On Neumaier's Enclosure Method for the Solution of Dissipative ODEs

Markus Neher
Institut für Angewandte Mathematik
Universität Karlsruhe

(joint work with Ken Jackson and Annie Yuk)

Introduction

Smooth IVP: $u' = f(t, u), \;\; u(t_0) \; = \; u_0.$

($u = (u_1, \ldots, u_n), \quad f = (f_1, \ldots, f_n)$).

Moore's enclosure method:

– Automatic computation of Taylor coefficients.

– Interval iteration: For $j := 1, 2, \ldots$:

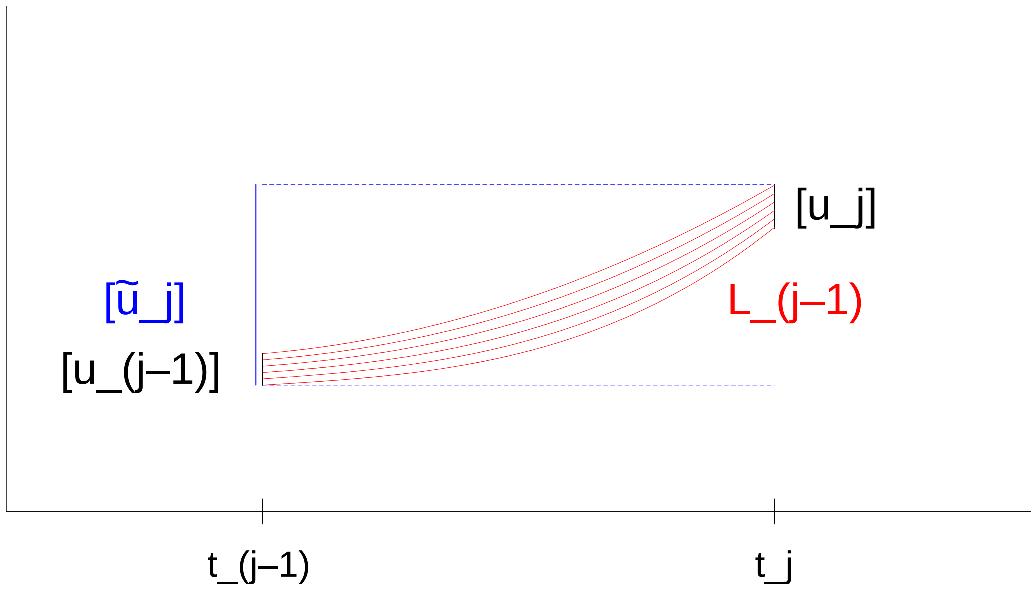
A priori enclosure: $[\widehat{u}_j] \supseteq u(t)$ for all $t \in [t_{j-1}, t_j]$ ("Algorithm I").

Truncation error: $[z_j] := \dfrac{h^{(m+1)}}{(m+1)!} f^{(m)}([t_{j-1}, t_j], [\widehat{u}_j])$.

$$u(t_j) \in [u_j] := [u_{j-1}] + \sum_{k=1}^{m} \frac{h^k}{k!} f^{(k-1)}(t_{j-1}, [u_{j-1}]) + [z_j]$$

("Algorithm II").

Moore's Enclosure Method



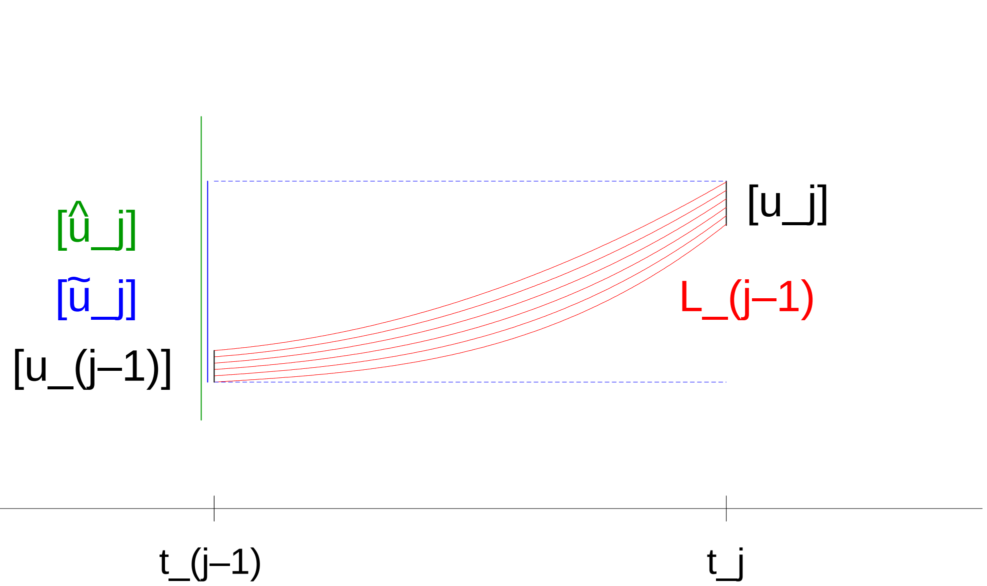
$$t_j := t_{j-1} + h, \quad j = 1, 2, \ldots.$$

$$L_{j-1} := \{u(t) \mid u' = f(t,u),\ u(t_{j-1}) = u_{j-1} \in [u_{j-1}],\ t \in [t_{j-1}, t_j]\}.$$

