Numerical Approximation for Singular Second Order Differential Equations

David Benko[1], Daniel C. Biles[2], Mark P. Robinson[3] and John S. Spraker[3]

[1]University of South Alabama

[2]Belmont University

[3]Western Kentucky University

Abstract: We consider numerical approximation of solutions of singular second order differential equations. In particular, we study the backward (or implicit) Euler method. **We prove results concerning consistency, global error and stability. We show that the global error is linear with respect to the step size.** Numerical results are also given**, which demonstrate the linear convergence and compare the numerical results with known approximations.**

Keywords: numerical approximation, singular differential equations, backward Euler method, implicit Euler method

SECTION 1: INTRODUCTION

In [1] the authors proved local existence of solutions, and under additional assumptions, uniqueness of solutions for the initial value problem

$$y'' + p(t)y' + q(t, y(t)) = 0, \ t \in (0,1] \qquad \text{(IVP1)}$$

$$y(0) = \alpha, \ y'(0) = \beta,$$

in which p might be singular at $t = 0$. In this paper, we continue our study of this problem by investigating numerical approximation of solutions. In section 2, we consider application of the backward Euler method (also known as the implicit Euler method). That method seems to be ideal for (IVP1), since evaluation of p at $t = 0$ for that method is not needed. Based roughly on the approach in [2], we prove consistency, stability and also a global error estimate. In section 3, we supply numerical results for these theorems. **We conclude the article by further discussing the results. For further background, the reader can find that t**he backward Euler method has been applied in a number of recent papers, including, for example, [3], [4], [5], [6] and [7].

SECTION 2: BACKWARD EULER METHOD

We will apply the backward (or implicit) Euler method to (IVP1) to approximate solutions in the case in which p might be singular. In this section we make the following assumptions:

A1) $p \geq 0$ on $(0,1]$ (see note below)
A2) q is continuous on $[0,1] \times \mathbf{R}$ and for each $t \in [0,1]$, $q(t,\cdot)$ is Lipschitz continuous with Lipschitz constant L
A3) a unique solution u exists on $[0,1]$ (see note below)
A4) $u \in C^3[0,1]$.

Note that we could replace A1) by the weaker assumption that p is only bounded below and the following proof will still work. Note also that we may replace assumption A3) with the hypotheses of any existence and uniqueness theorem for such problems, for example, in the case $\beta = 0$ one could apply Theorem 4.3 of [8]. Also, the uniqueness in A3) need not be assumed explicitly - it follows from the results in this section.

Recall that the backward (or implicit) Euler method for the initial value problem $x' = f(t,x), x(t_0) = x_0$ is of the form

$$x_{n+1} = x_n + hf(t_{n+1}, x_{n+1})$$

where $h = t_{n+1} - t_n$ and x and f may be vector-valued. Note that it avoids evaluating f at the singularity $t_0 = 0$, but requires an approximation routine for solving implicit nonlinear algebraic equations. In this section, we prove consistency and derive global error and stability results. The notation and results of this section were originally motivated, in part, by [2]. Note that when applied to second order scalar differential equations, [2] requires that $p(t)$ is of the form $\frac{a+tb(t)}{t}$, where b is continuous. Hence, our assumptions cover cases such as $p(t) = \frac{1}{\sqrt{t}}$, which [2] does not.

We first set up the necessary notation. Converting (IVP1) to a system as usual (letting $x = y$ and $w = y'$), we obtain

$$
\begin{aligned}
x' &= w \\
w' &= -p(t)w - q(t,x) \\
x(0) &= \alpha \\
w(0) &= \beta.
\end{aligned}
\tag{IVP2}
$$

Let (u,v) represent the true solution to (IVP2) (where $u = x, v = w$).

