Zdzisław Jackiewicz
Global error estimation in the numerical solution of retarded differential equations by Euler’s method

Aplikace matematiky, Vol. 28 (1983), No. 3, 177–185

Persistent URL: http://dml.cz/dmlcz/104024

Terms of use:

© Institute of Mathematics AS CR, 1983

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.
GLOBAL ERROR ESTIMATION IN THE NUMERICAL SOLUTION
OF RETARDED DIFFERENTIAL EQUATIONS
BY EULER’S METHOD

Zdzislaw Jackiewicz

(Received June 3, 1982)

1. INTRODUCTION

Consider the initial-value problem for the system of retarded ordinary differential
equations

\[ \begin{align*}
    y'(t) &= f_i(\bar{y}(\bar{a}(t))), & t &\in [a, b], \\
    y_i(t) &= g_i(t), & t &\in [a, \bar{a}],
\end{align*} \]

\( i = 1, 2, \ldots, s, \) where \( s \) is a positive integer. Here \( a \leq a < b, g_i \) are specified initial
functions and

\[ \bar{y}(\bar{a}(t)) = (y_1(\bar{a}_1(t)), \ldots, y_1(\bar{a}_{s_1}(t)), \ldots, y_s(\bar{a}_{s_1}(t)), \ldots, y_s(\bar{a}_{s_k}(t))). \]

Putting \( y = [y_1, \ldots, y_s]^T, \ y' = [y'_1, \ldots, y'_s]^T, \ f = [f_1, \ldots, f_s]^T, \ g = [g_1, \ldots, g_s]^T, \)
where \( T \) stands for transposition, we can rewrite (1) in the vector form:

\[ \begin{align*}
    y(t) &= f(t), & t &\in [a, \bar{a}],
\end{align*} \]

\( 1' \)

For \( x \in R^s \) denote by \( \|x\| \) the maximum norm. We assume the following:

\( H_1. \) The function \( f: R^K \to R^s, K = k_1 + k_2 + \ldots + k_s, \) is of class \( C^1 \) and there
exists a constant \( M < \infty \) such that

\[ \begin{align*}
    \|f(u)\| &\leq M, & \|f(u) - f(v)\| &\leq M\|u - v\|, \\
    \|Df(u)\| &\leq M, & \|Df(u) - Df(v)\| &\leq M\|u - v\|
\end{align*} \]

for \( u, v \in R^K. \)

\( H_2. \) The functions \( \bar{a}_{i,j}: [a, b] \to [\bar{a}, b], \ i = 1, 2, \ldots, s, \ j = 1, 2, \ldots, k_i, \) are Lip-
schitz-continuous with constant \( Q < \infty, \) i.e.,

\[ |\bar{a}_{i,j}(t_1) - \bar{a}_{i,j}(t_2)| \leq Q|t_1 - t_2| \]

for \( t_1, t_2 \in [a, b]. \)
Let a fixed \( h \in (0, h_0] \), \( h_0 > 0 \) be given. To compute an approximate solution \( y_h : [a, b] \to \mathbb{R} \), consider Euler's method defined by

\[
y_h(t_n + rh) = y_h(t_n) + rh f(\bar{y}_h(\bar{a}(t_n))),
\]
\[
y_h(t) = g_h(t), \quad t \in [a, a],
\]

\( n = 0, 1, \ldots, N - 1, \) \( rh \in [0, 1] \). \( Nh = b - a \), \( t_n = a + nh \). Here \( g_h \) is some continuous approximation to the initial function \( g \).

To obtain an estimate of the global error \( e_h(t) = y_h(t) - y(t) \) we use the method of Zadunaisky (see [8], [7]). This method consists in the following. We construct the pseudo-problem

\[
u'(t) = f(\bar{u}(\bar{a}(t))) + d_h(t), \quad t \in [a, b],
\]
\[
u(t) = g(t), \quad t \in [a, a],
\]
in such a way that the exact solution \( u \) of this problem is known in advance and the defect function \( d_h \) is "small". This construction will be described in § 2. Denote by \( e_h^* \) the global error committed in the numerical solution of (3) by (2). Then, under certain conditions, \( e_h^* \) is a good estimate of \( e_h \). This result is stated in § 2 and its proof is given in § 3. In § 4 some numerical examples are given.

