# On the Blunting Method in the Verified Integration of ODEs

N. S. Nedialkov* K. R. Jackson† M. Neher‡

April 27, 2009

**Abstract**

Verified methods for the integration of initial value problems (IVPs) for ODEs aim at computing guaranteed error bounds for the flow of an ODE and maintaining a low level of overestimation at the same time.

This paper is concerned with one of the sources of overestimation: a matrix-vector product describing a parallelepiped in the phase space. We analyze the so-called blunting method developed by Berz and Makino, which consists of a special choice of the matrix in this product. For the linear model problem

$$u' = Au, \quad u(0) = u_0 \in \boldsymbol{u}_0,$$

where $u \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times m}$, $m \geq 2$, and $\boldsymbol{u}_0$ is a given interval vector, we compare the convergence behavior of the blunting method with that of the well-known QR interval method.

For both methods, the amount of overestimation of the flow of the initial set depends on the spectral radius of some well-defined matrix. We show that under certain conditions, the spectral radii of the matrices that describe the excess propagation in the QR method and in the blunting method, respectively, have the same limits, so that the excess propagation in both methods is similar.

## 1 Verified Integration of IVPs

Consider the set of autonomous IVPs

$$u' = f(u), \quad u(t_0) = u_0 \in \boldsymbol{u}_0, \quad t \in \boldsymbol{t} = [t_0, t_{\text{end}}], \tag{1}$$

on some domain $D \subset \mathbb{R}^m$, where $f \in C^n(D)$, $f : D \to \mathbb{R}^m$. $\boldsymbol{u}_0$ is a given interval vector in the space variables, and $t_{\text{end}} > t_0$ is a given endpoint of the time interval. The ODE

*Department of Computing and Software, McMaster University, Hamilton, ON, L8S 4L7, Canada (nedialk@mcmaster.ca).

†Computer Science Department, University of Toronto, 10 King's College Rd, Toronto, ON, M5S 3G4, Canada (krj@cs.toronto.edu).

‡Institut für Angewandte und Numerische Mathematik, Universität Karlsruhe (TH), 76128 Karlsruhe, Germany (markus.neher@math.uni-karlsruhe.de).

is defined in the traditional way, but the initial value is allowed to vary. In applications, this variability is used for modeling uncertainties in initial conditions.

For each $u_0 \in \boldsymbol{u}_0$, the point IVP

$$u' = f(u), \quad u(t_0) \;=\; u_0$$

has a classical solution, which is denoted by $u(t; u_0)$. In the following, we assume that $u(t; u_0)$ exists and is bounded for all $t \in \boldsymbol{t}$ and for all $u_0 \in \boldsymbol{u}_0$.

Our goal when solving (1) is to calculate bounds on the flow of the interval IVP. For each $t \in \boldsymbol{t}$, we wish to calculate an interval $\boldsymbol{u}(t)$ such that

$$u(t; u_0) \in \boldsymbol{u}(t)$$

holds for all $u_0 \in \boldsymbol{u}_0$. The tube $\boldsymbol{u}(t)$, $t \in \boldsymbol{t}$, then contains all solutions of $u' = f(u)$ that emerge from $\boldsymbol{u}_0$.

In Fig. 1, the flow of an initial set $\boldsymbol{u}_0 \subset \mathbb{R}^m$ in the phase plane is shown for $m = 2$. The direction field of the ODE is described by red arrows. The initial set $\boldsymbol{u}_0 \subset \mathbb{R}^2$ at $t = 0$ shall be defined by the rightmost blue square. As time passes, the initial set moves to the left in the phase plane and is deformed along its path. At some time $t_1 > 0$, the transported flow has changed to the shape of the blue set in the middle of the picture, whereas the leftmost blue set shows the flow at some time $t_2 > t_1$.

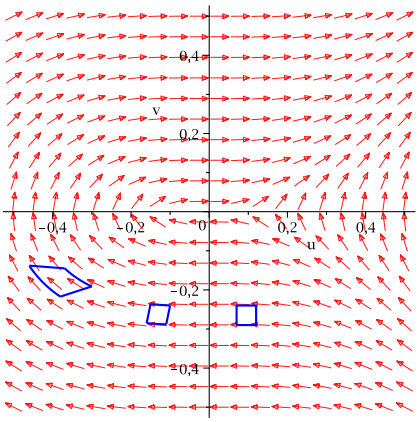


Figure 1: Flow of an initial set.

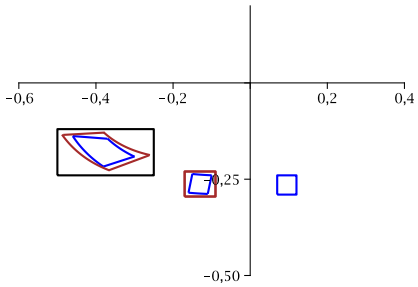


Figure 2: Interval enclosures for the flow.

A verified integration method for ordinary IVPs computes enclosure sets for the flow at distinct points $t_j$, $j = 0, \ldots, N$ [6, 8, 9, 10, 11, 14, 21, 23, 24, 26, 32, 33, 34, 35, 36, 39, 40, 41, 42, 43, 44, 45, 48, 49, 50]. In Fig. 2, the first two time steps of such an enclosure method are illustrated.

At $t = t_1$, the brown rectangle $\boldsymbol{u}_1$ is an enclosure set for the flow of $\boldsymbol{u}_0$ at $t = t_1$. Typically, $\boldsymbol{u}_1$ overestimates the flow. This wrapping of the flow into $\boldsymbol{u}_1$ introduces overestimation in any verified integration method. Instead of the exact flow at $t = t_1$, the enclosure set $\boldsymbol{u}_1$ is used as the initial set in the second integration step,

$$u' = f(u), \quad t \in [t_1, t_2], \quad u(t_1) \in \boldsymbol{u}_1.$$

Then, at $t = t_2$, the flow of $\boldsymbol{u}_1$ is enclosed into some set $\boldsymbol{u}_2$, depicted by the black square in Fig. 2. We observe that the flow of the original initial set $\boldsymbol{u}_0$ at $t = t_2$ is contained in $\boldsymbol{u}_2$, but that there is already much overestimation in this picture.

