The Legacy of Bernhard Riemann
After One Hundred and Fifty Years
ALM 35, pp. 69–78

Riemann's Saddle-point Method and the Riemann-Siegel Formula

M. V. Berry*

Abstract

Riemann's way to calculate his zeta function on the critical line was based on an application of his saddle-point technique for approximating integrals that seems astonishing even today. His contour integral for the remainder in the Dirichlet series for the zeta function involved not an isolated saddle, nor a saddle near a pole or an end-point or several coalescing saddles, but the configuration, unfamiliar even now, of a saddle close to an infinite string of poles. Riemann evaluated the associated integral exactly, and the resulting Riemann-Siegel formula underlies ways of computing the Riemann zeros and one of the physical approaches to the Riemann hypothesis.

2000 Mathematics Subject Classification: 01A55, 11M26, 11Y35, 30B50, 30E15, 34E05, 41A60.

Keywords and Phrases: zeta function, steepest descent, asymptotics.

Contents

*H H Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, UK.

1 Introduction

This book marks Riemann's death 150 years ago by celebrating his many achievements. I want to focus on one startlingly original aspect of his research that was never published in his lifetime, namely his application of the saddle-point method to the calculation of his eponymous zeta function. The underlying story is well known [1]. Nearly seventy years after Riemann died, Siegel [2] (reprinted in [3]) reconstructed his calculation from incomplete notes discovered posthumously, and elucidated what is now called the Riemann-Siegel formula.

The formula involves a contour integral, that Riemann approximated by a variant of his saddle-point method. In its simplest and most familiar form [4, 5], this involves integrands dominated by an exponential containing a large parameter. The approximation consists in expanding the integrand about its critical point on the integration path – the complex saddle (stationary point of the exponent) - leading in lowest order to a Gaussian integral that is easily evaluated. This version of the method is usually attributed to Debye [6], although he pointed out that he learned it from an 1863 paper by Riemann, published posthumously [7] with additional material by Schwarz. Riemann had concentrated on the approximation of certain hypergeometric functions with complex variables; this seems a specialised application, but it was clear that Riemann understood that the technique can be applied in more general cases. Between Riemann's and Debye's paper, the method was discovered independently by Nekrasov [8].

Sometimes, the technique is called the method of steepest descents This refers to rotating the integration contour to pass through the saddle in the direction for which the real part of the exponent decreases fastest. If instead the contour is rotated (by $\pi/4$ at the saddle) so that the real part of the exponent remains constant – that is, the integrand is oscillatory – the same technique is called the stationary-phase method. In this form, it was anticipated by Stokes [9] (reprinted in [10]) and later developed by Kelvin [11].

Nowadays, we are familiar with a number of generalizations of the saddle-point method [4]: as one or more parameters are varied, the saddle can coincide with an end-point of the integral, or a pole or branch-point, or several saddles can coalesce (chapter 36 of [12]). In these variants, the approximation depends on identifying a special function for which the local behaviour can be captured exactly. For the simplest saddle-point method, this is the Gaussian integral; for the saddle near an end-point or a pole, it is the error function; for two coalescing saddles, it is the Airy function [13]. Further extensions include deforming the integrands to get approximations that remain uniformly valid into the ordinary saddle-point regime far from the coalescences [14–16], understanding the high orders of the expansions [17, 18], and multiple integrals [4, 19].

In Riemann's approximation of the zeta function, his genius was to apply the saddle-point method to a situation that to my knowledge nobody else has considered in the intervening 150 years: a saddle close to an infinite string of poles. To explain this, it is necessary to reproduce previously published material [1, 2]. I will do so in a streamlined way ('Siegel-Edwards lite').