2. GLOBAL ERROR ESTIMATION

Assume that \( N \) is even and consider a piecewise polynomial interpolation of degree two to the numerical solutions \( \{y_{1,0}(t_n)\}_{n=0}^{N} \), \( i = 1, 2, \ldots, s \). In vector notation this can be written as

\[
P(t) = P^m(t) = a_0^m + (t - t_{2m})(a_1^m + (t - t_{2m+1})a_2^m), \quad t \in [t_{2m}, t_{2m+2}].
\]

Here, \( a_j^m, j = 0, 1, 2 \), are divided differences given by

\[
a_0^m = [t_{2m}, y_{h}], \quad y_{h}(t_{2m}),
\]
\[
a_1^m = [t_{2m}, t_{2m+1}; y_{h}] = f(\bar{y}_h(\bar{a}(t_{2m}))),
\]
\[
a_2^m = [t_{2m}, t_{2m+1}, t_{2m+2}; y_{h}] = \frac{1}{2h} [f(\bar{y}_h(\bar{a}(t_{2m+1}))) - f(\bar{y}_h(\bar{a}(t_{2m})))].
\]

Consider now the pseudo-problem defined by

\[
u'(t) = f(\bar{u}(\bar{a}(t))) + d_h(t), \quad t \in [t_{2m}, t_{2m+2}],
\]
\[
u(t) = g(t), \quad t \in [a, a],
\]

where

\[
d_h(t) = P(t) - f(P(\bar{a}(t))), \quad t \in [t_{2m}, t_{2m+2}].
\]

By \( u'(t_{2m}) \) and \( P'(t_{2m}) \) we mean the right hand side derivatives. It is obvious that \( P \)
is the continuous solution of this problem. The method (2) applied to (4) takes the form
\[ u_h(t_n + rh) = u_h(t_n) + rh[f(u_h(\bar{z}(t_n))) + d_h(t_n)], \]
\[ u_h(t) = g_h(t), \quad t \in [a, a], \]
n = 0, 1, \ldots, N - 1, r \in [0, 1]. Put \( e_h^*(t) = u_h(t) - P(t) \). We have the following.

**Theorem.** Assume that \( H_1 \) and \( H_2 \) hold. Then \( e_h(t) = e_h^*(t) + 0(h^2) \) as \( h \to 0 \).

This theorem generalizes some of the results obtained by Frank [2] and Frank/Ueberhuber [3] for ordinary differential equations. In [6] a similar result was obtained for Volterra integro-differential equations. The proof of this theorem is given in the next section and, as in [6], consists in checking if the method (2) possesses the "property (E)" defined by Stetter [7] (see also [8]).

3. THE PROOF OF THEOREM

We assume throughout this section that the conditions \( H_1 \) and \( H_2 \) are fulfilled and that \( N \) is even. Similarly as in [2] the proof is divided into a sequence of Lemmas.

**Lemma 1.** There exists a constant \( A < \infty \) independent of \( m \) and \( h \) such that \( \|a_m^m\| \leq A \) for \( m = 0, 1, \ldots, N/2 - 1; j = 0, 1, 2. \)

**Proof.** The proof for \( j = 0 \) and \( j = 1 \) is obvious. For \( j = 2 \), using \( H_1 \), we obtain
\[ \|a_m^m\| \leq \frac{M}{2h} \|\bar{y}_h(\bar{z}(t_{2m+1})) - \bar{y}_h(\bar{z}(t_{2m}))\|. \]

It is easy to see that the function \( y_h \) is Lipschitz-continuous with constant \( M \). This yields
\[ \|a_m^m\| \leq \frac{M^2}{2h} \|\bar{z}(t_{2m+1}) - \bar{z}(t_{2m})\| \leq \frac{M^2Q}{2h} |t_{2m+1} - t_{2m}| = \frac{1}{2}M^2Q. \]

Here, \( \bar{z}(t) = (\alpha_1(t), \ldots, \alpha_{1,k_1}(t), \ldots, \alpha_s(t), \ldots, \alpha_{s,k_s}(t)) \).

