Infinitely many Stokes smoothings in the gamma function

By M. V. Berry
H. H. Wills Physics Laboratory, Tyndall Avenue, Bristol BS8 1TL, U.K.

The Stokes lines for $\Gamma(z)$ are the positive and negative imaginary axes, where all terms in the divergent asymptotic expansion for $\ln \Gamma(z)$ have the same phase. On crossing these lines from the right to the left half-plane, infinitely many subdominant exponentials appear, rather than the usual one. The exponentials increase in magnitude towards the negative real axis (anti-Stokes line), where they add to produce the poles of $\Gamma(z)$. Corresponding to each small exponential is a separate component asymptotic series in the expansion for $\ln \Gamma(z)$. If each is truncated near its least term, its exponential switches on smoothly across the Stokes lines according to the universal error-function law. By appropriate subtractions from $\ln \Gamma(z)$, the switching-on of successively smaller exponentials can be revealed. The procedure is illustrated by numerical computations.

1. Introduction

My purpose is to draw attention to, and explore in detail, a peculiarity of the asymptotics of $\Gamma(z)$ for large $|z|$. Motivated by Stirling's formula, we write

$$\Gamma(z) = \sqrt{(2\pi)}\, z^{z-\frac{1}{2}} \exp\{-z\} \exp\{g(z)\} \tag{1}$$

and study $g(z)$ for large $|z|$. In the right half-plane, i.e. when $|\arg z| < \frac{1}{2}\pi$, $g(z) \to 0$ as $|z| \to \infty$. This follows from the formal (because divergent) asymptotic expansion

$$g(z) = \sum_{r=0}^{\infty} \frac{B_{2r+2}}{2(r+1)(2r+1)\, z^{2r+1}}, \tag{2}$$

where B_n are the Bernoulli numbers (Gradshteyn & Ryzhik 1980). For real z the series is alternating, because successive even Bernoulli numbers have opposite signs. Thus all terms have the same phase on the positive and negative imaginary axes, so these are the Stokes lines of $\Gamma(z)$. According to the now well-understood Stokes phenomenon (Stokes 1864; Dingle 1973), we expect a small exponential to appear across each of these lines; for simplicity, and without loss of generality, we henceforth consider just the positive imaginary axis.

The unusual feature of $\Gamma(z)$ is that not one but infinitely many small exponentials appear. That this must happen can be shown by the following version of an argument by Paris & Wood (1991), based not on asymptotics but the fundamental reflection formula

$$\Gamma(-z) = -[\pi/z \sin(\pi z)]\, \Gamma(z). \tag{3}$$

Substitution into (1) gives, after a short calculation,

Proc. R. Soc. Lond. A (1991) **434**, 465–472
Printed in Great Britain

$$g(z) = -g(z\exp\{-\mathrm{i}\pi\}) - \ln(1-\exp\{2\pi\mathrm{i}z\})$$
$$= -g(z\exp\{-\mathrm{i}\pi\}) + \sum_{n=1}^{\infty} \frac{\exp\{2\pi\mathrm{i}nz\}}{n}. \tag{4}$$

Continuation has produced an infinite string of exponentials, which in the upper half-plane are subdominant relative to the series (2) which is of order z^{-1}, multiplying $\exp(0)$. Thus $g(z)$ cannot be the odd function that (2) would imply if the series converged, and the vanishing of $g(z)$ as $|z| \to \infty$ cannot continue to hold throughout the left half-plane. In particular, as z approaches the negative real axis (anti-Stokes line) from above, that is when $\arg z = \pi - \epsilon$, the exponentials add to generate the poles of $\Gamma(z)$.

We expect all these subdominant exponentials to arise out of the remainder when the divergent series (2) is truncated. If the truncation is optimal, that is near the least term, we further expect the exponentials to appear where they are smallest, that is across the Stokes line $\arg z = \frac{1}{2}\pi$, and to switch on smoothly according to the universal error-function law (Berry 1989). Paris & Wood (1991) proved that the leading subdominant exponential ($n = 1$ in equation (4)) is indeed born in this way. Here it will be shown how all the smaller exponentials appear similarly, and how their smooth births can be detected in $\Gamma(z)$ by a sequence of increasingly delicate subtractions.

