Topics in Interval Analysis

EDITED BY

E. HANSEN

OXFORD
AT THE CLARENDON PRESS
1969

Oxford University Press, Ely House, London W. 1

GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
CAPE TOWN SALISBURY IBADAN NAIROBI LUSAKA ADDIS ABABA
BOMBAY CALCUTTA MADRAS KARACHI LAHORE DACCA
KUALA LUMPUR SINGAPORE HONG KONG TOKYO

PRINTED IN GREAT BRITAIN

Contents

8 · Ordinary Differential Equations

1. Introduction

LET there be given a system of n ordinary differential equations of first order

$$y' = f(t, y) \tag{1}$$

with the initial condition

$$y(t_a) = s; \qquad s \in R^n \tag{2}$$

so that the solution of (1) and (2) is

$$\tilde{y}(t; t_a, s) \tag{3}$$

for $t \geqslant t_a$. Usually a numerical approximation for $\tilde{y}(t; t_a, s)$ is sought. If interval methods are used, condition (2) can be generalized. Instead of a point $s \in R^n$, a set $W_a \subset R^n$ can be used so that the result is a set of solutions. This set of solutions will be denoted by

$$\overline{W}(t) = \{z : z = \tilde{y}(t; t_a, s),\ s \in W_a\} \tag{4}$$

so that $\overline{W}(t_a) = W_a$.

When using interval arithmetic it is advantageous if W_a can be described by an interval vector or by a product of an interval vector and a point matrix. Intervals will be denoted by the symbol $\lfloor\ \rfloor$, so that $\lfloor c \rfloor$ is an interval and $\lfloor w \rfloor$ is an interval vector.

Using Taylor series, Moore [1] obtains very good results in the numerical integration of ordinary differential equations. This chapter describes a new numerical process that can be realized in several different ways. We will call it 'Three-Process Method' or 3PM.

2. The 3PM process

Assume we are given a system (1) or n ordinary differential equations and a set of initial values W_a^* at the time t_a. Our problem is to construct a set W_b^{**} so that

$$W_b^{**} \supseteq W_b^*, \tag{5}$$

$$W_b^* = \overline{W}^*(t_b), \tag{6}$$

and

$$\overline{W}^*(t) = \{z : z = \tilde{y}(t; t_a, s),\ s \in W_a^*\} \tag{7}$$

where

$$t_b = t_a + h, \quad h > 0. \tag{8}$$

The step length can be determined by process (I) given below.

Three procedures will be described for solving the problem. The procedures are defined by three sub-problems that must be solved. There are many numerical realizations for each procedure so that it is possible to find a large number of combined realizations for the whole 3PM.

2.1. *Process (I)*

Assume we are given an initial time value t_a and an initial set W_a^* at the time t_a. Also given is an integer $k \geqslant 0$.

Our problem is to determine a step length $h > 0$ and an interval polynomial with vector coefficients

$$\lfloor \hat{P}(t-t_a) \rfloor = \sum_{\nu=0}^{k} \lfloor p_\nu \rfloor (t-t_a) \tag{9}$$

so that for all $t \in [t_a, t_b]$,

$$\lfloor \hat{P}(t-t_a) \rfloor \supseteq \overline{W}^*(t). \tag{10}$$

Here t_b is determined by (8) and $\overline{W}^*(t)$ by (7). In most cases it is sufficient to construct $\lfloor \hat{P} \rfloor$ for $k = 0$.

2.2. *Process (II)*

Assume that at the time t_a there is given an initial point $\dot{V}_a^*$ with $\dot{V}_a^* \in W_a^*$. The dot over a variable denotes that the variable is a single value and not a set. Our problem is to find a set V_b^{**} with

$$V_b^{**} \supseteq \dot{V}_b^* \tag{11}$$

where

$$\dot{V}_b^* = \tilde{y}(t_b; t_a, \dot{V}_a^*). \tag{12}$$

Obviously process (II) contains the ordinary problem of numerical integration, starting with a single initial point. For the performance of process (II), the result of process (I) must be used.

