Convergence acceleration by orthogonal polynomials

Dirk Laurie

University of Stellenbosch

Luminy, 29 September 2009

Sequence or series

Let $s_n,\ n = 1, 2, 3, \ldots$ be a sequence in a vector space $\mathcal{S}$.
Often $\mathcal{S}$ is a field, in fact usually $\mathbb{R}$ or $\mathbb{C}$.

- $s_n \longrightarrow s$, not very fast.
- $a_k = \Delta s_k = s_{k+1} - s_k, \qquad s_n = \sum_{k=0}^{n-1} a_k.$
- Definition implies $s_0 = 0$.

For some sequences, defining $s_0 = 0$ when it was not given, is nonsense.

The convergence acceleration matrix

$s_{k,n}$ depends on precisely $s_k, s_{k+1}, s_{k+2}, \ldots, s_{k+n}$.

Thus $s_{k,0} = s_k$.

Not all methods define $s_{k,n}$ for all possible pairs (k, n).

$$\begin{bmatrix} 0 & s_{0,1} & s_{0,2} & \cdots & s_{0,n-1} & s_{0,n} \\ s_{1,0} & s_{1,1} & s_{1,2} & \cdots & s_{1,n-1} & \\ \vdots & \vdots & \vdots & \iddots & & \\ s_{n-2,0} & s_{n-2,1} & s_{n-2,2} & & & \\ s_{n-1,0} & s_{n-1,1} & & & & \\ s_{n,0} & & & & & \end{bmatrix}$$

Transformed sequences

The first (if s_0 is allowed) or second row,

$$s'_k = s_{0,k}, \qquad s''_k = s_{1,k-1},$$

is in practice a good compromise between a single value and the full acceleration matrix.

Linearity and convergence

Linear acceleration methods $t_k = \alpha s_k + u_k \implies t_{k,n} = \alpha s_{k,n} + u_{k,n}$

Convergence rate A real number such that $\lim_{k\to\infty} r^{-k} a_k$ exists and is nonzero. (Signed convergence radius.)

Logarithmic convergence if $r = 1$.

For this talk,

$$-1 \leqslant r < 1.$$

Recursion formula

All linear methods can be written as

$$s_{k,j} = s_{k+1,j-1} + \frac{r_{k,j}}{1 - r_{k,j}}(s_{k+1,j-1} - s_{k,j-1}),$$

where $r_{k,j}$ is preassigned.

Not the same as saying there is a simple formula for $r_{k,j}$.

$$\begin{array}{ccc} s_{k,j-1} & \longrightarrow & s_{k,j} \\ s_{k+1,j-1} & \nearrow & \end{array}$$

Choice of $r_{k,j}$

If $r^{-k}a_k$ is constant, $r_{k,1} = r \implies s_{k,j} = s, \; j \geqslant 1$.

Generalized Euler method Choose $r_{k,j} = r_j$ equal to the (known, expected or guessed) convergence rate of $s_{\cdot,j-1}$.

Original Euler method If $r < 0$, choose $r_{k,j} = -1$ always.

An example when linear acceleration works spectacularly well

Let s_k be the perimeter of the regular polygon with $n_k = 3 \,.\, 2^k$ sides inscribed in the unit circle. ARCHIMEDES $[-240 \pm \varepsilon]^*$

$$s_k = n_k \sin(\pi/n_k)$$

Since the power series for $(\sin \pi x)/x$ contains only even powers, take $r_j = 4^{-j}$. Case $j = 1$: HUYGENS [1654]

*A palimpsest containing the words "Archimedes scripsit anno CCXL ante natem Christi" is believed to be a fake, possibly because it is in Latin, not in Greek.

Acceleration by Richardson extrapolation

$$s_{k,j} = s_{k+1,j-1} + \frac{4^{-j}}{1-4^{-j}}(s_{k+1,j-1} - s_{k,j-1}).$$

k	s_k	s''_k
1	3.00000000000000	3.00000000000000
2	3.10582854123025	3.14110472164033
3	3.13262861328124	3.14159245389765
4	3.13935020304687	3.14159265357789
5	3.14103195089051	3.14159265358979

s_k has approximately $0.6(k+1)$ correct digits. ($\log_{10} 4$)

s''_k has approximately $0.6k(k+1)$ correct digits.

What can we learn from this example?