**Lemma 2.** \( \|d_h(t)\| = O(h) \) as \( h \to 0 \) for \( t \in [a, b] \).

**Proof.** For \( t \in [t_{2m}, t_{2m+2}) \) we get
\[ d_h(t) = (P^m)'(t) - f(P^m(\bar{z}(t))) = a_1^m + a_2^m[(t - t_{2m}) + (t - t_{2m+1})] - f(\bar{y}_h(\bar{z}(t_{2m})) + P^m(\bar{z}(t)) - \bar{y}_h(\bar{z}(t_{2m}))) = a_2^m[(t - t_{2m}) + (t - t_{2m+1})] - D f(\eta(t)) (P^m(\bar{z}(t)) - \bar{y}_h(\bar{z}(t_{2m}))), \]
where \( \eta(t) \in R^K \) lies between \( \bar{P}(\bar{z}(t)) \) and \( \bar{y}(\bar{z}(t_{2m})) \). In view of Lemma 1 and \( H_1 \)
we obtain
\[ \| d_h(t) \| \leq 2Ah + M \| \tilde{P}^m(\tilde{x}(t)) - \tilde{y}_h(\tilde{x}(t_{2m})) \|. \]
We have to estimate the quantities \( |P_i(a_{i,j}(t)) - y_{i,h}(a_{i,j}(t_{2m}))| \) for \( i = 1, 2, \ldots, s \), \( j = 1, 2, \ldots, k_i \). For any \( i, j \), \( a_{i,j}(t) \in [t_{2v}, t_{2v+2}] \) for some \( v = v(i,j) \leq m \). We have
\[
\begin{align*}
|P_i(a_{i,j}(t)) - y_{i,h}(a_{i,j}(t_{2m}))| &= |a_{i,0}^i + (a_{i,j}(t) - t_{2v}) (a_{i,1}^i + (a_{i,j}(t) - t_{2v+1}) a_{i,2}^i) - y_{i,h}(a_{i,j}(t_{2m}))| \\
&\leq |y_{i,h}(t_{2v}) - y_{i,h}(a_{i,j}(t_{2m}))| + 2h(A + Ah) \leq 2hM + 2hA + O(h^2).
\end{align*}
\]
Finally,
\[
\| \tilde{P}^m(\tilde{x}(t)) - \tilde{y}_h(\tilde{x}(t_{2m})) \| = O(h) \quad \text{and} \quad \| d_h(t) \| = O(h) \quad \text{as} \quad h \to 0.
\]

**Lemma 3.** Denote by \( e \) the solution of the problem
\begin{equation}
(5) \quad e'(t) = Df(y(a(t))) e(a(t)) - iy''(t), \quad t \in [a, b],
\end{equation}
\begin{equation}
(6) \quad y(t_n + rh) = y(t_n) + rhf(y(a(t_n))) + \mu(t_n, r, h),
\end{equation}
where \( y \) is the solution of (1). Then \( e_h(t_n + rh) = he(t_n + rh) + O(h^2) \) as \( h \to 0 \).

