that b and c are actually simple functions of t_0. The parameter a is chosen so that the approximation is exact at t_0. So, once t_0 is selected, the only choice is K, the number of zeros to be matched on either side of t_0. The result is shown in the top plot in Figure 5.3, where we superimpose $Z(t)$ and its approximation based at $t_0 = 5447.86$ using 20 matched zeros.. The bottom plot shows the same idea based at $t_0 = \gamma_{7010} \approx 7273.70$, which has scale factor $a = 0.9130$. In both cases one sees that the behavior near t_0 is determined by the nearby zeros and a local scale factor (which depends on t_0).

In the top plot in Figure 5.3 we can interpret the scale factor $a = 1.0991$ as the magnitude of the carrier wave at $t_0 = 5447.86$, and similarly for the scale factor $a = 0.9130$ at $t_0 = 7273.70$. Those scale factors are different, which shows that the magnitude of $Z(t)$ is not solely due to the local zero spacing and a single global factor.
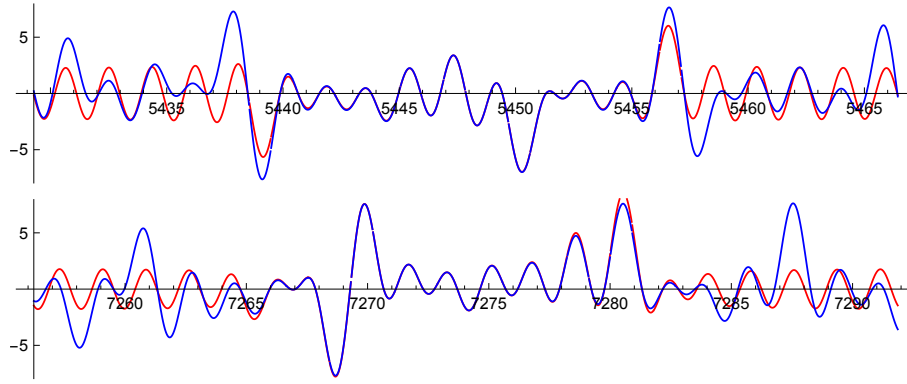
15

Figure 5.3 The function $Z(t)$ together with a local approximation as in (5.3) with $K = 10$. The top graph has t near 5447.86 and the local scale factor is 1.0991. The bottom graph has t near 7273.70 and the scale factor is 0.913.

Here we are focusing on (5.3) because it explicitly separates the contribution of the local zeros from the contribution of the carrier wave. There are other ways to approximate the ζ-function, see for example Subsection 10.2.

5.4 Some caveats

The scale factors in Figure 5.3 are not the carrier wave in the sense of Bombieri-Hejhal, because Montgomery's heuristic calculation suggests that at the small height of that example the carrier wave should only be approximately constant in a very narrow window. In Figure 5.3 we have fit to the zeros in a much wider window. Nevertheless, those plots illustrate that the nearby zeros and a single local scale factor determine the local behavior of $Z(t)$.

Note 5.4 The scale factor at a point is not well-defined: it depends on K, the number of zeros, or more precisely the width of the window, over which one has done the fit. If the window covers an enormous number of zeros then the local scale factor will just be 1, because of the long-range rigidity of the zeros; see Subsection 8.2. If the function changes scale in the window, i.e., the carrier wave is not approximately constant in the window, then the scale factor is not providing useful information. This is discussed further in Note 8.6.

It remains to justify that the global scale factor (independent of t_0) is $A_Z = 2$. That comes from the overall factor of 2 in the Riemann-Siegel formula (4.1).

The parameters b and c in (5.3) are also extraneous: instead of $A_Z \cos(bt+c)$ we could use the first term in the Riemann-Siegel formula:

$$2 \cos(\theta(t)).$$

Indeed $\theta'(5447.86) \approx 3.38255$, which is the value for b in (5.4). Using $2 \cos(\theta(t))$ in (5.3) does not change the point illustrated by Figure 5.3.

In order to really "see" the carrier wave, one must go to enormously larger values of t. That is not computationally feasible in the L-function world, but it is easy in the random matrix world. In Section 6 and Section 7 we review the connections between the ζ-function and the characteristic polynomials of random unitary matrices, returning to carrier waves in Section 8.

5.5 The highest tone

Many observations about the zeta-function can be traced to the fact that $S(t)$, the error term in the zero counting function $N(t)$, grows very slowly and is zero

16

on average. Here we collect some additional consequences.

Principle 5.5 *The Z-function, interpreted as a sound wave made up of separate tones, has a highest frequency component which is the loudest tone and which is separated in frequency from the lower tones. In particular, the Z-function is (approximately) a band-limited function with discrete support.*

Specifically, at height t the highest frequency is $\frac{1}{2}\log(t/2\pi)$, the factor of $1/2$ coming from the fact that sin or cos have two zeros in each period. Each subsequent frequency is lower by $\frac{1}{2}\log(2)$, $\frac{1}{2}\log(3)$, ..., and its contribution is smaller by a factor $1/\sqrt{2}$, $1/\sqrt{3}$,.... These observations come straight from the Riemann-Siegel formula (4.1), and are illustrated in Figure 5.6, which shows the Fourier transform of $Z(t)$ for $t \approx 10^6$. Note that $\log(10^6/2\pi) \approx 11.978$.
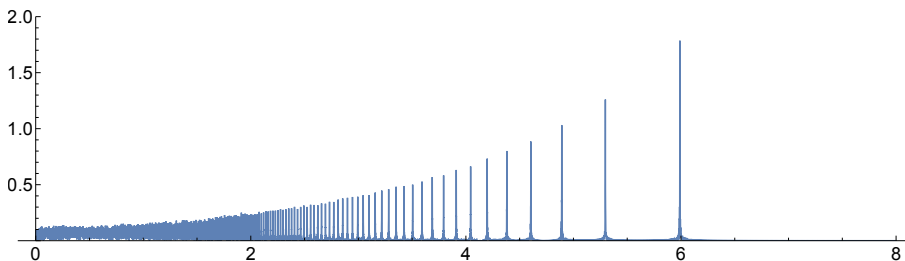

Figure 5.6 The Fourier transform of $Z(t)$, computed numerically, for $10^6 \leq t \leq 10^6 + 5000$.

The Z-function is not actually band-limited, a fact which is relevant to the phenomenon of superoscillations, see [10, 11]. However, it is close enough to band-limited that band-limited interpolation has become an indespensible tool in large-scale computations of the zeros [79, 17].

Principle 5.7 *The highest tone of the Z-function comes from the Γ-factor in the functional equation, and the lower tones come from the Dirichlet coefficients. The Γ-factor places the zeros in precise well-spaced locations, and then the Dirichlet coefficients make adjustments.*

To see that interpretation, view the Riemann-Siegel formula (4.1) as arising from the expression $Z(t) = X(\frac{1}{2} + it)^{-1/2}\zeta(\frac{1}{2} + it)$ for the Z-function in terms of the zeta-function, see (3.6), combined with the **approximate functional equation**, (4.12.4) of [101],

$$\zeta(s) = \sum_{1 \leq n \leq x} \frac{1}{n^s} + X(s) \sum_{1 \leq n \leq y} \frac{1}{n^{1-s}} + \text{ correction terms}, \qquad (5.5)$$

where $xy = t/2\pi$. Setting $x = y$, combining terms, and using Stirling's formula for the Γ-functions in $X(\frac{1}{2} + it)$ yields the Riemann-Siegel formula. The $n = 1$ terms contribute $2\Re X(\frac{1}{2} + it)^{-1/2}$, which is bounded by 2 and has regularly spaced zeros with the same counting function as the zeros of the zeta function (but with error term bounded by 1). Those zeros interlace the Gram points. We see that Gram's law is a consequence of a more general principle:

Principle 5.8 The first term dominates at low height. *When the analytic conductor of an L-function is small, the first term in the Riemann-Siegel formula has a strong influence on the location of the critical zeros.*

We used the term "analytic conductor" so that the principle applies more widely. For the ζ-function at $s = \frac{1}{2} + it$, the analytic conductor is proportional to $\log(2+|t|)$. Figure 5.9 illustrates the principle by graphing the ζ-function and and the first term in its Riemann-Siegel formula, over the interval $[1000, 1050]$. The Gram points are the locations of the local maxima and local minima of the

red curve which oscillates between -2 and 2. Gram's law (see Subsection 14.4 for further discussion) follows from the expectation that at low height, the zeros of the ζ-function will be close to the zeros of the first term.
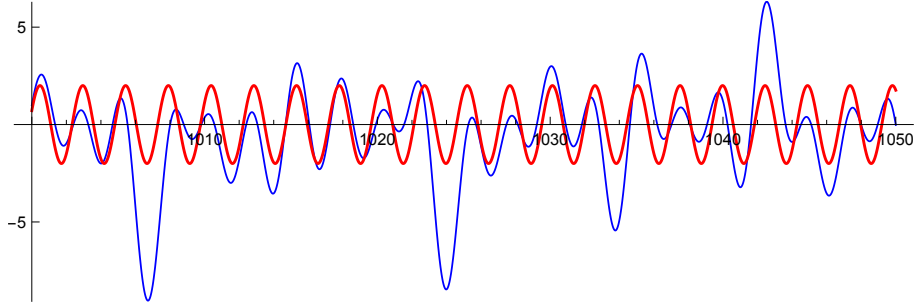

Figure 5.9 The functions $Z(t)$ and $2\cos(\vartheta(t))$ on the interval $1000 \le t \le 1050$.

There is a Riemann-Siegel formula for an L-function of degree d, with the modification that xy is of size t^d, the numerators of the two halves of the approximate functional equation are a_n and $\overline{a}_n$, respectively, and $X(s)$ is replaced by $\varepsilon X(s)$. For large t there is highest frequency term, with frequency asymptotically $\frac{d}{2}\log t$, involving only the data in the functional equation. And just like in the classical case, initially the Dirichlet coefficients make small adjustments to the location of the zeros arising from the Γ-factors. The Γ-factors have a surprising influence on the initial zeros, particularly for nonarithmetic L-functions, whose trivial zeros have nonzero imaginary part. That influence can be captured by working directly with $X(\frac{1}{2}+it)^{-1/2}$ instead of using the asymptotics from Stirling's formula. See [47].

6 A random model for the ζ-function

A graph of $Z(t)$ gives the impression of randomness: the function wiggles, and there is no apparent pattern to those wiggles. Obviously $Z(t)$ is not random, because it is a specific function which has no randomness in its definition. But we can make a random function in the following way. Fix an interval $\Upsilon = [t_0, t_1] \subset \mathbb{R}$. If $T \in \mathbb{R}$ is random, then $Z_T(t) := Z(T+t)$ is a random function on Υ.

If we had another set of random functions $\mathcal{Z}_T$ on Υ, and it was possible to prove theorems about $\mathcal{Z}_T$, and if furthermore we had reason to believe that Z_T and $\mathcal{Z}_T$ had similar properties, then we could turn theorems about $\mathcal{Z}_T$ into conjectures about Z_T. That would be illuminating, particularly if precise conjectures about Z_T were in short supply. That is how random matrix theory made fundamental contributions to the study of L-functions.

6.1 Self-reciprocal polynomials and the functional equation

Before getting into the details, let's consider a certain class of polynomials which have constant term 1:

$$f(z) = 1 + a_1 z + a_2 z^2 + \cdots + a_{N-1} z^{N-1} + a_N z^N. \qquad (6.1)$$

The polynomial f is **self-reciprocal** if

$$a_j = a_N \overline{a_{N-j}} \qquad \text{for} \qquad 0 \le j \le N, \qquad (6.2)$$

18

or equivalently

$$f(z) = \mathcal{X}(z)\overline{f}(z^{-1}) \quad \text{where} \quad \mathcal{X}(z) = a_N z^N, \tag{6.3}$$

or equivalently

$$\mathcal{Z}_f(\theta) = \mathcal{X}(e^{i\theta})^{-\frac{1}{2}} f(e^{i\theta}) \quad \text{is real if} \quad \theta \in \mathbb{R}. \tag{6.4}$$

An important consequence of any of these conditions is that

$$f(re^{i\theta}) = 0 \quad \text{if and only if} \quad f(r^{-1}e^{i\theta}) = 0. \tag{6.5}$$

In (6.3), $\overline{f}$ is the **Schwarz reflection** of f: $\overline{f}(z) := \overline{f(\overline{z})}$. The parallel with the functional equation for the ζ-function is evident if one compares (6.3) to (3.3), and (6.4) to (3.6). At the risk of belaboring the point, in both cases there is an involution of the plane, the set of zeros is fixed by the involution, the **Riemann Hypothesis** is the assertion that every zero is fixed by the involution, and there is a Z-function which is real on the set of points which are fixed by the involution. The unit circle is the analogue of the critical line for self-reciprocal polynomials. Thus, if a collection of polynomials is claimed to be a random model for the ζ-function, those polynomials must be self-reciprocal.

6.2 Random polynomials and trigonometric polynomials

We seek random self-reciprocal polynomials which can serve as a model for $\zeta(s)$ or equivalently $Z(t)$. How to choose the randomness? There are three reasonable options.

The first option is to choose the coefficients of $f(z)$ so that they satisfy (6.2). The second is to recognize that Z_f is a trigonometric polynomial (i.e. a finite Fourier series), so its coefficients can be chosen randomly with only the requirement that they are real. The third option is to choose the zeros so that they satisfy (6.5). We will argue that only the third of those options is reasonable.

Figure 6.1 shows $Z(t)$ at height 1.5×10^{30}, in a region where nothing special is happening. (Data courtesy of Bober and Hiary [17][59].)
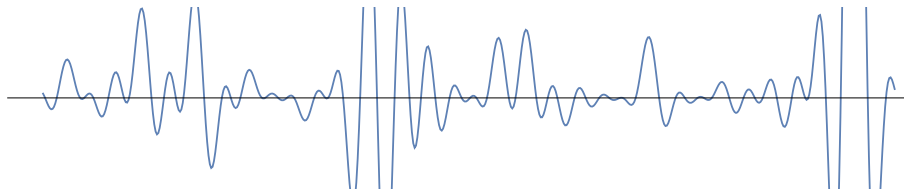
Figure 6.1 $Z(t)$ on an interval of width 2π near 1.5×10^{30}

We wish create random polynomials which capture the type of randomness visible in Figure 6.1. But what does that mean? The following features are worthy of attention.

(A) Do all the zeros appear to be real?

(B) How often does the graph get big? Almost equivalently, how often are there large gaps between zeros?

(C) How often are pairs of zeros close together? How often are there clusters of close together zeros?

(D) Are the maxima and minima generally around the same size, with only occasional much larger or smaller local extrema?

19

Item D is obviously a leading question. The answer is "yes".

Item A refers to RH. We can't escape the fact that RH is true within the realm where we can experiment, so the random polynomials we seek must satisfy RH. That is, we seek **unitary polynomials**, which is standard terminology for polynomials with all zeros on $|z| = 1$.

The requirement of generating unitary polynomials effectively rules out the first two options for choosing random polynomials. If the coefficients are chosen randomly, then one can choose small coefficients to obtain a perturbation of $z^N - 1$. The zeros will be on the unit circle, but they will be close to regularly spaced and the function will not have randomness similar to Figure 6.1. If the coefficients are random and large, then the polynomial will usually not be unitary (although it may have a large proportion of its zeros on the unit circle). Choosing the coefficients randomly and then conditioning on having all zeros on the unit circle turns out to produce the wrong type of randomness [44]. (It gives the COE, not the CUE.)

By process of elimination we conclude that the polynomials we seek must arise by having their zeros randomly generated, and on the unit circle. Such polynomials have arisen in mathematical physics, which we describe next.

Note 6.2 It would be interesting to define a natural-looking ensemble of random self-reciprocal polynomials which are unitary most of time, but not always. If such an ensemble was consistent with RH being true within the realm of current computation, and it predicted a height at which RH would start to fail, that could throw doubt on RH. Indeed, one of the great shortcomings of RH skepticism is the lack of conjectures for: the height at which RH fails, how often zeros would be off the line, and how their real parts would be distributed. Principle 2.2 implies that even if RH fails, 100% of the zeros are on the critical line.

6.3 The circular β-ensembles

The **circular β-ensemble** consists of N random points $e^{i\theta_1}, \ldots, e^{i\theta_N}$ on the unit circle, equipped with the joint probability density function

$$\frac{\Gamma(1 + \beta/2)^N}{\Gamma(1 + N\beta/2)(2\pi)^N} \prod_{1 \leq j < k \leq N} |e^{i\theta_j} - e^{i\theta_k}|^\beta. \qquad (6.6)$$

This probability space is denoted CβE(N). We use $\langle f(\theta_1, \ldots, \theta_N) \rangle$ for the **expected value** of f averaged over CβE(N).

The CβE(N) is defined for any $\beta \geq 0$ and any positive integer N. We can obtain a random unitary (hence, self-reciprocal) polynomial by taking the points to be the zeros of a polynomial with constant term 1, as in (6.1).

If we write the θ_j in increasing order, then it is clear that $\langle \theta_{j+1} - \theta_j \rangle = 2\pi/N$. Almost equally clear is that the PDF of $\theta_{j+1} - \theta_j$ vanishes to order β at 0. For that reason β is referred to as the **order of repulsion** between the points.

As $\beta \to 0$ the repulsion disappears and the points become independently uniformly distributed on the circle. As $\beta \to \infty$ the points approach equal spacing (known as the **picket fence** distribution), with the only remaining randomness being the angle of an initial point which determines the location of all the others. Thus, as $\beta \to 0$ the particles behave like a gas, and as $\beta \to \infty$ they crystallize. For this reason, β is sometimes referred to as **inverse temperature**.

Between those temperature extremes there are three values which have been studied extensively in the mathematical physics literature: $\beta = 1$, 2, and 4. Those are more commonly known as the COE, CUE, and CSE, where the

O/U/S stand respectively for Orthogonal/Unitary/Symplectic. Examples of random polynomials $\mathcal{Z}(\theta)$ from each of those ensembles are shown in Figure 6.3. Note that the graphs all have the same vertical scale, ranging from -6 to 6. The graphs show the effect of β on the frequency of small zero gaps, large zero gaps, and large values. Those examples have $N = 68$, that choice ensuring the polynomials have the same number of zeros in a span of 2π as the Z-function shown in Figure 6.1. Our next goal is to persuade that one of those ensembles provides a good model for the ζ-function.

Figure 6.3 Example random polynomials from CβE(68) with, reading top to bottom, $\beta = 1, 2$, and 4.

Hopefully one of the plots in Figure 6.3 looks like it has similar randomness to the Z-function in Figure 6.1, because otherwise we are out of ideas. The bottom plot, $\beta = 4$, is ruled out in multiple ways. The 4th order repulsion causes the zeros to be too rigidly spaced, which in turn prevents the function from having sufficiently many large or small values. Polynomials from the CSE are pleasant but rather boring, lacking the pizzazz of the Riemann ζ-function.

The top plot in Figure 6.3 can be ruled out by carefully examining the features of the function. The most obvious property is that the COE example is large much more often: consider the width of the cutoff at the large maxima. The COE example also has more small zero gaps, both individually (small local extrema) and in clusters.

The CSE is too rigid to serve as a model for the ζ-function, and the COE is too flexible. The Goldilocks zone is occupied by the CUE which is conjectured to provide, with appropriate adjustments and caveats, a model for the zeros of the ζ-function. We clarify the connection, and provide more terminology and historical details, in Section 7.

Warning 6.4 The CUE has the same distribution as the eigenvalues of Haar-random matrices from the compact unitary group $U(N)$. But the COE and CSE, and the GOE, GUE, and GSE described in Subsection 7.1, do not correspond to Haar measure on classical matrix groups with similar names. The eigenvalues of Haar-random matrices from the unitary symplectic and unitary orthogonal groups all have quadratic repulsion for the bulk of their eigenvalues, with different behavior for the eigenvalues close to 1. These play a role in modeling the low-lying zeros of certain families of L-functions, see [70, 87].

6.4 Where does the CUE come from?

To do calculations involving the CUE, such as computing an expectation $\langle f \rangle$, all one needs is the joint law of the zeros: (6.6) with $\beta = 2$. That is, it does not matter how the points are generated. But to construct an example, such as the plots in Figure 6.3, one needs a way to produce actual sets of points with the appropriate distribution. Here are some ways:

1. *Brownian motion.*

   Have N points start at the origin and undergo non-intersecting Brownian motion until they hit the unit circle. Those points will be distributed according to (6.6). (There are Brownian motion models for the CUE and other β-ensembles. See [38].)

2. *A product of tridiagonal matrices.*

   Let Ξ_k be certain independent random 2×2 matrices, depending on a parameter β, defined in [72]. Let

   $$L = \mathrm{diag}(\Xi_0, \Xi_2, \ldots, \Xi_{\lfloor N/2 \rfloor}) \quad \text{and} \quad M = \mathrm{diag}(\Xi_{-1}, \Xi_1, \ldots, \Xi_{\lfloor N/2 \rfloor - 1}).$$

   The eigenvalues of the tridiagonal matrix LM, equivalently ML, will be distributed according to the $\mathrm{C}\beta\mathrm{E}(N)$.

3. *The unitary group $U(N)$.*

   Let $A \in U(N)$ be chosen randomly with respect to Haar measure. The eigenvalues of A will be distributed according to (6.6) with $\beta = 2$.

The final option is the easiest to implement, but see Mezzadri's paper [76] which describes possible pitfalls.

To summarize:

Principle 6.5 *The connection between random matrices and the local statistics of zeros of the ζ-function is not actually about random matrices: all that really matters is the $\beta = 2$ repulsion (6.6) on the zeros/eigenvalues. Random matrices, either Haar-random from $U(N)$ or products of tridiagonal matrices [72], just happen to be convenient ways to generate such distributions.*

However, for historical reasons, and for lack of a good alternate term, it is common to refer to the N points in the $\mathrm{C}\beta\mathrm{E}(N)$ as "eigenvalues", even if they do not arise from a matrix.

The description we have given is incomplete and our claim of a connection between the CUE and the ζ-function is perhaps not persuasive, so in Section 7 we provide more details about random matrix eigenvalues, random unitary polynomials, and the connection with L-functions. Then in Section 8 we return to our main theme and explore carrier waves in random polynomials from the CUE.

7 RMT and L-functions: terminology and history

The $\mathrm{C}\beta\mathrm{E}$ were not how Random Matrix Theory (RMT) first arose in mathematical physics. RMT came about in the 1950s as a way to understand complicated quantum mechanical systems. For example: the energy levels of a large atomic nucleus. The justification is that the Hamiltonian of the system is a large and complicated matrix with certain symmetries, and so the eigenvalues of the Hamiltonian should have similar statistical properties to a large random matrix with the same symmetries.

The connection with L-functions arose in three phases.

1. *The* GUE *era.*

   The original formulation in the early 1970's concerned the limiting local statistics of eigenvalues/zeros. By *limiting* we refer to the rescaled eigenvalues in the large matrix limit. The matrices are Hermitian so the eigenvalues are real. See Subsection 7.1 for more details.

2. *The classical compact group era.*

   In the late 1990's it was realized that certain compact unitary groups of matrices could be associated with families of L-functions. The size of the matrices is a natural function of the conductor of the L-functions (so one could model at finite height), the particular matrix group is determined from properties of the L-functions, and the characteristic polynomial can be used to model the values of the L-functions. See Subsection 7.3 for more details.

3. *The recipe era.*

   In the early 2000's a heuristic, inspired by RMT but not actually making use of random matrices, was developed which reproduces virtually every conjecture arising from RMT. In particular, the heuristic for ratios can be used to obtain all correlation functions of the zeros. See [33]. Furthermore, the heuristic produces the complete main term and not just the leading order behavior. The heuristic goes by the name "the recipe". See the end of Subsection 7.5 for a brief mention.

Despite the fact that the recipe has largely replaced RMT as a tool for making conjectures about L-functions, random matrices are essential for our discussion of carrier waves, because characteristic polynomials provide concrete examples which we can visualize.