2. A convergent series of exactly terminated asymptotic expansions

We begin with formal manipulations of the asymptotic series (2). Expressing the Bernoulli numbers in terms of the Riemann zeta function, replacing this by its (convergent) Dirichlet series and interchanging summations, we obtain successively

$$g(z) = 2z \sum_{r=0}^{\infty} \frac{(-1)^r \zeta(2r+2)(2r)!}{(2\pi z)^{2r+2}}$$
$$= 2z \sum_{n=1}^{\infty} \sum_{r=0}^{\infty} \frac{(-1)^r (2r)!}{(2\pi z n)^{2r+2}}. \tag{5}$$

Thus $g(z)$ is a series, labelled by n, of asymptotic expansions whose terms, labelled by r, take the standard form of factorials divided by powers. The divergent tail of the series labelled n will generate the nth subdominant exponential in (4). Anticipating this, and following Dingle (1973), the quantities raised to powers, namely

$$nF \equiv -2\pi\mathrm{i}nz, \tag{6}$$

will be interpreted as 'singulants', that is the difference between the dominant exponent (zero) and the nth subdominant exponent $2\pi\mathrm{i}nz$ (cf. (4)). The singulants are positive real on the Stokes line $\arg z = \frac{1}{2}\pi$.

To convert the formal expression (5) into an exact explicit representation of $g(z)$, we truncate each asymptotic series at an arbitrary point N_n, and apply Borel summation to the divergent tail. Thus we obtain the convergent series of terminated expansions

$$g(z) = \sum_{n=1}^{\infty} \left\{ \frac{1}{\mathrm{i}\pi n^2 F} \sum_{r=0}^{N_n-1} \frac{(2r)!}{(nF)^{2r}} + \frac{\exp(-nF)}{n} M(nF, N_n) \right\}, \tag{7}$$

where M denotes the 'terminant' integral

$$M(F,\,N) \equiv \frac{1}{\mathrm{i}\pi}\int_0^\infty \mathrm{d}u \frac{u^{2N}}{1-\mathrm{i}\epsilon - u^2}\exp\{F(1-u)\} \tag{8}$$

(where ϵ is a real positive infinitesimal ensuring that the contour passes above the pole at $u = 1$). Appendix A gives an alternative derivation of (7), less illuminating but not involving the summation of divergent series. The special case where all truncations N_n are the same was previously obtained formally by Dingle (1973, p. 435), also using Borel summation, and proved by Paris & Wood (1991), using a Mellin-transform representation of $\Gamma(z)$.

Crossing the Stokes line with arg z increasing through $\frac{1}{2}\pi$ corresponds to Im F increasing through zero. Optimal (i.e. least term) truncation of the nth series in (7) corresponds to choosing

$$N_n = \mathrm{Int}\,(\tfrac{1}{2}n|F|) = \mathrm{Int}\,(n\pi|z|). \tag{9}$$

The appropriate approximation of M, obtained as explained by Berry (1989) (after replacing $1-u^2$ by $2(1-u)$), is

$$\left.\begin{aligned} &M(F,N) \approx \tfrac{1}{2}[1+\mathrm{Erf}\,(\mathrm{Im}\,F/\surd(2\,\mathrm{Re}\,F))],\\ &\text{if } N \approx \tfrac{1}{2}|F|,\, |\mathrm{Im}\,F| \ll \mathrm{Re}\,F,\, \mathrm{Re}\,F \gg 1, \end{aligned}\right\} \tag{10}$$

where Erf denotes the error function. As ImF increases, M approaches unity, confirming that the exponentials in (7) can be identified as those in (4).