For linear extrapolation methods, the following two properties tend to occur together.

- Very precise information about error expansion.
- Very fast convergence of transformed sequence.

Are there ‘universal’ linear methods?

- No.
- Not in the sense that they work in all cases.
- Yes, if r is known and $r \neq 1$. (That’s not universal!)
- Well, $r = -1$ includes the slowest convergent alternating series.
- Actually, methods based on assuming $r = -1$ work fairly well for other $r < 0$.

The case $r = -1$ (alternating series)

Model problem:

$$s_k = 1 - 1/2 + 1/3 - 1/4 + \cdots + a_k,$$
$$a_k = (-1)^{k+1}/k$$

$$s = \log 2 \doteq 0.69314718055994\overline{5}$$

Euler's method

$$\text{For model problem} \quad s_{1,15} = 0.69314728193932\,0, \qquad \text{error} = 1.0\,10^{-7},$$
$$\text{But} \quad s_{6,10} = 0.69314717536153\,1, \qquad \text{error} = 5.2\,10^{-9}.$$

Coincidence, or is it always better to stop earlier?

Van Wijngaarden's method (or something equivalent to it)

Take the subsequence $s_{1+k,2k}$, $k = 1, 2, \ldots,$ of Euler's method.

Equivalently, if only one value is needed: given $3n + 1$ terms of an alternating series with $r = -1$, throw away the first n of the partial sums and apply Euler's method to the remainder.

Obviously this idea should not be repeated until only one or two terms are left! Why does it work the first time round?

Totally alternating series

Totally monotone sequence:

a_k and its differences $\Delta a_k = a_{k+1} - a_k$, $\Delta^2 a_k, \ldots,$ are monotone.

Totally alternating series:

$s_k = \sum_{m=0}^{k-1} (-1)^m b_m$ where b_k is totally monotone.

Hausdorff's Theorem, generalized:

A series with convergence rate r is totally monotone when $0 \leqslant r < 1$, or totally alternating when $-1 \leqslant r \leqslant 0$, if and only if its terms can be represented as $a_k = \displaystyle\int_0^r t^k w(t)\,dt$, where w does not change sign.

Application to convergence acceleration

[CRVZ] H. Cohen, F. Rodriguez Villegas, D. Zagier, Experimental Mathematics **9** (2000) 3–12.

Assume that $s_0 = 0$ is valid. Partial sums and limit:

$$s_k = \sum_{j=0}^{k-1} \int_0^r t^j w(t)\,dt = \int_0^r \frac{(1-t^k)w(t)\,dt}{1-t}$$

$$s = \int_0^r \frac{w(t)\,dt}{1-t}$$

Accelerated values and error

General linear: $s_{k,n} = c_{0,n}s_k + c_{1,n}s_{k+1} + \cdots + c_{n,n}s_{k+n}$

Define $p_n(t) = c_{0,n} + c_{1,n}t + \cdots + c_{n,n}t^n$

$$\implies s_{k,n} = \int_0^r \frac{(p_n(1) - t^k p_n(t))w(t)\,dt}{1-t}$$

Acceleration leaves a constant series invariant $\implies$ $p_n(1) = 1$.

$$s - s_{k,n} = \int_0^r \frac{t^k p_n(t)}{1-t} w(t)\,dt$$

Error bounds

For $-1 \leqslant r < 1$:

$$s - s_{k,n} = \int_0^r \frac{t^k p_n(t)}{1-t} w(t)\,dt$$

$$|s - s_{k,n}| \leqslant \left| \frac{r^k}{1-r} \int_0^r p_n(t) w(t)\,dt \right|$$

$$\left| 1 - \frac{s_{0,n}}{s} \right| \leqslant \|p_n\|_r$$

$$\|p_n\|_r = \max_{t \in I_r} |p_n(t)| \quad \text{where } I_r = [\min(0,r), \max(0,r)].$$

Euler's and Van Wijngaarden's method revisited

Take $r = -1$.

For Euler's method,

$$p_n(t) = ((1+t)/2)^n \implies \|p_n\|_{-1} = 2^{-n}.$$

For Van Wijngaarden's method, for n divisible by 3,

$$p_n(t) = \left(t(1+t)^2/4\right)^{n/3} \implies \|p_n\|_{-1} = \left|p_3\left(-\tfrac{1}{3}\right)\right|^{n/3} = 3^{-n}.$$

How to choose p_n

$$p_n(1) = 1$$

because a constant sequence must be unchanged when “accelerated”.

