SIAM REVIEW
Vol. 47, No. 3, pp. 553–575

Hyperasymptotics and the Linear Boundary Layer Problem: Why Asymptotic Series Diverge*

John P. Boyd†

Abstract. The simplest problem with boundary layers, $\epsilon^2 u_{xx} - u = -f(x)$, is used to illustrate (i) why the perturbation series in powers of ϵ is asymptotic but divergent, (ii) why the optimally truncated expansion is "superasymptotic" in the sense that that error is proportional to $\exp(-[\text{constant}]/\epsilon)$, and (iii) how to obtain an improved "hyperasymptotic" approximation.

Key words. perturbation methods, asymptotic, hyperasymptotic, exponential smallness, beyond-all-orders perturbation theory

AMS subject classifications. 34E05, 40G99, 41A60, 65B10

DOI. 10.1137/S003614450444436X

> One remarkable fact of applied mathematics is the ubiquitous appearance of divergent series, hypocritically renamed asymptotic expansions. Isn't it a scandal that we teach convergent series to our sophomores and do not tell them that few, if any, of the series they meet will converge? The challenge of explaining what an asymptotic expansion is ranks among the outstanding taboo problems of mathematics.
>
> —Gian-Carlo Rota, in *Indiscrete Thoughts* (1997), p. 222

1. Introduction.

Divergent asymptotic series are one of the foundations of applied mathematics. Feynman diagrams (particle physics), Rayleigh–Schrödinger perturbation series (quantum chemistry), boundary layer theory and the derivation of soliton equations (fluid mechanics), and even numerical algorithms like the "nonlinear Galerkin" method [4, 22] are examples. Unfortunately, classic texts like van Dyke [33], Nayfeh [26], and Bender and Orszag [1], which are very good on the *mechanics* of divergent series, largely ignore two important questions. First, why do some series diverge for all nonzero ϵ where ϵ is the perturbation parameter? And how can one break the "error barrier" when the error of an optimally truncated series is too large to be useful? In the last couple of decades, the century-old theory of asymptotics has morphed into a triad of asymptotic/superasymptotic/hyperasymptotic theory.

This article is a case study that illuminates these questions and the new theoretical ideas. The problem is an ordinary differential equation whose solution has boundary

*Received by the editors June 12, 2004; accepted for publication (in revised form) January 8, 2005; published electronically July 29, 2005. This work was supported by the National Science Foundation through grant OCE9986368.
http://www.siam.org/journals/sirev/47-3/44436.html

†Department of Atmospheric, Oceanic and Space Science, University of Michigan, 2455 Hayward Avenue, Ann Arbor, MI 48109 (jpboyd@engin.umich.edu, http://www.engin.umich.edu:/∼ jpboyd/).

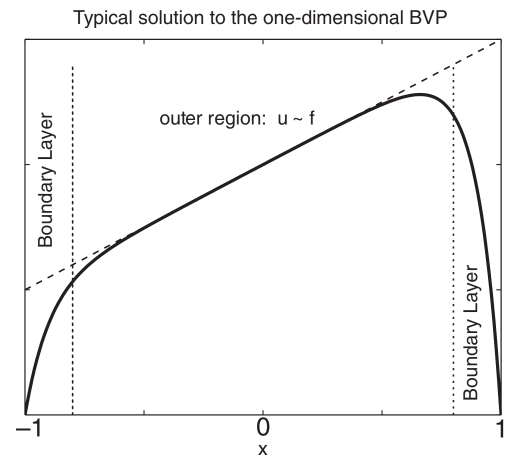

Fig. 1.1 *The solution to $\epsilon^2 u_{xx} - u = -f(x)$ has three regions when $\epsilon << 1$. Over most of the interval (dubbed the "outer region" in matched asymptotics perturbation theory), $u \approx f$. The outer region is flanked by two narrow boundary layers, each of width $O(\epsilon)$.*

layers when the perturbation parameter ϵ is small:

$$\epsilon^2 \, u_{xx} - u = -f(x), \tag{1.1}$$

$$u(\pm 1) = 0, \tag{1.2}$$

where subscript x denotes differentiation with respect to that coordinate. This problem can be approximately solved by the method of matched asymptotic expansions [11]. Unfortunately, matching does not provide an obvious, intuitive explanation for the divergence of the series in powers of ϵ.

Although very simple, this boundary value problem contains narrow boundary layers (of width $O(\epsilon)$) as illustrated schematically in Figure 1.1. As such, it is an exemplar for high-Reynolds number fluid flows, which usually contain similar boundary layers, and more generally for other linear and nonlinear problems that contain narrow layers of rapid variation and multiple length scales.

The general solution to a linear, second-order ordinary differential equation can always be written as the sum of two homogeneous solutions plus a particular solution:

$$u(x) = A \exp(-[x+1]/\epsilon) + B \exp([x-1]/\epsilon) + u_{part}(x;\epsilon), \tag{1.3}$$

where A and B are constants determined by applying the boundary conditions and $u_{part}(x;\epsilon)$ is the (as yet unknown) particular solution.

There are multiple warning solutions that all-is-not-well in the limit $\epsilon \to 0$:

1. The homogeneous solutions are not analytic in ϵ at $\epsilon = 0$.
2. Because $\lim_{\epsilon\to 0} \epsilon^{-k} \exp(-x/\epsilon) = 0$ for all finite k of either sign, all the derivatives of the homogeneous solutions vanish at the origin, implying that these functions have only the trivial (and useless) power series

$$\exp(-[x+1]/\epsilon) \sim 0 + 0 \cdot \epsilon + 0 \cdot \epsilon^2 + \cdots \tag{1.4}$$

for all $x > -1$, and similarly for the other homogeneous solution.

3. The second derivative disappears at $\epsilon = 0$, thereby lowering the order of the differential equation.

Fortunately, the inhomogeneous terms are known *analytically* for this problem—a good thing, too, since these functions have no useful ϵ power series.

Instead, we shall concentrate on the subtler difficulties that arise in finding a particular integral. As emphasized in undergraduate classes, the particular integral is not unique; any solution of the form of the *general* solution is a "particular integral," that is, a solution of the inhomogeneous differential equation. However, it is possible to find a unique particular integral, dubbed u_{slow}, which varies only on the slow $O(1)$ length scale of $f(x)$. Because it varies slowly, it not obvious why there should be any difficulties in expanding it as a power series in ϵ. Nevertheless, the perturbation series for u_{slow} is usually asymptotic but divergent. Why does it diverge? It is this puzzle that we try to resolve in the rest of the article.

First, though, a brief defense of asymptotic series.

2. The Relevance of Asymptotic Series. Applied mathematics has greatly expanded in the last few decades through a mixture of supercomputing and theoretical advances. The dark side of this progress is that the new and trendy has crowded out much of the old and unfashionable. In the words of one reviewer, "the entire subject of asymptotic expansions is taught less frequently these days, having been supplanted by a set of computer codes. (I am not asserting that this trend is a good idea, but simply that it exists.)" Another reviewer noted that students ask, "Is this topic only classical mathematics, or does it still find widespread use? Has computation replaced the need for these methods?"

Unfortunately, asymptotics is usually taught very badly when taught at all. When a student asks, "What does one do when x is larger than the radius of convergence of the power series?", the response is a scowl and a muttered, "asymptotic series!", followed by a hasty scribbling of the inverse power series for a Bessel function. "But of course, that's all built-in to MATLAB, so one never has to use it anymore."

Humbug! First, asymptotic series are accurate precisely in those extreme parameter ranges where brute-force computation fails. In fluid mechanics, for example, number-crunching or arithmurgy[1] is easy only when the Reynolds number is low. Boundary layer theory, which is an asymptotic expansion in inverse powers of the Reynolds number, is most successful when the Reynolds number is large—a regime where numerical methods succeed only by exploiting the insights of the boundary layer through high-grid densities near the boundaries. Our chosen boundary value problem (1.1) illustrates this: the required number of grid points on a uniform grid is proportional to $1/\epsilon$, regardless of the algorithm, whereas the multiple scales series becomes more accurate as $\epsilon \to 0$. Arithmurgy hasn't replaced asymptotics; rather, number-crunching and asymptotic series are *complementary* and *mutually enriching*.

Second, most physics and engineering problems are a blend of multiple length and time scales. Indeed, SIAM has launched a new journal entitled *Multiscale Modeling and Simulation*. So-called reductive perturbation theory can often encapsulate the relevant scales in a simplified differential equation—but such reductions are merely asymptotic series.

The Korteweg–de Vries equation of water wave theory, for example, is a one-space-dimensional model derived by multiple scales perturbation theory from the three-

[1] "Arithmurgy" is a descriptive term for scientific computation; it means literally "number-working" from the Greek $\alpha\rho\iota\theta\mu o\sigma$, "number," and $\epsilon\rho\gamma o\sigma$, "working."

space-dimensional hydrodynamic equations when the wave amplitude is small and the waves are long relative to the water depth. The nonlinear Schrödinger (NLS) equation, and coupled systems of NLS equations, are powerful tools for understanding pulses of light down a fiber optics cable: space again reduced in dimension from 3 to 1 through the black art of perturbation theory. Indeed, all of classical physics is a multiple scales lowest order approximation to the quantum realm. And quantum mechanics in turn is a multiple scales approximation to a more complete theory—string theory? M-brane theory?—as yet dimly grasped.

