Electronic Transactions on Numerical Analysis.
Volume 34, pp. 14-19, 2008.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

A NOTE ON NUMERICALLY CONSISTENT INITIAL VALUES FOR HIGH INDEX DIFFERENTIAL-ALGEBRAIC EQUATIONS*

CARMEN ARÉVALO†

Dedicated to Víctor Pereyra on the occasion of his 70th birthday

Abstract. When differential-algebraic equations of index 3 or higher are solved with backward differentiation formulas, the solution can have gross errors in the first few steps, even if the initial values are equal to the exact solution and even if the stepsize is kept constant. This raises the question of what are *consistent* initial values for the difference equations. Here we study how to change the exact initial values into what we call *numerically consistent* initial values for the implicit Euler method.

Key words. high index differential-algebraic equations, consistent initial values

AMS subject classifications. 65L05

1. Introduction.

A differential-algebraic equation (DAE) has the form

$$F(t, x, \dot{x}) = 0,$$

where the matrix $\partial F/\partial \dot{x}$ is singular. Here we shall consider the differential-algebraic equation of the form

$$\dot{p} = U(t, v) \tag{1.1a}$$
$$\dot{v} = F(t, p, v) + G(t, p, v)\,\lambda \tag{1.1b}$$
$$0 = R(t, p), \tag{1.1c}$$

where $t \in \mathbb{R}$, $p \in \mathbb{R}^n$, $v \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^s$, and $U : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^n$, $F : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $G : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{m\times s}$ and $R : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^s$. Assume that $s \leq \min(m, n)$ in order to avoid an over-determined system. The system (1.1) has index 3 if $\partial R/\partial p \cdot \partial U/\partial v \cdot G$ is nonsingular for all $t \in [t_0, T]$. For simplicity we take $t_0 = 0$.

The algebraic variables appear linearly as in (1.1b) in important classes of physical problems. This condition is fulfilled, for example, by the Euler-Lagrange equations of multibody mechanics, which have applications in biomechanics, the dynamics of machinery, robotics, and vehicle design.

To solve this type of DAE, several techniques have been considered. One proposition has been to solve the system in its original formulation using a backward differentiation formula (BDF), as implemented in the DAE solver DASSL [4]; but such a variable-step-size, variable-order code based on BDF methods presents some essential difficulties when solving higher index DAEs, especially in the accuracy of the algebraic variables [2].

The initial values (p_0, v_0, λ_0) are said to be *consistent* if the DAE has a differentiable solution $(p(t), v(t), \lambda(t))$ in the interval $[0, T]$ such that $(p(0), v(0), \lambda(0)) = (p_0, v_0, \lambda_0)$.

Brenan and Engquist [3] defined *numerically consistent starting values to order* $k+1$ for the k-step BDF applied to (1.1) as starting values $[p_{k-1}, \ldots, p_0]$, $[v_{k-1}, \ldots, v_0]$, $[\lambda_{k-1}, \ldots, \lambda_0]$

* Received March 31, 2008. Accepted August 6, 2008. Published online on December 13, 2008. Recommended by Godela Scherer.

†Numerical Analysis, Centre for Mathematical Sciences, Lund University, Box 118, SE-221 00 Lund, Sweden. (Carmen.Arevalo@na.lu.se)

ETNA
Kent State University
etna@mcs.kent.edu

such that

$$\|p_j - p(t_j)\| \le K_1 h^{k+1} \tag{1.2a}$$
$$\|v_j - v(t_j)\| \le K_2 h^{k+1} \tag{1.2b}$$
$$\|R(t_j, p_j)\| \le K_3 h^{k+2} \tag{1.2c}$$

for some constants K_1, K_2, K_3 and $j = 0, 1, \ldots, k-1$. They proved that the k-step BDF ($k = 1, \ldots, 6$) with constant step size h converges globally with $O(h^k)$ accuracy to the solutions, but only after after $k+1$ steps, and provided the method uses numerically consistent starting values to order $k+1$.