We assume that the reader is familiar with real interval arithmetic [1, 17, 36, 37, 46]. In this paper, the set of compact real intervals is denoted by

$$\mathbb{IR} = \{\, \boldsymbol{x} = [\underline{x}, \overline{x}] \mid \underline{x}, \overline{x} \in \mathbb{R},\ \underline{x} \le \overline{x} \,\}.$$

Intervals are denoted by boldface. Lower-case letters are used for denoting scalars and vectors. Matrices are denoted by upper-case letters. The *midpoint* and the *width* of an interval $\boldsymbol{x}$ are denoted by $\mathrm{m}(\boldsymbol{x}) := (\overline{x} + \underline{x})/2$ and $\mathrm{w}(\boldsymbol{x}) := \overline{x} - \underline{x}$, respectively. The set of all $m$-dimensional interval vectors is denoted by $\mathbb{IR}^m$.

This paper is organized as follows. The general interval Taylor method for ordinary IVPs is outlined in Section 2. In section 3, several interval methods for a linear model problem are compared. The blunting method for the model problem is analyzed in Section 4. Some numerical examples are presented.

# 2 Enclosure Methods for IVPs Based on Taylor Expansion

In this section, we review interval Taylor methods for the interval IVP (1),

$$u' = f(u), \quad u(t_0) = u_0 \in \boldsymbol{u}_0, \tag{2}$$

on some domain $D \subset \mathbb{R}^m$, where $f \in C^n(D)$, $f : D \to \mathbb{R}^m$, $t \in [t_0, t_{\mathrm{end}}]$ for some $t_{\mathrm{end}} > t_0$.

The basic idea of Taylor methods is to approximate each solution of (2) by its Taylor polynomial:

$$u(t) \approx u(t_0) + \sum_{k=1}^{n-1} \frac{u^{(k)}}{k!}(t - t_0)^k. \tag{3}$$

The derivatives in (3) are computed by differentiating $u' = f(u)$ with respect to $t$. From the chain rule, we obtain

$$u'' = \left(\frac{\partial}{\partial u} f(u)\right) u' = f(u) \frac{\partial}{\partial u} f(u).$$

Higher order derivatives of $u$ are computed similarly. Letting

$$f^{[0]}(u) = u, \quad f^{[k]}(u) = \frac{1}{k}\left(\frac{\partial f^{[k-1]}}{\partial u} f\right)(u) \quad \text{for} \quad k \ge 1,$$

values of these derivatives can be computed recursively from by automatic differentiation from initial values $u_0$ or $\boldsymbol{u}_0$, respectively, and the well-known formulas for the Taylor coefficients of sums, products, quotients, and the standard functions [15, 36, 47].

In verified integration, it is not sufficient to compute an approximation to the solution. Taylor's theorem, however, opens a door for computing not only the approximate solution in (3), but also a rigorous error bound with it. According to Taylor's theorem,

$$u(t) = u(t_0) + \sum_{k=1}^{n-1} f^{[k]}(t - t_0)^k + \frac{u^{(n)}(\tau)}{n!}(t - t_0)^n$$

for some $\tau \in (t_0, t)$. Now suppose that for some $t_1 > t_0$, componentwise bounds $\underline{u}, \overline{u}$ fulfilling

$$\underline{u} \leq u(t) \leq \overline{u} \text{ for all } t \in [t_0, t_1]$$

are known. Then the interval arithmetic evaluation of $u^{(n)}([\underline{u}, \overline{u}])$ yields rigorous lower and upper bounds for the truncation error in (3), and thus also lower and upper bounds for the solutions of (2), for $t \in [t_0, t_1]$.

Suitable values $\underline{u}, \overline{u}$ can be computed by Picard iteration [36]. Using the Picard-Lindelöf operator and the Banach fixed-point theorem, one can show that if $h_0$ and $\boldsymbol{v}_0 \supseteq \boldsymbol{u}_0$ satisfy

$$\tilde{\boldsymbol{u}}_0 = \boldsymbol{u}_0 + [0, h_0] f(\boldsymbol{v}_0) \subseteq \boldsymbol{v}_0,$$

then (2) has a unique solution $u(t; u_0)$ for each $u_0 \in \boldsymbol{u}_0$. Moreover, $u(t; u_0) \in \tilde{\boldsymbol{u}}_0$ holds for all $t \in [t_0, t_0 + h_0]$ and all $u_0 \in \boldsymbol{u}_0$. The interval bound $\tilde{\boldsymbol{u}}_0$ is called a coarse enclosure of the flow of $\boldsymbol{u}_0$, for $t \in [t_0, t_1]$, where $t_1 = t_0 + h_0$.

Picard iteration is easily performed, but it has the disadvantage that if often restricts the step size of verified integration methods. More advanced schemes for computing coarse enclosures are discussed in [12, 26].

Once a coarse enclosure is available, it can be used for obtaining improved bounds on the flow by some kind of refinement procedure. For example, inserting $\tilde{\boldsymbol{u}}_0$ in the formula for the truncation error of the Taylor series in (3), the inclusion

$$u(t_1) \in \boldsymbol{u}_1 = \boldsymbol{u}_0 + \sum_{k=1}^{n-1} h_0^k f^{[k]}(\boldsymbol{u}_0) + h_0^n f^{[n]}(\tilde{\boldsymbol{u}}_0) \tag{4}$$

is obtained. Usually, $\boldsymbol{u}_1$ is a tighter enclosure for the flow of (2) at $t = t_1$ than the coarse enclosure $\tilde{\boldsymbol{u}}_0$. Nevertheless, the refinement procedure (4) is not very useful for practical computations, because the latter formula is width increasing in time. We have

$$\mathrm{w}(\boldsymbol{u}_1) = \mathrm{w}(\boldsymbol{u}_0) + w\left(\sum_{k=1}^{n-1} h_0^k f^{[k]}(\boldsymbol{u}_0) + h_0^n f^{[n]}(\tilde{\boldsymbol{u}}_0)\right) \geq \mathrm{w}(\boldsymbol{u}_0).$$

Tighter enclosures can be achieved by applying the mean value form to $f^{[k]}$ in (4). For any $\widehat{u}_0 \in \boldsymbol{u}_0$ it holds that

$$u(t_1) \in \widehat{u}_0 + \sum_{k=1}^{n-1} h_0^k f^{[k]}(\widehat{u}_0) + h_0^n f^{[n]}(\tilde{\boldsymbol{u}}_0) + \left(I + \sum_{k=1}^{n-1} h_0^k J\big(f^{[k]}(\boldsymbol{u}_0)\big)\right)(\boldsymbol{u}_0 - \widehat{u}_0), \tag{5}$$

where $I$ denotes the identity matrix and $J\big(f^{[k]}\big)$ is the Jacobian of $f^{[i]}$. As was mentioned before, values of these Jacobians can be computed via automatic differentiation.