Indeed, the need to resolve vast ranges in amplitude and scale is built into our senses of sight, sound, and touch. Two 19th century physiologists observed what is now known as the Weber–Fechner law: the magnitude of a subjective sensation increases proportional to the logarithm of the stimulus intensity. For example, our eyes function in both dim starlight and the glare of the noonday sun, a million times brighter, but we don't perceive a difference of a factor of a million because our brain scales intensity *logarithmically.* Similarly, our eyes (at least in youth) are able to vary their focus to perceive a grain of sand at one extreme and a mountain range at the other.

Scientific computation obeys its own kind of Weber–Fechner law: for most classes of phenomena, the billionfold increase in speed from ENIAC (a speed of several kiloflops, used for the first numerical weather forecast in 1950) to the multiteraflop supercomputers of today has not, alas, produced a billionfold increase in our knowledge of the atmosphere, or of any other complex phenomena with multiple scales. Instead, the following is self-evident.

PROPOSITION 2.1 (logarithmic law of arithmurgy). *Computational insight into nature increases logarithmically with flop-rates and memory.*

The logarithmic law of arithmurgy implies that brute-force computation will *never* solve all of science's interesting problems. Direct numerical simulation (DNS) of three-dimensional turbulence at very high Reynolds numbers requires supercomputers with a billion processors, atom-sized memory cells, and interconnects faster than the speed of light! But the renormalization group theory [36]—a multiple scales theory with ideas borrowed from quantum field theory—is not so constrained. There are many competitive strategies for turbulence, some flaunting their connections with asymptotics, some disguising these connections, but all at least implicitly using asymptotic thinking to bridge the gap between computation and the power of multiscaled nature.

To understand, for a complex phenomenon like turbulence, what is hidden beneath divergent series is obviously very difficult, full of thorns for our grandchildren. However, the hyperasymptotics revolution has given new insights into the strengths and limits of multiple scales phenomena, and in the rest of this article, we shall look at an elementary but illuminating example.

3. Asymptotic, Superasymptotic, and Hyperasymptotic. The perturbation series derived in the next section is typical of so-called singular perturbation methods such as the method of multiple scales, the method of matched asymptotic expansions, and the method of steepest descent: it is asymptotic but *diverges* for all values of the perturbation parameter ϵ.

A divergent series can still be useful if it satisfies Poincaré's definition of "asymptotic."

DEFINITION 3.1 (asymptoticity; see [1]). *A power series is* asymptotic *to a function $f(\epsilon)$ if, for* fixed N *and* sufficiently small ϵ*, the error $E(\epsilon; N)$ falls as fast as*

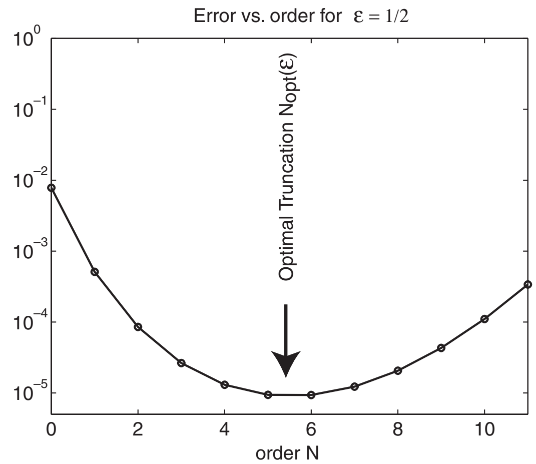

Fig. 3.1 *A plot of errors versus perturbation order for a typical function. The chosen function is* $\rho(\epsilon) \equiv (12/\epsilon)\left\{\log(\Gamma(1/\epsilon)) + (1/\epsilon - 1/2)\log(\epsilon) + 1/\epsilon - (1/2)\sqrt{2\pi}\right\}$. *This has the asymptotic series* $\rho(\epsilon) \sim 1 - (1/30)\epsilon^2 + (1/105)\epsilon^4 - (1/140)\epsilon^6 + \cdots$. *The plot shows the errors for various truncations of the asymptotic series N for* $\epsilon = 1/2$. *The optimal truncation* N_{opt} *is about halfway between fifth and sixth order for this particular function and this particular value of* ϵ.

ϵ^{N+1}, *that is,*

$$E(\epsilon; N) \equiv |f(\epsilon) - \sum_{j=0}^{N} a_j\,\epsilon^j| \sim O\left(\epsilon^{N+1}\right) \quad as\ \epsilon \to 0. \tag{3.1}$$

$O(\ )$ is the usual "Landau gauge" symbol that denotes that the quantity to the left of the asymptotic equality is bounded in absolute value by a constant times the function inside the parentheses on the right.

Asymptotic series are useful because the definition guarantees that the Nth order approximation is always accurate if ϵ is sufficiently small. If the series is *divergent*, then it is *generic* that, for sufficiently small ϵ, the error for *fixed* ϵ falls to a minimum as the truncation N increases before inexorably rising and rising without limit as $N \to \infty$ [1]. A typical example is Figure 3.1. The good news for this example is that the minimum error is smaller than 10^{-5}—quite acceptable for almost all engineering and scientific applications. The fall-and-then-rise of error motivates the following definitions.

Definition 3.2 (optimal truncation and optimal error). *For the asymptotic series of a given function, the* "optimal truncation" $N_{opt}(\epsilon)$ *is that value of* N *where the error* $E(\epsilon; N)$ *has a* minimum *for* fixed ϵ. *The error at the minimum is the* "optimally truncated error" *or* "optimal error" *for short,* $E_{opt}(\epsilon) \equiv \min_N E(\epsilon; N)$.

The neologism "superasymptotic" describes how E_{opt} varies with ϵ. The left panel of Figure 3.2 plots the error-versus-order curves for many ϵ and then connects the minima of each curve by the heavy solid line, which is therefore a plot of $E_{opt}(\epsilon)$. The right panel is a plot of N_{opt} against $1/\epsilon$. It is in fact typical, and not merely an accident of this particular example, that the optimal error falls as the *exponential* of the optimal truncation order or, equivalently, of the reciprocal of the small parameter

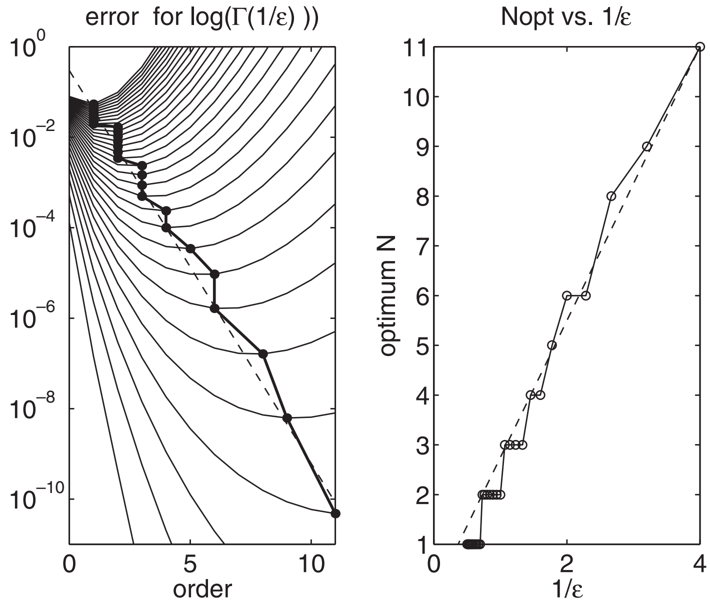

Fig. 3.2 *Left: errors versus perturbation order N for $\rho(z = 1/\epsilon)$ for $\epsilon = j/16, j = 1, 2, \ldots, 32$. The dots, connected by a heavy solid curve, mark the optimal truncation (minimal error) for each value of ϵ. The dashed curve is a crude fit: $E_{opt} \approx 0.3\, \exp(-2\, N_{opt}(\epsilon))$. Right panel: N_{opt} plotted versus $1/\epsilon$. The dashed line is the crude linear fit, $N_{opt} \approx 2.75/\epsilon$.*

ϵ; the (fitted-by-eye) dashed curves for the illustrated example are

$$E_{opt} \approx 0.3 \exp(-2N_{opt}) \approx 0.3 \exp(-5.5/\epsilon). \tag{3.2}$$

Poincaré's definition is all about *powers* of ϵ, but the minimum error is in fact an *exponential*, something obviously quite different. Sir Michael Berry and his collaborator Christopher Howls therefore defined the following.

DEFINITION 3.3 (superasymptotic). *An* optimally truncated *asymptotic series is a "*superasymptotic*" approximation. The error is typically an* exponential *function of the* reciprocal *of the perturbation parameter* ϵ [3, 2].

In recent years, as explained in the books [30, 5, 12, 13, 17, 19, 23, 32, 15, 35, 21, 27] and the reviews [6, 28], a rather large variety of tools for improving upon the superasymptotic approximation have been developed.

DEFINITION 3.4. *A* hyperasymptotic *approximation in the broad sense is one that achieves higher accuracy than a superasymptotic approximation. A* hyperasymptotic *approximation in the narrow sense achieves this improvement by adding one or more terms of a* second *asymptotic series, with different scaling assumptions, to the optimal truncation of the original asymptotic expansion* [2]. *(With another rescaling, this process can be iterated by adding terms of a third asymptotic series, and so on.) Hyperasymptotic methods are also called "*exponential asymptotics*" and "*asymptotics-beyond-all-orders.*"*

Sequence acceleration and sum acceleration methods are powerful tools for improving upon a superasymptotic approximation, but some authors do not like to label them as "hyperasymptotic" because these methods (i) are very useful for convergent series, too, and (ii) have a history that is much older than the modern surge of interest in hyperasymptotic methods that directly attack the cause of the divergence.

