Avoiding and Computing Singularities in Polynomial Homotopy Continuation

Jan Verschelde

University of Illinois at Chicago
Department of Mathematics, Statistics, and Computer Science
http://www.math.uic.edu/~jan
https://github.com/janverschelde
janv@uic.edu

Symbolic-Numeric Computing seminar, 6 November 2023
CUNY Graduate Center, NYU Courant, NC State University

Supported by NSF grant DMS 1854513 and a 2023 Simons Travel Award.

Outline

1. Introduction
   - polynomial homotopy continuation
   - the ratio theorem of Fabry and Padé approximants
   - analytic continuation and extrapolation methods

2. A Priori Step Size Control
   - linearization to compute Taylor series
   - the Fabry-Hesse-Newton-Padé predictor
   - cost analysis and computational results

3. Extrapolating Towards Singular Solutions
   - logarithmic convergence
   - Richardson extrapolation on two illustrative examples
   - the Rho algorithm and going past the last pole

Avoiding and Computing Singularities

polynomial homotopy continuation

A polynomial homotopy is a family of polynomial systems, where the systems in the family depend on one parameter.

For example, the homotopy

$$h(\mathbf{x}, t) = \gamma(1-t)g(\mathbf{x}) + t\,f(\mathbf{x}) = \mathbf{0}, \quad \text{random } \gamma \in \mathbb{C},$$

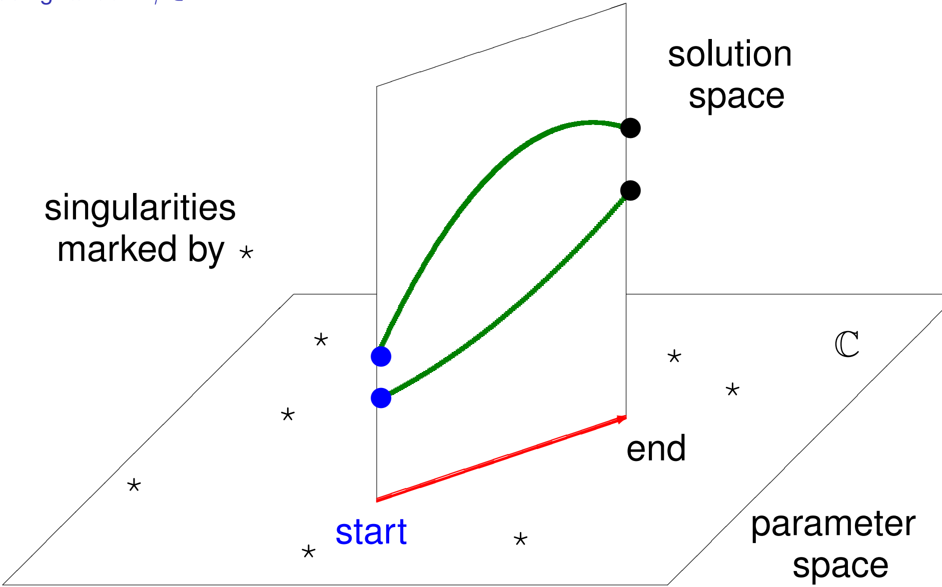
connects

- the target system $f(\mathbf{x}) = \mathbf{0}$, at $t = 1$, to the
- the start system $g(\mathbf{x}) = \mathbf{0}$, at $t = 0$.

Continuation methods apply path tracking algorithms to track solution paths $x(t)$ starting at solutions of $g(\mathbf{x}) = \mathbf{0}$ and ending at the solutions of $f(\mathbf{x}) = \mathbf{0}$.

parameter continuation schematic in $\mathbb{C}$

using random $\gamma \in \mathbb{C}$



step size control

Consider homotopies in one single parameter and assume

1. no singularity on each path, and
2. no diverging paths,

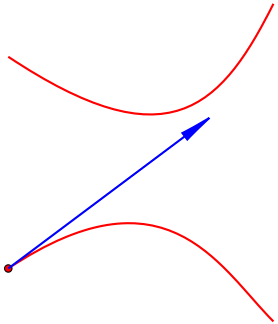
except perhaps at the end of the path.

Problem: determine the step size of the path tracker.

