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.

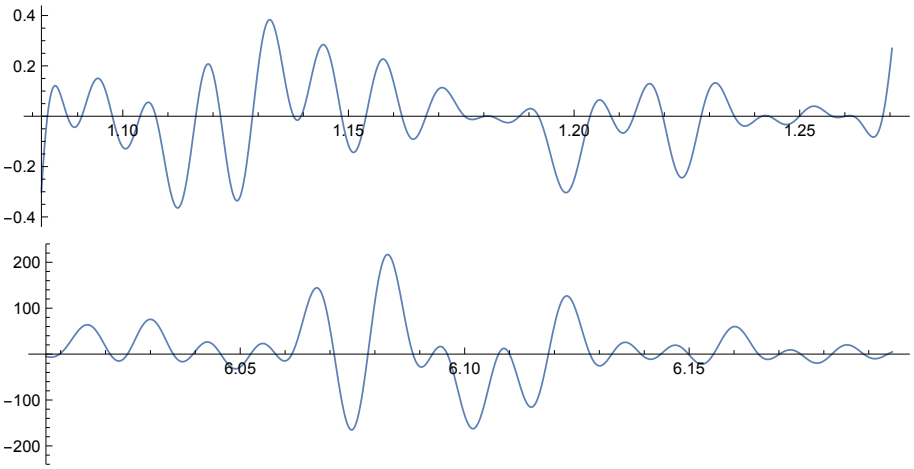

Figure 8.1 Graphs of $\mathcal{Z}_B(\theta)$ over two intervals of width $2\pi \times 0.03$.

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.

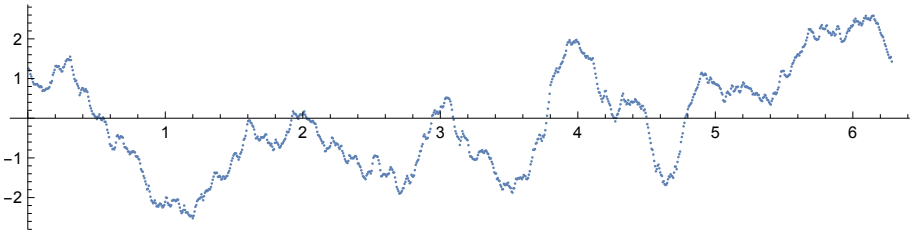

Figure 8.2 The carrier wave for $\mathcal{Z}_B(\theta)$, on a logarithmic scale, calculated using $K = 10$ in (5.3).

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 **jth 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 $\cdots \le \gamma_{-2} \le \gamma_{-1} < 0 < \gamma_1 \le \gamma_2 \le \cdots$.

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*

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

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.

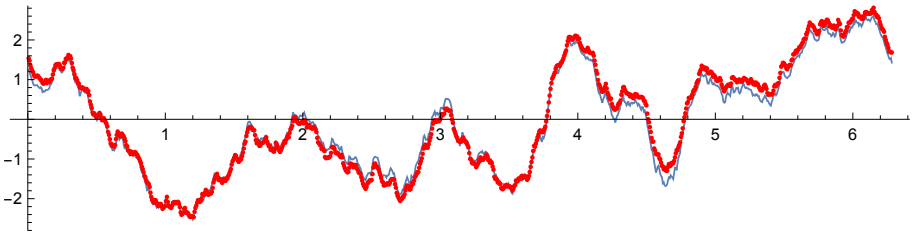

Figure 8.5 The thicker red graph is the local density of zeros (the density wave) measured by (8.2) with $J = 11$, for the degree 1000 polynomial $\mathcal{Z}_B(\theta)$. The thinner blue graph is the (logarithm of the) carrier wave for the same polynomial, calculated with $K = 10$. The similarity of 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 $\cdots \le \gamma_{-2} \le \gamma_{-1} < 0 < \gamma_1 \le \gamma_2 \le \cdots$. 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

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

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

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

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

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

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

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 $\prod_{j\geq 1}(1 - t^2/(j-\frac{1}{2})^2)$ 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 $\mathcal{Z}_B(\theta)$.

