whelm this advantage on the plus side. As Gram puts it, equilibrium between plus and minus values of Re ζ will be achieved† only very slowly as t increases. Thus the positivity of Re ζ at Gram points, and consequently the alternation of Gram points with zeros of Z, is likely to persist for some distance beyond $t = 50$.

To locate the Gram points computationally is quite easy. In the first place, 11 of them can be read off from Table IV, namely, the 11 sign changes of Im ζ which are not roots ρ. Since these points are the points where $\vartheta(t)$ is a multiple of π, since $\vartheta(t)$ is increasing for $t \geq 10$ (at least), and since the Gram point near 18 must surely be the solution of $\vartheta(t) = 0$ [because $\vartheta(18) = 0.08$], it is clear that the last 10 of these 11 Gram points are solutions of the equation

$$\vartheta(t) = n\pi$$

for $n = -1, 0, 1, 2, 3, 4, 5, 6, 7, 8$. [The first one, near 3.4, is also a solution for $n = -1$, as can be seen quite easily by setting $t = 3.4$ in (2).] Thus the first Gram point beyond the range of Table IV is a solution of $\vartheta(t) = 9\pi$. Since $\vartheta(t) = 8\pi$ occurs near $t = 48.7$ and since $\vartheta'(t) \sim \frac{1}{2}\log(t/2\pi) \sim \frac{1}{2}\log 8 = \frac{3}{2}\log 2 = 1.0$ in this region, $\vartheta(t) = 9\pi$ should occur near $48.7 + \pi \sim 51.8$. Now (2) gives $\vartheta(51.8) = 28.344$, whereas $9\pi = 28.274$; thus ϑ is too large by 0.070 at 51.8. In the course of computing $\vartheta(51.8)$ one finds that the derivative $\vartheta'(t) \sim \frac{1}{2}\log(t/2\pi)$ is about 1.05. Therefore t should be decreased to about $51.8 - (0.070/1.05) = 51.8 - 0.0666 \sim 51.734$. This value is correct to three places as can be checked by substituting it into (2) and observing that the result is 9π to three places. In this way any Gram point can be found with any desired degree of accuracy with relatively little computation.

In summary, if the nth Gram point g_n is defined to be the unique real number satisfying $\vartheta(g_n) = n\pi$, $g_n \geq 10$ $(n = 0, 1, 2, \ldots)$, then g_n can be computed and the above arguments give some reason to believe that Re $\zeta(\frac{1}{2} + ig_n)$ will be positive for n well beyond the limit $n = 8$ of Table IV. As long as this remains true it follows that there is at least one root ρ on the line segment from $\frac{1}{2} + ig_{n-1}$ to $\frac{1}{2} + ig_n$ and, in all likelihood, exactly one root.

This program of Gram's was followed by Hutchinson [H11] who computed all the values g_n up to $g_{137} = 300.468$ and determined the sign of Re $\zeta(\frac{1}{2} + ig_n)$ for each of them. He found that there were two exceptions‡ to the rule

†Gram states: "Si cela est juste, on peut inférer que l'équilibre ne s'établira que peu à peu, de sorte que la même régle sur la répartition des α [the roots on the line] par rapport aux γ [the Gram points] se maintiendra aussi pour les α suivantes les plus rapprochées de α_{15}." Actually Gram was wrong in believing that equilibrium would be achieved at all because as Titchmarsh proved [T4], the average value of $\zeta(\frac{1}{2} + ig_n)$ is 2. See the concluding remarks of Section 11.1.

‡According to Haselgrove [H8] this should not have been a surprise since he says that Bohr and Landau proved in 1913 that there are infinitely many exceptions. However, this seems to be an error on Haselgrove's part. Titchmarsh [T5] proved the existence of infinitely many exceptions in 1935 and he, with his extensive knowledge of the literature, believed this to be a new result.

$\operatorname{Re} \zeta(\frac{1}{2} + ig_n) > 0$, namely, $n = 126$ and $n = 134$. He found, moreover, that these exceptions are slight in the sense that the corresponding values of $\operatorname{Re} \zeta$ are only slightly less than zero and that if the points are shifted only a little bit—g_{126} from 282.455 up to 282.6 and g_{134} from 295.584 down to 295.4—then the sign of $\operatorname{Re} \zeta$ becomes positive. It is easily checked that these two shifts do not change the sign of $\cos \vartheta(g_n)$ and therefore by the same argument as before [namely, $\cos \vartheta(g_n)$ alternates in sign, $Z(g_n) \cos \vartheta(g_n)$ is always positive, and therefore $Z(g_n)$ alternates in sign] it follows that *there is at least one root* ρ *on the line segment from* $\frac{1}{2} + ig_{n-1}$ *to* $\frac{1}{2} + ig_n$ $(n = 1, 2, \ldots, 137)$ *when* g_{126} *and* g_{134} *are shifted as above.* This locates at least 138 roots ρ (there is one between $\frac{1}{2}$ and $\frac{1}{2} + ig_0$); and it is at least plausible that each of these segments contains only one so that there are only 138 in all. By methods explained in the next section, Hutchinson was in fact able to show that there are exactly 138 roots ρ in the range $0 \leq \operatorname{Im} s \leq g_{137}$, counted with multiplicities, which proves then that *all roots* ρ *in the range* $0 \leq \operatorname{Im} s \leq 300$ *lie on the line* $\operatorname{Re} s = \frac{1}{2}$ *and all of them are simple zeros of* ξ.

Hutchinson called the tendency of the zeros of Z to alternate with the Gram points g_n *Gram's law.* Since Gram stated only that this pattern would persist beyond $n = 8$ and seemed to doubt that it would persist indefinitely, this is a rather poor choice of terminology in that a "law" is usually something which is true without exception. Hutchinson knew that Gram's "law" was not a "law" in this sense when he proposed the name, and therefore he clearly did not use the word in the way it is usually used in mathematics today. Nonetheless the term "Gram's law" has won acceptance in the literature and it will be used in what follows. In the range covered by present-day calculations, extending up to the three-and-a-half-millionth Gram point, the exceptions to Gram's law are surprisingly slight (see Chapter 8).

6.6 TECHNIQUES FOR COMPUTING THE NUMBER OF ROOTS IN A GIVEN RANGE

Gram in the course of earlier work with $\xi(\frac{1}{2} + it)$ had succeeded in computing the Taylor series coefficients of its logarithm with great accuracy. Since

$$\begin{aligned}\log \xi\left(\frac{1}{2} + it\right) &= \sum_{\operatorname{Re} \alpha > 0} \log\left(1 - \frac{t^2}{\alpha^2}\right) + \log \xi\left(\frac{1}{2}\right) \\ &= \log \xi\left(\frac{1}{2}\right) - \left(\sum_{\operatorname{Re} \alpha > 0} \frac{1}{\alpha^2}\right)t^2 - \left(\frac{1}{2} \sum_{\operatorname{Re} \alpha > 0} \frac{1}{\alpha^4}\right)t^4 - \cdots,\end{aligned}$$

this gives the numerical values of $\Sigma \alpha^{-2n}$, where α has the meaning given it by Riemann, namely, $\rho = \frac{1}{2} + i\alpha$ is the generic root of ξ (see Section 1.18). Now α^{-2n} decreases very rapidly as α increases, so the series $\Sigma \alpha^{-2n}$ is dominated by its first few terms. By comparing $\Sigma\alpha^{-10}$ with the sum extended over 15 roots he had already located on the line, Gram was able to show that there are no other roots in the range $0 \leq \operatorname{Im} s \leq 50$. However, this method rapidly becomes unworkable as the number of roots considered increases, so to extend the computations beyond 10 or 15 roots a new method was required. Such a method was found by Backlund [B1] around 1912.

Backlund's method is based on Riemann's observation that if $N(T)$ denotes the number of roots ρ in the range $0 < \operatorname{Im} s < T$, then†

$$N(T) = \frac{1}{2\pi i}\int_{\partial R} \frac{\xi'(s)}{\xi(s)}\,ds,$$

where R is a rectangle of the form $\{-\epsilon \leq \operatorname{Re} s \leq 1 + \epsilon,\ 0 \leq \operatorname{Im} s \leq T\}$, where ∂R is the boundary of R oriented in the usual counterclockwise direction, where it is assumed that T is such that there are no roots ρ on the line $\operatorname{Im} s = T$, and where $N(T)$ counts the roots with multiplicities. By symmetry and the fact that ξ is real on the real axis, this can be rewritten as

$$N(T) = \frac{1}{2\pi}\cdot 2\operatorname{Im}\left[\int_C \frac{\xi'(s)}{\xi(s)}\,ds\right],$$

where C is the portion of ∂R from $1 + \epsilon$ to $\frac{1}{2} + iT$. Using the definition $\xi(s) = \pi^{-s/2}\Pi(\frac{1}{2}s - 1)\frac{1}{2}s(s-1)\zeta(s)$ and the fact that the logarithmic derivative of a product is the sum of the logarithmic derivatives puts this in the form

$$\frac{1}{\pi}\operatorname{Im}\left\{\int_C \frac{d}{ds}\left[\log \pi^{-s/2}\Pi\left(\frac{s}{2} - 1\right)\right] ds\right\}$$
$$+ \frac{1}{\pi}\operatorname{Im}\left\{\int_C \frac{d}{ds}[\log s(s-1)]\,ds\right\} + \frac{1}{\pi}\operatorname{Im}\left\{\int_C \frac{\zeta'(s)}{\zeta(s)}\,ds\right\}.$$

The first two terms, being integrals of derivatives, can be evaluated using the fundamental theorem of calculus; the first is $\pi^{-1}\vartheta(T)$ because it is π^{-1} times the imaginary part of $\log \pi^{-s/2}\Pi(\frac{1}{2}s - 1)$ at $s = \frac{1}{2} + iT$ when this log is defined to be real on the positive real axis, and the second is 1 because it is π^{-1} times the imaginary part of the log of $(\frac{1}{2} + iT)(\frac{1}{2} + iT - 1) = -T^2 - \frac{1}{4}$ when $\log s(s-1)$ is taken to be real for $s > 1$. Thus

$$N(T) = \frac{1}{\pi}\vartheta(T) + 1 + \frac{1}{\pi}\operatorname{Im}\int_C \frac{\zeta'(s)}{\zeta(s)}\,ds.$$

Backlund observed that *if it can be shown that* $\operatorname{Re}\zeta$ *is never zero on* C, *then this formula suffices to determine* $N(T)$ *as the integer nearest to* $\pi^{-1}\vartheta(T) + 1$; this follows simply from noting that if $\operatorname{Re}\zeta$ is never 0 on C, then the curve $\zeta(C)$ never leaves the right halfplane so that $\log\zeta$ is defined all along $\zeta(C)$

†This follows from the "argument principle" of complex analysis or, more directly, from termwise integration of the uniformly convergent (see Section 3.2) series $\xi'(s)/\xi(s) = \Sigma(s-\rho)^{-1}$ using the Cauchy integral fromula.

and gives an antiderivative of ζ'/ζ on C whose imaginary part lies between $-\pi/2$ and $\pi/2$, which by the fundamental theorem shows that the last term above has absolute value less than $\frac{1}{2}$.

Backlund was able to prove by this method that $N(200) = 79$. By methods similar to Gram's he was also able to locate 79 changes of sign of $Z(t)$ on $0 < t < 200$ and thus to prove that *all roots* ρ *in the range* $0 < \operatorname{Im} s < 200$ *are on the line* $\operatorname{Re} s = \frac{1}{2}$ *and all are simple zeros of* ξ. This extended Gram's result from 50 to 200. In 1925 Hutchinson [H11] extended the same result to 300; that is, he proved that all roots ρ in the range $0 < \operatorname{Im} s < 300$ are simple zeros on $\operatorname{Re} s = \frac{1}{2}$. It was explained in the previous section how Hutchinson was able to prove that ξ has at least 138 zeros on the line $\operatorname{Re} s = \frac{1}{2}$ in the range $0 < \operatorname{Im} s < g_{137} = 300.468$ so that, in view of the above observations and in view of the fact that $\pi^{-1}\vartheta(g_{137}) + 1 = 137 + 1 = 138$, the proof of Hutchinson's theorem is reduced to proving that $\operatorname{Re} \zeta$ is never zero on the broken line segment from $1\frac{1}{2}$ to $1\frac{1}{2} + ig_{137}$ to $\frac{1}{2} + ig_{137}$ (which is the curve C when $\epsilon = \frac{1}{2}$, $T = g_{137}$). This Hutchinson did as follows.