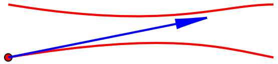
- Too small: inefficient.
- Too large: jump off the path, possibly onto another path.

the path jumping problem

Curves are far apart, with high curvature:



Curves are close to each other, with low curvature:



alternatives to numerical continuation

Numerical continuation applies an adaptive step size control
in a predictor-corrector method with double precision arithmetic.

Alternatives to numerical continuation:

- intervals, parallelotopes, or ball arithmetic
  [Kearfott and Xing, 1994], [Martin, Goldsztejn, Granvilliers, Jermann, 2013], [Lecerf and van der Hoeven, 2016];
- symbolic deformation methods
  [Jeronimo, Matera, Solernó, and Waissbein, 2009], [Hauenstein, Safey El Din, Schost, Vu, 2021];
- certified homotopy tracking
  [Beltrán and Leykin, 2013], [Xu, Burr, and Yap, 2018].

(The above list of references contains just a sample.)

Padé approximants as predictors

1. H. Schwetlick and J. Cleve. **Higher order predictors and adaptive steplength control in path following algorithms.** *SIAM Journal on Numerical Analysis*, 24(6):1382–1393, 1987.
2. A. Trias. **The holomorphic embedding load flow method.** In *2012 IEEE Power and Energy Society General Meeting*, pages 1–8. IEEE, 2012.
3. A. Trias and J. L. Martin. **The holomorphic embedding loadflow method for DC power systems and nonlinear DC circuits.** *IEEE Transactions on Circuits and Systems*, 63(2):322–333, 2016.

The holomorphic embedding load flow method takes the poles of the Padé approximants into account in its step size control.

Avoiding and Computing Singularities

1. Introduction
   - polynomial homotopy continuation
   - the ratio theorem of Fabry and Padé approximants
   - analytic continuation and extrapolation methods

2. A Priori Step Size Control
   - linearization to compute Taylor series
   - the Fabry-Hesse-Newton-Padé predictor
   - cost analysis and computational results

3. Extrapolating Towards Singular Solutions
   - logarithmic convergence
   - Richardson extrapolation on two illustrative examples
   - the Rho algorithm and going past the last pole

detecting nearby singularities

Applying the ratio theorem of Fabry, we can detect singular points based on the coefficients of the Taylor series.

Theorem (the ratio theorem, Fabry 1896)

If for the series $x(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_n t^n + c_{n+1} t^{n+1} + \cdots$*,*
we have $\displaystyle\lim_{n \rightarrow \infty} c_n / c_{n+1} = z$*, then*

- z *is a singular point of the series, and*
- *it lies on the boundary of the circle of convergence of the series.*

Then the radius of this circle is $|z|$*.*

The ratio c_n / c_{n+1} is the pole of Padé approximants of degrees $[n/1]$ (n is the degree of the numerator, with linear denominator).

the ratio theorem of Fabry and Padé approximants

Consider $n = 3$, $x(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3 + c_4 t^4$.

$$[3/1] = \frac{a_0 + a_1 t + a_2 t^2 + a_3 t^3}{1 + b_1 t}$$

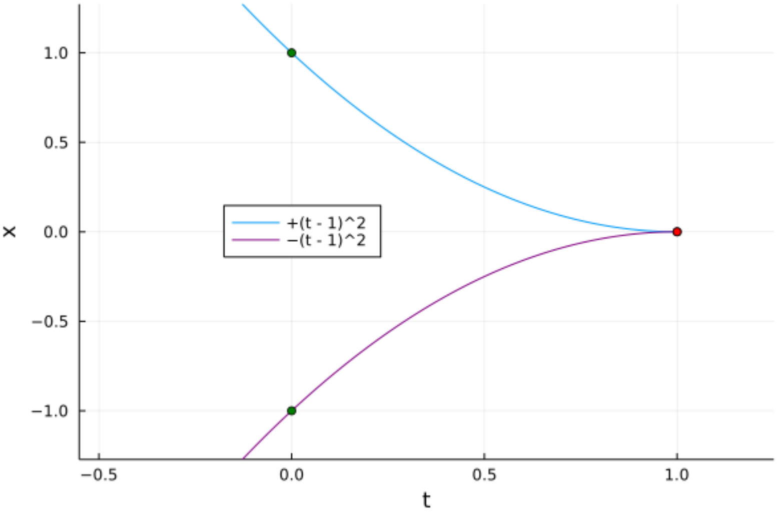
$$\begin{array}{rcl} (c_0 + c_1 t + c_2 t^2 + c_3 t^3 + c_4 t^4)(1 + b_1 t) & = & a_0 + a_1 t + a_2 t^2 + a_3 t^3 \\ c_0 + \quad c_1 t + \quad c_2 t^2 + \quad c_3 t^3 + \quad c_4 t^4 & & \\ + b_1 c_0 t + b_1 c_1 t^2 + b_1 c_2 t^3 + b_1 c_3 t^4 & = & a_0 + a_1 t + a_2 t^2 + a_3 t^3 \end{array}$$

