ASC Report No. 13/2010

Convergence of Collocation Schemes for Nonlinear Index 1 DAEs with a Singular Point

Alexander Dick, Othmar Koch, Roswitha März, Ewa Weinmüller

Institute for Analysis and Scientific Computing
Vienna University of Technology — TU Wien
www.asc.tuwien.ac.at ISBN 978-3-902627-03-2

Most recent ASC Reports

12/2010 *Markus Aurada, Samuel Ferraz-Leite, Petra Goldenits, Michael Karkulik, Markus Mayr, Dirk Praetorius*
Convergence of Adaptive BEM for some Mixed Boundary Value Problem

11/2010 *Mario Bukal, Ansgar Jüngel, Daniel Matthes*
Entropies for Radially Symmetric Higher-Order Nonlinear Diffusion Equations

10/1010 *Karl Rupp, Ansgar Jüngel, Karl-Tibor Grasser*
Matrix Compression for Spherical Harmonics Expansions of the Boltzmann Transport Equation for Semiconductors

9/2010 *Markus Aurada, Samuel Ferraz-Leite, Dirk Praetorius*
Estimator Reduction and Convergence of Adaptive BEM

8/2010 *Robert Hammerling, Othmar Koch, Christa Simon, Ewa B. Weinmüller*
Numerical Solution of singular Eigenvalue Problems for ODEs with a Focus on Problems Posed on Semi-Infinite Intervals

7/2010 *Robert Hammerling, Othmar Koch, Christa Simon, Ewa B. Weinmüller*
Numerical Solution of Singular ODE Eigenvalue Problems in Electronic Structure Computations

6/2010 *Markus Aurada, Michael Feischl, Dirk Praetorius*
Convergence of Some Adaptive FEM-BEM Coupling

5/2010 *Marcus Page, Dirk Praetorius*
Convergence of Adaptive FEM for some Elliptic Obstacle Problem

4/2010 *Ansgar Jüngel, Josipa-Pina Milišić*
A Simplified Quantum Energy-Trans port Model for Semiconductors

3/2010 *Georg Kitzhofer, Othmar Koch, Gernot Pulverer, Christa Simon, Ewa Weinmüller*
The New MATLAB Code bvpsuite for the Solution of Singular Implicit BVPs

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

E-Mail: admin@asc.tuwien.ac.at
WWW: http://www.asc.tuwien.ac.at
FAX: +43-1-58801-10196

ISBN 978-3-902627-03-2

© Alle Rechte vorbehalten. Nachdruck nur mit Genehmigung des Autors.

Convergence of Collocation Schemes for Nonlinear Index 1 DAEs with a Singular Point

A. Dick, O. Koch, R. März, and E. Weinmüller

June 8, 2010

Abstract

We analyze the convergence behavior of collocation schemes applied to approximate solutions of BVPs in nonlinear index 1 DAEs, which exhibit a critical point at the left boundary. Such a critical point of the DAE causes a singularity in the inherent nonlinear ODE system. In particular, we focus on the case when the inherent ODE system is singular with a *singularity of the first kind* and apply polynomial collocation to the *original DAE system*. We show that for a well-posed boundary value problem in DAEs having a sufficiently smooth solution, the global error of the collocation scheme converges with the so-called stage order. Due to the singularity, superconvergence does not hold in general. The theoretical results are supported by numerical experiments.

1 Introduction

Motivated by numerous important applications from physics, see [3], [4], [13], [27], chemistry, cf. [6], [25], [26], mechanics, [5], ecology, see [20] and [23], or economy [7], [8], [14], a lot of interest and effort was put into the development of efficient numerical algorithms for the approximate solution of BVPs in explicit ODEs exhibiting singularities. Such problems are often given in the form,

$$x'(t) = \frac{1}{t^\alpha} M(t)x(t) + h(x(t), t), \quad t \in (0, 1], \tag{1a}$$

$$B_0 x(0) + B_1 x(1) = \beta, \quad x \in C[0, 1], \tag{1b}$$

where $\alpha \geq 1$, x is an m-dimensional real function, M is a smooth $m \times m$ matrix, h is a smooth function mapping into $\mathbb{R}^m$, and $\beta \in \mathbb{R}^p$, $B_0, B_1 \in \mathbb{R}^{p\times m}$. For $\alpha = 1$ the problem is called singular with a singularity of the first kind, for $\alpha > 1$ it is essentially singular (singularity of the second kind). Research activities in related fields, like the computation of connecting orbits in dynamical systems ([24]), or singular Sturm-Liouville problems ([1]), also benefit from new findings for problems of the form (1). The first attempt to extend collocation techniques developed in the context of singular explicit ODEs to DAEs was discussed for linear index 1 DAEs in [19]. To avoid repetitions we refer the reader to the latter paper for an overview of the state of research on the numerical treatment for singular ODEs.