In particular, for the implicit Euler method the state variables p and v have $O(h)$ accuracy after the first step, but the error in the algebraic variable λ is $O(1)$, even when the initial values are exact. By approximating the errors in the algebraic variable, a correction mechanism was devised in order to obtain $O(h)$ accuracy even in the initial k steps [2]. These formulas correct the errors locally and produce $O(h)$ accurate algebraic variables.

Our view is that these $O(1)$ errors are caused by initial values that are inconsistent with the difference equations. We attempt to redefine *numerically consistent* initial values as initial values that are *consistent with the difference equations*, as opposed to those consistent with the differential equation, as in (1.2). We explain the $O(1)$ errors in the algebraic variables as a result of starting the BDF solver with initial values that are consistent with the differential equation, but not with the difference equations. Once this is established, we develop a scheme to construct numerically consistent initial values. We illustrate our results by solving the same problems that appear in [1].

2. Numerically consistent initial values. Initial values that are numerically consistent with the difference equation generated by an order p method should produce $O(h^p)$ accurate solutions.

DEFINITION 2.1. *The set* $(x_0, x_1, \ldots, x_{k-1})$ *is a* numerically consistent initial value *at* $t = t_0$ *for a differential equation* $f(t, x, \dot{x}) = 0$*, solved with a* k*-step multistep method of order* p*, if the method initiated with* $(x_0, x_1, \ldots, x_{k-1})$ *generates the approximation* x_k *such that*

$$x(t_k) - x_k = O(h^p). \tag{2.1}$$

For differential-algebraic equations (DAEs) in semi-explicit form

$$\dot{x} = f(t, x, \lambda)$$
$$0 = g(t, x, \lambda)$$

of index less than 2, consistent initial conditions are also numerically consistent; but for higher index DAEs, consistent initial conditions do not necessarily produce an approximation satisfying (2.1). For index 3 DAEs, in fact, BDF methods approximate the algebraic variables with $O(1)$ errors after the first step, even if the initial values are exact.

3. Euler-Lagrange equations and the implicit Euler formula. The index 3 formulation of constrained multibody systems is

$$\dot{p} = v \tag{3.1a}$$
$$M\dot{v} = F(p, v) + G(p)^T \lambda \tag{3.1b}$$
$$0 = R(p), \tag{3.1c}$$

ETNA
Kent State University
etna@mcs.kent.edu

where M is the positive definite mass matrix, $G = \partial R/\partial p$, and $G(p)M^{-1}G(p)^T$ is non-singular.

To simplify notation, we will denote $G(p_1)$ by G_1 and $F(p_1, v_1)$ by F_1. When the implicit Euler formula is applied to (3.1) using exact initial values, the algebraic variable λ_1 has an $O(1)$ error $\epsilon^\lambda = \lambda(t_1) - \lambda_1$ that can be approximated with $O(h)$ accuracy by [2]

$$\delta\lambda = (G_1M^{-1}G_1^T)^{-1}G_1M^{-1}F_1 + \lambda_1.$$

The implicit Euler method requires initial values only for state variables p and v. Taking the initial condition

$$p_0 = p(0) + O(h^3) \tag{3.2a}$$
$$v_0 = v(0) - hM^{-1}G_1^T\delta\lambda + O(h^2), \tag{3.2b}$$

the first step is

$$\begin{aligned} p_1 &= p_0 + hv_1 \\ v_1 &= v_0 + h(I - B)M^{-1}F_1 + hM^{-1}G_1^T\delta\lambda \\ 0 &= R(p_1), \end{aligned}$$

where the matrix B is dependent on p_1 and is defined by

$$B = M^{-1}G_1^T(G_1M^{-1}G_1^T)^{-1}G_1.$$

The matrix B is a projector and $B \cdot M^{-1}G_1^T = M^{-1}G_1^T$.

