Perturbation Methods using Backward Error

Robert M. Corless

GO20, Gozo, Malta, May 2024

Western University, Canada

Slides available at `rcorless.github.io`; please download them

Joint work with Nic Fillion

Maple Transactions

Announcing Maple Transactions

a "Diamond" class open access journal with no page charges

Now listed by DBLP

mapletransactions.org

There must be “a few books” already on Perturbation Methods

Figure 1: RMC and Don Quixote, in Alcalá de Henares, 2017

Why, on Earth, write another?

Fools rush in where angels fear to tread.
—Alexander Pope, *An essay on criticism*

- There are only two other books that use backward error [6, 7]
- We claim backward error is *very* useful for perturbation methods*
- We think computer algebra is still under-utilized nowadays, although there are some works that use it systematically
- Even though scientific computing has progressed *far* beyond perturbation methods, there is still a need for them.

* This fact may seem obvious in retrospect. We contend that the obstacle of hand labour has discouraged it in practice till now.

The goal: short lucid formulae

Numerical solution and graphs (and animations) are truly valuable, but sometimes a short lucid formula can tell you just as much as an hour with a simulator and visualization tools can.

This *depends* on the scientist (or student!) understanding the terms in the formula, of course!

The Watson–Wong–Wyman lemma

Suppose [8]

$$I(x) = \int_{t=0}^{b} f(t)e^{-xt}dt \tag{1}$$

exists and is finite for $x > 0$. We will take $x > 0$ and $b > 0$, and allow the case $b = \infty$ which will occasionally require explicit mention.

Suppose now that $\phi_k(t)$ is an asymptotic sequence for $k = 0, 1, 2, \ldots$ as $t \to 0^+$, which implies that $\phi_{k+1}(t) = o(\phi_k(t))$ as $t \to 0^+$, and suppose moreover that each $\phi_k(t) \geq 0$ for all $t \geq 0$. Suppose that

$$f(t) \sim \sum_{n=0}^{\infty} a_n \phi_n(t), \quad t \to 0^+ \,. \tag{2}$$

WWW lemma continued

Suppose also that $\psi_k(x) := \int_{t=0}^{c} \phi_k(t) \exp(-xt) dt$ (here c might be b, or ∞, or any convenient nonzero upper limit) is an asymptotic sequence for $k = 0, 1, 2, \ldots$ as $x \to \infty$. Finally, suppose that the $\psi_k(x)$ decay to zero more slowly than $\exp(-\alpha x)$ for any $\alpha > 0$. That is,

$$e^{\alpha x} \psi_k(x) \to \infty \tag{3}$$

as $x \to \infty$, for any integer $k \geq 0$ and for any real $\alpha > 0$.

Then

$$\int_0^b f(t) e^{-xt} dt \sim \sum_{n=0}^{\infty} a_n \psi_n(x), \quad x \to +\infty \,. \tag{4}$$

Why?

I wrote a short procedure (see the Jupyter notebook "A stronger Watson's lemma") for computation of asymptotic approximations to

$$\int_0^\infty f(t)e^{-xt}\,dt \tag{5}$$

using Watson's lemma. The code turned out to be able to compute expansions for which Watson's lemma did not apply: in particular, when the expansion of $f(t)$ at $t = 0^+$ contained logarithms, or contained exponentially small terms such as $\exp(-1/t)$. Maple uses *generalized* series, by default.

Wong and Wyman had explored this already in 1972, but it's not really until computer algebra caught up that it's become useful.

Aside: Generalized series

Since at least [2] Maple has computed *generalized* series approximations with its `series` command. That means that it allows its answers to contain certain other "elementary" terms, not just powers of the main variable. For instance, consider

$$x^{\sin x} = 1 + \ln(x)\,x + \frac{1}{2}\ln(x)^2\,x^2 + \left(-\frac{\ln(x)}{6} + \frac{\ln(x)^3}{6}\right)x^3 + \mathrm{O}\left(x^4\right) \quad (6)$$

which has a hierarchy of terms involving powers of x and powers of $\ln x$ in it; many such series are triangular in that not all powers of $\ln x$ occur at each power of x. The above was generated by the command `series( x^sin(x), x, 4 )`. The O symbol therefore includes slower-growing terms with the hidden constant.