3. Exposing successive smoothings by subtraction

In order to see the smooth switching-on of the nth small exponential in (7), it is necessary to subtract from $g(z)$ all larger exponentials, and all larger terms of the asymptotic series in powers of $1/(nF)$. To achieve the latter, it is necessary to regroup the terms in the double sum over n and r, as follows:

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{r=0}^{N_n-1}\ldots &= \sum_{r=0}^{N_1-1}\sum_{n=1}^{\infty}\ldots + \sum_{r=N_1}^{N_2-1}\sum_{n=2}^{\infty}\ldots + \sum_{r=N_2}^{N_3-1}\sum_{n=3}^{\infty}\ldots \\ &= \sum_{m=1}^{\infty}\sum_{r=N_{m-1}}^{N_m-1}\sum_{n=m}^{\infty}\ldots \quad (\text{where } N_0 \equiv 0). \end{aligned} \tag{11}$$

From (7), the sums over n have the form

$$\sum_{n=m}^{\infty}\frac{1}{n^{2r+2}} = \zeta(2r+2) - \sum_{n=1}^{m-1}\frac{1}{n^{2r+2}} \equiv \zeta_m(2r+2). \tag{12}$$

Thus

$$g(z) = \frac{1}{\mathrm{i}\pi F}\sum_{m=1}^{\infty}\sum_{r=N_{m-1}}^{N_m-1}\frac{(2r)!}{F^{2r}}\zeta_m(2r+2) + \sum_{m=1}^{\infty}\frac{\exp(-mF)}{m}M(mF,\,N_m). \tag{13}$$

This reordered expression is still convergent, and holds for any choice of truncations N_m.

Now we make the optimal choice (9), and seek to expose the nth small exponential by subtracting from $g(z)$ the larger subdominant exponentials $m = 1$ through $n-1$ and the asymptotic series labelled $m = 1$ through n. Thus it is natural to define

$$\begin{aligned} L_n(z) &\equiv n \exp(nF) \\ &\times \left[g(z) - \frac{1}{\mathrm{i}\pi F} \sum_{m=1}^{n} \sum_{r=N_{m-1}}^{N_m - 1} \frac{(2r)!}{F^{2r}} \zeta_m(2r+2) - \sum_{m=1}^{n-1} \frac{\exp(-mF)}{m} M(mF, N_m) \right] \\ &= [M(nF, N_n) + R_n(F)]. \end{aligned} \tag{14}$$

The remainder

$$R_n(F) = n \exp(nF) \sum_{m=n+1}^{\infty} \left[\frac{\exp(-mF)}{m} M(mF, N_m) + \frac{1}{\mathrm{i}\pi F} \sum_{r=N_{m-1}}^{N_m - 1} \frac{(2r)!}{F^{2r}} \zeta_m(2r+2) \right] \tag{15}$$

must vanish asymptotically, that is as $\operatorname{Re} F \to \infty$, if the subtraction is to succeed. We consider separately the terms involving M and those involving ζ_m. Obviously the terms in the first class vanish asymptotically, because the decay of all the exponentials involving m is faster than the growth of that involving n, and the terminant integrals M remain bounded (cf. (10)). In the second class of terms, the sum over r is dominated by its first term $m = n+1$, $r = N_n$. Using (9) and estimating $\zeta_m(2r+2)$ by $m^{-(2r+2)}$, we find that these terms also vanish asymptotically, as

$$\exp\{-n[|F| \ln(1+1/n) + |F| - \operatorname{Re} F]\}. \tag{16}$$

These estimates are confirmed by a more refined analysis.

Taking the asymptotic limit of M on the right-hand side of (14), and defining

$$z \equiv |z| \exp(\mathrm{i}\theta), \tag{17}$$

we therefore find that across the Stokes line $\theta = \frac{1}{2}\pi$ the nth small exponential switches on like

$$L_n(z) \xrightarrow[\text{as } |z| \to \infty]{} \tfrac{1}{2}[1 + \operatorname{Erf}\{(\theta - \tfrac{1}{2}\pi)\surd(\pi n |z|)\}]. \tag{18}$$

This is our main result. It shows that all the subdominant exponentials switch on smoothly in the universal manner. Because of the increasing singulants the higher switchings get sharper (over a θ range proportional to $1/\surd n$). In evaluating the limit of L_n, it is important not to substitute the error functions for the functions M, in the definition (14), corresponding to the larger subdominant exponentials $m < n$, because the subtractions will succeed only if the Ms are controlled to exponential accuracy, whereas the error function is an approximation valid only up to terms of order $1/\surd|F|$ (Berry 1989; Olver 1990). Paris & Wood (1991) prove the special case of (18) with $n = 1$.