The alternative label "exponential asymptotics" is used as a synonym for "hyperasymptotic" because to improve upon the superasymptotic approximation, one must calculate terms which are proportional to an *exponential* of $1/\epsilon$. The "beyond-all-orders" term arises because $\exp(-[\text{constant}]/\epsilon)$ decreases to zero faster than any finite *power* of ϵ, and therefore effects proportional to such an exponential are completely missed by an ϵ-power series and lie "beyond all orders" in powers of ϵ.

4. Perturbation Series for the Particular Integral. If $\epsilon \ll 1$, it is plausible that the second derivative in the differential equation, which is multiplied by ϵ^2, will be small compared to the undifferentiated term. This gives the lowest order approximation

$$u^{(0)} = f(x). \tag{4.1}$$

This implicitly assumes that $u_{part}(x;\epsilon)$ varies on roughly the same independent-of-ϵ length scale as the forcing $f(x)$; the lowest order approximation shows that this is a self-consistent approximation.

The homogeneous solutions, $\exp(\pm x/\epsilon)$, vary on a "fast" $O(1/\epsilon)$ length scale. The particular integral is not unique because of the freedom to add arbitrary multiples of the homogeneous solutions to any particular integral to generate a new solution of the inhomogeneous differential equation. However, the assumption that the particular integral is *slowly varying* picks out a *unique* particular integral. If we call this $u_{slow}(x;\epsilon)$, then the *general* particular integral is

$$u_{part}(x;\epsilon) = u_{slow}(x;\epsilon) + A\exp(x/\epsilon) + B\exp(-x/\epsilon), \tag{4.2}$$

where A and B are arbitrary constants. However, only when $A = B = 0$ do we obtain a particular integral which varies *only* on the *slow* length scale and therefore is consistent with the neglect of the second derivative in the differential equation.

Because the perturbation scheme exploits the "slow" $O(1)$ length scale in a problem that has a second, much faster length scale, this technique is called the "method of multiple scales." Higher approximations can be found by substituting the expansion

$$u_{slow}(x) \sim \sum_{j=0}^{\infty} \epsilon^{2j}\, u^{(2j)}(x) \tag{4.3}$$

into the differential equation and matching powers, yielding

$$u_{slow} = \sum_{j=0}^{\infty} \epsilon^{2j}\, \frac{d^{2j} f}{dx^{2j}}. \tag{4.4}$$

This derivation implicitly ignores the boundary layers of $u(x)$. This does not cause problems because the particular integral $u_{slow}(x)$ does not necessarily have boundary layers. It is only the homogeneous solutions which vary rapidly within an $O(\epsilon)$ distance of the endpoints and impart similar behavior to $u(x)$.

5. Why the Perturbation Series Diverges, I. It is possible to give an intuitive if nonrigorous explanation for divergence directly from the perturbation series. (A more precise argument using a Fourier integral representation will be given later.)

Define

$$E(\epsilon;N) \equiv u_{slow}(x;\epsilon) - \sum_{j=0}^{N} \epsilon^{2j}\, u^{(2j)}. \tag{5.1}$$

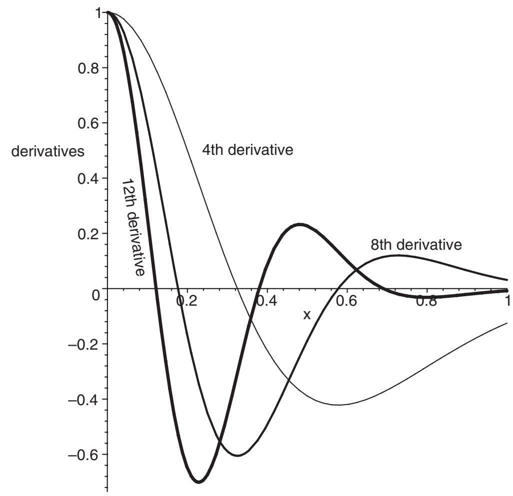

Fig. 5.1 *The fourth, eighth, and twelfth derivatives of $f = 1/(1 + x^2)$, scaled by dividing by the maximum of each function. The theme of the graph is that the derivative of a function almost always varies more rapidly as the order of the derivative increases. Because the function and all its even derivatives are symmetric with respect to the origin, only positive x is illustrated.*

Without approximation, this solves

$$\epsilon^2 \frac{d^2 E}{dx^2} - E = \epsilon^{2N+2} \frac{d^{2N+2} f}{dx^{2N+2}}(x). \tag{5.2}$$

The usual multiple scales procedure, as done earlier, is to neglect the $\epsilon^2 E_{xx}$ term on the grounds that E varies only on the slow $O(1)$ length scale of $f(x)$.

Heuristically, multiple scales theory breaks down because, for fixed ϵ and sufficiently high order N, the inhomogeneous term is *not* varying slowly and there are no longer two disparate length scales. It is inconsistent to neglect the second derivative compared to E itself when the inhomogeneous term in (5.2) and therefore E are varying on an $O(1/\epsilon)$ length scale.

The reason that the right-hand side of (5.2) is varying rapidly is that *differentiation* is an *antismoothing* operation: for almost all $f(x)$, the $2N$th derivative oscillates faster and faster as $N \to \infty$.

We shall not offer a rigorous proof of this assertion, but instead show Figure 5.1. The function $f = 1/(1+x^2)$ falls monotonically from its peak at the origin. Nevertheless, its derivatives oscillate faster and faster as the order of the derivative increases: the fourth derivative has one root for positive x, the eighth derivative two, the twelfth derivative three, and so on. The trend is obvious: as the derivative order increases, it is possible, for any ϵ, to find a sufficiently high-order derivative that falls from its maximum at the origin to its first root in a distance as small as ϵ, implying an $O(1/\epsilon)$ length scale.

As we shall see more precisely using the Fourier integral representation later, the statement that $f(x)$ is a slowly varying function is never entirely true except for special cases, such as when $f(x)$ is a polynomial. Instead, all functions with singularities at finite complex x have within them some arbitrarily fast variation which is explicit in the Fourier transform of the function, or a windowed version of the function, and which is developed, like a photograph emerging from the developer fluid in a darkroom, by the repeated differentiations that generate the right-hand side of the perturbative equation (5.2). The few exceptions to this principle will be discussed in later sections.

6. Fourier Integral Solution.

> The shortest path between two truths in the real domain passes through the complex domain.
>
> —Jacques Hadamard (1865–1963)

Deeper insight comes by doing something that at first blush is a little goofy: expanding the problem from its physical domain, $x \in [-1, 1]$, to the entire real line. To an engineer especially, this seems bizarre: if the boundary layer problem is from fluid mechanics, then the fluid is only on $x \in [-1, 1]$, and we are extending a fluid problem to regions where there is no fluid!

However, an analytic function $f(x)$ is not merely defined on the interval where it solves a physical problem, but throughout the entire complex plane. Often, as noted by Hadamard, one must leave the physical domain to understand the mathematics.

For example, power series in x are rarely applied to solve boundary problems. The reason is that off-interval singularities often restrict the radius of convergence of the power series to only a part of the physical domain. The functions $f = 1/(x^2+1/4)$ and $f = 1/(x + 3/2)$ are both analytic everywhere on the interval $x \in [-1, 1]$, but neither has a power series which converges over the whole interval.

In a similar spirit, we shall examine our boundary layer problem on the whole real line so as to exploit the power of Fourier integral methods. As explained in the appendix and more fully in [7, 8, 9], it always possible to extend an arbitrary $f(x)$ from an interval to the entire real line in such a way that the extended function $\tilde{f}(x)$ has a well-behaved Fourier transform and also equals $f(x)$ on the physical interval, $x \in [-1, 1]$.

Leaving the details to those references, we shall henceforth assume that such an extension has been performed, if necessary, and therefore that the function $f(x)$ in our boundary value problem has the Fourier transform $F(k)$ defined by

$$F(k) \equiv \frac{1}{2\,\pi} \int_{-\infty}^{\infty} f(x)\ \exp(-i\,k\,x)\,dx. \tag{6.1}$$

It can then be verified by direct substitution that the all-important particular integral is given *exactly* by the inverse Fourier transform

$$u_{slow}(x) = \int_{-\infty}^{\infty} \frac{F(k)}{1+\epsilon^2\,k^2}\ \exp(i\,k\,x)\,dk. \tag{6.2}$$

The perturbative approximation (4.3), derived earlier by simple substitution and matching of powers, can also be derived from the Fourier integral. The trick is to expand the denominator in the integrand of the integral representation of the particular integral, (6.2), and then exploit the identity

$$\frac{d^{2j} f}{dx^{2j}} = \int_{-\infty}^{\infty} (-k^2)^j\,F(k)\ \exp(i\,k\,x)\,dk. \tag{6.3}$$

6.1. Integral Representation of the Error. By substituting the familiar identity for the partial sums of a geometric series

$$\frac{1}{1+\epsilon^2 k^2} = \sum_{j=0}^{N} (-1)^j k^{2j} \epsilon^{2j} + (-1)^{N+1} \frac{k^{2N+2}\epsilon^{2N+2}}{1+\epsilon^2 k^2} \tag{6.4}$$

into the Fourier transform integral, one obtains the exact integral representation of the error

$$E(\epsilon; N) = (-1)^{N+1} \epsilon^{2N+2} \int_{-\infty}^{\infty} \frac{k^{2N+2} F(k)}{1+\epsilon^2 k^2} \exp(i\, k\, x)\, dk. \tag{6.5}$$

We now have the tools to explain the divergence of the perturbation series and also to prove superasymptoticity, as done in turn in the next two sections.