In the first place, it is easily shown that $\operatorname{Re} \zeta$ is not zero anywhere on the line $\operatorname{Re} s = 1\frac{1}{2}$. One need only observe that for $\operatorname{Re} s = \sigma > 1$

$$\begin{aligned} |\operatorname{Im} \log \zeta(s)| \le |\log \zeta(s)| &= \left| \int_0^\infty x^{-s}\, dJ(x) \right| \\ &\le \int_0^\infty x^{-\sigma}\, dJ(x) = \log \zeta(\sigma) \end{aligned}$$

and that the value of $\zeta(1\frac{1}{2})$, which can be found by Euler–Maclaurin summation, is less than $e^{\pi/2}$. Hence $\operatorname{Im} \log \zeta(s)$ on $\operatorname{Re} s = 1\frac{1}{2}$ lies between $-\pi/2$ and $\pi/2$, and $\zeta(s)$ cannot lie on the imaginary axis. Thus it remains to prove only that $\operatorname{Re} \zeta$ is never zero on the line segment from $\frac{1}{2} + ig_{137}$ to $1\frac{1}{2} + ig_{137}$. Hutchinson's method of doing this is simply to examine the real parts of the individual terms of the expansion

$$\zeta(s) = \sum_{n=1}^{N-1} n^{-s} + \frac{N^{1-s}}{s-1} + \frac{1}{2}N^{-s} - s\int_N^\infty \bar{B}_1(x)x^{-s-1}\, dx \tag{1}$$

where $s = \sigma + ig_{137}$ $(\frac{1}{2} \le \sigma \le 1\frac{1}{2})$ and $N = 51$. The real part of the first sum consists of 50 terms, say $h_n(\sigma)$, namely,

$$h_n(\sigma) = n^{-\sigma} \cos(g_{137} \log n) \qquad (\tfrac{1}{2} \le \sigma \le 1\tfrac{1}{2}, \quad n = 1, 2, \ldots, 50).$$

For fixed n the sign of $h_n(\sigma)$ is the same for all values of σ, and the objective is to prove that the positive terms dominate the negative terms throughout the interval $\frac{1}{2} \le \sigma \le 1\frac{1}{2}$ with enough left over to dominate the remaining terms as well. Hutchinson found that 21 of the terms $h_n(\sigma)$ are negative, namely, those with $n = 3, 4, 6, 8, 11, 13, 15, 16, 17, 20, 21, 27, 28, 30, 32, 34, 36, 37, 40, 41, 42$. Now the larger n is, the more rapidly the absolute value of $h_n(\sigma)$ decreases as σ increases, so any negative terms dominated by the first two positive terms h_1, h_2 at $\sigma = \frac{1}{2}$ will remain dominated by them for

$\sigma > \frac{1}{2}$. Simple computation shows that they dominate the first four negative terms, that is,

$$[h_1(\tfrac{1}{2}) + h_2(\tfrac{1}{2})] + [h_3(\tfrac{1}{2}) + h_4(\tfrac{1}{2}) + h_6(\tfrac{1}{2}) + h_8(\tfrac{1}{2})] > 0;$$

so the same is true for $\sigma > \frac{1}{2}$. The next negative term is h_{11}, and there are four more positive terms h_5, h_7, h_9, h_{10} which precede it. These four suffice to dominate the next seven negative terms, that is,

$$\begin{aligned}&[h_5(\sigma) + h_7(\sigma) + h_9(\sigma) + h_{10}(\sigma)]\\ &\quad + [h_{11}(\sigma) + h_{13}(\sigma) + h_{15}(\sigma) + h_{16}(\sigma) + h_{17}(\sigma) + h_{20}(\sigma) + h_{21}(\sigma)] > 0\end{aligned}$$

for $\sigma \geq \frac{1}{2}$. This is proved as before, by proving the inequality computationally for $\sigma = \frac{1}{2}$ and observing that the positive terms, having lower indices, decrease in absolute value more slowly than the negative terms. The next negative term is h_{27}, which is preceded by nine more positive terms. These nine suffice to dominate *all* the remaining 10 negative terms, which leaves 14 positive terms with which to dominate the real parts of the remaining three terms of formula (1) for $\zeta(s)$. The first of these three terms is $(s-1)^{-1}N^{1-s}$, the real part of which is

$$\begin{aligned}&\operatorname{Re}\left[\frac{\sigma - 1 - iT}{(\sigma-1)^2 + T^2} \cdot \frac{\exp(-iT\log N)}{N^{\sigma-1}}\right]\\ &\qquad = \frac{(\sigma-1)\cos(T\log N) - T\sin(T\log N)}{[(\sigma-1)^2 + T^2]N^{\sigma-1}},\end{aligned}$$

where $N = 51$, $T = g_{137}$, and $\frac{1}{2} \leq \sigma \leq 1\frac{1}{2}$. Thus, for σ in this range, the absolute value of this term is at most

$$N^{-\sigma}\frac{\frac{1}{2}|\cos(T\log N)| + T\,|\sin(T\log N)|}{T^2N^{-1}}$$

from which it is easily shown that the next positive term h_{29} is more than enough to dominate it, that is,

$$h_{29}(\sigma) + \operatorname{Re}\frac{N^{1-s}}{s-1} > 0$$

for $\frac{1}{2} \leq \sigma \leq 1\frac{1}{2}$. The next term $\frac{1}{2}N^{-s}$ of (1) has a positive real part when $N = 51$, $s = \sigma + ig_{137}$, so only the last term of (1) remains and there are still 13 positive terms with which to dominate it. Integration by parts in the usual manner puts the last term of (1) in the form

$$\begin{aligned}&\frac{B_2}{2!}sN^{-s-1} + \frac{B_4}{4!}s(s+1)(s+2)N^{-s-3} + \cdots\\ &\qquad + \frac{B_{2\nu}}{(2\nu)!}s(s+1)\cdots(s+2\nu-2)N^{-s-2\nu+1} + R,\end{aligned}$$

where, by Backlund's estimate of the remainder (Section 6.4),

$$|R| \leq \frac{|s+2\nu+1|}{\sigma+2\nu+1}\left|\frac{B_{2\nu+2}}{(2\nu+2)!}s(s+1)\cdots(s+2\nu)N^{-s-2\nu-1}\right|.$$

In the case under consideration $s \sim \sigma + 300i$, $\frac{1}{2} \leq \sigma \leq 1\frac{1}{2}$, and $N = 51$. Hutchinson sets $2\nu = 50$ and notes that $|s + k|/N < 6$ for $k = 0, 1, 2, \ldots, 51$ so that the modulus of the last term of (1) is less than

$$\frac{|B_2|}{2!}6 \cdot 51^{-\sigma} + \frac{|B_4|}{4!}6^3 51^{-\sigma} + \cdots + \frac{|B_{50}|}{50!}6^{49}51^{-\sigma} + 6\frac{|B_{52}|}{52!}6^{51}51^{-\sigma}.$$

Now the first few values of $|B_{2\nu}| \cdot 6^{2\nu-1}/(2\nu)!$ are

$$\frac{|B_2|}{2} \cdot 6 = \frac{1}{2}, \qquad \frac{|B_6|}{6!}6^5 = \frac{6^5}{720 \cdot 42} = \frac{9}{35},$$
$$\frac{|B_4|}{4!} \cdot 6^3 = \frac{6^3}{24 \cdot 30} = \frac{3}{10}, \qquad \frac{|B_8|}{8!}6^7 = \frac{6^7}{8!30} = \frac{81}{350},$$

and the ratio of two successive values is, by Euler's formula for $\zeta(2k)$ [(2) of Section 1.5], equal to

$$\begin{aligned}\frac{|B_{2k+2}|}{(2k+2)!}6^{2k+1}\frac{(2k)!}{|B_{2k}|}6^{-2k+1} &= \frac{\zeta(2k+2)}{2^{2k+1}\pi^{2k+2}} \cdot \frac{2^{2k-1}\pi^{2k}}{\zeta(2k)} \cdot 6^2 \\ &= \frac{6^2}{4\pi^2}\frac{\zeta(2k+2)}{\zeta(2k)}.\end{aligned}$$

As $k \longrightarrow \infty$ this ratio approaches $6^2/4\pi^2 = (3/\pi)^2$. The first few ratios 3/5, 6/7, 9/10 are less than $6^2/4\pi^2$ and increase as k increases, a phenomenon which persists for all k as can be seen from the fact that

$$\begin{aligned}&\log\zeta(2k+2) - 2\log\zeta(2k) + \log\zeta(2k-2) \\ &\quad = \int_0^\infty [x^{-2k-2} - 2x^{-2k} + x^{-2k+2}]\, dJ(x) \\ &\quad = \int_0^\infty x^{-2k}(x^{-2} - 2 + x^2)\, dJ(x) \\ &\quad = \int_0^\infty x^{-2k}(x^{-1} - x)^2\, dJ(x) \geq 0,\end{aligned}$$

$$\log\zeta(2k+2) - \log\zeta(2k) \geq \log\zeta(2k) - \log\zeta(2k-2),$$

$$\frac{\zeta(2k+2)}{\zeta(2k)} \geq \frac{\zeta(2k)}{\zeta(2k-2)}.$$

Thus the ratios are all less than $(3/\pi)^2$, which gives the bound

$$\left[\frac{1}{2} + \frac{3}{10} + \frac{9}{35} + \frac{9}{35}\left(\frac{3}{\pi}\right)^2 + \frac{9}{35}\left(\frac{3}{\pi}\right)^4 + \cdots + \frac{9}{35}\left(\frac{3}{\pi}\right)^{44}\right]51^{-\sigma}$$
$$+ 6 \cdot \frac{9}{35}\left(\frac{3}{\pi}\right)^{46}51^{-\sigma}$$

for the modulus of the last term of (1). When the geometric series is summed and the resulting number is estimated for $\sigma = \frac{1}{2}$, it is found to be decidedly less than 1.4. On the other hand, Hutchinson found that the remaining 13 positive terms have the sum 1.492 when $\sigma = \frac{1}{2}$. Thus, since $51^{-\sigma}$ decreases as

σ increases more rapidly than any of the 13 positive terms do, it follows that these 13 positive terms dominate the last term of (1) on $\frac{1}{2} \le \sigma \le 1\frac{1}{2}$, and the proof that $\operatorname{Re}\zeta > 0$ on the line segment from $\frac{1}{2} + ig_{137}$ to $1\frac{1}{2} + ig_{137}$ is complete.

Hutchinson states that by using the same methods he was also able to show that $\operatorname{Re}\zeta$ does not vanish on the line segment from $\frac{1}{2} + ig_{268}$ to $1\frac{1}{2} + ig_{268}$. This implies of course that $N(g_{268}) = \pi^{-1}\vartheta(g_{268}) + 1 = 269$ so, since $g_{268} = 499.1575$, it constitutes one half of a proof that all roots ρ in the range $0 < \operatorname{Im} s < (500 - \epsilon)$ are simple zeros on the line $\operatorname{Re} s = \frac{1}{2}$, that is, it constitutes one half of an extension of the previous result from 300 to just below 500. However, Hutchinson was apparently unable to complete the proof by locating another $269 - 138 = 131$ changes of sign in $\zeta(\frac{1}{2} + it)$ for $300 < t < 500$. The evaluation of $\zeta(\frac{1}{2} + it)$ by Euler–Maclaurin summation is very lengthy for t in this range, and the project of performing enough such evaluations to locate the required 131 sign changes was more than Hutchinson was willing, or perhaps able, to undertake. It is just as well that he did not undertake it because only a few years later a much shorter method of evaluating $\zeta(\frac{1}{2} + it)$, and hence of finding the required 131 sign changes, was discovered. This method, which uses the Riemann–Siegel formula, is the subject of the next chapter.

6.7 BACKLUND'S ESTIMATE OF $N(T)$

As in the preceding section, let $N(T)$ denote the number of roots ρ in the range $0 < \operatorname{Im} s < T$. Then Riemann's estimate of $N(T)$ [see (d) of Section 1.19] is the statement that the relative error in the approximation

(1) $$N(T) \sim \frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi}$$

is less than a constant times T^{-1} as $T \to \infty$. Since the right side is greater than a constant times $T \log T$ as $T \to \infty$, this statement will follow if it can be shown that the absolute error in (1) is less than a constant times $\log T$. But the formulas

$$N(T) = \pi^{-1}\vartheta(T) + 1 + \pi^{-1} \operatorname{Im} \int_C \frac{\zeta'(s)}{\zeta(s)}\, ds,$$

$$\vartheta(T) = \frac{T}{2} \log \frac{T}{2\pi} - \frac{T}{2} - \frac{\pi}{8} + \frac{1}{48T} + \frac{7}{5760T^3} + \cdots$$

of Sections 6.6 and 6.5, respectively, combine to give

$$N(T) - \left[\frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi}\right] = \left[-\frac{1}{8} + \frac{1}{48\pi T} + \cdots\right] + 1 + \pi^{-1} \operatorname{Im} \int_C \frac{\zeta'(s)}{\zeta(s)}\, ds$$

which shows that the magnitude of the absolute error in (1) is less than

$$1 + \pi^{-1}\left|\operatorname{Im}\int_C \frac{\zeta'(s)}{\zeta(s)}\,ds\right| \tag{2}$$

for large T. Here C denotes the broken line segment from $1\frac{1}{2}$ to $1\frac{1}{2} + iT$ to $\frac{1}{2} + iT$, and it is assumed that $\zeta(s)$ is not zero on C, which is the same as to say that T is not a discontinuity of the step function $N(T)$. Now it suffices to prove Riemann's estimate (1) for such T, and for such T the integral in (2) can be evaluated by the fundamental theorem

$$\operatorname{Im}\int_C \frac{\zeta'(s)}{\zeta(s)}\,ds = \operatorname{Im}\log\zeta\left(\frac{1}{2} + iT\right)$$

when the log on the left is defined by analytic continuation along C. The generalization of the statement that this integral lies between $-\pi/2$ and $\pi/2$ if $\operatorname{Re}\zeta$ has no zeros on C is the statement that if $\operatorname{Re}\zeta$ has n zeros on C, then the integral lies between $-(n + \frac{1}{2})\pi$ and $(n + \frac{1}{2})\pi$. Thus the error in (1) has absolute value less than

$$1 + n + \tfrac{1}{2},$$

and in order to prove Riemann's estimate (1), it suffices to prove that *there is a constant K such that the number n of zeros of* $\operatorname{Re}\zeta$ *on the line segment from* $\frac{1}{2} + iT$ *to* $1\frac{1}{2} + iT$ *is at most* $K\log T$ *for all sufficiently large* T. Backlund was able to prove this theorem by a simple application of Jensen's theorem and was thereby able to give a much simpler proof [B1] of Riemann's estimate (1) than von Mangoldt's original proof [M3]. His proof is as follows.

Let $f(z) = \frac{1}{2}[\zeta(z + 2 + iT) + \zeta(z + 2 - iT)]$. Since $\zeta(\bar{s}) = \overline{\zeta(s)}$, the function f is identical with $\operatorname{Re}\zeta(z + 2 + iT)$ for real z, so the number n in question is equal to the number of zeros of $f(z)$ on the interval $-1\frac{1}{2} \le z \le -\frac{1}{2}$ of the real axis. Now consider Jensen's formula

$$\log|f(0)| + \sum\log\left|\frac{R}{z_i}\right| = \frac{1}{2\pi}\int_0^{2\pi}\log|f(Re^{i\theta})|\,d\theta \tag{3}$$

(see Section 2.2). Since $f(z)$ is analytic in the entire z-plane except for poles at $z + 2 \pm iT = 1$, that is, poles at $z = -1 \pm iT$, Jensen's formula applies whenever $R \le T$. Consider the case $R = 2 - \epsilon$ where ϵ is chosen so that f has no zeros on the circle $|z| = R$. The second sum on the left side of Jensen's formula (3) is a sum of positive terms and the terms corresponding to the n zeros in question are all at least $\log|(2 - \epsilon)/1\frac{1}{2}| = \log\frac{1}{3}(4 - 2\epsilon)$; hence

$$\log|f(0)| + n\log\tfrac{1}{3}(4 - 2\epsilon) \le \log M$$

where M is the maximum value of $|f(z)|$ on $|z| = 2 - \epsilon$. As $\epsilon \downarrow 0$, this gives

$$n \le \frac{\log|M/f(0)|}{\log\frac{4}{3}}$$

as an upper bound for n where M is an upper bound for $|f(z)|$ on $|z| = 2$.