It turns out that polynomial collocation retains its very advantageous convergence properties, even when the singularity is present, and therefore it serves as a robust and dependable basic solver *for singular ODEs*. As we have expected, collocation shows similar properties in the context of singular linear DAEs, see [19]. The aim of this article is to extend these results to the nonlinear case.

The open domain MATLAB code `bvpsuite` has been designed to solve general implicit systems of ODEs which may have arbitrary order including zero. In particular, algebraic constraints are also permitted and therefore, DAEs are in the scope of the code. [16], [17], [19] give numerical experiments and comparisons with existing software. We stress that in the present paper, we only use `bvpsuite` executed on coherently refined grids in order to illustrate the convergence order of the involved collocation schemes. We do not attempt to compare the available software for DAE systems here. As a matter of fact, `bvpsuite` has been used to work out the related conjectures before proving them. We recall that the dependable performance of the code is ensured by a strict analysis only for singular problems of the form (1) with $\alpha = 1$[1]. For $\alpha > 1$ the convergence theory of the collocation applied to (1) is an open and extremely challenging question. The additional difficulty in case of DAEs is due to their *implicit nature*.

Much progress has been made concerning DAE theory and applications, but there are still many questions left open. In particular, the numerical treatment of critical points and singularities is just emerging. Encouraged by the positive results for the linear case [19], we approach here the singular nonlinear index 1 DAE systems.

DAEs *with properly stated leading term*, were introduced and studied for example in [2, 9, 21]. This enables a proper and natural description of the involved solution derivatives. To this end, one considers DAEs written in the form

$$f((D(t)x(t))', x(t), t) = 0, \quad t \in [a, b]. \tag{2}$$

One of the advantages of this precise description of the problem structure is that there exists an *inherent explicit regular ODE* uniquely determined by the problem data, see [9, 10]. Under mild assumptions, DAEs in standard form can be reformulated to have properly stated leading terms. For DAEs with properly stated leading terms arising in applications, see [9].

In [22], *linear DAEs* with properly stated leading term and type $1A$-critical points have been analyzed. This means that after decoupling the system using the matrix chain technique developed in [2] into the differential and algebraic components, the related inherent ODE exhibits a singularity of the first or second kind. The singularities discussed here are the counterparts to the $1A$-critical points for *nonlinear DAEs*.

Recall that according to [19], for linear systems of DAEs with a singularity of the first kind and appropriately smooth problem data, the stage order of the collocation scheme is retained, provided that the so-called canonical projector remains bounded. This means that the global error of the collocation scheme with k collocation points is $O(h^k)$ uniformly in t, where h denotes the constant stepsize. We observe order reductions if the canonical projector becomes unbounded. In this article we will formulate the respective convergence results for BVPs in nonlinear index 1 DAEs with a singularity of the first kind.

The paper is structured in the following way. In Section 2, we describe the problem setting and show how the analytical system can be decoupled into the differential and algebraic components. The problem data is given in such a way that the inherent ODE exhibits a singularity of the first kind at $t = 0$. Collocation methods are introduced and their *convergence behaviour at collocation points* is analyzed in Section 3, while their *uniform convergence* is discussed in Section 4. In Section 5, we consider a quasi-linear DAE system. Here, the aim is to formulate sufficient conditions in terms of the original problem data leading to the decoupled system from Section 2. The analytical results are illustrated by numerical experiments in Section 6.