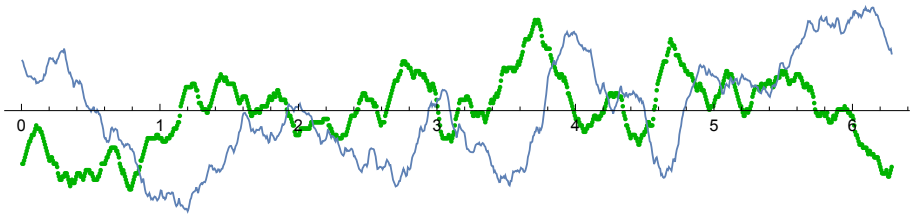

Figure 8.8 The thinner blue graph is the carrier wave of $\mathcal{Z}_B(\theta)$ from Figure 8.2 and the thicker green graph is the $\mathcal{S}_B(\theta)$ for $\mathcal{Z}_B(\theta)$. The graphs illustrate Principle 8.7: $\mathcal{S}_B(\theta)$ is increasing/decreasing in regions where $\mathcal{Z}_B(\theta)$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

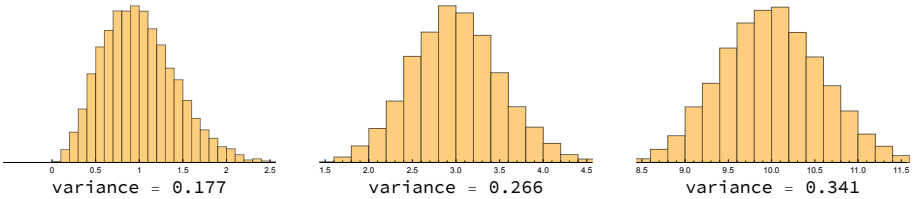

Figure 8.9 The normalized 1st, 3rd, and 10th neighbor spacing for γ_n for $10^{12} \le n < 10^{12} + 10000$, along with the variance of each sample.

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],

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

which is easier to handle theoretically but perhaps more difficult to grasp intuitively.

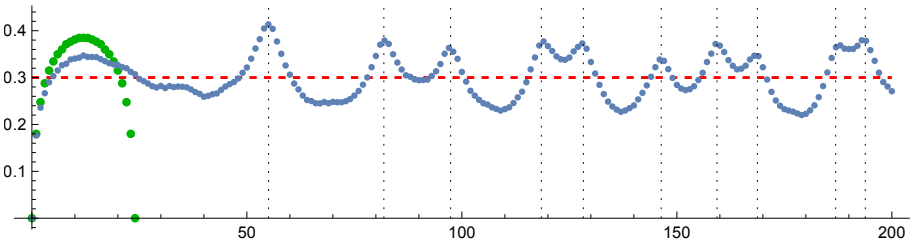

Figure 8.10 The blue dots show the variance of the jth nearest-neighbor spacing, $0 \le j \le 200$, for zeros γ_n with $10^{12} \le n < 10^{12} + 10000$. The height of the horizontal red dashed line is twice the average variance of the samples, which would be the variance of the neighbor gaps if the 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.

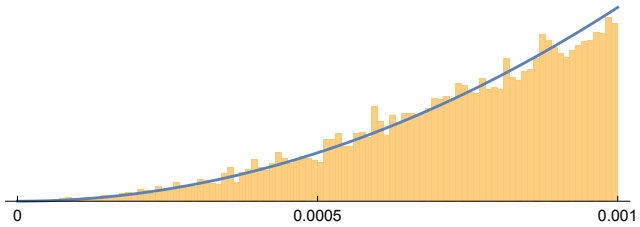

Figure 8.12 Histogram of the 9333 smallest normalized neighbor gaps among the first 10^{13} zeros of the ζ-function, compared to the predicted density $\frac{\pi^2}{3}x^2$. Data from Xavier Gourdon [53].

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:

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

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

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

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 $p_2(x) \sim e^{-\pi^2x^2/8}$ 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.

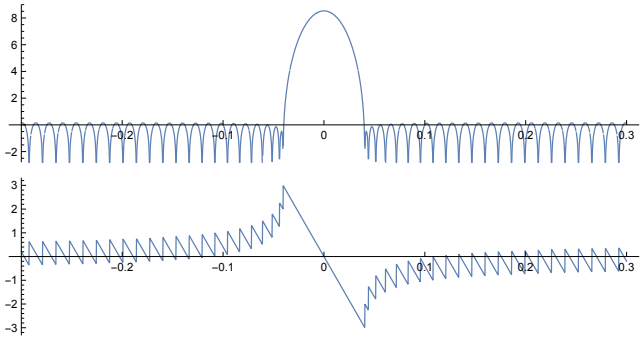

Figure 9.4 Graphs of $\mathcal{Z}(\theta)$ and $\mathcal{S}(\theta)$ for a degree 74 polynomial with a zero gap of 6 times the average and all other zeros in their most likely configuration with respect to Haar measure (6.6).

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.