**Proof.** Define the local error \( \mu(t_n, r, h) \) of the method (2) at the point \( t_n + rh \) by
\[
\mu(t_n, r, h) = y''(t_n) \frac{r^2h^2}{2} + O(h^3) \quad \text{as} \quad h \to 0.
\]
Subtracting (6) from (2) we get
\[
e_h(t_n + rh) = e_h(t_n) + rh\left[f(y(a(t_n))) - f(y(\tilde{a}(t_n)))\right] - \frac{1}{2}r^2h^2 y''(t_n) + O(h^3).
\]
Routine manipulations yield
\[
e_h(t_n + rh) = e_h(t_n) + rh\left[D f(y(\tilde{a}(t_n))) \right. \\
\left. + \frac{1}{2}D^2 f(\tilde{y}(\tilde{a}(t_n))) \tilde{e}_h(\tilde{a}(t_n)) \right] - \frac{1}{2}r^2h^2 y''(t_n) + 0(h^3) =
\]
\[
e_h(t_n) + rhD f(y(\tilde{a}(t_n))) \tilde{e}_h(\tilde{a}(t_n)) - \frac{1}{2}r^2h^2 y''(t_n) + 0(h^3).
\]
Let \( e_h(t_n + rh) = e_h(t_n + rh)/h \). Then
\begin{equation}
(7) \quad e_h(t_n + rh) = e_h(t_n) + rh\left[D f(y(\tilde{a}(t_n))) \tilde{e}_h(\tilde{a}(t_n)) - \frac{1}{2}r y''(t_n)\right] + O(h^2).
\end{equation}

Putting \( e_h(t) = 0 \) for \( t \in [a, a] \) we can look at (6) as the result of applying to the equation (5) some numerical method with additional error of order two. Similarly as in [6] it is easy to check that this method is consistent with order one. Consequently, it follows from Theorem 5 of [5] that \( \| e_h(t_n + rh) - e(t_n + rh) \| = O(h) \) as \( h \to 0 \) or \( e_h(t_n + rh) = h e(t_n + rh) + O(h^2) \), which is our claim.
Lemma 4. Denote by $e^*$ the continuous solution of the problem

\begin{align*}
(e^*)'(t) &= Df(P(a(t)j)e^*(S(t)) - P'(t),
& t \in [t_{2m+1}, t_{2m+2}),
& e^*(t) = 0, \quad t \in [a, a],
& m = 0, 1, \ldots, N/2 - 1,
& \text{where } P \text{ is the solution of (4).}
\end{align*}

Then $e^*(t_n + rh) = \frac{h}{e^*}(t_{n} + rh) + O(h^2)$ as $h \to 0$ for $n = 0, 1, \ldots, N - 1, \ r \in [0, 1]$.

Proof. The proof of this lemma is similar to that of Lemma 3 is therefore omitted.

The next lemma is a generalization of Gronwall’s inequality.

Lemma 5. Assume that $w_i(t) \geq 0$, $i = 1, 2, \ldots, s$, $t \in [a, a]$ and

\begin{align*}
w_i(t) &\leq B \int_a^t \sum_{i=1}^s \sum_{j=1}^{k_i} w_i(z_{i,j}(x)) \, dx + C, \quad t \in [a, b],
\end{align*}

where $B$ and $C$ are nonnegative constants. Then

\begin{align*}
w_i(t) &\leq C \exp(BK(t - a)), \quad t \in [a, b].
\end{align*}

Proof. It follows from the theory of integral inequalities that $w_i(t) \leq W_i(t)$, $t \in [a, b]$, where $W_i$ are functions satisfying the equations

\begin{align*}
W_i(t) &= B \int_a^t \sum_{i=1}^s \sum_{j=1}^{k_i} W_i(z_{i,j}(x)) \, dx + C, \quad t \in [a, b],
W_i(t) &= w_i(t), \quad t \in [a, a].
\end{align*}

It is easy to see that the functions $W_i$ are nondecreasing for $t \in [a, b]$. This yields

\begin{align*}
W_i(t) &\leq B \int_a^t \sum_{i=1}^s \sum_{j=1}^{k_i} W_i(x) \, dx + C = B \int_a^t \sum_{i=1}^s k_i W_i(x) \, dx + C, \quad t \in [a, b].
\end{align*}

Now, after simple calculations, the result follows from Gronwall’s inequality.

Lemma 6. $\|y(t) - P(t)\| = 0(h)$ and $\|y'(t) - P'(t)\| = 0(h)$ as $h \to 0$ for $t \in [a, b]$.