2 Outline derivation of Riemann-Siegel remainder integral

Riemann approximated $\zeta(s)$ high on the critical line $s = 1/2 + \mathrm{i}t$, i.e. t real and $t >> 1$. For this is is necessary to start by analytically continuing the Dirichlet series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \tag{2.1}$$

which converges only for $\mathrm{Re}s > 1$. On the critical line, it is convenient to approximate not the complex function $\zeta(1/2 + \mathrm{i}t)$ but the function $Z(t)$, which the functional equation for $\zeta(s)$ guarantees is real for real t:

$$Z(t) = \exp(\mathrm{i}\theta(t))\zeta\left(\frac{1}{2} + \mathrm{i}t\right), \text{ where } \theta(t) = \arg\left(\Gamma\left(\frac{1}{4} + \frac{1}{2}\mathrm{i}t\right)\right) - \frac{1}{2}t\log\pi. \tag{2.2}$$

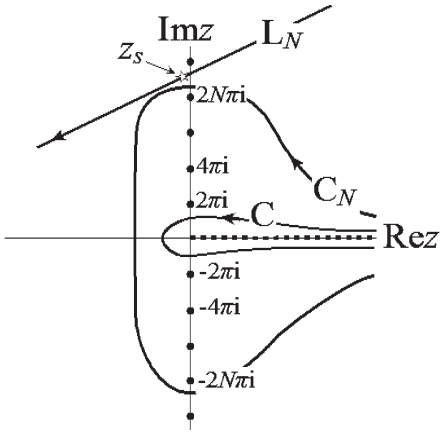

Figure 1: Complex plane z of the integrand, and integration contours, in the Riemann-Siegel integrals. Dots indicate the poles, the star indicates the saddle, and the dashed line is the branch cut.

In the Dirichlet series (2.1), the first N terms are retained, and the continuation is accomplished modifyng the terms $n > N$ using the following form of the Hankel integral [12] for the gamma function:

$$1 = -\frac{\Gamma(s)}{2\pi\mathrm{i}} \int_C \mathrm{d}z \frac{\exp(-z)}{(-z)^s} \quad (s \neq 0, -1, -2\cdots). \tag{2.3}$$

Here the z plane is cut along the positive real axis and the contour is shown in figure 1. Elementary manipulations now give a representation of $Z(t)$ in which the tail of the Dirichlet series is resummed:

$$\begin{aligned} Z(t) &= \sum_{n=1}^{N} \frac{\exp(\mathrm{i}(\theta(t) - t\log n))}{\sqrt{n}} \\ &\quad - \frac{\exp(\mathrm{i}\theta(t))\Gamma\left(\frac{1}{2} - \mathrm{i}t\right)}{2\pi\mathrm{i}} \int_C \mathrm{d}z \frac{(-z)^{-\frac{1}{2}+\mathrm{i}t}\exp(-Nz)}{\exp(z) - 1}. \end{aligned} \tag{2.4}$$

This is the required analytic continuation.

The next step is to expand the contour to C_N (figure 1), to capture the first $N > 0$ poles on both the positive imaginary and the negative imaginary axes. This leads to

$$\begin{aligned} Z(t) &= \sum_{n=1}^{N} \frac{\exp(\mathrm{i}(\theta(t) - t\log n))}{\sqrt{n}} \\ &\quad + f(t)\sum_{n=1}^{N} \frac{\exp(\mathrm{i}(-\theta(t) + t\log n))}{\sqrt{n}} + R_N(t), \end{aligned} \tag{2.5}$$

with the remainder now given by

$$R_N(t) = -\frac{\exp(\mathrm{i}\theta(t))\Gamma\left(\frac{1}{2} - \mathrm{i}t\right)}{2\pi\mathrm{i}} \int_{C_N} \mathrm{d}z \frac{(-z)^{-\frac{1}{2}+\mathrm{i}t}\exp(-Nz)}{\exp(z) - 1}. \tag{2.6}$$

Gamma function manipulations (reflection and duplication formulas [12]) give the prefactor of the second sum as

$$f(t) = 2\exp\left(2\mathrm{i}\theta(t)\right)\Gamma\left(\frac{1}{2} - \mathrm{i}t\right)(2\pi)^{-\frac{1}{2}+\mathrm{i}t}\cos\left(\frac{1}{2}\pi\left(\frac{1}{2} - \mathrm{i}t\right)\right) = 1. \tag{2.7}$$

Therefore the N terms of the second series in (2.5) are the complex conjugates of their counterparts in the first series, and

$$Z(t) = 2\sum_{n=1}^{N} \frac{\cos\left(\theta(t) - t\log n\right)}{\sqrt{n}} + R_N(t). \tag{2.8}$$