4. Numerical illustration

We computed $\Gamma(z)$, and the subtractions $L_n(z)$ defined by (14), for $0 \leqslant \theta = \arg z \leqslant \pi$ and two values of the singulant modulus. These were $|F| = 5$, corresponding to $|z| = 0.79577\ldots$, and $|F| = 10$, corresponding to $|z| = 1.59155\ldots$. Such small values were chosen deliberately, to illustrate how quickly the limit (18) is approached.

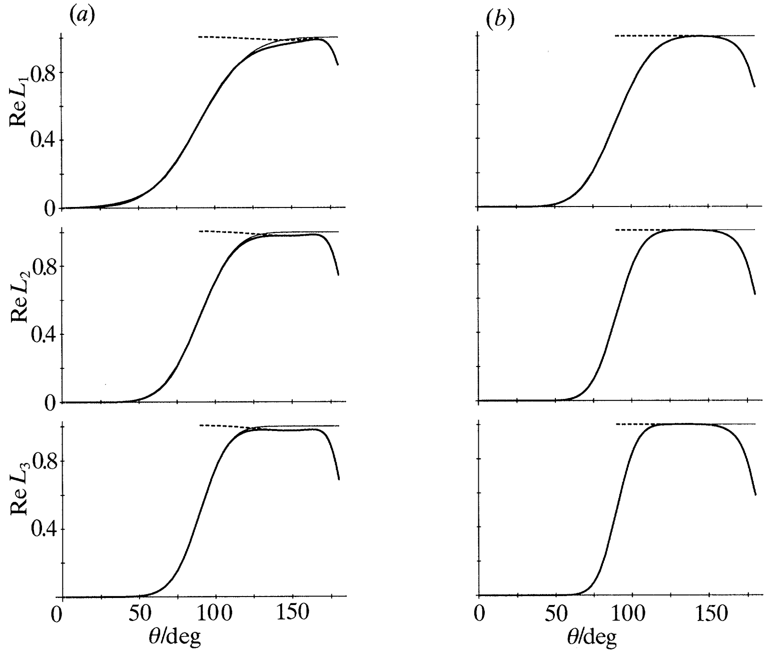

Figure 1. First three subtractions of the gamma function, for (*a*) $|F| = 5$, (*b*) $|F| = 10$. The thick lines show $\operatorname{Re} L_n$, computed from equation (14), for $n = 1$, 2, and 3. The thin lines show the corresponding error-function predictions, computed from equation (18). The dashed lines show the approximation (19) which should apply near the anti-Stokes line $\theta = \pi$ (the apparent straightness in (*b*) of these lines, and also of the thin lines, is an illusion resulting from the larger value of $|F|$). The abscissa is $\theta = \arg z$, so that $\theta = 0°$ corresponds to the positive real z axis, $\theta = 90°$ to the positive imaginary z axis (Stokes line), and $\theta = 180°$ to the negative real z axis (anti-Stokes line).

To compute the terminant integrals (8), the denominator in the integrand was expanded in partial fractions, to give two integrals which were then evaluated in terms of exponential integrals as explained in Appendix B of Berry & Howls (1990).

Figure 1 shows the graphs of $\operatorname{Re} L_n$ (thick line), and the error function (thin line) predicted by equation (18), for the first three subtractions, that is $n = 1$, 2, and 3. The agreement is very close before and across the Stokes line, indicating unambiguously that the error function correctly describes the smoothing not only for the leading subdominant exponential but for all the smaller ones. Note that when $|F| = 5$ we are accurately calculating the factorial of an argument with *negative* real part using asymptotics that becomes exact in the limit of large positive real part! The correctness of the subtraction procedure based on (14) is confirmed by the fact that for $F = 10$, $n = 3$ it successfully uncovers a buried exponential of magnitude $\exp(-30)$ hidden behind two larger ones.