Proof. Integrating (1') and (4) we obtain

\begin{align*}
y(t) &= y(a) + \int_a^t f(\bar{y}((x))) \, dx, \quad t \in [a, b],
P(t) &= P(a) + \int_a^t f(\bar{P}(\bar{a}(x))) \, dx + \int_a^t d(x) \, dx, \quad t \in [a, b].
\end{align*}

Subtracting these equations and using $H_1$ we get

\begin{align*}
\|y_i(t) - P_i(t)\| &\leq \int_a^t M \sum_{i=1}^s \sum_{j=1}^{k_i} |y_i(z_{i,j}(x)) - P_i(z_{i,j}(x))| \, dx + C,
\end{align*}

181
where \( C = (b - a) \sup \|d_i(x)\| : x \in [a, b] \). Putting \( w_i(t) = |y_i(t) - P_i(t)| \), we obtain from Lemma 5 that

\[
w_i(t) \leq C \exp\left(MK(b - a)\right),
\]

\( i = 1, 2, \ldots, s \). This proves the first part of the lemma. The second part follows from the inequality

\[
\|y'(t) - P'(t)\| \leq M\|\dot{y}(\bar{a}(t)) - P(\bar{a}(t))\| + \|d_i(t)\|, \quad t \in [a, b].
\]

Lemma 7. \( e^*(t) = e(t) + O(h) \) as \( h \to 0 \) for \( t \in [a, b] \).

Proof. Integrating (5) and (8) and subtracting the resulting equations we obtain

\[
|e_i(t) - e_i^*(t)| \leq \int_a^t \left| Df_i(\bar{y}(\bar{a}(x))) \bar{e}(\bar{a}(x)) - Df_i(P(\bar{a}(x))) \bar{e}^*(\bar{a}(x)) \right| dx +
\]

\[
+ \frac{1}{2} \left( |y'(t) - P'(t)| + |y'(a) - P'(a)| \right), \quad t \in [a, b].
\]

Putting \( E = \sup \|\bar{e}^*(\bar{a}(x))\| : x \in [a, b] \) we get

\[
|Df_i(\bar{y}(\bar{a}(x))) \bar{e}(\bar{a}(x)) - Df_i(P(\bar{a}(x))) \bar{e}^*(\bar{a}(x))| \leq
\]

\[
\leq |Df_i(\bar{y}(\bar{a}(x))) \bar{e}(\bar{a}(x)) - Df_i(P(\bar{a}(x))) \bar{e}^*(\bar{a}(x))| +
\]

\[
+ |Df_i(\bar{y}(\bar{a}(x))) \bar{e}^*(\bar{a}(x)) - Df_i(P(\bar{a}(x))) \bar{e}^*(\bar{a}(x))| \leq
\]

\[
\leq M\|\bar{e}(\bar{a}(x)) - \bar{e}^*(\bar{a}(x))\| + ME\|\bar{y}(\bar{a}(x)) - P(\bar{a}(x))\|.
\]

Hence, in view of Lemma 6,

\[
|e_i(t) - e_i^*(t)| \leq M \int_a^t \sum_{i=1}^s \sum_{j=1}^{k_i} \left| e_i(\bar{x}_{i,j}(x)) - e_i^*(\bar{x}_{i,j}(x)) \right| dx + O(h)
\]

as \( h \to 0 \). Now the desired conclusion follows from Lemma 5.

Proof of Theorem. The theorem follows immediately from Lemmas 3, 4, and 7. Compare also the proof of Theorem 2 in [6].

4. NUMERICAL EXAMPLES

Example 1 (Hill [4]).

\[
y'(t) = -\left[y(t)/(1 + 2t)^2\right]^{(1+2t)^2}, \quad t \in [0, 1],
\]

\[
y(0) = 1.
\]

The exact solution is \( y(t) = -\exp(t) \).