7.1 The GUE era

The matrices in this setting are Hermitian, so the eigenvalues are real. The random entries are real, complex, or quaternionic depending on the type of physical system being modeled, and the randomness is Gaussian. As in the circular ensembles, the three Gaussian ensembles are named GOE, GUE, and GSE, where again O/U/S stands for Orthogonal/Unitary/Symplectic. Again there is a single parameter β which describes the eigenvalue repulsion, with $\beta = 1$, 2, or 4 for each of those cases, respectively. And again the way the eigenvalues (or energy levels) are generated could be Brownian motion [38], or tridiagonal matrices [39], or the original formulation of random Hermitian matrices with a given symmetry. The only concern for our purposes is that one has N points on the real line with joint probability density function:

$$\frac{\Gamma(1+\beta/2)^N}{(2\pi)^{N/2}\prod_{j=1}^N \Gamma(1+j\beta/2)} \exp\left(-\tfrac{1}{2}\sum_{j=1}^N \lambda_j^2\right) \prod_{1\leq j<k\leq N} |\lambda_j - \lambda_k|^\beta. \qquad (7.1)$$

A similarity with the CβE is the order β repulsion between points. A difference is that the points are on a line, with the factor $\exp(-\tfrac{1}{2}\lambda_j^2)$ keeping more of the points near the origin and preventing the points from wandering too far away. There is a tension between the factor keeping the eigenvalues close to the origin and the order β repulsion which tries to keep them apart. The result is that $|\lambda_j| < \sqrt{2\beta N}$ for most j and the rescaled points $\lambda_j/\sqrt{\beta N}$ are distributed according to the **Wigner semicircle law** $\pi^{-1}\sqrt{2-x^2}$.

The similarities outweigh the differences. The GβE and CβE belong to the same **universality class** which is characterized only by order β repulsion between points. Let's be specific about which properties are universal. Define the rescaled eigenvalues $\tilde{\lambda}$ so that $\langle \tilde{\lambda}_{j+k} - \tilde{\lambda}_j \rangle = k$ for all fixed k. Consider any **local statistic**, that is, a statistic only involving a finite number of differences of rescaled eigenvalues. Examples of local statistics are the normalized nearest neighbor spacing

$$p_2(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \left\langle \frac{\#\{\tilde{\lambda}_{j+1} - \tilde{\lambda}_j \in [x, x + \Delta x]\}}{N} \right\rangle, \tag{7.2}$$

the pair correlation function

$$R_2(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \left\langle \frac{\#\{\tilde{\lambda}' - \tilde{\lambda} \in [x, x + \Delta x]\}}{N} \right\rangle, \tag{7.3}$$

or more generally the joint distribution of $(\tilde{\lambda}_{j_1} - \tilde{\lambda}_{j_0}, \ldots, \tilde{\lambda}_{j_k} - \tilde{\lambda}_{j_0})$ for any j_0 and $j_1 < \ldots < j_k$. Note that the focus is on the relationship between an eigenvalue and its neighbors, not on their absolute location.

The quantities above depend on N, but we suppress that notation because the leading order behavior is independent of N.

Principle 7.1 Universality. *For each β, all β-ensembles have the same limiting local eigenvalue statistics as $N \to \infty$.*

Principle 7.1 comes from the fact that the β-repulsion term is the most important feature of the measure. Here is a less hand-wavy explanation in the case of the CUE(N) and the GUE(N). Both ensembles are examples of a **determinental point process**. This means there is a **kernel function** $K_N(x, y)$ such that the n-correlation function of the eigenvalues can be expressed as the determinant of an $n \times n$ matrix with entries involving $K_N(x, y)$. For example, the pair correlation function is given by

$$R_{2,N}(x - y) = \begin{vmatrix} 1 & K_N(x, y) \\ K_N(y, x) & 1 \end{vmatrix}. \tag{7.4}$$

Thus, the kernel function determines all the correlation functions, so it also determines all the local statistics (because those can be expressed in terms of the correlation functions, possibly involving a complicated inclusion-exclusion argument).

For GUE(N) the kernel function is

$$K^{\mathrm{GUE}(N)}(x, y) = \sqrt{N} \frac{\psi_N(x) \psi_{N-1}(y) - \psi_{N-1}(x) \psi_N(y)}{x - y}, \tag{7.5}$$

where $x, y \in \mathbb{R}$. Here $\psi_n(x) = He_n(x) e^{-x^2/4} / \sqrt{\sqrt{2\pi} n!}$, where $He_n(x)$ is a Hermite polynomial. Change variables $(x, y) \mapsto (4x/\sqrt{N}, 4y/\sqrt{N})$ for the normalized neighbor spacing.

For CUE(N) the kernel function is

$$K^{\mathrm{CUE}(N)}(\theta, \phi) = \frac{1}{N} \frac{\sin(N(\theta - \phi)/2)}{\sin((\theta - \phi)/2)}, \tag{7.6}$$

where $\theta, \phi \in \mathbb{R} \mod 2\pi$. Change variables $(\theta, \phi) \mapsto (2\pi\theta/N, 2\pi\phi/N)$ for the normalized neighbor spacing. To leading order those kernels are the same, because by properties of Hermite polynomials, both (normalized) large N limits are $\sin(\pi x)/(\pi x)$.

The difference between the kernel functions of the GUE(N) and CUE(N) is $O(1/N)$, so the same holds for all correlation functions. Figure 7.2 shows the pair correlation for each ensemble for $N = 50$, and their difference (with the GUE(N) pair correlation being larger in a neighborhood of 0).
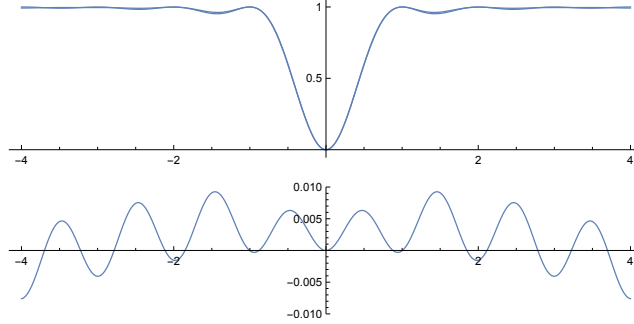

Figure 7.2 Pair correlation for GUE(50) and CUE(50), and their difference.

Note that by (7.4) and the limiting values of the kernel functions for the GUE and CUE, we find that both have the limiting normalized pair correlation function:

$$R_2(x) = 1 - \frac{\sin^2(\pi x)}{(\pi x)^2}. \tag{7.7}$$

That pair correlation function was the key to discovering the connection between random matrix eigenvalues and the zeros of the ζ-function.

7.2 Zero statistics for L-functions

The RMT revolution in number theory began when Montgomery [77] determined partial information about the pair correlation of the zeros of the ζ-function. The theorem he proved, combined with some heuristics about the prime numbers, led him to conjecture that asymptotically the pair correlation function was

$$R_{2,\zeta}(x) = 1 - \frac{\sin^2(\pi x)}{(\pi x)^2}. \tag{7.8}$$

Here

$$R_{2,\zeta}(x) := \lim_{\Delta x \to 0} \frac{1}{\Delta x} \lim_{T \to \infty} \frac{1}{N(T)} \sum_{0 < \gamma, \gamma' \leq T} \#\{\tilde{\gamma}' - \tilde{\gamma} \in [x, x + \Delta x]\} \tag{7.9}$$

is the ζ-function analogue of (7.3). Montgomery's conjecture became more significant when he met Freeman Dyson, who informed him that (7.8) was the limiting pair correlation of eigenvalues of the GUE, as shown in (7.7). Combined with extensive computations by Odlyzko [81], this established:

Principle 7.3 The GUE Hypothesis, aka the Montgomery-Odlyzko law. *The limiting local statistics of the zeros of the ζ-function are the same as the limiting local statistics of the GUE.*

The "limiting statistics" above refer to $T \to \infty$ and $N \to \infty$.

Consequences of the GUE Hypothesis.   The GUE Hypothesis has many useful consequences:

1. Montgomery's pair correlation conjecture (7.8).

2. 100% of the zeros of the ζ-function are simple.

3. The PDF $p_2(x)$, see (7.2), of the nearest neighbor spacing $\tilde{\gamma}_{j+1} - \tilde{\gamma}_j$ is given by an explicit expression, which is well-approximated in the bulk by the **Wigner surmise**

$$\frac{32}{\pi^2} x^2 e^{-\frac{4}{\pi} x^2},\tag{7.10}$$

which is the exact expression for GUE(2). The limiting behavior of $p_2(x)$ for GUE(n) as $n \to \infty$ is $\pi^2 x^2/3$ as $x \to 0$ and $\exp(-\pi^2 x^2/8)$ as $x \to \infty$.

The next two items are consequences of those limiting behaviors.

4. The smallest gaps between consecutive zeros satisfy

$$\tilde{\gamma}_{j+1} - \tilde{\gamma}_j \sim \gamma_j^{-\frac{1}{3}+o(1)},\tag{7.11}$$

that is,

$$\liminf_{j\to\infty} \frac{\log(\gamma_{j+1} - \gamma_j)}{\log \gamma_j} = -\frac{1}{3}.\tag{7.12}$$

It is possible to make a more precise statement. See Subsection 8.7.

5. The largest gaps satisfy

$$\tilde{\gamma}_{j+1} - \tilde{\gamma}_j \sim c_{large}\sqrt{\log \gamma_j},\tag{7.13}$$

where perhaps $c_{large} = 1/\sqrt{32}$ [8]. Or perhaps c_{large} is a bit smaller. See Subsection 8.7.

Information about small gaps between zeros has implications for the class number problem [32], which was Montgomery's original motivation for studying the pair correlation function [77].

Shortcomings of the GUE **Hypothesis.** The predictions of the GUE Hypothesis are spectacular, but it was clear from the beginning that only part of the story was being revealed. Odlyzko's computations showed a discrepancy with the predictions. The discrepancy decreased at larger heights, as would be expected since the GUE Hypothesis only refers to the limiting behavior, but clearly there was a need for a more precise model.

Another shortcoming is that the prime numbers are nowhere to be seen in the GUE Hypothesis. How can it be that the truth about the ζ-function has nothing to do with the prime numbers?

Finally, and most relevant to this present work, the GUE Hypothesis says little or nothing about the actual values of the ζ function. This appeared in Odlyzko's calculations (which computed the values of the ζ-function and not just the zeros), which found a large discrepancy between the computed value distribution and the limiting Gaussian distribution from Selberg's central limit theorem. The convergence to Gaussian is slow, see Figure 7.5, so it was not surprising to see a discrepancy. But the nature of the discrepancy is not explained by the GUE Hypothesis. It took until the year 2000 to address that shortcoming.

7.3 After 30 years of GUE: classical compact groups and the Keating-Snaith Law

In the late 1990's two innovations increased the influence of RMT on number theory. The first was due to Katz and Sarnak [68] who found that naturally arising collections of L-functions have a **symmetry type** which is given by a classical compact matrix group. The possible symmetry types are Unitary,

Symplectic, and Orthogonal, with Orthogonal splitting into two cases depending on whether the sign of the functional equation is always $+1$ or equally likely $+1$ or -1. The random matrices in this context are elements of the compact unitary groups $U(N)$, $Sp(N)$, and $O(N)$, where the randomness is uniform with respect to Haar measure. Note that $U(N)$ with Haar measure is the same as the CUE(N), but the others are completely different than their mathematical physics counterparts having similar names. In fact, all of those classical compact groups have the same bulk eigenvalue statistics: the only difference arises with the eigenvalues near 1. Numerical evidence [87], and much subsequent work, supports the conjecture that the symmetry type of the family of L-functions determines the distribution of zeros near the critical point.

Keating and Snaith [69][70] made the second key leap when they recognized that the characteristic polynomial could be used to model the L-function itself. Here the **characteristic polynomial** of the $N \times N$ matrix A is written in the slightly nonstandard form

$$\Lambda_A(z) = \det(I - zA) \tag{7.14}$$
$$= \prod_{1 \le j \le N} (1 - ze^{i\theta_j}),$$

which is in keeping with our normalization (6.1) for self-reciprocal polynomials. Here the $e^{i\theta_j}$ are the eigenvalues of A. The analogue of the Z-function is

$$\mathcal{Z}_A(\theta) = ((-1)^N \det A)^{-\frac{1}{2}} e^{-iN\theta/2} \det(I - e^{i\theta}A)$$
$$= ((-1)^N \det A \, e^{iN\theta})^{-\frac{1}{2}} \Lambda_A(e^{i\theta}), \tag{7.15}$$

which is real for $\theta \in \mathbb{R}$. We write $\mathcal{S}_A$ for the error term in the zero counting function:

$$\#\{\theta_j \in [0, X]\} = \frac{1}{2\pi}X + C_A + \mathcal{S}_A(X)$$

where C_A is chosen so that $\mathcal{S}_A$ is 0 on average. The number C_A is analogous to the constant $7/8$ in the zero counting function of the ζ-function, see (3.9).

We confine our discussion to the case of the ζ-function, which constitutes a unitary family, meaning that $\zeta(\frac{1}{2} + it)$ is modeled by the characteristic polynomial of Haar-random matrices from $U(N)$.

Principle 7.4 The Keating-Snaith Law. *Model $\zeta(\frac{1}{2} + it)$ for $t \approx T$ by $\Lambda_A(e^{i\theta})$ for Haar-random $A \in U(N)$, where $N = \log(T/2\pi)$.*

A key component of Principle 7.4 is the identification $N = \log(T/2\pi)$, where the equality means "is an integer close to". This is justified by "equating the density of zeros". When calculating asymptotics, where the only realistic hope is to make leading-order predictions, one can take $N = \log T$. But for numerical comparison involving small numbers, the more precise choice of N is helpful. See Subsection 8.5 for further discussion of $N = \log T$.

The first great success of Principle 7.4 was explaining the apparent discrepancy between numerical data and Selberg's central limit theorem. Keating and Snaith proved a central limit theorem for $\log \Lambda_A(e^{i\theta})$, exactly analogous to Selberg's result. Odlyzko's data, taken at height 1.52×10^{19}, does not look Gaussian, and is not even symmetric. But analogous data for $\Lambda_A(e^{i\theta})$ for Haar-random $A \in U(42)$ gives a close fit with the ζ-function data. See Figure 1 in [69]. The "42" comes from $\log(1.52 \times 10^{19}/(2\pi)) \approx 42.33$. (Analogous to Principle 5.2, it is the carrier wave which obeys the Keating-Snaith central limit theorem.)
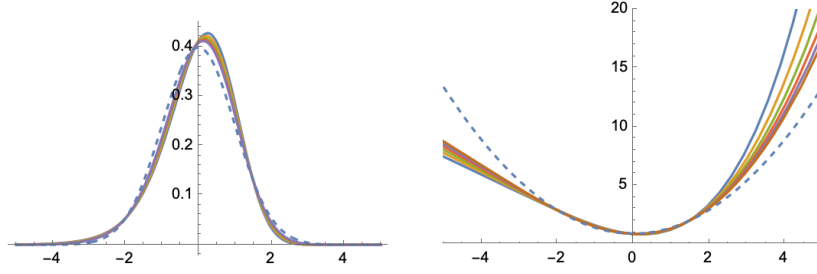
Figure 7.5 On the left is the PDF of $\log|\Lambda_A(e^{i\theta})|$ for random $A \in U(N)$ for $N = 25, 50, 100, 200, 400,$ and 800, each normalized to have unit variance. On the right is the negative logarithm of those PDFs. In both cases the dashed line is the limiting Gaussian distribution.

The second great success came from predicting the mysterious factor g_k in the conjectured moments (13.2) of the ζ-function. Previously it was known that $g_1 = 1$ and $g_2 = 2$. It had been conjectured [29, 31] that $g_3 = 42$ and $g_4 = 24024$. What Keating and Snaith found was (they compute an exact expression, but only the asymptotic is relevant for this discussion):

$$\left\langle |\Lambda_A(e^{i\theta})|^{2k} \right\rangle_{U(N)} \sim g_{U,k} N^{k^2}, \tag{7.16}$$

as $N \to \infty$, where

$$g_{U,k} = k^2! \prod_{j=0}^{k-1} \frac{j!}{(j+k)!}. \tag{7.17}$$

They found the beautiful equality $g_{U,k} = g_k$ for the known and conjectured values for $k = 1, 2, 3, 4$.

We make explicit some of the assumptions in Principle 7.4.

Principle 7.6 The Keating-Snaith Law, part 2. *To model* $\zeta(\frac{1}{2} + it)$ *throughout the interval* $[T, 2T]$, *choose* e^N *Haar-random characteristic polynomials from* $U(N)$, *where* $N \approx \log T/2\pi$.

In other words, the ζ-function is treated as being independent on each separate interval of length 2π. One could argue that $e^N/2\pi$ random matrices is a better choice, but that makes no difference in practice. Principle 7.6 is one of the many ways to obtain the conjectured maximum values (3.13) and (3.12) of $Z(t)$ and $S(t)$ [43].

The success of the Keating-Snaith Law is our justification to use characteristic polynomials from $U(N)$, equivalently $\text{CUE}(N)$, to explore the behavior of $Z(t)$ beyond the realm accessible by direct computation. We address some natural questions concerning the connection between RMT and the ζ-function before examining carrier waves in characteristic polynomials in Section 8.

Limitations of the Keating-Snaith law. The GUE Hypothesis is a statement about limiting behavior, and the Keating-Snaith law can be interpreted as a first order correction. Here we discuss the prospects of making a more precise version of the Keating-Snaith law.

It is obvious that random matrices cannot capture the subtle behavior of the ζ-function because the matrices do not know anything about the prime numbers. For example, when characteristic polynomials were first used to conjecture moments of the ζ-function, see (13.1), the arithmetic factor was inserted in an ad-hoc manner. A way to make the moment conjecture more natural is the "hybrid model" [52], which expresses the ζ-function as a product with two components: one involving the zeros and one involving the primes. See

Assuming a statistical independence of those two components, the leading term in conjectured moments arises in a natural way.

It is natural to wonder if a more precise hybrid model could produce lower order terms in the conjectured moments? Since the zero statistics of the ζ-function and of characteristic polynomials agree only to leading order, a more precise model would either have to produce eigenvalue statistics which incorporate arithmetic effects, or the arithmetic factor would somehow have to compensate for the incorrect lower order terms in the factor involving the zeros. Neither option seems plausible. The difficulty is further compounded by the fact that the two pieces cannot be treated as independent — because by inspection the main terms in the conjectured moments do not factor as a product. We say a bit more about this in Subsection 10.2.

A more detailed hybrid model incorporating the interactions between the primes and the zeros would be interesting, but it seems unlikely that such a construction exists. Even without that, the recipe already provides precise information about the moments and the zeros of the ζ-function. In particular, we know a lot about the lower order corrections to the Keating-Snaith law. In every case we find that the lower order corrections have a dampening effect. For example, the arithmetic factor in the conjectured moments, a_k, is very small. Another example is that the spacing of zeros of the ζ-function is more rigid than the eigenvalues of random matrices, see Subsection 8.6. This has the effect of making extreme gaps, or dense clusters of zeros, slightly less likely. To summarize:

Principle 7.7 *Lower-order corrections to the Keating-Snaith law have a dampening effect: the statistical behavior of the ζ-function and its zeros tend to be slightly less extreme than the analogous quantities for random unitary matrices.*

7.4 Modeling $Z(t)$ vs. modeling $\zeta(s)$

Our original motivation was to model $Z(t)$, but our description of the Keating-Snaith law referred to modeling $\zeta(s)$. To clarify:

Principle 7.8 *Model $Z(t)$ for $t \in \mathbb{R}$ by $\mathcal{Z}_A(\theta)$. Model $\zeta(\frac{1}{2} + a + it)$ for $a \in \mathbb{C}$ by $\Lambda(e^{i\theta - a})$.*

Some consequences of this model arise from the symmetries in the CβE measure (6.6):

Lemma 7.9 *The* CβE(N) *is invariant under the operations*

$$\{\theta_j\} \leftrightarrow \{\theta_j + \theta\} \tag{7.18}$$

$$\{\theta_j\} \leftrightarrow \{-\theta_j\}. \tag{7.19}$$

Note that the GβE, (7.1), has a symmetry analogous to (7.19), but no translation invariance as in (7.18).

By (7.18), the expected value of expressions like $\mathcal{Z}_A(\theta)\mathcal{Z}_A(\theta+\alpha)$ or $|\Lambda_A(e^{i\theta})|^k$ are independent of θ, so one typically sets $\theta = 0$. See for example (7.20) and (7.21). Also, since $\mathcal{Z}_A(\theta + 2\pi/N) = -\mathcal{Z}_{A'}(\theta)$ with A and A' equally likely, the distribution of $\mathcal{Z}_A(\theta)$ is symmetric (but the distribution of $\Lambda_A(\theta)$ is not: it has average value 1.) Thus, Principle 3.2 is a theorem for characteristic polynomials.

By (7.19), any (finite) sequence of gaps between eigenvalues is as likely as that sequence in reverse. Another way to say it is that $\mathcal{Z}_A(-\theta) = \mathcal{Z}_{A'}(\theta)$ with A and A' equally likely. This explains some conjectures from [96], see Subsection 14.5.

Modeling $Z(t)$ by $\mathcal{Z}_A(\theta)$ should seem natural: both are real-valued functions on $\mathbb{R}$, and (with $N \approx \log t/2\pi$) have the same average spacing between their zeros.

Modeling $\zeta(s)$ by $\Lambda_A(z)$ is slightly more subtle: one is a function which (conjecturally) has its zeros on the line $\sigma = \frac{1}{2}$, and the other has its zeros on the circle $|z| = 1$. How to translate between those worlds? The answer is to identify the half-plane $\sigma > \frac{1}{2}$ with the interior of the unit disc $|z| < 1$. That choice makes sense for several reasons; two of the author's favorites are: $\lim_{s \to +\infty} \zeta(s) = 1$ and $\lim_{z \to 0} \Lambda_A(z) = 1$; and all zeros of $\zeta(s)$ lying on $\sigma = \frac{1}{2}$ implies [98] all zeros of $\zeta'(s)$ lie in $\sigma > \frac{1}{2}$, while all zeros of $\Lambda_A(z)$ lying on $|z| = 1$ implies (by Gauss-Lucas) all zeros of $\Lambda_A'(z)$ lie in $|z| < 1$.

The most natural mapping from $\sigma > \frac{1}{2}$ to $|z| < 1$ is $s \mapsto e^{\frac{1}{2}-s}$, as in Principle 7.8.

7.5 Analogies, and lack thereof

We have seen that, conjecturally, to leading order the zeros of the ζ-function have the same limiting local statistics as the GUE or CUE. When modeling the ζ-function at height T we use Haar-random characteristic polynomials from $U(N)$ under the matching $N \leftrightarrow \log T/2\pi$. One of three things can happen.

Case 1. Agreement to leading order. For example, the author conjectured [40]

$$\frac{1}{T} \int_0^T \frac{\zeta(\frac{1}{2}+a+it)\zeta(\frac{1}{2}+b-it)}{\zeta(\frac{1}{2}+c+it)\zeta(\frac{1}{2}+d-it)} \, dt \sim \frac{(a+d)(b+c)}{(a+b)(c+d)} - T^{-a-b}\frac{(c-a)(d-b)}{(a+b)(c+d)}, \tag{7.20}$$

for $a, b, c, d \ll 1/\log T$, as $T \to \infty$. For characteristic polynomials we have the theorem [61][27]

$$\left\langle \frac{\Lambda(e^{-a})\overline{\Lambda(e^{-b})}}{\Lambda(e^{-c})\overline{\Lambda(e^{-d})}} \right\rangle_{U(N)} \sim \frac{(a+d)(b+c)}{(a+b)(c+d)} - e^{-N(a+b)}\frac{(c-a)(d-b)}{(a+b)(c+d)}, \tag{7.21}$$

as $N \to \infty$. Those expressions fit the analogy in Principle 7.8 and are identical under $N \leftrightarrow \log T$. Both expressions can be made more precise; they have different lower order terms, with the primes playing an important role for the ζ-function. (See [33], Section 4.)

In Section 5 we described the result of Bombieri and Hejhal [20] that if a set of Z-functions satisfy RH, then a linear combination has 100% of its zeros on the critical line. The same holds for characteristic polynomials [7]: if $a_1, \ldots, a_n$ are nonzero real numbers and $A_1, \ldots, A_n$ are chosen independently and Haar-random from $U(N)$, then the expected number of zeros of $\sum a_n \mathcal{Z}_{A_n}$ on the unit circle is $N - o(N)$.