We solve for b_1 in the term for t^4: $c_4 + b_1 c_3 = 0 \Rightarrow b_1 = -c_4/c_3$.

The denominator of $[3/1]$ is $1 - c_4/c_3 t$. The pole of $[3/1]$ is c_3/c_4.

an example not covered by Fabry's theorem

$$h(x,t) = x^2 - (t-1)^4 = (x-(t-1)^2)(x+(t-1)^2) = 0$$



Avoiding and Computing Singularities

1. Introduction
   - polynomial homotopy continuation
   - the ratio theorem of Fabry and Padé approximants
   - analytic continuation and extrapolation methods
2. A Priori Step Size Control
   - linearization to compute Taylor series
   - the Fabry-Hesse-Newton-Padé predictor
   - cost analysis and computational results
3. Extrapolating Towards Singular Solutions
   - logarithmic convergence
   - Richardson extrapolation on two illustrative examples
   - the Rho algorithm and going past the last pole

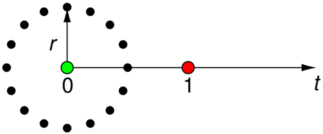
analytic continuation and extrapolation methods

How many terms in the Taylor series are needed
to accurately locate a singularity?

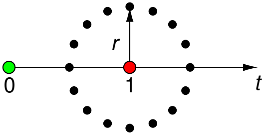
- Peter Wynn. **The rational approximation of functions which are formally defined by a power series expansion.** *Math. Comp.* 14(70): 147–186, 1960.
- Peter Henrici. **An algorithm for analytic continuation.** *J. SIAM Numer. Anal.*, 3(1): 67–78, 1966.
- Claude Brezinski and Michela Redivo Zaglia. **Extrapolation Methods.** Elsevier, 1991.
- Avrim Sidi. **Practical Extrapolation Methods. Theory and Applications.** Cambridge University Press, 2003.
- Lloyd Nicholas Trefethen. **Approximation Theory and Approximation Practice.** SIAM, 2013.

difference with Cauchy integrals

- Use function values around the regular point at 0 to compute the coefficients of the Taylor series:



- Use function values around the singularity at 1 to compute the coefficients of the Laurent series:



In both cases, what is a good step size r?

Avoiding and Computing Singularities

linearization

Working with truncated power series, computing modulo $O(t^d)$, is doing arithmetic over the field of formal series $C[[t]]$.

Linearization: consider $\mathbb{C}^n[[t]]$ instead of $\mathbb{C}[[t]]^n$. Instead of a vector of power series, we consider a power series with vectors as coefficients.

Solve $\mathbf{Ax} = \mathbf{b}$, $\mathbf{A} \in \mathbb{C}^{n \times n}[[t]]$, $\mathbf{b}, \mathbf{x} \in \mathbb{C}^n[[t]]$.

$$
\begin{array}{rcl}
\mathbf{A} & = & A_0 t^a + A_1 t^{a+1} + \cdots , \\
\mathbf{b} & = & \mathbf{b}_0 t^b + \mathbf{b}_1 t^{b+1} + \cdots \\
\mathbf{x} & = & \mathbf{x}_0 t^{b-a} + \mathbf{x}_1 t^{b-a+1} + \cdots
\end{array}
$$

where $A_i \in \mathbb{C}^{n \times n}$ and $\mathbf{b}_i, \mathbf{x}_i \in \mathbb{C}^n$.

