Finding positively invariant sets of ordinary differential equations using interval global optimization methods

Mihály Csaba Markót
University of Vienna, Vienna, Austria

Zoltán Horváth
Széchenyi István University, Győr, Hungary

SCAN 2014, Würzburg, September 21–26, 2014.

Positive invariance

The considered IVP:

$$\begin{aligned} u'(t) &= f(u(t)), \qquad \forall t \geq 0 \\ u(0) &= u_0, \end{aligned}$$

where $f : \mathbb{R}^N \to \mathbb{R}^N$ is continuously differentiable.

We assume that there exists a unique solution $u(t; u_0)$ for all u_0.

Definition

A set $\mathcal{C} \subseteq \mathbb{R}^N$ is called positively invariant *for f if $\forall u_0 \in \mathcal{C}$ $u(t; u_0) \in \mathcal{C}$ holds $\forall t \geq 0$.*

Positive invariance, cont'd

Theorem

(Explicit Euler condition) A nonempty, convex, closed set $\mathcal{C}$ is positively invariant for f if there exists a real, positive ε constant, such that the containment relation $v + \varepsilon f(v) \in \mathcal{C}$ holds $\forall v \in \mathcal{C}$.

$\Rightarrow$ In this case we say that $\mathcal{C}$ *is positively invariant for f w.r.t. ε*.

Lemma

If $\mathcal{C}$ is nonempty, convex, and compact, and $f \not\equiv 0$, then the set of ε values resulting in positive invariance is $(0, \varepsilon_{max}]$ for some $\varepsilon_{max} \in \mathbb{R}$.

Discrete positive invariance

Definition (restricted to one-step schemes)

Let us given f, u_0, and a $\tau > 0$ stepsize, and let us denote the considered integration scheme (e.g., from the RK family) by F, i.e.

$$u_{n+1} = u_n + \tau F(u_n, \tau, f), \ \ i = 0, \dots$$

A nonempty, convex, closed set $\mathcal{C} \subseteq \mathbb{R}^N$ is called discrete positively invariant *(d.p.i.) for F w.r.t.* the stepsize constant $\tau^* > 0$, *if*

$$\forall \tau \in (0, \tau^*], \ \forall u_0 \in \mathcal{C}, \ \forall u_n \in \mathcal{C} \ \Rightarrow \ u_{n+1} \in \mathcal{C}, n = 0, \dots$$

$\Rightarrow$ If $\mathcal{C}$ is nonempty, convex, and compact, and $F \not\equiv 0$, then the set of τ^* values resulting in d.p.i. is $(0, \tau_{max}]$ for some $\tau_{max} \in \mathbb{R}$.

The considered integration schemes

1. Explicit Euler method:
   - $u_{n+1} = u_n + \tau F(u_n, \tau, f) = u_n + \tau f(u_n)$
   - verifying $u_{n+1} \in \mathcal{C}$ for all $u_n \in \mathcal{C}$ means that
     - $\mathcal{C}$ is d.p.i. for F w.r.t. τ, AND
     - $\mathcal{C}$ is positively invariant for f w.r.t. $\varepsilon = \tau$

2. Explicit forms of Rosenbrock W-methods:
   - s-stage Rosenbrock method:

$$\begin{array}{rcl} u_{n+1} & = & u_n + \sum_{i=1}^{s} b_i k_i, \\ k_i & = & \tau f(u_n + \sum_{j=1}^{i-1} \alpha_{ij} k_j) + \tau Q \sum_{j=1}^{i} \gamma_{ij} k_j, \quad i = 1, \ldots, s, \end{array} \tag{1}$$

     where τ is the stepsize, $Q = f'(u_n)$, and γ_{ij}, α_{ij}, and b_i are the determining coefficients
   - **Rosenbrock W-method**: $Q = f'(u_0)$ or $Q \approx f'(u_0)$.

The considered integration schemes (cont'd)

2. Explicit forms of Rosenbrock W-methods:
   - we need an explicit form $u_{n+1} = u_n + \tau F(u_n, \tau, f)$
   - it is essential to make some transformation on the formulas of the iterative scheme to reduce the interval overestimation as much as possible
   - F can be created from the AMPL model of the (improved) iterative formulas by using the ampl2dag converter of the COCONUT Environment
   - verifying $u_{n+1} \in \mathcal{C}$ for all $u_n \in \mathcal{C}$ means that $\mathcal{C}$ is d.p.i. for F ($=$ the respective s-stage Rosenbrock-W scheme) w.r.t. τ

Interval arithmetic

Notation:

- $\mathbb{I}$: the set of real, compact intervals. **Boldface** symbols are used to denote one- and multidimensional intervals (*boxes*).
- The lower and upper bound of $\boldsymbol{x}$: $\inf(\boldsymbol{x}), \sup(\boldsymbol{x})$.
- The arithmetic operators $\circ \in \{+, -, \cdot, /\}$ and elementary functions $\varphi : \mathbb{R} \to \mathbb{R}$ are defined for interval arguments, so that
  - $\boldsymbol{x} \circ \boldsymbol{y} \supseteq \{x \circ y \,|\, x \in \boldsymbol{x}, y \in \boldsymbol{y}\}$, $\boldsymbol{\varphi}(\boldsymbol{x}) \supseteq \{\varphi(x) \,|\, x \in \boldsymbol{x}\}$,
  - for computer implementations, **the computed enclosures are mathematically correct even in the presence of floating point errors**.
- Compound functions ($\boldsymbol{f} : \mathbb{I}^N \to \mathbb{I}$) can be built just as for the real case (naive interval arithmetic). However, the result of such interval function evaluations usually **overestimate the real range**.