Case 2. Agreement to leading order, after inserting an arithmetic factor. This situation appears in the moments of L-functions, as illustrated in (13.2) and (7.16)

Case 3. Failure of the analogy. For L-functions, RH implies the Lindelöf Hypothesis (LH), see Conjecture 9.7. But for unitary polynomials the analogue of RH is true but the analogue of LH can fail. For example, the unitary polynomial $(z+1)^N$ achieves the value 2^N on the unit circle, which is not $\ll e^{\varepsilon N}$ for all $\varepsilon > 0$.

A less extreme example is a degree N polynomial with all zeros in half of the unit circle: it takes values as large as $\sqrt{2}^N$. Somewhat less obviously: LH fails for a degree N unitary polynomial which has no zeros in an arc of a fixed length. The relationship between large zero gaps and the size of the function is central to the main theme of this paper, and is the topic of Section 9. A related situation is key to one of our refutations, see Subsection 12.3.

How to have confidence in the predictions? Given the range of possibilities above, how can one to give credence to any specific prediction?

The 3rd case, complete failure, may seem alarming but in fact it is of little concern. The goal is to model statistically. By Principle 7.4, at height T the ζ-function is modeled by Haar-random characteristic polynomials from $U(N)$ with $N = \log(T/2\pi)$. More specifically, by Principle 7.6, e^N random matrices are chosen from $U(N)$. That may seem like a large number, but one would never expect to encounter a polynomial with all its zeros in half of the circle. Examining the CUE(N) measure, (6.6) with $\beta = 2$, we see that a polynomial with all zeros in half of the circle would have many of the terms in that product be less than half the average. Therefore such a polynomial is e^{-cN^2} times less likely than typical, so it has little chance of appearing in a sample of size e^N if N is large. The same argument applies in the case of polynomials with no zeros in an arc of fixed width.

Thus, combining both parts of the Keating-Snaith law finds that the Lindelöf Hypothesis is predicted by random matrix theory.

Cases 1 and 2 are actually the same: the arithmetic factor in the first case happens to be 1. Those arithmetic factors are well-understood, having arisen long before RMT entered the picture. In any specific case one can identify the arithmetic factor.

Note that this discussion is mostly moot, because tools like **the recipe** [26] and **the ratios conjecture** [27] directly produce expressions which include any arithmetic factors and furthermore include all main terms. In particular, Case 1 and Case 2 above are handled by the ratios conjecture and the recipe, respectively. However, RMT remains useful for the purpose of this paper because it enables us to generate examples for building intuition about $Z(t)$ for large t.

8 Waves in a unitary polynomial

As defined in (7.15), the rotated characteristic polynomial $\mathcal{Z}_A(\theta)$ for Haar-random $A \in U(N)$ is a model for $Z(t)$ for t of size e^N. If we let $N = 1000$, we obtain a model for the zeta function at height $e^{1000} \approx 10^{434}$. That height is sufficient to "see" the carrier waves in the characteristic polynomial. The analogy between L-functions and characteristic polynomials suggests that carrier waves in the ζ-function at that height should be comparable.

We will illustrate carrier waves by picking *one* large random matrix and exploring its characteristic polynomial in detail. The goal is provide intuition for carrier waves — these examples will not show extreme behavior. We will see that carrier waves occur within each individual random unitary polynomial.

8.1 An example of a carrier wave

Suppose $B \in U(1000)$ is a specific (randomly chosen) matrix, which is fixed and will be used in all the illustrations below. The function $\mathcal{Z}_B(\theta)$ is real-valued, periodic with period 2π, and in principle it can be graphed. However, that

graph would convey little useful information because 1000 wiggles across the width of a piece of paper is beyond the resolution of what can be printed or seen. Instead, we offer two other ways to graph that function and see the carrier wave.

Figure 8.1 shows graphs of $\mathcal{Z}_B(\theta)$ over two intervals of width $2\pi \times 30/1000$. Note the vertical scales in the graphs.
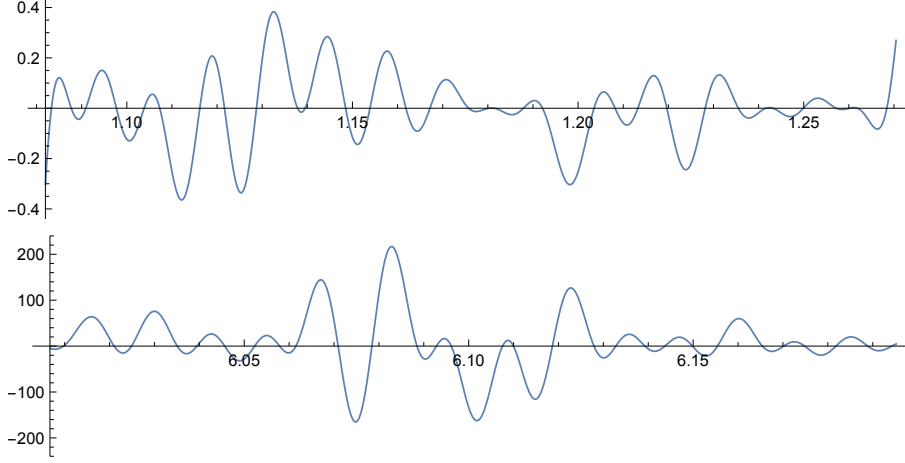

Figure 8.1 Graphs of $\mathcal{Z}_B(\theta)$ over two intervals of width $2\pi \times 0.03$.

Figure 8.1 illustrates the effect of the carrier wave. Both graphs are part of the *same* polynomial, so the global scale factors are equal. The variation in the local zero spacing has a similar effect in both regions, yet there is a factor of 500 difference in the graphs. That factor of 500 is due to the carrier wave.

We can use the idea of (5.3) to measure the carrier wave in $\mathcal{Z}_B(\theta)$, with the obvious modification of replacing $\cos(x)$ by a function with equally spaced zeros on the unit circle: $c(\theta) = z^{N/2} - z^{-N/2}$, where $z = e^{i\theta}$. At a given point $e^{i\theta_0}$, we move the nearby zeros of $c(\theta)$ to match the zeros of the polynomial, and then choose a scale factor so the functions are equal at $e^{i\theta_0}$. That scale factor is the carrier wave at $e^{i\theta_0}$. The resulting carrier wave for our example function $\mathcal{Z}_B(\theta)$, sampled at a large number of points around the circle, is shown in Figure 8.2. In the notation of (5.3), we set $K = 10$, meaning that 10 zeros on either side of a given point were used to calculate the carrier wave. Note: the vertical axis is on a logarithmic scale.
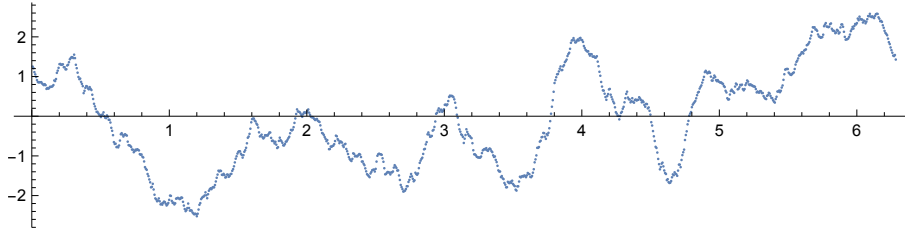
Figure 8.2 The carrier wave for $\mathcal{Z}_B(\theta)$, on a logarithmic scale, calculated using $K = 10$ in (5.3).

Figure 8.2 indicates that near $e^{1.2i}$ the polynomial is wiggling with small amplitude, and near $e^{6.1i}$ it is wiggling with large amplitude. That is how the examples were chosen for Figure 8.1.

The carrier wave is the primary factor in determining the large values, but the local spacing will still have an effect. If one focuses on the maximum value for a specific characteristic polynomial, the variation in the local spacing will

cause that maximum to occur within a larger than average gap. This can be seen in the lower plot in Figure 8.1, where the maximum near 6.083 occurs in a gap of 1.6 times the average. See Figure 5 in [50] for a detailed numerical study of this phenomenon. At $N = 1000$ the carrier wave is more important than the local variation. The polynomial in Figure 8.1 has 20 gaps which are more than twice the average, the largest of which is more than 2.45 times the average. But the local maximum in those gaps is smaller than in the 1.6 gap in the lower graph in Figure 8.1 because the carrier wave in those other regions is smaller.

8.2 The density of zeros is more important than individual gaps

A close examination of the two plots in Figure 8.1 reveals another key fact. The two plots cover equal widths, namely 30 times the average gap between zeros. But the first plot has 33 zeros, while the second plot has only 29 zeros. What we have observed is:

Principle 8.3 *The main contribution to the carrier wave is the difference between the density of nearby zeros and the average density: the local density is negatively correlated with the size of the function.*

Thus, it is the gaps between zeros which determine the size of the function, but it is collections of nearby gaps which are the dominant factor: not individual gaps.

Principle 8.3 would be more useful if we had a definition of "local density" and a way to measure it. We will develop an expression based on the gaps between widely spaced zeros.

In the top graph of Figure 8.1 the function is small over a wide region. Subsection 5.1 described the scale of the graph as arising from the zeros outside that region. Since the zeros in that region are closer together than average, one can view this as being caused by the zeros on either side pushing in towards that region, compressing those zeros. Equivalently, one expects the distance between a zero to the left of that region and a zero to the right of that region to be smaller than expected (based on the number of intervening zeros and the average gap size). If γ, γ' are zeros, and

$$\tilde{\gamma}' - \tilde{\gamma} = j + \delta(\gamma, \gamma') \tag{8.1}$$

where j is the expected normalized gap between γ and γ', then we will call $\delta(\gamma, \gamma')$ the j**th neighbor discrepancy**. Thus, when the zeros are more dense in a region, the jth neighbor discrepancy for pairs of zeros either side of that region, should tend to be negative.

To turn the neighbor discrepancy into a measure of the density, we need to determine the relative weights to give to all the neighbor gaps spanning the region. Consider a function $F(t)$ in a region centered on 0, and renumber the zeros of F as $\cdots \leq \gamma_{-2} \leq \gamma_{-1} < 0 < \gamma_1 \leq \gamma_2 \leq \cdots$.

Principle 8.4 *The logarithm of the carrier wave is given to leading order by the* **density wave**, *expressed as a weighted sum of the jth neighbor discrepancies by*

$$\sum_{j \geq J} \frac{\delta(\gamma_{-j}, \gamma_j)}{j}. \tag{8.2}$$

The sum (8.2) converges, by Dirichlet's test.

Figure 8.5 compares the density wave of zeros for the same degree 1000 polynomial used in the above examples, to the logarithm of the carrier wave

from Figure 8.2. Here the density wave is measured by (8.2) with $J = 11$. Note that the thin blue graph was calculated using 20 zeros centered around a given point, while the thicker red graph was calculated using the complementary set of 980 zeros. The agreement between those two functions, calculated using different sets of zeros, is remarkable.
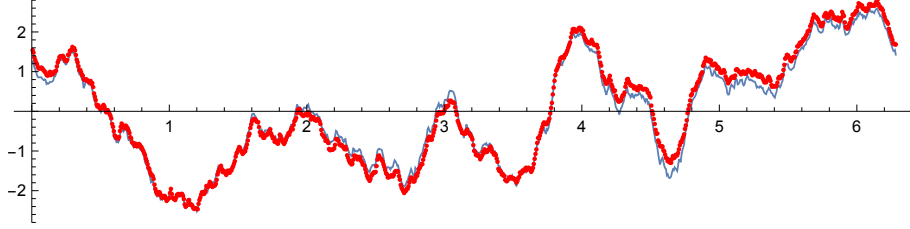

Figure 8.5 The thicker red graph is the local density of zeros (the density wave) measured by (8.2) with $J = 11$, for the degree 1000 polynomial $\mathcal{Z}_B(\theta)$. The thinner blue graph is the (logarithm of the) carrier wave for the same polynomial, calculated with $K = 10$. The similarity of those functions illustrates Principle 8.3 and Principle 8.4.

Note 8.6 The formula in Principle 8.4 depends on J, so the density of the zeros at a point is not uniquely defined; the same shortcoming applies to the local scale factor (i.e., the carrier wave) which depends on the parameter K, as observed in Note 5.4. However, those are the same shortcomings: an interval of zeros is used to determine the value of the carrier wave. If the zeros used to calculate the pressure wave are the complement of the zeros chosen to calculate the carrier wave, as in Figure 8.5, then it is perhaps not surprising in retrospect that those neighbor gaps provide an accurate measure of the density of zeros they contain, and so are highly correlated with the carrier wave.

An unresolved issue is how to identify the appropriate window. In Subsection 11.3 we suggest that the carrier wave for degree N polynomials covers around $\log N$ zeros, which is around 7 for $N = 1000$, so the size window we are using in these examples is not unreasonable. Larger examples will require a wider window.

If N is sufficiently large then the number of zeros in the interval used to calculate the wave will itself be large, suggesting some type of fractal waves-within-waves phenomenon. Directly exhibiting such behavior, even in the random matrix world, seems out of reach, although some of the ideas in Section 11 should be relevant.

Justification of Principle 8.4. We now justify (8.2). Recall that we are considering a function F with zeros $\cdots \leq \gamma_{-2} \leq \gamma_{-1} < 0 < \gamma_1 \leq \gamma_2 \leq \cdots$. Let κ be the expected density of zeros near 0. For small t we have

$$F(t/\kappa) = \prod_{j=1}^{\infty} \left(1 - \frac{t}{\tilde{\gamma}_j}\right)\left(1 - \frac{t}{\tilde{\gamma}_{-j}}\right). \tag{8.3}$$

In (8.3) write $\tilde{\gamma}_j = j \mp \frac{1}{2} + \delta_j$, where $\mp\frac{1}{2}$ has the opposite sign of j. That choice assures that δ_j is 0 on average. We have

$$F(t/\kappa) = \prod_{j=1}^{\infty} \left(1 - \frac{t}{j - \frac{1}{2} + \delta_j}\right)\left(1 - \frac{t}{-j + \frac{1}{2} + \delta_{-j}}\right) \tag{8.4}$$

$$= \prod_{j=1}^{\infty}\left(1 - \frac{t}{j - \frac{1}{2}} + \frac{t\delta_j}{(j - \frac{1}{2})^2} - \frac{t\delta_j^2}{(j - \frac{1}{2})^3} + \cdots\right) \tag{8.5}$$

34

$$\times \left(1 + \frac{t}{j - \frac{1}{2}} + \frac{t\delta_{-j}}{(j - \frac{1}{2})^2} + \frac{t\delta_{-j}^2}{(j - \frac{1}{2})^3} + \cdots \right) \tag{8.6}$$

$$= \prod_{j=1}^{\infty} \left(1 - \frac{t^2 + t(\delta_j + \delta_{-j})}{(j - \frac{1}{2})^2} + \frac{t^2(\delta_j - \delta_{-j}) - t(\delta_j^2 - \delta_{-j}^2)}{(j - \frac{1}{2})^3} + \cdots \right) \tag{8.7}$$

$$= \prod_{j=1}^{\infty} \left(1 - \frac{t^2}{(j - \frac{1}{2})^2} + \frac{t^2(\delta_j - \delta_{-j})}{(j - \frac{1}{2})^3} + \cdots \right) \tag{8.8}$$

$$= \prod_{j=1}^{\infty} \left(1 - \frac{t^2}{(j - \frac{1}{2})^2}\left(1 + \frac{\delta_j - \delta_{-j}}{j - \frac{1}{2}}\right) + \cdots \right). \tag{8.9}$$

On the next-to-last line we moved terms involving $\delta_j + \delta_{-j}$ and $\delta_j^2 - \delta_{-j}^2$ into the lower order terms because those expressions do not detect departures from the average zero spacing. Now recognize $\prod_{j \geq 1}\left(1 - t^2/(j - \frac{1}{2})^2\right)$ as $\cos(\pi t)$ and factor it out of the above expression. The terms with $j \geq J$ give the carrier wave. Taking the logarithm, expanding, and replacing $j - \frac{1}{2}$ by j, because $j \geq J$ is large, we have (8.2) because $\delta_j - \delta_{-j} = \delta(\gamma_{-j}, \gamma_j)$.

8.3 The carrier wave and $S(t)$

The relationship between zero density and the size of the carrier wave can also be expressed in terms of $S(t)$. Since $S(t)$ is the error term in the zero counting function, we see that $S(t)$ will be generally increasing in a region where the zeros are more dense than the local average, and vice-versa. By "generally" we mean when averaged across a range of several zeros, because of course $S(t)$ is decreasing wherever it is continuous, and jumps by 1 unit at each critical zero. Thus we have:

Principle 8.7 *In regions where $Z(t)$ is particularly large, $S(t)$ will on average be decreasing. In regions where $Z(t)$ is particularly small, $S(t)$ will on average be increasing.*

Figure 8.8 shows $\mathcal{S}_B(\theta)$, superimposed on the logarithm of the carrier wave, for $\mathcal{Z}_B(\theta)$.
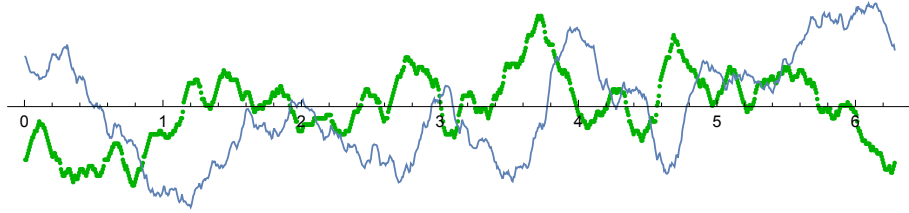
Figure 8.8 The thinner blue graph is the carrier wave of $\mathcal{Z}_B(\theta)$ from Figure 8.2 and the thicker green graph is the $\mathcal{S}_B(\theta)$ for $\mathcal{Z}_B(\theta)$. The graphs illustrate Principle 8.7: $\mathcal{S}_B(\theta)$ is increasing/decreasing in regions where $\mathcal{Z}_B(\theta)$ is particularly small/large.

Principle 8.7 explains why carrier waves are not observed in plots of $Z(t)$: $S(t)$ remains small throughout the realm accessible by computers, so there is no opportunity for it to increase or decrease enough to create a visible carrier wave.

Note that the relationship described in Principle 8.7 does not contradict Selberg's result that $\log|Z(t)|$ and $S(t)$ are statistically independent. However, it does suggest (negative) correlations in lower order terms and in the tails.

The failure to observe carrier waves computationally is also explained by Principle 8.3. We cloak this in fancy language in the next subsection.

8.4 Carrier waves are an emergent phenomenon

The conclusion to all the discussion in this section is that carrier waves are an **emergent phenomenon**. That is defined as an observable property of a system which arises from the interaction of many components of the system, but is not a property of those individual components. In the case at hand, the components are the zeros and the interaction is the repulsion between the zeros. The emergent property is variations in the density of zeros.

An emergent property cannot be understood merely by examining the components and their interactions. The study of microscopic water droplets interacting with gaseous nitrogen and oxygen does not reveal the different types of clouds, and cannot tell you that a particular cumulus cloud is shaped like a rabbit.

Emergent phenomena can only exist in a large system with many interacting components. Points made earlier in this section explain why carrier waves can only occur in high degree polynomials. By Principle 8.3, the carrier wave arises from fluctuations in the density of the zeros. The concept of density does not make sense unless one has more than just a few points. Once one has enough points to meaningfully talk about their density, those points must exist in a larger system in order to allow the local density to be significantly different than average. And that system must be yet larger in order for the repulsion between points to allow enough flexibility that significant fluctuations in density are likely. Thus, carrier waves can only exist within a large system.

The explanation above describes why carrier waves do not appear in unitary polynomials until the degree is large. But what about the ζ-function, which has infinitely many zeros? Isn't that a large system, even if one restricts to the range accessible by computers?

The explanation lies in the rigidity of the zero spacing, i.e., the slow growth of $S(t)$, interpreted as the error term in the zero counting function. Since the zeros must stay close to their expected location, there is limited opportunity to create a region where there are significantly more, or significantly fewer zeros. Not only does $S(t)$ grow slowly, it also changes sign frequently. It is as if the ζ-function not only resets itself after every interval of length 2π, there is an anti-correlation between extreme values on one interval and the next. Thus, a sequence of independent characteristic polynomials slightly overestimates the frequency of large values. In other words, the arithmetic factors used to adjust random matrix predictions, such as a_k in the $2k$th moment of the ζ-function, see (13.2), tend to be very small. This is an aspect of Principle 7.7.

8.5 A hidden choice in the Keating-Snaith law

The matching $N = \log(T/2\pi)$ is designed to allow the random characteristic polynomial $\mathcal{Z}_A(\theta)$ with $A \in U(N)$ to model $Z(t)$ for $t \approx T$. The justification is "matching the density of zeros", but that justification contains a hidden choice. If instead we had $A \in U(\kappa N)$ and considered $\mathcal{Z}_A(\theta/\kappa)$, we again would have two functions with the same average zero gap. Maybe there is a better choice than $\kappa = 1$? It is tempting to think that a larger value of κ would be better, because the standard choice imposes the absurdity $Z(t) = Z(t + 2\pi)$, not that $Z(t) = Z(t + 2\pi\kappa)$ for some $\kappa > 1$ is particularly less absurd.

Choosing a different value for κ does not work, and carrier waves are the reason. A rescaled characteristic polynomial $\mathcal{Z}_A(\theta/10)$ for $A \in U(10N)$ does not look like the concatenation of 10 different $\mathcal{Z}_A(\theta)$ for $A \in U(N)$. The rescaled polynomial will have much larger carrier waves, because the system is larger so there is more scope for the zeros to slosh back-and-forth and create regions of higher and lower density. A side-effect will be that moments of the polynomials

will not accurately predict moments of the ζ-function: the maximum from a $10\times$ larger system will typically be much larger than the maximum among 10 different smaller systems.

Thus, in some ways (such as the lack of large carrier waves at low height), the ζ-function behaves like a sequence of small systems, not a single large system. But actually it is specific properties of the large ζ-system which is the underlying reason the carrier waves start out small and are confined to short intervals. This is explained in the next section.

8.6 Spectral rigidity

Another way to see that the ζ-function cannot be modeled by a small number of large characteristic polynomials involves the analogue of what is called **spectral rigidity** [12] in physics. For the ζ-function, that terminology is just another way of saying that $S(T)$, the error term in the zero counting function $N(T)$, grows very slowly. We examine this in the context of the neighbor spacing between zeros, in particular the normalized jth nearest neighbor spacing $\tilde{\gamma}_{n+j} - \tilde{\gamma}_n$. Figure 8.9 shows the distribution of the 1st, 3rd, and 10th normalized neighbor spacings for 10000 consecutive zeros at height 10^{12}. Data taken from [80], which is the same data used in Berry's original study [12] of spectral rigidity in the context of the ζ-function.

variance = 0.177   variance = 0.266   variance = 0.341
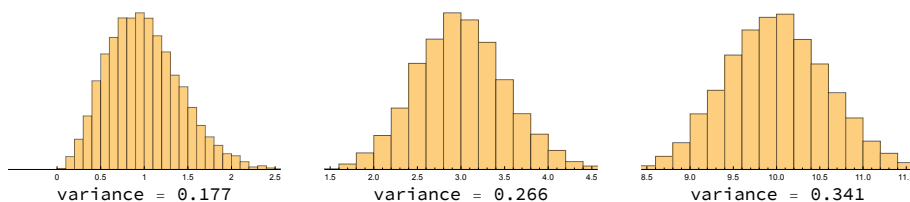
Figure 8.9 The normalized 1st, 3rd, and 10th neighbor spacing for γ_n for $10^{12} \le n < 10^{12} + 10000$, along with the variance of each sample.