What polynomial is as small as possible over $[0, r]$ but satisfies this constraint?

Shifted Chebyshev polynomials

$$p_n(t) = P_n(t) = k_n T_n(-1 + 2t/r),$$

where k_n is chosen so that $p_n(1) = 1$. GUSTAFSON [1978]

When $r = -1$, $\|P_n\|_{-1} = 5.828^{-n}$.
$(3 + 2\sqrt{2} \doteq 5.828)$

NB: Good convergence only along rows. Nothing special about $t^k p_n(t)$ when k increases and n is fixed.

$$
\begin{array}{c}
0.666666666666667 \\
0.705882352941177 \\
0.693602693602694 \\
0.693240901213172 \\
0.693150956487263 \\
0.693148308759757 \\
0.693147230444045 \\
0.693147198701394 \\
0.693147181406835 \\
0.693147180897730 \\
0.693147180576273
\end{array}
$$

$|s - s_{0,11}| = 1.6\,10^{-11}$. Compare with $5.828^{-11} = 3.8\,10^{-9}$.

[CRVZ]: even better

If w is assumed to be smooth, integration by parts is justified. $\rightsquigarrow$ One can select $p_n(t)$ from the matrix of Zagier polynomials [PARI/GP]

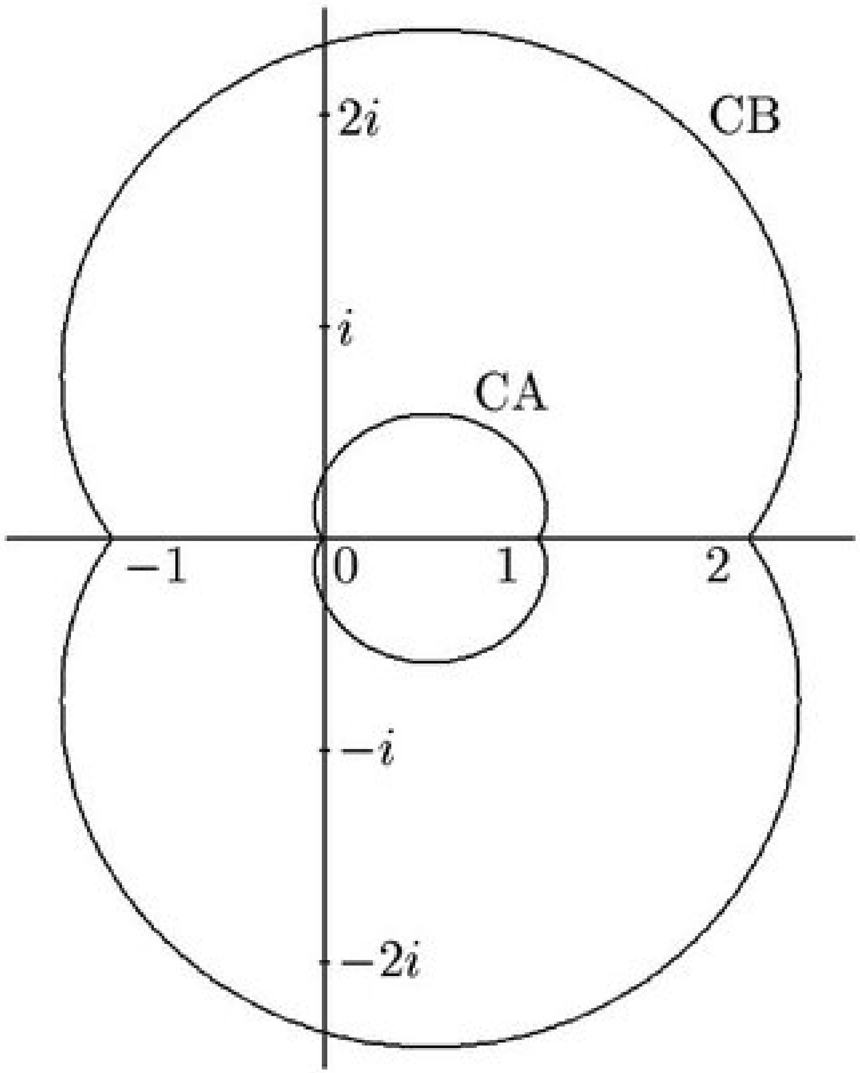
$$P_{m,n} = k_{m,n} \nabla^{2m}(n^{m+1} P_n),$$

$$\text{where} \quad \nabla^2 P_n = P_n - P_{n-2}$$

Two subsequences recommended by [CRVZ]:

A: $m = n - 1$. We need $P_{-k} = P_k$. Remember $P_k(t) = \cos(k\theta(t))$.