This representation, exact for all values of t and N, is the starting point of the derivation of the Riemann-Siegel formula. The series (called 'the main sum', or 'the approximate functional equation') is real, and so is $Z(t)$. Therefore $R_N(t)$ is also real, though this is not obvious from the expression (2.6). The reappearance, in the tail of the Dirichlet series, of the complex conjugates of the first N terms of the series, is an example of the general asymptotic phenomenon of resurgence, in which the high orders of a divergent series can be expressed in terms of the low orders of the series [17, 20, 21]. The main sum itself is quite accurate: even the term $n = 1$ possesses zeros with the correct asymptotic densities, and the remaining $N - 1$ terms shift the approximate zeros close to their exact locations.

3 Leading-order remainder

Now the theory moves from exactness to approximation for large t. The exponent in the numerator of the integrand in (2.6) possesses a saddle, at z_s, given by

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\left(-\frac{1}{2} + \mathrm{i}t\right)\log z - Nz\right) = 0 \quad \Rightarrow z = z_s = \frac{\mathrm{i}t}{N} - \frac{1}{2N}. \tag{3.1}$$

We want this saddle to lie close to the contour C_N, so that it dominates the integral for R_N. This can be achieved by choosing

$$N = \left[\sqrt{\frac{t}{2\pi}}\right], \tag{3.2}$$

in which [...] denotes the integer part (floor function). Denoting the fractional part of $\sqrt{(t/2\pi)}$ by p, i.e.

$$\sqrt{\frac{t}{2\pi}} = N + p \quad (0 \le p < 1), \tag{3.3}$$

the location of the saddle can be written as

$$z_s = \frac{2\pi\mathrm{i}\,(N+p)^2}{N} - \frac{1}{2N} \to 2\pi\mathrm{i}\,(N+2p) \text{ as } N \to \infty. \tag{3.4}$$

As N increases, z_s approaches the imaginary axis.

Figures 2(a,b) show the normalized modulus of the numerator of the integrand in (2.6) along the straight line through the saddle in the direction of steepest descent there, namely

$$M(x) = \left|\left(\frac{z}{z_0}\right)^{-\frac{1}{2}+\mathrm{i}t} \exp\left(-N\,(z-z_0)\right)\right|, \text{ for } z = z_0 + x\exp\left(\frac{1}{4}\mathrm{i}\pi\right), \tag{3.5}$$

where $z_0 = z_s$, compared with the Gaussian approximation (quadratic expansion about z_0). The Gaussian fit is almost perfect, even for the relatively small value illustrated.

A point not emphasized in the Riemann-Siegel literature is that it is only for $0 < p < 1/2$ that the large N location (3.4) of the saddle lies between the poles N and $N+1$, so that C_N can pass through it; for $1/2 < p < 1$, the saddle lies above the pole $N+1$. But this does not matter: it is not necessary for the C_N to pass exactly through the saddle. In fact it is convenient to expand the integral not through z_s but through the location z_c of the pole N:

$$z_c = 2\pi\mathrm{i}N = z_s - 4\pi\mathrm{i}p \text{ for } N >> 1. \tag{3.6}$$

As figures 2(c,d) illustrate, the modulus $M$of the integrand (i.e. (3.5) with $z_0 = z_c$, is still close to Gaussian. The maximum is shifted, and the fit for fixed N deteriorates as p increases, because the path passes further from the saddle. traversing regions of ascent where M is large. For all p, the fit improves as N increases.

Expanding about z_c to quadratic order, the remainder becomes, after more manipulations and using the large t (Stirling) approximation for the gamma function [1],

$$R_N\left(2\pi\,(N+p)^2\right) \underset{N>>1}{\to} \left(\frac{2\pi}{t}\right)^{1/4} (-1)^{N+1}\,\Psi(p), \tag{3.7}$$

in which

$$\Psi(p) = \frac{\exp\left(\mathrm{i}\left(\frac{1}{8}\pi - 2\pi p^2\right)\right)}{2\pi\mathrm{i}} \int_L \mathrm{d}u \frac{\exp\left(\frac{\mathrm{i}u^2}{4\pi} + 2pu\right)}{\exp(u) - 1}. \tag{3.8}$$