7. Why the Perturbation Series Diverges, II. The perturbation series in powers of ϵ can be derived from the Fourier integral representation of the particular integral,

$$u_{slow}(x) = \int_{-\infty}^{\infty} \frac{F(k)}{1+\epsilon^2 k^2} \exp(i\, k\, x)\, dk, \tag{7.1}$$

merely by expanding the denominator as a geometric series:

$$\frac{1}{1+\epsilon^2 k^2} = \sum_{n=0}^{\infty} (-1)^n \epsilon^{2n} k^{2n}. \tag{7.2}$$

Because $1/(1+\epsilon^2 k^2)$ has simple poles at $k = \pm i/\epsilon$, the geometric series converges only for $|k| < 1/\epsilon$. The reason that the resulting series for $u_{slow}(x;\epsilon)$ (usually) *diverges* is that the inside-the-integral geometric series has only a *finite* radius of convergence but the integral has an *infinite* range of integration. It follows that we are committing the mathematical crime of applying a series *outside* of its disk of convergence. No wonder the series for u_{slow} diverges!

The real surprise is that this mathematical felony has any usefulness at all! The reason that the asymptotic series is useful is indicated schematically in Figure 7.1.

Texts on Fourier integrals explain that if $f(x)$ is a well-behaved function that (i) decays as fast as $O(1/|x|^2)$ as $|x| \to \infty$ along the real line and (ii) is analytic in the strip $\Im(x) \in [-\mu, \mu]$ for some positive constant μ, then its transform $F(k)$ will decay as $\exp(-\mu|k|)$ or faster as $|k| \to \infty$. Thus, the schematic integrand in Figure 7.1 decays exponentially for large $|k|$ and must be plotted on a logarithmic scale. The crucial point is that when $\epsilon << 1$, the integrand of the Fourier representation of $u_{slow}(x;\epsilon)$ is *exponentially small* in $1/\epsilon$ everywhere outside the region of convergence of the geometric series. Our mathematical crime must be reclassified as a misdemeanor: the denominator series diverges only where the integrand is very, very small.

In the limit $N \to \infty$ for fixed ϵ, even this smallness cannot save it: the error in the geometric series becomes larger and larger without bound as $N \to \infty$ and so the series for $u_{slow}(x;\epsilon)$ must diverge, too. However, if we truncate the series at the Nth order where N is finite, the error due to the geometric series will be finite, too, and is multiplied by the exponential smallness of the integrand in the regions in the integration variable k where the truncated geometric series is inaccurate.

The optimal truncation of the series, $N_{opt}(\epsilon)$, is determined by the tug-of-war between two opposing tendencies. As N increases, the error in approximating the integrand for $|k| < 1/\epsilon$ *decreases*, while the error for larger $|k|$ (where the geometric series diverges) *increases*. The optimal truncation (and minimum error) represents the best compromise, for a given ϵ, between these two simultaneous effects of increasing N.

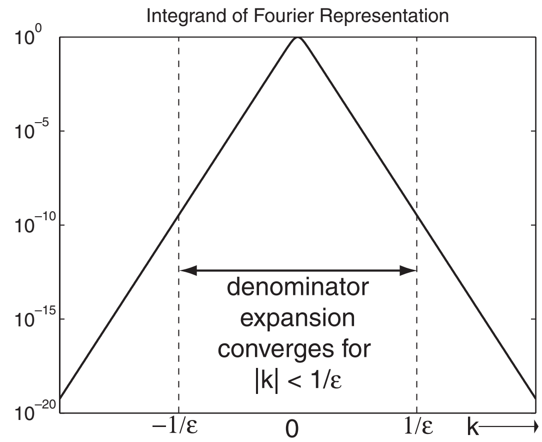

Fig. 7.1 *Schematic of a typical integrand for the Fourier integral representation of the error in the multiple scales perturbation theory. The geometric series for* $1/(1+\epsilon^2 k^2)$ *converges only within the range bounded by the two vertical dotted lines. Fortunately, the Fourier transform* $F(k)$ *of the inhomogeneous term in the differential equation,* $f(x)$*, usually falls* exponentially *fast, so the integrand is very tiny everywhere that the geometric series is divergent. This makes it possible to extract a "superasymptotic" approximation whose error is an exponential function of* $1/\epsilon$.

8. Proof of Superasymptoticity. These notions of "exponential smallness" and growing-and-decreasing errors can be quantified by using the error integral defined by (6.5). For fixed ϵ, the optimal truncation N_{opt} is that value of N which *minimizes* this error.

Unfortunately, the error integral usually cannot be evaluated in closed form: if we could, we would have no need of perturbation theory! However, it is fairly straightforward to *bound* the error, giving the following.

THEOREM 8.1 (superasymptoticity). *If the Fourier transform of the forcing* $f(x)$ *satisfies the bound*

$$|F(k)| \leq p \exp\left(-q^r k^r\right) \qquad \forall k, \tag{8.1}$$

where p, q, r *are positive constants, then the error in the asymptotic approximation to the particular integral* u_{slow} *of* $\epsilon^2 u_{xx} - u = f(x)$,

$$E(N;\epsilon) \equiv \left| u_p - \sum_{j=0}^{N} \epsilon^{2j} u^{(2j)} \right|, \tag{8.2}$$

satisfies, for sufficiently small ϵ, *the bound*

$$|E(N,\epsilon)| \leq 2\sqrt{2\pi}\frac{p}{r}\frac{1}{q^{r/2}}\,\epsilon^{r/2-1}\,\exp\left(-\frac{q^r}{\epsilon^r}\right) \tag{8.3}$$

when

$$N(\epsilon) \approx N_{opt}(\epsilon) \sim \frac{r q^r}{2\,\epsilon^r} + \left(\frac{r}{4} - \frac{3}{2}\right) + O(\epsilon^r). \tag{8.4}$$

Proof. To obtain a bound on the error for general N, not necessarily optimal, note that the error integral can be bounded by the new integral Λ obtained by (i) replacing the denominator of the error integral by 1 and (ii) replacing the Fourier transform in the error integral by its bound (8.1):

$$|E(N,\epsilon)| \leq 2\,p\,\Lambda(N,\epsilon), \tag{8.5}$$

where, using identity 3.478.1 on p. 342 of the fourth edition of [14],

$$\Lambda(N,\epsilon) \equiv \int_0^\infty \epsilon^{2N+2} k^{2N+2} \exp(-q^r k^r)\,dk = \frac{\epsilon^{2N+2}}{r\,q^{(2N+3)}}\,\Gamma\left(\frac{2N+3}{r}\right). \tag{8.6}$$

The optimal truncation can be estimated by differentiating the bound Λ with respect to N for fixed ϵ and finding the zero of $d\Lambda/dN$, which defines the value of $N = N_{opt}(\epsilon)$ that minimizes Λ. The derivative vanishes when

$$\Psi\left(\frac{2N+3}{r}\right) = \log(q^r/\epsilon^r), \tag{8.7}$$

where $\Psi(z) \equiv d\log(\Gamma)/dz$ is the so-called digamma function. The equation $\log(\alpha) = \Psi(z)$ can be approximately solved by $z \sim \alpha + (1/2) - (1/(24\,\alpha)) + O(\alpha^{-2})$, giving (8.4) for $N_{opt}(\epsilon)$.

Substituting this into Λ, applying the large-argument asymptotics for the gamma function, and doubling the result to bound the errors in the gamma series bounds the error by the right-hand side of (8.3) when N is equal to the estimate of $N_{opt}(\epsilon)$. Since the true error can only be smaller, this proves the theorem. □

The generic case is $r = 1$; that is, if $f(x)$ is analytic within the strip $|\Im(x)| \leq \mu$ for some constant μ, then the Fourier transform $F(k)$ will asymptotically decay proportional to $\exp(-\mu|k|)$. With optimal truncation, the error in the perturbation series is then bounded by a constant times $\exp(-\mu/\epsilon)$.

9. Examples.

9.1. A Good Word for Cases and Examples.

The trouble with general theorems is that they are, well, general. Useful as the proof of superasymptoticity is, it does not fully explore the full diversity of the solutions to our one-dimensional boundary value problem like the set of examples cataloged in Table 9.1.

9.2. *f(x)* Is a Polynomial: Termination.

If $f(x)$ is a polynomial of degree M, then there is always a particular solution $u_{slow}(x;\epsilon)$ which is a polynomial of the same degree. Let

$$f = \sum_{j=0}^{M} f_j x^j, \qquad u_{slow} = \sum_{j=0}^{M} a_j x^j. \tag{9.1}$$

Matching powers of x gives the backward recurrence

$$a_M = f_M, a_{M-1} = f_{M-1},\;\; a_j = f_j + \epsilon^2 (j+2)(j+1)\,a_{j+2},\;\; j = M-2, M-3, \ldots, 0. \tag{9.2}$$

An alternative derivation is to observe that for a polynomial of Mth degree, all derivatives of order higher than M are *zero*. This implies that the multiple scales perturbation theory must *terminate* with the ϵ^M term (if M is even) or the ϵ^{M-1} term (if M is odd).