The histograms in Figure 8.9 have the same horizontal scales, so one can see (even without the calculated variances) that the variance of the nth neighbor spacing is growing for $n \le 10$. However, it is clear that the variance cannot continue to grow at that rate, because the variance of the neighbor spacing is bounded by a small multiple of the variance $\langle S(T)^2 \rangle = \log \log T$. Figure 8.10 shows the variance of the jth nearest-neighbor spacing for $1 \le j \le 200$ for the same set of zeros. Note that the phenomenon illustrated in Figure 8.10 is usually expressed in terms of the **number variance** [12, 13],

$$V_T(L) = \text{Var}(N(T + L) - N(T)) = \text{Var}(S(T + L) - S(T)), \qquad (8.10)$$

which is easier to handle theoretically but perhaps more difficult to grasp intuitively.
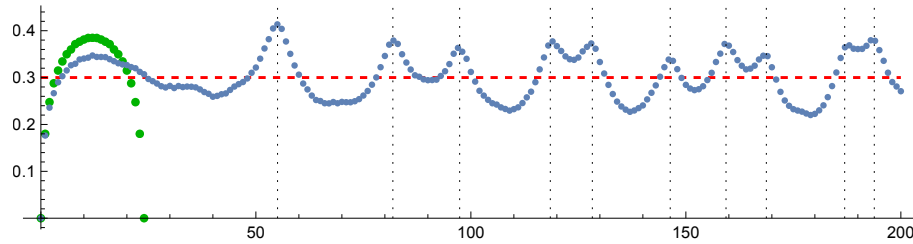
Figure 8.10 The blue dots show the variance of the jth nearest-neighbor spacing, $0 \le j \le 200$, for zeros γ_n with $10^{12} \le n < 10^{12} + 10000$. The height of the horizontal red dashed line is twice the average variance of the samples, which would be the variance of the neighbor gaps if the zeros were distributed independently. The vertical dotted lines are at $c\,\gamma_k$, with $c \approx 3.894$ and γ_j the height of the jth nontrivial zero of the ζ-function. The green dots are the variance of the jth nearest-neighbor spacing, $0 \le j \le 24$, for eigenvalues of random matrices in $U(24)$.

Note that the average value of $\tilde{\gamma}_n - n$ is $\frac{1}{2}$, and in the sample used in the above figures, has variance 0.15005. If the jth nearest neighbors were distributed independently, then the jth nearest neighbor gaps would have variance 2×0.15005, which is the horizontal line in Figure 8.10. Thus, the relative displacement of neighboring zeros (relative to their expected location) is sometimes positively correlated, and sometimes negatively correlated, across a span of many zeros.

The peaks in Figure 8.10 occur near the rescaled zeros of the ζ-function, where the scale factor is the density of the zeros at the height of the sample (in this example, $t \approx 2.67 \times 10^{11}$). This **resurgence** of the first few zeros occurs in many zero statistics [13, 33], typically manifested via the influence of the zeros on the size of $\zeta(1 + it)$.

The universality of Principle 7.1 is illustrated in Figure 8.10 by the approximate agreement, for very small gaps, between the ζ-zeros and eigenvalues variances. This is referred to as the **universal regime**. In the universal regime, no details about the ζ-function are relevant to the leading-order behavior. This idea appears again in Principle 8.11. In the non-universal regime, the prime numbers (equivalently, the low-lying zeros) have a strong effect. Only recently has this effect been treated rigorously [74].

The rigidity in the zeros spacing, equivalently the slow growth of $S(t)$, prevents the zeros from moving far from their expected location, which prevents large regions with a high or low density of zeros, which prevents large carrier waves at low height.

8.7 Extreme gaps

In Subsection 7.2 we described the GUE predictions for the largest and smallest gaps between zeros. We now elaborate on how those predictions are made, and indicate some issues which impact those predictions.

Small gaps. Subsection 7.1 mentioned the universality of the β-ensembles: the local statistics of the eigenvalues only depend (to leading order) on the degree of repulsion. This has implications for the neighbor spacing. From the CUE measure, (6.6) with $\beta = 2$, it follows immediately that the nearest-neighbor spacing vanishes to order 2, and (by integrating θ_j from θ_{j-1} to θ_{j+1}) the next-nearest-neighbor spacing vanishes to order 7. Thus, to leading order the pair correlation for the CUE, (7.8), describes the distribution of small gaps

between eigenvalues, and that prediction is not sensitive to any details about the particular system being discussed. That is,

Principle 8.11 *The PDF $p_2(x)$ of the normalized gaps between zeros of the ζ-function, or any other L-function, decays like $(\pi^2/3)x^2$ as $x \to 0$. That prediction is very likely to be accurate, with the prime numbers and other details particular to any specific L-function having little influence on the smallest gaps.*

Any skepticism about Principle 8.11 should be alleviated by Figure 8.12, which shows the distribution of the smallest normalized gaps among the first 10^{13} zeros of the ζ-function, compared to the $(\pi^2/3)x^2$ prediction. Note that Figure 8.12 concerns the universal regime, so there is no contradiction with our repeated assertion that numerical calculations of the ζ-function can give a misleading impression.
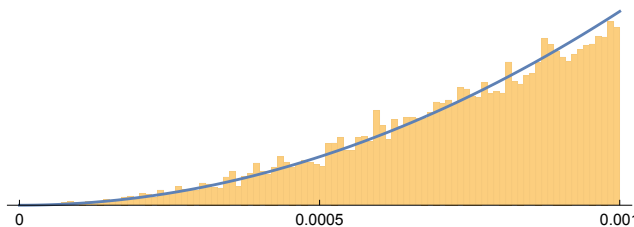

Figure 8.12 Histogram of the 9333 smallest normalized neighbor gaps among the first 10^{13} zeros of the ζ-function, compared to the predicted density $\frac{\pi^2}{3}x^2$. Data from Xavier Gourdon [53].

To estimate the size of the normalized smallest gaps among a sample of X gaps, calculate the gap size which has a 50% chance of appearing:

$$\text{solve} \quad X \int_0^{\xi_{min}} Ax^2 \, dx = \frac{1}{2} \quad \text{to get} \quad \xi_{min} = \left(\frac{AX}{3}\right)^{-\frac{1}{3}}, \qquad (8.11)$$

where $A = \frac{\pi^2}{3}$. So, among the gaps up to height T, one expects the smallest normalized gaps to be of size $T^{-\frac{1}{3}+o(1)}$. One can try to be more precise, but there probably is no meaning in the specific constant or power of $\log T$. There is meaning, however, to the exponent $-\frac{1}{3}$. A small gap between two zeros forces a small value for the derivative of the ζ-function at both of those zeros. An implication [62] is that the predicted discrete mean values

$$\frac{1}{N(T)} \sum_{0 < \gamma_j \leq T} |\zeta'(\tfrac{1}{2} + i\gamma_j)|^{2k} \sim c_k T^{2k(k+2)}, \qquad (8.12)$$

conjectured to hold for $k > -\frac{3}{2}$, cannot hold for $k < -\frac{3}{2}$ because (due to the smallest neighbor gaps) there are individual terms which are very large.

Large gaps. The situation for large gaps is likely to be significantly more complicated.

For large random unitary matrices we have $p_2(x) \sim e^{-\pi^2 x^2/8}$ as $x \to \infty$. For an individual matrix, the neighbor gaps are not independent because adjacent neighbor gaps are negatively correlated, and also they average to exactly 1. But, if one pretends those gaps are independent, then using the tail of p_2 the expected maximum among M normalized gaps would be $\sqrt{8 \log M}/\pi$. Setting $M = N$ gives the correct answer for the expected maximum among the gaps of an individual random matrix [8]. Thus, there is some justification to using the tail of the CUE neighbor spacing to conjecture the maximum gap over longer

ranges, and Figure 1 in [8] shows that such a prediction is supported by data involving 2×10^9 zeros of the ζ-function.

As illustrated in Figure 8.10, for more widely spaced zeros, where "widely" might only mean a few times the average zero spacing, the distribution of nth neighbor gaps for the ζ-function departs significantly from that of random unitary matrices: the ζ gaps are more constrained. Therefore, it would not be surprising if the largest normalized gaps between zeros of the ζ-function were smaller than the analogous normalized gaps between eigenvalues of random unitary matrices. It particular, it would not be surprising if the largest normalized neighbor gaps of the ζ-function (either as conjectured in [8], or from a naive interpretation of the tail of $p_2(x)$) were actually smaller than $\sqrt{8 \log T}/\pi$. See Principle 7.7.

Resolving such an issue with data is difficult because the distribution of large gaps decays rapidly. In this paper we only make use of the fact that the largest normalized gaps are likely to be $O(\sqrt{\log T})$.

9 Extreme values

So far we have considered typical large and small values. That is, values within the bulk of the distribution of $\log |Z(t)|$. The conclusion was that those large values are due to the carrier wave. Since the carrier wave arises from variations in the local density of zeros, once the wave is large it must stay large over a span of several zeros. Thus, the graph of the Z-function when it is particularly large looks the same as it does anywhere else — if you ignore the scales on the axes.

However, we have not yet ruled out the possibility that even larger values might arise from very rare events, such as an extremely large gap between zeros. The effect might look similar to what appears in the data of Bober and Hiary [17], as shown in Figure 3.4. We will explain why the largest neighbor gaps cannot cause the function to reach the largest values. First a trivial observation, following immediately from Principle 3.5 and (7.13):

Principle 9.1 *The largest gaps between zeros contribute* $O(\sqrt{\log t})$ *to* $S(t)$.

It takes more effort to establish the analogous result for $Z(t)$:

Principle 9.2 *The largest gaps between zeros contribute* $O(\sqrt{\log t})$ *to* $\log |Z(t)|$.

Since the largest values of $\log |Z(t)|$ are conjectured to be $\gg \sqrt{\log(t) \log \log(t)}$, Principle 9.2 tells us that the largest neighbor gaps are not responsible for the extreme values.

By Principle 8.4, a weighted sum of neighbor gaps (the density wave) is the leading factor in the size of $\log |Z(t)|$. It would be interesting to analyze that weighted sum to possibly give another way of predicting the extreme values. Because of spectral rigidity, Subsection 8.6, each neighbor discrepancy $\delta(\gamma_{-j}, \gamma_j)$ has a similar distribution. Can one recover the conjectured extreme values of $\log |Z(t)|$? When that sum is large, is it due to a handful of very large terms, or a large number of moderately sized terms? Analysis will be complicated by the fact that $\delta(\gamma_{-j}, \gamma_j)$ and $\delta(\gamma_{-j-k}, \gamma_{j+k})$ are positively correlated if k is small.

Principle 9.2 follows immediately from (7.13), see also Subsection 8.7, combined with:

Principle 9.3 *The size of $Z(t)$ at a point comes from two independent sources: the carrier wave, and the local arrangement of zeros. An individual large gap of g times the average zero spacing contributes $O(g)$ to the local maximum of $\log |Z(t)|$ within that large gap.*

Note that Principle 9.3 and Principle 8.4 both indicate that when zeros move apart by g units on the scale of the average zero spacing, the effect is a factor $e^{O(g)}$ on the size of the function. However, the justification in the case of a large nearest neighbor gap is quite different than the case of neighbor discrepancy for distant gaps. We have not been successful at handling the two cases in a unified way.

Subsection 9.1 is devoted to justifying Principle 9.3.

9.1 Zeros near a large gap

To understand the effect of a large gap on the size of the function, we need information about the gaps adjacent to the large gap. From Haar measure (6.6) on $U(N)$ we see that the expected size of the immediately adjacent gaps is smaller than average. Bober and Hiary [17] expressed it as: $S(t)$ tends to be increasing immediately before and immediately after a large gap. To isolate the effect of only the single large gap, we propose to focus on the case where *the nearby zeros are in their most likely configuration.* We will do this for zeros on the unit circle.

Determining the most likely configuration is an easy computer experiment. Begin by fixing two zeros with a chosen gap and distributing the other $N-2$ zeros on the circle anywhere outside that gap (equally spaced is a perfectly good starting configuration). Then repeatedly perturb the $N-2$ zeros to increase the measure (6.6), until the process stabilizes. The result will be a good approximation to the most likely configuration.

With $N = 74$ and a gap of 6 times the average, which is the random matrix analogue of $Z(t)$ in Figure 3.4, the most likely configuration is shown in Figure 9.4.
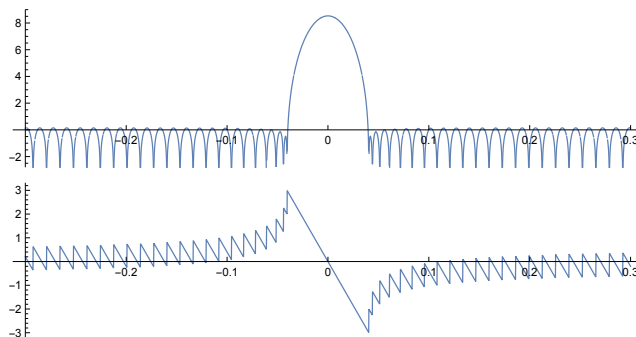
Figure 9.4 Graphs of $\mathcal{Z}(\theta)$ and $\mathcal{S}(\theta)$ for a degree 74 polynomial with a zero gap of 6 times the average and all other zeros in their most likely configuration with respect to Haar measure (6.6).

A similarity with Figure 3.4 is the general shape of $S(t)$ near the large gap. A difference is that the maximum in Figure 9.4 is smaller, which is due to the fact that the immediate neighbor gaps in Figure 3.4 are wider than the most likely configuration. The wider neighboring gaps in Figure 3.4 are not surprising, because the large gap in Figure 9.4 is not as large as possible in that range. That means there is room for more variation in the nearby zeros, so less likely configurations can occur.

We see that a large gap causes nearby gaps to be smaller. The key issue for justifying Principle 9.3 is: how far does that influence extend? That is, at what distance from the large gap do the neighbor gaps (approximately) return to their average? The answer is easy to express if we measure on the scale of the average gap between zeros.

Principle 9.5 *The normalized range of influence of a large gap is proportional to the normalized width of the gap.*

So, doubling the width of a large gap will double the range of its influence on nearby neighbor gaps.

To see why Principle 9.5 is true, consider the following thought experiment. Fix a large gap, and then distribute the other gaps in their most likely configuration. It is helpful to think of the zeros as particles with equal charge, obeying a force law such that the particles are in equilibrium when they are in their most likely configuration. Now double all the charges. The particles will still be in equilibrium. Then split every doubled charge into two equal charges, leave one in place, and move the other halfway toward the neighboring charge, with the movement away from the large gap. The new configuration will not be in equilibrium, but it will be nearly so. As the system settles into equilibrium (holding fixed the two particles forming the large gap), the particles will move very little, because they were already close to equilibrium. (The reader is invited to persuade themself that these equilibrium configurations are stable. Also, as the system settles back into equilibrium, the slight movement of the particles brings them closer to the large gap.) The result is a new system where the large gap is the same on an absolute scale, as is the range of influence of that large gap. But in the new system, the large gap and its influence are twice as large on the scale of the average spacing. Thus, Principle 9.5.

Figure 9.6 provides another way to understand Principle 9.5. For 230 zeros on the circle, which corresponds to $Z(t)$ near 10^{100}, we used the procedure described at the start of this section to determine the most likely configuration with a gap of 5, 10, or 15 times the average. The first plot in Figure 9.6 shows the successive normalized neighbor gaps in each of the three cases. As expected, the immediate neighbor gap is small, and is smaller when the large gap is bigger. When the large gap is 5 times the average, the immediate neighbor gap is approximately 0.32 times the average. When the large gap is 15 times the average, the immediate neighbor gap is approximately 0.1 times the average. The gaps grow, approaching 1, the normalized average gap, when farther away from the large gap. And as expected, when the large gap is larger, it takes longer for the neighbor gaps to approach 1.

The second plot in Figure 9.6 is made from the same data, but this time the horizontal axis is scaled by the size of the large gap. We see that the three graphs are virtually indistinguishable. In other words, measured on the scale of the large gap, the decay of the influence of that gap is independent of the size of the gap. By 3 times the width of the large gap, there is very little influence on the neighbor gaps.
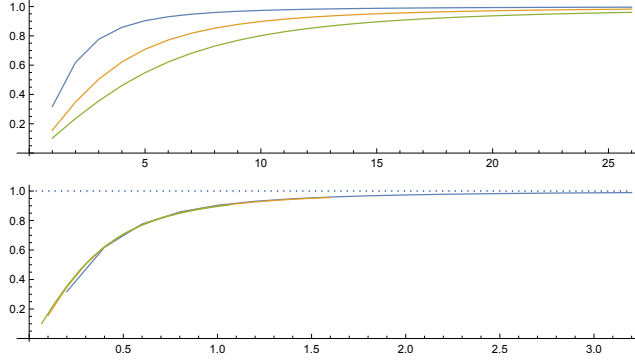
Figure 9.6 The typical normalized gap between successive zeros adjacent to a large gap of 5, 10, or 15 times the average spacing. In the plot on the top the horizontal scale is the index of the gap. On the bottom the horizontal scale is the index divided by the size of the large gap.

We use Principle 9.5 to justify Principle 9.3. The idea is similar to (5.3). Start with a function having equally spaced zeros, and then move some of those zeros to a new location. In this case it is easiest to start with $f_0(z) = z^N + 1$ and then move the zeros with $-\frac{\pi}{N}g < \theta_j < \frac{\pi}{N}g$ to create a normalized gap of size g, placing those zeros in their most likely location. Denote the new value of θ_j by $\hat{\theta}_j$, and call the resulting function $f_g(z)$. We have

$$f_g(z) = (z^N + 1) \prod_{-\pi g \leq j \leq \pi g} \frac{1 - e(\hat{\theta}_j)}{1 - e(\theta_j)}.$$

By Principle 9.5, the zeros which were moved will have $|\hat{\theta}_j| \leq Kg/N$ where K is some absolute constant, independent of g and N. Therefore

$$\begin{aligned}
f_g(1) &\ll \prod_{-\pi g \leq j \leq \pi g} \frac{\frac{2\pi Kg}{N}}{\frac{2\pi(2j-1)}{2N}} \\
&\ll \frac{K^{2g}g^{2\pi g}}{(\pi g)!^2} \\
&= e^{O(g)},
\end{aligned}$$

the last step following from Stirling's formula. That establishes Principle 9.3.

We have seen that a large gap has essentially the same effect on $S(t)$ and on $\log|Z(t)|$. Yet, it was almost trivial to handle $S(t)$, and fairly complicated to handle $\log|Z(t)|$. Why is it simple to understand the effect of large gaps on the imaginary part of $\log\zeta(\frac{1}{2} + it)$ but not on the real part? Is there a simple way to deal with both cases? Perhaps not: the discussion at the end of Section 1.2 of [74] describes some essential differences between $S(t)$ and $\log|\zeta(\frac{1}{2} + it)|$, the most important being that $\log|\zeta(\frac{1}{2} + it)|$ is unbounded in a neighborhood of its discontinuities, which "cause technical difficulties".

9.2 More intuition for the need for carrier waves

We describe yet another way to see that carrier waves are primarily responsible for the size of $Z(t)$.

Suppose the typical large value of $Z(t)$ were primarily due to large gaps, such as in Figure 3.4. Combining Selberg's theorem on the distribution of $\log|Z(t)|$ with Principle 9.3 would imply that a positive proportion of t occur in a gap of relative size $\sqrt{\log\log(t)}$. In other words, the probability that a gap

43

is larger than $\sqrt{\log\log(t)}$ times the average, must be $\gg 1/\sqrt{\log\log(t)}$. This would correspond to the distribution of nearest-neighbor spacings decreasing much more slowly than conjectured. That is contrary to our observations and not consistent with the conjectures on neighbor spacing arising from RMT.

The situation becomes more clear when we consider small values. Obviously the zeros on the critical line cause very small (i.e.., large negative) values for $\log|Z(t)|$. It is not difficult to see (consider $\log|\sin x|$, for example) that the neighborhoods of the zeros cannot account for $\log|Z(t)|$ being $\ll -\sqrt{\log\log(t)}$ for a positive proportion of t. Therefore the typical small values must arise from the function staying small between neighboring zeros.

Suppose two neighbor zeros are separated by g times the average spacing, with g very small. By considering $x(g-x)$ or $\sin(x)-1+g^2$, one expects the local maximum between those zeros to scale as g^2. Thus, if the local zero spacing is the main cause of those small values, then a positive proportion of t must occur within a gap of size $o(1)$ times the local average. That is absurd because it requires more than 100% of the gaps to be very small! Thus, the large proportion of small values required by Selberg's result must come from a conspiracy involving many gaps, with most of those gaps not particularly small.

9.3 The Lindelöf Hypothesis and self-reciprocal polynomials

As mentioned in Subsection 7.5, an apparent failure in the analogy between the ζ-function and random unitary polynomials is the Lindelöf Hypothesis:

Conjecture 9.7 The Lindelöf Hypothesis (LH). *If $\varepsilon > 0$ then*

$$Z(t) = O_\varepsilon(t^\varepsilon) \tag{9.1}$$

as $t \to \infty$, or in other notation,

$$Z(t) = e^{o(\log t)}. \tag{9.2}$$

The analogue for a unitary matrix $A \in U(N)$ is

$$\max_{|z|=1} \mathcal{Z}_A(z) = e^{o(N)}. \tag{9.3}$$

That estimate need not hold: if all of the eigenvalues of A are in $\Re z \geq 0$ then $|\mathcal{Z}_A(-1)| \geq \sqrt{2}^N$.

We suggest that when one is discussing LH (therefore implicitly not assuming RH because RH implies LH), the comparison should be with self-reciprocal polynomials — which are not necessarily unitary. With that in mind, consider Backlund's [5] equivalence (see [101] Section 13.5). Let

$$N(\sigma, T) = \#\{\rho = \beta + i\gamma \ : \ \zeta(\rho) = 0, \ \beta > \sigma, \ 0 < t \leq T\}. \tag{9.4}$$

Theorem 9.8 Backlund [5]. *The Lindelöf Hypothesis true if and only if*

$$N(\sigma, T+1) - N(\sigma, T) = o_\sigma(\log T) \tag{9.5}$$

for all $\sigma > \frac{1}{2}$.

Backlund's equivalence (9.5) says that for any fixed $A > 0$ and any fixed-width strip around the critical line, a negligible proportion of the zeros with $T \leq \gamma \leq T + A$ lie outside that strip.

One of the implications in Theorem 9.8 holds for self-reciprocal polynomials. If P is a polynomial, let

$$N(\delta, P) = \#\{\rho = re^{i\theta} \ : \ P(\rho) = 0, \ r < e^{-\delta}\}. \tag{9.6}$$

Proposition 9.9 *Suppose P is a self-reciprocal polynomial with*

$$\max_{|z|=1} |P(z)| = e^{o(\deg P)}. \tag{9.7}$$

Then

$$N(\delta, P) = o_\delta(\deg P) \tag{9.8}$$

for all $\delta > 0$. In other words, a negligible proportion of the zeros lie outside any fixed-width annulus around $|z| = 1$.

Proof. Suppose $\delta > 0$ and factor P as

$$P(z) = Q_\delta(z)Q^\delta(z) \tag{9.9}$$

where the zeros of Q_δ are the zeros of P in the annulus $e^{-\delta} < |z| < e^{\delta}$. Suppose the degree of Q^δ is $2M$. The maximum of $Q^\delta(z)$ on $|z| = 1$ is decreased if the zeros of Q^δ are moved to have equally spaced arguments and all have absolute value $e^{-\delta}$ or e^{δ}. Thus, it is sufficient to consider the case

$$Q^\delta(z) = (z^M - e^{-\delta M})(z^M - e^{\delta M}) = z^{2M} - (e^{-\delta M} + e^{\delta M})z^M + 1. \tag{9.10}$$

By (9.10) we have $|Q^\delta(z)| \geq e^{\delta M} - 3$ on all of $|z| = 1$. Since $Q_\delta(z)$ cannot be small on all of $|z| = 1$ (because it is 1 on average), by (9.7) we have $M = o(\deg P)$, as claimed. ∎

The converse of Proposition 9.9 is not true, and adapting the above proof fails because $Q_\delta(z)$ in (9.9) could be large on $|z| = 1$. Thus:

Principle 9.10 *There are two different ways a self-reciprocal polynomial can be* very large *on $|z| = 1$: there may be a large number of zeros far from the unit circle, or there may be a large gap (or other irregularity) in the arguments of the zeros.*

For the ζ-function, the second option does not occur, because the slow growth of $S(t)$, equivalently the small error term in the zero counting function $N(t)$, prevents extreme irregularities in the distribution of zeros.

By "very large" we mean a violation of (the polynomial analogue of) LH.

In Subsection 12.3 we will suggest that one of the claimed arguments against RH is actually an argument *for* the possibility of improving bounds on the size of extreme gaps between zeros of the ζ-function, based on bounds for the size of $Z(t)$ and its derivatives.

