Defect-based a posteriori error estimation for index 1 DAEs with a singularity of the first kind

W. Auzinger, H. Lehner and E. Weinmüller

Institute for Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstrasse 8–10/101, 1040 Wien, Austria

Abstract. We show how the QDeC estimator, an efficient and asymptotically correct a posteriori estimator for collocation solutions to ODEs can be extended to differential algebraic equations (DAEs), where a singularity of the first kind is admitted.

Keywords: Error estimation, differential algebraic equations, singularity of the first kind, collocation, defect correction
PACS: 02.60.Cb, 02.60.Lj, 02.70.Jn

1. PROBLEM SETTING

We consider linear systems of DAEs of index 1,

$$A(t)(D(t)x(t))' + B(t)x(t) = g(t), \quad t \in (0,T], \tag{1}$$

with appropriately smooth data $A(t) \in \mathbf{R}^{m\times n}$, $B(t) \in \mathbf{R}^{m\times m}$ and constant $D(t) = D \in \mathbf{R}^{n\times m}$. For our analysis we assume that (1), augmented with a correct number or initial conditions, is well-posed and admits a smooth solution $x^*(t)$. We assume

$$m > n, \quad \ker A(t) = \{0\} \ \ \forall\, t \in (0,T], \quad \text{and} \quad \operatorname{im} D = \mathbf{R}^n. \tag{2}$$

Conditions (2) imply that $(AD)(t) \in \mathbf{R}^{m\times m}$ is singular, with $\operatorname{rank}(AD)(t) \equiv n$ for $t \in (0,T]$.

The requirement that $D(t)$ be constant is not a real restriction, as any such system with varying $D(t)$ can be rewritten by introducing a new variable $u(t) = D(t)x(t)$, resulting in a dilated system of the type (1) for $\hat{x}(t) := (x(t),u(t))^T$ with a constant matrix $\hat{D}(t) \equiv \hat{D}$, see [4].

We consider collocation solutions $p(t)$ for (1), defined by the relations

$$A(t_{ij})(Dp)'(t_{ij}) + B(t_{ij})p(t_{ij}) = g(t_{ij}), \tag{3}$$

where $p(t)$ is represented by a polynomial $p_i(t)$ of degree $\le s$ on each subinterval $[\tau_i, \tau_{i+1}]$, and

$$p_{i-1}(\tau_i) = p_i(\tau_i), \quad \tau_0 = 0,\ h_i := \tau_{i+1} - \tau_i > 0,\ \tau_N = T, \quad t_{ij} := \tau_i + c_j h_i,\ 0 < c_1 < \ldots c_j < \ldots < c_s = 1, \tag{4}$$

for $i = 0\ldots N-1$, $j = 1\ldots s$. Note, in particular, that $c_s = 1$ is essential for our analysis. We also assume that s is even, which will be necessary to guarantee the asymptotic correctness of our error estimator to be defined in Section 2. We also denote $h := \max_{i=0\ldots N-1} h_i$.

Index 1 systems of the type (1) can be decomposed into an *inherent ODE* and a purely algebraic equation (cf. (11)–(14) below). In this paper, we will specifically address DAEs involving a singularity in the differential component. In particular, the inherent ODE is assumed to take the form

$$y'(t) = \frac{M(t)}{t}\, y(t) + f(t), \quad t \in (0,T], \tag{5}$$

featuring a singularity of the first kind, with sufficiently smooth data $M(t)$ and $f(t)$, see [6]. The application of collocation methods to such a class of DAEs was studied in [7].

Note: In [4], the non-singular problem type was considered. In particular, a detailed proof of Theorem 3.1 below is given in [4] for the case of an index 1 DAE with a regular inherent ODE.

2. DEFECT-BASED ERROR ESTIMATION

In the context of regular and singular ODEs, a method for computing an a-posteriori estimate ε of the global error $e := p - x^*$ of a polynomial collocation solution p was proposed in [2] and implemented in [1]. This method is based on the defect correction principle [10, 11]. In particular, for a special realization of the defect, an efficient, asymptotically correct error estimator, the *QDeC estimator*, was designed and analyzed in [2, 3] for collocation solutions on arbitrary grids. These techniques, originally designed for the ODE case, are now extended to the DAE context. This turns out not to be straightforward because of the coupling between differential and algebraic components.