Another example

$$\int_0^{\pi} t^{\sin t} e^{-xt}\, dt = \frac{1}{x} + \frac{1-\gamma-\ln(x)}{x^2} + \frac{1+\frac{\pi^2}{6}+\ln(x)^2+\gamma^2-3\gamma+\ln(x)\,(-3+2\gamma)}{x^3} + O\left(\frac{1}{x^4}\right) \tag{7}$$

which is made possible to compute in Maple with my "Watson" program, by means of the generalized series just mentioned. Checking at (just by chance) $x = 113$ gives 0.00852250 for the integral and 0.00852275 for the series; reasonable agreement.

“The method of exact solution”

My code says that $\int_0^\infty \exp(-xt)/(1-\exp(-1/t))dt$ is asymptotic to

$$\frac{\sqrt{\pi}\,\mathrm{e}^{-2\sqrt{x}}}{x^{\frac{3}{4}}} + \frac{1}{x} + \frac{3\sqrt{\pi}\,\mathrm{e}^{-2\sqrt{x}}}{16x^{\frac{5}{4}}} - \frac{15\sqrt{\pi}\,\mathrm{e}^{-2\sqrt{x}}}{512x^{\frac{7}{4}}} + O\left(\frac{1}{x^{\frac{9}{4}}}\right) \tag{8}$$

as $x \to \infty$. Notice the mixture of algebraically small and exponentially small terms. This works because Maple knows

$$\int_0^\infty \mathrm{e}^{-\frac{1}{t}}\mathrm{e}^{-tx}\,dt = \frac{2K_1\left(2\sqrt{x}\right)}{\sqrt{x}} \tag{9}$$

where K_1 is a Bessel K function.

Small backward error is not always needed

For quadrature problems, and Laplace transforms in particular, small forward error does not require small backward error (e.g. $\Delta f(t)$ could be large in a very narrow interval; or large for some large t, which gets squashed by $\exp(-xt)$).

However, it can be *enough* to explain success: small backward error $\Delta f(t)$ does imply small (absolute) forward error for quadrature, so there's that. This is because quadrature is well-conditioned: $|\Delta I| \leq \|\Delta f\|_1$ so the condition number is just 1. [Oscillatory integrals and relative error are a different matter.]

Is it obvious that these asymptotic series are the exact integrals of perturbed integrands? Maybe to all of us here today, but maybe not, too; and certainly not to all potential readers of the book, some of whom might be students. This part needs to be explained in detail, more than is currently in the draft. Paradoxically, this is easier for the solutions of ODE.

A classical ODE

The "Renormalization Group Method" [3] makes short work of many nonlinear oscillator problems. Consider the Rayleigh equation

$$\ddot{y} - \varepsilon\dot{y}\left(1 - \frac{4}{3}\dot{y}^2\right) + y = 0\,. \tag{10}$$

"This RG method works, although it is somewhat inefficient since it first obtains the naive expansion. . ."

— Robert E. O'Malley [4, p. 187]

The Renormalization Group method in a nutshell

- Choose N and compute the regular solution to $O(\varepsilon^{N+1})$ with secular terms in it, starting with $y_0(t) = 2A\cos(t) = A\exp(it)$+c.c.
- Gather up the term $Ay_A(t)e^{it}$ (we want the coefficient $y_A(t) = 1 + O(\varepsilon)$ of the resonant term)
- Replace the initial amplitude A by the time-dependent amplitude $R(t)$ determined by solving the equation $f(A,R,\varepsilon) = R^2 - A^2(\Re(y_A)^2 + \Im(y_A)^2) = 0$ perturbatively (regular perturbation works!)
- Determine the differential equations for $R(t)$ and $\theta(t)$ via

$$\frac{R'(t)}{R(t)} + i\theta'(t) = \frac{y_A'(t)}{y_A(t)} \tag{11}$$

  Set $t = 0$ after differentiation in the right-hand side, and replace every A by the R that you found; cf Lie group–Lie algebra exponential.
- Redo the computation with initial approximation $y(t) = 2R(t)\cos(t + \theta(t))$, using the differential equations for R and θ as you go along.

Results

It turns out $\theta'(t) = 0$. Then if $y = 2R(t)\cos t + \varepsilon R^3(t)\cos 3t/3$ where

$$\frac{d}{dt}R(t) = \frac{\varepsilon}{2}R(t)\left(1 - 4R^2(t)\right) \tag{12}$$

(an equation we may expect students to be able to solve by hand; note the stable limit cycle at $R = 1/2$) then $y(t)$ exactly solves

$$\ddot{y} - \varepsilon\dot{y}\left(1 - \frac{4}{3}\dot{y}^2\right) + y = r(t) = \varepsilon^2 v(t) \tag{13}$$

where $v(t)$ is precisely known (we have a formula for it, which we can inspect explicitly, but it's a little long to present here). Moreover, $v(t)$ is bounded for all time t.