9.4 Carrier waves and mollifiers

A **mollifier** is a Dirichlet polynomial $M(s)$ which is, in a sense we will make precise, an approximation to $1/\zeta(s)$. In many applications, such as Levinson's method [73, 24], one considers mean values like

$$I(M, T) = \int_0^T |\zeta(\tfrac{1}{2} + it)M(\tfrac{1}{2} + it)|^2 dt \tag{9.11}$$

where M is a Dirichlet polynomial of length T^θ. One would like $I(M, T) \sim cT$ for some $c > 0$. If $M = 1$ then $I(1, T) \sim cT \log T$, so we see that a mollifier

must be small where the ζ-function is large. That is, $M(s)$ "mollifies" the wild behavior of the ζ-function.

It is clear that an approximation to $1/\zeta(s)$ is a good candidate for a mollifier. We will use the idea of carrier waves to motivate the use of more complicated mollifiers. These will be based on (approximations to the reciprocal of) linear combinations of the ζ-function and its derivatives.

Rephrasing the main idea of carrier waves, see Subsection 5.1, for most X, if t is small we have

$$\zeta(X+t) \approx e^{A(X)} P_X\Big(t\,\frac{\log X}{2\pi}\Big). \tag{9.12}$$

Here $A(X)$ is normally distributed and changes slowly, $P_X(x)$ is a polynomial with zeros that are spaced 1 apart on average, and the distribution of P_X changes slowly with X. In particular, $A(X)$ is responsible for the size of the ζ-function, and it is the reason $I(1,T)$ is of size $T \log T$. Thus, the mollifier must counteract the $e^{A(X)}$ factor, with P_X and the zeros from P_X playing a secondary role.

Because $A(X)$ changes slowly and the distribution of P_X also changes slowly, we have that the kth derivative of the ζ-function is the same size as the ζ-function, times a factor of $\log^k X$:

$$\zeta^{(k)}(X+t) \approx e^{A(X)} \Big(\frac{\log X}{2\pi}\Big)^k P_X^{(k)}\Big(t\,\frac{\log X}{2\pi}\Big). \tag{9.13}$$

Thus, for any real numbers $a_0, \ldots, a_K$, and with $L = \log T/2\pi$, a Dirichlet polynomial approximation to the reciprocal of

$$\sum_{k=0}^{K} a_k \frac{1}{L^k} \zeta^{(k)}(s) \tag{9.14}$$

is a viable mollifier for the ζ-function, and it is also a viable mollifier for any expression of the form (9.14). This idea lies behind various improvements to Levinson's method [48, 83].

10 The primes

When describing the relationship between different objects, it is common to use anthropomorphic language, or to speak in terms of cause and effect. For example, one of the principles below is "the carrier wave arises from the prime numbers". The intent is not to describe a causal relationship, but rather to provide intuition — intuition which possibly may eventually lead one to new insight or a new proof.

A previous principle asserted that the carrier wave arises from fluctuations in the density of zeros. Have we contradicted ourself, and if so, which assertion is "correct"?

Both principles are valid because both help provide intuition for a mental model for the behavior of the zeta-function. Equalities like

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1} = \frac{e^{(\log(2\pi)-1-\frac{\gamma}{2})s}}{2(s-1)\Gamma(1+\frac{s}{2})} \prod_\rho (1 - s/\rho)e^{s/\rho}, \tag{10.1}$$

where the first product is over the primes and the second is over the zeros, suggest that it is not meaningful to ask whether the primes cause the zeros or the zeros cause the primes. But if one has knowledge, or intuition, about one

of those objects, then it is reasonable to ask what that says about the other. This is where analogies between different areas help further the subject. In the previous parts of this paper we primarily used knowledge and conjecture about the zeros as a starting point to understand the ζ-function. In this section and in Section 11 we take the primes as the starting point.

10.1 Where are the primes?

The Riemann ζ-function and its zeros were introduced because of their connection to the prime numbers, yet the previous several sections of this paper had little mention of the primes.

So, where are the primes?

Principle 10.1 *The carrier wave arises from the prime numbers.*

That principle does not contradict Principle 8.3, which says that the carrier wave arises from the relative density of the zeros. Both principles have explanatory power. Which one is more illuminating will depend on the particular question at hand.

We partially justify Principle 10.1 by Selberg's theorem (3.15) that $\log |\zeta(\frac{1}{2} + it)|$ has a Gaussian distribution. The starting point of the proof is to write

$$\log |\zeta(s)| = \sum_{\text{small primes } p} \Re \frac{1}{p^s} + \sum_{\text{zeros } \rho \text{ near } s} \log |s - \rho| + \text{error terms.} \quad (10.2)$$

The hard work is showing that, most of the time, the sum over zeros and the error terms are of size $O(1)$. Thus, only the primes contribute to the size of the ζ-function. One should expect a central limit theorem to hold for the sum over primes, because the $\log p$ are linearly independent over the rationals. Thus, the variance should be $\sum_{p \leq x} p^{-1} \sim \log \log x$ which is larger than the size of the error terms.

Selberg's theorem is not an adequate justification for Principle 10.1, because by definition a carrier wave must stay large over the span of several zeros. That does not follow from the proof sketched above. In Section 11 we describe more recent work which extracts additional information about the randomness of the ζ-function from the primes.

Principle 5.1 described the behavior of $Z(t)$ as arising from a combination of local and global factors. We rephrase that in light of Principle 10.1 (also see Subsection 8.6).

Principle 10.2 *The short range behavior of the zeros is universal and follows random matrix statistics. The long range behavior of the zeros is dictated by the prime numbers.*

In the next subsection we quantify what is meant by "short range" and "long range" above.

10.2 The hybrid model

The ζ-function can be expressed in terms of only zeros, and in terms of only primes. Here we describe the **hybrid model**, in which one can adjust the relative contribution of the zeros and the primes.

Gonek, Hughes, and Keating[52] proved that if $s = \sigma + it$, with $0 \leq \sigma \leq 1$ and $|t| \geq 2$, then for $X > 2$ and K any positive integer,

$$\zeta(s) = P_X(s) Z_X(s) \left(1 + O\left(\frac{X^{2-\sigma+K}}{(|t| \log X)^K} \right) + O(X^{-\sigma} \log X) \right), \quad (10.3)$$

where

$$P_X(s) := \exp\left(\sum_{n \leq X} \frac{\Lambda(n)}{n^s \log n}\right), \tag{10.4}$$

Here $\Lambda(n)$ is the von-Mangoldt function, which is supported on prime powers, and

$$Z_X(s) := \exp\left(-\sum_\rho U((s-\rho)\log X)\right). \tag{10.5}$$

Here the ρ are non-trivial zeros of $\zeta(s)$ and $U(z) = \int_0^\infty u(x)E_1(z\log x)\ dx$, where $E_1(z) = \int_z^\infty \frac{e^{-w}}{w}\ dw$ is the exponential integral and u is any smooth function supported in $[e^{1-1/X}, e]$.

The parameter X controls the relative influence of the primes and the zeros. If X is large, there are many primes in $P_X(s)$, and only the zeros very close to s are relevant to the product in $Z_X(s)$, while if X is small, the zeros further away from s make a contribution to $Z_X(s)$, but the number of primes in $P_X(s)$ is diminished.

One might expect Z_X and P_X to behave somewhat independently, and there are two bits of evidence to support that idea. First, Gonek, Hughes, and Keating [52] propose the "splitting conjecture": if $X \to \infty$ with $X = O(\log^{2-\epsilon} T)$ then the $2k$th moment of $\zeta(\frac{1}{2}+it)$ splits as the product of the $2k$th moment of $P_X(\frac{1}{2}+it)$ and $Z_X(\frac{1}{2}+it)$ separately. They rigorously compute the moments of $P_X(\frac{1}{2}+it)$, obtaining the expected arithmetic factor.

Second: Farmer, Gonek, and Hughes [43] calculate the (conjectural) extreme value of the product $P_X(\frac{1}{2}+it)Z_X(\frac{1}{2}+it)$ for $T \leq t \leq 2T$, using a random model for each piece separately. They find that, for a wide range of X, the expected maximum of the product is independent of X, even though the maximum each piece separately depends on both X and T.

Both of the above results concern the leading order behavior, and there seems to be no reason to believe that any type of random model for the ζ-function can accurately predict lower order behavior. However, recent work of Sawin [93] concerns a random matrix model in the function field case, in which the lower order terms of the moments have a structure similar to the conjectured moments of the ζ-function.

11 Randomness not involving the zeros

In this section we briefly describe how expressions like

$$\log\zeta(\tfrac{1}{2}+it) \quad\leftrightarrow\quad \sum_p \frac{1}{\sqrt{p}}p^{-it} \tag{11.1}$$

can be used to understand the statistical behavior of the ζ-function on the critical line. To a traditional analytic number theorist, quantities like (11.1) hurt the eyes and churn the stomach. But to a physicist or a probabilist, such expressions are a meaningful starting point for establishing interesting results.

It has been noted that characteristic polynomials do not know about the primes. However, in the characteristic polynomial world there is an analogous expression [62] to (11.1) for the characteristic polynomial of a matrix $A \in U(N)$:

$$\log\Lambda_A(e^{i\theta}) = \sum_n \frac{1}{\sqrt{n}}\frac{\text{Tr }A^n}{\sqrt{n}}e^{i\theta}. \tag{11.2}$$

It is a theorem [34] that for $1 \leq n \leq k$, with k fixed, as $N \to \infty$ the $\operatorname{Tr} A^n / \sqrt{n}$ are i.i.d Gaussian.

In both worlds we have a weighted sum of i.i.d. random variables, so we would like to know their variance. For the L-functions, the variance involves $\sum_p p^{-1}$, which diverges like $\log \log T$. For the characteristic polynomials we have $\sum_n n^{-1}$, which diverges like $\log N$. The Keating-Snaith Law $N = \log T$ is a clue that we should expect similar types of randomness in the two worlds.

The perspective we take here is a metric space X, a circle or line our context, with each $x \in X$ having an associated random (generalized) function F_x. In other words, a random field. Of particular concern will be the relationship between the random functions at different points. If the covariance satisfies

$$\langle F_x, F_{x'} \rangle = -\log |x - x'| + g(x, x') \tag{11.3}$$

where g is smooth, we say that the field is **log-correlated**.

To get into the spirit of this section, we encourage the reader to heuristically manipulate (11.1) to show that $\log \zeta(\frac{1}{2} + it) \log \zeta(\frac{1}{2} - is) \approx -\log |t - s|$, therefore $\log \zeta(\frac{1}{2} + it)$ is log-correlated. Adapted from the Appendix in [49], the key steps are to argue that only the diagonal terms contribute, and then to recognize the resulting sum as $\log \zeta(1 + i(t - s))$.

For details about the topics below, see [6, 91, 92, 49].

11.1 Log-correlation and short-range maximum values

Fyodorov, Hiary, and Keating [51], and Fyodorov and Keating [49], considered the maximum values of the ζ-function, not the global maxima, but over intervals of bounded length:

$$\zeta_{\max}(L; T) = \max_{T \leq t \leq T+L} \log |\zeta(\tfrac{1}{2} + it)|. \tag{11.4}$$

For comparison with characteristic polynomials, the interesting case is $L = 2\pi$.

The log-correlation of $\zeta(\frac{1}{2} + it)$, in other words the fact that the ζ-function at nearby points are not independent, causes the maximum values to be slightly smaller than one might expect.

The FHK conjecture [51] is

$$\zeta_{\max}(2\pi; T) \sim \log \log(T/2\pi) - C \log \log \log(T/2\pi) + X_T, \tag{11.5}$$

where

- $C = \frac{3}{4}$,

- X_T is a random variable which has a limiting distribution, X, as $T \to \infty$, and

- The PDF of X decays like xe^{-x} as $x \to \infty$.

Upper and lower bounds of size (11.5) are established in [1, 57] and [2], respectively.

Note that (11.5) is more than just a conjecture for the expected maximum value on an interval: it is a conjecture for the distribution of the maxima on such intervals.

The FHK conjecture relates to several areas of mathematical physics, and has an analogous statement for the CUE(N) under the usual identification $N = \log(T/2\pi)$. See Section 8 of [97] and Section 2 of [6] and references therein. We limit our discussion to the relationship between the FHK conjecture (11.5) and carrier waves.

Why the FHK conjecture is surprising. Much of the discussion below follows Section 8 of Soundararajan [97].

Consider the random matrix analogue of $\zeta_{\max}(2\pi; T)$:

$$\mathcal{Z}_{\max}(U) = \max_{\theta \in [0, 2\pi]} \log |\mathcal{Z}_U(e^{i\theta})|. \tag{11.6}$$

One can view $\mathcal{Z}_{\max}(U)$ as arising from M random θ samples of the random variable $\log |\mathcal{Z}_U(e^{i\theta})|$, for some sample size M. If those samples were independent, and we use the fact that $\log |Z_U(e^{i\theta})|$ has a Gaussian distribution with variance $\sqrt{\frac{1}{2} \log N}$ (Keating and Snaith's analogue of Selberg's theorem), then $\mathcal{Z}_{\max}(U)$ should typically be around

$$\sqrt{\tfrac{1}{2} \log N} \left(\sqrt{2 \log M} - \log \sqrt{4\pi \log M} / \sqrt{2 \log M} \right). \tag{11.7}$$

That leaves us to decide on the number of samples M. Clearly we cannot choose $M > N$ because $\mathcal{Z}_U$ is determined by the N eigenvalues of U. If we choose $M = N$, a seemingly reasonable choice because $|\mathcal{Z}_U(e^{i\theta})|$ has N local maxima and we want to know the largest, we find that (11.7) equals

$$\log N - \tfrac{1}{4} \log \log N. \tag{11.8}$$

Setting $N = \log T / 2\pi$ we contradict the $C = \frac{3}{4}$ part of conjecture (11.5). That is the first way in which the FHK conjecture is surprising. In particular, their conjecture says that the short range maximum values tend to be slightly smaller than one would expect from a naive view of Selberg's theorem.

The second surprising aspect is the size of the tail of the random variable X. The maximum among independent choices from a Gaussian follow the **Gumbel distribution**, which decays like e^{-x}, whereas the FHK conjecture posits decay like xe^{-x}.

The conclusion is that the N choices from $|Z_U(e^{i\theta})|$ are not independent, which Fyodorov-Hiary-Keating [51, 49] phrase as long-range correlations. That is another way of describing the carrier waves, which we try to make more explicit in Subsection 11.3.

Non-implications for extreme values. It is tempting to use the FHK conjecture (11.5) as a step toward a conjecture for the extreme values, by invoking Principle 7.6 and using the conjectured tail xe^{-x} for X. That would predict much larger values for the ζ-function than suggested in [43], but that reasoning is flawed because of an implicit switching of limits. The limiting distribution of X_T as $T \to \infty$ may decrease like xe^{-x}, but the distribution of X_T has an additional factor $e^{-x^2 / \log \log T}$. See [3]. That factor dominates the tail when the sample size is large.

11.2 Gaussian multiplicative chaos

We have seen that $\log \zeta(\frac{1}{2} + it)$ is log-correlated. And even better: by Selberg's theorem it is (to leading order) Gaussian, so we have a Gaussian log-correlated field. Its (suitably normalized) exponential is an example of **Gaussian multiplicative chaos**, GMC, a topic which has received considerable attention due to its connection to turbulence, mathematical finance, quantum gravity, and elsewhere. Here we briefly describe the work of Saksman and Webb; see [91, 92] for details.

We have been viewing $\zeta(\frac{1}{2} + it)$ as random by choosing $t \in [T, 2T]$ uniformly. That approach treats the *values* of the ζ-function as random. Now we want to generate a random *function* from the ζ-function.

Let $\omega \in [0,1]$ be uniformly distributed, and consider the random function

$$\mu_T(x) := \zeta(\tfrac{1}{2} + ix + i\omega T) \qquad \text{for} \qquad x \in (0,1). \qquad (11.9)$$

The goal is to understand the limiting statistics of μ_T as $T \to \infty$.

The main result of [91] is that as $T \to \infty$ the random function μ_T converges to the **randomized zeta function** $\zeta_{\mathrm{rand}}(\tfrac{1}{2} + ix)$, where

$$\zeta_{\mathrm{rand}}(s) = \prod_p (1 - p^{-s} e^{2\pi i\theta_p})^{-1} \qquad (11.10)$$

with the θ_p i.i.d. uniform on $(0,1)$. Note that ζ_{rand} is a generalized function. The convergence of $\mu_T(x)$ to $\zeta_{\mathrm{rand}}(\tfrac{1}{2} + ix)$ is convergence in law with respect to the strong topology of the Sobolev space $W^{-\alpha,2}(0,1)$ for any $\alpha > \tfrac{1}{2}$. Furthermore, we have

$$\zeta_{\mathrm{rand}}(\tfrac{1}{2} + ix) = g(x)v(x) \qquad (11.11)$$

where v is a Gaussian multiplicative chaos distribution and g is a random smooth function which almost surely has no zeros. (There are additional conditions on v and g, see [91, 92].) A similar factorization holds for random characteristic polynomials, where the GMC factor is the same but the smooth part is slightly different. In both Selberg's theorem and the Keating-Snaith analogue for characteristic polynomials, the value distribution is Gaussian, but only to leading order. Similarly, ζ_{rand} is a perturbation of a GMC object.

Since the randomized zeta function is a generalized function, and not an actual function or a measure, it is not clear how to visualize it or how to have it shed light on the "valleys within valleys" mentioned below.

11.3 How wide is the carrier wave?

Our discussion of carrier waves has not indicated their characteristic scale. That is, over what span of zeros does the logarithm of the function typically change very little? Four sources of information about that question have been mentioned. We will phrase things in terms of $M/\log T$, where M is the typical number of zeros over which the size of $\log|\zeta(\tfrac{1}{2} + it)|$ does not change.

Montgomery's original speculations suggested that M might be as large as $\exp(\delta_T \log\log T)$, for some function $\delta_T \to 0$. The material below suggests that δ_T might need to decrease at least as quickly as $\log\log\log T / \log\log T$.

Bombieri and Hejhal [20] show that M can be taken to be any fixed number. That was sufficient for their application. It may be that a detailed examination of their proof allows M as large as $\log\log(T)^\kappa$ for any $\kappa < \tfrac{1}{4}$.

We now show that the FHK conjecture (11.5) suggests $M = \log\log T$. We will invoke (11.7). Suppose Y is the span over which the (logarithm of) the function typically changes very little, measured on the scale of the average zero spacing. Independent choices from $\mathcal{Z}_U(e^{i\theta})$ must have θ separated by Y/N, to cover the circle with $M = N/Y$ independent choices. Making that substitution in (11.7) gives main terms

$$\log N - \frac{2\log Y + \log\log N}{4}. \qquad (11.12)$$

Equating to (11.5) with $C = \tfrac{3}{4}$ and $N = \log(T/2\pi)$ we find $Y = \log\log(T/2\pi)$, as claimed.

Moments can provide information about how quickly the size of the ζ-function can change. For fixed $a, b > 0$ consider

$$I_+(a,b;T) := \int_2^T \zeta\big(\tfrac{1}{2} + \tfrac{a}{\log T} + it\big)\overline{\zeta\big(\tfrac{1}{2} + \tfrac{a}{\log T} + \tfrac{ib}{\log T} + it\big)}\,dt. \qquad (11.13)$$

By a theorem of Ingham [64],

$$I_+(a, b; T) \sim \frac{1 - e^{-2a-ib}}{2a + ib} T \log T. \tag{11.14}$$

If the size of $\zeta(\frac{1}{2} + it)$ were typically the same as $\zeta(\frac{1}{2} + ib/\log t + it)$, then the size of $I_+(a, b; T)$ would not depend on b. But it does, and in fact if b is not bounded then the size of the integral changes.

Does this suggest that the scale of the carrier wave, M, cannot grow with T? That is not a valid conclusion, because the carrier wave concerns the bulk of the distribution, and moments are sensitive to rare large values. Indeed, points where the $\zeta(\frac{1}{2} + it)$ is of size at most $\log \log(t)^A$, which happens 100% of the time, do not contribute to the main term of $I_+(a, b; T)$ when a and b are bounded. Thus, moments cannot tell us about the scale of the carrier wave.

11.4 Small values of the wave, and valleys within valleys...

It is easy to lose sight of the fact that the carrier wave also concerns very small values, which occupy just as much space as the large values. Modify (11.13), but to be sensitive to small values:

$$I_-(a, b; T) := \int_2^T \zeta\left(\tfrac{1}{2} + \tfrac{a}{\log T} + it\right)^{-1} \overline{\zeta\left(\tfrac{1}{2} + \tfrac{a}{\log T} + \tfrac{ib}{\log T} + it\right)^{-1}} dt. \tag{11.15}$$

By the ratios conjecture [27],

$$I_-(a, b; T) \sim \frac{15}{\pi^2} \frac{1}{2a + ib} T \log T. \tag{11.16}$$

For the analogous conjecture for an arbitrary primitive L-function $L(s)$, replace the constant $15/\pi^2$ by $\sum |\mu_L(n)|^2/n^2$, where $L(s)^{-1} = \sum \mu_L(n) n^{-s}$.

The same arguments as for I_+ apply here: the typical small values are not relevant to the size of I_-, and the atypically small values do not persist over the span of many zeros. Indeed, it may seem initially surprising that the shapes of I_+ and I_- are so similar. It is not clear what this suggests about the bias in the value distribution at low heights, which is skewed toward small values as illustrated in Figure 7.5.

It would be interesting to adapt the approach of FHK, or use some other method, to model the small values of $\log |\zeta(\frac{1}{2} + it)|$ on an interval. The zeros prevent considering that question literally, so some modification would be needed. Considering values at local maxima probably runs afoul of small maxima from occasional small zero gaps. Averaging over a small interval, a few widths of the average zero gaps, should capture the wave but may be difficult to handle analytically. Or perhaps moving away from the critical line, on a scale comparable to the average zero gap as in (11.15), is the right way to think about it.

For the example carrier waves in this paper, the first step was to generate a large random matrix and then find all of its eigenvalues. That enabled us to exhibit behavior far beyond what can be computed for the ζ-function, but it limits our ability to produce very large examples. To quote Fyodorov and Keating [49] (slightly out of context), the carrier wave should exhibit a hierarchical structure of "valleys within valleys within valleys". It would be interesting to get a glimpse of this structure, which would require directly generating example carrier waves corresponding to much larger random matrices. We mention this again at the end of Subsection 14.4. It is not clear whether on such a large scale there would be qualitative differences between the L-function and the characteristic polynomial worlds.

12 The claimed reasons to doubt RH: properties of $Z(t)$

Several of the claimed reasons to doubt RH involve analytic properties of $Z(t)$. We address those reasons first, calling upon the Principles from the previous sections in this paper.

We have described how the analogy between the ζ-function and the characteristic polynomials of random unitary matrices has provided insight to the behavior of the ζ-function. We now call on a fact which has been implicit in our previous discussions and now will play a decisive role.

> RH is true for characteristic polynomials of random unitary matrices.

That is, the zeros of unitary polynomials lie on the unit circle.

If we are confronted with information about the ζ-function, and the analogous information also applies to characteristic polynomials, then what can we conclude? Certainly we cannot conclude that RH may be false, because that conclusion would also apply to unitary polynomials, contradicting the fact that the analogue of RH is true. Similarly, every numerical calculation of the ζ-function has found all zeros on the critical line, therefore any information arising from explicit computations cannot throw doubt on RH. We summarize those observations:

Principle 12.1 *Any fact which directly translates to a statement about unitary polynomials cannot be used as evidence against RH.*

Principle 12.2 *Any fact arising from numerical computations of the ζ-function, except for an actual counterexample, cannot be used as evidence against RH.*

We will see that these principles, coupled with material from earlier sections of this paper, immediately refutes some of the arguments against RH. However, it is important to keep in mind that those arguments against RH were reasonable at the time they were made: it is only subsequent discoveries which have revealed new ideas which allow these principles to be applied.

12.1 Reason 1: Lehmer's phenomenon

This reason for doubting RH comes from Section 2 of [66].

Computing zeros has been a fundamental part of exploring the ζ-function, going back to Riemann who calculated approximations to the first few zeros. When D.H. Lehmer computed the first $15,000$ zeros in 1956, he found that $\tilde{\gamma}_{6710} - \tilde{\gamma}_{6709} < 0.054$. Those zeros are unusually close, in a sense which can be made precise. Such zeros are now called a **Lehmer pair**.