B: $m = \lfloor n/2 \rfloor$.



Error bounds for the CRVZ polynomials

[CRVZ] says:

If w is analytic in the small region,

Case A has error 7.89^{-n} and

Case B has error 9.56^{-n}.

If w is analytic in the large region,

Case A has error 17.93^{-n} and

Case B has error 14.41^{-n}.

De Boor classification* of Zagier polynomials

A cautious person will only use $m = 0$. (Chebyshev polynomials)

A reasonable person will use $m = \lfloor n/2 \rfloor$. (Case B)

An adventurous person will use $m = n - 1$. (Case A)

* *The need to think analytically . . . has been replaced by psychoanalytic introspection into the user's own personality.* — Phil Davis

CRVZ polynomials (Case A) in action on model problem

0.666666666666667
0.686274509803922
0.693693693693694
0.693176196711771
0.693145649808004
0.693147065004871
0.693147178638311
0.693147180715762
0.693147180574906
0.693147180560469
0.693147180559937

$|s - s_{0,11}| = 8.2\,10^{-15}$. Compare with $17.93^{-11} = 1.6\,10^{-14}$.

Why optimize bounds?

The error estimate is

$$s - s_{k,n} = \int_0^r \frac{t^k p_n(t)}{1-t} w(t)\, dt$$

Can't we do better than to optimize error bounds?

Acceleration by orthogonal polynomials

It's just a moment integral, after all.

Remember, $r < 0$, so $(1-t)^{-1}$ is harmless.

$$s - s_{k,n} = \int_0^r \frac{t^k p_n(t)}{1-t} w(t)\,dt$$

Why not rather use the shifted Legendre polynomials?

$$
\begin{array}{l}
0.666666666666667 \\
0.692307692307692 \\
0.693121693121693 \\
0.693146417445483 \\
0.693147157853040 \\
0.693147179886528 \\
0.693147180540013 \\
0.693147180559356 \\
0.693147180559928 \\
0.693147180559945 \\
0.693147180559945
\end{array}
$$

$|s - s_{0,10}| = 5.5\,10^{-16}$.

For CRVZ Case A we have $|s - s_{0,10}| = 5.2\,10^{-13}$.