However, it is clear from figure 1 that the agreement degrades on approach to the anti-Stokes line $\theta = \pi$. The reason is not connected with the Stokes smoothing, but indicates the failure of the subtraction procedure. This relies on the rapid decrease of $|\exp(-nF)|$ for increasing n, and fails because the exponentials are all of comparable magnitude near $\theta = \pi$ and cannot be separated. To calculate the correct form for L_n in this region, we note that equation (14) simplifies considerably, as follows. The terminants M can be set equal to unity (cf. (10)), the asymptotic series (double sum over m and r) can be neglected (because $|F|$ is large), and $g(z)$ can be set equal to the sum of exponentials (cf. (4) with $g(z\exp\{-\mathrm{i}\pi\})$ neglected because $z\exp\{-\mathrm{i}\pi\}$ is near the positive real axis when z is near the negative real axis) generated by the Stokes phenomenon. Thus

$$L_n(z) \xrightarrow[\arg z \to \pi]{} n\exp(nF)\left[-\ln\{1-\exp(-F)\} - \sum_{m=1}^{n-1}\frac{\exp(-mF)}{m}\right]. \tag{19}$$

The real part of this quantity is plotted as the dashed lines in figure 1; these coincide with the thick lines as $\theta \to \pi$, and evidently give very accurate approximations to the L_n calculated from (14).

5. Concluding remarks

In the hierarchy of functions with divergent asymptotic expansions, it is common for the late terms to be increasingly well approximated by a factorial divided by a power. Among such functions, $\Gamma(z)$ occupies a peculiar position, intermediate between the simplest and the most general cases. In the simplest cases (exemplified by the exponential integral and the error function), the approximation is exact: *every* term is a factorial divided by a power, and Borel summation terminates the truncated series exactly, with an integral belonging to a class which includes (8) (Dingle 1973) describing the appearance of a subdominant exponential across a Stokes line.

In the most general case (exemplified by typical solutions of one-dimensional Schrödinger equations, or integrals of exponentials with several saddles), the 'factorial divided by power' approximation is simply the first term of a *divergent* series for the late terms. This happens when the subdominant exponential is itself the first term of an asymptotic series, rather than standing alone as in the simplest cases. Such linking of dominant and subdominant divergent series is 'resurgence', and is the basis of very accurate 'hyperasymptotic' schemes (Berry & Howls 1990, 1991), based for example on iterated Borel summation.

What is distinctive about $\Gamma(z)$ is that each term (labelled r) in its asymptotic expansion (5) is a *convergent* infinite series of factorials divided by powers. As we have seen, each component asymptotic series can be resummed and describes a different subdominant exponential, whose simultaneous appearances across the Stokes line can nevertheless be separately monitored by the subtraction procedure we have described.

The gamma function has other peculiarities, which become evident on considering the problem of calculating the factorial of a large integer k to high accuracy by truncating the asymptotic expansion (5) near its least term. This would require knowledge of $(2r)! \approx [\mathrm{Int}\,|F|]! \approx [\mathrm{Int}\,(2\pi k)]!$, that is the factorial of an integer considerably larger than k itself. (This seems a defect of the approximation, but it might be possible to turn it to advantage by using the series in reverse, to calculate $[\mathrm{Int}\,(2\pi k)]!$ given $k!$.)

Even if the series (5) could be employed to its least term (for example with the later terms more roughly approximated) the optimal absolute error in $k!$ would increase with k, for the simple reason that the exponentially small error in $g(z)$ is multiplied by the factorially large prefactor in (1). This optimal error is

$$\epsilon(k) \equiv |k! - (k!)_{\mathrm{approx}}| \approx k^{-1}\sqrt{(2/\pi)}\exp\{k(\ln k - 1 - 2\pi)\}. \tag{20}$$

The minimum value of $\epsilon(k)$ is approximately $\exp\{-(\exp(2\pi)+2\pi)\} \approx 10^{-235}$, and occurs for $k \approx \exp(2\pi) \approx 535$. $\epsilon(k)$ reaches unity when $k \approx 1463$ (slightly greater than