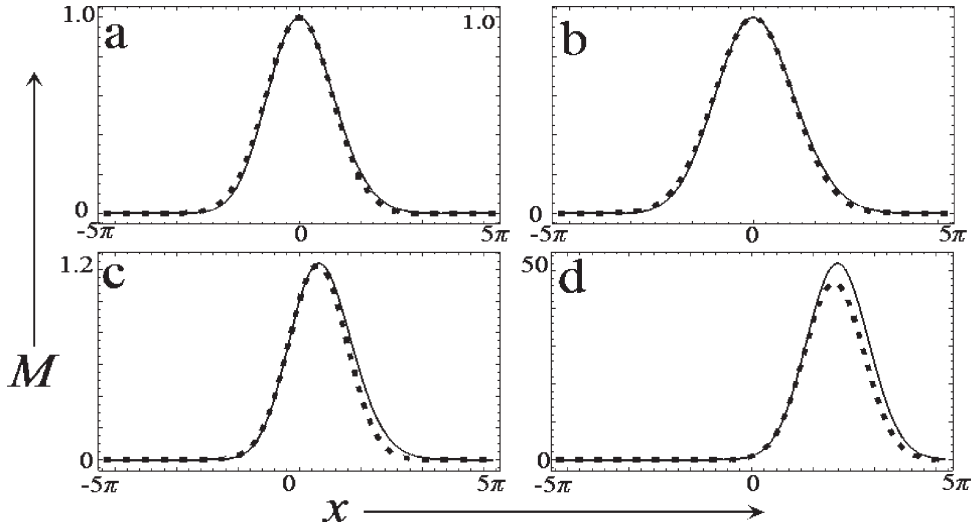

Figure 2: Full curves: (a,b) modulus M (equation (3.5)) of the Riemann-Siegel integral (2.6) along straight 45° (locally steepest) path L_N through the saddle z_s (equation (3.4)), with $N = 3$, for (a): $p = 0.2$ and (b): $p = 0.8$. (c,d): as (a), (b) for the path through z_c (equation (3.6)) crossing the pole N; for (c), $N = 3$, and for (d), $N = 20$. Dashed curves: corresponding Gaussian approximations.

The contour L passes from upper right to lower left, crossing the imaginary axis between $u = 0$ and $u = 2\pi\mathrm{i}$. Careful analysis [1] shows that the segments of the contour C_N thus neglected give contributions negligible in comparison with that from L.

Thus Riemann identified the dominant contribution to $R_N(t)$ as an integral whose numerator is Gaussian and whose denominator is a string of poles. As explained elsewhere [1, 2], he was able to evaluate this integral exactly. The result is

$$\Psi(p) = \frac{\cos 2\pi \left(p^2 - p - \frac{1}{16}\right)}{\cos 2\pi p} \tag{3.9}$$

(the derivation involves two relations between $\Psi(p)$ and $\Psi(p + 1/2)$). Despite appearances, this is an entire function: the zeros $p = (n+1/2)\pi$ of the denominator are cancelled by those of the numerator.

Thus, almost casually, without fanfare, in an achievement unmatched from his day to ours, Riemann established the leading-order correction to the main sum in (2.8), so that

$$\begin{aligned} Z(t) &= 2\sum_{n=1}^{N} \frac{\cos\left(\theta\left(t\right) - t\log n\right)}{\sqrt{n}} \\ &\quad + \left(\frac{2\pi}{t}\right)^{1/4} (-1)^{N+1} \frac{\cos 2\pi \left(p^2 - p - \frac{1}{16}\right)}{\cos 2\pi p} + O\left(\frac{1}{t^{1/2}}\right). \end{aligned} \tag{3.10}$$

Note that this expression is real, indicating that the saddle-point approximation preserves the essence of the functional equation.