Now $|f(0)| = |\operatorname{Re}\zeta(2+iT)| \geq 1 - 2^{-2} - 3^{-2} - 4^{-2} - \cdots = 1 - [\zeta(2) - 1] = 2 - (\pi^2/6) > \frac{1}{4}$, so this gives

$$n \leq \frac{1}{\log 4 - \log 3} \cdot \log 4M \leq \text{const}\log M + \text{const},$$

and to prove the theorem, it suffices to show that $\log M$ grows no faster than a constant times $\log T$. Now

$$\begin{aligned} M &= \max|\tfrac{1}{2}\zeta(2e^{i\theta}+2+iT) + \tfrac{1}{2}\zeta(2e^{i\theta}+2-iT)| \\ &\leq \max|\zeta(2e^{i\theta}+2+iT)|, \end{aligned}$$

so it suffices to estimate the growth of $|\zeta(s)|$ in the strip $0 \leq \operatorname{Re}s \leq 4$. But Backlund's estimate of the remainder R in

$$\zeta(s) = 1/(s-1) + \tfrac{1}{2} + R$$

gives (see Section 6.4)

$$|\zeta(s)| \leq \left|\frac{1}{s-1}\right| + \frac{1}{2} + \frac{|s+1|}{\sigma+1}\frac{B_2}{2!}|s|.$$

Hence for $s = 2e^{i\theta} + 2 + iT$,

$$\begin{aligned} M &\leq \frac{1}{T-2} + \frac{1}{2} + \frac{T+2+5}{0+1}\frac{1}{12}\cdot(T+2+4) \\ &\leq \text{const}\,T^2 \end{aligned}$$

for large T and the theorem follows.

By refining this estimate carefully, Backlund was able to obtain the specific estimate

$$\begin{aligned} &\left|N(T) - \left(\frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} - \frac{7}{8}\right)\right| \\ &\qquad < 0.137\log T + 0.443\log\log T + 4.350 \end{aligned}$$

for all $T \geq 2$. (See Backlund [B3].)

6.8 ALTERNATIVE EVALUATION OF $\zeta'(0)/\zeta(0)$

Euler–Maclaurin summation can be used to prove the formula $\zeta'(0)/\zeta(0) = \log 2\pi$ of Section 3.8 as follows:

$$\begin{aligned} \zeta(s) - \frac{1}{s-1} &= \sum_{n=1}^{\infty} n^{-s} - \int_1^{\infty} x^{-s}\,dx \\ &= \sum_{n=1}^{N-1} n^{-s} + \sum_{n=N}^{\infty} n^{-s} - \int_1^N x^{-s}\,dx - \int_N^{\infty} x^{-s}\,dx \\ &= \sum_{n=1}^{N-1} n^{-s} - \int_1^N x^{-s}\,dx + \tfrac{1}{2}N^{-s} - s\int_N^{\infty} \bar{B}_1(x)x^{-s-1}\,dx \end{aligned}$$

at first for Re $s > 1$ but then by analytic continuation for Re $s > 0$. For $s = 1$ the right side is

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N-1} - \log N + \frac{1}{2N} - \int_N^\infty \bar{B}_1(x)x^{-2}\,dx$$

which approaches Euler's constant γ (see Section 3.8) as $N \longrightarrow \infty$. Thus the Taylor series expansion of $(s-1)\zeta(s)$ around $s = 1$ begins $(s-1)\zeta(s) = 1 + \gamma(s-1) + \cdots$ and γ is the logarithmic derivative of $(s-1)\zeta(s)$ at $s = 1$. But the functional equation in the form (4) of Section 1.6 gives

$$(s-1)\zeta(s) = -\Pi(1-s)(2\pi)^{s-1}2\sin(s\pi/2)\zeta(1-s)$$

so logarithmic differentiation of both sides at $s = 1$ gives

$$\begin{aligned}\gamma &= -\frac{\Pi'(0)}{\Pi(0)} + \log 2\pi + \frac{\pi}{2}\frac{\cos(\pi/2)}{\sin(\pi/2)} - \frac{\zeta'(0)}{\zeta(0)}\\ &= \gamma + \log 2\pi - \frac{\zeta'(0)}{\zeta(0)}\end{aligned}$$

and the result follows.

Chapter 7

The Riemann–Siegel Formula

7.1 INTRODUCTION

In 1932 Carl Ludwig Siegel published an account [S4] of the work relating to the zeta function and analytic number theory found in Riemann's private papers in the archives of the University Library at Göttingen [R1a]. This was an event of very great importance in the history of the study of the zeta function not only because the work contained new and important information, but also because it revealed the profundity and technical virtuosity of Riemann's researches. Anyone who has read Siegel's paper is unlikely to assert, as Hardy did in 1915 [H3a], that Riemann "could not prove" the statements he made about the zeta function, or to call them, as Landau [L3] did in 1908, "conjectures." Whereas the eight-page resumé *Ueber die Anzahl*. . . , the only work which Riemann published on this subject, could possibly be interpreted as a series of remarkable heuristic insights, Siegel's paper shows clearly that there lay behind it an extensive analysis which may have lacked detailed error estimates but which surely did not lack extremely powerful methods and which in all likelihood was based on a very sure grasp of the magnitudes of error terms even when they were not explicitly estimated.

The difficulty of Siegel's undertaking could scarcely be exaggerated. Several first-rate mathematicians before him had tried to decipher Riemann's disconnected jottings, but all had been discouraged either by the complete lack of any explanation of the formulas, or by the apparent chaos in their arrangement, or by the analytical skill needed to understand them. One wonders whether anyone else would ever have unearthed this treasure if Siegel had not. It is indeed fortunate that Siegel's concept of scholarship derived from the older tradition of respect for the past rather than the contemporary style of novelty.

There are two topics covered in the paper, the one an asymptotic formula for the computation of $Z(t)$ and the other a new representation of $\zeta(s)$ in terms of definite integrals. This chapter is devoted mainly to the asymptotic formula for $Z(t)$, which is known as the Riemann–Siegel formula. The majority of the chapter, Sections 7.2–7.5, consists of the derivation of this formula. Some computations using the formula are given in Section 7.6, error estimates are discussed in Section 7.7, and the relation of the formula to the Riemann hypothesis is discussed in Section 7.8. Finally, in Section 7.9, the new representation of $\zeta(s)$ is derived.

7.2 BASIC DERIVATION OF THE FORMULA

Recall Riemann's formula for $\zeta(s)$ which "remains valid for all s," namely, the formula

$$\zeta(s) = \frac{\Pi(-s)}{2\pi i} \int_{+\infty}^{+\infty} \frac{(-x)^s}{e^x - 1} \cdot \frac{dx}{x}, \tag{1}$$

where the limits of integration indicate a contour which begins at $+\infty$, descends the real axis, circles the singularity at the origin once in the positive direction, and returns up the positive real axis to $+\infty$ [see (3) of Section 1.4] and where $(-x)^s = \exp[s \log(-x)]$ is defined in the usual way for $-x$ not on the negative real axis. There are two ways that finite sums can be split off from (1), the first being to use

$$\frac{e^{-Nx}}{e^x - 1} = \sum_{n=N+1}^{\infty} e^{-nx}$$

in place of $(e^x - 1)^{-1} = \sum e^{-nx}$ in the derivation of (1) to find

$$\zeta(s) = \sum_{n=1}^{N} n^{-s} + \frac{\Pi(-s)}{2\pi i} \int_{+\infty}^{+\infty} \frac{e^{-Nx}(-x)^s}{e^x - 1} \cdot \frac{dx}{x} \tag{2}$$

and the second being to change the contour of integration to a curve C_M which circles the poles $\pm 2\pi i M, \pm 2\pi i(M-1), \ldots, \pm 2\pi i$ as well as the singularity 0 of the integrand (say C_M is the path which descends the real axis from $+\infty$ to $(2M+1)\pi$, circles the boundary of the disk $|s| \leq (2M+1)\pi$ once in the positive direction, and returns to $+\infty$) and, using the residue theorem, to find

$$\zeta(s) = \Pi(-s)(2\pi)^{s-1} 2 \sin\frac{\pi s}{2} \sum_{n=1}^{M} n^{-(1-s)} + \frac{\Pi(-s)}{2\pi i} \int_{C_M} \frac{(-x)^s}{e^x - 1} \cdot \frac{dx}{x} \tag{3}$$

as in the derivation of the functional equation of ζ in Section 1.6. The first proof of the functional equation (see Section 1.6) amounts simply to showing that if $\operatorname{Re} s > 1$, then the integral in (2) approaches zero as $N \to \infty$ (because it is $\sum_{N+1}^{\infty} n^{-s}$) whereas if $\operatorname{Re} s < 0$, then the integral in (3) approaches zero as $M \to \infty$ (by the estimate of Section 1.6). For s in the critical strip $0 < \operatorname{Re} s < 1$, however, neither the integral in (2) nor the integral in (3) can be neglected.

The techniques by which (2) and (3) were derived can easily be combined to give

$$\text{(4)}\qquad \begin{aligned}\zeta(s) = \sum_{n=1}^{N} n^{-s} + \Pi(-s)(2\pi)^{s-1}2\sin\frac{\pi s}{2}\sum_{n=1}^{M} n^{-(1-s)}\\ + \frac{\Pi(-s)}{2\pi i}\int_{C_M}\frac{(-x)^s e^{-Nx}}{e^x-1}\cdot\frac{dx}{x}\end{aligned}$$

which can be put in a more symmetrical form by multiplying by $\frac{1}{2}s(s-1)\Pi(\frac{1}{2}s-1)\pi^{-s/2}$ and using the identities of the factorial function needed for the derivation of the symmetrical form of the functional equation [see (5) of Section 1.6] to find

$$\text{(5)}\qquad \begin{aligned}\xi(s) &= (s-1)\Pi\left(\frac{s}{2}\right)\pi^{-s/2}\sum_{n=1}^{N} n^{-s}\\ &+ (-s)\Pi\left(\frac{1-s}{2}\right)\pi^{-(1-s)/2}\sum_{n=1}^{M} n^{-(1-s)}\\ &+ \frac{(-s)\Pi(1-s/2)\pi^{-(1-s)/2}}{(2\pi)^{s-1}2\sin(\pi s/2)\cdot 2\pi i}\int_{C_M}\frac{(-x)^s e^{-Nx}}{e^x-1}\cdot\frac{dx}{x}\end{aligned}$$

for all N, M, s. The case of greatest interest is the case $s = \frac{1}{2} + it$ where t is real. Then, in view of the symmetry between s and $1 - s$, it is natural to set $N = M$. If $f(t)$ denotes $(-\frac{1}{2} + it)\Pi((\frac{1}{2} + it)/2)\pi^{-(1/2+it)/2}$, the formula then becomes

$$\begin{aligned}\xi\left(\frac{1}{2}+it\right) &= f(t)\sum_{n=1}^{N} n^{-(1/2)-it} + f(-t)\sum_{n=1}^{N} n^{-(1/2)+it}\\ &+ \frac{f(-t)}{(2\pi)^{(1/2)+it}2i\sin[\frac{1}{2}\pi(\frac{1}{2}+it)]}\int_{C_N}\frac{-(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x-1}.\end{aligned}$$

Now by definition $Z(t)$ satisfies $\xi(\frac{1}{2} + it) = r(t)Z(t)$, where (see Section 6.5)

$$\begin{aligned}r(t) &= \exp\left[\operatorname{Re}\log\Pi\left(\frac{1}{2}s-1\right)\right]\pi^{-(1/4)}\cdot\frac{-t^2-\frac{1}{4}}{2}\\ &= \exp\left[\log\Pi\left(\frac{1}{2}s-1\right)\right]\pi^{-(1/4)}\cdot\frac{s(s-1)}{2}\exp\left[-i\operatorname{Im}\log\Pi\left(\frac{1}{2}s-1\right)\right]\\ &= \Pi\left(\frac{s}{2}\right)(s-1)\pi^{-(1/4)}\,e^{-i\vartheta(t)}\pi^{-(1/2)it}\\ &= f(t)e^{-i\vartheta(t)},\end{aligned}$$

$$f(t) = r(t)e^{i\vartheta(t)}$$

($s = \frac{1}{2} + it$), so a factor $r(t) = r(-t)$ can be canceled from all terms above.

Using $\vartheta(-t) = -\vartheta(t)$ and the simplification $2i\sin(\pi s/2) = e^{-i\pi s/2}(e^{i\pi s} - 1) = e^{-i\pi/4}e^{t\pi/2}(e^{i\pi/2}e^{-t\pi} - 1) = -e^{-i\pi/4}e^{t\pi/2}(1 - ie^{-t\pi})$ then puts the formula in the form

$$Z(t) = \sum_{n=1}^{N} n^{-1/2}\cdot 2\cos[\vartheta(t) - t\log n] + \frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1 - ie^{-t\pi})}\int_{C_N}\frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x - 1}$$

for all real t. Although the series on the right diverges as $N \longrightarrow \infty$, the terms do decrease in size, which gives some reason to believe that if N is suitably chosen, the approximation

$$Z(t) \sim 2\sum_{n=1}^{N} n^{-1/2}\cos[\vartheta(t) - t\log n]$$

might have some merit. The study of this approximation is of course equivalent to the study of the remainder term

$$\frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1 - ie^{-t\pi})}\int_{C_N}\frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x - 1}. \tag{6}$$

The Riemann–Siegel formula is† a technique for the approximate numerical evaluation of this integral and hence of $Z(t)$ and $\zeta(\frac{1}{2} + it)$.