2.3. *Process (III)*

Assume there is given a decomposition of W_a^* as the sum of the point $\dot{V}_a^*$ and a set U_a^* so that

$$W_a^* \subseteq \dot{V}_a^* + U_a^*, \quad 0 \in U_a^*. \tag{13}$$

Our problem is to find a set U_b^{**} with the property

$$U_b^{**} \supseteq U_b^* \tag{14}$$

where

$$U_b^* = W_b^* - \dot{V}_b^*. \tag{15}$$

The set U_b^* can be interpreted as the image of the perturbations U_a^* of $\dot{V}_a^*$ if integration is performed from t_a to t_b. For process (III) the result of process (I) must be used.

2.4. *Composition of processes (I), (II), and (III)*

By composition of processes (I), (II), and (III) the whole 3PM can be constructed in a simple way. This is illustrated in diagram (16) (p. 94).

From (16) it can be seen that 3PM is a one-step method. It differs from other integration methods in the separation of the integration into the two processes (II) and (III).

Some realizations of (I), (II), and (III) will now be considered.

3.1. *Realization of process (I)*

Moore ([1], p. 131 et seq.) gives a realization of process (I) for $k = 0$. This realization works very well. A realization for $k > 0$ can be found by using the Picard–Lindelöf iteration.

There may exist an interval vector $\lfloor w_a^* \rfloor$ with the property

$$W_a^* \subseteq \lfloor w_a^* \rfloor .$$

By using Moore's method for $k = 0$, an interval polynomial of degree $k = 0$ can be obtained with

$$\lfloor \hat{P}_0(t-t_a) \rfloor = \lfloor p_0 \rfloor \tag{17}$$

so that, for $t_b = t_a + h$ $(h > 0)$,

$$\overline{W}^*(t) \subseteq \lfloor p_0 \rfloor \tag{18}$$

for all $t \in [t_a, t_b]$. The Picard–Lindelöf iteration now leads to

$$\lfloor \hat{P}_{\nu+1} \rfloor := \lfloor w_a^* \rfloor + \int_0^t \lfloor f(t_a+\tau, \lfloor \hat{P}_\nu(\tau) \rfloor) \rfloor \, d\tau \tag{19}$$

for $\nu = 0, 1, 2, \dots$.

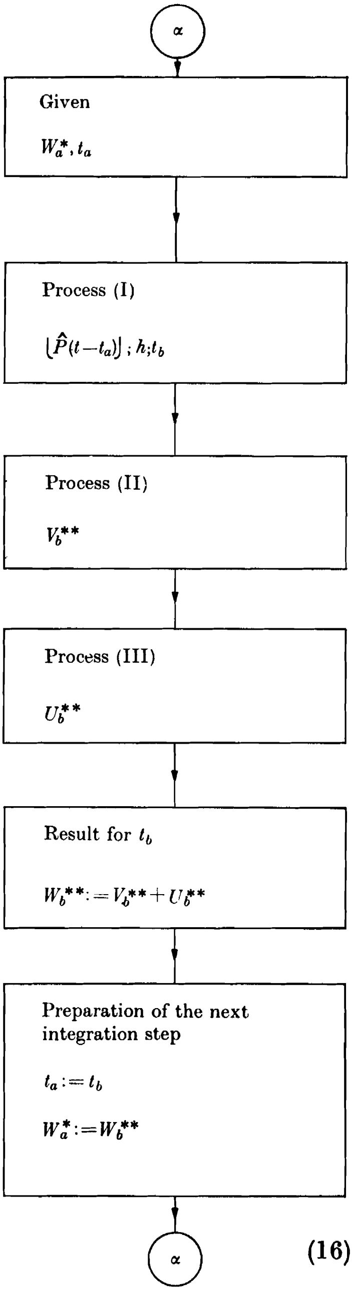
In (19) $\lfloor f \rfloor$ is a representation for f in interval arithmetic operations. By using interval polynomials for the integration (19), it is easy to get $\lfloor \hat{P}_{\nu+1} \rfloor$. But it is necessary to have the ability to limit the degree of $\lfloor \hat{P}_{\nu+1} \rfloor$. This can be done by *vergröberung*. After each step, $\nu \to \nu+1$, and the degree of $\lfloor \hat{P}_{\nu+1} \rfloor$ can be reduced to $\nu+1$. In this way—after k steps using (19)—we obtain $\lfloor \hat{P} \rfloor = \lfloor \hat{P}_k \rfloor$, where

$$\overline{W}^*(t) \subseteq \lfloor \hat{P}(t-t_a) \rfloor \tag{20}$$

for all $t \in [t_a, t_b]$. In practice it is sufficient to have the result (17) for $k = 0$. Only the cases $k = 1$ or $k = 2$ may also be of practical interest.