We denote the global errors at t_1 by $\epsilon^p = p(t_1) - p_1$, $\epsilon^v = v(t_1) - v_1$ and $\epsilon^\lambda = \lambda(t_1) - \lambda_1$. The $O(h)$ accurate correction term $\delta\lambda$ may be expressed in terms of the exact value of the algebraic variable as

$$\delta\lambda = \frac{1}{2}((G_1M^{-1}G_1^T)^{-1}G_1M^{-1}F_1 + \lambda(t_1)).$$

By expanding $p(t)$ and $v(t)$ about t_1 and evaluating at $t = 0$ we obtain

$$\begin{aligned} \epsilon^v &= hM^{-1}G_1^T(\delta\lambda + \epsilon^\lambda) + O(h^2) \\ \epsilon^p &= h\epsilon^v - \frac{h^2}{2}M^{-1}(F(p(t_1), v(t_1)) + G(p(t_1))^T\lambda(t_1)) + O(h^3). \end{aligned}$$

These expressions lead to

$$\epsilon^p = \frac{h^2}{2}M^{-1}(G_1^T(G_1M^{-1}G_1^T)^{-1}G_1M^{-1}F_1 - F(t_1)) + h^2M^{-1}G_1^T\epsilon^\lambda + O(h^3),$$

and multiplying by G_1 we can conclude that

$$G_1\epsilon^p = h^2G_1M^{-1}G_1^T\epsilon^\lambda + O(h^3).$$

If we expand $R(p)$ about p_1 and use the fact that $\epsilon^p = O(h^2)$ together with $R(p(t_1)) = 0$ and $R(p_1) = 0$, we finally obtain $G_1\epsilon^p = O(h^4)$, thus implying that

$$\epsilon^\lambda = O(h).$$

In other words, the initial condition (3.2) is numerically consistent for (3.1) solved with the implicit Euler method.

ETNA
Kent State University
etna@mcs.kent.edu

The initial value (3.2b) may be written as

$$v_0 = v(0) + B(v(0) - v_1) + O(h^2). \tag{3.3}$$

Equation (3.3) in the form

$$\frac{v_0 - v(0) + O(h^2)}{h} = -B\frac{(v_1 - v(0))}{h} \tag{3.4}$$

can be interpreted as an indication that the initial values for v consistent with the differential equation are also consistent with the difference equation only in the case that $\dot{v}(0)$ is in the null space of G_1. In practical terms, this would require that $\lambda(t_0) = -(G_0 M^{-1} G_0^T)^{-1} G_0 M^{-1} F_0 + O(h)$.

As the change in initial values for the state variable v is $O(h)$, it will not affect the accuracy of the solution in any of the state variables.

4. The more general case. The more general problem we consider here is

$$\dot{p} = U(t, q) \tag{4.1a}$$
$$\dot{q} = F(t, p, q) + G(t, p, q)\Lambda \tag{4.1b}$$
$$0 = R(t, p). \tag{4.1c}$$

This is an index 3 DAE if $R_p \cdot U_q \cdot G$ is non-singular.

According to [2], the algebraic variable Λ is estimated by the first step of the implicit Euler method with an $O(1)$ error $\epsilon^\Lambda = \Lambda(t_1) - \Lambda_1$, which is approximated with $O(h)$ accuracy by

$$\delta\Lambda = (R_p U_q G)^{-1} R_p [U_q (F + G\Lambda) U_t],$$

where all functions are evaluated at $t = t_1$. The first implicit Euler step is

$$\begin{aligned} p_1 &= p_0 + hU(t_1, q_1) \\ q_1 &= q_0 + h(F(t_1, p_1, q_1) + G(t_1, p_1, q_1)\Lambda_1) \\ 0 &= R(t_1, p_1) \end{aligned}$$

with numerically consistent initial condition

$$p_0 = p(0) + O(h^3) \tag{4.2a}$$
$$q_0 = q(0) - hG\delta\Lambda + O(h^2). \tag{4.2b}$$

We may write (4.2b) as

$$q_0 = q(0) - AU_q(q_1 - q(0)) - hAU_t + O(h^2)$$

with $A = G(R_p U_q G)^{-1} R_p$. The matrix AU_q is a projector with $AU_q G = G$.