Bound constrained interval global optimization

- The problem setting:

$$\begin{array}{ll} \min & f(x), \\ \text{s.t.} & x \in \boldsymbol{x}_0, \end{array}$$

  where $\boldsymbol{x}_0 \in \mathbb{I}^N$ is the search box, and $f : \mathbb{R}^N \to \mathbb{R}$ is twice continuously differentiable on $\boldsymbol{x}_0$.
- We need **complete and rigorous global search**: compute *mathematically correct* interval enclosures for all global minimizers and the global minimum.
- We employed the `coco_gop_ex` interval B&B solver (Markót and Schichl, COCONUT Environment, Uni. Vienna).

The basic problem considered in the talk

Problem 1

Let us given a twice cont. diff. function $F : \mathbb{R}^N \to \mathbb{R}^N$*, a* **box** $\boldsymbol{v} \in \mathbb{I}^N$*, and a* $\tau \geq 0$*. Decide whether the containment relation*

$$v + \tau F(v) \in \boldsymbol{v}$$

holds for all $v \in \boldsymbol{v}$*.*

Thus, we have to verify a containment property

- **for all points of** $\boldsymbol{v}$
- **with mathematical rigor**.

Solving Problem 1 with interval global optimization

For boxes, the condition $v + \tau F(v) \in \boldsymbol{v}$ can be decomposed into $v_i + \tau F_i(v) \in \boldsymbol{v}_i,\ i = 1, \dots, N$.

Lemma

Let us given F, $\boldsymbol{v} \in \mathbb{I}^N$, and $\tau \geq 0$. Then the following two conditions are equivalent:

(a) $v + \tau F(v) \in \boldsymbol{v}, \quad \forall v \in \boldsymbol{v}$;

(b) **the global minima of the 2n bound constrained GO problems below are all nonnegative:**

$$\min \ v_i + \tau F_i(v) - \inf(\boldsymbol{v}_i), \quad s.t.\ v \in \boldsymbol{v}; \qquad i = 1, \dots, N,$$

$$\min \ \sup(\boldsymbol{v}_i) - (v_i + \tau F_i(v)), \quad s.t.\ v \in \boldsymbol{v}; \qquad i = 1, \dots, N.$$

Further problems to tackle

Problem 2

Given F and $\boldsymbol{v} \in \mathbb{I}^N$, find the maximal $\tau \geq 0$, such that $\boldsymbol{v}$ is d.p.i. for F w.r.t. τ.

This problem is easily solved to a pre-given precision by solving a sequence of Problem 1 (iterative refinement).

Problem 3

Let us given F, $\tau \geq 0$, and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{I}^N$, $\boldsymbol{a} \subseteq \boldsymbol{b}$. Determine a box $\boldsymbol{v}$ such that $\boldsymbol{a} \subseteq \boldsymbol{v} \subseteq \boldsymbol{b}$ and $\boldsymbol{v}$ is d.p.i. for F w.r.t. τ.

- We developed an algorithm that finds the **smallest** such $\boldsymbol{v}$.
- The algorithm is based on iteratively inflating those bounds of $\boldsymbol{a}$ for which the respective GO problem still has a negative global minimum.
- One inflating iteration consists of finding the smallest zero and the smallest fixed point of a one-dimensional function in a closed interval (easy to obtain with interval arithmetic tools).

Test problem #1: the Robertson reaction model

- describes the kinetics of an autocatalytic reaction
- The ODE:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}' = \begin{pmatrix} -k_1 y_1 + k_3 y_2 y_3 \\ k_1 y_1 - k_2 y_2^2 - k_3 y_2 y_3 \\ k_2 y_2^2 \end{pmatrix},$$

  where $y_i, i = 1, 2, 3$ are the concentrations of the components, and k_1, k_2, k_3 are the reaction rate constants.
- In the present setting $k_1 = 0.04, k_2 = 3 \cdot 10^7, k_3 = 10^4$.
- We investigate (discrete) positive invariance in the neighborhood of the equilibrium point $y = (0, 0, 1)$.
- Since $y_1 + y_2 + y_3 = 1$, we transformed the system into a 2-dimensional one by the substitution $y_3 = 1 - y_1 - y_2$.