4 Higher orders and concluding remarks

Riemann did more than calculate the leading-order remainder: he developed a systematic expansion scheme, to derive higher orders as a series in powers of $(2\pi/t)^{1/4}$, involving derivatives of $\Psi(p)$. I do not describe these higher corrections here, because there is a more direct procedure, stimulated by an idea of Keating [22] and elaborated elsewhere [23]. This is based directly on the Dirichlet series (2.1), from which the remainder, correcting the main sum, is given formally by

$$R_N(t) = \exp\left(\mathrm{i}\theta(t)\right) \sum_{N+1}^{\infty} \frac{1}{n^{\frac{1}{2}+\mathrm{i}t}} - \exp\left(-\mathrm{i}\theta(t)\right) \sum_{1}^{N} \frac{1}{n^{\frac{1}{2}-\mathrm{i}t}}. \tag{4.1}$$

The procedure consists of expanding each sum about its limit N, and also $\theta(t)$, for $t >> 1$. The resulting terms are all real, as they must be, and coincide with those obtained rigorously by Riemann and Siegel. Moreover, the formal procedure enables the high-order behaviour of the series to be established, indicating that it diverges factorially, although in a slightly unfamiliar way [23].

In the main sum (2.8), the upper limit N depends on t according to (3.2). Therefore the main sum is a discontinuous function of t. But the exact function $Z(t)$ is continuous and also smooth, so one role of the Riemann-Siegel correction terms is to systematically reduce the discontinuities in the value and derivatives of the main sum. There are several more sophisticated expansions of $Z(t)$ [23-27], in which the discontinuities are smoothed.

All methods currently used for numerical calculations of the zeta function and its zeros are based on the Riemann-Siegel formula. The principal difficulty is computing the main sum, but this has been overcome in several ways [28, 29].

The Riemann-Siegel formula has a physical interpretation. One approach to the Riemann hypothesis is based on the conjecture that $Z(t)$ is the spectral determinant (characteristic polynomial) of a quantum Hamiltonian operator whose classical counterpart is a chaotic dynamical system [30-32]. On this analogy, the heights t of the Riemann zeros correspond to quantum energy levels. This quantum system, and its classical counterpart, have not been identified, but several of their properties are known. In particular, the Riemann-Siegel main sum is the counterpart of an expansion of the quantum spectral determinant as a sum over combinations of periodic orbits of the classical system [30, 33, 34], and the divergence of the series for the remainder $R_N(t)$ suggests the nature of the complex classical periodic orbits [23]. Further insights into the conjectured underlying classical dynamical system might be hidden in the detailed form of the series of Riemann-Siegel corrections, involving $\Psi(t)$.

If Riemann had not left hints of his way of calculating $Z(t)$ in his Nachlass, and if Siegel had not discovered and deciphered Riemann's notes, it is likely that we would still be unaware of their formula today. The Riemann-Siegel formula underlies one approach to the Riemann hypothesis, it is implicated in connections between quantum mechanics chaotic dynamics, and the prime numbers, and it is employed in computations of the Riemann zeros. Riemann's achievement, innovative in so many ways, in particular his application of the saddle-point method to an integral with a string of poles, still seems magical.

References

[1] Edwards, H. M., 2001, *Riemann's Zeta Function*, Dover Publications, Mineola, New York.

[2] Siegel, C. L., 1932, Über Riemanns Nachlass zur analytisches Zahlentheorie *Quellen und Studien zur Geschichte der Mathematik, Astronomie, und Physik* **2**, 45-80.

[3] Siegel, C. L., 1966, *Carl Ludwig Siegel Gesammelte Abhandlungen*, Springer-Verlag, Berlin.

[4] Wong, R., 1989, *Asymptotic approximations to integrals*, Academic Press, New York and London.

[5] de Bruijn, N. G., 1958, *Asymptotic Methods in Analysis*, North-Holland, reprinted by Dover books 1981, Amsterdam.

[6] Debye, P., 1909, Näherungsformeln für die Zylinderfunktionen für grosse Werte des Arguments und unbeschränkt veränderliche Werte des Index *Math. Ann* **67**, 535-558.

[7] Riemann, B., 1863, *Sullo svoglimento del quoziente di due serie ipergeometriche in frazione continua infinita in Complete works, 2nd. ed, pp 424-430* eds., Dover 1963, New York.

[8] Petrova, S. S. & Solov'ev, A. D., 1997, The Origin of the Method of Steepest Descent, *Historia Mathematica* **24**, 361-375.