block linear algebra

Computing the first d terms of the solution of $\mathbf{Ax} = \mathbf{b}$:

$$\begin{array}{l} \left(A_0 t^a + A_1 t^{a+1} + A_2 t^{a+2} + \cdots + A_d t^{a+d} \right) \\ \cdot \ \left(\mathbf{x}_0 t^{b-a} + \mathbf{x}_1 t^{b-a+1} + \mathbf{x}_2 t^{b-a+2} + \cdots + \mathbf{x}_d t^{b-a+d} \right) \\ = \mathbf{b}_0 t^b + \mathbf{b}_1 t^{b+1} + \mathbf{b}_2 t^{b+2} + \cdots + \mathbf{b}_d t^{b+d}. \end{array}$$

Written in matrix format:

$$\left[\begin{array}{ccccc} A_0 & & & & \\ A_1 & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ A_d & A_{d-1} & A_{d-2} & \cdots & A_0 \end{array} \right] \left[\begin{array}{c} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_d \end{array} \right] = \left[\begin{array}{c} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_d \end{array} \right].$$

If A_0 is regular, then solving $\mathbf{Ax} = \mathbf{b}$ is straightforward.

error analysis of a lower triangular block Toeplitz solver

Solving $(A_0 + A_1 t + A_2 t^2 + \cdots + A_d t^d)(x_0 + x_1 t + x_2 t^2 + \cdots + x_d t^d)$
$= (b_0 + b_1 t + b_2 t^2 + \cdots + b_d t^d)$

leads to a lower triangular block system:

$$\begin{bmatrix} A_0 & & & & \\ A_1 & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ A_d & A_{d-1} & A_{d-2} & \cdots & A_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_d \end{bmatrix}.$$

Cost to solve: $O(n^3) + O(dn^2)$.

Let κ be the condition number of A_0. Let $\|A_0\| = \|x_0\| = 1$, $\|x_d\| \approx \rho^d$.

In our context, $\rho \approx 1/R$, where R is the convergence radius.

If $\|A_d\| \approx \rho^d$, then $\dfrac{\|\Delta x_d\|}{\|x_d\|} \approx \kappa^{d+1} \epsilon_{\rm mach}$, and accuracy is lost.

order of series, accuracy and precision

$$\exp(t) = \sum_{k=0}^{d-1} \frac{t^k}{k!} + O(t^d)$$

k	$1/k!$	recommended precision	eps
7	2.0e-004	double precision okay	2.2e-16
15	7.7e-013	use double doubles	4.9e-32
23	3.9e-023	use double doubles	
31	1.2e-034	use quad doubles	6.1e-64
47	3.9e-060	use octo doubles	4.6e-128
63	5.0e-088	use octo doubles	
95	9.7e-149	need hexa doubles	5.3e-256
127	3.3e-214	need hexa doubles	

GPUs capable of teraflop performance can compensate the cost overhead caused by quad double arithmetic.

Avoiding and Computing Singularities

1. Introduction
   - polynomial homotopy continuation
   - the ratio theorem of Fabry and Padé approximants
   - analytic continuation and extrapolation methods

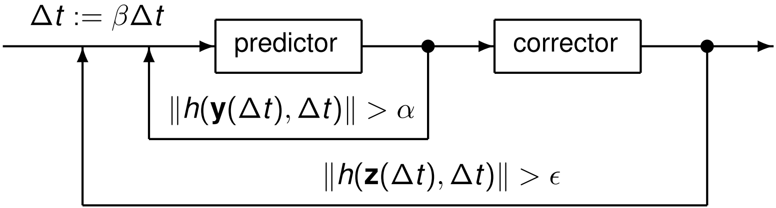
2. A Priori Step Size Control
   - linearization to compute Taylor series
   - the Fabry-Hesse-Newton-Padé predictor
   - cost analysis and computational results

3. Extrapolating Towards Singular Solutions
   - logarithmic convergence
   - Richardson extrapolation on two illustrative examples
   - the Rho algorithm and going past the last pole

a posteriori and a priori step size control

We solve a polynomial system with a homotopy $h(\mathbf{x}, t) = \mathbf{0}$.

An *a posteriori* step size control uses feedback loops.



Extreme choices for α and ϵ (not recommended):

- If $\alpha \leq \epsilon$, then the corrector is not needed.
- If $\alpha = \infty$, then the first feedback loop does never happen.

Setting 0.5 for β cuts the step size Δt in half.

An *a priori* step size control does not need feedback loops.

estimating the distance to the nearest path

Consider a Taylor series expansion of the homotopy at one path, truncated after degree 2, to estimate the distance to the nearest path.