The existence of Lehmer pairs is a claimed reason to doubt RH. The justification is that if a real holomorphic function $f(t)$ has two real zeros very close together, then $f(t) + \varepsilon$ will have a pair of complex zeros for some small ε. For example, near the Lehmer pair noted above, $Z(t) - 0.004$ has a pair of complex zeros. So, in a sense, the Lehmer pairs suggest:

Principle 12.3 *If the Riemann Hypothesis is true, it is "barely true".*

The claimed argument is: Lehmer pairs indicate that RH is barely true, therefore we should be skeptical of RH.

The above argument contradicts Principle 12.1. The gaps between neighboring zeros of random unitary polynomials can be arbitrarily small. That is, arbitrarily close Lehmer pairs exist for random unitary polynomials: the normalized nearest-neighbor spacing is supported on all of $(0, \infty)$. Furthermore,

the calculations of Odlyzko [81] and subsequent computations indicate that the distribution of small gaps between zeros of the ζ-function closely match the (suitably scaled) distribution of small gaps between random unitary eigenvalues. Thus, the data on Lehmer pairs support the connection between ζ-zeros and the GUE (or equivalently the CUE because the connection only concerns leading-order behavior). But by Principle 12.1 the data provide no reason to doubt RH.

Other ways in which RH is barely true. There are other ways in which RH is "barely true". The function $Z(t) + \delta$ has complex zeros for any nonzero (positive or negative) $\delta \in \mathbb{R}$, and the reason has nothing to do with Lehmer pairs. Carrier waves (the small values, not the large values, see Subsection 9.2) cause $Z(t)$ to stay small on intervals containing many zeros, without those zeros being close together. Thus, for any $\delta > 0$ there will exist intervals I containing arbitrarily many zeros such that $|Z(t)| < \delta$ on I. Therefore $Z(t) + \delta$ and $Z(t) - \delta$ will have many positive local minima (respectively, negative local maxima) on I, and thus will have many pairs of non-real zeros.

Here is yet another way. RH is equivalent to $\Lambda \leq 0$, where Λ is the de Bruijn-Newman constant [78]. Rodgers and Tao [85] showed that $\Lambda \geq 0$, extended by Dobner [37] to all L-functions. So $\Lambda = 0$ is the only possibility consistent with RH.

Thus, RH is at best "barely true" in multiple ways. That is just a fact one has to live with. It does not tip the scales for or against RH.

"The Riemann Hypothesis is not an analysis problem". That is a quote from Brian Conrey, although he suggests that others have expressed a similar sentiment. There are several ways to understand its meaning, one of which comes from the fact that RH is at best barely true. There are many functions which are known to have all (or almost all) of their zeros on the real line or the unit circle. Examples are sine and cosine, Bessel and other special functions, Eisenstein series, period polynomials, and many families of orthogonal polynomials. What all those examples have in common is that the techniques come from analysis and not number theory, and asymptotically the zeros are close to equally spaced. In other words, the analogue of RH is not barely true: it remains true under a small perturbation. That is a feature of analysis techniques, and that is why those techniques are not going to prove RH.

We are not saying analysis techniques are not important or useful. Indeed, some of the refutations of reasons to doubt RH are purely analysis arguments. But if there is a proof of RH, the deep facts in that proof will not come from analysis.

The above discussion suggests it would be inappropriate to refer to the fact that sin, cos, Bessel functions, and some orthogonal polynomials have only real zeros, or that period polynomials and some other orthogonal polynomials have all zeros on the unit circle, as "the Riemann Hypothesis" for those functions. Such terminology overinflates the depth of such results.

12.2 Reason 2: Large values of $Z(t)$

In Section 5 of [66] the claimed reason to doubt RH directly addresses carrier waves. Specifically, it is noted that RH combined with conjectures about the maximum size of $Z(t)$ indicate that (this is quoted from [66]).

...the graph of $Z(t)$ will consist of tightly packed spikes, which will
be more and more condensed as t increases, with larger and larger
oscillations. This I find hardly conceivable.

The author agrees that it is hard to believe, and the difficulty of believing it was specifically mentioned in the introduction to Section 5. But believe it we must, because that is how carrier waves work. Also, what the graph looks like is just a matter of perspective: a snapshot of the Z-function covering a range of 200 zeros will look like tightly packed spikes no matter the size of the function and whether or not RH is true.

Thus, this reason for doubting RH is not valid. Indeed, a significant portion of the zeros are known to be on the critical line, so the "tightly packed spikes" must exist whether or not RH is true. The appearance of the graph is merely a matter of perspective: you can either see a pleasantly undulating curve, or tightly packed spikes, depending on the choice of scales on the axes.

12.3 Reason 3: Derivatives near a large maximum

The approach of Blanc [14] is based on the following interesting formula [15]. Suppose f is a smooth function on an open interval containing $[-a, a]$ for some $a > 0$, and suppose $\mathbf{x} = \{x_0, \ldots, x_n\}$ are distinct zeros of f in that interval. Then

$$\sum_{k=1}^{r} \Psi_{2k-1}(\mathbf{x}, a) f^{(2k-1)}(a) - \sum_{k=1}^{r} \Psi_{2k-1}(\mathbf{x}, -a) f^{(2k-1)}(-a) \qquad (12.1)$$
$$= \int_{-a}^{a} \Psi_{2r-1}(\mathbf{x}, t) f^{(2r)}(t) \, dt.$$

The exact form of Ψ_ℓ is tangential to our discussion, but we include it for completeness:

$$\Psi_\ell(\mathbf{x}, t) = \frac{(4a)^\ell}{(\ell+1)!} \sum_{k=0}^{n} \mu_k(\mathbf{x}) \left(B_{\ell-1}\left(\tfrac{1}{2} + \frac{x + x_k}{4a} \right) + B_{\ell-1}\left(\left\{ \frac{x - x_k}{4a} \right\} \right) \right)$$

where

$$\mu_k(\mathbf{x}) = 2^{-n} \prod_{\substack{0 \le j \le n \\ j \ne k}} \left(\sin\left(\pi \frac{x_k}{2a} \right) - \sin\left(\pi \frac{x_j}{2a} \right) \right)^{-1}.$$

Here B_ℓ is the Bernoulli polynomial and $\{x\} = x - [x]$ is the fractional part of x. The relevant fact about μ_k is that $\sum_k \mu_k = 0$.

Blanc's approach involves applying (12.1) in the context shown in Figure 12.4, where we have repurposed data from Bober and Hiary [17] which previously appeared in Figure 3.4.
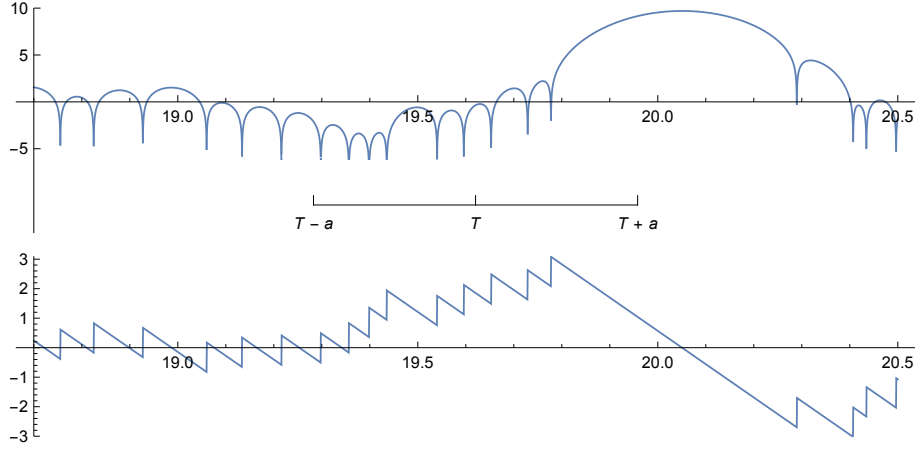
Figure 12.4 The function $\log|Z(t)|$ near a local extremum caused by a large zero gap, and $S(t)$ in the same region. The parameters T and a are chosen so that $Z'(T+a)$ is a (large) local extremum, $Z'(T-a)$ is a (presumably small) local extremum, and $S(T+a)$ and $S(T-a)$ have opposite sign.

Blanc chooses T and a so that $T+a$ lies in a large zero gap, $|Z(T+t)|$ has a very large maximum a bit to the right of a, and $T+a$ is a local extremum of $Z'(T+a)$. (Those choices are illustrated in Figure 12.4, although it is difficult to see an inflection point of $Z(t)$ on the graph of $\log|Z(t)|$. Note how the large zero gap causes the maximum of $Z'(t)$ to shift toward the maximum of $Z(t)$.) Furthermore assume $S(T+t)$ changes sign in $(-a,a)$. Blanc also chooses a so that $Z'(T-a)$ is a local extremum, although that does not seem to be essential.

The setup above lets us describe Blanc's main idea. With $f(t) = Z(T+t)$, the quantity $f^{(2k-1)}(a)$ in (12.1) will be large for small k. For larger k there is no reason for $f^{(2k-1)}(a)$ to be large, and for all k there is no reason for $f^{(2k-1)}(-a)$ to be large. Thus, some of the quantities on the left side of (12.1) will be very large. Table 12.5 lists the quantities in (12.1) in the context of Figure 12.4. Similar to Table 1 of [14], we write

$$\beta_\ell^\pm = (-1)^{(\ell-1)/2}\Psi_\ell(\mathbf{x}, T \pm a)\theta'(T)^\ell \tag{12.2}$$

$$d_\ell^\pm = (-1)^{(\ell-1)/2}Z^\ell(T \pm a)/\theta'(T)^\ell, \tag{12.3}$$

so the left side of (12.1) is $\sum \beta_\ell^+ d_\ell^+ - \sum \beta_\ell^- d_\ell^-$.

Table 12.5 Quantities appearing in (12.1) using data from Bober and Hiary [17] as shown in Figure 12.4

ℓ	β_ℓ^+	d_ℓ^+	$\beta_\ell^+ d_\ell^+$	β_ℓ^-	d_ℓ^-	$\beta_\ell^- d_\ell^-$
1	0.000896	3209.249	2.877	-0.00561	0.24181	-0.001
3	0.004657	625.137	2.911	-0.01944	0.37736	-0.007
5	0.014213	182.535	2.549	-0.04274	0.52865	-0.022
7	0.033762	67.231	2.296	-0.07822	0.70713	-0.055
9	0.069663	28.657	1.996	-0.13156	0.91859	-0.120
11	0.132016	13.376	1.765	-0.21285	1.17131	-0.249

We can obtain a contradiction if we can show that the left side of (12.1) is large and the right side is small. For the data from Figure 12.4, the right side of (12.1) is approximately 13.95852, which is consistent with Table 12.5. That merely serves as a check on the numerical computations, because RH is true in the range of those data.

If we were only considering characteristic polynomials, then by Principle 12.1

we cannot obtain such a contradiction. Therefore we need another ingredient which is specific to L-functions. This is provided by another interesting result of Blanc [14].

Recall that by (4.2) we have $\theta'(t) \sim \frac{1}{2} \log \frac{t}{2\pi}$.

Theorem 12.6 Blanc [14]. *Suppose T is sufficiently large, $\delta > \frac{1}{2}$, and $a = o(T/\theta'(T))$. If*

$$\delta\theta'(T) \leq 2K \leq 2\theta'(T)^2 \tag{12.4}$$

then

$$|Z^{(2K)}(t)| \leq 2\zeta(\tfrac{1}{2} + \delta)\theta'(T)^{2K} \tag{12.5}$$

for $t \in [T - a, T + a]$.

Blanc [14] states the case $\delta = \frac{3}{2}$ and estimates $2\zeta(2) < 4$. It might be worth investigating what choice of δ is optimal for applications of Theorem 12.6.

By Theorem 12.6 we can choose r so that the right side of (12.1) is small. Therefore, if the setup described above makes the left side of (12.1) large, we have obtained a contradiction.

The analogue of Theorem 12.6 is not true for unitary polynomials, so we explore the implications of (12.5) in some detail before returning to Blanc's argument.

Note that (12.5) is not true for any fixed K as $T \to \infty$, because that would contradict known Ω-results and conjectures for $Z^{(n)}(t)$. We now explain why, more generally, the bound (12.5) is not true for K far outside the range (12.4). We do this by exploiting properties of repeated differentiation.

Everyone wants to be cosine. Differentiating a function like $Z(t)$ can be viewed as an operation on its set of zeros, replacing the zeros of $Z(t)$ by the zeros of $Z'(t)$. The relevant fact is that differentiation causes the zeros to become more evenly spaced [46][71], with large gaps becoming smaller and small gaps becoming larger. Thus, differentiation damps the irregularities in the size of the function: very large maxima (measured relative to the size of nearby maxima) become less extreme. Under repeated differentiation the zero gaps are becoming more regular, eventually approaching equality. For sufficiently large k the kth derivative is locally approximately a function with equally spaced zeros. In other words, for t in any fixed interval, if k is sufficiently large, then

$$Z^{(k)}(t) \approx A_k \cos(B_k t + C_k) \tag{12.6}$$

throughout that interval. Berry [9] termed this phenomenon **cosine is a universal attractor**.

If the number of derivatives is too small for the zeros to be approximately equally spaced, then the variation in the zero spacing will prevent the bound (12.5) from holding. So (12.5) will not hold if K is much smaller than the lower bound in (12.4).

For a moderate number of derivatives, the parameter B_k in (12.6) directly relates to the local density of zeros. Since the local density of zeros of $Z(t)$ varies slowly, when the kth derivative initially approaches cosine we have $B_k \approx \theta'(T)$, as we previously saw in (5.4). That fact, and the chain rule, explains why powers of $\theta'(T)$ appear on the right side of (12.5), and the bound is optimal [16].

Finally, if we take "too many" derivatives then the local density of zeros begins increasing, so B_k starts to increase and therefore (12.5) will not hold if K is much larger than the upper bound in (12.4).

Some unitary polynomials approach cosine more slowly. We have already alluded to the fact that Theorem 12.6 does not hold for all unitary

polynomials. More precisely, the analogue of the bound (12.5) does hold, but the number of derivatives needed to realize that bound can be much larger than the range (12.4). To see this, it is sufficient to consider self-reciprocal polynomials in **polar form**:

$$\mathcal{Z}(z) = z^{\frac{N}{2}} + a_{N-1} z^{\frac{N-1}{2}} + \cdots + \overline{a}_{N-1} z^{-\frac{N-1}{2}} + z^{-\frac{N}{2}} \tag{12.7}$$

with differentiation operator $D = z\frac{d}{dz}$. Note also that the analogue of $\theta'(T)$ is $N/2$.

We see that

$$
\begin{aligned}
D^{2K}\mathcal{Z}(z) =\ & \left(\frac{N}{2}\right)^{2K} z^{\frac{N}{2}} + \left(\frac{N-1}{2}\right)^{2K} a_{N-1} z^{\frac{N-1}{2}} + \\
& \cdots + \left(\frac{N-1}{2}\right)^{2K} \overline{a}_{N-1} z^{-\frac{N-1}{2}} + \left(\frac{N}{2}\right)^{2K} z^{-\frac{N}{2}} \\
=\ & \left(\frac{N}{2}\right)^{2K} z^{-\frac{N}{2}}\bigg(z^N + (1-N^{-1})^{2K} a_{N-1} z^{N-1} + \\
& \cdots + (1-2N^{-1})^{2K}\overline{a}_{N-2} z^2 + (1-N^{-1})^{2K}\overline{a}_{N-1} z + 1 \bigg) \\
=\ & \left(\frac{N}{2}\right)^{2K} z^{-\frac{N}{2}}\left(z^N + 1 + o(1) \right) \tag{12.8}
\end{aligned}
$$

as $K \to \infty$. By (12.8) the roots of $D^{2K}\mathcal{Z}(z)$ are approaching the Nth roots of 1, so in particular they are approaching equal spacing. It was not even necessary to assume the original self-reciprocal polynomial had all its zeros on the unit circle. Yet the analogue of Theorem 12.6 fails without further restrictions on the polynomials because the required number of derivatives may be much larger than the analogue of the lower bound in (12.4). One can see this from (12.8) because $(1-N^{-1})^{2K}$ is not small if $K = O(N)$.

For unitary polynomials the analogue of Theorem 12.6 has counterexamples similar to the failure of the analogue of the Lindelöf hypothesis for some unitary polynomials. Consider $z^{-N/2}(z+1)^N$, with differentiation operator $D = z\frac{d}{dz}$. If we take κN D-derivatives, with $\kappa < 1$, the polynomial will still have a high multiplicity zero at $z = -1$. That will force the other zeros to be spread out, and so the function will be large. Even if we take N derivatives, so all the zeros are simple, it will still take many more derivatives for the zeros to become approximately equally spaced. In particular, by considering only the $N - \sqrt{N}$th coefficient of $(z+1)^N$, we see that $D^N z^{-N/2}(z+1)^N$ is large at $z = 1$.

From the above discussion we see that the analogue of Theorem 12.6 does not hold for characteristic polynomials, so Blanc's argument, which we analyze next, does not fall victim to Principle 12.1.

Analysis of Blanc's argument. Summarizing the ideas so far: we want to assume a very large gap between zeros, causing a very large isolated local maximum of $Z(t)$. With the choices of T and a as in Figure 12.4, the quantity $Z^{(2k-1)}(a)$ in (12.1) will be very large for small k. This might lead to a contradiction if the parameter r allows Theorem 12.6 to apply, because then the right side of (12.1) will be small. If the left side of (12.1) is large and the right side is small, then we have obtained a contradiction.

Given a large gap between zeros, and choosing r so that Theorem 12.6 applies, how can having $Z^{(2k-1)}(a)$ be very large for small k fail to lead to a contradiction? There are two ways.

1. $\Psi_1(\mathbf{x}, a)$ might be very small, so actually the left side of (12.1) is small. (For example, see Table 12.5)

2. Some other terms on the left side might be large and of the opposite sign, causing cancellation.

If the above possibilities do not apply (so we do obtain a contradiction), what has been contradicted? Have we contradicted RH, or something else? We will argue that such a contradiction, if it happens, has nothing to do with RH but rather it comes from assuming a particularly large gap between zeros. In other words, the theorems of Blanc may actually contain hidden information which might be used to provide an improved upper bound on gaps between zeros of the ζ-function.

Blanc addresses Item 2 above by noting that the setup in Figure 12.4 specifically avoids the likelihood that other terms are large. Figure 2 in [14] provides good evidence for the intuition that β_ℓ^+ should be small for large ℓ.

To investigate the size of $\Psi_\ell(\mathbf{x}, a)$, we consider a scenario similar to Figure 12.4 at height 10^{100}, which Figure 2 in [14] suggests is sufficient to test the method. Since $\log 10^{100} \approx 230.26$, we can use degree 230 unitary polynomials. Given a large zero gap, we can place the other zeros in their most likely location as described in Subsection 9.1. An example of this, with a gap 6 times the average in a degree 74 polynomial, is shown in Figure 9.4. Since everything is explicit, we can evaluate every quantity in (12.1). Table 12.7 lists the terms in (12.2) and (12.3) for degree 230 polynomials with a large gap of 5, 10 and 15 times the average spacing, with all other zeros close to their most likely configuration.

Table 12.7

ℓ	$\beta_\ell^+(5)$	$d_\ell^+(5)$	$\beta_\ell^+(10)$	$d_\ell^+(10)$	$\beta_\ell^+(15)$	$d_\ell^+(15)$
1	0.00099	363.5	2.5×10^{-6}	446988.0	1.9×10^{-8}	4.9×10^8
3	0.00329	95.2	0.000016	57835.6	1.9×10^{-7}	4.2×10^7
5	0.00654	37.7	0.000057	11436.1	1.0×10^{-6}	5.6×10^6

The main takeaway from Table 12.7 is that β_ℓ^+ is small for small ℓ when d_1^+ is large. Thus, we must question the possibility of obtaining a contradiction from (12.1). Specifically, we suggest that there is no evidence to support Blanc's assertion (bottom of first column on page 5 of [14]) "the sum of the first $\beta_{2\ell-1}^+ d_{2\ell-1}^+$ is probably large".

The comments above do not address a key element of Blanc's argument, based on (12.5), for which the analogous result for unitary polynomials is not true. Translating (12.5) to the realm of self-reciprocal polynomials gives the statement: in the context of the polar form (12.7), with $\mathcal{Z}^{2K}(z) := D^{2K}\mathcal{Z}(z)$,

$$\text{if } K \geq \frac{3}{8}N \text{ then } \max_{|z|=1} |\mathcal{Z}^{2K}(z)| \leq 3.29\left(\frac{N}{2}\right)^{2K}. \qquad (12.9)$$

That bound on $|\mathcal{Z}^{2K}(z)|$ holds if K is sufficiently large, but not in general for $K \ll N$. In particular, for the degree 230 polynomial with a zero gap 15 times the average used in Table 12.7, one must take 197 derivatives in order to achieve the estimate in (12.9). The polynomial with gap 10 times the average requires 123 derivatives. This suggests to the author that the ingredients in Blanc's argument do not throw doubt on RH, but instead point to a possible new method for bounding the size of large gaps between zeros of the ζ-function. We propose:

Problem 12.8 Suppose $\mathcal{Z}$ is a polar unitary polynomial of degree N with a zero gap of κ times the average. For $C > 2$, determine a lower bound on $\eta = \eta(\kappa, C)$ such that if $2K < \eta N$ then

$$\max_{|z|=1} |\mathcal{Z}^{(2K)}(z)| > C\Big(\frac{N}{2}\Big)^{2K}.$$

$\square$

The techniques used to solve that problem might translate into a method for improving the upper bound on gaps between zeros of the ζ-function.

Further reflections on characteristic polynomials. There is an apparent discrepancy in our discussion of large gaps between zeros. On the one hand we have described how random matrix theory provides a precise conjecture on the maximum gap size. On the other hand, we have described properties of $Z(t)$ which suggest constraints on large gaps between zeros, and those constraints do not apply to characteristic polynomials.

These apparent discrepancies might be resolved by examining in more detail how characteristic polynomials are used to model $Z(t)$. By Principle 7.6, to model the largest gaps one chooses around e^N independent random matrices in $U(N)$, where $N \approx \log T$. Based on the enormous success of using RMT to shed light on L-functions, it it plausible that the analogue of Theorem 12.6 holds almost surely (in the probabilistic sense) for e^N Haar-random matrices chosen from $U(N)$.

13 Claimed reasons to doubt RH: other properties

The final two claims we consider involve averages of high powers of $Z(t)$, and properties of Dirichlet series which are not L-functions.

13.1 Reason 4: Error terms in moment formulas, and maybe the Lindelöf Hypothesis is also false

The argument in Section 4 of [66] is based on the meager knowledge we have about moments of the ζ-function:

$$\int_0^T Z(t)^2 \, dt = T \log T + (2\gamma - 1 - \log 2\pi)T + O(T^{E_1 + \varepsilon})$$
$$\int_0^T Z(t)^4 \, dt = T P_2(\log T) + O(T^{E_2 + \varepsilon})$$

where P_2 is a certain degree 4 polynomial. The conjectured optimal error terms are $E_1 = \frac{1}{4}$ and $E_2 = \frac{1}{2}$.

For higher moments, the conjecture [28] is:

$$\int_0^T Z(t)^{2k} \, dt = T P_k(\log T) + O(T^{E_k + \varepsilon}) \tag{13.1}$$

$$\sim g_k a_k T \log^{k^2} T, \tag{13.2}$$

as $T \to \infty$, for some $E_k < 1$, where $P_k(x)$ is a polynomial of degree k^2 with leading coefficient $g_k a_k$, where a_k is given explicitly, and g_k is an integer. That conjecture is open for $k \geq 3$.

Why moments? Moments of the ζ-function were introduced by Hardy and Littlewood [56] as an approach toward the Lindelöf Hypothesis 9.7.

RH implies LH, so it is natural to attack LH as an "easier" problem. LH is equivalent to

$$\int_0^T Z(t)^{2k}\, dt \ll_k T^{1+\varepsilon} \tag{13.3}$$

for all integer $k > 0$, which is why Hardy and Littlewood investigated moments.