Maybe that was just lucky

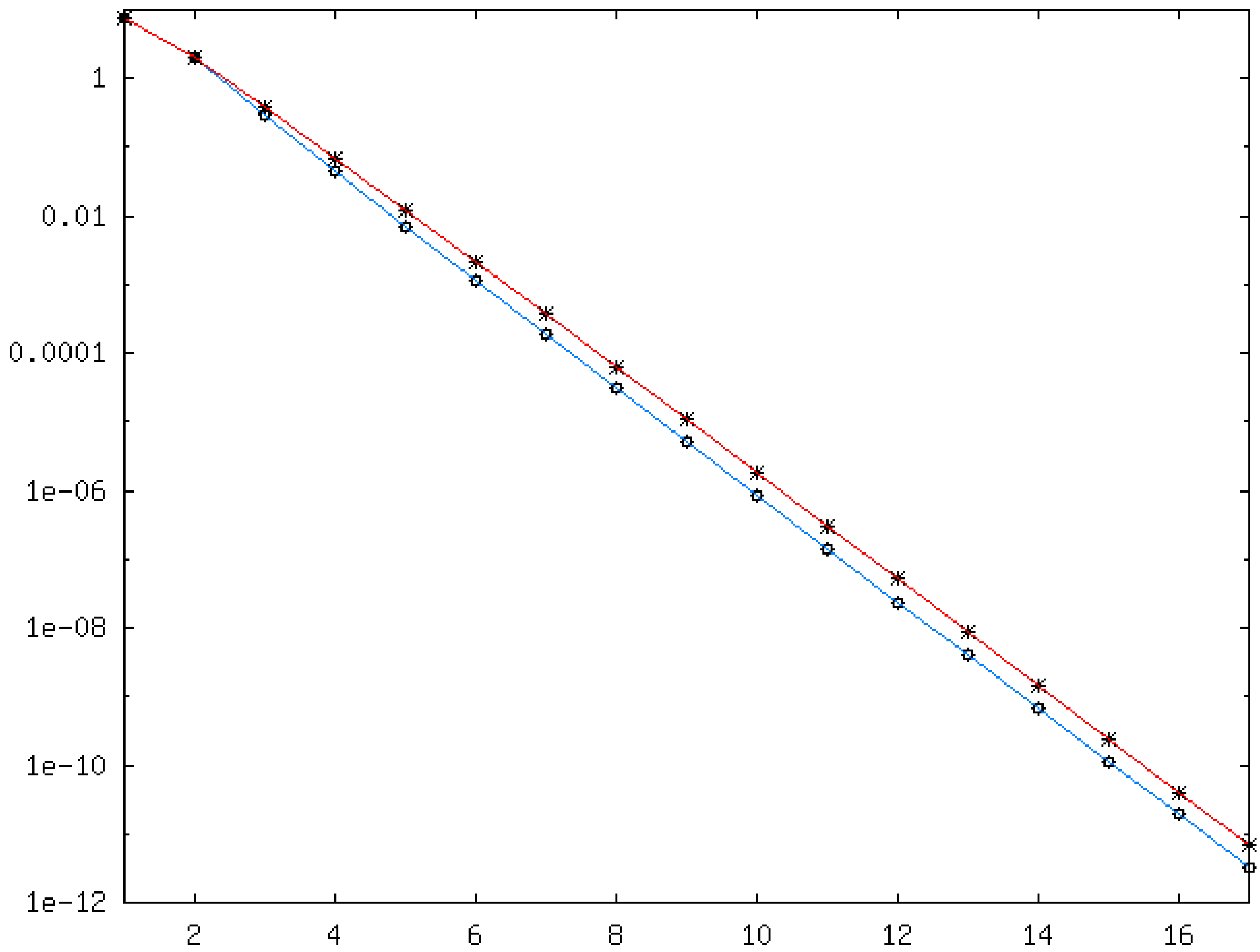
Example:

$$\eta(\beta, r) = \frac{1}{\beta} - \frac{r}{1+\beta} + \frac{r^2}{2+\beta} - \frac{r^3}{3+\beta} + \cdots$$

with $r = 0.94$, $\beta = 0.125$.

Plot the error for Legendre polynomials, scaled to $[-r, 0]$.

Close to, but not as good as, Chebyshev polynomials scaled to $[-1, 0]$.


1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
2
4
6
8
10
12
14
16

Why were the Legendre polynomials so very good?

The first example had

$$a_k = (-1)^k/(k+1) = -\int_0^{-1} t^k \, dt.$$

Thus $w(t) = 1$.