We now construct the backward Euler numerical scheme for (IVP2). Divide the interval $[0,1]$ into N equal subintervals, where $N \in \mathbf{N}$ and let $h = 1/N$. Let $x_0 = \alpha$ and $w_0 = \beta$. For $i = 0,1,2,\ldots,N-1$, let

$$
\begin{aligned}
x_{i+1} &= x_i + hw_{i+1} \\
w_{i+1} &= w_i + h(-p(t_{i+1})w_{i+1} - q(t_{i+1},x_{i+1}))
\end{aligned}
\tag{1}
$$

and we thus have

$$\frac{x_{i+1}-x_i}{h} - w_{i+1} = 0 \tag{2a}$$

$$\frac{w_{i+1}-w_i}{h} - (-p(t_{i+1})w_{i+1} - q(t_{i+1},x_{i+1})) = 0. \tag{2b}$$

This leads to the following definition. Let $\mathbf{X} = (X_0,X_1,\ldots,X_N)$ and $\mathbf{W} = (W_0,W_1,\ldots,W_N)$ be arbitrary vectors. Define $F_h : \mathbf{R}^N \times \mathbf{R}^N \to \mathbf{R}^N \times \mathbf{R}^2$ by

$$F_h(\mathbf{X},\mathbf{W}) = \left(\frac{X_{i+1}-X_i}{h} - W_{i+1}, \frac{W_{i+1}-W_i}{h} - (-p(t_{i+1})W_{i+1} - q(t_{i+1},W_{i+1})) \right)$$

for $i = 0,1,\ldots,N-1$. Define $\mathbf{u} = (u(t_0),u(t_1),\ldots,u(t_N))$ and $\mathbf{v} = (v(t_0),v(t_1),\ldots,v(t_N))$ and let $\|F_h(\mathbf{u},\mathbf{v})\| = \max\limits_{i\in\{0,1,\ldots,N\}}\{|F_h(\mathbf{u},\mathbf{v})_{i,1}|,|F_h(\mathbf{u},\mathbf{v})_{i,2}|\}$. **Our first result is the following.**

Theorem 2.1 (consistency): There exists $F \geq 0$ such that $\|F_h(\mathbf{u},\mathbf{v})\| \leq Fh$ for any $N \in \mathbf{N}$.

Proof: Let $M_1 = \max\limits_{0\leq t\leq 1}|u''(t)|$ and $M_2 = \max\limits_{0\leq t\leq 1}|u'''(t)|$. We have, for $i = 0,1,\ldots N-1$ and for some $\tau_i \in [t_i,t_{i+1}]$

$$|F_h(\mathbf{u},\mathbf{v})_{i,1}| =$$
$$\left| \frac{u(t_{i+1}) - u(t_i)}{h} - v(t_{i+1}) \right| =$$
$$|u'(\tau_{i+1}) - u'(t_{i+1})| \le M_1 h.$$

Similarly, we can show $|F_h(\mathbf{u},\mathbf{v})_{i,2}| \le M_2 h$. We then have $\|F_h(\mathbf{u},\mathbf{v})\| \le Fh$, where $F = \max\{M_1, M_2\}$. ■

In order to prove a global error result, we will need to make use of the following two lemmas.

Lemma 2.1: Assume $h < 1/\sqrt{2L}$, where L is the Lipschitz constant from A2). For $i = 0,1,\ldots,N$, let $R_h(\mathbf{u},\mathbf{v})_i = (u(t_i), v(t_i))$ and let $(\varepsilon_i, \delta_i) = R_h(\mathbf{u},\mathbf{v})_i - (x_i, w_i)$. Then,

$$|\varepsilon_{i+1}| \le \frac{|\varepsilon_i| + h|\delta_i| + Fh^2 + Fh^3}{1 - Lh^2}, \tag{3}$$
$$|\delta_{i+1}| \le |\delta_i| + Fh^2 + Lh|\varepsilon_{i+1}|.$$

Proof: We have for $i = 0,1,\ldots,N$

$$x_i = R_h(\mathbf{u},\mathbf{v})_{i,1} - \varepsilon_i$$
$$w_i = R_h(\mathbf{u},\mathbf{v})_{i,2} - \delta_i.$$

Substituting into (2a), we have

$$0 = \frac{x_{i+1} - x_i}{h} - w_{i+1} =$$
$$\frac{-\varepsilon_{i+1} + \varepsilon_i}{h} + \delta_{i+1} + \frac{R_h(\mathbf{u},\mathbf{v})_{i+1,1} - R_h(\mathbf{u},\mathbf{v})_{i,1}}{h} - R_h(\mathbf{u},\mathbf{v})_{i+1,2} =$$
$$-\frac{\varepsilon_{i+1} - \varepsilon_i}{h} + \delta_{i+1} + F_h(\mathbf{u},\mathbf{v})_{i,1}. \tag{4}$$