3.2. *Realization of process (II)*

For the realization of process (II), only a point-integration need be done, so that nearly all one-step integration methods are available if the remainder can be written down. The simplest realization is given by the Taylor series (see Moore [1]).



(16)

A point $\dot v_a^*$ must be chosen within $\lfloor w_a^* \rfloor$. Then $\dot v_a^*$ is a realization of $\dot V_a^*$. The Taylor series through terms of second order now has the form

$$\lfloor v_b^{**} \rfloor := \dot v_a^* + h \lfloor f(t_a, \dot v_a^*) \rfloor + \frac{h^2}{2!} \lfloor f'\{[t_a, t_b], \lfloor \hat P([0,h]) \rfloor\} \rfloor \tag{21}$$

where V_b^{**} is realized by

$$V_b^{**} = \lfloor v_b^{**} \rfloor . \tag{22}$$

Now it is true that

$$\dot V_b^* \in \lfloor v_b^{**} \rfloor \tag{23}$$

because in the remainder of (21) the whole set of solutions $\overline{W}^*(t)$ is included in $\lfloor \hat P([0,h]) \rfloor$.

3.3 *Realization of process (III)*

It is very important to have a good realization of process (III). Only then is it possible to get small bounds for the error propagation. By linearization of the given differential equation in the neighbourhood of $\tilde y(t; t_a, \dot V_a^*)$ and by using the theory of matricants a realization can be found. The interval vector

$$\lfloor \breve u_a^* \rfloor := \lfloor w_a^* \rfloor - \dot v_a^* \tag{24}$$

is a description of U_a^*. Interval functions $\lfloor l_{ij} \rfloor$ must be obtained which contain the functional matrix of f for each element so that

$$L(t,y) = \{l_{ij}(t,y)\}, \tag{25}$$

$$l_{ij}(t,y) = \frac{\partial f_i(t,y)}{\partial y_j}, \tag{26}$$

$$\lfloor l_{ij}(t,y) \rfloor \ni l_{ij}(t,y) \quad \text{for } t_a \leqslant t \leqslant t_b. \tag{27}$$

Let $\lfloor p_0 \rfloor$ be the result of process (I) for $k = 0$. Then

$$\lfloor \gamma \rfloor := \lfloor L([t_a, t_b], \lfloor p_0 \rfloor) \rfloor \tag{28}$$

must be computed. Now $\lfloor \gamma \rfloor$ contains all matrices L with arguments in the set $\overline{W}^*(t)$ for $t \in [t_a, t_b]$. Hence it follows by the theory of matricants that for

$$\lfloor Q \rfloor := \dot{f} + \sum_{\nu=1}^{\infty} \frac{h^\nu}{\nu!}(\lfloor \gamma \rfloor)^\nu, \tag{29}$$

$$\lfloor \check{u}_b^{**} \rfloor := \lfloor Q \rfloor \lfloor \check{u}_a^* \rfloor \tag{30}$$

so that for the interval vector $\lfloor \check{u}_b^{**} \rfloor$, the relation

$$\lfloor \check{u}_b^{**} \rfloor \supseteq U_b^* \tag{31}$$

holds. The machine calculation of $\lfloor Q \rfloor$ can be done by interval arithmetic without difficulty.