2 Problem Specification and Analytical Results

We discuss convergence properties of collocation for DAEs, where the so-called *inherent ODE* may have a singularity of the first kind. In this section we analyze the error of collocation methods applied to a nonlinear boundary value problem for a system of DAEs given in the following form:

$$f((D(t)x(t))', x(t), t) = 0, \quad t \in (0, 1], \tag{3a}$$

$$B_0 D(0)x(0) + B_1 D(1)x(1) = \beta, \tag{3b}$$

[1] Clearly, this also holds for regular problems with $\alpha = 0$.

where $f(y,x,t) \in \mathbb{R}^m$, $D(t) \in \mathbb{R}^{n\times m}$, $y \in \mathbb{R}^n$, $x \in \mathcal{D}$, with $\mathcal{D} \subseteq \mathbb{R}^m$ open, $t \in [0,1]$, $n \le m$. The data f, f_y, f_x, D are assumed to be at least continuous on their definition domains. Moreover, we require that

$$\ker f_y(y,x,t) = 0, \quad (y,x,t) \in \mathbb{R}^n \times \mathcal{D} \times (0,1], \tag{4}$$

$$\mathcal{R}(D(t)) = \mathbb{R}^n, \quad t \in [0,1]. \tag{5}$$

Conditions (5) and (4) mean that the matrix $D(t)$ has constant full row rank n on the closed interval while $f_y(y,x,t)$ has full column rank n on $\mathcal{D} \times (0,1]$, respectively. At $t = 0$ the matrix $f_y(y,x,t)$ may undergo a rank drop. The structure (4) and (5) means that the system (3a) has a properly stated leading term on $\mathbb{R}^n \times \mathcal{D} \times (0,1]$, cf. [9]. We consider solutions in the function space given below,

$$C^1_D([0,1],\mathbb{R}^m) := \{x \in C([0,1],\mathbb{R}^m) : Dx \in C^1([0,1],\mathbb{R}^n)\}.$$

This setting includes classical singular boundary value problems of the form (1) with $m = n$, $D(t) = I$ and $f(y,x,t) = t^\alpha y - M(t)x - h(x,t)$. In this paper, we are interested in $n < m$.
The structure of the boundary conditions given in (3b) which are necessary and sufficient for (3) to be well-posed, will be specified later.

We now define

$$N_0(t) := \ker D(t), \quad t \in [0,1], \tag{6}$$

and note that owing to the properties of the leading term, cf. (4), (5),

$$\ker f_y(y,x,t)D(t) = N_0(t), \quad (y,x,t) \in \mathbb{R}^n \times \mathcal{D} \times (0,1].$$

Let us denote by Q_0 a continuous pointwise projector function onto $\ker D$, $Q_0(t)^2 = Q_0(t)$, $\mathcal{R}(Q_0(t)) = \ker D(t)$, $t \in [0,1]$, and let $P_0 := I - Q_0$. Moreover, let us define

$$G_0(y,x,t) := f_y(y,x,t)D(t), \quad (y,x,t) \in \mathbb{R}^n \times \mathcal{D} \times [0,1], \tag{7}$$

$$G_1(y,x,t) := G_0(y,x,t) + f_x(y,x,t)Q_0(t), \quad (y,x,t) \in \mathbb{R}^n \times \mathcal{D} \times [0,1]. \tag{8}$$

In the following we discuss DAEs (3a) which are regular with tractability index 1 on $\mathbb{R}^n \times \mathcal{D} \times (0,1]$. Consequently, $G_1(y,x,t)$ is nonsingular on $\mathbb{R}^n \times \mathcal{D} \times (0,1]$. However, we permit a singular behavior of $G_1(y,x,t)$ for $t \to 0$, causing a singularity of the first kind in the associated *inherent ODE*. To this end, we suppose that

$$tG_1(y,x,t)^{-1} \tag{9}$$

has a continuous extension onto $R^n \times \mathcal{D} \times [0,1]$.
We introduce finally, the pointwise generalized inverse D^- of D uniquely defined by the following requirements:

$$D^-DD^- = D^-, \;\; DD^-D = D, \;\; DD^- = I, \;\; D^-D = P_0 \tag{10}$$

which need to hold pointwise on $[0,1]$. Note that D^- is also continuous on $[0,1]$.