Among other things, this shows that our computation was free of blunders. We did this up to $N = 16$, by the way, essentially for fun (the code took 35 minutes to do that case).

The residual $r(t) = \varepsilon^2 v(t)$

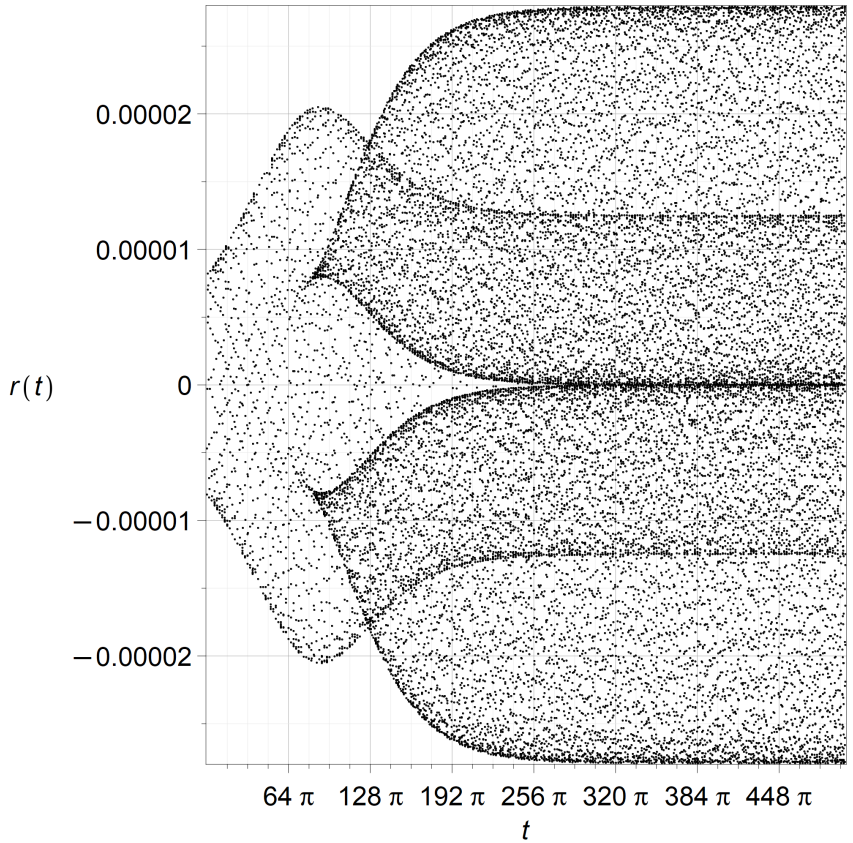

Figure 2: $R_0 = 0.15$ and $\varepsilon = 0.01$.

The effects of a small residual

The Rayleigh equation is (near its limit cycle) extremely well-conditioned, although phase error can accumulate algebraically. The residual $r(t)$, which is small for small $\varepsilon > 0$, has very little effect on the solution—which is as it should be, because real physical oscillators are frequently subject to small forcing oscillations (shaky ground, shaky table, and the like).

Our highest-order computation, $O(\varepsilon^{17})$, had uniformly very small residual, which at the limit cycle was well below rounding error size when $\varepsilon = 0.2$. With $N = 13$ and $\varepsilon = 0.4$, the residual was about 10^{-8}.

Resonant terms

Curiously enough, that residual contains resonant terms, in that the coefficients of $\sin t$ and $\cos t$ are not zero! This might be surprising, but one can compute a *structured* backward error, at the limit cycle, that looks like this:

$$\ddot{y} - \varepsilon\dot{y}\left(1 - (1+\varepsilon^2 A)\frac{4}{3}\dot{y}^2\right) + (1+\varepsilon^2 B)y = r_1(t) = \varepsilon^2 v_1(t) \qquad (14)$$

for some bounded quantities A and B and where now the new residual $r_1(t)$ contains no resonant terms. That is, we might interpret the residual as containing a slight change in frequency and in damping, as well as a small forcing.

In fact we find $A = -1/32 + O(\varepsilon^2)$ and $B = 1/8 + O(\varepsilon^2)$ at the limit cycle.

Absence of secularity

The normal treatment of secularity assumes that one knows ahead of time that the reference solution to the model problem is bounded. Sometimes that's obvious physically, but sometimes one has to prove boundedness (e.g. with the Duffing equation one finds a conserved integral).