Bounding such moments proved to be difficult: 100 years after Hardy and Littlewood we only know (13.3) for $k = 1$ and 2. But instead of becoming a dead-end in terms of proving LH, moments became a topic of independent interest. Other "families" of L-functions were introduced, with various types of moments. The focus shifted to finding precise formulas for the main terms in moments, not just on estimating the order of magnitude. Random matrix theory played a major role in understanding the leading order behavior and the structure of those moments, although ultimately number-theoretic heuristics provided the most precise conjectures. See [69][25][70][26][36].

Analysis of the argument. The claimed argument against RH arises from the error term in the $2k$th moment. Based on the conjectured best possible values $E_1 = \frac{1}{4}$ and $E_2 = \frac{1}{2}$ Ivić suggests [65][66] that the best possible error term in the $2k$th moment is $O(T^{k/4+\varepsilon})$. There is slightly more to Ivić's reasoning, such as the apparent inability of available techniques to provide bounds on high moments, but the main input is the conjectured values for E_1 and E_2.

If the error term (13.1) truly is of size $T^{k/4+\varepsilon}$, meaning that (13.1) is false with $k/4$ replaced by a smaller number, then (13.3) is also false. Therefore LH is false, so RH is false. That is the essence of that argument against RH.

Basing an argument on such limited data for small k is problematic, and the conjecture of $E_k = k/4$ is not based on any underlying principles. Suppose someone instead conjectured $E_k = \frac{3}{4} - 2^{-k}$ for $k \geq 1$. That also is a simple formula which fits the limited data, and it has the added benefit of implying LH. But without any good evidence, that conjecture should also not be taken seriously. Thus, this proposed reason to doubt RH is not based on a mathematical foundation, but on naive pattern matching.

Before the reader criticizes the author for being unkind, let us note that this author fully recognizes the folly of speculating on the size of error terms, and indeed is guilty of similarly unwise speculation. In a foundational paper on moments of L-functions [26] this author and coauthors conjecture that in great generality the error term in any conjectured higher moment is $E_k = \frac{1}{2}$. The conjecture of Ivić is as pessimistic as possible, and the conjecture in [26] is as naively optimistic as possible. Unfortunately, that optimistic error term is not always correct. Specifically, in some cases [35][36] there is a secondary main term of size $X^{\frac{3}{4}}$. It may even be that similar secondary main terms are common. The lesson is that speculating on the error term in moment conjectures is difficult, and it is folly to draw conclusions from such error term conjectures.

The structure of moments. Immediately after making a poorly founded conjecture about error terms, Ivić goes on [65][66] to make the perceptive speculation:

> the shape of the asymptotic formula...changes when $k = 4$.

That speculation has been borne out by subsequent work on the structure of moments.

The $2k$th moment (13.1) is only known for $k = 0$, 1, and 2. Conjectures for larger moments are a fairly recent development. One reason those moments are difficult is that the natural way to proceed is via an "approximate functional equation" for $Z(t)^k$, which is a sum of two Dirichlet polynomials of length $t^{k/2}$. (The Riemann-Siegel formula (4.1) can be viewed as an example: there are $\approx t^{\frac{1}{2}}$ terms in the sum.) When used to evaluate $\int_0^T Z(t)^{2k}\, dt$ the approximate functional equation will have length $T^{k/2}$. The difficulty is the lack of tools to handle Dirichlet polynomials where the length of the polynomial is greater than the length of the integral.

Conrey and Ghosh [29] developed a heuristic approach which gave a conjecture when $k = 3$, and Conrey and Gonek [31] pushed the method to its limit to make a conjecture when $k = 4$. That truly is the limit of the method, because when applied to the case $k = 5$ those methods produce an answer which is negative! Thus, Ivić was correct: some new phenomenon appears, beginning at $k = 4$. But that new phenomenon has more implications for that main term than for the error term.

13.2 Reason 5: The Deuring-Heilbronn phenomenon

The Riemann ζ-function is the simplest of an infinite family of functions known as "L-functions". The next simplest examples are the Dirichlet L-functions, $L(s, \chi)$, where χ is a primitive Dirichlet character. RH for all $L(s, \chi)$ is known as the **Generalized Riemann Hypothesis**.

Informally, by **L-function** we mean a Dirichlet series with a functional equation and an Euler product of a particular form. To give a more precise definition, there are two primary approaches. The *axiomatic approach* considers L-functions as analytic functions with certain properties. Within the axiomatic approach at one extreme we have the Selberg class [95][30], which has a minimal set of general axioms. At the other extreme are the tempered balanced analytic L-functions [45] which have a large number of precise axioms.

The *structural approach* describes L-functions as arising from arithmetic or automorphic objects. For the perspective in this paper, the important fact is that conjecturally all approaches describe the same set of functions, and *all such L-functions have an analogue of the Riemann Hypothesis*. The collection of all such Riemann Hypotheses is called the **grand Riemann Hypothesis**.

Under any of the above definitions, each L-function has a functional equation analogous to (3.3), a ξ_L-function analogous to (3.4), and a Z_L-function analogous to (3.6). Each Z_L-function is a real valued function which wiggles with the same type of randomness as $Z(t)$. It also has carrier waves as described in the earlier parts of this paper.

Linear combinations. The Deuring-Heilbronn phenomenon concerns functions which are not L-functions, but rather are linear combinations of L-functions. As we will see in Principle 13.2, a nontrivial linear combination of L-functions never satisfies RH, in the worst possible way.

Consider two L-functions which are being added, and focus on a small interval of the real line (say, where one expects to see a few dozen zeros of each Z-function). If one of the Z-functions happens to be much larger than the other over that entire interval, then the sum looks just like the larger Z-function. In particular, if the larger Z-function satisfies RH, then their sum satisfies RH on that interval. That is the typical situation and this was the motivation for developing the idea of carrier waves: the logarithm of each Z-function has a Gaussian distribution, and because of the carrier waves its size does not change too rapidly. Thus, 100% of the time one of them is much larger than the

other and stays larger over a span of many zeros. Under some mild additional hypotheses, that is how Bombieri and Hejhal [20] proved that if two L-functions have the same functional equation and each satisfies RH, then their sum has 100% of its zeros on the critical line.

But, 100% is not everything. As the two Z-functions take turns being the largest, occasionally they are about the same size over the span of a few zeros. When that happens, and we are adding two *different* functions that look like the top graph in Figure 3.1 and are similar in size, then not all of the zeros of the sum will be real. In other words:

Principle 13.1 *A nontrivial linear combination of L-functions will not satisfy the Riemann Hypothesis.*

Naturally occurring examples. The above discussion makes it seem obvious that a nontrivial linear combination of L-functions will never satisfy RH. But the situation was not always so clear. Let $\chi_{5,2}$ be the Dirichlet character mod 5 with $\chi(2) = i$, and define θ by $\tan\theta = (\sqrt{10-2\sqrt{5}}-2)/(\sqrt{5}-1)$. The **Deuring-Heilbronn function** is given by

$$F_{DH}(s) = \tfrac{1}{2}\sec\theta\left(e^{-i\theta}L(s,\chi_{5,2}) + e^{i\theta}L(s,\overline{\chi}_{5,2})\right) \qquad (13.4)$$

which satisfies the functional equation

$$\xi_{DH}(s) := 5^{\frac{1}{2}s}\pi^{-\frac{1}{2}s}\Gamma(\tfrac{1}{2}s + \tfrac{1}{2})F_{DH}(s) = \xi_{DH}(1-s). \qquad (13.5)$$

Except for the sign, that is the same functional equation satisfied by $L(s,\chi_{5,2})$. However $F_{DH}(s)$ does not satisfy RH: its first zero off the critical line is near $0.8085 + 85.699i$. Furthermore [101] it has infinitely many zeros in $\sigma > 1$, although it does not seem that an explicit example has ever been computed.

We have two Dirichlet series, $F_{DH}(s)$ and $L(s,\chi_{5,2})$, which satisfy similar functional equations. We know that one of them does not satisfy RH, so why should the other? That is the essence of the argument in Section 3 of [66]: why should we believe RH for one function when we know it is false for another, and those functions (quoting from [66]) "share many common properties"?

The Euler product. The standard answer to the above question (as noted in [66]) is that the functions are fundamentally different because $L(s,\chi_{5,2})$ is an L-function, but $F_{DH}(s)$ is not — because it does not have an Euler product. It is an exercise to check that every axiom in the Selberg class [95] is essential: omit any one of them and there will be functions which do not satisfy RH. The Euler product condition is not more important than the other axioms, but it is important, and without it one has no expectation of RH.

The Deuring-Heilbronn function is one of many examples constructed by taking linear combinations of L-functions. It is an exercise to show that a nontrivial linear combination of Euler products is not an Euler product. Thus, any such linear combination is not expected to satisfy RH, even if the constituent functions do satisfy RH. This explains why examples like $F_{DH}(s)$ do not cast doubt on RH. However, there is a bit more to be said if we consider the locations of the zeros of those functions.

Zeros outside the critical strip. The Deuring-Heilbronn function failed RH in the worst possible way: it has zeros in the region where the Dirichlet series converges absolutely. In fact it has $\gg T$ zeros in $\sigma > 1$ up to height T, but that is not any extra information because Dirichlet series are almost periodic functions (of t) in $\sigma > 1$: so once there is one zero in that region, there must

be $\gg T$ zeros in that region. Those zeros in $\sigma > 1$ are a general phenomenon, so we obtain a refinement of Principle 13.1:

Principle 13.2 *A nontrivial linear combination of L-functions will have infinitely many zeros in $\sigma > 1$.*

Principle 13.2 is a theorem for Dirichlet L-functions [90], for automorphic L-functions [22], and in an axiomatic setting [86]. Those results are interesting because one has a vector space of Dirichlet series, all of which satisfy the same functional equation. Conjecturally only the functions in the span of individual basis elements (i.e. the actual L-functions) satisfy RH, yet every other element of the space fails RH in the worst possible way. This highlights the key role of the Euler product.

Why exactly is the Euler product so important? Maybe it isn't. In terms of the analytic properties of the L-function, the most obvious consequence is that the Euler product prevents the L-function from having any zeros in $\sigma > 1$. So, is it the Euler product that matters, or the lack of zeros in $\sigma > 1$?

14 Discussion of the Mistaken Notions

We revisit the topics in Section 4 in the context of the Principles. Specifically, we re-examine the mistaken notions that the largest values of the ζ-function arise from particularly large zero gaps, that those large values occur when many of the initial terms in the Riemann-Siegel formula have the same sign, that large values are places to look for possible counterexamples to RH, and that the Gram points are special.

14.1 Large values and large gaps

Mistaken Notion 4.3, equating the largest values of the ζ-function with the largest gaps, is the core idea for three of the Notions. As has already been adequately addressed by the discussion of carrier waves, see Principle 5.2, that Notion does not accurately describe the typical large values, nor the truly largest values.

However, Mistaken Notion 4.3 *does* describe the behavior in the range of current computations, and sometimes a wrong idea can lead to the right answer. Kotnik [67] sought computational evidence that the known Ω-result (3.14) was not the true order of growth. The method was to find the sequence of largest values of $|\zeta(\frac{1}{2} + it)|$ for small t. Those computations did make it appear plausible that the true order of growth is faster than the Ω-result. Since those large values arise from large gaps, the computations were not revealing the true order of growth. But by Principle 9.2, the maxima in the large gaps should be $\gg e^{c\sqrt{\log t}}$, which is larger than the current best Ω-result.

14.2 Finding large values

Mistaken Notion 4.4 asserts that the largest values of the ζ-function occur when many initial terms in the Riemann-Siegel formula have the same sign. The Notion makes sense: the first few terms are the largest, so if you want the entire sum to be large, it is efficient to focus on the initial terms. That argument ignores the fact that there are $\gg t^{\frac{1}{2}}$ terms of size $\gg t^{-\frac{1}{4}}$. Even a small bias in such a large number of terms could contribute much more than the initial terms. Indeed, one of the arguments for the conjectured maximum size of the

ζ-function is based on a random model for the tail of that sum, which is larger than can be obtained from a strong bias in the initial terms.

Nevertheless, Mistaken Notion 4.4 is how the largest known values of the ζ-function have been found [17][100]. It is possible to take that idea too far. Consider

$$F(t, g(t)) = 2 \sum_{n < g(t)} n^{-\frac{1}{2}} \cos(\theta(t) - t \log n). \qquad (14.1)$$

The Riemann-Siegel formula has $g(t) = \sqrt{t/2\pi}$. One might expect that $F(t, t^\varepsilon)$ is a good approximation for the large values of $Z(t)$ because (assuming the Lindelöf Hypothesis) a similarly short sum is a good approximation to $\zeta(s)$ in $\sigma > \frac{1}{2}$. Kotnik [67] considered $F(t, t^\delta)$ for $\delta = \frac{1}{3}$ and $\delta = \frac{1}{4}$. Those functions are faster to evaluate than the Riemann-Siegel formula, and this sped up the computations by more quickly locating regions where $Z(t)$ was more likely to be large.

Tihanyi [99] took it a step too far by considering $F(t, \log t/2\pi)$. Since $F(t, g(t)) \ll \sqrt{g(t)}$, Tihanyi's approximation has no hope of capturing the largest values. Yet, using $F(t, \log t/2\pi)$ as a filter (the same strategy as Kotnik [67]) turned out to be useful for finding many points where $Z(t)$ was larger than 1000, and in subsequent work [100], found a value of $Z(t)$ even larger than that found by Bober and Hiary [17]. So, a Notion may be wrong or misleading, yet still serve as the seed for a new discovery.

Although it is out of reach to compute in regions where the carrier wave, and not the local zero gaps, provide the main contribution to the size of $Z(t)$, it would still be of theoretical interest to find heuristics for locating truly large values.

14.3 Trying to refute RH

If RH were not true, how would one go about searching for a pair of zeros off the line? Mistaken Notion 4.5 suggests that a large gap might "push some zeros off the line". That is contrary to the idea behind Principle 12.1: random unitary polynomials can have large zero gaps — conjecturally with the same size and frequency as the ζ-function — which influence the nearby gaps but do not destroy the fact that the polynomial is unitary.

We offer two more arguments against Mistaken Notion 4.5. The first is similar to an idea used in the discussion of carrier waves in Section 5: run the argument in reverse. Suppose there is a pair of zeros off the line. Would one expect there to be a particularly large gap nearby? Surely not. That pair off the line, if they were not too far from the line, would only cause a slightly larger than average gap in zeros on the critical line, and furthermore the function would be small (not large) within that zero gap.

The second argument concerns the fact that no plausible suggestion has been made for how RH would fail if it were false. Maybe the zeros in $\sigma > \frac{1}{2}$, if they exist, all lie on $\sigma = \frac{3}{4}$? Is that a particularly absurd suggestion? Zeros that far from the line have a very small influence on the behavior of $Z(t)$ — in particular there is no associated large value or large zero gap. It is perfectly reasonable to seek fame and fortune by trying to computationally disprove RH. But searching for zeros off the line should be accompanied by, and tailored to, a model for how the failure will occur. A scenario in which RH fails, and those failures are typically associated to a large zero gap, is arguably even less plausible than having all non-critical zeros on $\sigma = \frac{3}{4}$.

The carrier wave frequently causes the Z-function to stay very small over a wide span of zeros. In that region a slight perturbation would cause many zeros to fall off the critical line. Maybe attempts to computationally disprove RH

should begin by looking for regions where the initial terms in the Riemann-Siegel formula mostly cancel to nothing?

14.4 Gram's law, as it was originally formulated

Mistaken Notion 4.6 contains two statements: that Gram points are special, and that Gram's law is helpful for understanding properties of the zeros. First we address Gram's law. For reasons why Gram points are not special, see Subsection 14.5.

Gram's law, as usually stated, misses a key point of Gram's original observation. What Gram noted is that the zeros and the Gram points "separate each other". In other words, in the range computed by Gram not only does the nth Gram interval (g_n, g_{n+1}) contain exactly one zero, that zero is the $n + 2$nd zero. (The $+2$ arises because the definition of Gram point has $g_0 \approx 17.84$.) We propose:

Reformulation 14.1   The Precise Gram's Law. *The **precise Gram's law holds** for the **Gram interval** (g_m, g_{m+1}) if that interval contains the $m + 2$nd zero of the Z-function and no other zeros.*

That version of Gram's law also follows from Principle 5.8.

Gram's observation was based on the first 15 zeros, but note that he suggested the interlacing would not continue to hold indefinitely. In fact, Gram's law (either version) fails after the 126th zero, $\gamma_{126} \approx 279.2$, because there is a large zero gap and the 127th zero falls in the 126th Gram interval. The Gram interval (g_{3357}, g_{3358}) contains exactly one zero, but the precise Gram's law fails because the previous interval contains two zeros. That is the first time there is a difference between the two versions of the law. There are 16 differences among the first $10,000$ zeros.

That the precise Gram's law holds initially can be seen either as a consequence of $S(t)$ being small, or that the first term in the Riemann-Siegel formula is providing a good approximation for the zeros. A similarity between those two reasons is that the sign and the Γ factor in the functional equation give rise both to the first term in the Riemann-Siegel formula, and also to the main part of the counting function of the zeros. The Dirichlet series gives rise to the other terms, and also to $S(t)$.

However, we suggest that $S(t)$ provides a more fundamental insight into the behavior of the ζ-function and its zeros. A large height, the Riemann-Siegel formula needs more terms, but it is not immediately obvious that this causes the precise Gram's law to fail most of the time. On the other hand, at large height $|S(t)|$ typically is big, which implies that the precise Gram's law holds rarely:

Observation 14.2 A necessary (but not sufficient) condition for the precise Gram's law to hold on a Gram interval (g_n, g_{n+1}) is that $|S^+(g_n)| \leq 1$.

The frequency of that occurrence at height t can be estimated using Selberg's central limit theorem. It is clear that in the limit it occurs 0% of the time. The slow growth of $S(t)$ makes this effect difficult to observe: even for characteristic polynomials of random matrices in $U(1000)$, as in the examples in Section 8, the precise Gram's law still holds more than 40% of the time.

The fact that Gram's law (the non-precise version) holds around 66% of the time should be viewed as an aspect of the random matrix model of zeros of the ζ-function: the normalized nearest neighbor spacing of zeros/eigenvalues, see the leftmost plot in Figure 8.9, is concentrated around 1. So it is reasonably likely that *any* interval of normalized length 1 contains exactly one zero/eigenvalue

(although at low height a Gram interval is more likely than a random interval, see [55] or apply Principle 5.8).

The precise formulation of Gram's law describes a rigidity which only holds initially.

Principle 14.3 The sloshing water model. *At larger heights the zeros* ***slosh back-and-forth***. *That term is meant to evoke what happens when one tries to walk while carrying a wide but shallow pan of water. Large groups of zeros are shifted to the left or the right of their expected location, even though locally all sets of zeros behave similarly.*

Because of the sloshing, $S(t)$ is large (in absolute value) most of the time. Long runs of zeros miss their intended Gram intervals: many in a row falling to the right, and later, many in a row falling to the left. (But locally the zeros are still obeying random matrix statistics, so a Gram interval (or any other interval of that length) will contain exactly one zero more than 66% of the time.) The Z-function is wiggling with great amplitude for a while, and then wiggling with small amplitude. The size of the carrier wave is independent of whether the nearby zeros arrive before or after their Gram interval, that is, independent of the size or the sign of $S(t)$. All of these phenomena are manifestations of the density wave, a wave which cannot exist if the precise Gram's law were true too often.

Problem 14.4 Develop a random model for the density wave, equivalently the carrier wave, preferably as a theorem in the case of the CUE.   □

A good solution to that problem will allow generating examples from systems far larger than possible if one must first produce a large random matrix and then find its eigenvalues.

A solution to that problem must grapple with the issue that the definition of "density wave" depends on the range over which the density is being measured. A solution to that problem, along the lines of directly creating a density wave as in Figure 8.5, could be a step toward the next level of understanding the behavior of the ζ-function a large height. By that we mean: in the first level we recognize that the local distribution of zeros is random; that randomness is now understood. This is occurs in both the L-function and the characteristic polynomial worlds. The next level is carrier/density waves. This is beyond the reach of the L-function world, and observable but not understood in the random matrix world. Once carrier waves are understood, as in Problem 14.4, then it is natural to ask whether those waves are sitting in a yet larger structure. The analogy looks like

$$\text{local zero spacing : carrier wave :: carrier wave : ?????.} \qquad (14.2)$$

If one could directly produce density/carrier waves, then perhaps those can be generated without needing to directly produce a large random matrix. This may exhibit interesting (self similar?) structures, as mentioned in Subsection 11.4.

14.5 Gram points are of transient interest

A recent paper of Shanker [96] used computational data for the ζ-function to make two conjectures about Gram intervals, and to suggest that "Gram points have interesting properties which distinguish them from random points on the critical line". We examine that work in the context of the Principles, reaching a different conclusion.

The conjectures in [96] are:

1. If $\{g_n\}$ are the Gram points, then the distribution of $Z(g_{2n})$ is the negative of the distribution of $Z(g_{2n+1})$.

2. The properties of the *Z*-function on a set of consecutive Gram points is the same if that set is reversed.

   For example, the probability that the number of zeros between consecutive Gram points is $3, 1, 4, 1, 5, 9$ is the same as the probability that the number of zeros is $9, 5, 1, 4, 1, 3$.

We will see that both conjectures are true for random unitary matrices, therefore both conjectures are immediate consequences of the Keating-Snaith Law.

The second part of [96] uses two sets of data to suggest that Gram points are special. Say that a Gram point g_n is **good** if $(-1)^n Z(g_n) > 0$ and **bad** otherwise. Table 14.5 contains data from Table 2 in [96].

Table 14.5 The proportion of good and bad Gram intervals containing $0, 1, 2$, or 3 zeros, and the combined proportion ignoring whether the interval is good or bad. Data from 10^7 zeros near $t = 10^{28}$. The bottom row is the limiting prediction from the GUE Hypothesis (Principle 7.3).

	$m = 0$	1	2	3	size
good	0.1097	0.7829	0.1071	0.00016	7373998
bad	0.3083	0.3838	0.3006	0.00709	2626002
all	0.1619	0.6781	0.1579	0.00198	10000000
GUE	0.1702	0.6614	0.1664	0.00186	

The second set of data concerns consecutive Gram intervals which contain either $2, 1, \dots, 1, 0$ or $0, 1, \dots, 1, 2$ zeros. Those **Gram blocks** are called **Type I** and **Type II**, respectively. As mentioned above, Type I and Type II, of the same length, occur with equal probability. However, what is considered in [96] are intervals which are shifted from Gram intervals by $k\delta$ where δ is the length of a Gram interval and $k \in \{-0.2, -0.1, 0, 0.1, 0.2\}$. Table 14.6 contains part of Table 5 from [96].

Table 14.6 The ratio Type II/Type I intervals for displaced Gram points, for 10^7 zeros near $t = 10^{28}$.

Length	-0.2δ	-0.1δ	0	0.1δ	0.2δ
2	2.268	1.504	0.999	0.663	0.441
3	3.624	1.896	0.998	0.526	0.274
4	5.588	2.367	1.001	0.426	0.178
5	8.849	2.923	1.011	0.343	0.115
6	14.373	3.728	1.008	0.266	0.070

The third conjecture in [96] is that in each row of Table 14.6, the product of the entries for $k\delta$ and $-k\delta$ equals 1. For characteristic polynomials, that conjecture is immediate from the formula for Haar measure; see below for a more complete explanation.

The numbers Table 14.5 were used to suggest that, while the behavior of all Gram points seems close to the limiting GUE prediction, the difference between the good and bad intervals casts doubt on the random matrix prediction. The numbers in Table 14.6 were used to suggest that the Gram points are special, distinguishing them from typical points having the same spacing. We will reach different conclusions by invoking the Keating-Snaith Law and performing similar experiments with random unitary polynomials.

Gram point for unitary polynomials. Recall that Gram points for the ζ-function are the extrema of the first term in the Riemann-Siegel formula (4.1).