It is well known that properties of the linearized problem are crucial in the analysis of the nonlinear setting. Therefore, we assume that a solution $x_\star \in C^1_D([0,1],\mathbb{R}^m)$ of (3) exists and introduce the linearization of (3). The linearization of (3a) has the following form:

$$A_\star(t)(D(t)z(t))' + B_\star(t)z(t) = g(t), \quad t \in (0,1], \tag{11}$$

where

$$A_\star(t) = f_y((D(t)x_\star(t))', x_\star(t), t), \;\; B_\star(t) = f_x((D(t)x_\star(t))', x_\star(t), t), \quad t \in [0,1].$$

This linear DAE is regular with tractability index 1 on the interval $(0,1]$. It was demonstrated in [2] that with the above assumptions the solutions of the DAE (11), see also [19], can be decoupled on $(0,1]$ into the *differential components* Dz and the *algebraic components* Q_0z. While Dz satisfies the explicit inherent ODE,

$$(D(t)z(t))' + \underbrace{D(t)G_{\star 1}^{-1}(t)B_\star(t)D(t)^-}_{=:-\frac{1}{t}M_\star(t)} D(t)z(t) = D(t)G_{\star 1}^{-1}(t)g(t), \quad t \in (0,1], \tag{12}$$

where

$$G_{\star 1}(t) = A_\star(t)D(t) + B_\star(t)Q_0(t) = G_1((D(t)x_\star(t))', x_\star(t), t),$$

the algebraic components are given by

$$Q_0(t)z(t) = -Q_0(t)G_{\star 1}^{-1}(t)B_\star(t)D(t)^- D(t)z(t) + Q_0(t)G_{\star 1}^{-1}(t)g(t) \tag{13}$$

and the solutions of (11) can be expressed as

$$z(t) = D(t)^- D(t)z(t) + Q_0(t)z(t), \quad t \in (0,1]. \tag{14}$$

We note that $M_\star$ is continuous on $[0,1]$.

In analogy to the theory of explicit ODEs [15], we say that the solution $x_\star$ of boundary value problem (3) is *isolated* if and only if the linearization of boundary value problem (3),

$$A_\star(t)(D(t)z(t))' + B_\star(t)z(t) = 0, \quad t \in (0,1], \tag{15a}$$

$$B_0D(0)z(0) + B_1D(1)z(1) = 0, \tag{15b}$$

has only the trivial solution. It this case the boundary value problem (3) is said to be *well-posed*.

We now turn back to the nonlinear DAE (3a) and decouple it into the inherent ODE and the algebraic constraints. We first introduce the notation,

$$u_\star(t) := D(t)x_\star(t), \;\; w_\star(t) := Q_0(t)x_\star(t) + D(t)^-(D(t)x_\star(t))', \quad t \in [0,1]$$

as well as the function

$$F(w,u,t) := f(D(t)w, D(t)^-u + Q_0(t)w, t),$$

where $w \in \mathbb{R}^m$, $u \in \mathbb{R}^n$, $t \in (0,1]$ such that $D(t)^-u + Q_0(t)w \in \mathcal{D}$.

We observe that $F(w_\star(t), u_\star(t), t) = 0$, $t \in [0,1]$ and

$$F_w(w_\star(t), u_\star(t), t) = f_y((D(t)x_\star(t))', x_\star(t), t)D(t) + f_x((D(t)x_\star(t))', x_\star(t), t)Q_0(t) = G_{\star 1}(t), \quad t \in [0,1].$$

$G_{\star 1}(t)$ is nonsingular for $t > 0$. Consequently, it follows from the Implicit Function Theorem that for $t \in (0,1]$, the equation $F(w,u,t) = 0$ is locally equivalent to $w = \omega(u,t)$, where the function ω together with its partial derivative ω_u are continuous and

$$\omega(u_\star(t), t) = w_\star(t), \;\; \omega_u(u_\star(t), t) = -G_{\star 1}^{-1}(t)B_\star(t)D(t)^-, \quad t \in (0,1], \tag{16}$$

for details see [9].