The distance $\|\Delta \mathbf{z}\|$ to the nearest path is estimated by

$$\eta = \frac{2\sigma_n(J_h)}{\sqrt{\sigma_{1,1}^2 + \sigma_{2,1}^2 + \cdots + \sigma_{n,1}^2}} \lesssim \|\Delta \mathbf{z}\|,$$

where

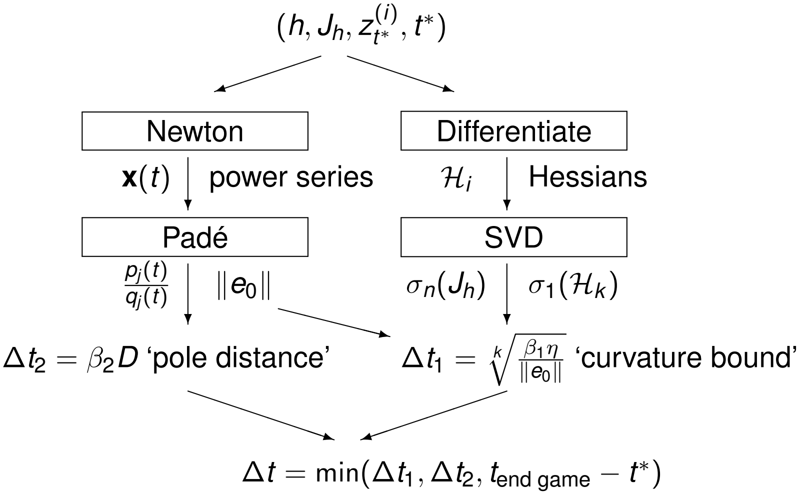
- $\sigma_n(J_h)$ is the smallest singular value of the Jacobian matrix,
- $\sigma_{i,1}$ is the largest singular value of the Hessian matrix at the i-th polynomial in the homotopy h.

With Padé approximants p_i/q_i we compute an estimate for the error e_0:

$$\left\| x(\Delta t) - \left(\frac{p_1(\Delta t)}{q_1(\Delta t)}, \ldots, \frac{p_n(\Delta t)}{q_n(\Delta t)} \right) \right\| \approx \|e_0\| \, |\Delta t|^k,$$

where k is determined by the degrees of the Padé approximants.

schematic summary of a priori step size control



The values β_1 and β_2 are experimentally defined tolerances.

Avoiding and Computing Singularities

cost analysis

For n variables,

- the cost of the linear algebra is $O(n^4)$,
- the cost to differentiate and evaluate n Hessians is 2n times the cost of computing the Jacobian,
- for power series truncated at degree n, the cost overhead factor of Newton's method is $O(n \log(n))$.

Relative to a posteriori step size control,
the cost overhead of a priori step size control is $O(n \log(n))$.

Use parallel computers to offset the cost overhead.

computational results

Available in PHCpack since v2.4.72, released 1 September 2019.
To track a large number of paths, a static workload distribution message passing implementation ran on a 44-core workstation.

Two benchmarks:

1. 1,048,576 paths defined by 20 quadrics, one linear equation, the `katsura-20` benchmark from computational physics. About 66 solutions have a large condition number of about 10^7.

   HOM4PS-2.0para [Li, Tsai, Parallel Computing 2009] reported 4 path jumpings in their runs on `katsura-20`.

2. 1,594,297 paths defined by 13 cubic equations, in `noon-13`, arising in a model of a neural network.

All runs were done in double precision, no path jumpings occurred.

Homogeneous coordinate formulations are important.

the 184,756 paths to all cyclic 11-roots

The cyclic 11-roots problem is a sparse polynomial system

- in 11 variables with 181,756 isolated solutions;
- the mixed volume of the Newton polytopes equals 181,756.

The start system is a system with the same Newton polytopes, but with randomly generated complex coefficients.

A run with `phc` on some difficult path shows:

- Around $t = 0.5$, the coordinates take extreme values, suggesting a diverging path.
- But there is no nearby pole at $t = 0.5$ and `phc -u` can complete without the bound involving the Jacobian and Hessians.

These computations are confirmed with the program `Padé.jl`, Julia code written by Simon Telen and Marc Van Barel.

work still in progress...