Similarly, we have

$$0 = \frac{w_{i+1} - w_i}{h} - (-p(t_{i+1})w_{i+1} - q(t_{i+1}, x_{i+1})) =$$
$$-\frac{\delta_{i+1} - \delta_i}{h} - p(t_{i+1})\delta_{i+1} + F_h(\mathbf{u},\mathbf{v})_{i,2} - q(t_{i+1}, R_h(\mathbf{u},\mathbf{v})_{i+1,1}) + q(t_{i+1}, R_h(\mathbf{u},\mathbf{v})_{i+1,1} - \varepsilon_{i+1}). \tag{5}$$

Let $-l_{i+1}$ be the sum of the last three terms in (5). The right-hand side of (5) becomes

$$-\frac{\delta_{i+1}-\delta_i}{h}-p(t_{i+1})\delta_{i+1}-l_{i+1}. \tag{6}$$

Thus, (4) and (6) yield

$$-\frac{\varepsilon_{i+1}-\varepsilon_i}{h}+\delta_{i+1}+F_h(\mathbf{u},\mathbf{v})_{i,1}=0 \tag{7a}$$

$$-\frac{\delta_{i+1}-\delta_i}{h}-p(t_{i+1})\delta_{i+1}-l_{i+1}=0, \tag{7b}$$

for $i=0,1,2,\ldots,N-1$ and $\varepsilon_0=0,\ \delta_0=0$. Solving (7b) for δ_{i+1}, we obtain $\delta_{i+1}=\frac{\delta_i-hl_{i+1}}{1+hp(t_{i+1})}$, and substituting this expression into (7a) we obtain

$$\varepsilon_{i+1}=\varepsilon_i+h\frac{\delta_i-hl_{i+1}}{1+hp(t_{i+1})}+hF_h(\mathbf{u},\mathbf{v})_{i,1}. \tag{8}$$

Also, recalling the definition of l_{i+1}, we have (using Theorem 2.1 and assumption A2)

$$\begin{aligned}|l_{i+1}| &= |F_h(\mathbf{u},\mathbf{v})_{i,2}-q(t_{i+1},R_h(\mathbf{u},\mathbf{v})_{i+1,1})+q(t_{i+1},R_h(\mathbf{u},\mathbf{v})_{i+1,1}-\varepsilon_{i+1})|\\ &\le Fh+L|\varepsilon_{i+1}|.\end{aligned} \tag{9}$$

We now have from (8), A1), Theorem 2.1 and (9)

$$\begin{aligned}|\varepsilon_{i+1}| &= \left|\varepsilon_i+h\frac{\delta_i-hl_{i+1}}{1+hp(t_{i+1})}+hF_h(\mathbf{u},\mathbf{v})_{i,1}\right|\\ &\le |\varepsilon_i|+h|\delta_i-hl_{i+1}|+h|F_h(\mathbf{u},\mathbf{v})_{i,1}|\\ &\le |\varepsilon_i|+h|\delta_i|+h^2|l_{i+1}|+Fh^2\\ &\le |\varepsilon_i|+h|\delta_i|+h^2(Fh+L|\varepsilon_{i+1}|)+Fh^2.\end{aligned}$$

Solving the above for $|\varepsilon_{i+1}|$, we obtain

$$|\varepsilon_{i+1}|\le\frac{|\varepsilon_i|+h|\delta_i|+Fh^2+Fh^3}{1-Lh^2},$$

where we used the assumption that $h < 1/\sqrt{2L}$. Similarly, we can derive a corresponding inequality for $|\delta_{i+1}|$:

$$|\delta_{i+1}| \leq |\delta_i| + Fh^2 + Lh|\varepsilon_{i+1}|,$$

establishing Lemma 2.1. ■

Lemma 2.2: Let $\alpha_1 = \alpha_1(h) = \frac{1}{1-Lh^2}$ for $h \in (0,H]$, where $H \equiv \min\left\{1, 1/\sqrt{2L}\right\}$. Choose $G > 0$ such that $\alpha_1 + \alpha_1 h \leq 1 + Gh$ and $1 + Lh\alpha_1 + Lh^2\alpha_1 \leq 1 + Gh$ for all $h \in (0,H]$. Define $F_2 = F(1 + GH)$. Then,

$$|\varepsilon_i|, |\delta_i| \leq h\frac{F_2}{G}[(1+Gh)^i - 1]$$

for all $i = 0, 1, \ldots, N$.