Letting

$$\boldsymbol{S}_0 = I + \sum_{k=1}^{n-1} h_0^k J\big(f^{[k]}(\boldsymbol{u}_0)\big), \quad \boldsymbol{z}_1 = h_0^n f^{[n]}(\tilde{\boldsymbol{u}}_0),$$

we obtain an enclosure for the flow of (2):

$$u(t_1; u_0) \in \boldsymbol{u}_1 = \widehat{u}_0 + \sum_{k=1}^{n-1} h_0^k f^{[k]}(\widehat{u}_0) + \boldsymbol{z}_1 + \boldsymbol{S}_0(\boldsymbol{u}_0 - \widehat{u}_0). \tag{6}$$

This iteration has been introduced by Moore [34, 35, 36]. It is called the direct method. Its main disadvantage is that the interval vector $\boldsymbol{S}_0(\boldsymbol{u}_0 - \widehat{u}_0)$ may significantly overestimate the set

$$\mathcal{S} = \{S_0(u_0 - \widehat{u}_0) \mid S_0 \in \boldsymbol{S}_0,\ u_0 \in \boldsymbol{u}_0\}.$$

To obtain a valid enclosure method for the flow of (2), it would be sufficient to compute a bound on $\mathcal{S}$. Wrapping $\mathcal{S}$ into $\boldsymbol{S}_0(\boldsymbol{u}_0 - \widehat{u}_0)$ may result in significant overestimation of $\mathcal{S}$, which also causes overestimation in the initial set of the next integration step (cf. Section 1).

The so-called *wrapping effect* was first observed by Moore in 1965 [35]; a recent analysis has been given by Lohner [25]. To reduce wrapping in interval methods for the verified integration of IVPs, several schemes have been proposed in the literature [35, 14, 21, 23, 24]. On a general scale, interval arithmetic has also been extended with symbolic computations [13, 18, 27] to diminish the wrapping-effect. Starting in the 1990s, Berz and his group developed a rigorous multivariate Taylor arithmetic [2, 27, 30]. Taylor model arithmetic has been defined in [2, 5, 27, 29, 30]. A software implementation of Taylor model arithmetic has been developed by Berz and Makino [3, 28] in the COSY Infinity package [4]. Using COSY Infinity, Taylor models have been applied with success to a variety of problems, including global optimization [38], verified multidimensional integration [7], and the verified solution of ODEs and DAEs [6, 16].

The major source of overestimation in (6) is in the matrix vector product $\boldsymbol{S}_0(\boldsymbol{u}_0 - \widehat{u}_0)$. The amount of overestimation is effectively reduced if this product is not performed explicitly, but if $\mathcal{S}$ is propagated in the verified integration scheme as a parallelepiped instead of an interval. This is achieved by introducing local coordinate systems described by matrices $B_j$. Letting $\widehat{u}_0 = \mathrm{m}(\boldsymbol{u}_0)$, $\boldsymbol{r}_0 = \boldsymbol{u}_0 - \widehat{u}_0$, $B_0 = I$, choosing some nonsingular matrices $B_j$, and performing a sequence of integration steps, formula (6) is turned into the interval iteration

$$\left.\begin{aligned} \widehat{u}_j &= \widehat{u}_{j-1} + \sum_{k=1}^{n-1} h_{j-1}^k f^{[k]}(\widehat{u}_{j-1}) + \mathrm{m}(\boldsymbol{z}_j), \\ \boldsymbol{u}_j &= \widehat{u}_{j-1} + \sum_{k=1}^{n-1} h_{j-1}^k f^{[k]}(\widehat{u}_{j-1}) + \boldsymbol{z}_j + (\boldsymbol{S}_{j-1}B_{j-1})\boldsymbol{r}_{j-1}, \end{aligned}\right\} \quad j = 1, 2, \ldots, \tag{7}$$

in which $\widehat{u}_j$ is some approximate point solution for the central IVP in (2), $\boldsymbol{z}_j$ denotes the (usually small) local error of the $j$-th integration step, and the global error $\boldsymbol{r}_j$ is propagated according to

$$\boldsymbol{r}_j = \left(B_j^{-1}(\boldsymbol{S}_{j-1}B_{j-1})\right)\boldsymbol{r}_{j-1} + B_j^{-1}(\boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j)). \tag{8}$$

There is freedom of choice for the matrices $B_j$. In Moore's direct method, $B_j = I$ is used in all steps. It is known that this method only performs well in the absence of rotations. The parallelepiped method [14, 23] is obtained by choosing $B_j = \mathrm{m}(\boldsymbol{S}_{j-1}B_{j-1})$. This method is only suitable for pure rotations, because otherwise the matrices $B_j$ tend to become singular after a few steps.

The hitherto most successful general scheme for reducing the wrapping effect is Lohner's QR method [23, 24]. In the QR method, $B_j$ is chosen as the orthogonal matrix $Q$ in the QR factorization of $\mathrm{m}(\boldsymbol{S}_{j-1}B_{j-1})$. Thus, the matrices $B_j$ are well-conditioned

for all time and their inversion is achieved by transposition. In [40] it has been shown that for linear autonomous systems, the bound on the global error of the QR method is not much bigger than the global error of the corresponding approximate Taylor method.

As an alternative to QR factorization, one could modify the matrices $B_j$ of the parallelepiped method such that their condition number must not exceed some suitable bound. This variant of the parallelepiped method was already mentioned by Lohner in [23], but not implemented in his thesis. Berz and Makino [8, 32] developed this idea into a practical algorithm which they named blunting method.

The purpose of this paper is the analysis of the blunting method and a comparison of its asymptotic behavior with the asymptotic behavior of the QR method. The authors have not been successful in developing an analysis for nonlinear or time-dependent linear systems. We only consider the model problem of a linear system with constant coefficients. Therefore, our analysis does not cover the full story, but on the other hand, methods that perform badly for the model problem are likely to fail for other problems as well.