[9] Stokes, G. G., 1847, On the numerical calculation of a class of definite integrals and infinite series, *Trans. Camb. Phil. Soc.* **9**, 379-407.

[10] Stokes, G. G., 1883, *Mathematical and Physical Papers*, Unversity Press, Cambridge.

[11] Kelvin, L., 1887, On the waves produced by single impulse on water of any depth, or in a dispersive medium, *Philos. Mag.* **23**, 252-255.

[12] DLMF,2010, *NIST Handbook of Mathematical Functions*, Cambridge University Press, Cambridge http://dlmf.nist.gov.

[13] Airy, G. B., 1838, On the intensity of light in the neighbourhood of a caustic, *Trans. Camb. Phil. Soc.* **6**, 379-403.

[14] Chester, C., Friedman, B. & Ursell, F., 1957, An extension of the method of steepest descents, *Proc. Camb. Phil. Soc.* **53**, 599-611.

[15] Duistermaat, J. J., 1974, Oscillatory Integrals, Lagrange Immersions and Unfolding of Singularities, *Communs Pure App. Math.* **27**, 207-281.

[16] Berry, M. V., 1976, Waves and Thom's theorem, *Advances in Physics* **25**, 1-26.

[17] Dingle, R. B., 1973, *Asymptotic Expansions: their Derivation and Interpretation*, Academic Press, New York and London.

[18] Berry, M. V. & Howls, C. J., 1991, Hyperasymptotics for integrals with saddles, *Proc. Roy. Soc. Lond.* **A434**, 657-675.

[19] Howls, C. J., 1997, Hyperasymptotics for multidimensional integrals, exact remainder terms and the global connection problem, *Proc. Roy. Soc. Lond.* **A453**, 2271-2294.

[20] Écalle, J., 1985, *Les fonctions résurgentes (3 volumes).* Pub. math. Orsay

[21] Berry, M. V. & Howls, C. J., 2015, *Divergent series: taming the tails in The Princeton Companion to Applied Mathematics* , ed. Higham, N., Princeton University Press, Princeton, pp634-640

[22] Keating, J. P., 1993, *The Riemann zeta-function and quantum chaology in Quantum Chaos* eds. Casati, G., Guarneri, I. & Smilansky, U., North-Holland, Amsterdam, pp. 145-185.

[23] Berry, M. V., 1995, The Riemann-Siegel formula for the zeta function: high orders and remainders, *Proc.Roy.Soc.Lond.* **A450**, 439-462.

[24] Paris, R. B., 1994, An asymptotic representation for the Riemann zeta function on the critical line, *Proc. Roy. Soc. Lond.* **A446**, 565-587.

[25] Paris, R. B. & Cang, S., 1997, An exponentially-improved Gram-type formula for the Riemann zeta function, *Methods. Appl. Anal.* **4**, 326-338.

[26] Paris, R. B. & Cang, S., 1997, An asymptotic representation for $\zeta(1/2 + \mathrm{i}t)$, *Methods. Appl. Anal.* **4**, 449-470.

[27] Kuznetsov, A., 2007, On he Riemann-Siegel formula, *Proc. R. Soc. A* **463**, 2557-2568.

[28] Odlyzko, A. M. & Schönhage, A., 1988, Fast Algorithms for Multiple Evaluations of the Riemann Zeta Function, *Trans. Amer. Math. Soc.* **309**, 797-809.

[29] Hiary, G. A., 2011, Fast methods to compute the Riemann zeta function, *Ann. Math.* **174**, 891-946.

[30] Berry, M. V., 1986, *Riemann's zeta function: a model for quantum chaos? in Quantum chaos and statistical nuclear physics* eds. Seligman, T. H. & Nishioka, H., Vol. 263, pp. 1-17.

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

[32] Berry, M. V., 2008, Three quantum obsessions, *Nonlinearity* **21**, T19-T26.

[33] Berry, M. V. & Keating, J. P., 1992, A new approximation for $\zeta(1/2 + \mathrm{i}t)$ and quantum spectral determinants, *Proc. Roy. Soc. Lond.* **A437**, 151-173.

[34] Keating, J. P. & Sieber, M., 1994, Calculation of spectral determinants, *Proc. Roy. Soc. Lond.* **A447**, 413-437.