Newton polytopes, mixed volumes, and homotopies

- The Newton polytope of a polynomial is the convex hull of the exponents of those monomials appearing with nonzero coefficient.
- A regular subdivision Δ of the polytopes defines homotopies, starting at solutions of systems supported on the faces of Δ.

Theorem (Bernshteĭn's theorems, 1975)

Let P be the Newton polytopes of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$. Let $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$.
An initial form system of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ has faces of P as Newton polytopes.

1. *The mixed volume $V(P) \geq \#$isolated solutions of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$.*
2. *If $V(P) > \#$isolated solutions of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ in ${\mathbb{C}^*}^n$, then $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ has initial form systems with solutions in ${\mathbb{C}^*}^n$.*

- $V(P)$ is a generically sharp upper bound. For systems with fewer solutions, faces of Newton polytopes certify diverging paths.

Avoiding and Computing Singularities

the location problem

At the end of a solution path, we may have a singular solution.

At a singular point, the matrix of all partial derivatives is not full rank.

The *location problem* asks to detect the value of the parameter in the homotopy where a singular point occurs.

How many terms in the Taylor series are needed to solve the location problem?

Taylor series of roots of a polynomial homotopy

Consider the *monomial homotopy*

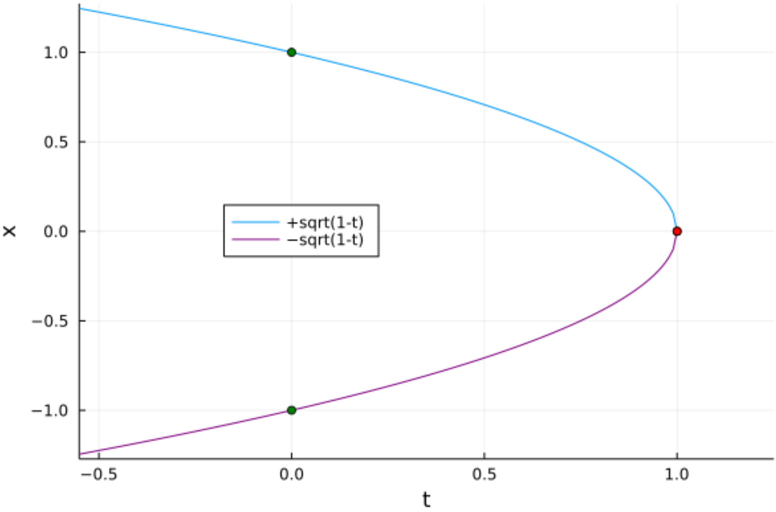
$$h(x,t) = x^2 - 1 + t = 0,$$

where x is the variable and t the parameter.

- At $t = 0$, the solutions are $x = \pm 1$.
- At $t = 1$, we have the double root $x = 0$.

In this test problem, starting at $t = 0$,
we compute 1 as the nearest singularity.

paths defined by $h(x,t) = x^2 - 1 + t = 0$



convergence of the coefficient ratios

Proposition (convergence of the coefficient ratios, CASC 2022)

Assume $x(t)$ is a series which satisfies the conditions of the ratio theorem of Fabry, with a radius of convergence equal to one. Let c_n be the coefficient of t^n in the series, then

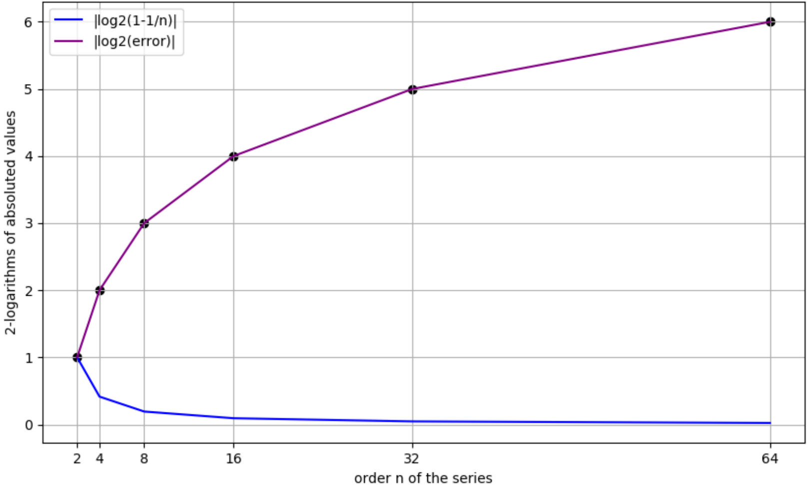
$$\left| 1 - \frac{c_n}{c_{n+1}} \right| \text{ is } O(1/n)$$

for sufficiently large n.