But here if we can find a solution (as here) with a uniformly bounded residual, this provides its own proof of validity.

The argument looks a bit circular, but it's not. We used the RG method to find a solution that had no secular terms in it, and hence which had a bounded residual, which proved the solution valid (in the context of modelling errors).

Morrison's counterexample

In [4, pp. 192–193], we find a discussion of the equation

$$y'' + y + \varepsilon(y')^3 + 3\varepsilon^2(y') = 0\,. \tag{15}$$

O'Malley's solution, there and in [5], is incorrect. We claim that had he computed a residual, he would have identified the blunder*.

* By "blunder" we mean arithmetic error, or algebra error, no more. It's just that the word "error" is a bit overused in this field already. Also, I feel quite a bit of trepidation in pointing out this blunder: O'Malley was a giant of perturbation methods. But we are certain that our solution is correct.

RG Solution

All we need do is to change the differential equation in our Jupyter notebook script, and alter the interrogations of the solution afterward. At $N = 2$, we get

$$\begin{aligned} z(t) =& 2R(t)\cos(t+\theta(t)) + \frac{\varepsilon R(t)^3 \sin(3t+3\theta(t))}{4} \\ &+ \varepsilon^2 \left(\frac{27R(t)^5 \cos(3t+3\theta(t))}{32} - \frac{3R(t)^5 \cos(5t+5\theta(t))}{32} \right) \end{aligned} \tag{16}$$

with

$$\dot{R}(t) = -\frac{3\varepsilon}{2} R(t) \left(R(t)^2 + \varepsilon \right) \tag{17}$$

and

$$\dot{\theta}(t) = \frac{9}{16} R^4(t) \varepsilon^2 \,. \tag{18}$$

With this, we get a uniformly small residual, which is small even compared to the decaying amplitude.

Aging spring, Lengthened pendulum

Some physical problems have natural secular terms in them. For instance, consider the "aging spring" [1]:

$$\ddot{y} + e^{-\varepsilon t} y = 0\,. \tag{19}$$

Cheng and Wu claimed to have used the two-scale method* to get their putative solution $\exp(\varepsilon t/4)\sin(2(1-\exp(-\varepsilon t/2))/\varepsilon)$, which looks bizarre to me. But, its residual is

$$\frac{1}{16}\varepsilon^2 \mathrm{e}^{\varepsilon t/4} \sin\left(\frac{2(1-\mathrm{e}^{-\varepsilon t/2})}{\varepsilon}\right)\,. \tag{20}$$

*Hoo boy. I think you have to *abuse* the method. I don't really know what they did; maybe I reconstructed their computation. But because the residual is small, it wouldn't have mattered if they had used dowsing to find the solution.

Better backward error

Notice that the residual in equation (20) is just $\varepsilon^2 Y(t)/16$ where $Y(t)$ is the two-scale solution. This means that $Y(t)$ is the *exact* solution to

$$y'' + \left(e^{-\varepsilon t} - \frac{\varepsilon^2}{16} \right) y(t) = 0 \,. \tag{21}$$

This is an equation that we can *directly* interpret in terms of the original model.

Notice that the spring constant becomes *negative* when $\exp(-\varepsilon t) = \varepsilon^2/16$, or $t = -2\ln(\varepsilon/4)/\varepsilon$. We thus learn that the approximate solution is likely not valid for t larger than this, *in a way that is consonant with the mathematical modelling*. [Cheng and Wu say that this equation is used in some kind of quantum application.]