# 3 Taylor Methods for a Linear Model Problem

We now consider the linear model problem

$$\begin{aligned} u' &= Au, \qquad (A \in \mathbb{R}^{m\times m},\ m \geq 2) \\ u(0) &= u_0 \in \boldsymbol{u}_0. \end{aligned} \tag{9}$$

The interval Taylor method with constant order $n$ and stepsize $h$ for solving (9) consists of the iteration

$$u_j := Tu_{j-1} + z_j, \quad j = 1, 2, \ldots,$$

where

$$T = T_{n-1}(hA) = \sum_{\nu=0}^{n-1} \frac{(hA)^\nu}{\nu!},$$

and $z_j$ denotes the local error.

We enclose the flow of (9) by

$$\begin{aligned} u(t_j; \boldsymbol{u}_0) &= \{u(t_j; u_0) \mid u_0 \in \boldsymbol{u}_0\} \\ &\subseteq \{u_j + S_j w + B_j r \mid w \in \boldsymbol{u}_0 - \mathrm{m}(\boldsymbol{u}_0),\ r \in \boldsymbol{r}_j\}, \end{aligned}$$

where $u_j, w, r \in \mathbb{R}^m$, $\boldsymbol{r}_j \in \mathbb{IR}^m$, $S_j, B_j \in \mathbb{R}^{m\times m}$ and $B_j$ must be nonsingular. The set

$$\{u_j + S_j w \mid w \in \boldsymbol{u}_0 - \mathrm{m}(\boldsymbol{u}_0)\}$$

approximates the flow $\{u(t_j; u_0) \mid u_0 \in \boldsymbol{u}_0\}$ at $t = t_j$. The global error is guaranteed to be contained in the set

$$\{B_j r \mid r \in \boldsymbol{r}_j\}.$$

We first formulate the interval methods mentioned in the previous section for the model problem and then develop an error analysis for the blunting method.

## 3.1 Propagation of the Global Error

In all interval methods that we study in this paper, the iteration is started with $u_0 = \mathrm{m}(\boldsymbol{u}_0)$, $\boldsymbol{r}_0 = 0$, $S_0 = B_0 = I$. The propagation of the global error is described by the iteration (8),

$$\boldsymbol{r}_j = (B_j^{-1} T B_{j-1})\boldsymbol{r}_{j-1} + B_j^{-1}\big(\boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j)\big).$$

For minimal overestimation, $B_j$ should be chosen such that the inclusion

$$\{TB_{j-1}r + z \mid r \in \boldsymbol{r}_{j-1},\, z \in \boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j)\} \subseteq \{B_j r \mid r \in \boldsymbol{r}_j\} \tag{10}$$

is as a tight as possible.

There is a straightforward geometric interpretation of the inclusion (10). The sum of a parallelepiped (containing the accumulated global error of the previous integration steps) and an interval box (containing the local error of the current integration step) is enclosed into a new parallelepiped, as illustrated by Fig. (3.1).

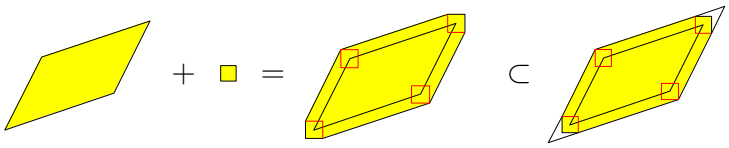

Figure 3: Propagation of the global error.

## 3.2 Wrapping Effect

In Moore's direct method [34, 35, 36], $B_j = I$ is used for all $j$, so that the global error is propagated according to

$$\boldsymbol{r}_j = T\boldsymbol{r}_{j-1} + \boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j).$$

The chosen coordinates are optimal for the local error, which is simply added to the global error. However, if $T$ describes a rotation in phase space, then the volume of $T\boldsymbol{r}_{j-1}$ can be much larger than the volume of $\boldsymbol{r}_{j-1}$.

The parallelepiped method [14, 23] uses $B_j = TB_{j-1}$. Thus,

$$\boldsymbol{r}_j = \boldsymbol{r}_{j-1} + T^{-j}\big(\boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j)\big).$$

The chosen coordinates are optimal for the global error. The local errors are usually small compared to the global error, so that a certain amount of overestimation of the local error may be tolerable. In the parallelepiped method, the local coordinate system is suitable for the local error, provided that $\mathrm{cond}(T^j)$ is not too big. However, in the presence of shear, $T^j$ becomes singular for $j \to \infty$. In practice, the matrices $B_j$ often become ill-conditioned, resulting in large overestimations and finally in the breakdown of the method.

Lohner [23, 24] proposed the following modification of the parallelepiped method. The QR factorization of $TB_{j-1}$ is computed, $Q_jR_j = TB_{j-1}$, and $B_j$ is chosen as the orthogonal matrix $Q_j$:

$$\boldsymbol{r}_j = R_j\boldsymbol{r}_{j-1} + Q_j^T\big(\boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j)\big).$$

The coordinate system used in the QR method is not optimal for the global error, but rotations occurring in $T$ are accounted for. To minimize overestimation in the QR method, the columns of $TB_{j-1}$ must be ordered according to decreasing length in the orthogonalization process [23, 24].

Since $Q_j$ is an orthogonal matrix, $B_j$ is well-conditioned for all $j$. So far, the QR method has been the most successful method for simultaneously treating rotation, contraction, and shear. However, there is a famous example by Kühn [21] showing that even the QR method may fail exponentially.

In his thesis [23], Lohner mentions an alternative to the QR method, by modifying the matrices $B_j$ in the parallelepiped method such that their condition number must not exceed some suitable bound. Lohner never pursued this any further, but Berz and Makino [8, 32] developed a practical algorithm which they named the blunting method. They also compute the QR factorization of $TB_{j-1}$, but then form a linear combination of $Q_j$ and $TB_{j-1}$:

$$\boldsymbol{r}_j = (B_j^{-1}TB_{j-1})\boldsymbol{r}_{j-1} + B_j^{-1}\big(\boldsymbol{z}_j - \mathrm{m}(\boldsymbol{z}_j)\big), \quad B_j = TB_{j-1} + \varepsilon Q_j.$$

Here, $\varepsilon$ is a diagonal matrix of small positive entries $\varepsilon_{ll}$, $l = 1, 2, \ldots, m$.

In numerical experiments, it has been observed that for certain choices of the blunting factors $\varepsilon_{ll}$, the local coordinates are sometimes better suited for the global error than the coordinates used in the QR method. Like the QR method, the blunting method can simultaneously handle rotation, contraction, and shear, but the overestimation depends on the blunting factors.