Proof: Define $z_0 = 0$ and $z_{i+1} = (1+Gh)z_i + F_2h^2$ for $i = 0, 1, \ldots, N-1$. Solving this recursion relation, we obtain $z_i = \frac{(1+Gh)^i - 1}{G}F_2h$. We now show inductively that $|\varepsilon_i|, |\delta_i| \leq z_i$ for $i = 0, 1, \ldots, N$. Note that $|\varepsilon_0|, |\delta_0| = 0 = z_0$. Assume for some $k \in \{0, 1, \ldots, N-1\}$ that $|\varepsilon_k|, |\delta_k| \leq z_k$. We now have, from Lemma 2.1 and the definition of α_1

$$\begin{aligned} |\varepsilon_{i+1}| &\leq \alpha_1(|\varepsilon_i| + h|\delta_i| + Fh^2 + Fh^3) \\ &\leq \alpha_1(z_i + hz_i) + \alpha_1(F + Fh)h^2 \\ &\leq (1+Gh)z_i + F_2h^2 = z_{i+1}. \end{aligned}$$

We also have using Lemma 2.1

$$\begin{aligned} |\delta_{i+1}| &\leq |\delta_i| + Lh|\varepsilon_{i+1}| + Fh^2 \\ &\leq |\delta_i| + Lh\alpha_1(|\varepsilon_i| + h|\delta_i| + Fh^2 + Fh^3) + Fh^2 \\ &\leq z_i + Lh\alpha_1(z_i + hz_i) + Lh\alpha_1(Fh^3 + Fh^2) + Fh^2 \\ &\leq (1+Gh)z_i + F_2h^2 = z_{i+1}. \quad ■ \end{aligned}$$

Corollary 2.1: $|\varepsilon_i|, |\delta_i| \leq h\frac{F_2}{G}[e^G - 1]$ for $i = 0, 1, \ldots, N$, where G and F_2 are as specified in Lemma 2.2.

Proof: From Lemma 2.2, we have
$|\varepsilon_i| \leq h\frac{F_2}{G}[(1+Gh)^N - 1] = h\frac{F_2}{G}\left[(1+\frac{G}{N})^N - 1\right] \leq h\frac{F_2}{G}[e^G - 1]$
and the same for $|\delta_N|$. ■

From Corollary 2.1, we easily obtain a global error result.

Corollary 2.2 (global error): Let $H \equiv \min\left\{\sqrt{\frac{1}{2L}}, 1\right\}$. Then, $|\varepsilon_i| \leq K_1 h$ and $|\delta_i| \leq K_2 h$ for $i = 0, 1, \ldots, N$, where K_1, K_2 are independent of h and i, for any $h \in (0, H)$.

The next question we address is that of stability for this method.

Theorem 2.2 (stability): Let $\gamma_1, \gamma_2 \in \mathbf{R}$ be given. Let $\psi_0 = \alpha + \gamma_1$, $\omega_0 = \beta + \gamma_2$ and for $i = 0, 1, \ldots, N-1$, let

$$\frac{\psi_{i+1} - \psi_i}{h} - \omega_{i+1} = 0,$$

$$\frac{\omega_{i+1} - \omega_i}{h} + p(t_{i+1})\omega_{i+1} + q(t_{i+1}, \psi_{i+1}) = 0.$$

Then, there exists a $K_3 \geq 0$ such that for all $i = 0, 1, 2, \ldots, N$

$$|\psi_i - x_i| \leq K_3 \max(|\gamma_1|, |\gamma_2|),$$

$$|\omega_i - w_i| \leq K_3 \max(|\gamma_1|, |\gamma_2|),$$

where x_i, w_i are the approximations defined earlier, with $x_0 = \alpha$, $w_0 = \beta$.