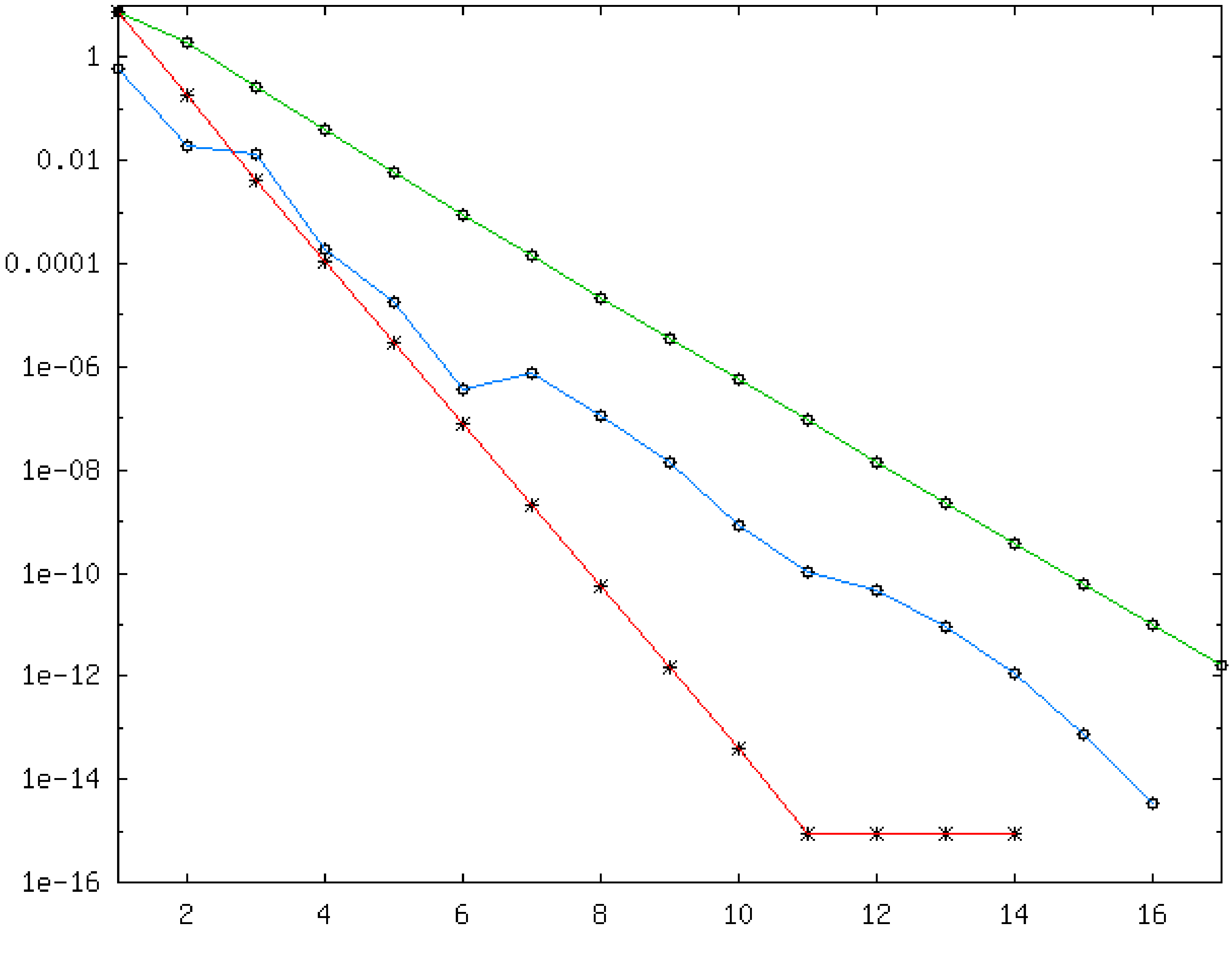
No surprise that $\int_0^{-1} p_n(t)(1-t)^{-1} w(t) \, dt$ is very small when p_n is orthogonal with respect to w!

Perfect polynomials for the second example

Notice that $(k+\beta)^{-1} = \int_0^1 t^k t^{\beta-1}\,dt.$

So a good choice for the acceleration polynomials p_n may be the Jacobi polynomials $P_n^{0,\beta-1}$, shifted to $[0,-r]$.