The aging spring is ill-conditioned

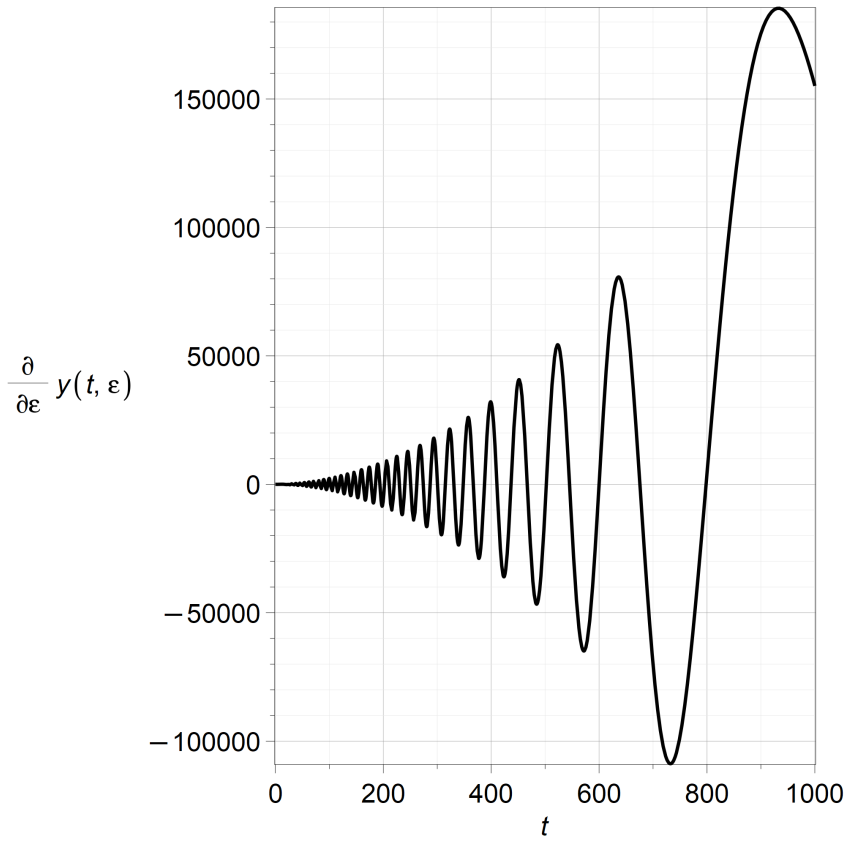

Figure 3: Taking the derivative with respect to ε shows that the solution is ill-conditioned. $\varepsilon = 1/100$ here.

Perturbation vs Exact Solution

The analysis of the aging spring just performed—exhibiting an approximate solution that is the exact solution of a nearby problem of similar type, together with a residual and a condition number—tells us at least as much information as the exact solution in terms of Bessel functions does.

We have identified an important issue, namely the sensitivity of the solution to changes in the problem, that will still be important for the exact (reference) solution.

"Small" vs "Small enough"

Since Backward Error Analysis requires the *context* of the original problem to be taken into account, this explicitly allows us to consider whether the computed residual (or other backward error form) is actually small compared to other neglected effects.

This is *not* mathematics! Mathematics abstracts, as far as possible, with the goal of making its results and predictions independent of context.

If in our actual modelling instance, the small parameter was (say) $\varepsilon = 0.34$ and the residual was bounded by (say) 10^{-3}, then we *would need to know the problem context* to know if the residual was small compared to effects already neglected.

This is the crux of the matter.

Once this is settled, we can consider conditioning: are such small effects amplified to the point where we lose all predictability or control, or are we ok?

An important tangent: Published blunders

Off the top of my head, blunders in published perturbation computations have been exhibited by

- John P. Boyd (a hyperasymptotic expansion)
- Robert E. O'Malley (Morrison's counterexample)
- Émile Mathieu (in his 1868 paper which defined what are now called Mathieu functions)
- Bender & Orszag (a plain multiple scales computation, fixed in later editions)

at least. I claim that had they computed a final residual, they would have detected their blunders. Given that *all* of the above are/were experts, and therefore it's clear that the rest of us make blunders at least as frequently, I claim that residuals are even more necessary for us.

Perturbation is just taking derivatives

The simplicity of a perturbation computation hides its importance. We are investigating what happens if a small part of the model changes.

This is itself a fundamental question of science. It's not surprising that the old techniques are still valuable; maybe it's a surprise just how valuable they can be.

That said, nowadays one can do a heck of a lot with a simulation window and a slider bar.

Thank you for listening.

This work supported by NSERC, and by the Spanish MICINN.

I am also happy to announce that SIAM has offered Nic and me a contract for this book, and we are to deliver it to them by December. Your feedback today will help to improve the book, and we will acknowledge you all.

Book text available at `https://github.com/rcorless/rcorless.github.io/blob/main/PerturbationBEABook.pdf`. Please download it and read it and send me (or Nic) your comments, by June 30 if possible.

Let's open the topic for discussion.

References

[1] Hung Cheng and Tai Tsun Wu. **"An aging spring".** In: *Studies in applied Mathematics* 49.2 (1970), pp. 183–185.

[2] Keith O Geddes and Gaston H Gonnet. **"A new algorithm for computing symbolic limits using hierarchical series".** In: *International Symposium on Symbolic and Algebraic Computation*. Springer. 1988, pp. 490–495.

[3] Eleftherios Kirkinis. **"The Renormalization Group: A perturbation method for the graduate curriculum".** In: *SIAM Review* 54.2 (2012), pp. 374–388.