$\exp(2\pi+1)$). For larger values of k, the absolute error exceeds unity. This apparent inability to calculate $k!$ efficiently with an error less than unity frustrates one possible method of testing large k for primality, based on Wilson's theorem (Baker 1984). This states that k is prime if and only if it is a solution of the equation

$$(k-1)! = -1 \pmod{k}. \tag{21}$$

Finally, it is worth noting that the exponential refinements described here are obscured if one uses the saddle-point method to approximate $\Gamma(z)$, rather than $\ln\Gamma(z)$. Starting from the familiar representation

$$\Gamma(z) = [\exp(2\pi i z)-1]^{-1}\int_{\mathrm{C}} \mathrm{d}t\, t^{z-1}\exp(-t), \tag{22}$$

where C is the Hankel contour, and changing to a new variable defined by $t \equiv \exp(u)$, we find an infinite string of saddles at $u = \ln z + 2\pi i n$. The contour links valleys at $u = \infty$ and $u = \infty + 2\pi i$. When $\arg z < \frac{1}{2}\pi$, this can be deformed to a steepest path through two of these saddles, whose contributions combine to cancel the prefactor in (22), leaving in lowest-order Stirling's formula (equation (1) with $g = 1$). When $\arg z > \frac{1}{2}\pi$ the steepest path passes through only one saddle, leaving the prefactor uncancelled to give the poles for z negative real. Thus by this method a single Stokes phenomenon suffices to generate the poles, and as Paris & Wood (1991) suggest, recovery of the smaller exponentials, so easy with $\ln\Gamma(z)$, might require hyperasymptotics (Berry & Howls 1991).

I thank Richard Paris and Alastair Wood for introducing me to this problem and kindly showing me their paper before publication, Christopher Howls for a careful reading of the manuscript and several corrections, and Jonathan Keating for a helpful suggestion.

Appendix A. Alternative derivation of equation (7)

We use 'Binet's second representation' for $g(z)$ (Dingle 1973, pp. 65–66), namely

$$g(z) = 2\int_0^\infty \mathrm{d}t\,\frac{\arctan(t/z)}{\exp(2\pi t)-1}. \tag{A 1}$$

With suitable choice of branch for arctan, this is valid throughout the required region of $\arg z$. Expanding the denominator and integrating once by parts gives

$$g(z) = 2\sum_{n=1}^{\infty}\frac{1}{(2\pi n)^2 z}\int_0^\infty \mathrm{d}v\,\frac{\exp(-v)}{1+v^2/(2\pi n z)^2}, \tag{A 2}$$

where the contour always passes above the pole in the right half-plane. Writing

$$\frac{1}{1+v^2/(2\pi n z)^2} = \sum_{r=1}^{N_n-1}\left(\frac{v}{-2\pi i n z}\right)^{2r} + \frac{(-v/2\pi i n z)^{2N_n}}{1+v^2/(2\pi n z)^2} \tag{A 3}$$

and making a simple change of variable, we at once obtain (7).

References

Baker, A. 1984 *A concise introduction to the theory of numbers*. Cambridge University Press.
Berry, M. V. 1989 *Proc. R. Soc. Lond.* A **422**, 7–21.
Berry, M. V. & Howls, C. J. 1990 *Proc. R. Soc. Lond.* A **430**, 653–668.

Berry, M. V. & Howls, C. J. 1991 *Proc. R. Soc. Lond.* A**434**. (In the press.)

Dingle, R. B. 1973 *Asymptotic expansions: their derivation and interpretation*. New York and London: Academic Press.

Gradshteyn, I. S. & Ryzhik, I. M. 1980 *Table of integrals, series and products* New York and London: Academic.

Olver, F. W. J. 1990 In *Proc. Int. Symp. Asymptotic and Computational Analysis* (ed. R. Wong), pp. 329–355. New York: Marcel Dekker.

Paris, R. B. & Wood, A. D. 1991 *J. Comp. Appl. Math.* (In the press.)

Stokes, G. G. 1864 *Trans. Camb. phil. Soc.* **10**, 106–128. (Reprinted in *Mathematical and physical papers by the late Sir George Gabriel Stokes* (Cambridge University Press 1904), vol. IV, pp. 77–109.)

Received 12 April 1991; accepted 23 May 1991