# 4 The Blunting Method

In the QR method, we perform a QR factorization $TB_{j-1} = Q_jR_j$ and select $B_j = Q_j$. For the model problem, this choice leads to the simultaneous iteration $Q_jR_j = TQ_{j-1}$ [40]. In the blunting method, we select $B_j$ from

$$\begin{aligned} &TB_{j-1} = Q_j^*R_j^* \quad \text{(QR factorization of } TB_j\text{)},\\ &\widehat{B}_j = TB_{j-1}D_j + Q_jG_j,\\ &B_j = \widehat{B}_jF_j. \end{aligned}$$

$D_j$ is a diagonal matrix such that $TB_{j-1}D_j$ is normalized (each column is of length 1 in $\|\cdot\|_2$). $G_j$ is a diagonal matrix with positive blunting factors [8, 32]. $F_j$ is a diagonal matrix such that $B_j = \widehat{B}_jF_j$ is normalized. Letting

$$V_j = (R_j^*D_j + G_j)F_j,$$

obtain the simultaneous iteration

$$Q_j^*(R_j^*V_{j-1}^{-1}) = TQ_{j-1}^*.$$

Choosing $Q_0 = Q_0^* = I$ (where $I$ is the identity matrix), the relations between the respective matrices in the QR and in the blunting methods are

$$Q_j = Q_j^*, \quad R_j = R_j^*V_{j-1}^{-1}.$$

We are interested in the excess propagation in

$$(B_j^{-1}TB_{j-1})\boldsymbol{r}_{j-1}.$$

In the QR method, we have $Q_j^TTQ_{j-1} = R_j$, whereas the blunting method reads

$$B_j^{-1}TB_{j-1} = V_j^{-1}Q_j^TQ_jR_j^* = V_j^{-1}R_j^* = V_j^{-1}R_jV_{j-1}.$$

Since the width of $\boldsymbol{r}_j$ is

$$\mathrm{w}(\boldsymbol{r}_j) = |B_j^{-1}TB_{j-1}|\mathrm{w}(\boldsymbol{r}_j) + |B_j^{-1}|w(\boldsymbol{z}_j),$$

the excess propagation depends on the spectral radius of a certain matrix. In the QR method, this matrix is [40]

$$H_{j,i} = |Q_j^{-1}TQ_{j-1}|\,|Q_{j-1}^{-1}TQ_{j-2}|\cdots|Q_{i+1}^{-1}TQ_i| = |R_j|\,|R_{j-1}|\cdots|R_{i+1}|,$$

whereas in the blunting method, it is

$$\begin{aligned} P_{j,i} &= |B_j^{-1}TB_{j-1}|\,|B_{j-1}^{-1}TB_{j-2}|\cdots|B_{i+1}^{-1}TB_i| \\ &= |V_j^{-1}R^jV_{j-1}|\,|V_{j-1}^{-1}R_{j-1}V_{j-2}|\cdots|V_{i+1}^{-1}R_{i+1}V_i|. \end{aligned}$$

Now we consider the case when $T$ has eigenvalues $\lambda_i$ of distinct magnitudes, i.e. $|\lambda_1| > |\lambda_2| > \cdots|\lambda_n| > 0$. In the QR method, the diagonal of $|R_j|$ converges to $(|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|)$, as $j \to \infty$. As $j$ becomes sufficiently large, $v_{j,i}$ behaves like

$$\big(\,|\lambda_1|^{j-i+1}, |\lambda_2|^{j-i+1}, \ldots, |\lambda_n|^{j-i+1}\,\big).$$

The diagonal of $P_{j,i}$ behaves like

$$\big(\,|\lambda_1|^{j-i+1}, \alpha_{i,j}^{(2)}|\lambda_2|^{j-i+1}, \ldots, \alpha_{i,j}^{(n)}|\lambda_n|^{j-i+1}\,\big),$$

where

$$\alpha_{i,j}^{(k)} = \frac{(V_i)_{k,k}}{(V_j)_{k,k}}.$$

Now let $\alpha_k := \alpha_{1,j}^{(k)}$. Similar error propagations hold in the QR and in the blunting methods, if $\alpha_k \approx 1$ for $k = 2, \ldots, n$. In the case of a scalar blunting factor $\varepsilon$, explicit bounds for $\alpha_k$ are obtained, namely

$$\frac{\varepsilon}{1+\varepsilon} \le \alpha_k \le 1 + \frac{1}{\varepsilon}.$$

For $\varepsilon = 10^{-3}$, we have

$$10^{-3} \approx \frac{10^{-3}}{1+10^{-3}} \le \alpha_k \le 1001,$$

whereas for $\varepsilon = 1$, the estimation is

$$0.5 \le \alpha_k \le 2.$$

At least for $\varepsilon \approx 1$, the QR and in the blunting methods have the same asymptotic behavior.

### 4.1 Condition numbers of $B_j$

The matrices $B_j$ in the blunting method are nonsingular, but little is known about their condition numbers. Using scalar blunting factors again, from $B_j = Q_j V_j$ for some orthogonal matrix $Q_j$, we obtain

$$\left\|V_j^{-1}\right\|_{1,\infty} \leq \frac{(1+1/\varepsilon)^{n-1}}{\varepsilon}.$$

For $\varepsilon = 10^{-3}$, this bound reads

$$\left\|V_j^{-1}\right\|_{1,\infty} \leq 10^3(10^3+1)^{n-1},$$

which is huge even for moderate values of $n$. For $\varepsilon = 1$, the more practical bound

$$\left\|V_j^{-1}\right\|_{1,\infty} \leq 2^{n-1}$$

is obtained, which ensures that the matrices $B_j$ are well-conditioned for $\varepsilon \approx 1$.

### 4.2 Remarks

The blunting method and the QR method both work well for our simple test problem $u' = Au$ (assuming $T$ has eigenvalues of distinct magnitude). The suggested blunting factor $10^{-3}$ [8, 32] may not always be a good choice. It seems reasonable for $u' = Au$ to start with small blunting factors and increase them as $j$ increases. The optimal choice of the blunting factors is still an open question.

At present, we do not know how to analyze the case that $T$ has two or more eigenvalues of the same magnitude (this includes the important case the $T$ has a pair of complex conjugate eigenvalues) or how to accommodate permutations in the QR and blunting methods. Also, the quality of upper bounds for cond( $B_j$ ), which sometimes are huge, is not known. These questions will be the subject of future research.