[4] Robert E. O'Malley. **Historical Developments in Singular Pertubations.** Springer, 2014.

References ii

[5] Robert E. O'Malley and Eleftherios Kirkinis. **"A Combined Renormalization Group-Multiple Scale Method for Singularly Perturbed Problems".** In: *Studies in Applied Mathematics* 124.4 (2010), pp. 383–410.

[6] Anthony John Roberts. **Model emergent dynamics in complex systems.** SIAM, 2014.

[7] Donald R. Smith. **Singular-perturbation Theory.** Cambridge University Press, 1985.

[8] R Wong and M Wyman. **"Generalization of Watson's Lemma".** In: *Canadian Journal of Mathematics* 24.2 (1972), pp. 185–208.

Using computation to illustrate formulae

Let's try to understand which is bigger, the term $\exp(-1/\varepsilon)$ or any algebraic term ε^j. L'Hopital's rule shows that as $\varepsilon \to 0^+$ the exponential is transcendentally smaller than any ε^j. But what happens if we ask when the two are equal?

$$e^{-1/\varepsilon} = \varepsilon^j \tag{22}$$

exactly when $\varepsilon_{-1} = e^{W_{-1}(-1/j)}$ (on the left) and when $\varepsilon_0 = e^{W_0(-1/j)}$. Here W_{-1} and W_0 are the two real branches of the Lambert W function. [Short, lucid formulae, just what we want*.]

So ε^j is *smaller* than $\exp(-1/\varepsilon)$ if $\varepsilon_{-1} < \varepsilon < \varepsilon_0$. Paradoxically, this is most of the interval, for large j!

* Heh. $\varepsilon_{-1} \sim 1/(j \ln j)$ and $\varepsilon_0 \sim 1 - 1/j$ might be easier to understand!

When is that?

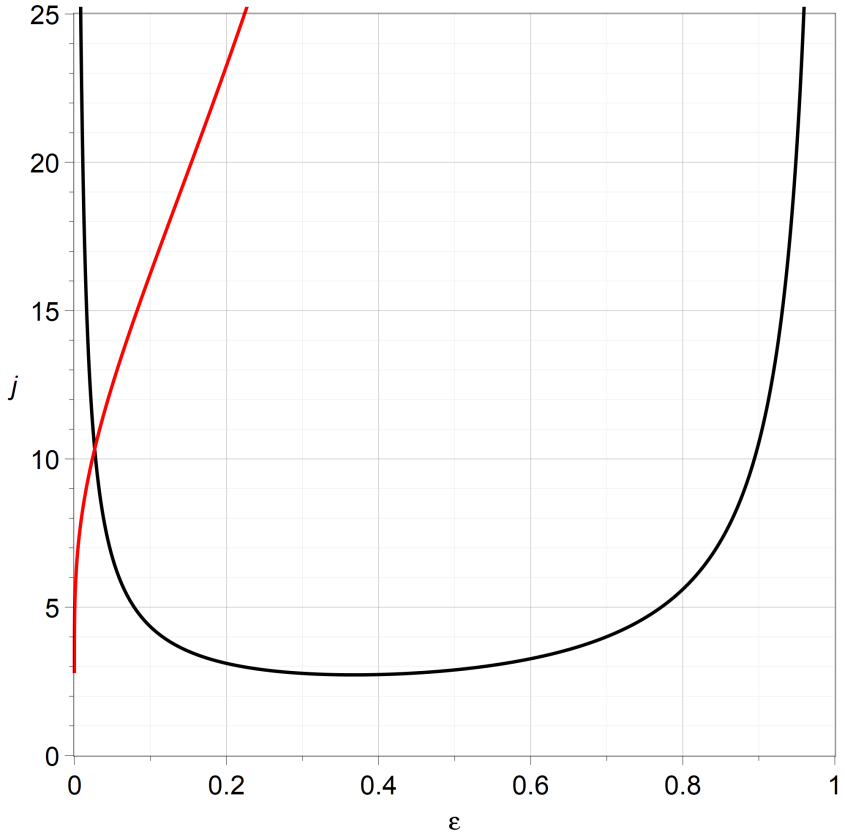

Figure 4: For values of j above this curve, $\varepsilon^j < \exp(-1/\varepsilon)$. That is, the "exponentially small" term is more important! Left of the red line is lost to rounding error in double precision. Note $\exp(-1/\varepsilon) = 2^{-54}$ already when $\varepsilon = \ln(2)/54 \approx 0.0267$.