The essence of Riemann's technique for evaluating the integral (6) is a standard technique for the approximate evaluation of definite integrals known as the *saddle point method* or the *method of steepest descent* (see,‡ for example, Jeffreys and Jeffreys [J2]). Consider the modulus of the integrand

$$\frac{(-x)^{-(1/2)+it}e^{-Nx}}{e^x - 1}. \tag{7}$$

As long as the contour of integration stays well away from the zeros of the denominator $x = 0, \pm 2\pi i, \pm 4\pi i, \ldots$, this modulus is at most a constant times the modulus of the numerator, so in looking for places where the modulus of the integrand is large, it suffices to consider places where the modulus of the numerator is large. Now this modulus is $e^{\phi(x)}$, where

$$\phi(x) = \operatorname{Re}\{(-\tfrac{1}{2} + it)\log(-x) - Nx\}.$$

†The assumption $\operatorname{Re} s = \frac{1}{2}$ is made only for the sake of convenience; the entire analysis of (6) applies, with only slight modifications, to the last term of (4). Siegel carries through this analysis and makes the assumption $\operatorname{Re} s = \frac{1}{2}$ only as the last step. The assumption $N = M$ is a natural concomitant of the assumption $\operatorname{Re} s = \frac{1}{2}$, but if, for example, $\operatorname{Re} s > \frac{1}{2}$, then s is nearer the range where the series with N converges, and it would presumably be better to take $N > M$. Siegel indicates the method for dealing with the case $N \neq M$, but he does not carry it through.

‡Jeffreys and Jeffreys attribute the method to Debye, although Debye himself acknowledges (see [D1]) that the method occurs in a posthumously published fragment of Riemann [R1, pp. 405–406]. Be that as it may, the widespread use of the method in the theory of Bessel functions and in theoretical physics dates from Debye's rediscovery of it in 1910.

Since ϕ is a harmonic function, it has no local maxima or minima, but it does have a saddle point at the unique point where the derivative of $(-\frac{1}{2} + it)\log(-x) - Nx$ is zero, namely, at the point $(-\frac{1}{2} + it)/N$. Let α denote this point. In the vicinity of α the function ϕ can be written in the form

$$\begin{aligned}
\phi(x) &= \operatorname{Re}\Big\{\Big(-\frac{1}{2} + it\Big)\log(-\alpha) \\
&\quad + \Big(-\frac{1}{2} + it\Big)\log\Big(1 + \frac{x-\alpha}{\alpha}\Big) - N\alpha - N(x-\alpha)\Big\} \\
&= \text{const} + \operatorname{Re}\Big\{\Big(-\frac{1}{2} + it\Big) \\
&\quad \times \Big[\frac{x-\alpha}{\alpha} - \frac{1}{2}\Big(\frac{x-\alpha}{\alpha}\Big)^2 + \frac{1}{3}\Big(\frac{x-\alpha}{\alpha}\Big)^3 - \cdots\Big] - N(x-\alpha)\Big\} \\
&= \text{const} + \operatorname{Re}\Big\{-\frac{1}{2}\Big(-\frac{1}{2} + it\Big)\Big(\frac{x-\alpha}{\alpha}\Big)^2 \\
&\quad + \frac{1}{3}\Big(-\frac{1}{2} + it\Big)^3 - \cdots\Big\} \\
&= \text{const} - \frac{1}{2}\operatorname{Re}\Big\{\frac{N^2(x-\alpha)^2}{-\frac{1}{2} + it}\Big\} \\
&\quad + \text{terms in } (x-\alpha)^3, (x-\alpha)^4, \ldots .
\end{aligned}$$

If x passes through α along the line $\operatorname{Im}\log(x - \alpha) = \frac{1}{2}\operatorname{Im}\log(-\frac{1}{2} + it)$ where $(x - \alpha)^2/(-\frac{1}{2} + it)$ is real and positive, then ϕ has a local maximum at α and, consequently, the modulus of the integrand (7) has a local maximum at α. (On the other hand, if x passes through α along the line perpendicular to this one, then ϕ has a local *minimum* at α; thus the method is to cross the saddle point at α along the line of steepest descent, which gives the method its name.) Thus, if it can be arranged that the path of integration passes through α in this way and never enters regions away from α where the integrand is large, then the integral will have been concentrated into a small part of the total path of integration and this short integral will be approximable by local methods.

Now if t is large, then the saddle point $\alpha = (-\frac{1}{2} + it)/N$ lies near the positive imaginary axis, whatever value of N is chosen. But the path of integration C_N (recall $M = N$) crosses the positive imaginary axis between $2\pi Ni$ and $2\pi(N + 1)i$, so in order for C_N to pass near the saddle point it is necessary to have $(-\frac{1}{2} + it)/N \sim 2\pi Ni$ or $N^2 \sim (-\frac{1}{2} + it)/2\pi i \sim t/2\pi$. This motivates the choice of N as the integer part of $(t/2\pi)^{1/2}$, that is, $N = [(t/2\pi)^{1/2}]$ is the largest integer less than $(t/2\pi)^{1/2}$. Then the saddle point α lies near $it(t/2\pi)^{-1/2} = i(2\pi t)^{1/2}$, which is between $2\pi Ni$ and $2\pi(N + 1)i$ as desired. Since $\operatorname{Im}\log(-\frac{1}{2} + it) \sim \pi/2$, the path of integration should pass through the saddle point along a line of slope approximately 1 and, because of the configuration of C_N, it should go from upper right to lower left.

Let a denote the approximate saddle point $a = i(2\pi t)^{1/2}$ and let L denote the line of slope 1 through a directed from upper right to lower left. Then, in summary, the saddle point method suggests that *the integral* (6) *is approximately equal to the integral of the same function over a segment of L containing a, and this latter integral can be approximated by local methods* using the fact that the modulus of the integrand has a saddle point near a.

7.3 ESTIMATION OF THE INTEGRAL AWAY FROM THE SADDLE POINT

Let $t > 0$ be given, let $a = i(2\pi t)^{1/2}$, and let L be the line of slope 1 through a directed from upper right to lower left. The objective of this section is to show that if L_1 is a suitable segment of L, then the remainder term (6) of Section 7.2 is accurately estimated by the approximation

(1)
$$\frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1-ie^{-t\pi})}\int_{C_N}\frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x-1}$$
$$\sim \frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1-ie^{-t\pi})}\int_{L_1}\frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x-1},$$

where, as in Section 7.2, N is the integer part of $(t/2\pi)^{1/2}$ and C_N is a contour which begins at $+\infty$, circles the poles $\pm 2\pi iN, \pm 2\pi i(N-1), \ldots, \pm 2\pi i$, and the singularity 0 once in the positive direction and returns to $+\infty$. [The value of t must be assumed to be such that $(t/2\pi)^{1/2}$ is not an integer so that a is not a pole of the integrand.] More precisely, the objective is to show that the error in the approximation (1) is very small for t at all large and approaches zero as $t \to \infty$.

Approximations to the right side of (1) will be given in the following section, 7.4. These approximations will be derived using termwise integration of a power series in $(x - a)$ whose radius of convergence is $|a|$ and for this reason it will be advantageous to take L_1 to lie well within this radius of convergence, say the portion of L which lies within $\frac{1}{2}|a|$ of a. Thus *L_1 will be the directed line segment from $a + \frac{1}{2}e^{i\pi/4}|a|$ to $a - \frac{1}{2}e^{i\pi/4}|a|$*, where $a = i(2\pi t)^{1/2}$.

With this choice of L_1, the path of integration C_N can be taken to be $L_0 + L_1 + L_2 + L_3$, where L_0 is the (infinite) portion of L which lies above and to the right of L_1, where L_2 is the vertical line segment from the lower left end of L_1 to the line $\{\operatorname{Im} x = -(2N + 1)\pi\}$, and where L_3 is the (infinite) portion of $\{\operatorname{Im} x = -(2N + 1)\pi\}$ to the right of the lower end of L_2. If $\operatorname{Re} x$

is very large, then the very small term e^{-Nx} dominates the integrand and it is easily seen that this is a valid choice of C_N even though L_0 and L_3 do not approach infinity along the positive real axis. With these definitions, then, the approximation (1) to be proved becomes

$$(2)\qquad \frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1-ie^{-t\pi})} \times \int_{L_j} \frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x-1} \sim 0 \qquad (j=0,2,3).$$

These three approximations will be considered in turn.

The case $j=0$ of (2): The modulus of the numerator of the integrand is, as before, $e^{\phi(x)}$, where $\phi(x)=\operatorname{Re}\{(-\frac{1}{2}+it)\log(-x)-Nx\}$. The presence of a saddle point of ϕ near a, where L nearly passes over a maximum of ϕ, suggests that ϕ increases as x descends L_0 toward its terminal point and that $e^{\phi(x)}$ has its maximum on L_0 at this terminal point. This is easily confirmed by differentiating $\phi=\operatorname{Re}[(-\frac{1}{2}+it)\log(-a-ke^{i\pi/4})-N(a+ke^{i\pi/4})]$ with respect to k for k real and greater than or equal to $\frac{1}{2}|a|$ to find

$$\begin{aligned}
\frac{d\phi}{dk} &= \operatorname{Re}\left\{\left(-\frac{1}{2}+it\right)(a+ke^{i\pi/4})^{-1}e^{i\pi/4}-Ne^{i\pi/4}\right\}\\
&= -\frac{1}{2}\operatorname{Re}\{(a+ke^{i\pi/4})^{-1}e^{i\pi/4}\}\\
&\quad + \operatorname{Re}\left\{\left(\frac{it}{a}\right)\left(1+\frac{k}{a}e^{i\pi/4}\right)^{-1}e^{i\pi/4}\right\} - N\operatorname{Re}\{e^{i\pi/4}\}\\
&\le \frac{1}{2}|a+ke^{i\pi/4}|^{-1} + \left(\frac{t}{2\pi}\right)^{1/2}\operatorname{Re}\left\{\left(1+\frac{k}{|a|}e^{-i\pi/4}\right)^{-1}e^{i\pi/4}\right\}\\
&\quad - N\left(\frac{\sqrt{2}}{2}\right).
\end{aligned}$$

With $u=k|a|^{-1}\ge\frac{1}{2}$ this can be written in the form

$$\begin{aligned}
\frac{d\phi}{dk} &\le \frac{1}{2}|a|^{-1} + \left(\frac{t}{2\pi}\right)^{1/2}\operatorname{Re}\{(1+ue^{-i\pi/4})^{-1}e^{i\pi/4}\}\\
&\quad - \left[\left(\frac{t}{2\pi}\right)^{1/2}-1\right]\operatorname{Re}\{e^{i\pi/4}\}\\
&= \frac{1}{2}(2\pi t)^{-1/2} + \left(\frac{t}{2\pi}\right)^{1/2}\operatorname{Re}\{(1+ue^{-i\pi/4})^{-1}(e^{i\pi/4}-e^{i\pi/4}-u)\} + \frac{\sqrt{2}}{2}\\
&= \frac{1}{2}(2\pi t)^{-1/2} - \left(\frac{t}{2\pi}\right)^{1/2}\operatorname{Re}\{(u^{-1}+e^{-i\pi/4})^{-1}\} + \frac{\sqrt{2}}{2}.
\end{aligned}$$

The middle term is at most a positive constant times $-(t/2\pi)^{1/2}$, hence $d\phi/dk$ is negative on all of L_0 whenever t is at all large, as was to be shown. Thus the modulus of the numerator is at most

$$\begin{aligned}
&\exp \operatorname{Re}\left\{\left(-\frac{1}{2}+it\right)\log\left(-a-\frac{1}{2}|a|e^{i\pi/4}\right)-N\left(a+\frac{1}{2}|a|e^{i\pi/4}\right)\right\}\\
&\quad=\exp\left[-\frac{1}{2}\log\left|a+\frac{1}{2}ae^{-i\pi/4}\right|\right.\\
&\qquad\left.-t\operatorname{Im}\log\left(-a-\frac{1}{2}ae^{-i\pi/4}\right)-N\cdot\frac{1}{2}(2\pi t)^{1/2}\frac{\sqrt{2}}{2}\right]\\
&\quad=|a|^{-1/2}\left|1+\frac{1}{2}e^{-i\pi/4}\right|^{-1/2}\\
&\qquad\times\exp\left[-t\operatorname{Im}\log\left(-i-\frac{1}{2}e^{i\pi/4}\right)-N\cdot\frac{1}{2}(\pi t)^{1/2}\right]\\
&\quad\le(2\pi t)^{-1/4}\exp\left\{-t\left[-\frac{\pi}{2}-\operatorname{Arctan}\left(\frac{\frac{1}{2}\cdot\frac{1}{2}\sqrt{2}}{1+\frac{1}{2}\cdot\frac{1}{2}\sqrt{2}}\right)\right]\right.\\
&\qquad\left.-\left[\left(\frac{t}{2\pi}\right)^{1/2}-1\right]\frac{1}{2}(\pi t)^{1/2}\right\}\\
&\quad=(2\pi t)^{-1/4}e^{t\pi/2}\exp\left\{t\operatorname{Arctan}\left(\frac{1}{2\sqrt{2}+1}\right)-\frac{t}{2\sqrt{2}}\right\}\exp\left[\frac{1}{2}(\pi t)^{1/2}\right]\\
&\quad\le e^{t\pi/2}e^{-t/11}\exp\left[\frac{1}{2}(\pi t)^{1/2}\right]
\end{aligned}$$

because

$$\operatorname{Arctan}\left(\frac{1}{2\sqrt{2}+1}\right)-\frac{1}{2\sqrt{2}}\le\frac{1}{2\sqrt{2}+1}-\frac{1}{2\sqrt{2}}<-\frac{1}{11}$$

and because it can be assumed that $2\pi t>1$. Thus the integral to be estimated has modulus at most

$$\begin{aligned}
&\left|\frac{e^{-i\vartheta(t)}e^{-t\pi/2}e^{t\pi/2}e^{-t/11}\exp[\frac{1}{2}(\pi t)^{1/2}]}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1-ie^{-t\pi})}\right|\left|\int_{L_0}\frac{|dx|}{|e^x-1|}\right|\\
&\quad\le\frac{e^{-t/11}\exp[\frac{1}{2}(\pi t)^{1/2}]}{(2\pi)^{1/2}(1-e^{-t\pi})}\int_{(1/2)(\pi t)^{1/2}}^{\infty}\frac{e^{-u}\sqrt{2}\,du}{1-e^{-u}}\\
&\quad\le\frac{e^{-t/11}\exp[\frac{1}{2}(\pi t)^{1/2}]}{(2\pi)^{1/2}(1-e^{-t\pi})}\cdot\frac{\sqrt{2}}{1-\exp[-\frac{1}{2}(\pi t)^{1/2}]}[-e^{-u}]\Big|_{(\pi t)^{1/2}/2}^{\infty}\\
&\quad=\frac{e^{-t/11}}{\pi^{1/2}}\cdot\frac{1}{(1-e^{-t\pi})\{1-\exp[-\frac{1}{2}(\pi t)^{1/2}]\}}.
\end{aligned}$$

The second term differs negligibly from 1 when t is at all large and the integral (2) in the case $j=0$ is comfortably less than $e^{-t/11}$.