### 4.3 Example: Influence of the blunting factors

We conclude this paper with a numerical experiment. For selected scalar blunting factors, we show the resulting overestimation of a linear model problem. If the blunting factors are small, then the blunting method resembles the parallelepiped method. For larger blunting factors, the blunting method behaves very similar to the QR method.

In Figures 4-6, the dashed lines belong to the parallelepiped method with resetting. The solid lines denote the enclosure sets computed with the QR method. The enclosure sets of the blunting method are displayed in red.

## Conclusion

We have studied the blunting method in the verified integration of ordinary IVPs. For a linear model problem, we showed that the blunting method performs asymptotically as the QR method, if the matrix $T$ in the Taylor method has eigenvalues of distinct magnitude. An extension of our analysis to linear problems with eigenvalues of the same magnitude and to nonlinear problems will be the subject of future research.

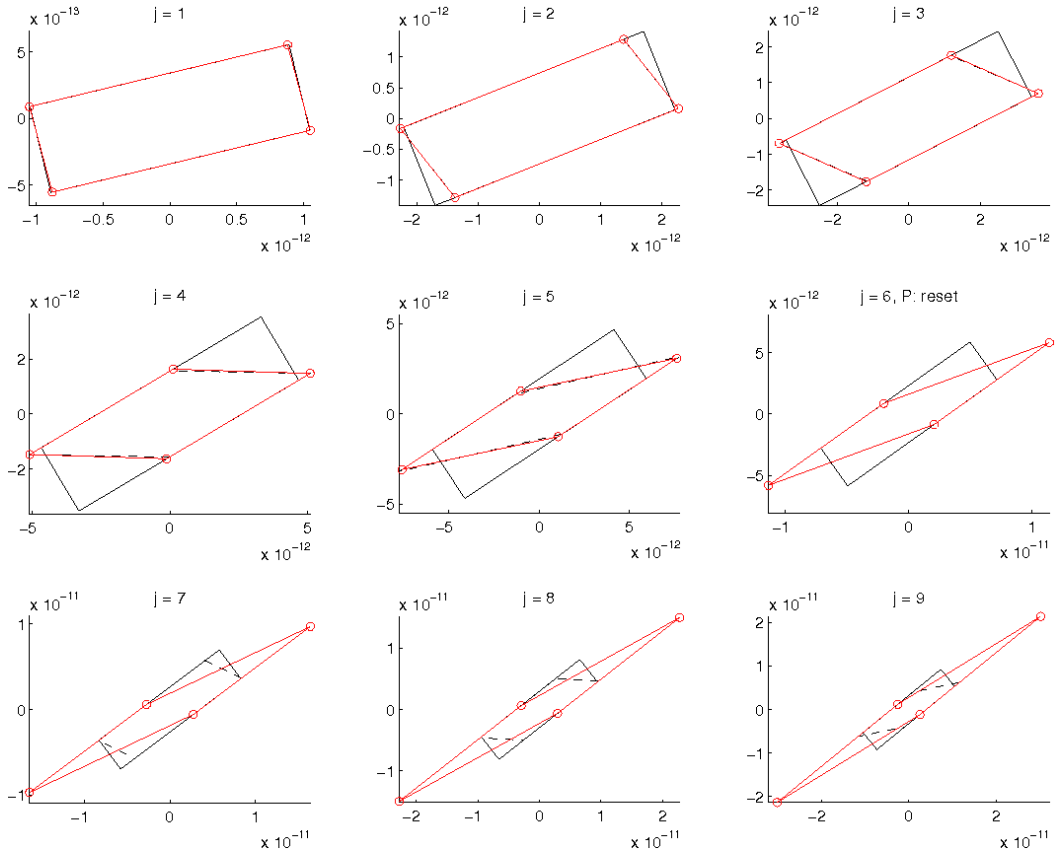


Figure 4: Blunting factor 1.0E-2.

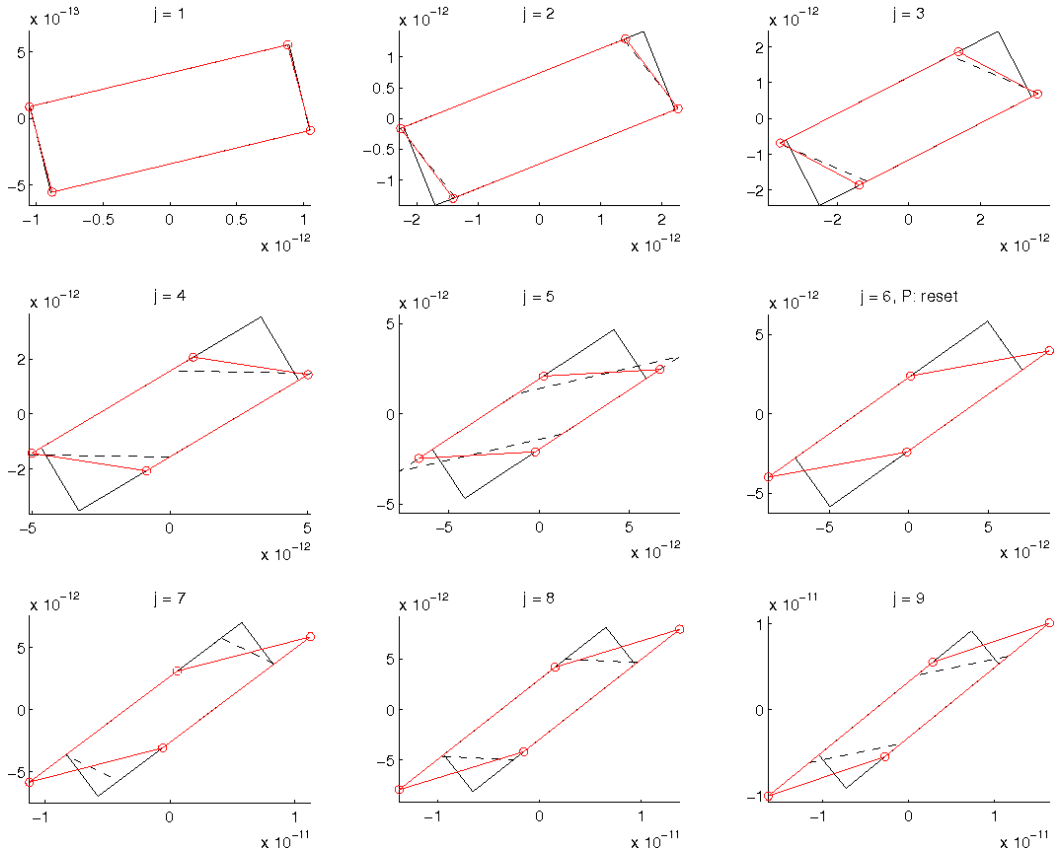


Figure 5: Blunting factor 1.0E-1.