$[\tilde{u}_j] \in \mathbb{IR}^n$: a priori enclosure such that $\forall u \in L_{j-1}$:

$$u(t) \in [\tilde{u}_j] \ \text{ for all } \ t \in [t_{j-1}, t_j].$$

A priori Enclosure



Fixed point iteration: Determine $[\widehat{u}_j]$ such that

$$[u_{j-1}] + [0,h]\, f([t_{j-1}, t_j], [\widehat{u}_j]) \subseteq [\widehat{u}_j].$$

$\Rightarrow$ Step size restrictions: Explicit Euler steps.

Logarithmic Norm

Let $A \in \mathbb{R}^{n \times n}$ and let $\|.\|$ be a norm in $\mathbb{R}^n$. Then the norm of A is given by

$$\|A\| := \sup_{u \neq 0} \frac{\|Au\|}{\|u\|},$$

and the logarithmic norm of A is defined as

$$\mu(A) := \limsup_{h \to +0} \frac{\|I + hA\| - 1}{h}.$$

Properties of $\mu(A)$

- $\max\{\text{Re}\,\lambda \mid \lambda \in \text{Spec}\,A\} =: \alpha(A) \leq \mu(A) \leq \|A\|$
- $\|.\| = \|.\|_2 \Rightarrow \mu(A) = \alpha(\frac{1}{2}(A + A^T))$,

  $\|.\| = \|.\|_\infty \Rightarrow \mu(A) = \max_i\{A_{ii} + \sum_{j \neq i} |A_{ij}|\}$.

Neumaier's Enclosure Method

Theorem. [Neumaier 1993] *Let $F : [0, \bar{t}] \times \mathbb{R}^n \to \mathbb{R}^n$, let $S \in \mathbb{R}^{n \times n}$ be invertible, and let $p : [0, \bar{t}] \to \mathbb{R}^n$ be an approximate solution of the IVP*

$$(\star) \quad x'(t) = F(t, x(t)), \quad x(0) = x_0,$$

in the sense that

$$\|S^{-1}(x_0 - p(0))\| \le \delta,$$

$$\|S^{-1}(F(t, p(t)) - p'(t))\| \le \varepsilon \quad \textit{for} \;\; t \in [0, \bar{t}].$$

If

$$\mu\left(S^{-1}\frac{\partial F}{\partial x}(t, x)S\right) \le \mu \quad \textit{for} \;\; t \in [0, \bar{t}], \;\; x \in \mathbb{R}^n,$$

then $(\star)$ has a solution $x : [0, \bar{t}] \to \mathbb{R}^n$ satisfying

$$\|S^{-1}((x(t) - p(t))\| \le \varphi(t) := \delta\, e^{\mu t} + \varepsilon\, t\, \textit{exp}_1(\mu t) \quad \textit{for} \;\; t \in [0, \bar{t}],$$

where

$$\textit{exp}_1(\tau) := \begin{cases} (e^\tau - 1)/\tau & \textit{for} \;\; \tau \ne 0, \\ 1 & \textit{for} \;\; \tau = 0. \end{cases}$$

Neumaier's Enclosure Method

- $\|.\| = \|.\|_2$ in the following.
- δ: upper bound on the (weighted) initial error of p.

  $\|S^{-1}(x_0 - p(0))\| \le \delta$ describes an ellipsoid (a ball with radius δ, if $S = I$).
- ε: upper bound on the (weighted) defect of p.
- μ: global upper bound on the logarithmic norm of [a similar matrix to] the Jacobian $\Rightarrow$ most restrictive bound.
- $\|S^{-1}(x(t) - p(t))\| \le \varphi(t)$ describes an ellipsoidal tube with center $p(t)$.
- For $\mu < 0$, $\mu\, t \ll 0$, we have $\varphi(t) \approx \delta\, e^{\mu t} + \varepsilon / |\mu|$.

  Initial error is damped exponentially.

Neumaier's Enclosure Method

- Flexible: Any sufficiently smooth approximate solution that allows rigorous bounding of the defect can be used.

- Good long-term bound if the *uniform dissipation condition*

$$\sup_{0<t<\bar{t},x\in\mathbb{R}^n} \mu\left(S^{-1}\frac{\partial F}{\partial x}(t,x)S\right) < 0$$

  is satisfied.

