

Richardson Extrapolation

Philipp K. Janert
(Dated: April 2006)

In this paper, we look at *Richardson Extrapolation* — a clever technique to overcome the pernicious finite-differencing error in numerical analysis.

PREREQUISITES

Note: This paper requires familiarity with derivatives and Taylor's Theorem.

FINITE DIFFERENCING

The main concept of mathematical analysis is the notion of the *limit process*. It arises for instance in the definition of the derivative as limit of the difference quotient:

$$f'(x) = \lim_{\epsilon \rightarrow 0} \frac{f(x + \epsilon) - f(x)}{\epsilon}$$

or in the definition of the integral as the limit of Riemann-Sums:

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{b-a}{n} f\left(a + k \frac{b-a}{n}\right)$$

The crucial insight of mathematical analysis was the realization that the expressions on the left-hand-side of these equations may well exist (provided $f(x)$ is sufficiently well-behaved), even as $\epsilon \rightarrow 0$ or $n \rightarrow \infty$. Much of basic calculus deals with finding analytical expressions for these limits for specific choices of $f(x)$ — this is where rules such as $(x^n)' = nx^{n-1}$ come from. (Here and in the following we are just going to use derivatives to demonstrate the ideas — application to integration is quite straightforward.)

But what to do when $f(x)$ is not in one of the handy tables of “standard” functions? Well, we can approximate the derivative numerically: just take a really small value for ϵ and evaluate the difference quotient, right? Not quite.

There are two different problems with this naive approach. The first should be obvious to anyone who has done even a little numerical analysis. If we make ϵ small, $f(x + \epsilon)$ will be close to $f(x)$, so we end up subtracting two almost equal numbers from one another, with the resulting loss of precision due to round-off errors. (Imagine $f(x + \epsilon)$ and $f(x)$ to differ only in the last 6 digits of their 16-digit (**double**) representation. Subtracting one from the other then leads to a loss of the first 10 (!) digits in the result.) Consequentially, there is a minimum value for ϵ , below which the quality of the result degrades due to round-off.

But there is another problem, which may be less obvious. For any finite (non-vanishing) value of ϵ , $(f(x + \epsilon) - f(x)) / \epsilon$ does not equal $f'(x)$. The limit is only attained when ϵ actually reaches 0: as long as ϵ is non-zero, we are simply “not there yet”.

To see the consequences of this effect, let's consider the Taylor series expansion of $f(x)$ around x :

$$f(x + \epsilon) = f(x) + \epsilon f'(x) + \frac{\epsilon^2}{2!} f''(x) + \frac{\epsilon^3}{3!} f'''(x) + \dots$$

By re-arranging, we easily find:

$$f'(x) = \frac{f(x + \epsilon) - f(x)}{\epsilon} - \frac{\epsilon}{2!} f''(x) + \mathcal{O}(\epsilon^2)$$

In other words, when approximating f' through its difference quotient, we incur a *finite differencing error* which is linear in ϵ . Since we can't choose ϵ to be very small without ruining the result through round-off error, we seem to be caught in a bind.

Before discussing the solution out of this dilemma, let's first look at a straightforward way to improve the order of our approximations. Write:

$$f(x - \epsilon) = f(x) - \epsilon f'(x) + \frac{(-\epsilon)^2}{2!} f''(x) + \frac{(-\epsilon)^3}{3!} f'''(x) + \dots$$

(note the signs!), and now subtract $f(x - \epsilon)$ from $f(x + \epsilon)$ and divide by 2ϵ :

$$f'(x) = \frac{f(x + \epsilon) - f(x - \epsilon)}{2\epsilon} - \frac{\epsilon^2}{3!} f'''(x) + \mathcal{O}(\epsilon^4)$$

Now the leading correction is *quadratic* in ϵ , since all terms containing odd powers cancel. If we assume that ϵ may be around 10^{-6} in a double-precision calculation, this reduces the finite-differencing error from 10^{-6} to 10^{12} . Much more acceptable!

But it seems that this is the best that we can do — or not?

REACHING FOR ZERO

The important insight is that we know more about the behavior of these quantities: we know that the finite-differencing error behaves as a linear function of ϵ in the first case and as a quadratic function in the second.

With this in mind, we can rewrite the expressions above as follows:

$$\frac{f(x + \epsilon) - f(x)}{\epsilon} = f'(x) + C_1(x)\epsilon + \mathcal{O}(\epsilon^2)$$

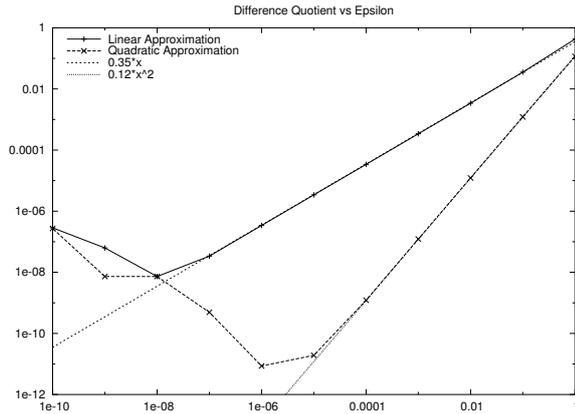


FIG. 1: Absolute value of Difference Quotient as function of ϵ as double-log plot (the exact value of the derivative has been subtracted).