The case $j=2$ of (2): On L_2 the real part of x is constant, say $-b$, where $b=-\operatorname{Re}(a-\frac{1}{2}e^{i\pi/4}|a|)=\frac{1}{2}\cdot\frac{1}{2}\sqrt{2}\cdot(2\pi t)^{1/2}=\frac{1}{2}(\pi t)^{1/2}$. The denominator of the integrand on L_2 then has modulus at least $1-e^{-b}$ which is greater than $\frac{1}{2}$ for t at all large. The numerator of the integrand has modulus at most

$$|(-x)^{-(1/2)+it}e^{-Nx}| \le (\max|x|^{-1/2})\{\max\exp[-t\operatorname{Im}\log(-x)]\}\exp\left[\left(\frac{t}{2\pi}\right)^{1/2}\cdot b\right].$$

The maximum value of $|x|^{-1/2}$ occurs at the point where L_2 crosses the real axis, at which point it is $b^{-1/2}$. The maximum value of $\exp[-t\operatorname{Im}\log(-x)]$ occurs at the minimum value of $\operatorname{Im}\log(-x)$ at the initial point of L_2, where

$$\operatorname{Im}\log(-x) = \operatorname{Im}\log\left(-i + \frac{1}{2}e^{i\pi/4}\right) = -\frac{\pi}{2} + \operatorname{Arctan}\left(\frac{\frac{1}{2}\frac{1}{2}\sqrt{2}}{1-\frac{1}{2}\frac{1}{2}\sqrt{2}}\right),$$
$$\exp[-t\operatorname{Im}\log(-x)] = e^{t\pi/2}\exp\left[-t\operatorname{Arctan}\frac{1}{2\sqrt{2}-1}\right].$$

Finally, $(t/2\pi)^{1/2}b = t/2\sqrt{2}$, so the modulus of the numerator is at most $b^{-1/2}e^{t\pi/2}e^{-kt}$ where $k = \operatorname{Arctan}(1/2\sqrt{2}-1) - (1/2\sqrt{2})$. By direct numerical evaluation it is found that $k < \frac{1}{8}$ so the integrand has modulus at most $2b^{-1/2}e^{t\pi/2}e^{-t/8}$ on L_2. Since the length of the path of integration L_2 is less than $2|a| = 2(2\pi t)^{1/2} = 4\sqrt{2}\,b$, this shows that (2) is at most

$$\frac{e^{-t\pi/2}}{(2\pi)^{1/2}(1-e^{-t\pi})}\cdot 2b^{-1/2}e^{t\pi/2}e^{-t/8}\cdot 4\sqrt{2}\,b \le ke^{-t/8}t^{1/4},$$

where the constant k is about $8\sqrt{2}(\frac{1}{2}\pi^{1/2})^{1/2}(2\pi)^{-1/2} < 5$. Thus for $t \ge 100$ the modulus of (2) in this case is much less than in the case $j = 0$.

The case $j = 3$ of (2): On L_3 the imaginary part of x is identically equal to $-(2N+1)\pi$, hence the denominator of the integrand is $-e^{\operatorname{Re}x} - 1$ which has modulus at least 1. The least value of $|x|$ on L_3 is $(2N+1)\pi$ so $(-x)^{-1/2}$ has modulus less than $(2N+1)^{-1/2}\pi^{-1/2}$ on L_3. The least value of $\operatorname{Im}\log(-x)$ on L_3 is greater than $\operatorname{Im}\log(1+i) = \pi/4$ so $(-x)^{it}$ has modulus at most $e^{-t\pi/4}$. Thus (2) has modulus at most

$$\frac{e^{-t\pi/2}}{(2\pi)^{1/2}(1-e^{-t\pi})}\frac{e^{-t\pi/4}}{(2N+1)^{1/2}\pi^{1/2}}\int_{-(1/2)(\pi t)^{1/2}}^{\infty} e^{-Nu}\,du$$
$$= \frac{e^{-3\pi t/4}}{\pi(1-e^{-t\pi})\sqrt{2}\,(2N+1)^{1/2}}\,\frac{\exp[\frac{1}{2}N(\pi t)^{1/2}]}{N}$$
$$\le e^{-3\pi t/4}\exp\left[\frac{1}{2}\left(\frac{t}{2\pi}\right)^{1/2}(\pi t)^{1/2}\right] \le e^{-t}$$

for $t > 2\pi$. Thus this term is entirely negligible compared to $e^{-t/11}$ when t is large.

In summary, the above very crude estimates suffice to show that the error in the approximation (1) is considerably less than $e^{-t/11}$ for $t \ge 100$.

7.4 FIRST APPROXIMATION TO THE MAIN INTEGRAL

The estimates of the preceding section show that, with an absolute error which is very small and which decreases very rapidly as t increases, the remainder term R in the formula

$$Z(t) = 2 \sum_{n^2<(t/2\pi)} n^{-1/2} \cos[\vartheta(t) - t \log n] + R$$

is approximately equal to a definite integral

$$(1) \qquad R \sim \frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1 - ie^{-\pi t})} \int_{L_1} \frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x - 1},$$

where N is the integer part of $(t/2\pi)^{1/2}$ and where L_1 is a line segment in the complex x-plane which has slope 1, length $(2\pi t)^{1/2}$, and midpoint on the imaginary axis at $i(2\pi t)^{1/2}$. [It is assumed that $(t/2\pi)^{1/2}$ is not an integer so the integrand is not singular on L_1.] The objective of this section is to develop a first approximation to the value of this definite integral.

The argument which led to the above integral, namely, the technique of ignoring the denominator $e^x - 1$ and applying the saddle point method to the numerator, suggests that the path of integration should pass through the saddle point $\alpha = (-\frac{1}{2} + it)/N$ and the numerator should be expanded in powers of $(x - \alpha)$. This has two disadvantages, the first being that α depends on the discrete variable N and the second being that α has a small real part which complicates the computations. Instead set $a = i(2\pi t)^{1/2} = 2\pi i(t/2\pi)^{1/2}$ and expand the numerator in terms of $(x - a)$. This gives

$$\begin{aligned}
&\exp\{(-\tfrac{1}{2} + it)\log(-a) + (-\tfrac{1}{2} + it) \\
&\qquad \times \log[1 + (x - a)/a] - Na - N(x - a)\} \\
&= (-a)^{-(1/2)+it}e^{-Na} \\
&\qquad \times \exp\{[(-\tfrac{1}{2} + it)a^{-1} - N](x - a) \\
&\qquad - (-\tfrac{1}{2} + it)\cdot\tfrac{1}{2}(x - a)^2a^{-2} + \cdots\}.
\end{aligned}$$

Now the coefficient of $(x - a)$ in the exponential is approximately $it/i(2\pi t)^{1/2} - N = (t/2\pi)^{1/2} - N = p$, where p is the fractional part of $(t/2\pi)^{1/2}$. The coefficient of $(x - a)^2$ is approximately $-it\cdot\frac{1}{2}/(-2\pi t) = i/4\pi$. The coefficients of $(x - a)^3, (x - a)^4, \ldots$ are approximately $\pm(1/n)(it)/[i(2\pi t)^{1/2}]^n = \text{const}\, t^{(-n+2)/2}$ and are therefore small for large t. Thus it is natural to write the numerator of the integrand in the form

$$(-a)^{-(1/2)+it}e^{-Na}e^{p(x-a)}e^{i(x-a)^2/4\pi}g(x - a)$$

because then $g(x - a)$ is the exponential of the power series

$$-i\frac{(x - a)^2}{4\pi} - p(x - a) - N(x - a) + \left(-\frac{1}{2} + it\right)\log\left(1 + \frac{x - a}{a}\right)$$

whose coefficients are all small when t is large; the expansion $g(x-a) = \sum_{n=0}^{\infty} b_n(x-a)^n$ has radius of convergence $|a|$ because $x = 0$ is the only singularity of the function it defines, its constant term b_0 is 1, and its remaining coefficients $b_1, b_2, \ldots$ are small for large t. Thus the integral in (1) becomes

$$\frac{e^{-i\vartheta(t)}e^{-t\pi/2}(-a)^{-(1/2)+it}e^{-Na}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1-ie^{-t\pi})}\int_{L_1}\frac{e^{i(x-a)^2/4\pi}e^{p(x-a)}\sum_{n=0}^{\infty} b_n(x-a)^n\,dx}{e^x-1}. \tag{2}$$

The factor $\exp[i(x-a)^2/4\pi]$ is real on L (where, as before, L is the line of which L_1 is a segment), has a maximum of 1 at $x = a$, and decreases rapidly as x moves away from a (for example, at the ends of L_1 it is $\exp\{i[\pm\frac{1}{2}e^{i\pi/4}(2\pi t)^{1/2}]^2/4\pi\} = e^{-t/8}$ which is very small for large t), so this integral is highly concentrated near $x = a$ where the b_0 term dominates. Thus the integral above is approximately

$$\int_L \frac{e^{i(x-a)^2/4\pi}e^{p(x-a)}\,dx}{e^x-1}. \tag{3}$$

Riemann was able to evaluate this definite integral in closed form, and hence, since the factors in front of the integral can be evaluated numerically, he was able to find a numerical approximation to the value of R in (1).

Before evaluating the integral (3) it is advantageous to simplify the expression (2) by taking the change of variable $x = u + 2\pi iN$, $x - a = u + 2\pi iN - 2\pi i(t/2\pi)^{1/2} = u - 2\pi ip$, where p is the fractional part of $(t/2\pi)^{1/2}$. Then (2) takes the form

$$\begin{aligned}
&\frac{e^{-i\vartheta(t)}(e^{i\pi/2})^{it}[-i(2\pi t)^{1/2}]^{-(1/2)+it}e^{-N2\pi i(N+p)}}{(2\pi)^{(1/2)+it}(e^{i\pi/2})^{-1/2}(1-ie^{-t\pi})} \\
&\qquad\times\int_{\Gamma_1}\frac{e^{i(u-2\pi ip)^2/4\pi}e^{p(u-2\pi ip)}\sum b_n(u-2\pi ip)^n\,du}{e^u-1} \\
&= \frac{e^{-i\vartheta(t)}[(t/2\pi)^{1/2}]^{-(1/2)+it}}{2\pi(-i)(1-ie^{-t\pi})}e^{-2\pi iN^2-2\pi iNp-\pi ip^2-2\pi ip^2} \\
&\qquad\times\int_{\Gamma_1}\frac{e^{iu^2/4\pi}e^{2pu}\sum b_n(u-2\pi ip)^n\,du}{e^u-1} \\
&= \left(\frac{t}{2\pi}\right)^{-1/4}\left(\frac{t}{2\pi}\right)^{it/2}e^{-i\vartheta(t)-i\pi(N+p)^2-i\pi N^2-2\pi ip^2} \\
&\qquad\times\frac{1}{(1-ie^{-t\pi})(-1)2\pi i}\int_{\Gamma_1}\frac{e^{iu^2/4\pi}e^{2pu}\sum b_n(u-2\pi ip)^n\,du}{e^u-1}
\end{aligned}$$

where Γ_1 is the line of slope 1 and length $(2\pi t)^{1/2}$, whose midpoint is $2\pi ip$, directed from upper right to lower left. Set

$$U = \frac{\exp\{i[(t/2)\log(t/2\pi)-(t/2)-(\pi/8)-\vartheta(t)]\}}{1-ie^{-t\pi}}.$$

Then, by the formula for $\vartheta(t)$ [(1) of Section 6.5], U is very near 1 for t large and, since $(N+p)^2 = t/2\pi$ and $(-1)^{N^2} = (-1)^N$, (2) takes the form

$$(4) \qquad \left(\frac{t}{2\pi}\right)^{-1/4} U e^{i\pi/8}(-1)^{N-1} e^{-2\pi i p^2} \frac{1}{2\pi i}\int_{\Gamma_1} \frac{e^{iu^2/4\pi} e^{2pu} \sum b_n (u-2\pi i p)^n\, du}{e^u - 1}.$$

Riemann proved that

$$(5) \qquad e^{i\pi/8} e^{-2\pi i p^2} \cdot \frac{1}{2\pi i}\int_{\Gamma} \frac{e^{iu^2/4\pi} e^{2pu}\, du}{e^u - 1} = \frac{\cos 2\pi(p^2 - p - \frac{1}{16})}{\cos 2\pi p},$$

where Γ is a line of slope 1, directed from upper right to lower left, which crosses the imaginary axis between 0 and $2\pi i$. This shows that to a first approximation the remainder R in (1) is

$$(6) \qquad R \sim (-1)^{N-1}\left(\frac{t}{2\pi}\right)^{-1/4} \frac{\cos 2\pi(p^2 - p - \frac{1}{16})}{\cos 2\pi p}$$

which can, of course, be evaluated when t is given (recall that N is the integer part of $(t/2\pi)^{1/2}$ and p the fractional part—the apparent singularities at $p = \frac{1}{4}, \frac{3}{4}$ are discussed below) so that (1) can be used to give a first approximation to $Z(t)$. This is the first term of the Riemann–Siegel formula, the later terms of which will be developed in the next section. The remainder of this section is devoted to the proof of Riemann's formula (5).