- Problem: Computation of μ.
  - Choice of S.
  - Estimation of $\mu\left(S^{-1}\frac{\partial F}{\partial x}(t,x)S\right)$ for all $x \in \mathbb{R}^n$.
  - Estimation of $\mu\left(S^{-1}\frac{\partial F}{\partial x}(t,x)S\right)$ in a tube with center $p(t)$ is also difficult.

Choice of S

- Let
$$H := \frac{\partial F}{\partial x}(0, p(0))$$

- If available, use [the real and imaginary parts of] a full set of eigenvectors of H for the columns of S.

- If S is ill-conditioned, replace eigenvectors by independent basis vectors of invariant subspaces of H.

Relative Errors vs. Absolute Errors

- Well-conditioned ellipsoids

  $\Rightarrow$ Absolute errors of all components are of the same size.

  $\Rightarrow$ Their relative errors may differ strongly.

- Componentwise enclosures of AWA, COSY-VI, etc. are usually accurate with respect to relative precision.

  $\Rightarrow$ Scaling of [the components of the solution of] the IVP.

  $\Rightarrow$ Local a posteriori refinement.

Rotating Eigenbases

Wrapping: The ellipsoid

$$\|S_k^{-1}x\| \le r_k$$

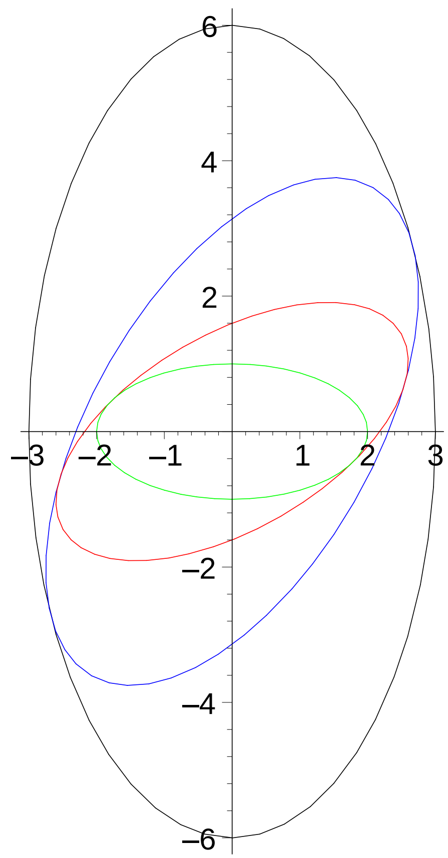
is contained in

$$\|S_{k+1}^{-1}x\| \le r_{k+1}$$

iff

$$r_{k+1} \ge \|S_{k+1}^{-1}S_k\| \, r_k$$

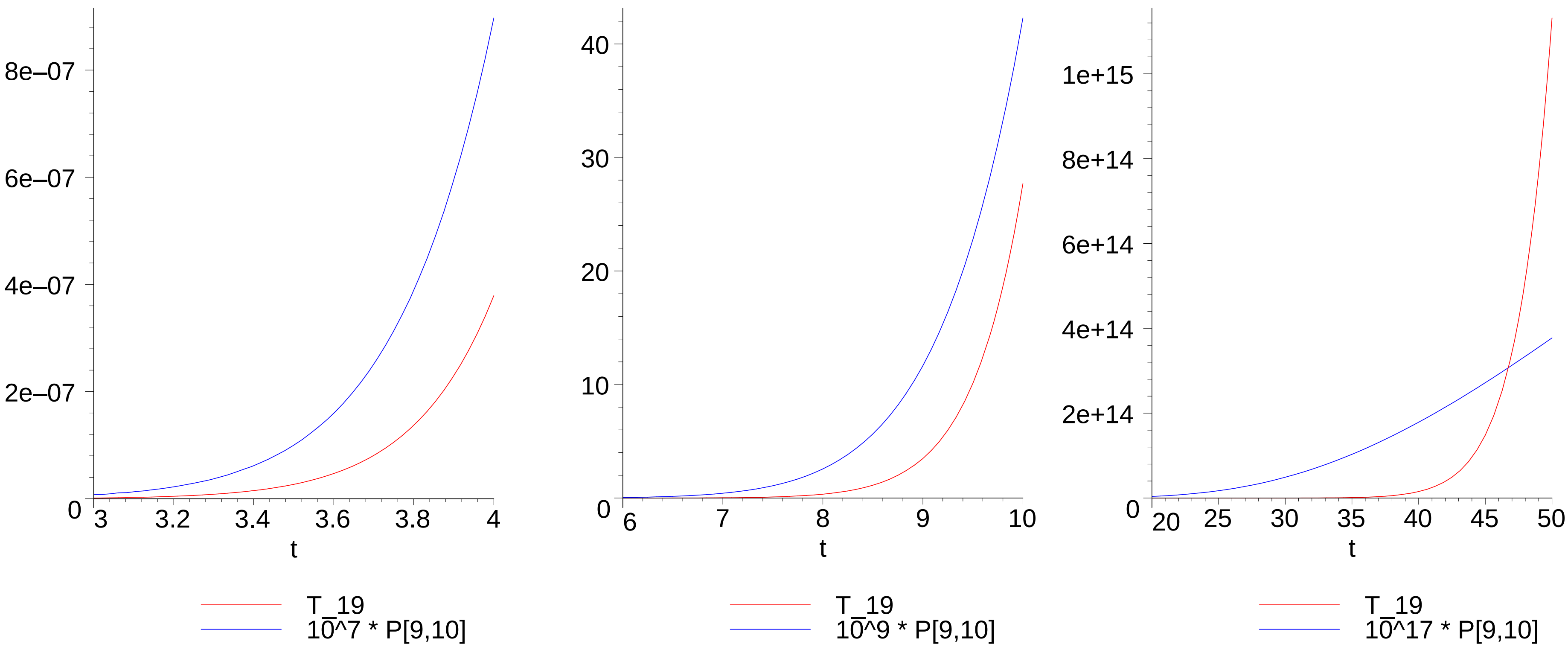
holds.