A naive application of the classical procedure proposed in [10] would be based on the *pointwise defect*

$$d(t) := A(t)(Dp)'(t) + B(t)p(t) - g(t) \tag{6}$$

of the approximate solution $p(t)$. However, as pointed out in [2] in the ODE context, the resulting method for error estimation does not lead to useful results. In particular, in our collocation setting (3) we have $d(t_{ij}) \equiv 0$ at the collocation nodes, hence nontrivial defect information is not obtained by evaluation at these points. In our QDeC method, a more natural and robust *modified defect* is used instead (see [2] for a discussion of this choice in the ODE context). It is defined via approximate integral means of $d(t)$ between each pair of successive collocation nodes,

$$\bar{d}_{ij} := \sum_{k=0}^{s} \alpha_{jk}\, d(t_{ik}) = \frac{1}{h_{ij}} \int_{t_{i,j-1}}^{t_{ij}} d(t)\, dt + \mathscr{O}(h^{s+1}), \quad h_{ij} := t_{ij} - t_{i,j-1}, \tag{7}$$

for $i = 0 \ldots N-1$, $j = 1 \ldots s$, where the α_{jk} are appropriate quadrature coefficients. Note that due to $d(t_{ij}) \equiv 0$ at the collocation nodes, each sum in (7) actually reduces to a single non-vanishing term for $k = 0$. In this way, the appropriately weighted non-vanishing defect values $d(\tau_i)$ at the left endpoints of the collocation intervals are communicated to the collocation nodes and enter the equation defining the global error estimator, see (8) below.

Computing the defect for the type of DAEs considered here requires to reformulate the problem in such a way that the quantities involved are numerically well-defined. If the DAE system is initially specified in such a way that quantities of the type $\mathscr{O}(1/t)$ occur, multiplication by t obviously does not affect its solution but makes it numerically tractable for our purpose; in particular, $d(0)$ is well-defined after this rescaling.

In the linear case considered here, we define our a posteriori error estimate $\varepsilon_{ij} \approx e(t_{ij})$ as the solution of the Euler-type difference scheme

$$A(t_{ij})\,\frac{D\varepsilon_{ij} - D\varepsilon_{i,j-1}}{h_{ij}} + B(t_{ij})\varepsilon_{ij} = \bar{d}_{ij}, \tag{8}$$

with the inhomogeneous term given by (7), and with homogeneous initial conditions.

3. ANALYSIS OF ASYMPTOTIC CORRECTNESS FOR THE IVP CASE

We can show that the defect weighting according to (7), originally motivated by the local integration inherent in the (numerical) solution of an ODE, entails the full asymptotic quality of the global error estimate also for the algebraic components. We stress that the decoupling procedure used in the proof below is just an analytic device. What the theorem asserts is that the method can be applied, in a black-box fashion, to the DAE system as originally given.

Theorem 3.1 *While the global error of the collocation method* (3) *is of order s, i.e.*

$$e(t) = p(t) - x^*(t) = \mathscr{O}(h^s), \tag{9}$$

the QDeC estimate ε_{ij} *of the global error* (9) *defined via the modified defect* (7) *and the auxiliary scheme* (8) *is asymptotically correct, i.e.*

$$\varepsilon_{ij} - e(t_{ij}) = \mathscr{O}(h^{s+1}). \tag{10}$$

For the particular assumptions on the problem structure underlying our analysis see the following proof, which makes use of the results from [3, 4, 7].

Proof (outline). Our analysis is based on decoupling the index 1 problem (1) into its associated inherent ODE and a system of purely algebraic equations, according to the ideas from [5]. Let Q be a projector onto $\ker D$, and assume that

$$G(t) := A(t)D + B(t)Q \tag{11}$$

is nonsingular on $(0,T]$, which actually is the *defining property* for the index 1 case in the sense of 'tractability index', cf. e.g. [7]. Furthermore, let $D^- \in \mathbf{R}^{m\times n}$ be a generalized reflexive inverse of D (cf. [8, 9, 12]) chosen such that $D^-D = I_{m\times m} - Q$ and $DD^- = I_{n\times n}$. Furthermore we denote

$$N(t) := G^{-1}(t)B(t)D^- \in \mathbf{R}^{m\times n}. \tag{12}$$

Premultiplying the system (1) by $DG(t)^{-1}$ and $QG(t)^{-1}$, respectively, yields the decoupled system

$$(Dx)'(t) + DN(t)Dx(t) \;=\; DG^{-1}(t)g(t), \tag{13}$$

$$Qx(t) + QN(t)Dx(t) \;=\; QG^{-1}(t)g(t). \tag{14}$$

The solution of (1) can then be represented as

$$x(t) = (I_{m\times m} - Q)x(t) + Qx(t) = D^-(Dx)(t) + (Qx)(t). \tag{15}$$

A similar decoupling argument is applied to the discrete systems (3) and (8).