This is a dramatic counterexample to the folk wisdom that multiple scales perturbation series are always divergent!

Table 9.1 *Exemplary solutions. $S(x)$ is the Stieltjes function and $F(k)$ is the Fourier transform of $f(x)$.*

$f(x)$	$u_{slow}(x;\epsilon)$
$f_0+f_1x+f_2x^2$	$\left(f_0+\epsilon^2 f_2\right)+f_1x+f_2x^2$
$\cos(\kappa x)$	$\frac{\cos(kx)}{1+\epsilon^2\kappa^2}$
$\exp(ax)$	$u=\frac{1}{1-\epsilon^2 a^2}\exp(ax)$
$\mathrm{sinc}(x)=\frac{\sin(\pi x)}{\pi x}$	$\frac{1}{\pi}\int_0^\pi \frac{\cos(kx)}{1+\epsilon^2k^2}\,dk$
$\exp(-[1/2]x^2)$	$\frac{\sqrt{\pi}}{2\,\epsilon\,\sqrt{2}}\exp([1/2]/\epsilon^2)$ $\left\{\exp(-x/\epsilon)\left[1+\mathrm{erf}\left(\{x-1/\epsilon\}/2^{1/2}\right)\right]\right.$ $\left.+\exp(x/\epsilon)\left[1-\mathrm{erf}\left(\{x+1/\epsilon\}/2^{1/2}\right)\right]\right\}$
$\frac{4}{1+x^2}$	$\frac{1}{1+ix}\left\{S\left(-\frac{i\epsilon}{1+ix}\right)+S\left(\frac{i\epsilon}{1+ix}\right)\right\}$ $+\frac{1}{1-ix}\left\{S\left(-\frac{i\epsilon}{1-ix}\right)+S\left(\frac{i\epsilon}{1-ix}\right)\right\}$
$\exp(-\lvert x\rvert)$	$\frac{\exp(-\lvert x\rvert)-\epsilon\exp(-\lvert x\rvert/\epsilon)}{1-\epsilon^2}$
General f	$\int_{-\infty}^{\infty}\frac{F(k)}{1+\epsilon^2k^2}\exp(ikx)dk$

9.3. *f* Is a Cosine or Trigonometric Polynomial: Convergence. If $f(x)$ is a trigonometric function, then the explicit solution is

$$(9.3)\qquad f(x)=\cos(\kappa x)\qquad\rightarrow\qquad u_{slow}(x;\epsilon)=\frac{\cos(\kappa\,x)}{1+\epsilon^2\kappa^2}.$$

If this particular solution is expanded in ϵ, the fact that the only singularities are at $\epsilon=\pm i/\kappa$ implies that the ϵ-power series will *converge* for all $|\epsilon|<1/\kappa$.

It is trivial to extend this reasoning to a trigonometric polynomial, whose series will also have a finite radius of convergence.

However, the ϵ-series for an *infinite* Fourier series diverges. For example, with p a positive constant less than 1,

$$(9.4)\qquad f=1+2\sum_{j=1}^{\infty}p^j\,\cos(jx)\qquad\rightarrow\qquad u_{slow}=1+\sum_{j=1}^{\infty}p^j\,\frac{1}{1+\epsilon^2j^2}\,\cos(jx).$$

If we expand u_{slow} in ϵ, the expansions for all Fourier components with $j>1/\epsilon$ will diverge. Because j increases without bound, there is *no finite* ϵ for which the expansions of *all* denominators will converge. It is only when the Fourier series is truncated at $j=M$ for some finite M that the ϵ-series has the finite convergence radius $1/M$.

9.4. Bandlimited Forcing Functions: Convergence. Bandlimited functions are those whose Fourier transform $F(k)$ is identically zero for all $|k|>W$ for constant W, the bandwidth. A trigonometric polynomial is a bandlimited function; another is the so-called sinc function:

$$(9.5)\qquad f(x)=\mathrm{sinc}(x)=\frac{\sin(\pi x)}{\pi x}\qquad\rightarrow\qquad u_{slow}(x;\epsilon)=\frac{1}{\pi}\int_0^\pi\frac{\cos(kx)}{1+\epsilon^2k^2}\,dk.$$

The integral can be evaluated analytically as a messy sum of four cosine and sine integrals, but what is significant is that the integral for u_{slow} has a *finite* range of integration. The series u_{slow} for all bandlimited $f(x)$ of bandwidth W will converge for $|\epsilon| < 1/W$.

9.5. f = 4/(1 + x^2): Factorial Divergence. If $f(x)$ has a pole, branch point, or other singularity at a finite distance from the real axis, then the series for our choice of particular integral will *diverge* at a *factorial rate*, by which we mean that the coefficient of ϵ^j grows as $j!$. A simple example is

$$f(x) = \frac{4}{1+x^2} \qquad \rightarrow \qquad u_{slow}(x;\epsilon) = 4\int_0^\infty \frac{\exp(-k)}{1+\epsilon^2 k^2}\cos(kx)\,dk. \tag{9.6}$$

This can be written as a sum of four Stieltjes functions of complex argument as given in Table 9.1. At $x = 0$, the odd powers of ϵ cancel, but the even powers reinforce to give

$$u(0) \sim 4\sum_{j=0}^{\infty}(2j)!\;(-1)^j\;\epsilon^{2j}. \tag{9.7}$$

Thus, at $x = 0$, the factorial divergence can be explicitly displayed; there is a similar divergence for all x.

This is the generic situation: $f(x)$ will have a singularity somewhere in the finite complex plane, and the coefficient of ϵ^{2j} grows as $(2j)!$ for all x.

9.6. f = exp($-x^2$/2): Factorial-of-Half-Order Divergence. The rate of divergence depends not only on the fact that the power series of the denominator of the Fourier transform integral has a finite radius of convergence; it also depends on how much or how little amplitude the integrand has outside the radius of convergence.

$$f = \exp\left(-\frac{1}{2}x^2\right) \;\rightarrow\; u_{slow}(x;\epsilon) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\exp(-k^2/2)}{1+\epsilon^2\,k^2}\cos(kx)\,dk. \tag{9.8}$$

The crucial point is that the Fourier transform of $f(x)$ decays as a Gaussian function of k, i.e., very fast, rather than as $\exp(-q|k|)$ for some $q > 0$. The crime of using the geometric series of $1/(1+\epsilon^2 k^2)$ beyond its radius of convergence $1/\epsilon$ is therefore greatly reduced because $\exp(-[1/2]k^2)$ is so tiny beyond the radius of convergence.

The integral for u_{slow} is given by 3.954.2 on p. 497 of [14] and the explicit form is shown in Table 9.1. Unfortunately, this is a little messy, but the special case of $x = 0$ is simple. At $x = 0$, the identity $\int_0^\infty \exp(-[1/2]k^2)k^{2j}dk = 2^{j-1/2}\Gamma(j+1/2)$ gives

$$u_{slow}(0;\epsilon) = \frac{\sqrt{\pi/2}}{\epsilon}\exp\left(\frac{1}{2\,\epsilon^2}\right)\left\{1-\operatorname{erf}\left(\frac{1}{\epsilon\sqrt{2}}\right)\right\} \sim \frac{1}{2\sqrt{\pi}}\sum_{j=0}^{\infty}(-1)^j\epsilon^{2j}\,2^j\,\Gamma\left(j+\frac{1}{2}\right). \tag{9.9}$$

This shows that even though $f(x)$ is an *entire* function with no singularities except at $x = \infty$, the perturbation series still diverges. However, the coefficient of ϵ^{2j} grows as roughly as the factorial of j instead of as $(2j)!$.

The superasymptoticity theorem, Theorem 8.1 (with $p = (2\pi)^{-1/2}$, $q = 2^{-1/2}$, and $r = 2$), implies that

$$N_{opt}(\epsilon) = \frac{1}{2\,\epsilon^2}, \qquad E(N;\epsilon) \le 2\sqrt{2\pi}\,\exp(-\epsilon^2/2). \tag{9.10}$$

9.7. $f(x) = \exp(-|x|)$: Breakdown at Finite Order. At the other extreme, when the forcing function is singular for *real* x, the infinite series may not even exist! An illustration is

$$f(x) = \exp(-|x|) \quad \rightarrow \quad u_p(x;\epsilon) = \frac{\exp(-|x|) - \epsilon \exp(-|x|/\epsilon)}{1 - \epsilon^2}. \tag{9.11}$$

The lowest order in the multiple scales series,

$$u_{slow} \sim u^{(0)} = \exp(-|x|) + O(\epsilon), \tag{9.12}$$

is indeed a consistent approximation. However, the perturbation theory *breaks down* at higher order; note that $\exp(-|x|/\epsilon)$ is small beyond all powers of ϵ for all nonzero x.

In general, whenever the forcing function f has only a finite number of bounded derivatives on the interval $x \in [-1, 1]$, sufficiently high-order terms in the multiple scales series will not exist, as described by the following theorem. Singularities of $f(x)$ for real x, but outside the interval $[-1, 1]$, can and should be removed by the windowing procedure described in the appendix. Thus, only singularities on $x \in [-1, 1]$ can slow the decay of the Fourier transform $F(k)$ if $\tilde{f}(x)$ in the theorem is interpreted as an intelligently extended-and-windowed version of $f(x)$.