Implementation I

- Use Taylor polynomial for $p(t)$ on $[t_k, t_k + h_k] =: I_k$.
- Compute S, μ and ε using TMA.
- Problem: Accuracy of the Taylor polynomial.

Taylor Polynomial vs. Padé Approximation of e^{-t}



Implementation II

- Use Padé approximation for $p(t)$ on $[t_k, t_k + h_k] =: I_k$ (requires AD of ODE).

- Assumption: Better approximation on larger domains.

- Disadvantage: Rigorous computation of ε?

  “Padé Model Arithmetic” is infeasible.

- Try the following:
  - Subdivide I_k into subintervals I_{kj}. Compute 1D Taylor model q_j for each component of p, on each subinterval (AD of a 1D rational function).
  - Use q_j to compute ε.
  - S may be different on each subinterval.

- Alternative: Runge-Kutta solver.

Maple Implementation

- Non-rigorous implementation of Neumaier's method for 2D linear IVP

$$x'(t) = A(t)x(t) + b(t)$$

- Three types of approximate solutions:
  - Local Padé approximation of exact global solution, obtained by symbolic computation.
  - Padé approximation of exact local solution, obtained by symbolic computation.
  - Numerical solution with Maple's **dsolve** command.
    - Method = "dverk78", a 7th/8th order Runge-Kutta method.
    - Reliable, but less accurate than AWA or COSY-VI.
    - Numerical differentiation to approximate