In the singular case considered here we have

$$DN(t) = -\frac{M(t)}{t} \tag{16}$$

assuming $M(t) \in C^{s+2}[0,T]$. To complete the specification of our assumptions, suppose that

$$t\,G^{-1}(t), \quad G^{-1}(t)g(t), \quad \text{and} \quad QG^{-1}(t)B(t) \tag{17}$$

admit continuous extensions to $[0,T]$, see [7].

For the error $\delta_{ij} := \varepsilon_{ij} - e_{ij}$ of the error estimate, we obtain the system of difference equations

$$\begin{aligned}\frac{D\delta_{ij} - D\delta_{i,j-1}}{h_{ij}} &= \frac{1}{t_{ij}} M(t_{ij})D\delta_{ij} \; - \\ &- \sum_{k=0}^{s} \alpha_{jk}\big(\tfrac{1}{t_{ik}}M(t_{ik})De(t_{ik}) - \tfrac{1}{t_{ij}}M(t_{ij})De(t_{ij})\big) - \sum_{k=0}^{s} \alpha_{jk}\big(DG^{-1}(t_{ik}) - DG^{-1}(t_{ij})\big)d(t_{ik}) + \mathscr{O}(h^{s+1}).\end{aligned} \tag{18}$$

In [3] it has been shown that the first sum on the right-hand side of (18) asymptotically behaves like

$$\frac{M(0)}{t_{ij}}\,\mathscr{O}(h^{s+1}) + \mathscr{O}(h^{s+1}) \tag{19}$$

if all t_{ik} are uniformly bounded from below by some $C\cdot h$. However, under the assumption $De(0) = 0$, which is natural for an IVP, this estimate applies even if $t_{00} = 0$ occurs when indeterminate forms are taken as limits.

We can infer from the theory for ODEs [2, 3] that $DG^{-1}(t_{ij})d(t_{ij}) = \mathscr{O}(h^s)$.

On the other hand, the collocation equations (3) in their decoupled form imply that $QG^{-1}(t_{ij})d(t_{ij}) = 0$ for $j = 0\ldots s$, owing to the fact that $t_{i+1,0} = t_{is}$ with our choice of collocation nodes and the assumption that the collocation solution exactly satisfies the algebraic conditions at the initial point $t = 0$. Hence, $d(t_{ij}) = \mathscr{O}(h^s)$ for all i,j (cf. [4]). Assuming that $tDG^{-1}(t)$ is sufficiently smooth, the second sum in (18) is also $\mathscr{O}(h^{s+1})$. Hence, again invoking a result from [3], $DG^{-1}(t_{ij})\delta_{ij} = \mathscr{O}(h^{s+1})$ for $i = 0\ldots N-1,\; j = 1\ldots s$.

Furthermore,

$$Q\delta_{ij} = -QND\delta_{ij} + QG^{-1}(t_{ij})\bar{d}_{ij}, \tag{20}$$

and

$$\begin{aligned}QG^{-1}(t_{ij})\bar{d}_{ij} \;&=\; \sum_{k=0}^{s} \alpha_{jk}\, QG^{-1}(t_{ij})d(t_{ik}) = \\ &=\; \sum_{k=0}^{s} \alpha_{jk}\big(QG^{-1}(t_{ij}) - QG^{-1}(t_{ik})\big)d(t_{ik}) + \sum_{k=0}^{s} \alpha_{jk}\, QG^{-1}(t_{ik})d(t_{ik}),\end{aligned} \tag{21}$$

where the first sum on the right hand side can be estimated as $\mathscr{O}(h^{s+1})$ analogously to (18), and the second sum vanishes because all $QG^{-1}(t_{ik})d(t_{ik})$ vanish as shown above.

Thus,

$$\delta_{ij} = D^-D\delta_{ij} + Q\delta_{ij} = \mathscr{O}(h^{s+1}), \tag{22}$$

which completes the proof of (10).

TABLE 1. Error of collocation and error of QDeC error estimate, together with observed orders

	1. IVP				**2. BVP**			
N	e_1	ord_{e_1}	$\varepsilon_1 - e_1$	$\mathrm{ord}_{\varepsilon_1 - e_1}$	e_2	ord_{e_2}	$\varepsilon_2 - e_2$	$\mathrm{ord}_{\varepsilon_2 - e_2}$
4	3.059e-05	4.0	7.578e-06	5.1	3.419e-05	4.0	7.555e-06	5.0
8	2.543e-06	3.6	2.335e-07	5.0	2.543e-06	3.8	2.384e-07	5.0
16	1.796e-07	3.8	7.429e-09	5.0	1.796e-07	3.8	7.685e-09	5.0
32	1.189e-08	3.9	2.408e-10	5.0	1.189e-08	3.9	2.504e-10	4.9