The last equation can be written as

$$\frac{q_0 - q(0)}{h} = -AU_q \frac{q_1 - q(0)}{h} - AU_t + O(h),$$

mirroring equation (3.4) when U is independent of t. In this case, the numerically consistent initial values for q are consistent with the differential equation if $AU_q \cdot \dot{q} = 0$, that is, only if $\dot{q}$ is in the null space of AU_q. In practical terms, this would require that $\Lambda(0) = -(R_p U_q G)^{-1} R_p U_q F \mid_{t=0}$.

ETNA
Kent State University
etna@mcs.kent.edu

5. Computational results. We illustrate our results with the following two index 3 DAEs from [1].

PROBLEM 5.1. The system of Euler-Lagrange equations

$$\ddot{x} = 2y + x\lambda \tag{5.1a}$$
$$\ddot{y} = -2x + y\lambda \tag{5.1b}$$
$$0 = x^2 + y^2 - 1 \tag{5.1c}$$

has the solution

$$x = \sin(1+t)^2 \qquad y = \cos(1+t)^2$$
$$u = 2(1+t)\cos(1+t)^2 \qquad v = -2(1+t)\sin(1+t)^2$$
$$\lambda = -4(1+t)^2.$$

PROBLEM 5.2. This problem is defined in the form (4.1) with

$$p = \begin{bmatrix} x & y & z \end{bmatrix}^T$$
$$q = \begin{bmatrix} u & v & w \end{bmatrix}^T$$
$$\Lambda = \begin{bmatrix} \lambda & \beta \end{bmatrix}^T$$
$$U = \begin{bmatrix} 2u & v & w-1 \end{bmatrix}^T$$
$$F = \begin{bmatrix} -y & 2x + y\sin t^2 - 4yt & 4zt^2 + 0.5\sin t^2 \end{bmatrix}^T$$
$$G = \begin{bmatrix} x & 0 & 2z \\ 0 & 2y & 1 \end{bmatrix}^T$$
$$R = \begin{bmatrix} x^2 + y^2 + z^2 - 1 & z - 0.5 \end{bmatrix}^T.$$

Its solution is

$$x = \frac{\sqrt{3}}{2}\cos t^2 \qquad y = \frac{\sqrt{3}}{2}\sin t^2 \qquad z = 0.5$$
$$u = -\frac{\sqrt{3}}{2}t\sin t^2 \qquad v = \sqrt{3}t\cos t^2 \qquad w = 1$$
$$\lambda = -2t^2 \qquad \beta = -0.5\sin t^2.$$

Problems 5.1 and 5.2 were solved with implicit Euler using $h = 0.0005$ and $h = 0.001$. After the first step the algebraic variable λ shows $O(1)$ errors. Problem 5.1 was solved with $t_0 = 0$, so $G(P_0)^T\lambda_0 = [\sin(1) \;\; \cos(1)]\,\lambda_0$. Therefore, $\lambda(0)$ would have to be close to zero (but is instead equal to -4) for the exact initial values to be numerically consistent. The numerically consistent initial values were calculated from (3.3) and the step was redone. The exact initial values $\begin{bmatrix} u & v \end{bmatrix} = \begin{bmatrix} 1.0806 & -1.6829 \end{bmatrix}$ were changed to $\begin{bmatrix} 1.0814 & -1.6824 \end{bmatrix}$ for $h = 0.0005$ and to $\begin{bmatrix} 1.0823 & -1.6819 \end{bmatrix}$ for $h = 0.0010$. This resulted in algebraic variables of $O(h)$ accuracy (5.1). For Problem 5.2 with $t_0 = 1$ and $h = 0.001$, the initial values for q were changed from $\begin{bmatrix} -0.72874 & 0.93583 & 1 \end{bmatrix}$ to $\begin{bmatrix} -0.72985 & 0.93931 & 1 \end{bmatrix}$, producing algebraic variables of $O(h)$ accuracy, as shown in Table 5.2.

The results presented in Figure 5.1 show the $O(1)$ errors when Problem 5.1 was solved using exact initial values, and the $O(h)$ errors in the algebraic variable obtained by using numerically consistent initial values.