If $\lfloor Q \rfloor$ contains a geometric rotation, the result $\lfloor \check{u}_b^{**} \rfloor$ may not be very good. Moore discusses this difficulty in his book [1]. At Bonn we have evaluated a special method for carrying out a mapping that is better than (30). It is assumed that $U_a^* = W_c^* - \dot{V}_a^*$ has the following representation:

$$\lfloor \check{u}_a^* \rfloor \supseteq \dot{T}_a \lfloor \theta_a \rfloor \supseteq U_a^*, \tag{32}$$

$$\lfloor w_a^* \rfloor \supseteq \dot{T}_a \lfloor \theta_a \rfloor + \dot{v}_a^* \supseteq W_a^*. \tag{33}$$

The product $\dot{T}_a \lfloor \theta_a \rfloor$ is defined in the sense of the 'united extension' (see Moore [1]):

$$\dot{T}_a \lfloor \theta_a \rfloor = \bigcup_{\theta \in \lfloor \theta_a \rfloor} \dot{T}_a \theta. \tag{33}$$

Process (I) can be realized independently of this representation by using $\lfloor w_a^* \rfloor$. Also $\lfloor Q \rfloor$ can be computed in the old way. Only the mapping (30) must be obtained in a more complicated way:

$$\begin{gathered} \lfloor Q \rfloor \subseteq \dot{Q}_1 + \lfloor Q_2 \rfloor, \qquad \dot{Q}_1 \in \lfloor Q \rfloor, \\ \dot{Q}_1 \dot{T}_a \subseteq \dot{T}_b + \lfloor T \rfloor, \\ \lfloor T \rfloor \lfloor \theta_a \rfloor + \lfloor Q_2 \rfloor \dot{T}_a \lfloor \theta_a \rfloor \subseteq \lfloor \theta \rfloor. \end{gathered} \tag{34}$$

Then
$$\lfloor Q \rfloor (\dot{T}_a \lfloor \theta_a \rfloor) \subseteq \dot{T}_b \lfloor \theta_a \rfloor + \lfloor \theta \rfloor = U_b^{**} \tag{35}$$

and
$$\dot{T}_b \lfloor \theta_a \rfloor + \lfloor \theta \rfloor \supseteq U_b^*. \tag{36}$$

Now the set U_b^{**} is not described by an interval vector. This implies that the following change should be made in (16):

$$W_b^{**} = V_b^{**} + U_b^{**} \subseteq \lfloor v_b^{**} \rfloor + \dot{T}_b \lfloor \theta_a \rfloor + \lfloor \theta \rfloor. \tag{37}$$

By separation of v_b^{**} in the form

$$\lfloor v_b^{**} \rfloor \subseteq \dot{v}_{1b}^{**} + \lfloor v_{2b}^{**} \rfloor \quad (\dot{v}_{1b}^{**} \in \lfloor v_b^{**} \rfloor), \tag{38}$$

the relation
$$W_b^{**} \subseteq \dot{v}_{1b}^{**} + \dot{T}_b \lfloor \theta_a \rfloor + \lfloor \theta_1 \rfloor \tag{39}$$

holds with
$$\lfloor \theta_1 \rfloor \supseteq \lfloor \theta \rfloor + \lfloor v_{2b}^{**} \rfloor. \tag{40}$$

By performing the operations indicated in (39) it is possible to get a description of W_b^{**} in the form (33). If $\lfloor S \rfloor \supseteq \dot{T}_b^{-1}$ then

$$\dot{T}_b \lfloor \theta_a \rfloor + \lfloor \theta_1 \rfloor \subseteq \dot{T}_b(\lfloor \theta_a \rfloor + \lfloor S \rfloor \lfloor \theta_1 \rfloor), \tag{41}$$

$$\lfloor \theta_b \rfloor \supseteq \lfloor S \rfloor \lfloor \theta_1 \rfloor + \lfloor \theta_a \rfloor, \tag{42}$$

so that

$$W_b^{**} \subseteq \dot{v}_{1b}^{**} + \dot{T}_b \lfloor \theta_b \rfloor \tag{43}$$

and (43) is of the form (33). For starting the next integration step all variables must be given a new notation:

$$\begin{aligned} \dot{v}_a^* &:= \dot{v}_{1b}^{**}, \\ \dot{T}_a &:= \dot{T}_b, \\ \lfloor \theta_a \rfloor &:= \lfloor \theta_b \rfloor, \\ \lfloor w_a^* \rfloor &\supseteq \dot{T}_a \lfloor \theta_a \rfloor + \dot{v}_a^* \supseteq W_a^*. \end{aligned} \tag{44}$$

By this more complicated method, very good results can be obtained. By a small modification it is possible to also get 'inside' interval vectors $\lceil \theta_a \rceil$ with the property that the set so described is contained in the set of all solutions.