4. NUMERICAL EXAMPLE

On $(0,T] = (0,1]$ we consider the system

$$t\,x_1'(t) + \begin{pmatrix} -11 & -18 \\ 12 & 19 \end{pmatrix} x_1(t) + \begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix} x_2(t) = \begin{pmatrix} t\,\mathrm{e}^t(\sin t + \cos t - 12\mathrm{e}^t \sin t - 15\cos t + 13.5 \\ -t\sin t + 13\mathrm{e}^t \sin t + 17\cos t - 16 \end{pmatrix}, \tag{23}$$

$$\begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} x_1(t) + \begin{pmatrix} 1 & 0 \\ 0 & 0.2 \end{pmatrix} x_2(t) = \begin{pmatrix} \mathrm{e}^t \sin t + 2\cos t - 2.5 \\ 2.2\mathrm{e}^t \sin t + 3\cos t - 3 \end{pmatrix}. \tag{24}$$

1. IVP, with initial conditions

$$2x_{11}(0) + 3x_{12}(0) = 0, \quad x_{11}(0) + x_{12}(0) = -0.5. \tag{25}$$

2. BVP, with boundary conditions

$$2x_{11}(0) + 3x_{12}(0) = 0, \quad x_{11}(1) + x_{12}(1) = \mathrm{e}\,\sin(1) + \cos(1) - 1.5. \tag{26}$$

The results displayed in Table 1 have been obtained by an implementation in Maple 13, using IEEE double arithmetic. We use collocation at equidistant nodes with $s = 4$, on $N = 2,4,8,16,32$ subintervals of length $h = 1/N$, and invoke the QDeC error estimator defined by (8). The values in Table 1 refer to the maxima of $\|e\|_\infty$ and $\|\varepsilon - e\|_\infty$ evaluated over all mesh points τ_i. The asymptotically correct behavior $\varepsilon - e = \mathscr{O}(h^{s+1})$ of the error estimate is evident.

Although the BVP case is not (yet) covered by the above analysis, the empirical results show an analogous behavior.

REFERENCES

1. W. Auzinger, G. Kneisl, O. Koch, E. Weinmüller, *SBVP 1.0 – A* MATLAB *solver for singular boundary value problems*, Technical Report: ANUM Preprint No. **2/02**, Vienna University of Technology, 2002.
2. W. Auzinger, O. Koch, E. Weinmüller, *Efficient collocation schemes for singular boundary value problems*, Numer. Algorithms **31**, pp. 5–25 (2002).
3. W. Auzinger, O. Koch, E. Weinmüller, *Analysis of a new error estimate for collocation methods applied to singular boundary value problems*, SIAM J. Numer. Anal. **42**, pp. 2366–2386 (2005).
4. W. Auzinger, H. Lehner, and E. Weinmüller, *Defect-based a-posteriori error estimation for index-1 DAEs*, Preprint: ASC Report No. **20/2007**, Institute for Analysis and Scientific Computing, Vienna University of Technology, 2007.
5. K. Balla, R. März, *A unified approach to linear differential algebraic equations and their adjoints*, J. Anal. Appl. **21 (3)**, pp. 783–802 (2002).
6. F. R. de Hoog, R. Weiss, *Collocation methods for singular boundary value problems*, SIAM J. Numer. Anal. **15**, pp. 198–217 (1978).
7. O. Koch, R. März, D. Praetorius, E. Weinmüller, *Collocation methods for index 1 DAEs with a singularity of the first kind*, to appear in Mathematics of Computation. Preprint: ASC Report No. **10/2008**, Institute for Analysis and Scientific Computing, Vienna University of Technology, 2008.
8. P. Kunkel, V. Mehrmann, *Differential-Algebraic Equations – Analysis and Numerical Solution*, EMS Publishing House, Zürich, 2006.
9. S. Schulz, *Four Lectures on Differential-Algebraic Equations*, Report Series **497**, Dept. of Mathematics, Univ. of Auckland, 2003.
10. H. J. Stetter, *The defect correction principle and discretization methods*, Numer. Math. **29**, pp. 425–443 (1978).
11. P. E. Zadunaisky, *On the estimation of errors propagated in the numerical integration of ODEs*, Numer. Math. **27**, pp. 21–39 (1976).
12. G. Zielke, *Motivation und Darstellung von verallgemeinerten Matrixinversen*, Beiträge zur Numerischen Mathematik **7**, pp. 177–218, 1979.