Example 2 (Bellman, Buell, Kalaba [1]).
\[ y'(t) = -y(t - \exp(-t) - 1) + \left[ \cos(t) + \sin(t - \exp(-t) - 1) \right], \quad t \in [0, 1], \]
\[ y(t) = \sin(t), \quad t \in [-2, 0]. \]
The solution is \( y(t) = \sin(t). \)

Example 3.
\[ y'(t) = -2 \tan\left(\frac{t}{2}\right) y^2(t/2), \quad t \in [0, 1] \]
\[ y(0) = 1. \]
The exact solution is \( y(t) = \cos(t). \)

Example 4.
\[ y'(t) = \exp\left(y(\alpha(t))\right)/(t^2 + 4t + 3), \quad t \in [0, 1], \]
\[ y(t) = \ln(2 + t), \quad t \in [-1/2, 0], \]
where \( \alpha(t) = t - 1/(2 + t). \) The solution is \( y(t) = \ln(2 + t). \)
The results of computations are given in the tables below, where \( E = e_h(b) - e_h^*(b). \) These results confirm the Theorem given in § 2.

<table>
<thead>
<tr>
<th>Table 1. Results for Example 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h )</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>( 2^{-2} )</td>
</tr>
<tr>
<td>( 2^{-3} )</td>
</tr>
<tr>
<td>( 2^{-4} )</td>
</tr>
<tr>
<td>( 2^{-5} )</td>
</tr>
<tr>
<td>( 2^{-6} )</td>
</tr>
<tr>
<td>( 2^{-7} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Results for Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h )</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>( 2^{-2} )</td>
</tr>
<tr>
<td>( 2^{-3} )</td>
</tr>
<tr>
<td>( 2^{-4} )</td>
</tr>
<tr>
<td>( 2^{-5} )</td>
</tr>
<tr>
<td>( 2^{-6} )</td>
</tr>
<tr>
<td>( 2^{-7} )</td>
</tr>
</tbody>
</table>

183
### Table 3. Results for Example 3

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_h(1)$</th>
<th>$e_h^*(1)$</th>
<th>$E/h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>-0.269 199</td>
<td>-0.160 093</td>
<td>1.74</td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>-0.126 507</td>
<td>-0.131 445</td>
<td>0.32</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>-0.057 637</td>
<td>-0.063 532</td>
<td>1.51</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>-0.027 118</td>
<td>-0.028 893</td>
<td>1.82</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>-0.013 114</td>
<td>-0.013 578</td>
<td>1.90</td>
</tr>
<tr>
<td>$2^{-7}$</td>
<td>-0.006 442</td>
<td>-0.006 561</td>
<td>1.93</td>
</tr>
</tbody>
</table>

### Table 4. Results for Example 4

<table>
<thead>
<tr>
<th>$h$</th>
<th>$e_h(1)$</th>
<th>$e_h^*(1)$</th>
<th>$E/h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{-2}$</td>
<td>-0.055 766</td>
<td>-0.030 374</td>
<td>0.41</td>
</tr>
<tr>
<td>$2^{-3}$</td>
<td>-0.030 568</td>
<td>-0.023 474</td>
<td>0.45</td>
</tr>
<tr>
<td>$2^{-4}$</td>
<td>-0.016 092</td>
<td>-0.014 165</td>
<td>0.49</td>
</tr>
<tr>
<td>$2^{-5}$</td>
<td>-0.008 262</td>
<td>-0.007 763</td>
<td>0.52</td>
</tr>
<tr>
<td>$2^{-6}$</td>
<td>-0.004 180</td>
<td>-0.004 056</td>
<td>0.51</td>
</tr>
<tr>
<td>$2^{-7}$</td>
<td>-0.002 088</td>
<td>-0.002 064</td>
<td>0.40</td>
</tr>
</tbody>
</table>

### References

Souhrn

ODHAD GLOBÁLNÍ CHYBY NUMERICKÉHO ŘEŠENÍ ZPOŽDĚNÍ DIFERENCIÁLNÍ ROVNICE EULEROVOU METODOU

ZDZISLAW JACKIEWICZ

V článku je použita metoda Zadunaiského k odhadu globální chyby vzniklé při numerickém řešení soustavy zpožděných diferenciálních rovnic Eulerovou metodou. Je uvedeno několik numerických příkladů.

Author’s address: Prof. Zdzislaw Jackiewicz, Department of Mathematics, University of Arkansas SE 301, Fayetteville, AR 72701, USA.