$$p'(t) \approx (p(t+h) - p(t-h))/(2h), \quad h = 1.0\mathrm{E} - 10.$$

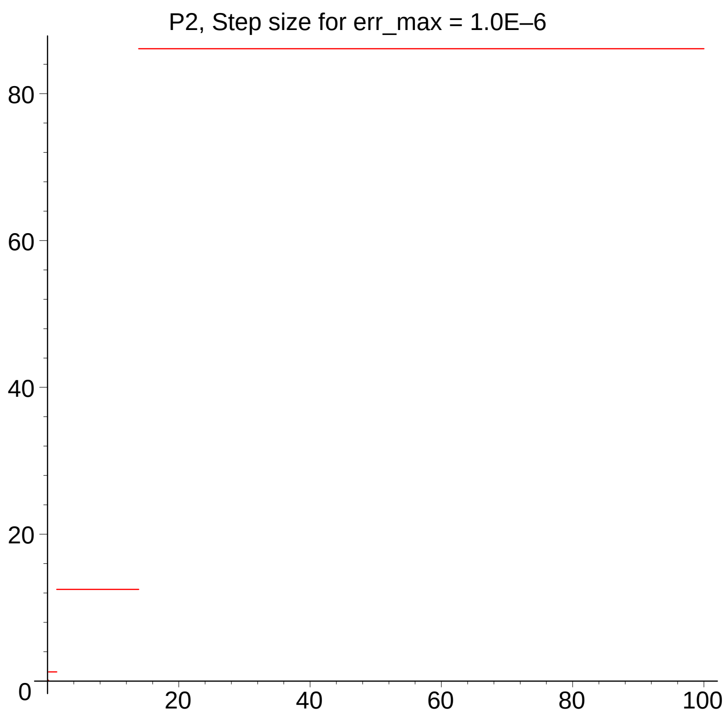
Numerical Examples

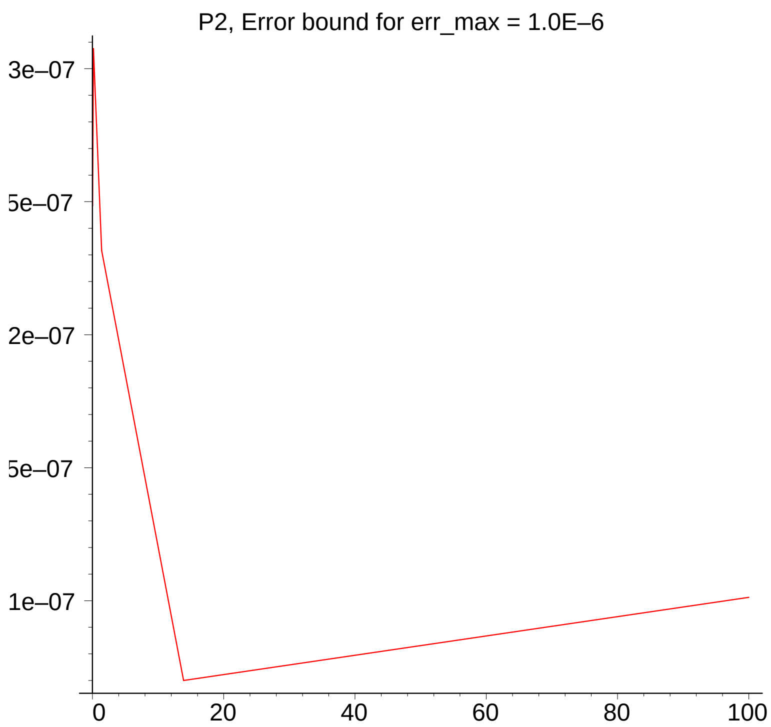
- Symbolic computation of S.
- Symbolic computation of μ. Not computed rigorously if $\mu = \mu(t)$.
- Non-rigorous defect computation by sampling with grid size $h_{k-1}/20$.
- Step size control via prescribed error bound on ε, such that

$$\frac{h_{k-1}}{20} \leq h_k \leq 10\, h_{k-1}.$$

Problem 1

$$A(t) = \begin{pmatrix} -1 & 0 \\ 0 & -1000 \end{pmatrix}, \quad b(t) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$





Problem 1

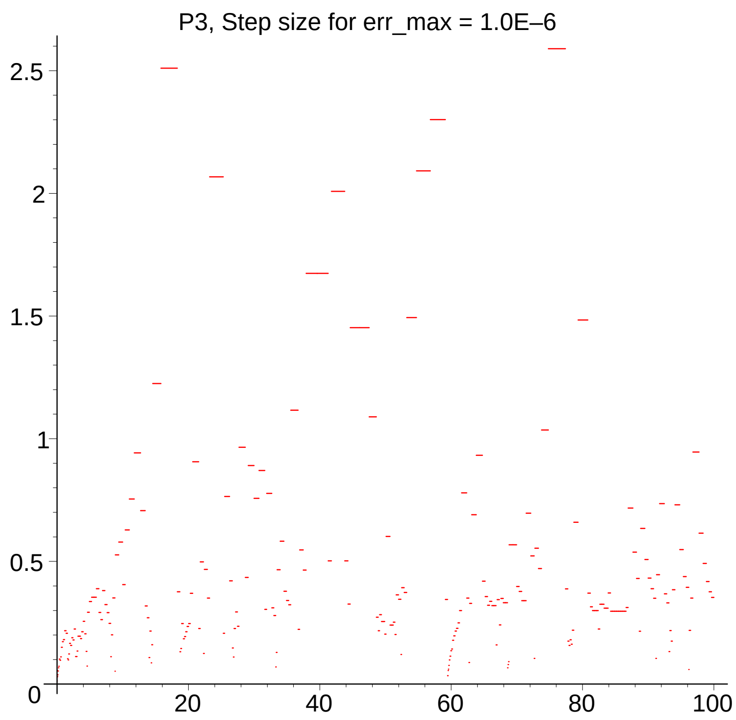
$$A(t) = \begin{pmatrix} -1 & 0 \\ 0 & -1000 \end{pmatrix}, \quad b(t) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

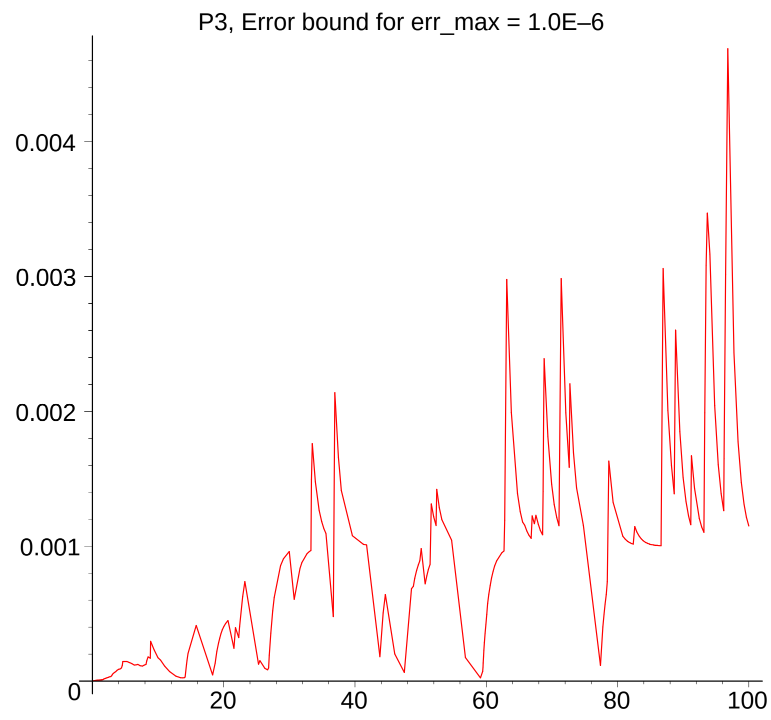
	AWA	COSY-VI	NM (numeric)	exact
t_{max}	100	10	100	—
Steps	100090	11429	5	—
Error	3.8E−16	2.2E−16	1.0E−7	1