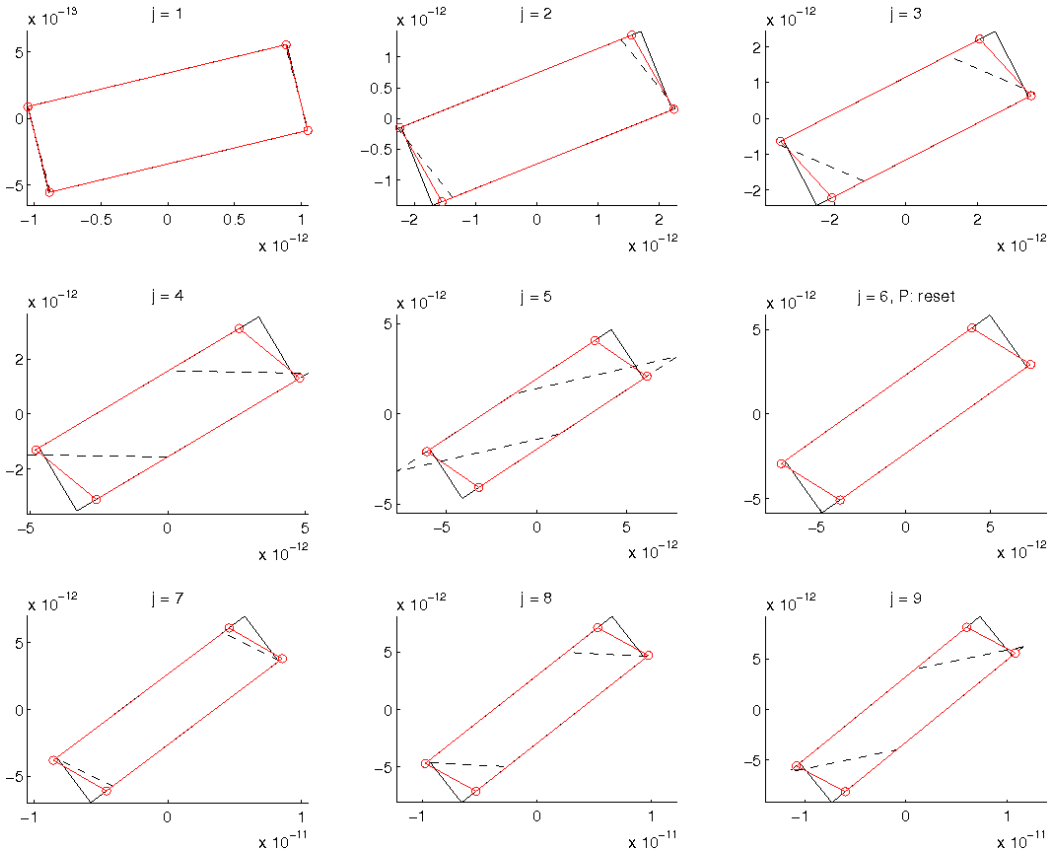


Figure 6: Blunting factor 1.0.

# References

[1] G. Alefeld and J. Herzberger. *Introduction to Interval Computations*. Academic Press, New York, 1983.

[2] M. Berz. From Taylor series to Taylor models. In *AIP Conference Proceedings 405*, pages 1–23, 1997.

[3] M. Berz. Cosy Infinity Version 8 reference manual. NSCL Technical Report MSUCL-1088, Michigan State University, 1998.

[4] M. Berz. COSY INFINITY. `http://bt.pa.msu.edu/index_files/cosy.htm`, August 2006.

[5] M. Berz and G. Hoffstätter. Computation and application of Taylor polynomials with interval remainder bounds. *Reliable Computing*, 4:83–97, 1998.

[6] M. Berz and K. Makino. Verified integration of ODEs and flows using differential algebraic methods on high-order Taylor models. *Reliable Computing*, 4:361–369, 1998.

[7] M. Berz and K. Makino. New methods for high-dimensional verified quadrature. *Reliable Computing*, 5:13–22, 1999.

[8] M. Berz and K. Makino. Suppression of the wrapping effect by Taylor model-based verified integrators: Long-term stabilization by shrink wrapping. *Int. J. Diff. Eq. Appl.*, 10:385–403, 2005.

[9] M. Berz and K. Makino. Performance of Taylor model methods for validated integration of ODEs. *Lecture Notes in Computer Science*, 3732:65–73, 2006.

[10] G. F. Corliss. Survey of interval algorithms for ordinary differential equations. *Appl. Math. Comput.*, 31:112–120, 1989.

[11] G. F. Corliss. Guaranteed error bounds for ordinary differential equations. In M. Ainsworth, J. Levesley, W. A. Light, and M. Marletta, editors, *Theory and Numerics of Ordinary and Partial Differential Equations*. Clarendon Press, Oxford, 1995.

[12] G. F. Corliss and R. Rihm. Validating an a priori enclosure using high-order Taylor series. In G. Alefeld, A. Frommer, and B. Lang, editors, *Scientific Computing and Validated Numerics*, pages 228–238. Akademie-Verlag, Berlin, 1996.

[13] J.-P. Eckmann, H. Koch, and P. Wittwer. A computer-assisted proof of universality in area-preserving maps. *Memoirs of the AMS*, 47(289), 1984.

[14] P. Eijgenraam. *The Solution of Initial Value Problems Using Interval Arithmetic*. Mathematical Centre Tracts No. 144, Stichting Mathematisch Centrum, 1981.

[15] A. Griewank. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2000.

[16] J. Hoefkens, M. Berz, and K. Makino. Verified high-order integration of DAEs and higher-order ODEs. In W. Krämer and J. Wolff von Gudenberg, editors, *Scientific Computing, Validated Numerics and Interval Methods*, pages 281–292. Kluwer, Dordrecht, Netherlands, 2001.

[17] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis*. Springer, London, 2001.