Remark: Note that for the linear case the function ω can be easily specified. We have

$$f(y,x,t) := A(t)y + B(t)x - g(t)$$

and

$$G_1(t) = A(t)D(t) + B(t)Q_0(t).$$

Consequently,

$$F(w,u,t) = A(t)D(t)w + B(t)(D(t)^- u + Q_0(t)w(t)) - g(t) = G_1(t)w + B(t)D(t)^- u - g(t),$$

and hence

$$\omega(u,t) = -G_1(t)^{-1}(B(t)D(t)^- u - g(t)), \quad u \in \mathbb{R}^n, \quad t \in (0,1].$$

We now use our so-called *decoupling function* $\omega : \mathcal{D}_\omega \times (0,1] \to \mathbb{R}^m$, where $\mathcal{D}_\omega \subseteq \mathbb{R}^n$ is an open set, to specify the inherent ODE associated with the nonlinear DAE (3a). Let x be a solution of (3a), and let us define

$$u(t) := D(t)x(t), \;\; w(t) := Q_0(t)x(t) + (D(t)x(t))'.$$

It follows that

$$D(t)w(t) = (D(t)x(t))', \;\; Q_0(t)w(t) = Q_0(t)x(t), \;\; P_0(t)x(t) = D(t)^- D(t)x(t) = D(t)^- u(t), \tag{17}$$

and

$$x(t) = D(t)^- u(t) + Q_0(t)w(t), \quad t \in [0,1]. \tag{18}$$

With the above notation, the original DAE (3a) can be rewritten as

$$f(D(t)w(t), D(t)^- u(t) + Q_0(t)w(t), t) = 0.$$

In case that the domain $\mathcal{D}_\omega$ is sufficiently large, the solution x can be represented in the following form:

$$x(t) = D(t)^- u(t) + Q_0(t)\omega(u(t),t), \quad t \in (0,1], \tag{19}$$

where u satisfies the inherent ODE,

$$u'(t) = D(t)\omega(u(t),t), \quad t \in (0,1]. \tag{20}$$

In order to apply the standard analysis for singular boundary value problems, cf. [11] and [18], we assume that the decoupling function ω satisfies

$$D(t)\omega(u,t) = \frac{1}{t}M(t)u + g(u,t), \quad u \in \mathcal{D}_\omega, \quad t \in (0,1], \tag{21}$$

where the $n \times n$ matrix function M and the function g are appropriately smooth on $[0,1]$ and $\mathcal{D}_\omega \times [0,1]$, respectively. Later on, in Section 5, for a large class of quasi-linear DAEs, we derive sufficient conditions for (21) to hold in terms of the original data of the DAE system.
The inherent ODE (20) is now augmented by the boundary conditions (3b),

$$B_0 u(0) + B_1 u(1) = \beta. \tag{22}$$

This yields the following boundary value problem:

$$u'(t) = \frac{1}{t}M(t)u(t) + g(u(t),t), \quad t \in (0,1], \tag{23a}$$

$$B_0 u(0) + B_1 u(1) = \beta. \tag{23b}$$

The linearization of the above boundary value problem reads, cf. (12),

$$\zeta'(t) = D(t)\omega_u(u_\star(t),t)\zeta(t) = \frac{1}{t}M_\star(t)\zeta(t), \quad t \in (0,1], \tag{24a}$$

$$B_0\zeta(0) + B_1\zeta(1) = 0. \tag{24b}$$

Note that the linear boundary value problem (15), for a system of DAEs, has only the trivial solution exactly when this is the case for the related boundary value problem for the inherent ODE (24), due to the solution representation (14). This means that the nonlinear boundary value problem (3), for a system of DAEs, is well-posed exactly when this holds for the related boundary value problem for the nonlinear inherent ODE (23).

We now specify the necessary and sufficient conditions for the linear ODE problem (24) to have only the trivial solution. First of all, it is clear that for the matrices B_0, B_1 in (24b) and (3b), $B_0, B_1 \in \mathbb{R}^{n\times m}$ has to be true. It was shown in [11] that the form of the boundary conditions (24b) which guarantee that (24) has a unique trivial solution depends on the spectral properties of the coefficient matrix $M_\star(0)$. Note that (21) implies

$$M_\star(t) = M(t) + t g_u(u,t), \quad u \in \mathcal{D}_\omega, \quad t \in (0,1],$$

and therefore $M_\star(0) = M(0)$.

To avoid fundamental modes of (24a) which have the form $\cos(\sigma \ln(t)) + \mathrm{i}\sin(\sigma \ln(t))$, we assume that zero is the only eigenvalue of $M(0)$ on the imaginary axis.