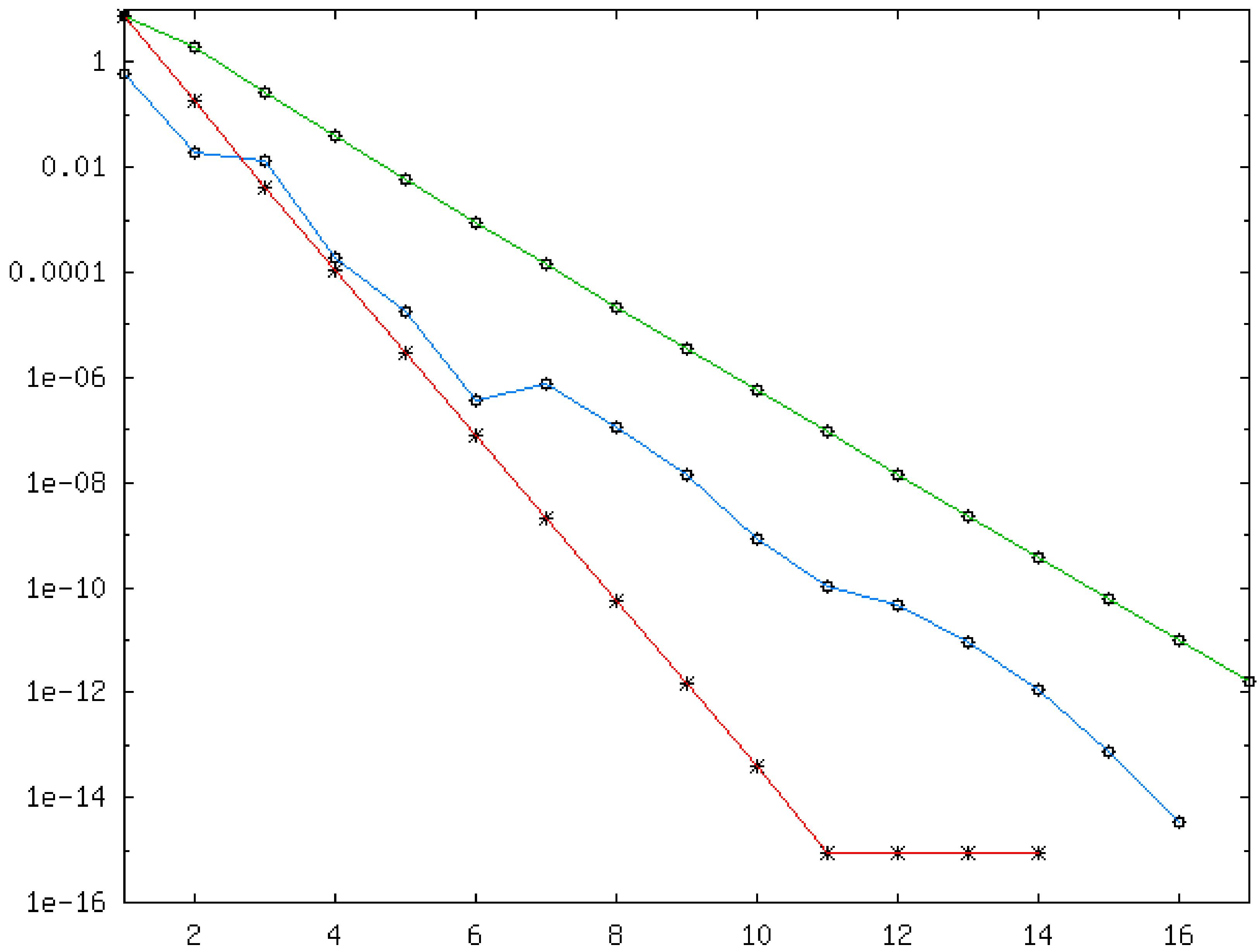
This time, we plot Chebyshev, CRVZ Case A, and Jacobi polynomials, all shifted to $[0,-r]$.