Let $\Psi(p)$ denote the left side of (5). Since $\exp(iu^2/4\pi)$ approaches zero very rapidly as $|u| \longrightarrow \infty$ in either direction along Γ, this integral converges for all p and defines an entire† function of the complex variable p. Let D denote the domain of the u-plane bounded by Γ and the line parallel to Γ which crosses the imaginary axis at the point which lies $2\pi i$ below the point where Γ crosses the imaginary axis. Then by the Cauchy integral formula

$$\frac{1}{2\pi i}\int_{\partial D} \frac{e^{iu^2/4\pi} e^{2pu}\, du}{e^u - 1} = \text{value at 0 of } e^{iu^2/4\pi} e^{2pu} \frac{u}{e^u - 1} = 1,$$

while on the other hand this integral over ∂D is

$$\begin{aligned} &\frac{1}{2\pi i}\int_{\Gamma} \frac{e^{iu^2/4\pi} e^{2pu}\, du}{e^u - 1} - \frac{1}{2\pi i}\int_{\Gamma} \frac{e^{i(u-2\pi i)^2/4\pi} e^{2p(u-2\pi i)}\, du}{e^{u-2\pi i} - 1} \\ &= e^{-i\pi/8} e^{2\pi i p^2}\Psi(p) - \frac{e^{-i\pi} e^{-4\pi i p}}{2\pi i}\int_{\Gamma} \frac{e^{iu^2/4\pi} e^{2pu+u}\, du}{e^u - 1} \\ &= e^{-i\pi/8} e^{2\pi i p^2}\Psi(p) + e^{-4\pi i p} e^{-i\pi/8} e^{2\pi i (p+1/2)^2}\Psi(p + \tfrac{1}{2}) \end{aligned}$$

so that equating the two expressions for the integral over ∂D gives a relation between $\Psi(p)$ and $\Psi(p + \frac{1}{2})$, namely,

$$(7) \qquad \begin{aligned} e^{i\pi/8} e^{-2\pi i p^2} &= \Psi(p) + e^{-4\pi i p} e^{2\pi i p} e^{i\pi/2}\Psi(p + \tfrac{1}{2}) \\ &= \Psi(p) + i e^{-2\pi i p}\Psi(p + \tfrac{1}{2}). \end{aligned}$$

†Thus the zeros of the denominator $\cos 2\pi p$ in (5) must be canceled by zeros in the numerator, and, indeed, if p is of the form (odd/4), then $p^2 - \frac{1}{16}$ is of the form (multiple of 8)/16 so $p^2 - p - \frac{1}{16}$ is also (odd/4) and the numerator is zero.

A second relationship between $\Psi(p)$ and $\Psi(p+\frac{1}{2})$ can be found by noting that when the integrals they contain are subtracted, the denominator $e^u - 1$ cancels to give

$$\begin{aligned} &e^{-i\pi/8}e^{2\pi ip^2}\Psi(p) - e^{-i\pi/8}e^{2\pi i(p+1/2)^2}\Psi(p+\tfrac{1}{2}) \\ &= \frac{1}{2\pi i}\int_\Gamma \frac{e^{iu^2/4\pi}[e^{2pu} - e^{2pu+u}]\,du}{e^u - 1} = -\frac{1}{2\pi i}\int_\Gamma e^{iu^2/4\pi}e^{2pu}\,du \\ &= -\frac{1}{2\pi i}\int_\Gamma e^{i(u-4\pi ip)^2/4\pi}e^{i4\pi p^2}\,du = e^{4\pi ip^2}\cdot K, \end{aligned}$$

where K is a constant independent of p; hence

$$\Psi(p) - e^{2\pi ip}\cdot i\cdot\Psi(p+\tfrac{1}{2}) = e^{i\pi/8}e^{2\pi ip^2}\cdot K. \tag{8}$$

With $p = \frac{1}{4}$ the two expressions (7) and (8) for $\Psi(\frac{1}{4}) + \Psi(\frac{3}{4})$ give

$$e^{i\pi/8}e^{-2\pi i/16} = e^{i\pi/8}e^{2\pi i/16}K, \qquad K = e^{-\pi i/4}.$$

If this is used in (8) and if $\Psi(p+\frac{1}{2})$ is eliminated between (7) and (8), one finds

$$e^{2\pi ip}\Psi(p) + e^{-2\pi ip}\Psi(p) = e^{2\pi ip}e^{i\pi/8}e^{-2\pi ip^2} + e^{-2\pi ip}e^{-i\pi/8}e^{2\pi ip^2}$$

which gives the desired expression (5) of $\Psi(p)$ as a quotient of cosines.

In summary, it has been shown that the remainder R has approximately the value (6). To deduce this approximation from the previous approximation (1), the series $\sum b_n(x-a)^n$ in (2) was truncated after the first term $b_0 = 1$, the factor U was replaced by 1, and the domain of integration of the resulting integral was extended to be the entire line of which it is a segment.

7.5 HIGHER ORDER APPROXIMATIONS

The only source of substantial error in the approximations of the preceding section is the truncation of the series $\sum b_n(x-a)^n = 1 + \cdots$ at the first term. In this section Riemann's method for obtaining higher order approximations using the higher order terms of this series will be described.

The computation of the individual coefficients b_n is not difficult. Recall that $\sum b_n(x-a)^n$ is by definition the exponential of the series

$$\begin{aligned} -[i(x-a)^2/4\pi] - (p+N)(x-a)& \\ + (-\tfrac{1}{2} + it)\log[1 + (x-a)/a]&, \end{aligned} \tag{1}$$

where $a = i(2\pi t)^{1/2}$, $p + N = (t/2\pi)^{1/2}$, $0 \le p < 1$, N an integer. Let $\omega = (2\pi/t)^{1/2}$. Then ω is small for t large and the coefficients of the series (1) can be expressed in terms of ω as

$$-\frac{i}{4\pi}(x-a)^2 - \omega^{-1}(x-a) + \left(-\frac{1}{2} + 2\pi i\omega^{-2}\right)\log\left[1 + \frac{\omega}{2\pi i}(x-a)\right]$$
$$= -\frac{1}{2}\frac{\omega}{2\pi i}(x-a) + \frac{1}{4}\left(\frac{\omega}{2\pi i}\right)^2(x-a)^2$$
$$+ \frac{1}{3}\left(\frac{\omega^2}{8\pi^2} + \frac{1}{2\pi i}\right)\left(\frac{\omega}{2\pi i}\right)(x-a)^3 + \cdots$$
$$+ (-1)^{n-1}\frac{1}{n}\left(\frac{\omega^2}{8\pi^2} + \frac{1}{2\pi i}\right)\left(\frac{\omega}{2\pi i}\right)^{n-2}(x-a)^n + \cdots.$$

Thus the coefficients of $(x-a)$, $(x-a)^2$ are monomials of degree 1,2, respectively, in ω, and the coefficients of the higher terms $(x-a)^n$ are binomials in ω whose terms are of degree $n-2$ and n. Since $\sum b_n(x-a)^n$ is the exponential of this series it follows immediately that *b_n is a polynomial in ω of degree at most n in which all terms have degree at least equal to the integer part of* $(n/3)$. [For example, to find b_{14} one could compute the coefficient of $(x-a)^{14}$ in the first 14 powers of the above series, divide by the appropriate factorial, and add. Many of the terms in the coefficient of $(x-a)^{14}$ would have degree 14 in ω; the terms of smallest degree would be those in which the first degree term in front of $(x-a)^3$ is used the maximum number of times, which in the case $n = 14$ will give terms of degree 6 in front of $(x-a)^3$ $(x-a)^3(x-a)^3(x-a)^3(x-a)^2 = (x-a)^{14}$.] The easiest way to compute the b_n explicitly is to make use of the fact that the derivative of the series (1) is

$$-\frac{i\cdot 2(x-a)}{4\pi} - \omega^{-1} + \left(-\frac{1}{2} + 2\pi i\omega^{-2}\right)\frac{1}{1+[(x-a)/a]}\cdot\frac{1}{a}$$
$$= \frac{(x-a)}{2\pi i} - \omega^{-1} + \left(-\frac{1}{2} + 2\pi i\omega^{-2}\right)\frac{1}{2\pi i\omega^{-1} + (x-a)}$$
$$= \frac{2\pi i\omega^{-1}(x-a)+(x-a)^2 - 2\pi i\omega^{-1}[2\pi i\omega^{-1}+(x-a)] + 2\pi i(-\frac{1}{2}+2\pi i\omega^{-2})}{2\pi i[(2\pi i\omega^{-1} + (x-a)]}$$
$$= \frac{(x-a)^2 - \pi i}{2\pi i[(x-a) + 2\pi i\omega^{-1}]}.$$

Thus the logarithmic derivative of $\sum b_n(x-a)^n$ is simply

$$\frac{(\sum b_n(x-a)^n)'}{\sum b_n(x-a)^n} = \frac{(x-a)^2 - \pi i}{2\pi i[(x-a) + 2\pi i\omega^{-1}]}.$$

Hence

$$2\pi i[(x-a) + 2\pi i\omega^{-1}][\textstyle\sum nb_n(x-a)^{n-1}] = [(x-a)^2 - \pi i][\textstyle\sum b_n(x-a)^n] \tag{2}$$

from which it is easy to derive a recursion relation among the b's which makes their computation possible. This computation is a bit long, however, and it will not be needed in what follows.

If the series $\sum b_n(x-a)^n$ is truncated at the Kth term rather than at the 0th term, then the argument of the preceding section leads to the higher order approximation

$$\frac{e^{-i\vartheta(t)}e^{-t\pi/2}}{(2\pi)^{1/2}(2\pi)^{it}e^{-i\pi/4}(1-ie^{-t\pi})}\int_{\Gamma_1}\frac{(-x)^{-(1/2)+it}e^{-Nx}\,dx}{e^x-1} \tag{3}$$
$$\sim(-1)^{N-1}\left(\frac{t}{2\pi}\right)^{-1/4}\cdot U[b_0c_0+b_1c_1+\cdots+b_Kc_K]$$

where

$$c_n=e^{i\pi/8}e^{-2\pi ip^2}\frac{1}{2\pi i}\int_\Gamma\frac{e^{iu^2/4\pi}e^{2pu}}{e^u-1}(u-2\pi ip)^n\,du.$$

Here N, p depend on t as before and Γ is the line of slope 1 through $2\pi ip$ oriented from upper right to lower left. The numbers c_n can be computed by expanding $(u-2\pi ip)^n$ and integrating termwise using the formula

$$\frac{d^k}{dp^k}[e^{2\pi ip^2}\Psi(p)]=\frac{d^k}{dp^k}\left[\frac{e^{i\pi/8}}{2\pi i}\int_\Gamma\frac{e^{iu^2/4\pi}e^{2pu}}{e^u-1}\,du\right]$$
$$=\frac{e^{i\pi/8}}{2\pi i}\int_\Gamma\frac{e^{iu^2/4\pi}e^{2pu}}{e^u-1}(2u)^k\,du.$$

In this way the c_n can be expressed as finite linear combinations of $\exp(2\pi ip^2)\,\Psi^{(k)}(p)$ $(k=0,1,\ldots,n)$ whose coefficients are polynomials in p ($\Psi^{(k)}$ is the kth derivative of Ψ). The easiest way to compute the c_n is to make use of the relationship

$$\Psi(p+y)=e^{i\pi/8}e^{-2\pi i(p+y)^2}\frac{1}{2\pi i}\int_\Gamma\frac{e^{iu^2/4\pi}e^{2(p+y)u}}{e^u-1}\,du$$
$$e^{2\pi iy^2}\Psi(p+y)=e^{i\pi/8}e^{-2\pi ip^2}\frac{1}{2\pi i}\int_\Gamma\frac{e^{iu^2/4\pi}e^{2pu}}{e^u-1}e^{2y(u-2\pi ip)}\,du$$
$$=\sum_{n=0}^\infty e^{i\pi/8}e^{-2\pi ip^2}\frac{1}{2\pi i}\int_\Gamma\frac{e^{iu^2/4\pi}e^{2pu}}{e^u-1}\frac{[2y(u-2\pi ip)]^n}{n!}\,du,$$
$$e^{2\pi iy^2}\sum_{m=0}^\infty\frac{\Psi^{(m)}(p)}{m!}y^m=\sum_{n=0}^\infty\frac{(2y)^n}{n!}c_n, \tag{4}$$

and to equate powers of y. The explicit expressions of the c_n which this relationship yields will not be needed in what follows. All that will be needed is the fact that c_n can be expressed as a linear combination of $\Psi(p), \Psi'(p), \ldots, \Psi^{(n)}(p)$ with coefficients which are independent of t.

In order to use the higher order approximation (3), it is necessary to be able to evaluate $b_0c_0+b_1c_1+\cdots+b_Kc_K$ for given t and K. Riemann devised the following method of accomplishing this without going through the computation of the b_n and c_n. Note that since the b_n are polynomials in ω, one can also regard $b_0c_0 + b_1c_1 + \cdots + b_Kc_K$ as a polynomial in ω and arrange it according to powers of ω. It is natural to do this for two reasons: first because ω is small for t large, so the importance of a term depends on the power of ω it contains; second because any given power of ω no longer occurs in $b_{K+1}, b_{K+2}, \ldots$ once K is sufficiently large, and hence the coefficient of any power of ω is independent of K for K sufficiently large. Riemann's method is a method for finding the coefficient of ω^k in $b_0c_0 + b_1c_1 + \cdots + b_Kc_K$ for large K. By the above observation about the form of the c_n, it is clear that this coefficient can be expressed as a linear combination of $\Psi(p)$, $\Psi'(p), \ldots, \Psi^{(n)}(p)$ with coefficients independent of t.

Let the implicit relation (4) satisfied by the c_n be multiplied by $\sum_{n=0}^{K} n!b_n(2y)^{-n}$. The right side becomes a power series in y in which a finite number of terms contain negative powers of y and in which, by the choice of the multiplier, the constant term (term in y^0) is $b_0c_0 + b_1c_1 + \cdots + b_Kc_K$. The left side becomes a product of three power series in y which, by the commutativity and associativity of multiplication of formal power series, is equal to $\sum \Psi^{(m)}(p)y^m/m!$ times

$$G(y) = \exp 2\pi i y^2 \sum_{n=0}^{K} n!b_n(2y)^{-n}.$$

If the coefficients of the nonpositive powers of y in $G(y)$ can be computed, then it will be a simple matter to find the constant term of its product with $\sum \Psi^{(m)}(p)y^m/m!$ and hence the desired expression $b_0c_0 + \cdots + b_Kc_K$. The essence of the argument below is to use the recurrence relation satisfied by the b_n to find a recurrence relation satisfied by the coefficients of G which makes it possible to compute the terms of G with nonpositive powers of y. More specifically, let the b_n be written as polynomials in ω and let the terms of G be rearranged in the order of powers of ω

$$G(y) = \sum_{j=0}^{\infty} A_j\omega^j$$

in which each A_j is a power series in y with a finite number of terms containing negative powers and in which the sum is actually finite because the largest power of ω in $b_0, b_1, \ldots, b_K$ is ω^K. The A_j depend on K, but if a particular positive integer ν is chosen and if K is large enough that $b_{K+1}, b_{K+2}, \ldots$ contain no terms in $\omega^0, \omega^1, \ldots, \omega^{\nu-1}$, then $A_0, A_1, \ldots, A_{\nu-1}$ are independent of K. The objective is to find the terms with nonpositive powers of y in A_0, $A_1, \ldots, A_{\nu-1}$. Computations will be carried out mod ω^ν so that G can be taken as

$$G(y) = e^{2\pi i y^2} \sum_{n=0}^{\infty} n!b_n(2y)^{-n} \qquad (\text{mod } \omega^\nu)$$

and the main step of the argument is to find a recurrence relation satisfied (mod ω^ν) by G which makes it possible to compute the A_j.