THEOREM 9.1. *If the Fourier transform $F(k)$ of the forcing function $\tilde{f}(x)$ in*

$$\epsilon^2 u_{xx} - u = -\tilde{f}(x) \tag{9.13}$$

has an algebraic rate of decay, i.e.,

$$F(k) \sim constant/|k|^m, \qquad k \rightarrow \pm\infty, \tag{9.14}$$

for some m, which is the case if $\tilde{f}(x)$ is singular for some real $x \in [-1, 1]$ such that only a finite number of bounded derivatives exists at the singular point [20], *then the multiple scales perturbation theory for the particular solution $u_{slow}(x;\epsilon)$ must break down. Term-by-term expansion of the Fourier integral solution yields convergent integrals only for the coefficients of ϵ^{2j} when $2j < m$.*

The proof follows from expanding the denominator of the Fourier integral representation, (6.2), and observing that the integrals arising term-by-term are divergent for sufficiently large degree.

9.8. Categories. Table 9.2 summarizes the behavior for different categories of functions. The assertion that multiple scales series are always factorially divergent is a *myth*. True, the series diverges for the *generic* case of an $f(x)$ which has singularities at some finite point in the complex x-plane. However, including other classes of $f(x)$, a broad spectrum of behavior is possible, ranging from termination to a finite radius of convergence to factorial-of-half-order divergence to factorial divergence to complete breakdown.

Table 9.2 *Solution classes and the properties of the asymptotic series.*

Type of $f(x)$	Perturbation series
Polynomial	Terminates
Trigonometric polynomial	Convergent
Bandlimited	Convergent
Gaussian	Factorial-of-half-order-divergent
Off-interval singularities	Factorially-divergent asymptotic
On-interval singularities	Breakdown at finite order

10. Hyperasymptotics: Breaking the Error Barrier for Divergent Series. The variety of hyperasymptotic techniques is very large, which is why [6] was rejected by *SIAM Review* without review because of excessive length. Some are very subtle, which is why [5] was rejected by a publisher on the grounds of insufficient subtlety.

In the broad sense, hyperasymptotic strategies include [6]

1. (second) asymptotic approximation of the error integral for the superasymptotic approximation;
2. resurgence schemes or resummation of late terms [12, 13, 3];
3. complex-plane matching of asymptotic expansions [29];
4. isolation strategies, or rewriting the problem so the exponentially small thing is the only thing [6];
5. special numerical algorithms, especially spectral methods [6];
6. hybrid numerical/analytical perturbative schemes [6, 18, 16];
7. sequence acceleration including Padé and Hermite–Padé approximants [1, 34, 31].

There is some overlap between these categories, but each is a whole family of methods and it would obviously take a book [5] to describe them all. Nevertheless, it is possible to give at least the flavor of this vast subject.

10.1. Necessity.

Divergent series converge faster than convergent series.

—George F. Carrier (1919–2002, National Medal of Science, 1998)

This seemingly self-contradictory statement was the product of 30 years of experience: ordinary power series usually require several terms to be useful. In contrast, often the leading term of an asymptotic expansion is rather accurate, even for not-so-small ϵ, and captures the *qualitative* character of the phenomenon even for ϵ sufficiently large that the quantitative accuracy deteriorates. Hyperasymptotics or even a superasymptotic approximation may then be useless.

Furthermore, a better tactic to obtain an answer for moderate ϵ is not to push the ϵ-power series to higher order, hyperasymptotically or otherwise, but instead to use a different perturbative expansion, such as one in powers of $1/\epsilon$ as illustrated in the next subsection.

Often the best strategy for hyperasymptotics is abstinence.

However, as reviewed in [6, 5, 30, 28], sometimes there are qualitatively distinct effects which are "beyond all orders" in the sense of being proportional to $\exp(-q/\epsilon)$. Such effects cannot be captured by an ϵ power series: going beyond the error barrier of the superasymptotic approximation is the only way to capture such effects.

For example, the equatorially trapped Kelvin wave in the ocean is unstable in the presence of mean currents. The imaginary part of the phase speed, which describes the growth rate of the waves, is an exponential function of $1/\epsilon$ where ϵ is the strength of the current. Therefore, the instability can be captured only by hyperasymptotics; it is completely missed by the asymptotic power series [10, 24, 25]. There are many other "hyperasymptotic-necessary" phenomena reviewed in [6, 5, 30, 28].

10.2. Perturbation Series for Large ϵ. A simple way to overcome the limits of a small-ϵ perturbation theory is to expand for *large* ϵ. Define the new perturbation parameter

$$\delta = 1/\epsilon^2. \tag{10.1}$$

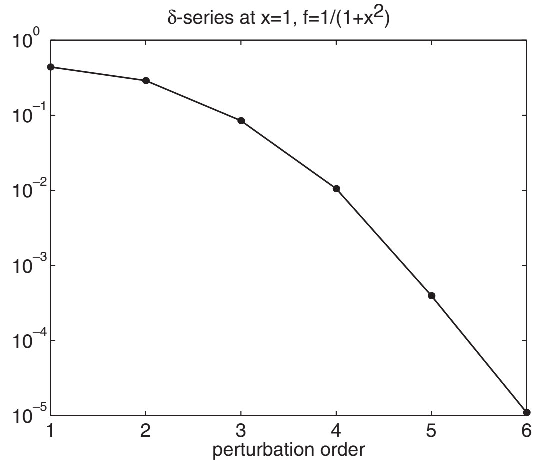

Fig. 10.1 *Corrections $d_n(x)$ for the δ series (expansion in powers of $1/\epsilon^2$ for large ϵ) as evaluated at $x = 1$ where each correction is a maximum. The inhomogeneous term $f = 1/(1 + x^2)$.*

If we expand $u_{xx} - \delta u = -\delta f(x)$ via

$$u = \sum_{j=1}^{\infty} d_n(x)\delta^n, \tag{10.2}$$

then at lowest order, $d_{1,xx} = -f(x)$, implying

$$d_1(x) = -\int^x dy \int^y f(z)dz. \tag{10.3}$$

Denote the kth iterated integral of $f(x)$ by $f^{(-k)}(x)$. The general series is

$$u = -\sum_{j=1}^{\infty} f^{(-2n)}(x)\delta^n. \tag{10.4}$$

As an example, when $f = 1/(1 + x^2)$,

$$\begin{aligned} u = {} & \delta \left\{ \frac{1}{2}\log(1+x^2) - x\arctan(x) \right\} \\ & + \delta^2 \left\{ -\frac{1}{6}x^3\arctan(x) - \frac{5}{12}x^2 - \frac{1}{3}\log(1+x^2) \right. \\ & \left. + \frac{1}{2}x\arctan(x) + \frac{1}{4}(1+x^2)\log(1+x^2) - \frac{1}{4} \right\} + \cdots . \end{aligned} \tag{10.5}$$

It is obviously much easier to analytically differentiate repeatedly than to integrate repeatedly. Still, for this example, not only can one evaluate the necessary integrals to all orders, but the series appears to converge even for ϵ as small as $1/10$. For $\epsilon = 1$ (i.e., $\delta = 1$) as illustrated in Figure 10.1, the maximum of each correction at high order is about one hundred times smaller than the maximum of its predecessor!

The value of switching to a convergent series in the reciprocal of the perturbation parameter is hardly new. Library software for evaluating Bessel functions, for example, has always combined an asymptotic series in powers of $1/x$ with a convergent power series for small x.

10.3. Hyperasymptotics from the Fourier Integral, I. The difference between the N-term truncation of the multiple scales series and $u_{slow}(x;\epsilon)$ is given *exactly* by the integral derived above as (6.5):

$$E(\epsilon;N) = (-1)^{N+1}\,\epsilon^{2N+2}\int_{-\infty}^{\infty}\frac{k^{2N+2}F(k)}{1+\epsilon^2\,k^2}\,\exp(i\;k\;x)\;dk. \tag{10.6}$$

In the superasymptotic approximation, this error integral is simply ignored and therefore approximated by zero. It follows that any nonzero approximation of $E(\epsilon;N)$ is hyperasymptotic!

The crucial observation is that the integrand of the error integral is very small for small $|k|$ because of the factor of k^{2N+2}. At larger k, the Fourier transform $F(k)$ of $f(x)$ will typically decay as $\exp(-\mu|k|)$, where μ is the distance of the closest singularities of the windowed function $\tilde{f}$ from the real x-axis. Thus, the integrand is the product of two factors with very different behavior: k^{2N+2} is growing very rapidly with $|k|$ while $F(k)$ is decaying exponentially fast with $|k|$. The integrand can be written as $\exp(\Phi)$, where

$$\Phi \equiv 2(N+1)\,\log(\epsilon) + 2(N+1)\log(k) - \mu|k| + ikx, \tag{10.7}$$

multiplied by other factors that vary only slowly with k and N.

The integral can be approximated by the "steepest descent" method [1]. For simplicity, assume that μ is real, $F(k)$ does not contain any rapidly oscillating factors, and specialize to $x = 0$; however, the steepest descent method is successful even when these are relaxed.

We approximate Φ by a local quadratic Taylor series in k in the vicinity of the "stationary point" k_s; the stationary point is defined to be a root of $d\Phi/dk$ for fixed ϵ and N:

$$k_s = \pm 2(N+1)/\mu. \tag{10.8}$$

There are two such stationary points; their contributions are added separately. Each contribution is dominated by $\exp(\Phi(k_s))$ plus other more slowly varying factors. One can therefore determine the optimum truncation $N_{opt}(\epsilon)$ merely by differentiating $\Phi(k_s(N);\epsilon,N)$ with respect to N and finding the root of the derivative: this gives

$$N_{opt}(\epsilon) \sim \frac{\mu}{2\epsilon} - 1 \tag{10.9}$$

plus corrections that depend on the slowly varying, nonexponential factors neglected at lowest order.