and

$$\frac{f(x + \epsilon) - f(x - \epsilon)}{\epsilon} = f'(x) + C_2(x)\epsilon^2 + \mathcal{O}(\epsilon^4)$$

Here, $f'(x)$ is the (unknown) exact derivative we want to calculate and C_1, C_2 are (also unknown) constants — they may (and in general will) depend on x , but they do *not* depend on ϵ . If we calculate the left-hand-side for several values of ϵ , we can extract those constants, and then *extrapolate* to $\epsilon = 0$. That's the idea behind Richardson Extrapolation.

Figure 1 shows the calculated value of the difference quotient as a function of ϵ for $f(x) = \sin(x)$, evaluated at $x = 3/4 (\approx \pi/4)$ in a double-logarithmic plot. (The known value of $f'(x) = \cos(x)$ has been subtracted and the absolute value of the expression taken to allow the data to be plotted in a double-log plot.)

Two things stand out about this graph: for relatively large values of ϵ ($10^{-6} \lesssim \epsilon \lesssim 1$), the leading order behavior is a very good approximation — as is shown by the good agreement with the straight lines (corresponding to x and x^2 , respectively). On the other hand, for ϵ less than about 10^{-6} , roundoff error increasingly starts to corrupt the result.

We can now determine the constants C_1 and C_2 by evaluating the expressions above for several values of ϵ , and thus eliminate them from the expression for f' . (Evaluate the difference quotient for two consecutive values of ϵ . Subtract the results from one another — this eliminates f' , leaving a single equation containing just one unknown: the constant C_1 or C_2 . Solve for C and plug it into the original equation, which can now be solved for f' .)

The result, after removing the the leading order finite-differencing error as described above, is shown in Figure 2 (again, to enable a double-log plot, we subtracted the known value of f' and took absolute values). Qualita-

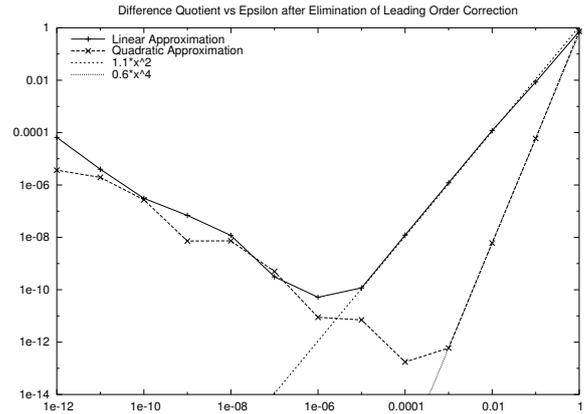


FIG. 2: Same as before, after eliminating the leading-order finite differencing error.

tively, this graph looks the same as the previous one: for “large” values of ϵ , we find a well-defined dependence on ϵ , for “small” ϵ the result is corrupted by round-off. What we are observing now is the next-to-leading order behavior. Note the improvement in the absolute size of the error over the previous graph!

Clearly, we can apply the same method as discussed above again to eliminate the error term to the next-highest order. And so on...

This method of successively eliminating finite differencing error in the numerical evaluation of derivatives can be automated, using “Neville’s” scheme for interpolating polynomials (references below).

QUESTIONS AND ALTERNATIVES

There are two main points to be taken from this paper: First of all, it may come as mild surprise how much predictive power rests in the commonplace expression “plus terms of order ϵ and higher” or its mathematical equivalent $\dots + \mathcal{O}(\epsilon)$. When doing analytical work, it is easy to get into a mind-set which regards error terms in expansions of all sorts effectively as “unknown garbage, hopefully small, certainly ignored”. Given this sort of attitude, it is startling to see how an expression really *does* depend on ϵ as indicated by the error term.

The second, much more practical take-away concerns the general concept behind Richardson Extrapolation. Although we have for simplicity’s sake concentrated on the evaluation of derivatives here, the same scheme (often under a different name) can be applied to any situation which involves finite-differencing. When applied to the numerical evaluation of definite integrals it is known as “Romberg Integration”, when applied to the numerical integration of ordinary differential equations it goes by the name of “Bulirsch-Stoer Method”. But the basic idea is even more general than that — whenever there is

some information about a problem that we have not yet used, we should be able to use it to improve our solution.

FURTHER READING

Applications of Richardson Extrapolation to problems in numerical analysis can be found in many books on numerics. **Numerical Recipes in C** by W. H. Press et.

al. contains a detailed discussion of the the evaluation of derivatives and an implementation using Neville's algorithm. The classic book by by J. Stoer and R. Bulirsch (**Introduction to Numerical Analysis**) contains many more advanced facets — such as the idea of using rational functions (instead of polynomials) for the extrapolation step.

© 2006 by Philipp K. Janert. All rights reserved.