ETNA
Kent State University
etna@mcs.kent.edu

TABLE 5.1

Absolute errors in the algebraic variable λ for Problem 5.1 using implicit Euler with step sizes 0.0005 and 0.0010. $O(1)$ errors and their corrected values are displayed in bold face.

t_j	$h = 0.0005$		$h = 0.001$	
	exact IV	numerically consistent IV	exact IV	numerically consistent IV
0.0005	**2.0040**	**0.004030**		
0.0010	0.0040085	0.0040085	**2.0080**	**0.0080120**
0.0015	0.0040185	0.0040185		
0.0020	0.0040286	0.0040286	0.0080341	0.0080341

TABLE 5.2

Absolute errors in the algebraic variable λ for Problem 5.2 using implicit Euler with step sizes 0.0005 and 0.0010. $O(1)$ errors and their corrected values are displayed in bold face.

t_j	$h = 0.0005$		$h = 0.001$	
	exact IV	numerically consistent IV	exact IV	numerically consistent IV
0.0005	**2.3973**	**0.0047995**		
0.0010	0.0056125	0.0056125	**2.3917**	**0.009586**
0.0015	0.0055573	0.0055573		
0.0020	0.0055028	0.0055028	0.011062	0.011062

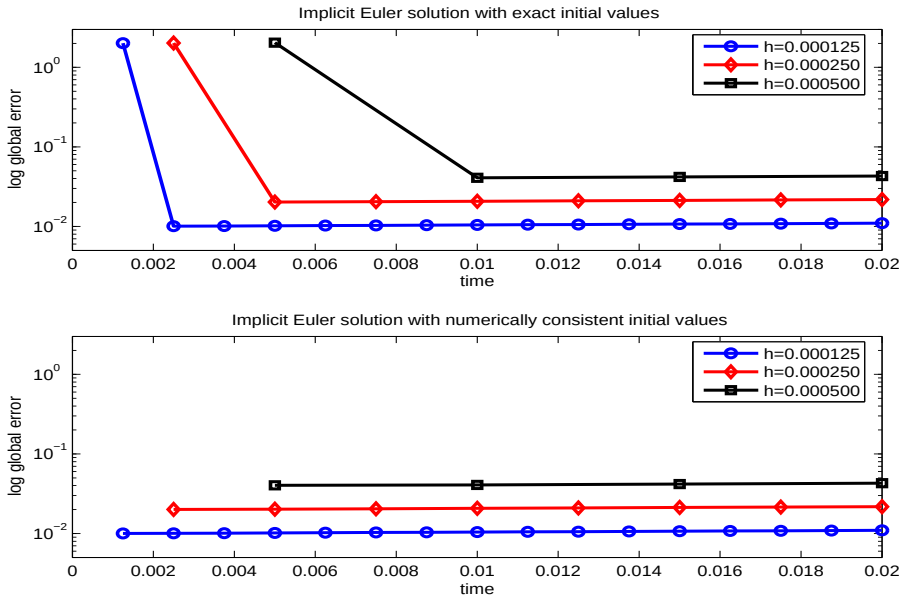

FIG. 5.1. *Implicit Euler used to solve an index 3 DAE test problem (see text) using numerically consistent initial values. Global errors with exact initial values (top) and numerically consistent initial values (bottom).*

REFERENCES

[1] C. ARÉVALO, *Matching the structure of DAEs and multistep methods*, Ph. D. thesis, Department of Computer Science, Lund University, Lund, Sweden, 1993.

[2] C. ARÉVALO AND P. LÖDSTEDT, *Improving the accuracy of BDF methods for index 3 differential-algebraic equations*, BIT, 35 (1995), pp. 297–308.

[3] K. BRENAN AND B. ENGQUIST, *Backward differentiation approximations of nonlinear differential/algebraic systems*, Math. Comp., 51 (1988), pp. 659–676.

[4] L. PETZOLD, *Differential/algebraic equations are not ODEs*, SIAM J. Sci. Statist. Comput., 3 (1982), pp. 367–384.