The steepest descent approximation with this value of N is then, to lowest order,

$$u_{slow}(0;\epsilon) \approx \sum_{j=0}^{[\mu/(2\epsilon)-1]}\epsilon^{2j}\,\frac{d^{2j}f}{dx^{2j}}(0)\;+\;2\sqrt{2\pi\,\mu}\sqrt{\epsilon}\,\exp(-\mu/\epsilon), \tag{10.10}$$

where the square brackets in the upper limit of the sum denote rounding to the integer nearest $\mu/(2\epsilon)$, the optimal truncation. The steepest descent method works equally well for nonzero x, but the stationary points become complex-valued and the path of integration must be deformed off the real axis. Nevertheless, the steepest descent method for all x can be extended into a second asymptotic series, also divergent, that greatly improves the accuracy of the superasymptotic approximation.

10.4. Hyperasymptotics from the Fourier Integral, II. There is an alternative method that does not require complex-valued arithmetic and also shows that hyperasymptotics is a suite of methods rather than a single technique. The key observation for the alternative is that given that the error integral (10.6) has a $(2N+2)$th-order zero at $k=0$, it is very silly to expand the denominator $1/(1+\epsilon^2 k^2)$ about $k=0$: the denominator series is most accurate where the integrand has no amplitude!

Equations (10.8) and (10.9) collectively show that the maximum of the error integral, when $N = N_{opt}(\epsilon)$, occurs at

$$k_{max} = \frac{1}{\epsilon} \tag{10.11}$$

independent of the small parameter μ that describes how rapidly $F(k)$ decays with $|k|$. Although the stationary point k_s is complex-valued when $x \neq 0$, the *real part* of the stationary point, and therefore the maximum of the integrand for real x, is still given by (10.8). Thus the formula for k_{max} is also independent of x.

Without approximation, we can rewrite the denominator and then expand in powers of a shifted variable that is zero when $k = k_{max} = 1/\epsilon$:

$$\frac{1}{1+\epsilon^2 k^2} = \frac{1}{2\left(1+(1/2)\left\{\epsilon^2 k^2 - 1\right\}\right)} = \frac{1}{2}\sum_{j=0}^{\infty}\left(-\frac{1}{2}\right)^j \left\{\epsilon^2 k^2 - 1\right\}^j . \tag{10.12}$$

The transformed expansion is very simple because when the new geometric series is truncated after, say, the Mth term, the truncated sum is a polynomial in $(\epsilon^2 k^2 - 1)$ which, by rearrangement, is also a polynomial of degree $2M$ in powers of ϵk. The term-by-term integrals are therefore exactly the same as for the original asymptotic series. The difference is that these integrals enter with weight functions which are different from 1 because of the expand-in-a-different-variable-and-truncate step.

For example, with the second truncation set at $M = 3$ and using $z \equiv \epsilon^2 k^2 - 1$,

$$\frac{1}{2\left(1+(1/2)z\right)} \approx \frac{1}{2} - \frac{1}{4}z + \frac{1}{8}z^2 - \frac{1}{16}z^3 = \frac{15}{16} - \frac{11}{16}\epsilon^2 k^2 + \frac{5}{16}\epsilon^4 k^4 - \frac{1}{16}\epsilon^6 k^6. \tag{10.13}$$

Inserting this into the error integral (10.6) and using the identity that $(-1)^j k^{2j} F(k)$ is the Fourier transform of the $2j$th derivative of $f(x)$ gives

$$u_{slow} = \sum_{j=0}^{N} \epsilon^{2j} \frac{d^{2j} f}{dx^{2j}} + \frac{15}{16}\frac{d^{2N+2} f}{dx^{2N+2}} + \frac{11}{16}\frac{d^{2N+4} f}{dx^{2N+4}} + \frac{5}{16}\frac{d^{2N+6} f}{dx^{2N+6}} + \frac{1}{16}\frac{d^{2N+8} f}{dx^{2N+8}}. \tag{10.14}$$

This is exactly the same form as the standard multiple scales series except that the last four terms are multiplied by weights.

As explained in much more detail in [6], the device of improving the convergence of a series by expanding in a shifted variable is the Euler sum acceleration. Although we have justified this method by hyperasymptotic thinking, the Euler acceleration has been applied to convergent series for more than two centuries, long before Poincaré formulated his definition of asymptotic series, and even longer before the modern conceptualization of hyperasymptotics. The method is very simple in application: for each choice of M such that $(M+1)$ terms are weighted, there is a fixed set of numerical weights [6].

The Euler acceleration adds a second asymptotic-but-divergent series to the superasymptotic approximation. The Euler-accelerated series is also divergent because

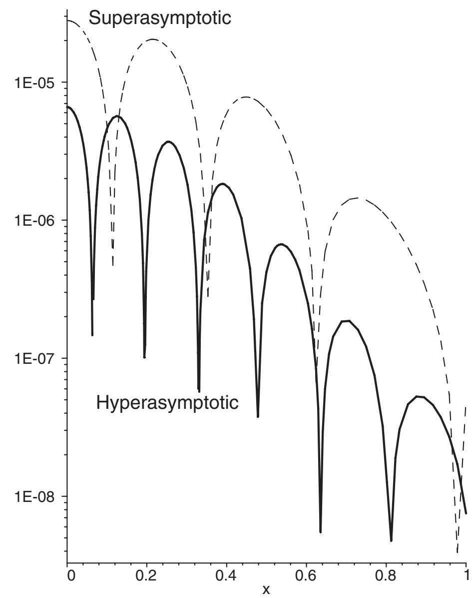

Fig. 10.2 *The errors for the superasymptotic approximation (dashed) and four-term hyperasymptotic improvement (solid) for $f = 1/(1+x^2)$ with $\epsilon = 1/12$ and, as the truncation of the standard asymptotic series, $N = 6$, which is the optimal truncation for this function for this ϵ.*

again the interval of integration in the error integral is infinite, and the expansion in powers of $(1/2)(\epsilon^2 k^2 - 1)$ has only a finite radius of convergence in k.

Figure 10.2 illustrates the improvement for a typical case.

11. Summary. Divergent series are ubiquitous in applied mathematics. Explanations of why series diverge and how the series can be improved have hitherto been confined to specialized monographs.

Our contention is that divergent series are not "the devil's invention," as Abel labeled them two centuries ago. Rather, it is possible to find examples of divergence and hyperasymptotics that can be explained using undergraduate mathematics. Of course, like any other rich field, there are depths and subtleties that can be probed only at the graduate level and beyond. But that is hardly a justification for the pedagogical crime of inflicting asymptotic series on students with definitions, rules, and heuristics, but without explanation.

The one-dimensional boundary value problem with boundary layers is a good introduction to the Why of divergence, and also to its cure.

Appendix. Obtaining a Nice Fourier Transform by Windowing. Even if $f(x)$ is an analytic function on $x \in [-1, 1]$, it may not have a nice Fourier transform $F(k)$; a counterexample is $f = x^4/(1 + x^2)$, which is unbounded as $|x| \to \infty$.

Theorem A.1 (windowing). *If $f(x)$ is a function which is analytic on the real interval $x \in [-\delta - 1, 1 + \delta]$ for some positive δ, however small, then for an arbitrarily small error tolerance α, one can always find a function $\tilde{f}$ that has a well-behaved Fourier transform $F(k)$, decaying exponentially fast as $|k| \to \infty$, and also such that*

$$|f(x) - \tilde{f}(x)| \le \alpha \qquad \forall\, x \in [-1, 1]. \tag{A.1}$$

This function is constructed as

$$\tilde{f}(x) \equiv f(x)\,\mathcal{W}(x), \tag{A.2}$$

where $\mathcal{W}$ is a function that is approximately or exactly equal to one on $x \in [-1,1]$ and rapidly diminishes to zero as $|x| \to \infty$.

Proof. The first crucial point is that if the window function is constructed such that

$$|\mathcal{W}(x) - 1| \leq \alpha / \max_{x\in[-1,1]}(|f(x)|), \qquad x \in [-1,1], \tag{A.3}$$

then the difference between f and $\tilde{f}$ will have to be less than α everywhere on the target interval. The second key idea is that if $\mathcal{W}$ is a smooth function that decays as $|x| \to \infty$, and if also f is sufficiently smooth for x on and near the interval $x \in [-1,1]$, then $\tilde{f}$ will have the same properties. The usual Fourier integral theorems then guarantee that $\tilde{f}$ will have a well-behaved Fourier transform $F(k)$.

There are many choices for the window function such as

$$\mathcal{W} \equiv \frac{1}{2}\left\{\operatorname{erf}(\lambda[x-1-\sigma]) - \operatorname{erf}(\lambda[x+1+\sigma])\right\}, \tag{A.4}$$

where λ and σ are user-choosable constants. Because this is an entire function, this "erf-window" does not introduce any additional singularities into $\tilde{f}$ except those present in f. Although $\tilde{f}$ never exactly equals f on $x \in [-1,1]$, the difference can be made arbitrarily small by choosing a sufficiently large λ.