Proof: For $i = 0, 1, \ldots, N$, let $a_i = \psi_i - x_i$ and $b_i = \omega_i - w_i$. Then, we have $a_0 = \gamma_1$, $b_0 = \gamma_2$,

$$\frac{a_{i+1} - a_i}{h} - b_{i+1} = 0, \text{ and}$$

$$\frac{b_{i+1} - b_i}{h} + p(t_{i+1})b_{i+1} + q(t_{i+1}, \psi_{i+1}) - q(t_{i+1}, x_{i+1}) = 0,$$

for $i = 0, 1, \ldots, N-1$.

By repeating the procedure in Lemma 2.1, we obtain similar inequalities for a_{i+1} and b_{i+1}:

$$|a_{i+1}| \leq \frac{|a_i| + h|b_i|}{1 - h^2 L}.$$

$$|b_{i+1}| \leq |b_i| + Lh|a_{i+1}|.$$

Now define H and G as in Lemma 2.2. Define $z_0 = \max(|\gamma_1|, |\gamma_2|)$ and $z_{i+1} = (1+Gh)z_i$ for $i = 0, 1, \ldots N-1$, thus $z_i = (1+Gh)^i z_0$. As in the proof of Lemma 2.2, we use induction to obtain

$$|a_i|, |b_i| \leq z_i$$

for all $i = 0, 1, \ldots, N$ and hence

$$|a_i| \leq (1+Gh)^i z_0 =$$

$$(1 + \frac{G}{N})^i \max(|\gamma_1|, |\gamma_2|) \leq$$

$$e^G \max(|\gamma_1|, |\gamma_2|)$$

and similarly $|b_i| \leq e^G \max(|\gamma_1|, |\gamma_2|)$. ■

SECTION 3: NUMERICAL RESULTS FOR THE BACKWARD EULER METHOD

We now provide some numerical results for the backward Euler method discussed in the previous section. **The first of these verifies the linear order of convergence.**

Example 3.1: We consider the nonlinear Lane-Emden equation with $n = 5$ [9]:

$$y'' + \frac{2}{t}y' + y^5 = 0,\ t \in (0,1]$$

$$y(0) = 1,\ y'(0) = 0.$$

The solution to this initial value problem is well known to be $u(t) = \sqrt{\frac{1}{1+\frac{1}{3}t^2}}$. We compare our approximations with the true solution for several values of the step size h, rounding to five decimal places **(see Table 3.1)**.

t	error for $h = .1$	error for $h = .05$	error for $h = .025$	approximation for $h = .025$
0	0	0	0	1
0.1	0.00162	0.00082	0.00041	0.99793
0.2	0.00308	0.00157	0.00079	0.99261
0.3	0.00430	0.00220	0.00111	0.98422
0.4	0.00522	0.00269	0.00136	0.97299
0.5	0.00581	0.00300	0.00153	0.95924
0.6	0.00605	0.00314	0.00160	0.94331
0.7	0.00596	0.00311	0.00159	0.92556
0.8	0.00558	0.00292	0.00150	0.90635
0.9	0.00496	0.00261	0.00134	0.88602
1.0	0.00415	0.00218	0.00112	0.86490

Table 3.1: Confirmation of convergence of order h

One can clearly see the convergence of order h as predicted by the results in the previous section.

The next two examples compare our numerical results with known approximate solutions, providing some confirmation of its accuracy.

Example 3.2: For the Lane-Emden equation with $n = 3$:

$$y'' + \frac{2}{t}y' + y^3 = 0, \ t \in (0,1]$$

$$y(0) = 1, \ y'(0) = 0,$$

a closed-form solution is not known, but in [10] the approximate closed-form solution $u(t) = \text{sech}\left(\frac{t}{\sqrt{3}}\right)$ is derived. We compare our approximation technique to this one for $h = .025$ **in Table 3.2.**

t	backward Euler method	Beech's approximation
0	1	1
0.1	0.99792	0.99834
0.2	0.99257	0.99337
0.3	0.98403	0.98519
0.4	0.97248	0.97391
0.5	0.95811	0.95973
0.6	0.94116	0.94286
0.7	0.92190	0.92355
0.8	0.90061	0.90206
0.9	0.87758	0.87868
1.0	0.85312	0.85372