[18] E. W. Kaucher and W. L. Miranker. *Self-Validating Numerics for Function Space Problems*. Academic Press, New York, 1984.

[19] R. Klatte, U. Kulisch, Ch. Lawo, M. Rauch, and A. Wiethoff. *C-XSC: A C++ Class Library for Extended Scientific Computing*. Springer, Berlin, 1993.

[20] R. Klatte, U. Kulisch, M. Neaga, D. Ratz, and Ch. Ullrich. *Pascal-XSC – Language Reference with Examples*. Springer, Berlin, 1992.

[21] W. Kühn. Rigorously computed orbits of dynamical systems without the wrapping effect. *Computing*, 61:47–67, 1998.

[22] U. Kulisch and W. L. Miranker. *Computer Arithmetic in Theory and Practice*. Academic Press, New York, 1981.

[23] R. Lohner. *Einschließung der Lösung gewöhnlicher Anfangs- und Randwertaufgaben und Anwendungen*. PhD thesis, Universität Karlsruhe, 1988.

[24] R. Lohner. Computation of guaranteed solutions of ordinary initial and boundary value problems. In J. R. Cash and I. Gladwell, editors, *Computational Ordinary Differential Equations*, pages 425–435. Clarendon Press, Oxford, 1992.

[25] R. Lohner. On the ubiquity of the wrapping effect in the computation of error bounds. In U. Kulisch, R. Lohner, and A. Facius, editors, *Perspectives of Enclosure Methods*, pages 201–217. Springer, Wien, 2001.

[26] K. Makino. *Rigorous analysis of nonlinear motion in particle accelerators*. PhD thesis, Michigan State University, 1998.

[27] K. Makino and M. Berz. Remainder differential algebras and their applications. In M. Berz, C. Bischof, G. Corliss, and A. Griewank, editors, *Computational Differentiation: Techniques, Applications and Tools*, pages 63–74. SIAM, Philadelphia, 1996.

[28] K. Makino and M. Berz. COSY INFINITY version 8. *Nuclear Instruments & Methods in Physics Research A*, 427:338–343, 1999.

[29] K. Makino and M. Berz. Efficient control of the dependency problem based on Taylor model methods. *Reliable Computing*, 5:3–12, 1999.

[30] K. Makino and M. Berz. Taylor models and other validated functional inclusion methods. *Int. J. Pure Appl. Math.*, 4:379–456, 2003.

[31] K. Makino and M. Berz. Suppression of the wrapping effect by Taylor model-based validated integrators. MSU Report MSUHEP 40910, Michigan State University, 2004.

[32] K. Makino and M. Berz. Suppression of the wrapping effect by Taylor model-based verified integrators: Long-term stabilization by preconditioning. *Int. J. Diff. Eq. Appl.*, 10:353–384, 2005.

[33] K. Makino and M. Berz. Suppression of the wrapping effect by Taylor model-based verified integrators: The single step. *Int. J. Pure Appl. Math.*, 36:175–197, 2006.

[34] R. E. Moore. The automatic analysis and control of error in digital computation based on the use of interval numbers. In L. B. Rall, editor, *Error in Digital Computation, Vol. I*, pages 61–130. John Wiley and Sons, New York, 1965.

[35] R. E. Moore. Automatic local coordinate transformations to reduce the growth of error bounds in interval computation of solutions of ordinary differential equations. In L. B. Rall, editor, *Error in Digital Computation, Vol. II*, pages 103–140. John Wiley and Sons, New York, 1965.

[36] R. E. Moore. *Interval Analysis*. Prentice Hall, Englewood Cliffs, N.J., 1966.

[37] R. E. Moore, R. B. Kearfott, and M. J. Cloud. *Introduction to Interval Analysis*. SIAM, Philadelphia, 2009.

[38] P. S. V. Nataraj and K. Kotecha. Global optimization with higher order inclusion function forms. Part 1: A combined Taylor-Bernstein form. *Reliable Computing*, 10:27–44, 2004.

[39] N. S. Nedialkov. *Computing rigorous bounds on the solution of an IVP for an ODE*. PhD thesis, University of Toronto, 1999.

[40] N. S. Nedialkov and K. R. Jackson. A new perspective on the wrapping effect in interval methods for initial value problems for ordinary differential equations. In U. Kulisch, R. Lohner, and A. Facius, editors, *Perspectives of Enclosure Methods*, pages 219–264. Springer, Wien, 2001.

[41] N. S. Nedialkov and K. R. Jackson. Some recent advances in validated methods for IVPs for ODEs. *Appl. Numer. Math.*, 42:269–284, 2003.

[42] N. S. Nedialkov, K. R. Jackson, and G. F. Corliss. Validated solutions of initial value problems for ordinary differential equations. *Appl. Math. Comput.*, 105:21–68, 1999.

[43] N. S. Nedialkov, K. R. Jackson, and J. Pryce. An effective high-order interval method for validating existence and uniqueness of the solution of an IVP for an ODE. *Reliable Computing*, 7:449–465, 2001.

[44] M. Neher. Geometric series bounds for the local errors of Taylor methods for linear $n$th order ODEs. In G. Alefeld, J. Rohn, S. Rump, and T. Yamamoto, editors, *Symbolic Algebraic Methods and Verification Methods*, pages 183–193. Springer, Wien, 2001.

[45] M. Neher, K. R. Jackson, and N. S. Nedialkov. On Taylor model based integration of ODEs. *SIAM J. Numer. Anal.*, 45:236–262, 2007.

[46] A. Neumaier. *Interval Methods for Systems of Equations*. Cambridge University Press, Cambridge, 1990.

[47] L. B. Rall. *Automatic Differentiation: Techniques and Applications, Lecture Notes in Computer Science, Vol. 120*. Springer, Berlin, 1981.

[48] R. Rihm. *Über Einschließungsverfahren für gewöhnliche Anfangswertprobleme und ihre Anwendung auf Differentialgleichungen mit unstetiger rechter Seite*. PhD thesis, Universität Karlsruhe, Karlsruhe, Germany, 1993.

[49] R. Rihm. Interval methods for initial value problems in ODEs. In J. Herzberger, editor, *Topics in Validated Computations*, pages 173–207. Elsevier, Amsterdam, 1994.

[50] H. J. Stetter. Validated solution of initial value problems for ODE. *Notes Rep. Math. Sci. Eng.*, 7:171–193, 1990.