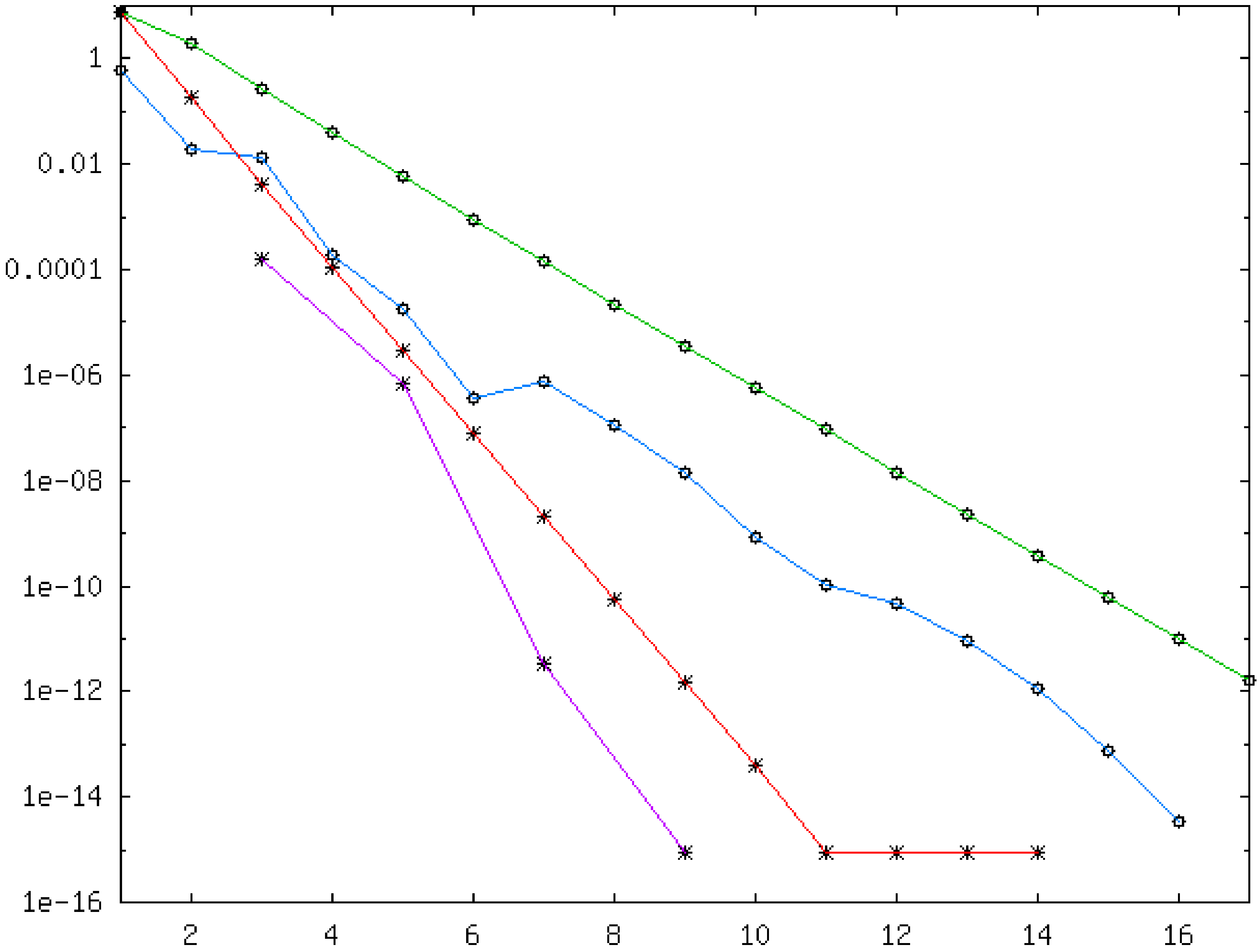

1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
2
4
6
8
10
12
14
16

Very good, but we can do even better

Note that the semilog graph for the Jacobi polynomials is straight.

This implies that the epsilon algorithm can accelerate them further.


1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
2
4
6
8
10
12
14
16


1
0.01
0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
2
4
6
8
10
12
14
16

Summary of numerical results

Number of terms required to reach machine accuracy

Original series	463
Chebyshev	21
CRVZ	17
Jacobi(0,-0.875)	11
Jacobi,epsilon	9

Recursive calculation of the acceleration matrix

Let the polynomials p_n, standardized so that $p_n(1) = 1$, satisfy a three-term recursion of the form

$$c_n p_{n+1} = (x - a_n)p_n - b_n p_{n-1}.$$

Then

$$c_n s_{k,n+1} = s_{k+1,n} - a_n s_{k,n} - b_n s_{k,n-1}.$$

$$\begin{array}{ccccc} s_{k,n-1} & \longrightarrow & s_{k,n} & \longrightarrow & s_{k,n+1} \\ & & s_{k+1,n} & \nearrow & \end{array}$$

Or work with monic polynomials

Put $p_n(x) = \hat{p}_n(x)/\hat{p}_n(1)$, where $\hat{p}_{n+1} = (x - \hat{a}_n)\hat{p}_n - \hat{b}_n\hat{p}_{n-1}$.

- Put $\hat{s}_{k,0} = s_k$ and compute $s_{k,n}$ by a similar recursion
- Also apply the same recursion starting from $w_{k,0} = 1$.
  Only $k = 0$ required since the columns are constant.
- Put $s_{k,n} = \hat{s}_{k,n}/w_{0,n}$.

Conclusion

- Acceleration by orthogonal polynomials can be spectacularly fast.
- A lot of information about the terms of the series is needed to achieve that speed.
  (A solution to the Hausdorff moment problem was available!)
- This will never be a general-purpose method.
  It is a small specialized tool that may sometimes be useful.
- Future work: a priori error bounds?