Table 3.2: Comparison of the backward Euler method with Beech's approximation

Example 3.3: For the Lane-Emden equation with $n = 1.5$:

$$y'' + \frac{2}{t}y' + y^{1.5} = 0,\ t \in (0,1]$$

$$y(0) = 1,\ y'(0) = 0,$$

a closed-form solution is not known, but in [11] the approximate closed-form solution $u(t) = \exp\left(\frac{-t^2}{6}\right)$ is derived. We compare our approximation technique to this one for $h = .025$ **in Table 3.3.**

t	backward Euler method	Fowler and Hoyle's approximation
0	1	1
0.1	0.99792	0.99834
0.2	0.99253	0.99336
0.3	0.98389	0.98511
0.4	0.97208	0.97369
0.5	0.95721	0.95919
0.6	0.93940	0.94177
0.7	0.91883	0.92158
0.8	0.89566	0.89883
0.9	0.87009	0.89372
1.0	0.84233	0.84648

Table 3.3: Comparison of the backward Euler method with Fowler and Hoyle's approximatio

SECTION 4: DISCUSSION

The backward Euler method is a natural choice for approximating solutions to initial value problems with a singularity at the intial time, since this scheme does not require evaluation at the initial time. The method also has the advantages of being well-known and having been applied in the past to a number of problems. We have applied this technique successfully with initial value problems of the form (IVP1), deriving results regarding consistency, global error and stability, which are the standard fundamental questions for numerical techniques for IVP's. The hypotheses for the result, given in A1-A4, are quite weak. The numerical results given in Section 3 help to verify the global error and the accuracy of the method. We suggest that the next direction for future research is to investigate application of other implicit approximation techniques to singular initial value problems of the form (IVP1). One could also study the effect of replacing the $p(t)y' + q(t,y(t))$ **terms with something of the form** $p(t)r(y') + q(t,y(t))$**,** $p(t,y') + q(t,y(t))$ **or or a single term** $q(t,y(t),y'(t))$**. Another question to address is that of allowing a singularity in the state variable at the initial state.**

REFERENCES

1. Biles, D. C., M. P. Robinson and J. S. Spraker, A generalization of the Lane-Emden equation, Journal of Mathematical Analysis and Applications, Vol. 273, 2002, p. 654-666.

2. Koch, O., P. Kofler and E. B. Weinmüller, The implicit Euler method for the numerical solution of singular initial value problems, Applied Numerical Mathematics, Vol. 34, 2000, p. 231-252.

3. Baňas, L., A numerical method for the Landau-Lifshitz equation with magnetostriction, Math. Methods Appl. Sci. Vol. 28, no. 16, 2005, p. 1939–1954.

4. Blajer, W. and K. Kolodziejczyk, A geometric approach to solving problems of control constraints: theory and a DAE framework, Multibody Syst. Dyn. Vol. 11, no. 4, 2004, p. 343–364.

5. Coyette, J.-P., K. Meerbergen and M. Robbé, Time integration for spherical acoustic finite-infinite element models, Internat. J. Numer. Methods Engrg. Vol. 64, no. 13, 2005, p. 1752–1768.

6. Cuesta, E. and C. Palencia, A numerical method for an integro-differential equation with memory in Banach spaces: qualitative properties, SIAM J. Numer. Anal., Vol. 41, no. 4, 2003, p. 1232–1241.

7. Smaoui, N., Nonlinear boundary control of the generalized Burger's equation, Nonlinear Dynam. Vol. 37, no. 1, 2004, p. 75–86.

8. Koch, O., P. Kofler and E. B. Weinmüller, Initial value problems for systems of ordinary first and second order differential equations with a singularity of the first kind, Analysis, Vol. 21, 2001, p. 373-389.

9. Chandrasekhar, S., "An Introduction to the Study of Stellar Structure," University of Chicago Press, Chicago, 1939.

10. Beech, M., An approximate solution for the polytrope $n = 3$, Astrophysics and Space Science, Vol. 132, 1987, p. 393-396.

11. Fowler, W., and F. Hoyle, Neutrino processes and pair formation in massive stars and supernovae, Astrophys. J. Suppl., Vol. 9, 1964, p. 201-319.

Corresponding author's address:
Daniel C. Biles

Department of Mathematics
Belmont University
1900 Belmont Blvd.
Nashville, TN 37212

e-mail addresses:
dbenko2007@yahoo.com
bilesd@math.belmont.edu
mark.robinson@wku.edu
john.spraker@wku.edu