Parameters:	All	$\mathrm{err}_{\mathrm{max}}$ = 1.0E−6
	AWA, COSY-VI	Order = 18
	COSY-VI	Diam = 1.0E−12
	COSY-VI	h_{min}: 0.000875 succeeds, 0.001 fails

Problem 2

$$A(t) = \begin{pmatrix} -1 & 0 \\ 0 & -1000 \end{pmatrix}, \quad b(t) = \begin{pmatrix} t \\ t \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$





Problem 2

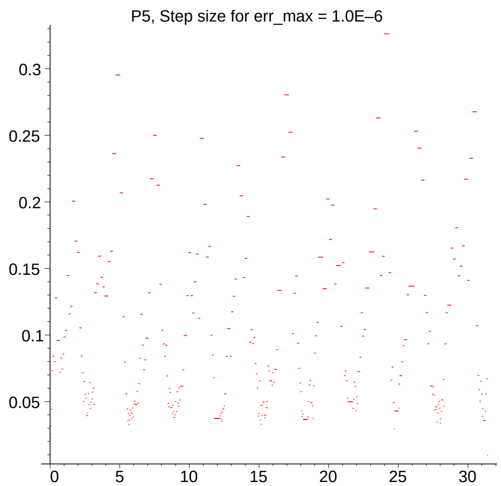
$$A(t) = \begin{pmatrix} -1 & 0 \\ 0 & -1000 \end{pmatrix}, \quad b(t) = \begin{pmatrix} t \\ t \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

	AWA	COSY-VI	NM (15 Digits)	NM (30 Digits)	exact
$t_{\max}$	100	10	100	100	—
Steps	130037	11429	246	3	—
Error	2.0E−11	2.0E−15	1.2E−3	4.2E−9	9.9E+1

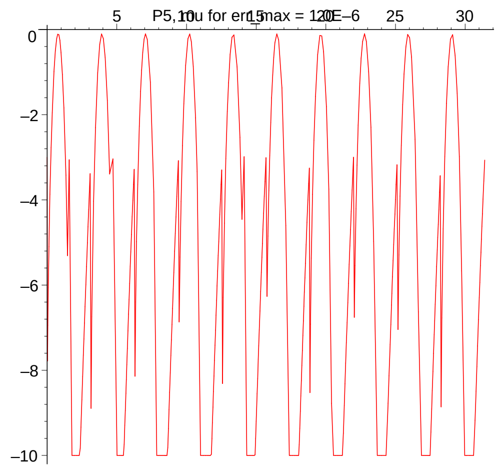
Parameters:	All	$\text{err}_{\max}$ = 1.0E−6
	AWA, COSY-VI	Order = 18
	COSY-VI	Diam = 1.0E−12
	COSY-VI	$h_{\min}$: 0.000875 succeeds, 0.001 fails

Problem 3

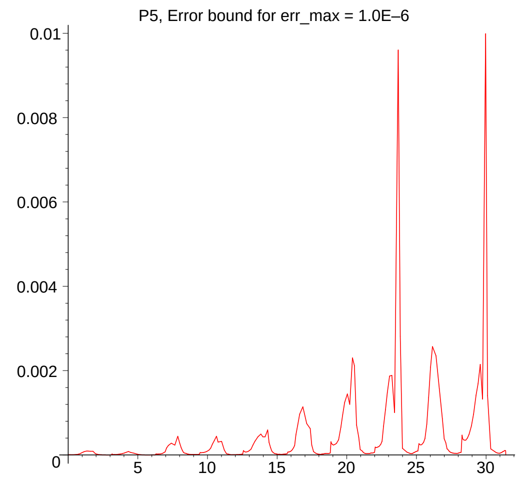
$$A(t) = \begin{pmatrix} -10 & 14\cos t \\ 14\sin t & -10 \end{pmatrix}, \quad b(t) = \begin{pmatrix} \exp(\sin(3t)) \\ t \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



Step size



μ



Error bound

Problem 3

$$A(t) = \begin{pmatrix} -10 & 14\,\cos t \\ 14\,\sin t & -10 \end{pmatrix}, \quad b(t) = \begin{pmatrix} \exp(\sin(3t)) \\ t \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