The good and the bad:

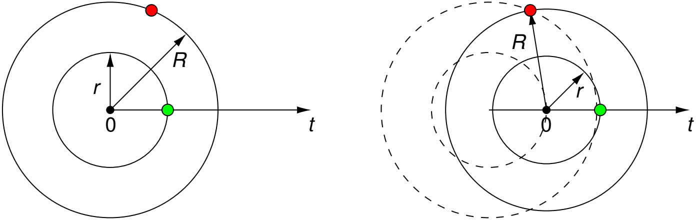
+ It confirms extensive computational experiments: using 8 terms of series are sufficient to avoid a singularity in the step size control.
- The $O(1/n)$ grows very slowly, e.g. $1/64 \approx 0.016$, $1/256 \approx 0.004$.

one extra bit of accuracy after each doubling of n

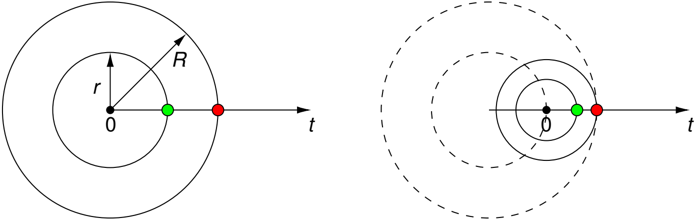


going past versus going towards a singularity

- Going past a singularity (the red dot):

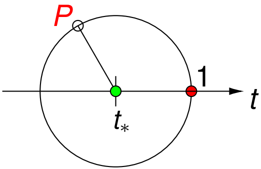


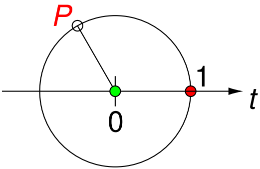
- Going towards a singularity (the red dot):



recentering and scaling the radius of convergence

At the critical distance to the last pole P:





At the right, after recentering the series at $t = 0$ and scaling, the distance to the closest singularity equals 1, as in the monomial homotopies case studies.

Definition (the last pole)

Given a solution path $\mathbf{x}(t)$ of a homotopy $h(\mathbf{x}, t) = \mathbf{0}$, *the last pole P* is a value for t such that

1. the matrix of all partial derivatives of $h(\mathbf{x}(P), P)$ is rank deficient,
2. of all poles, $\text{re}(P)$ is closest to the left of 1.

Avoiding and Computing Singularities

an example of Ojika, 1987

$$\mathbf{f}(x,y) = \left\{ \begin{array}{rcl} x^2 + y - 3 & = & 0 \\ x + 0.125y^2 - 1.5 & = & 0 \end{array} \right.$$

has a triple root at $(1, 2)$. Using a total degree start system with random γ, the t_0 after t_* was found at $t_0 = 0.956$.

After reconditioning, with order $n = 64$, the location is estimated at

$$1.0265192231142901 + 2.9197227799819557\text{E}{-05}\, I$$

and improved with Richardson extrapolation to

$$0.99997295801380 75 + 8.484367218447337\text{E}{-06}\, I,$$

which locates the singularity with an error of 10^{-6}.

Done with sympy 1.4, mpmath 1.1.0, and phcpy 1.1.1 (CASC 2022).

one fourfold cyclic 9-root

The cyclic 9-roots problem (solved by Faugère in 2001)
has several isolated roots of multiplicity four.

With the plain blackbox solver of PHCpack,
one path was selected that ended at one of the fourfold roots.

The t_0 after t_*, the location of the last pole, is $t_0 = 0.99832$.

After reconditioning, with $n = 32$, the convergence radius is

$$1.00000000099639 + 4.319265\text{E}{-09}\, I$$

and confirmed in double double precision.

Because of the close proximity to the singularity,
no extrapolation is necessary in this case.

Avoiding and Computing Singularities

the Rho Algorithm

Let c_n be the nth coefficient of the Taylor series of $x(t) = \sqrt{1-t}$.