Of course the desired relation must be deduced from the relation (2) satisfied by the b_n, which can be stated

$$2\pi i n b_n - 4\pi^2\omega^{-1}(n+1)b_{n+1} = b_{n-2} - \pi i b_n \tag{5}$$

for $n = 0, 1, 2, 3, \ldots$ (provided b_{-1}, b_{-2} are defined to be zero). The first step is to state these relations among the b_n as a differential equation satisfied by the formal power series

$$F(y) = \sum_{n=0}^{\infty} n!\, b_n (2y)^{-n}.$$

If the nth equation (5) is multiplied by $n!(2y)^{-n-1}$, then the first term on the left becomes the general term of the series $-\pi i DF$ (where D denotes formal differentiation with respect to y), the second term on the left becomes the general term of $-4\pi^2\omega^{-1}F$, the first term on the right becomes the general term of $\frac{1}{4}D^2(2y)^{-1}F = \frac{1}{8}D^2y^{-1}F$, and the second term on the right becomes the general term of $-(\pi i/2y)F$. More precisely

$$\pi i DF = -2\pi i \sum_0^{\infty} n!\, n b_n (2y)^{-n-1},$$

$$4\pi^2\omega^{-1}F = 4\pi^2\omega^{-1} + 4\pi^2\omega^{-1}\sum_0^{\infty} n!(n+1)b_{n+1}(2y)^{-n-1},$$

$$\tfrac{1}{8}D^2y^{-1}F = \sum_0^{\infty} n!\, b_{n-2}(2y)^{-n-1},$$

$$-\tfrac{1}{2}\pi i y^{-1}F = -\pi i \sum_0^{\infty} n!\, b_n (2y)^{-n-1},$$

so the relation (5) for $n = 0, 1, 2, \ldots$ is equivalent to the differential equation

$$\pi i DF + 4\pi^2\omega^{-1}F + \tfrac{1}{8}D^2y^{-1}F - \tfrac{1}{2}\pi i y^{-1}F = 4\pi^2\omega^{-1}$$

for F. Then integration by parts gives a differential equation satisfied by $G(y) = \exp(2\pi i y^2)F(y) \pmod{\omega^\nu}$ as follows.

$$\begin{aligned}
4\pi^2\omega^{-1}e^{2\pi i y^2} &= [\pi i DF + 4\pi^2\omega^{-1}F + \tfrac{1}{8}D^2y^{-1}F - \tfrac{1}{2}\pi i y^{-1}F]e^{2\pi i y^2} \\
&= \pi i DG - \pi i F(De^{2\pi i y^2}) + 4\pi^2\omega^{-1}G + \tfrac{1}{8}D^2y^{-1}G \\
&\quad -\tfrac{1}{8}(y^{-1}F)(D^2e^{2\pi i y^2}) - \tfrac{1}{4}(Dy^{-1}F)(De^{2\pi i y^2}) - \tfrac{1}{2}\pi i y^{-1}G \\
&= \pi i DG + 4\pi^2 yG + 4\pi^2\omega^{-1}G + \tfrac{1}{8}D^2y^{-1}G \\
&\quad -\tfrac{1}{8}(y^{-1}F)(-16\pi^2y^2e^{2\pi i y^2} + 4\pi i e^{2\pi i y^2}) \\
&\quad -\tfrac{1}{4}(-y^{-2}F + y^{-1}DF)(4\pi i y e^{2\pi i y^2}) - \tfrac{1}{2}\pi i y^{-1}G \\
&= \pi i DG + 4\pi^2 yG + 4\pi^2\omega^{-1}G + \tfrac{1}{8}D^2y^{-1}G + 2\pi^2 yG \\
&\quad -\tfrac{1}{2}\pi i y^{-1}G + \pi i y^{-1}G - \pi i DG + \pi i F(De^{2\pi i y^2}) \\
&\quad -\tfrac{1}{2}\pi i y^{-1}G \\
&= 6\pi^2 yG + 4\pi^2\omega^{-1}G + \tfrac{1}{8}D^2y^{-1}G - 4\pi^2 yG
\end{aligned}$$

and finally

$$e^{2\pi i y^2} = G + \tfrac{1}{2}\omega y G + \tfrac{1}{32}\omega\pi^{-2}D^2y^{-1}G$$

mod ω^ν. Since $G = \sum A_j\omega^j \pmod{\omega^\nu}$, this gives

$$e^{2\pi i y^2} = \sum_j A_j\omega^j + \sum_j \left(\frac{y}{2}A_j + \frac{1}{32}\pi^{-2}D^2y^{-1}A_j\right)\omega^{j+1};$$

so the terms in $\omega, \omega^2, \omega^3, \ldots$ cancel on the left, which implies

$$A_j = -\tfrac{1}{2}yA_{j-1} - \tfrac{1}{32}\pi^{-2}D^2y^{-1}A_{j-1} \qquad (j = 1, 2, 3, \ldots).$$

Since $A_0 = \exp(2\pi i y^2)$, this relation makes it possible to compute $A_1, A_2, A_3, \ldots$ in turn. Only the nonpositive powers of y in A_j are of interest, and to determine the coefficient of y^n in A_j it is only necessary to know the coefficients of y^{n+3} and y^{n-1} in A_{j-1}. Thus to determine all nonpositive terms in A_4, it is only necessary to begin with all terms through y^{12} in A_0 after which one easily computes

$$A_0 = 1 + 2\pi i y^2 - 2\pi^2y^4 - \frac{2^2\pi^3iy^6}{3} + \frac{2\pi^4y^8}{3} + \frac{2^2\pi^5iy^{10}}{3\cdot 5} - \frac{2^2\pi^6y^{12}}{3^2\cdot 5},$$

$$A_1 = -\frac{1}{2^4\pi^2y^3} - \frac{y}{2^3} - \frac{\pi i y^3}{2\cdot 3} + \frac{\pi^2y^5}{2^3} + \frac{\pi^3iy^7}{3\cdot 5} - \frac{\pi^4y^9}{2^2\cdot 3^2},$$

$$A_2 = \frac{5}{2^7\pi^4y^6} + \frac{1}{2^5\pi^2y^2} + \frac{i}{2^5\cdot 3\pi} + \frac{y^2}{2^6} + \frac{\pi i y^4}{2^4\cdot 3} - \frac{\pi^2y^6}{2^3\cdot 3^2},$$

$$A_3 = \frac{-5\cdot 7}{2^9\pi^6y^9} - \frac{1}{2^5\pi^4y^5} - \frac{i}{2^93\pi^3y^3} - \frac{1}{2^6\pi^2y} - \frac{7iy}{2^8\cdot 3\pi} + \frac{y^3}{2^7\cdot 3^2},$$

$$A_4 = \frac{5^2\cdot 7\cdot 11}{2^{13}\pi^8y^{12}} + \frac{7\cdot 11}{2^{10}\pi^6y^8} + \frac{5i}{2^{12}\cdot 3\pi^5y^6} + \frac{19}{2^{10}\pi^4y^4} + \frac{i}{2^{10}\cdot 3\pi^3y^2} + \frac{11\cdot 13}{2^{11}3^2\pi^2}.$$

Multiplying by $\sum \Psi^{(m)}(p)y^m/m!$ and taking the constant term of the result then gives as the coefficients of $\omega^0, \omega^1, \omega^2, \omega^3, \omega^4$ in $b_0c_0 + \cdots + b_Kc_K$ the expressions

$$\Psi(p),$$

$$-\frac{\Psi^{(3)}(p)}{2^4\pi^2 3!},$$

$$\frac{5\Psi^{(6)}(p)}{2^7\pi^4 6!} + \frac{\Psi^{(2)}(p)}{2^5\pi^2 2!} + \frac{i\Psi(p)}{2^5\cdot 3\pi},$$

$$-\frac{5\cdot 7\Psi^{(9)}(p)}{2^9\pi^6 9!} - \frac{\Psi^{(5)}(p)}{2^5\pi^4 5!} - \frac{i\Psi^{(3)}(p)}{2^9 3\pi^3 3!} - \frac{\Psi^{(1)}(p)}{2^6\pi^2},$$

$$\frac{5^2\cdot 7\cdot 11\Psi^{(12)}(p)}{2^{13}\pi^8 12!} + \frac{7\cdot 11\cdot\Psi^{(8)}(p)}{2^{10}\pi^6 8!} + \frac{5i\Psi^{(6)}(p)}{2^{12}\cdot 3\pi^5 6!} + \frac{19\Psi^{(4)}(p)}{2^{10}\pi^4 4!} + \frac{i\Psi^{(2)}(p)}{2^{10}\cdot 3\pi^3 2!} + \frac{11\cdot 13\cdot\Psi(p)}{2^{11}3^2\pi^2},$$

respectively, provided, of course, that $K \geq 12$.

Since the actual remainder $R = Z(t) - 2\sum_1^N \cos[\vartheta(t) - t\log n]$ is a real number the imaginary terms in the above formulas must have no significance. In fact if the factor

$$\begin{aligned}\exp\left\{i\left[\frac{t}{2}\log\left(\frac{t}{2\pi}\right) - \frac{t}{2} - \frac{\pi}{8} - \vartheta(t)\right]\right\} \\ &= \exp\left[-\frac{i}{48t} - \frac{7i}{5760t^3} + \cdots\right] \\ &= \exp\left[-\frac{i}{96\pi}\omega^2 + \cdots\right] \\ &= 1 - \frac{i}{96\pi}\omega^2 - \frac{1}{96^2\pi^2\cdot 2}\omega^4 + \cdots\end{aligned}$$

is taken into account, then the imaginary terms in front of ω^2 and ω^3 cancel and the coefficient of ω^4 becomes

$$\frac{1}{2^{23}\cdot 3^5\pi^8}\Psi^{(12)}(p) + \frac{11}{2^{17}\cdot 3^2\cdot 5\pi^6}\Psi^{(8)}(p) + \frac{19}{2^{13}3\pi^4}\Psi^{(4)}(p) + \frac{1}{2^7\pi^2}\Psi(p)$$

which not only eliminates the imaginary terms but also simplifies the coefficient of $\Psi(p)$.

In summary, *the remainder R in the formula*

$$Z(t) = 2\sum_{n^2<(t/2\pi)} n^{-1/2}\cos[\vartheta(t) - t\log n] + R$$

is approximately

$$\begin{aligned}R \sim (-1)^{N-1}\left(\frac{t}{2\pi}\right)^{-1/4} \\ \times\left[C_0 + C_1\left(\frac{t}{2\pi}\right)^{-1/2} + C_2\left(\frac{t}{2\pi}\right)^{-2/2} + C_3\left(\frac{t}{2\pi}\right)^{-3/2} + C_4\left(\frac{t}{2\pi}\right)^{-4/2}\right],\end{aligned}$$

where N is the integer part of $(t/2\pi)^{1/2}$, *p the fractional part, and*

$$C_0 = \Psi(p) = \frac{\cos 2\pi(p^2 - p - \frac{1}{16})}{\cos 2\pi p},$$

$$C_1 = -\frac{1}{2^5 3\pi^2}\Psi^{(3)}(p),$$

$$C_2 = \frac{1}{2^{11}\cdot 3^2\pi^4}\Psi^{(6)}(p) + \frac{1}{2^6\pi^2}\Psi^{(2)}(p),$$

$$C_3 = -\frac{1}{2^{16}\cdot 3^4\pi^6}\Psi^{(9)}(p) - \frac{1}{2^8\cdot 3\cdot 5\pi^4}\Psi^{(5)}(p) - \frac{1}{2^6\pi^2}\Psi^{(1)}(p),$$

$$\begin{aligned}C_4 = \frac{1}{2^{23}\cdot 3^5\pi^8}\Psi^{(12)}(p) + \frac{11}{2^{17}\cdot 3^2\cdot 5\pi^6}\Psi^{(8)}(p) \\ + \frac{19}{2^{13}\cdot 3\pi^4}\Psi^{(4)}(p) + \frac{1}{2^7\pi^2}\Psi(p).\end{aligned}$$

This† is the formula which Siegel found in Riemann's unpublished papers (see Fig. 2). The next section is devoted to some numerical applications of the formula and the following section to an analysis of the error.

7.6 SAMPLE COMPUTATIONS

To get some idea of the accuracy of the Riemann–Siegel formula, consider the computation of $Z(18)$, the value of which was found with three-place accuracy in Section 6.5. If $t = 18$, then $(t/2\pi)^{1/2} = 1.692\,569$, so in this case $N = 1$, $p = 0.692\,569$. The sum approximating $Z(18)$ consists of the single term $2\cos\vartheta(18) = 2\cos(0.080\,911) = 1.993\,457$. The denominator of $\Psi(p)$ is $\cos(2\pi \times 0.692\,569) = \cos(\pi + \frac{1}{3}\pi + 2\pi \times 0.025\,902) = (\sqrt{3}/2)\sin(0.162\,747) - \frac{1}{2}\cos(0.162\,747) = -0.353\,070$ and the numerator is $\cos 2\pi(p^2 - p - \frac{1}{16}) = \cos(-0.275\,417 \times 2\pi) = -\sin(0.159\,700) = -0.159\,022$; so

$$
\begin{aligned}
Z(18) &\sim 2\cos\vartheta(18) + (-1)^{1-1}(18/2\pi)^{-1/4}\Psi(0.692\,569) \\
&= 1.993\,457 + (0.768\,647)\frac{-0.159\,022}{-0.353\,070} \\
&= 1.993\,457 + 0.346\,197 = 2.339\,654
\end{aligned}
$$

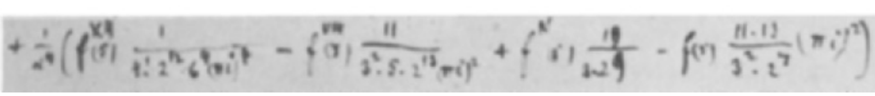
is the first approximation to $Z(18)$. Comparing this with the value 2.337 obtained in Section 6.5 shows that the Riemann–Siegel formula gives better than two-place accuracy even for this relatively small value of t and even when only the first approximation is used!