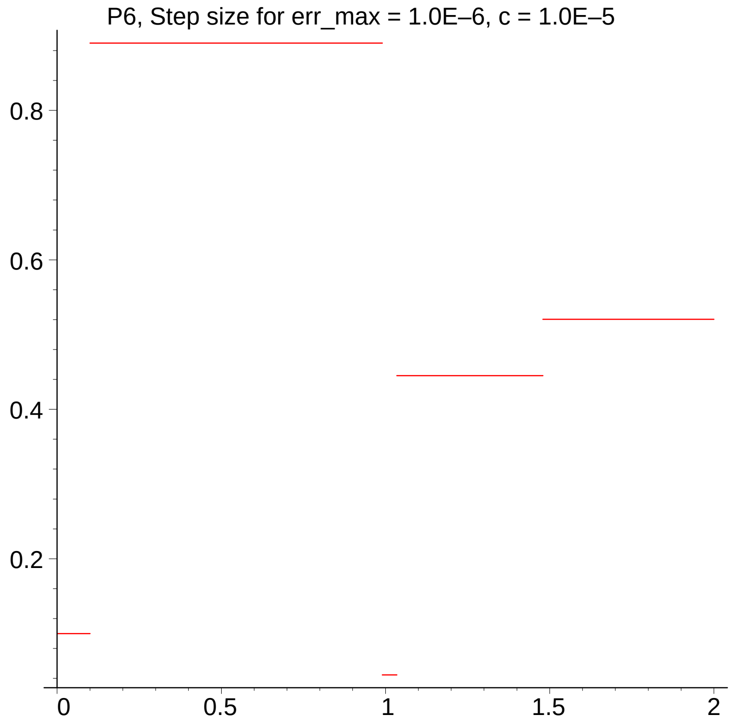
	AWA	COSY-VI	NM (numeric)	exact
Steps	940	807	362	—
Error	7.1E−12	1.4E−9	1.0E−4	3.5E+0

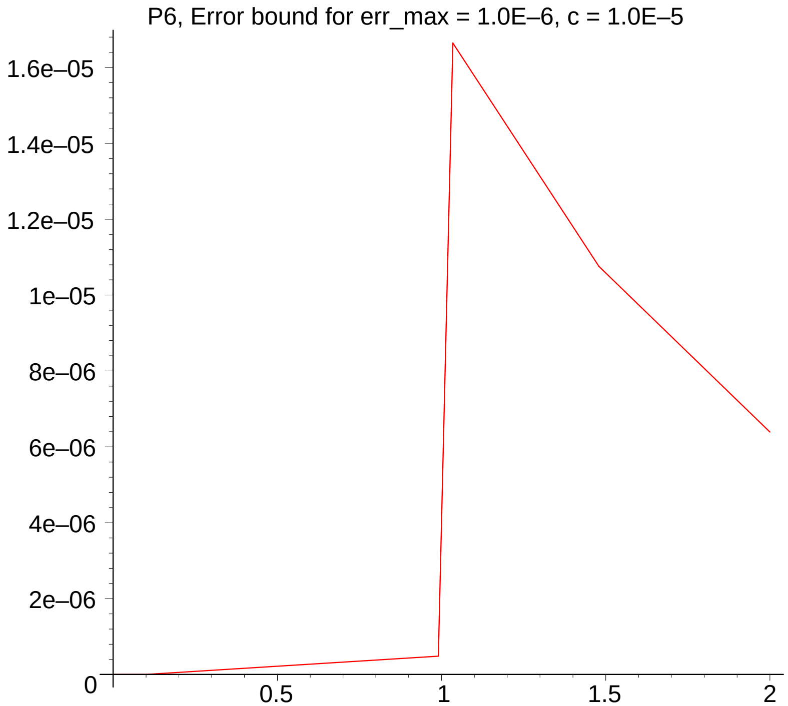
Parameters:	All	$t_{\max}$ = 10π, err$_{\max}$ = 1.0E−6
	AWA, COSY-VI	Order = 18

Problem 4

$$y'(t) = -y(t) + \frac{(t-2)^2 + c^2 - 1}{((t-1)^2 + c^2)^2}, \quad y(0) = 1; \qquad y(t) = \frac{1}{(t-1)^2 + c^2} + \delta e^{-t}$$

c = 1.0E$-$5:





Problem 4

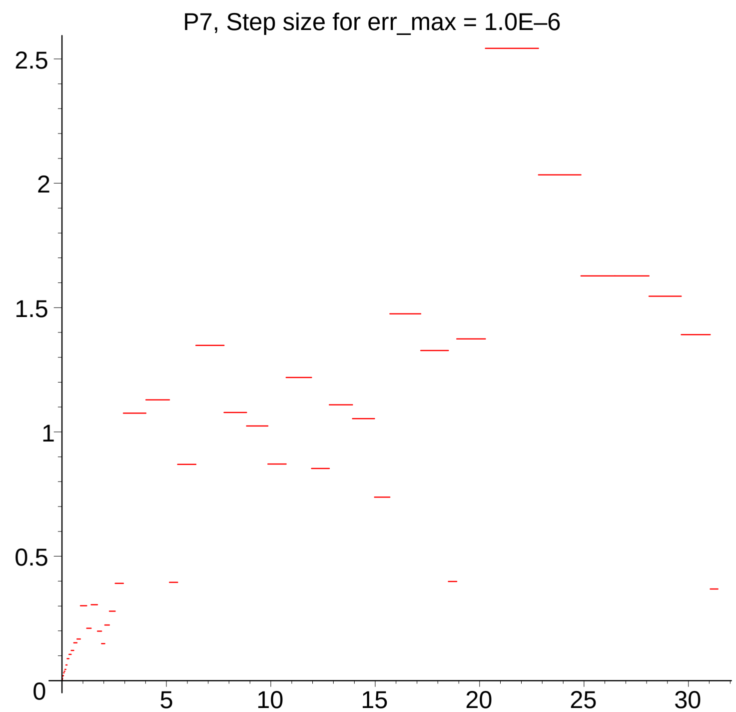
	c	AWA	COSY-VI	NM (30 Digits)
Steps	1E−2	202	89	3
Error	1E−2	5.4E−7	7.7E−6	3.9E−7
Steps	1E−4	423	192	3
Error	1E−4	2.2E−7	1.9E−5	4.1E−7
Steps	1E−5	544	1086	5
Error	1E−5	2.3E−7	3.8E+10	6.4E−6
Steps	1E−5	389	(err_{max}=1.0E−6)	
Error	1E−5	8.2E−1	(err_{max}=1.0E−6)	

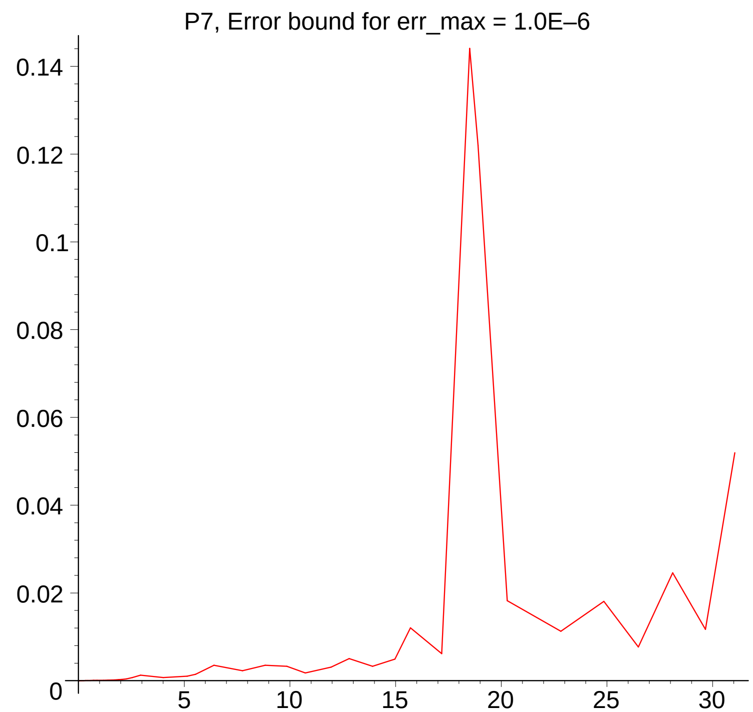
Parameters:	err_{max}	AWA = 1.0E−16, COSY-VI = 1.0E−12, NM = 1.0E−6
	AWA, COSY-VI	Order = 18
	COSY-VI	Diam = 1.0E−12; table value = c^2

Problem 5

$$A(t) = \begin{pmatrix} -1 & -99\frac{\cos t}{\sin^2 t + 0.1} \\ 0 & -100 \end{pmatrix}, \quad b(t) = \begin{pmatrix} \exp(\sin(3t)) \\ t \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Rotating eigensystem: $z_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad z_2 = \begin{pmatrix} \cos t \\ \sin^2 t + 0.1 \end{pmatrix}, \quad \mu \equiv 1$





Problem 5

$$A(t) = \begin{pmatrix} -1 & -99\frac{\cos t}{\sin^2 t + 0.1} \\ 0 & -100 \end{pmatrix}, \quad b(t) = \begin{pmatrix} \exp(\sin(3t)) \\ t \end{pmatrix}, \quad x(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

	AWA	COSY-VI	NM (numeric)	exact
Steps	4382	4181	50	—
Error	1.6E$-$12	1.7E$-$14	5.2E$-$2	-8.3E+1

Parameters:	All	$t_{\max}$ = 10π, err$_{\max}$ = 1.0E$-$6
	AWA, COSY-VI	Order = 18

Conclusion

- Neumaier's method for IVPs.
- Flexible: Any sufficiently smooth approximate solution can be used.
- Good long-term bound for dissipative systems.
- Crucial step: Computation of μ.
- Numerical examples: Large step sizes for dissipative systems.