As described in [55] we define the **Gram points** for a unitary polynomial in an analogous way. Recall that $\mathcal{Z}_A(\theta)$, defined in (7.15), is real for $\theta \in \mathbb{R}$. Therefore there exists θ_A such that

$$\mathcal{Z}_A(\theta) = 2\cos\left(\frac{N}{2}(\theta - \theta_A)\right) + \sum_{0 \le n < N/2} a_n \cos(n(\theta - \theta_A)), \qquad (14.3)$$

with $a_n \in \mathbb{R}$ and where for convenience we have assumed N is even. The leading term is analogous to the leading term $2\cos(\theta(t))$ in the Riemann-Siegel formula, so we define the **Gram points** for the matrix A to be the points where the leading term in (14.3) has a maximum or minimum: $(\theta_A + 2\pi n/N)_{n=0,1,2,\dots}$. Note that if $\det A = 1$ then $\theta_A = 0$. See [55] for a more detailed discussion.

Now it is straightforward to explain the first two conjectures in [96]. The first conjecture follows from Principle 3.2. The first two conjectures are immediate from the Keating-Snaith Law and Lemma 7.9, using the above definition of Gram point.

The third conjecture, concerning the statistics for a sequence of Gram intervals shifted by $k\delta$ and $-k\delta$ follows from Lemma 7.9 because any statistic for a sequence of Gram intervals shifted by $k\delta$ is formally the same as the reverse sequence shifted by $-k\delta$.

We now address the larger issue of interpreting the data in Table 14.5 and Table 14.6.

By the Keating-Snaith Law, we match the size of the matrices using the rule $N = \log(t/2\pi)$. In the context of the data from [96], this means that the appropriate matrix size is $N = 62$. Table 14.7 and Table 14.8 show the results from 10^7 random eigenvalues from matrices in $U(N)$ for $N = 62$, 250, and 1000, analogous to Table 14.5 and Table 14.6, respectively

Table 14.7 The proportion of good and bad Gram intervals containing $0, 1, 2$, or 3 zeros, and the combined proportion ignoring whether the interval is good or bad. Data from 10^7 eigenvalues of Haar-random matrices in $U(N)$ for $N = 62$, 250, and 1000.

N		$m = 0$	1	2	3	**size**
62	good	0.1111	0.7802	0.1083	0.00021	7295392
	bad	0.2996	0.4009	0.2924	0.00687	2704608
	all	0.1621	0.6776	0.1581	0.00201	10000000
250	good	0.1269	0.7484	0.1241	0.00050	6621923
	bad	0.2488	0.5031	0.2433	0.00461	3378077
	all	0.1681	0.6655	0.1643	0.00189	10000000
1000	good	0.1377	0.7262	0.1351	0.00081	6147041
	bad	0.2205	0.5610	0.2148	0.00356	3852959
	all	0.1696	0.6625	0.1658	0.00187	10000000

Table 14.8 The ratio Type II/Type I intervals for displaced Gram points, among 10^7 eigenvalues of Haar-random matrices in $U(N)$ for $N = 62$, 250, and 1000.

N	Length	-0.2δ	-0.1δ	0	0.1δ	0.2δ
62	2	2.245	1.498	1.001	0.670	0.448
	3	3.499	1.876	0.998	0.534	0.288
	4	5.048	2.257	1.003	0.444	0.198
	5	7.372	2.732	1.011	0.372	0.138
	6	9.756	3.155	0.994	0.318	0.100
250	2	1.730	1.322	1.002	0.762	0.579
	3	2.216	1.500	1.005	0.667	0.454
	4	2.673	1.638	0.998	0.611	0.373
	5	3.221	1.812	1.011	0.560	0.318
	6	3.683	1.987	1.018	0.524	0.273
1000	2	1.484	1.218	0.995	0.816	0.672
	3	1.750	1.337	1.000	0.749	0.570
	4	2.013	1.432	1.993	0.705	0.502
	5	2.148	1.447	0.982	0.664	0.451
	6	2.424	1.566	1.024	0.661	0.436

The data are consistent, therefore Gram points are not important.
Table 14.5 and Table 14.7 both show that the good and bad gram points behave differently, and the proportions in each table are similar. However, this is not evidence against the prediction that the distinction between good and bad gram points is meaningless in the long run. Indeed, a matrix size of $N = 62$ is small, particularly when some of the limiting behavior reveals itself on the scale of $\sqrt{\log N}$. The $N = 250$ and 1000 sections of Table 14.7 provide a nice illustration of the fact that, as $N \to \infty$, the good and bad intervals become equally likely. As described in Subsection 14.4, at large height most zeros have slid far away from their expected location, so it is meaningless to try to separate "good" from "bad". Yet the local spacing between zeros still looks the same everywhere, and that is the real reason that Grams law (but not the precise Grams law) continues to hold more than 66% of the time.

Table 14.8 shows that at $N = 62$, shifting the Gram points has a significant effect. But 62 is a small number, and the $N = 250$ and 1000 cases make it less surprising that as $N \to \infty$ every entry in that table will approach 1.

The data differ significantly. We have argued that the numbers in Table 14.5 and the $N = 62$ portion of Table 14.7 are similar, but in fact they differ significantly. The sample sizes are 10^7, so those numbers differ by more than 10 standard deviations. This is a consequence of Principle 7.7. The lower order terms mentioned in that principle can be (statistically) significant, particularly when we are sampling at a small matrix size such as $N = 62$. Note that the differences are in the direction indicated by that principle: for the random matrices larger gaps and clusters of zeros are more likely.

The differences between Table 14.6 and the $N = 62$ portion of Table 14.8 are even more pronounced. The differences are greatest for the unlikely events: there are 10^7 data points, but the lowest row of each table corresponding to length $= 6$ only occurs a few thousand times in the sample. So, those apparent discrepancies are also explained by the principle that for random matrices there are secondary terms which make extreme events more likely.

15 The chicken or the egg?

Which came first: the ζ-function or its zeros? That sounds like a meaningless question, but we suggest that the answer factors into whether or not one is skeptical of RH.

Suppose you believe the function is what matters, either the ζ-function given by its Dirichlet series or some other expression, or the Z-function given by the Riemann-Siegel formula. From that perspective, the zeros are determined by the function, and it would take an unlikely conspiracy for RH to be true. It might still bother you that RH is "barely true". The suggestion at the very end of Subsection 14.3 might be viewed as an intriguing idea worth exploring. A failure of RH could occur for accidental reasons.

Now suppose you believe it is the zeros which matter. The zeros determine the function, with the density wave setting the scale and the local arrangement providing lower-order adjustments. Every zero is important and meaningful: a purported list of zeros containing one tiny error would be detected (by Weil's explicit formula or some other means). If there were zeros off the line, they would be there to serve a specific purpose and their location would be meaningful. The suggestion at the end of Subsection 14.3 is just silly: the Z-function stays small over a wide region only because the density of zeros is higher.

It is hoped that this paper helps cultivate an appreciation for the explanatory power which comes from viewing the zeros as the basic object.

Bibliography

[1] Louis-Pierre Arguin, Paul Bourgade, and Maksym Radziwiłł, *The Fyodorov-Hiary-Keating Conjecture. I*, arXiv: 2007.00988; (2020)

[2] Louis-Pierre Arguin, Paul Bourgade, and Maksym Radziwiłł, *The Fyodorov-Hiary-Keating Conjecture. II*, arXiv: 2307.00982; (2023)

[3] Louis-Pierre Arguin, Guillaume Dubach, and Lisa Hartung, *Maxima of a Random Model of the Riemann Zeta Function over Intervals of Varying Length*, Ann. Inst. Henri Poincaré Probab. Stat. no. 1, **60** (2024) pp. 174-200. arXiv: 2103.04817;

[4] F. V. Atkinson, *The mean value of the zeta-function on the critical line*, Proc. London Math. Soc. no. 2, **47** (1941) pp. 174-200.

[5] R. Backlund, *Über die Beziehung zwischen Anwachsen und Nullstellen der Zetafunktion*, Öfversigt Finska Vetensk. Soc. no. 9, **61** (1918-19)

[6] E.C. Bailey and J.P. Keating, *Maxima of log-correlated fields: some recent developments*, J. Phys. A: Math. Theor. **55** no. 5, arXiv: 2106:15141; (2022)

[7] Yacine Barhoumi-Andréani, Christopher Hughes, Joseph Najnudel, and Ashkan Nikeghbali, *On the number of zeros of linear combinations of independent characteristic polynomials of random unitary matrices*, Int. Math. Res. Not. IMRN no. 23, (2015) pp. 12366-12404.

[8] Gérard Ben Arous and Paul Bourgade, *Extreme gaps between eigenvalues of random matrices*, Ann. Probab. **41** no. 4, (2013) pp. 2648-2681.

[9] M. V. Berry, *Universal oscillations of high derivatives*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **461** no. 2058, (2005) pp. 1735-1751.

[10] M. V. Berry, *Superoscillations and leaky spectra*, J. Phys. A: Math. Theor.

52 (2019) pp. 1-11.

[11] M. V. Berry, *Riemann zeros in radiation patterns: II. Fourier transforms of zeta*, J. Phys. A: Math. Theor. **48** (2015)

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

[13] M. V. Berry and J. P. Keating, *The Riemann zeros and eigenvalue asymptotics*, SIAM Review **41** (1999) pp. 236-266.

[14] Philippe Blanc, *A new reason for doubting the Riemann Hypothesis*, Exp. Math. **31** no. 1, (2022) pp. 88-92. 10.1080/10586458.2019.1586600

[15] Philippe Blanc, *An unexpected property of odd order derivatives of Hardy's function*, Publ. Inst. Math. (Beograd) (N.S.) **95** no. 109, (2014) pp. 173-188.

[16] Philippe Blanc, *Optimal upper bound for the maximum of the k-th derivative of Hardy's function*, JNT **154** (2015) pp. 105-117.

[17] Jonathan Bober and Ghaith Hiary, *New computations of the Riemann zeta function on the critical line*, Exp. Math. **27** no. 2, (2018) pp. 125-137.

[18] Andriy Bondarenko and Kristian Seip, *Extreme values of the Riemann zeta function and its argument*, Math. Ann. **372** no. 3-4, (2018) pp. 999-1015.

[19] E. Bombieri, *Problems of the Millennium: the Riemann Hypothesis*, https://www.claymath.org/wp-content/uploads/2022/05/riemann.pdf;

[20] E. Bombieri and D. Hejhal, *On the distribution of zeros of linear combinations of Euler products*, Duke Math. J. **80** no. 3, (1995) pp. 821-862.

[21] Andrew Booker, *Artin's Conjecture, Turing's Method, and the Riemann Hypothesis*, Experiment. Math. **15** no. 4, (2006) pp. 385-408.

[22] Andrew Booker and Frank Thorne, *Zeros of L-functions outside the critical strip*, Algebra Number Theory **8** no. 9, (2014) pp. 2027-2042.

[23] Richard P. Brent, *On the zeros of the Riemann zeta function in the critical strip*, Math. Comp. **33** no. 148, (1979) pp. 1361-1372.

[24] J. B. Conrey, *More than two fifths of the zeros of the Riemann zeta function are on the critical line*, J. Reine Angew. Math. **399** (1986) pp. 1-26.

[25] J. B. Conrey and D. W. Farmer, *Mean values of L-functions and symmetry*, Internat. Math. Res. Notices (2000) **17** pp. 883-908.

[26] J. B. Conrey, D. W. Farmer, J. P. Keating, M. O. Rubinstein, and N. C. Snaith, *Autocorrelation of random matrix polynomials*, Commun. Math. Phys **237** (2003) no. 3, pp. 365-395.

[27] J. B. Conrey, D. W. Farmer, and M. R. Zirnbauer, *Autocorrelation of ratios of L-functions*, Commun. Number Theory Phys **2** (2008) no. 3, pp. 593-636.

[28] J. B. Conrey and A. Ghosh, *Mean values of the Riemann zeta-function*, Mathematika **31** (1984) pp. 159-161.

[29] J. B. Conrey and A. Ghosh, *A conjecture for the sixth power moment of the Riemann zeta-function*, Int. Math. Res. Not. **15** (1998) pp. 775-780.

[30] J. B. Conrey and A. Ghosh, *On the Selberg class of Dirichlet series: small degrees*, Duke Math. J. **72** (1993) pp. 673-693.

[31] J. B. Conrey and S. M. Gonek, *High moments of the Riemann zeta-*

function, Duke Math. Jour. (2001) **107** pp. 577-604.

[32] J. B. Conrey and H. Iwaniec, *Spacing of zeros of Hecke L-functions and the class number problem*, Acta Arith. (2002) **103** no. 3, pp. 259-312.

[33] J. B. Conrey and N. C. Snaith, *Applications of the L-functions ratios conjectures*, Proc. Lond. Math. Soc. (2007) **94** no. 3, pp. 594-646.

[34] P. Diaconis and M. Shahshahani, *On the eigenvalues of random matrices*, J. Appl. Probab. **31A** (1994) pp. 49-62.

[35] A. Diaconu, *On the third moment of $L(\frac{1}{2}, \chi_d)$ I: The rational function field case*, J. Number Theory. **198** (2019) pp. 1-42.

[36] A. Diaconu, D. Goldfeld, and J. Hoffstein, *Multiple Dirichlet series and moments of zeta- and L-functions*, Compositio Math. **139** (2003) no. 3, pp. 297-360.

[37] Alexander Dobner, *A proof of Newman's conjecture for the extended Selberg class*, Acta Arith. **201** no. 1, pp. 29-62. (2021)

[38] Freeman Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, J. Mathematical Phys. **3** pp. 1191-1198. (1962)

[39] Ioana Dumitriu and Alan Edelman, *Matrix models for beta ensembles*, J. Math. Phys. **43** no. 11, pp. 5830-5847. (2002)

[40] David W. Farmer, *Long mollifiers of the Riemann zeta-function*, Mathematika **40** no. 1, pp. 71-87. (1993)

[41] David W. Farmer, *Jensen polynomials are not a plausible route to proving the Riemann Hypothesis*, Adv. Math. **411** (2022) 14 pages. Paper No. 108781 https://arxiv.org/abs/2008.07206;

[42] David W. Farmer, *When are the zeros of a polynomial distinct and real: a graphical view*, Preprint (2021) https://arxiv.org/abs/2010.15608;

[43] D.W. Farmer, S.M. Gonek, and C.P. Hughes, *The maximum size of L-functions*, J. Reine Angew. Math. **609** (2007) pp. 215-236.

[44] David W. Farmer, Francesco Mezzadri, and Nina Snaith, *Random polynomials, random matrices and L-functions*, Nonlinearity **19** (2006) no. 4, pp. 919-936.

[45] David W. Farmer, Ameya Pitale, Nathan Ryan, and Ralf Schmidt, *Analytic L-functions: definitions, theorems, and connections*, Bull. Amer. Math. Soc. (N.S.) **56** (2019) no. 2, pp. 261-280.

[46] David W. Farmer and Robert Rhoades, *Differentiation evens out zero spacings*, Trans. Amer. Math. Soc. **357** (2005) no. 9, pp. 3789-3811.

[47] David W. Farmer, Sally Koutsoliotas, Stefan Lemurell, and David P. Roberts, *Patterns in some landscapes of L-functions*, preprint;

[48] Shaoji Feng, *Zeros of the Riemann zeta function on the critical line*, J. Number Theory **132** (2012) no. 4, pp. 511-542.

[49] Yan Fyodorov and Jonathan Keating, *Freezing transition and extreme values: random matrix theory, $\zeta(\frac{1}{2} + it)$ and disordered landscapes*, Phil. Trans. R. Soc. A **372** (2014) no. 20120503, 32.

[50] Yan Fyodorov, Sven Gnutzmann, and Jonathan Keating, *Extreme values of CUE characteristic polynomials: a numerical study*, J. Phys. A **51** (2018) no. 46, 22.

[51] Yan Fyodorov, Ghaith Hiary, and Jonathan Keating, *Freezing Transition, Characteristic Polynomials of Random Matrices, and the Riemann Zeta-*

Function, Phys. Rev. Lett. **108** (2012) no. 170601,

[52] Gonek, S. M.; Hughes, C. P.; Keating, J. P., *A hybrid Euler-Hadamard product for the Riemann zeta function*, Duke Math. J. **136** no. 3, (2007) pp. 507-549.

[53] Xavier Gourdon, *Computation of zeros of the Zeta function*, preprint; http://numbers.computation.free.fr/Constants/Miscellaneous/zetazeroscompute.html (2004)

[54] J. Gram, *Sur les zéros de la fonction $\zeta(s)$ de Riemann*, Acta Math **27** (1903) pp. 289-304.

[55] Cătălin Hanga and Christopher Hughes, *Probabilistic models for Gram's Law*, preprint; arXiv: 1911.03190; (2020)

[56] G. H. Hardy and J. E. Littlewood, *Contributions to the theory of the Riemann zeta-function and the theory of the distribution of primes*, (1918) Acta Mathematica pp. 119-196. **41**

[57] A. Harper, *On the partition function of the Riemann zeta function, and the Fyodorov--Hiary--Keating conjecture*, preprint; arXiv: 1906.05783; (2019)

[58] Dennis Hejhal, *On the distribution of zeros of a certain class of Dirichlet series*, Internat. Math. Res. Notices no. 4, (1992) pp. 83-91.

[59] Ghaith Hiary, *Personal website*, https://people.math.osu.edu/hiary.1/ fastmethods.html (Accessed March 3, 2022)

[60] Ghaith Hiary and Andrew Odlyzko, *The zeta function on the critical line: numerical evidence for moments and random matrix theory models*, Math. Comp. **81** no. 279, (2012) pp. 1723-1752.

[61] A. Huckleberry, A. Püttmann, and M. R. Zirnbauer, *Haar expectations of ratios of random characteristic polynomials*, Complex Anal. Synerg. no. 1, (2016) **2** Paper No. 1, 73 pp..

[62] C.P. Hughes, J.P. Keating, and Neil O'Connell, *Random matrix theory and the derivative of the Riemann zeta function*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. no. 2003, **456** (2000) pp. 2611-2627.

[63] J. I. Hutchinson, *On the roots of the Riemann zeta-function*, Trans. Amer. Math. Soc. no. 27, (1925) pp. 49-60.

[64] A. E. Ingham, *Mean-value theorems in the theory of the Riemann zeta-function*, Proceedings of the London Mathematical Society no. 92, (1926) **27** pp. 273-300.

[65] A. Ivić, *On some results concerning the Riemann Hypothesis*, *Analytic Number Theory* LMS LNS no. 247, Cambridge University Press; Y. Motohashi, (1997) pp. 139-167.

[66] A. Ivić, *On some reasons for doubting the Riemann Hypothesis*, preprint; arXiv: 0311162; (2003)

[67] Tadej Kotnik, *Computational estimation of the order of $\zeta(1/2 + it)$*, Math. Comp. **73** no. 246, (2004) pp. 949-956.

[68] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, AMS Colloquium Publications **45** AMS, Providence, RI; (1999)

[69] J. P. Keating and N. C. Snaith, *Random matrix theory and $\zeta(\frac{1}{2} + it)$*, Comm. Math. Phys. **214** (2000) pp. 57-89.

[70] J. P. Keating and N. C. Snaith, *Random matrix theory and L-functions at $s = \frac{1}{2}$*, Comm. Math. Phys. **214** (2000) pp. 91-110.

[71] Haseo Ki, *The Riemann Ξi-function under repeated differentiation*, J. Number Theory **120** no. 1, (2006) pp. 120-131.

[72] Rowan Killip and Irina Nenciu, *Matrix models for circular ensembles.*, Int. Math. Res. Not. no. 50, (2004) pp. 2665-2701.

[73] Norman Levinson, *More than one third of zeros of Riemann's zeta-function are on $\sigma = 1/2$*, Advances in Math. **13** (1974) pp. 383-436.

[74] Meghann Moriah Lugar, Micah B. Milinovich, and Emily Quesada-Herrera, *On the number variance of zeta zeros and a conjecture of Berry*, Mathematika **69** 2 arXiv 2211.14918; (2023)

[75] M. Mehta, *Random Matrix Theory*, Academic Press. Boston; (1991)

[76] Francesco Mezzadri, *How to generate random matrices from the classical compact groups*, Notices Amer. Math. Soc. **54** no. 5, (2007) pp. 592-604.

[77] H.L. Montgomery, *The pair correlation of zeros of the Riemann zeta-function*, Proc. Symp. Pure Math. **24** (1973) pp. 81-93.

[78] Charles M. Newman, *Fourier transforms with only real zeros*, Proc. Amer. Math. Soc. **61** no. 2, (1977) pp. 245-251.

[79] A. Odlyzko, *The 10^{20}th zero of the Riemann zeta-function and 70 million of its neighbors.*, Preprint (1989)

[80] A. Odlyzko, *Zeros number $10^{12} + 1$ through $10^{12} + 10^4$ of the Riemann zeta function*, available at http://www.dtc.umn.edu/~odlyzko/zeta_tables;

[81] A. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Math. Comp. **48** (1987) pp. 273-308.

[82] Dave Platt and Tim Trudgian, *The Riemann hypothesis is true up to $3 \cdot 10^{12}$*, Bull. Lond. Math. Soc. **53** no. 3, (2021) pp. 792-797.

[83] Kyle Pratt, Nicolas Robles, Alexandru Zaharescu, and Dirk Zeindler, *More than five-twelfths of the zeros of ζ are on the critical line*, Res. Math. Sci. **7** no. 2, (2020) paper no. 2, 74.

[84] The PreTeXt project, *The PreTeXt document authoring system*, https://pretextbook.org

[85] Brad Rodgers and Terence Tao, *The de Bruijn-Newman constant is non-negative.*, Forum Math. Pi no. e6, **8** (2020) 62.

[86] Mattia Righetti, *Zeros of combinations of Euler products for $\sigma > 1$*, Monatsh. Math. **180** no. 2, (2016) pp. 337-356.

[87] Michael Rubinstein, *Low-lying zeros of L-functions and random matrix theory*, Duke Math. J. **109** no. 1, (2001) pp. 147-181.

[88] Michael Rubinstein, *Computational methods and experiments in analytic number theory*, *Recent perspectives in random matrix theory and number theory, London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press; no. 322, (2005) pp. 425-506.

[89] Z. Rudnick and P. Sarnak, *Zeros of principal L-functions and random matrix theory*, Duke Math. J. **81** (1996) pp. 269-322.

[90] Eric Saias and Andreas J. Weingartner, *Zeros of Dirichlet series with periodic coefficients*, Acta Arithmetica no. 4, **140** (2008) pp. 335-344.

[91] E. Saksman and C. Webb, *The Riemann zeta function and Gaussian*

multiplicative chaos: statistics on the critical line, Ann. Probab. no. 2, **48** (2020) pp. 2680-2754.

[**92**] E. Saksman and C. Webb, *On the Riemann zeta function and Gaussian multiplicative chaos*, Advancements in complex analysis -- from theory to practice Springer; (2020) pp. 473-496.

[**93**] Will Sawin, *A refined random matrix model for function field L-functions*, Preprint (2024) https://arxiv.org/abs/2409.02876;

[**94**] A. Selberg, *Contributions to the theory of the Riemann zeta-function*, Arch. Math. Naturvid. **48** no. 5, (1946) pp. 89-155.

[**95**] A. Selberg, *Old and new conjectures and results about a class of Dirichlet series*, *Proceedings of the Amalfi Conference on Analytic Number Theory (Maiori, 1989)* Univ. Salerno, Salerno; (1992)

[**96**] O. Shanker, *Good-to-bad Gram point ratio for Riemann zeta-function*, Experimental Math. **30** no. 1, (2021) pp. 76-85.

[**97**] K. Soundararajan, *The distribution of values of zeta and L-functions*, ICM--International Congress of Mathematicians. Vol. 2. Plenary lectures pp. 1260-1310. (2023) arXiv 2112.03389;

[**98**] A. Speiser, *Geometrisches zur Riemannschen Zetafunktion*, Math Ann. **110** (1935) pp. 514-521.

[**99**] Norbert Tihanyi, *Fast method for locating peak values of the Riemann zeta function on the critical line*, 2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing IEEE; (2014) pp. 91-96.

[**100**] Norbert Tihanyi, Atilla Kovács, and József Kovács, *Computing extremely large values of the Riemann zeta function*, J. Grid Computing **15** (2017) pp. 527-534.

[**101**] E. C. Titchmarsh, *The theory of the Riemann zeta-function. Second edition. Edited and with a preface by D. R. Heath-Brown*, The Clarendon Press, Oxford University Press, New York; (1986)

[**102**] T. Trudgian, *On the success and failure of Gram's law and the Rosser rule*, Acta Arith. **148** (2011) pp. 225-256.