Robertson model, EE method

D.p.i. sets for the Robertson 2-D model, for the regions $\boldsymbol{a} = [0, 10^q]^2,\ q = -6, \ldots, -12,\ \boldsymbol{b} = [0, 0.5]^2$. For all q, the second column of the table contains the found d.p.i. set for $\tau_{\text{start}} = 10^{-6}$. The third column contains the largest stepsize value for which the found set remains d.p.i.

q	$\mathcal{C} = \boldsymbol{v} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$	$\tau_{\max}$
-6	$[0, 1.999998 \cdot 10^{-1}]\ [0, 10^{-6}]$	$9.970099 \cdot 10^{-5}$
-7	$[0, 2.439024 \cdot 10^{-2}]\ [0, 10^{-7}]$	$9.997001 \cdot 10^{-5}$
-8	$[0, 2.493766 \cdot 10^{-3}]\ [0, 10^{-8}]$	$9.999700 \cdot 10^{-5}$
-9	$[0, 2.499375 \cdot 10^{-4}]\ [0, 10^{-9}]$	$9.999970 \cdot 10^{-5}$
-10	$[0, 2.499938 \cdot 10^{-5}]\ [0, 10^{-10}]$	$9.999996 \cdot 10^{-5}$
-11	$[0, 2.499994 \cdot 10^{-6}]\ [0, 10^{-11}]$	$9.999999 \cdot 10^{-5}$
-12	$[0, 2.499999 \cdot 10^{-7}]\ [0, 10^{-12}]$	$9.999999 \cdot 10^{-5}$

Robertson model, ROS1 scheme ($s = 1$)

- determining coefficients: γ, $\alpha_{21} = 1$, $b_1 = 1$
- for $\gamma = \gamma_- = 1 - \sqrt{2}/2$:

q	$\mathcal{C} = \boldsymbol{v} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$	$\tau_{\max}$
-6	$([0, 0.199998 \cdot 10^{-1}], [0, 10^{-6}])$	$1.411 \cdot 10^{-4}$
-9	$([0, 2.499375 \cdot 10^{-4}], [0, 10^{-9}])$	$1.414 \cdot 10^{-4}$
-12	$([0, 2.499999 \cdot 10^{-7}], [0, 10^{-12}])$	$1.414 \cdot 10^{-4}$

- for $\gamma = \gamma_+ = 1 + \sqrt{2}/2$:

q	$\mathcal{C} = \boldsymbol{v} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$	$\tau_{\max}$
-6	$([0, 0.199992 \cdot 10^{-1}], [0, 10^{-6}])$	$\geq 10^{12}$
-9	$([0, 2.499375 \cdot 10^{-4}], [0, 10^{-9}])$	$\geq 10^{12}$
-12	$([0, 2.499999 \cdot 10^{-7}], [0, 10^{-12}])$	$\geq 10^{12}$

Robertson model, ROS2 scheme ($s = 2$)

- determining coefficients: γ, $\gamma_{21} = -2\gamma$, $\alpha_{21} = 1$, $b_1 = b_2 = 0.5$
- for $\gamma = \gamma_- = 1 - \sqrt{2}/2$:

q	$\mathcal{C} = \boldsymbol{v} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$	$\tau_{\max}$
-6	$([0, 0.199997 \cdot 10^{-1}], [0, 10^{-6}])$	$2.391 \cdot 10^{-4}$
-9	$([0, 2.499375 \cdot 10^{-4}], [0, 10^{-9}])$	$2.414 \cdot 10^{-4}$
-12	$([0, 2.499999 \cdot 10^{-7}], [0, 10^{-12}])$	$2.414 \cdot 10^{-4}$

- for $\gamma = \gamma_+ = 1 + \sqrt{2}/2$:

q	$\mathcal{C} = \boldsymbol{v} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$	$\tau_{\max}$
-6	$([0, 0.199997 \cdot 10^{-1}], [0, 10^{-6}])$	$3.884 \cdot 10^{4}$
-9	$([0, 2.499375 \cdot 10^{-4}], [0, 10^{-9}])$	$4.855 \cdot 10^{7}$
-12	$([0, 2.499999 \cdot 10^{-7}], [0, 10^{-12}])$	$4.856 \cdot 10^{10}$

Conclusion

A current, often used assumption within the research community on positivity methods is that

- the $\tau_{max,S}$ value for a general scheme S is usually somewhere in the order of $\tau_{max,EE}$, or at most
- the ratio of the $\tau_{max,S}$ and $\tau_{max,EE}$ remains approximately constant as the equilibrium is approached

In contrast to that, on the Robertson model, for the 2-stage Rosenbrock-W scheme with $\gamma = \gamma_+$ we found that

- $\tau_{max,ROS2}$ is at least 8 orders of magnitudes larger than $\tau_{max,EE}$
- as the equilibrium point is approched by an order of magnitude, $\tau_{max,ROS2}$ (and $\tau_{max,ROS2}/\tau_{max,EE}$) also grows by an order of magnitude