4. Examples

To illustrate the preceding analysis, we consider two problems we have solved by the methods described. Our first example is the astronomical three-body problem. The differential equations used as the test problem were those for the three-dimensional problem sun–Jupiter–8th moon of Jupiter. This problem was integrated (see Krückeberg [4]) by Taylor series of order three with $h = 1/16$ for 1600 steps ($=$ 100 days) by 3PM. The results are

$$\begin{aligned} \lfloor y_1 \rfloor &= [-1{\cdot}28523007\underline{40},\ 1{\cdot}28523007\underline{03}], \\ \lfloor y_2 \rfloor &= [0{\cdot}8599642591,\ 0{\cdot}8599642\underline{669}], \\ \lfloor y_3 \rfloor &= [0{\cdot}30140620\underline{70},\ 0{\cdot}30140620\underline{88}], \\ \lfloor y_1' \rfloor &= [0{\cdot}9963448543,\ 0{\cdot}9963448\underline{606}], \\ \lfloor y_2' \rfloor &= [0{\cdot}5962818\underline{843},\ 0{\cdot}5962819067], \\ \lfloor y_3' \rfloor &= [-0{\cdot}60424862\underline{88},\ -0{\cdot}60424862\underline{28}]. \end{aligned}$$

For our second example, we consider the integration of $y'' = -y$. This very simple example is very interesting from the following point of view. The initial set is rotated through an angle $\zeta = 2\pi$ as the time variable runs from t to $t+2\pi$. It is difficult to find small bounds for the error propagation (see Moore [1]). Without any special technique, the bounds can be overestimated by a factor of about 500 for each rotation.

Moore has reduced this factor to about 16. Using the technique described above, it is possible to perform 12340 integration steps with $h = \pi/10$ and, after 617 rotations, get the following 'inside' and 'outside' interval vectors for the mapping of the starting 'window' $\lfloor \theta_0 \rfloor$:

$$\lfloor \theta_0 \rfloor = \begin{bmatrix} [-1, +1] \\ [-1, +1] \end{bmatrix},$$

$$\lceil \theta \rceil = \begin{bmatrix} [-0{\cdot}997, +0{\cdot}997] \\ [-0{\cdot}997, +0{\cdot}997] \end{bmatrix}; \qquad \lfloor \theta \rfloor = \begin{bmatrix} [-1{\cdot}002, +1{\cdot}002] \\ [-1{\cdot}002, +1{\cdot}002] \end{bmatrix}.$$

The results were computed using single word length in an interval version of Fortran, called Fortran-i, wherein Fortran expressions are automatically interpreted as interval arithmetic expressions.

REFERENCES

1. Moore, R. E. *Interval analysis.* Prentice-Hall, New Jersey (1966).
2. Nickel, K. Über die Notwendigkeit einer Fehlerschranken-Arithmetik für Rechenautomaten. *Num. Math.* **9** (1) 69–79 (1966).
3. Krückeberg, F. Zur numerischen Integration und Fehlererfassung bei Anfangswertaufgaben gewöhnlicher Differentialgleichungen. *Schriften des Rhein.-Westf. Institutes für Instrumentelle Mathematik an der Universität Bonn,* Nr. 1. Bonn (1961).
4. —— Zur numerischen Integration und Abschätzung des Integrationsfehlers bei gewöhnlichen Differentialgleichungen. *Tagungsbericht über mathematische Methoden der Himmelsmechanik und Astronautik* (*März 64*) *Oberwolfach-Berichte,* Heft 1, Herausgeber: Prof. Stiefel, Zürich.
5. Kulisch, U. and Apostolatos, N. Approximation der erweiterten Intervall-arithmetik durch die einfache Maschinenintervallarithmetik, *Computing* **2** (3) 181–94 (1967).