Then $f(n) = \dfrac{c_n}{c_{n+1}} = \dfrac{2(n+1)}{2n-1}$.

$f(k)$ converges logarithmically to 1, $f(64)$ has an error of `1.2e-02`.

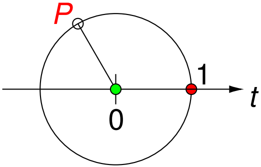
- Richardson extrapolation gives an error of `3.8e-08`,
- improved by repeated Aitken to an error of `2.3e-11`.

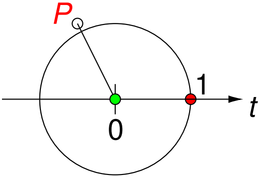
The Rho Algorithm (Wynn, 1955) needs 4 terms for a correct result.

The Rho Algorithm computes even order convergents of Thiele's interpolating continued fraction.
Thiele interpolation recovers $f(n)$ with just 4 points.

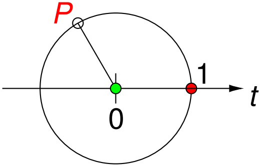
Past the Critical Distance

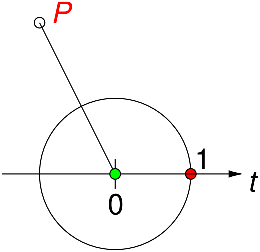
Just past the critical distance, the last pole P is at $-\frac{1}{2} + I$:





More past the critical distance, the last pole P is at $-1 + 2I$:





running the Rho Algorithm

The Rho Algorithm is applied on the sequence c_n/c_{n+1}, $k = 0, 1, \ldots d$,

- c_n is the nth coefficient of the Taylor series of $x(t)$, and
- $x(t) = \sqrt{a(1-t)(P-t)}$, where a is such that $x(0) = 1$.

The smallest error of the rho table for various P and d values:

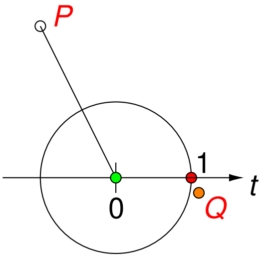
P	$d = 8$	$d = 16$	$d = 32$
$-1/2 + I$	5.0e-01	3.5e-01	1.4e-01
$-1/2 + 2I$	1.7e-01	9.8e-03	2.6e-05
$-1 + 4I$	2.5e-02	6.9e-05	6.3e-09
$-2 + 8I$	3.3e-03	5.3e-07	3.5e-11
$-4 + 16I$	4.1e-04	4.0e-09	2.4e-12

All calculations happened in double precision.
The coefficients c_n were computed with tolerance `1.0e-12`.

what about past 1?

The Rho Algorithm works well when the last pole P is sufficiently far.

But what about a pole Q with $\text{re}(Q) \geq 1$, close to 1?



Example: $P = -4 + 16I$ and $Q = 1 - I/2$.

The Rho Algorithm applied on the ratios of the coefficients of the series of $x(t) = \sqrt{a(1-t)(P-t)(Q-t)}$ (with a such that $x(0) = 1$) gives error `2.4e-02`.

Conclusions and Extrapolations

The application of the ratio theorem of Fabry,
combined with an estimate for the distance to the nearest path,
leads to a robust a priori step size control algorithm,
capable to track millions of solution paths.

The Rho Algorithm to compute the Fabry ratio works well
when applied to problems when there is no other pole near $t = 1$.

Poles Q with $\mathrm{re}(Q) \geq 1$ close to 1 cause problems.
Extrapolate then on the roots of the denominator of a [d/2] Padé
approximant for increasing values of d, the degree of the numerator.

references

- N. Bliss and J. Verschelde. **The method of Gauss–Newton to compute power series solutions of polynomial homotopies.** *Linear Algebra and its Applications*, 542:569–588, 2018.
- S. Telen, M. Van Barel, and J. Verschelde. **A Robust Numerical Path Tracking Algorithm for Polynomial Homotopy Continuation.** *SIAM Journal on Scientific Computing* 42(6):A3610–A3637, 2020.
- J. Verschelde and Kylash Viswanathan. **Locating the Closest Singularity in a Polynomial Homotopy.** In the *Proceedings of the 24th International Workshop on Computer Algebra in Scientific Computing (CASC 2022)*, pages 333–352. Springer-Verlag, 2022.