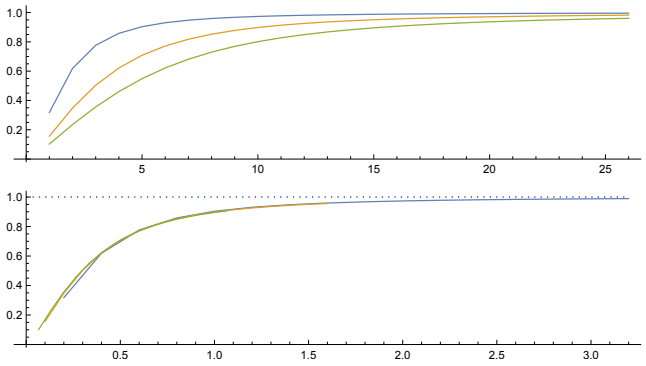

Figure 9.6 The typical normalized gap between successive zeros adjacent to a large gap of 5, 10, or 15 times the average spacing. In the plot on the top the horizontal scale is the index of the gap. On the bottom the horizontal scale is the index divided by the size of the large gap.

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

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

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

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

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

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 $a_0,\dots,a_K$, 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 $\sum_{p\le x} p^{-1} \sim \log\log x$ 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 $0\le\sigma\le 1$ and $|t|\ge 2$, 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

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

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 $T \le t \le 2T$, 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 $A \in U(N)$:

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

It is a theorem [34] that for $1 \le n \le k$, 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:

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

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

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

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** $\zeta_{\text{rand}}(\frac{1}{2} + ix)$, where

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

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 $\zeta_{\text{rand}}(\frac{1}{2} + ix)$ is convergence in law with respect to the strong topology of the Sobolev space $W^{-\alpha,2}(0,1)$ for any $\alpha > \frac{1}{2}$. Furthermore, we have

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

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 $\log|\zeta(\frac{1}{2} + it)|$ 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 $\kappa < \frac{1}{4}$.

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

$$\log N - \frac{2\log Y + \log\log N}{4}. \tag{11.12}$$

Equating to (11.5) with $C = \frac{3}{4}$ 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

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

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:

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

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

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

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

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

where

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

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.

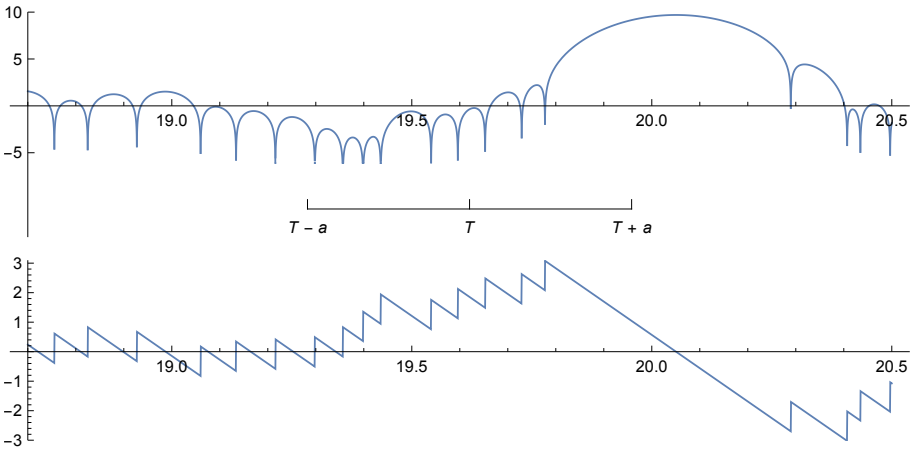

Figure 12.4 The function $\log|Z(t)|$ near a local extremum caused by a large zero gap, and $S(t)$ in the same region. The parameters T and a are chosen so that $Z'(T+a)$ is a (large) local extremum, $Z'(T-a)$ is a (presumably small) local extremum, and $S(T+a)$ and $S(T-a)$ have opposite sign.

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*

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

then

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

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

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

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 $\mathcal{Z}^{2K}(z) := D^{2K}\mathcal{Z}(z)$,

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

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

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

□

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:

$$\int_0^T Z(t)^2\,dt = T\log T + (2\gamma - 1 - \log 2\pi)T + O(T^{E_1+\varepsilon})$$
$$\int_0^T Z(t)^4\,dt = TP_2(\log T) + O(T^{E_2+\varepsilon})$$

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:

$$\int_0^T Z(t)^{2k}\,dt = TP_k(\log T) + O(T^{E_k+\varepsilon}) \tag{13.1}$$
$$\sim g_k a_k T \log^{k^2} T, \tag{13.2}$$

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 $k \geq 3$.