To use the higher order approximations, one must have some means of evaluating the more complicated functions $C_1, C_2, C_3, \ldots$ of Section 7.5. The simplest method of doing this is to compute the Taylor series coefficients of $\Psi(p)$, from which the Taylor series coefficients of the derivatives of Ψ and hence of the C's are easily computed. Since $0 < p < 1$, it is natural to ex-

†However, Siegel changed the coefficient of $\Psi(p)$ in C_4 from Riemann's value $11 \cdot 13(2^{11}3^2\pi^2)^{-1}$ to $(2^7\pi^2)^{-1}$ as on the preceding page. The above expression of the formula differs somewhat from both Riemann's expression of it and Siegel's (which differ from each other). Riemann expresses the series $\sum C_j\omega^j$ as a series $\sum B_j a^{-j}$, where $a = i(2\pi t)^{1/2}$ as before (α in Riemann's notation) and where, consequently, $B_j = (2\pi i)^j C_j$. Siegel expresses the series $\sum C_j\omega^j$ as a series $\sum A_j t^{-j/2}$, where, consequently, $A_j = C_j(2\pi)^{j/2}$. Moreover, Siegel expresses the A_j in terms of derivatives of the function

$$F(x) = \frac{\cos[x^2 + (3\pi/8)]}{\cos[(2\pi)^{1/2}x]} = \Psi\left[\frac{x}{(2\pi)^{1/2}} + \frac{1}{2}\right]$$

rather than in terms of derivatives of Ψ. His formulas can be deduced from those above using $\Psi^{(n)} = (2\pi)^{n/2}F^{(n)}$.


$$+\frac{1}{2^9}\left(f^{XII}(\xi)\,\frac{1}{4!\,2^{12}\,6^4(\pi i)^4} - f^{VIII}(\xi)\,\frac{11}{3^2\cdot 5\cdot 2^{13}(\pi i)^2} + f^{IV}(\xi)\,\frac{19}{3\cdot 2^9} - f(\xi)\,\frac{11\cdot 13}{3^2\cdot 2^7}(\pi i)^2\right)$$

pand these functions in powers of $p - \frac{1}{2}$. Since $\Psi(\frac{1}{2} + a) = \cos[2\pi a^2 + (3\pi/8)]/\cos 2\pi a$, the expansion of Ψ in powers of $a = p - \frac{1}{2}$ is a quotient of known even power series and, as such, is an even power series whose coefficients can be explicitly computed. Then $C_0 = \Psi, C_2, C_4, \ldots$ will be even power series and $C_1, C_3, C_5, \ldots$ odd power series whose coefficients are easily found. Haselgrove [H8] gives a table of coefficients of C_0, C_1, C_2, C_3, C_4 in powers of $(1 - 2p)$ which is reproduced as Table V. Using these coefficients with the above value of p and therefore with $1 - 2p = -0.385\,138$, one finds easily $C_0 = 0.450\,401$, $C_1 = -0.009\,207$, $C_2 = 0.004\,996$, $C_3 = -0.000\,316$, $C_4 = 0.000\,323$ from which $2\cos\vartheta + (t/2\pi)^{-1/4}[C_0 + (t/2\pi)^{-1/2}C_1 + \cdots + (t/2\pi)^{-2}C_4]$ is $2.336\,796 \sim Z(18)$. Since the C_4 term is a 3 in the fifth place, one might hope for four-place accuracy and indeed this answer is extremely close to Haselgrove's six-place value $Z(18) = 2.336\,800$. Thus the error estimates of Section 7.3 are much too generous and Riemann was in fact in possession of the means to compute $\zeta(\frac{1}{2} + it)$ with amazing accuracy.

Using the Riemann–Siegel formula it is quite easy to locate the first few roots of $\zeta(\frac{1}{2} + it)$ by computation. The main term $2\cos\vartheta(t)$ is zero near $t = 14.5$ where $\vartheta(t)$ is near $-\pi/2$, as a simple computation using formula (2) of Section 6.5 for $\vartheta(t)$ shows. For $t = 14.5$ simple estimates give $t/2\pi \sim 2.30$, $(t/2\pi)^{1/2} \sim 1.5$, $N = 1$, $p \sim \frac{1}{2}$. Since $\Psi(\frac{1}{2}) = -\cos(5\pi/8) \sim 0.38$ and $(t/2\pi)^{-1/4} \sim (2/3)^{1/2} \sim 0.8$, the first correction term $(t/2\pi)^{-1/4}\Psi(p)$ is about $(0.8)(0.38) \sim 0.30$; so to move toward a zero of Z, the value of t should be reduced to where the main term $2\cos\vartheta(t)$ is -0.30. The derivative of this term is $-2\vartheta'(t)\sin\vartheta(t) \sim -2\cdot\frac{1}{2}\log(t/2\pi)\cdot(-1) \sim 0.83$, so t should be reduced from 14.5 to about $14.5 - [(0.30)/(0.83)] = 14.14$. Thus there might well be a root between 14.1 and 14.2. Now $Z(14.1)$ and $Z(14.2)$ can be computed by exactly the same method as was used for $Z(18)$ above. The results are shown in Table VI. The size of the C_4 term suggests an accuracy of about four places, and this is more than confirmed by Haselgrove's tables, which give $Z(14.1) = -0.027\,463$, $Z(14.2) = +0.052\,045$. Thus there is definitely a root in the interval and linear interpolation would place it at $14.1 + h$, where $h/0.1 = (0.027\,466)/(0.027\,466 + 0.052\,042) \sim 0.345$ so the root is at about 14.1345. Further computations with t in this range show that the

---

Fig. 2 This is the sheet on which the Riemann–Siegel formula appears in Riemann's unpublished papers in the Göttingen University Library. (Here it is somewhat reduced in size.) The enlargement shows the final terms of the formula, which include the coefficient that Siegel simplified. The lack of coherent organization and of any explanation are typical of these papers, which include, along with the unexplained formulas, various random jottings such as the Chebyshev note on p. 5 and a computation of $\sqrt{2}$ to 38 decimal places. (Reproduced with the permission of the Niedersächsische Staats- und Universitätsbibliothek, Handschriftenabteilung, Göttingen.)

TABLE V

TABLE OF COEFFICIENTS[a]

s	$c_{0,s}$	s	$c_{1,s}$
0	+0·38268 34323 65089 77173	1	+0·02682 51026 28375 35
2	+ ·43724 04680 77520 44936	3	− 1378 47734 26351 85
4	+ ·13237 65754 80343 52333	5	− 3849 12504 82235 08
6	− 1360 50260 47674 18865	7	− 987 10662 99062 08
8	− 1356 76219 70103 58088	9	+ 331 07597 60858 40
10	− 162 37253 23144 46528	11	+ 146 47808 57795 42
12	+ 29 70535 37333 79691	13	+ 1 32079 40624 88
14	+ 7 94330 08795 21469	15	− 5 92274 87018 47
16	+ 4655 61246 14504	17	− 59802 42585 37
18	− 14327 25163 09551	19	+ 9641 32245 62
20	− 1035 48471 12314	21	+ 1833 47337 22
22	+ 123 57927 08384	23	− 44 67087 57
24	+ 17 88108 38577	25	− 27 09635 09
26	− 33914 14393	27	− 77852 89
28	− 16326 63392	29	+ 23437 63
30	− 378 51094	31	+ 1583 02
32	+ 93 27423	33	− 121 20
34	+ 5 22184	35	− 14 58
36	− 33506	37	+ 29
38	− 3412	39	+ 9
40	+ 58		
42	+ 15		
	Sum $= A$		Sum $= \frac{1}{8}\pi B - \frac{1}{4}A$

s	$c_{2,s}$	s	$c_{3,s}$	s	$c_{4,s}$
	0·0		0·00		0·00
0	+0518 85428 30293	1	+133 97160 907	0	+046 48338 9
2	+ 30 94658 38807	3	−374 42151 364	2	−100 56607 4
4	−1133 59410 78229	5	+133 03178 920	4	+ 24 04485 6
6	+ 223 30457 41958	7	+226 54660 765	6	+102 83086 1
8	+ 519 66374 08862	9	− 95 48499 998	8	− 76 57860 9
10	+ 34 39914 40762	11	− 60 10038 459	10	− 20 36528 6
12	− 59 10648 42747	13	+ 10 12885 828	12	+ 23 21229 0
14	− 10 22997 25479	15	+ 6 86573 345	14	+ 3 26021 5
16	+ 2 08883 92217	17	− 5985 366	16	− 2 55790 5
18	+ 59276 65493	19	− 33316 599	18	− 41074 6
20	− 1642 38384	21	− 2191 929	20	+ 11781 2
22	− 1516 11998	23	+ 789 089	22	+ 2445 6
24	− 59 07803	25	+ 94 147	24	− 239 2
26	+ 20 91151	27	− 9 570	26	− 75 0
28	+ 1 78157	29	− 1 876	28	+ 1 3
30	− 16164	31	+ 45	30	+ 1 4
32	− 2380	33	+ 22		
34	+ 54				
36	+ 20				
	Sum $= B/96\pi$				Sum $= -A/18432\pi^2$

[a]From Haselgrove [H8].

Main term	1.993 457
C_0 term	0.346 199
C_1 term	−0.004 181
C_2 term	0.001 341
C_3 term	−0.000 050
C_4 term	0.000 030
$Z(18)$	2.336 796

Computation of the approximation to $Z(18)$.

value of Z given by the Riemann–Siegel formula with terms through C_4 changes sign between 14.134 727 and 14.134 729. Thus the Riemann–Siegel formula permits the computation of the first root $\frac{1}{2} + i\alpha$, $\alpha = 14.134\,725\ldots$ (see Section 6.1) with at least five-place accuracy. If the C_5 and C_6 terms were used, it is possible that even greater accuracy might be achieved. *Riemann computed this root*, finding its value to be 14.1386; the error in his value results, as the above shows, not from the inherent inaccuracy of the Riemann–Siegel formula, but merely from the fact that he must have carried out only rough computations.

Riemann also took some steps toward proving that this root 14.1. . . is the *first* root. By (4) of Section 3.8

$$\sum_{\operatorname{Im}\rho>0}\left(\frac{1}{\rho}+\frac{1}{1-\rho}\right)=1+\frac{1}{2}\gamma-\frac{1}{2}\log\pi-\log 2.$$

TABLE VI

t	14.1	14.2
$\vartheta(t)$	−1.742 722	−1.702 141
p	0.498 027	0.503 330
$1-2p$	0.003 946	−0.006 660
$(t/2\pi)^{-1/2}$	0.667 545	0.665 190
$(t/2\pi)^{-1/4}$	0.817 034	0.815 591
Main term	−0.342 160	−0.261 934
C_0 term	0.312 671	0.312 129
C_1 term	0.000 058	−0.000 097
C_2 term	0.001 889	0.001 872
C_3 term	0.000 001	−0.000 002
C_4 term	0.000 075	0.000 074
$Z(t)$	−0.027 466	+0.052 042

Hence†

$$\sum_{\operatorname{Im}\rho>0} [1/\rho(1-\rho)] = 0.02309\ldots. \tag{1}$$

Now the root $\rho = \frac{1}{2} + i14.1\ldots$ already found accounts for about 0.005 of the sum of the right. If there were a root in the upper halfplane with smaller imaginary part than this one, then there would have to be *two* such roots, either because it would not be on the line $\operatorname{Re} s = \frac{1}{2}$, in which case there would be a symmetrical root on the opposite side of the line, or because it would be on the line, in which case $Z(t)$ would have to change sign a second time in order to be negative at 14.1 and at 0. Thus such a root would have to satisfy

$$1/\gamma^2 < \tfrac{1}{2}(0.018) = 0.009, \qquad \gamma > 10.$$

Using the Riemann–Siegel formula it is not difficult to see that such a root on the line is very improbable and therefore that the root just found is probably the first root on the line. If all 10 of the roots in the range $0 < \operatorname{Im}\rho < 50$ are located, they account for about 0.0136 of the total in (1) and therefore suffice to prove that the above root is indeed the root in the upper halfplane with the least imaginary part.

The next root on the line would be expected in the vicinity of the next zero of $\cos\vartheta(t)$, which occurs when $\vartheta(t) \sim \pi/2$, $t \sim 20.7$. Assuming that the Riemann–Siegel formula is accurate, it is easy to prove that $Z(t)$ does indeed change from $+$ back to $-$ near this point and to locate the root $21.022\ldots$ quite accurately. Of greater interest, however, is the *next* root, which occurs near $\vartheta(t) \sim 3\pi/2$, $t \sim 25.5$, because in this vicinity N increases from 1 to 2 and the approximation to $Z(t)$ passes through an apparent discontinuity. However, the discontinuity is illusory because if $(t/2\pi)^{1/2} = 2 - \epsilon$, then $N = 1$, p is nearly 1, and therefore $(-1)^{N-1}(t/2\pi)^{-1/4}\Psi(p)$ is $2^{-1/2}\cos(\pi/8)$. On the other hand, if $(t/2\pi)^{1/2} = 2 + \epsilon$, then the main sum contains a second term, namely $2\cdot 2^{-1/2}\cos[\vartheta(t) - t\log 2]$, and $(-1)^{N-1}(t/2\pi)^{-1/4}\Psi(p)$ changes sign to become $-2^{-1/2}\cos(\pi/8)$. But $\log 2 \sim \log(t/2\pi)^{1/2} = \frac{1}{2}\log(t/2\pi)$ so

$$\begin{aligned}\vartheta(t) - t\log 2 &\sim \frac{t}{2}\log\left(\frac{t}{2\pi}\right) - \frac{t}{2} - \frac{\pi}{8} + \frac{1}{48t} - \frac{t}{2}\log\left(\frac{t}{2\pi}\right)\\ &= -2^2\pi - \frac{\pi}{8} + \frac{1}{48\cdot 2^2\cdot 2\pi},\end{aligned}$$

$$\cos[\vartheta(t) - t\log 2] \sim \cos(\pi/8),$$

†The numerical value of Euler's constant γ can be found by using logarithmic differentiation of $\Pi(s) = s\Pi(s-1)$ to find $\gamma = -\Pi'(0)/\Pi(0) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + (1/n) - \Pi'(n)/\Pi(n)$ and by then using formula (4) of Section 6.3. As Siegel reports, Riemann wrote down the above constant to 20 decimal places 0.02309 57089 66121 03381. Thus, although Riemann did not prove that the series (4) of Section 1.10 *converges*, he did compute its sum to 20 decimal places. (I have not checked the accuracy of his answer.)