Another choice is to use a C^∞ window, that is, a function which is infinitely differentiable but not analytic. An example is

$$\mathcal{W} = \begin{cases} 1, & x \in [-1,1], \\ (1/2)\left\{1 + \operatorname{erf}\left(L\frac{y(|x|)}{\sqrt{1-y(|x|)^2}}\right)\right\}, & 1 < |x| < \Theta, \\ 0, & |x| > \Theta, \end{cases} \tag{A.5}$$

where L and Θ are positive user-choosable constants and where

$$y(x) \equiv -1 + 2\frac{\Theta - x}{\Theta - 1}. \tag{A.6}$$

Figure A.1 shows such a window; the graph of the erf-window is very similar.

The lack of analyticity implies a slower decrease of $F(k)$ with $|k|$ than for the erf-window. The advantage is that $\tilde{f} \equiv f$ without error everywhere on $x \in [-1,1]$. In addition, the window is identically zero for sufficiently large real $|x|$. Even if $f(x)$ has poles or other unbounded behavior at some finite x, $\tilde{f}$ will be bounded and well-behaved for all real x if the window is made sufficiently narrow. ∎

Thus, no real generality is lost by using the Fourier transform representation of the particular solution: through windowing, the necessary $F(k)$ can always be constructed if $f(x)$ is defined for all real x. If $f(x)$ is defined or known only on a finite interval, then it can be extended to the whole real axis as explained in [7].

Acknowledgments. I thank William Briggs and two anonymous reviewers for very detailed comments that greatly improved this work.

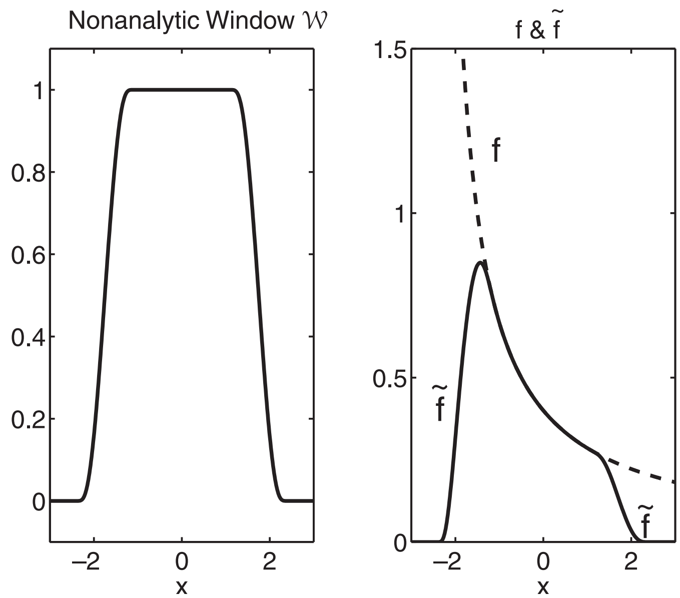

Fig. A.1 *Left: The C^∞ window $\mathcal{T}(x;\Theta,L)$ for $L=2$ and $\Theta=5/2$; the window is identically zero for all $|x|>\Theta$ and varies smoothly from one to zero in the regions $|x|\in[1,\Theta]$. Right panel (solid curve): a typical forcing function, $f=1/(5/2+x)$ and its windowed equivalent, $\tilde{f}$. Note that although f has an off-interval pole at $x=-2.5$, the windowed function $\tilde{f}$ is smooth for all real x and has a Fourier transform that decays exponentially fast with k.*

REFERENCES

[1] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1978.

[2] M. V. Berry, *Asymptotics, superasymptotics, hyperasymptotics*, in Asymptotics Beyond All Orders, H. Segur, S. Tanveer, and H. Levine, eds., Plenum, Amsterdam, 1991, pp. 1–14.

[3] M. V. Berry and C. J. Howls, *Hyperasymptotics*, Proc. Roy. Soc. London A, 430 (1990), pp. 653–668.

[4] J. P. Boyd, *Time-marching on the slow manifold: The relationship between the nonlinear Galerkin method and implicit timestepping algorithms*, Appl. Math. Lett., 7 (1994), pp. 95–99.

[5] J. P. Boyd, *Weakly Nonlocal Solitary Waves and Beyond-All-Orders Asymptotics: Generalized Solitons and Hyperasymptotic Perturbation Theory*, Math. Appl. 442, Kluwer Academic, Amsterdam, 1998.

[6] J. P. Boyd, *The Devil's invention: Asymptotics, superasymptotics and hyperasymptotics*, Acta Appl., 56 (1999), pp. 1–98.

[7] J. P. Boyd, *A comparison of numerical algorithms for Fourier extension of the first, second and third kinds*, J. Comput. Phys., 178, (2002), pp. 118–160.

[8] J. P. Boyd, *Approximation of an analytic function on a finite real interval by a bandlimited function and conjectures on properties of prolate spheroidal functions*, Appl. Comput. Harmonic Anal., 15 (2003), pp. 168–176.

[9] J. P. Boyd, *Fourier embedded domain methods: Extending a function defined on an irregular region to a rectangle so that the extension is spatially periodic and C^{oo}*, Appl. Math. Comput., 161 (2005), pp. 591–597.

[10] J. P. Boyd and A. Natarov, *A Sturm-Liouville eigenproblem of the Fourth Kind: A critical latitude with equatorial trapping*, Stud. Appl. Math., 101 (1998), pp. 433–455.

[11] G. F. Carrier and C. E. Pearson, *Ordinary Differential Equations*, Blaisdell, Waltham, MA, 1968.

[12] R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation*, Academic, New York, 1973.

[13] J. Écalle, *Les fonctions résurgentes*, Université de Paris-Sud, Paris, 1981.
[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed., Academic Press, New York, 1965.
[15] D. S. Jones, *Introduction to Asymptotics: A Treatment Using Nonstandard Analysis*, World Scientific, Singapore, 1997.
[16] J. B. Keller and M. J. Ward, *Asymptotics beyond all orders for a low Reynolds number flow*, J. Engrg. Math., 30 (1996), pp. 253–265.
[17] V. Kowalenko, M. L. Glasser, T. Taucher, and N. E. Frankel, *Generalised Euler-Jacobi Inversion Formula and Asymptotics Beyond All Orders*, London Math. Soc. Lecture Note Ser. 214, Cambridge University Press, Cambridge, UK, 1995.
[18] M. C. A. Kropinski, M. J. Ward, and J. B. Keller, *A hybrid asymptotic-numerical method for low Reynolds number flows past a cylindrical body*, SIAM J. Appl. Math., 55 (1995), pp. 1484–1510.
[19] J. C. Le Guillou and J. Zinn-Justin, eds., *Large-Order Behaviour of Perturbation Theory*, North-Holland, Amsterdam, 1990.
[20] M. J. Lighthill, *An Introduction to Fourier Analysis and Generalized Functions*, Cambridge University Press, New York, 1958.
[21] E. Lombardi, *Oscillatory Integrals and Phenomenon Beyond All Algebraic Orders*, Lecture Notes in Math. 1741, Springer-Verlag, New York, 2000.
[22] M. Marion and R. Témam, *Nonlinear Galerkin methods*, SIAM J. Numer. Anal., 26 (1989), pp. 1139–1157.
[23] V. P. Maslov, *The Complex WKB Method for Nonlinear Equations* I: *Linear Theory*, Birkhäuser, Boston, 1994.
[24] A. Natarov and J. P. Boyd, *Beyond-all-orders instability in the equatorial Kelvin wave*, Dyn. Atmos. Oceans, 33 (2001), pp. 181–200.
[25] A. Natarov and J. P. Boyd, *Shafer (Hermite-Padé) approximants for functions with exponentially small imaginary part with application to equatorial waves with critical latitude*, Appl. Math. Comput., 125 (2002), pp. 109–117.
[26] A. H. Nayfeh, *Perturbation Methods*, Wiley, New York, 1973.
[27] R. B. Paris and D. Kaminski, *Asymptotics and Mellin-Barnes Integrals*, Cambridge University Press, Cambridge, UK, 2001.
[28] R. B. Paris and A. D. Wood, *Stokes phenomenon demystified*, Bull. Inst. Math. Appl., 31 (1995), pp. 21–28.
[29] H. Segur and M. D. Kruskal, *On the nonexistence of small amplitude breather solutions in ϕ^4 theory*, Phys. Rev. Lett., 58 (1987), pp. 747–750.
[30] H. Segur, S. Tanveer, and H. Levine, eds., *Asymptotics Beyond All Orders*, Plenum, New York, 1991.
[31] A. V. Sergeev and D. Z. Goodson, *Summation of asymptotic expansions of multiple-valued functions using algebraic approximants: Application to anharmonic oscillators*, J. Phys. A, 31 (1998), pp. 4301–4317.
[32] B. Y. Sternin and V. E. Shatalov, *Borel-Laplace Transform and Asymptotic Theory: Introduction to Resurgent Analysis*, CRC Press, New York, 1996.
[33] M. Van Dyke, *Perturbation Methods in Fluid Mechanics*, 1st ed., Academic Press, Boston, 1964.
[34] E. J. Weniger, *Nonlinear sequence transformations for the acceleration of convergence and the summation of divergent series*, Comput. Phys. Rep., 10 (1989), pp. 189–371.
[35] J.-J. Xu, *Interfacial Wave Theory of Pattern Formation: Selection of Dendritic Growth and Viscous Fingering in Hele-Shaw Flow*, Springer-Verlag, New York, 1998.
[36] V. Yakhot, S. A. Orszag, S. Thangam, T. B. Gatski, and C. G. Speziale, *Development of turbulence models for shear flows by a double expansion technique*, Phys. Fluids A, 4 (1992), pp. 1510–1520.