Now, let S denote a projection onto the invariant subspace which is associated with eigenvalues of $M(0)$ which have positive real parts, and R a projection onto the kernel of $M(0)$. Finally, define

$$U := S + R, \quad V := I - U, \tag{25}$$

where I denotes the identity matrix in $\mathbb{R}^n$. In [11] it was shown that the boundary value problem (24) is well-posed if and only if the boundary conditions (24b) can equivalently be written as

$$V\zeta(0) = 0, \tag{26a}$$

$$R\zeta(0) = 0, \quad S\zeta(1) = 0. \tag{26b}$$

In the next section, we apply polynomial collocation to approximate solutions of (3) by means of an enlarged system,

$$f(u'(t), x(t), t) = 0, \tag{27a}$$

$$D(t)x(t) - u(t) = 0, \quad t \in (0,1], \tag{27b}$$

which can be brought into the form

$$\widehat{f}((\widehat{D}(t)\widehat{x}(t))', \widehat{x}(t), t) = 0, \quad t \in (0,1], \tag{28}$$

where $\widehat{x}(t) = (x(t), u(t))^T$ and

$$\widehat{f}(y, \widehat{x}, t) := \begin{pmatrix} f(y,x,t) \\ D(t)x - u \end{pmatrix}, \quad \widehat{D}(t) = \begin{pmatrix} 0 & I \end{pmatrix} =: \widehat{D}, \quad \widehat{x} = \begin{pmatrix} x \\ u \end{pmatrix} \in \mathcal{D} \times \mathbb{R}^n, \quad y \in \mathbb{R}^n.$$

Problem (28) is a regular DAE system with properly stated leading term and tractability index 1 on $(0,1]$. To see this, note that $\widehat{D}(t)$ is constant and therefore we define the related matrices $\widehat{G}_0(t)$, $\widehat{Q}_0$, and $\widehat{G}_1(t)$ as

$$\widehat{G}_0(y, \widehat{x}, t) := \begin{pmatrix} 0 & f_y(y,x,t) \\ 0 & 0 \end{pmatrix}, \quad \widehat{Q}_0 := \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$\widehat{G}_1(y, \widehat{x}, t) := \widehat{G}_0(y, \widehat{x}, t) + \begin{pmatrix} f_x(y,x,t) & 0 \\ D(t) & -I \end{pmatrix} \widehat{Q}_0 = \begin{pmatrix} f_x(y,x,t) & f_y(y,x,t) \\ D(t) & 0 \end{pmatrix},$$

respectively. Moreover,

$$\ker \widehat{G}_1 = \{z \in \mathbb{R}^{m+n},\ z = (z_1, z_2)^T;\ z_1 = Q_0 w,\ z_2 = Dw,\ w \in \ker G_1\}$$

for all $t \in (0,1]$, which means that $\widehat{G}_1$ is nonsingular for $t > 0$. The enlarged DAE (28) has exactly the same inherent ODE as the original DAE (3a) does.

3 Collocation Methods – Convergence at Collocation Points

For the theoretical discussion of collocation methods, we define meshes

$$\Delta := (\tau_0, \tau_1, \dots, \tau_N),$$

and $h_i := \tau_{i+1} - \tau_i,\ i = 0, \dots, N-1,\ \tau_0 = 0,\ \tau_N = 1$. For reasons of simplicity, we restrict the discussion to equidistant meshes, $h_i = h,\ i = 0, \dots, N-1$. However, the results also hold for nonuniform meshes which have a limited variation in the stepsizes. For collocation, k distinct points $t_{i,j} := \tau_i + h_i\rho_j,\ j = 1, \dots, k$, are inserted in each subinterval (τ_i, τ_{i+1}). Since we want to focus on Gaussian points, we restrict ourselves to interior collocation points, where $\rho_1 > 0$ and $\rho_k < 1$. A grid with equidistant interior collocation points is illustrated in Figure 1.

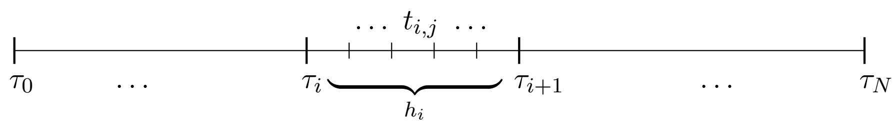

Figure 1: The computational grid

Now, let us denote by $\mathcal{B}_k$ the Banach space of continuous, piecewise polynomial functions $q \in \mathbb{P}_k$ of degree $\leq k,\ k \in \mathbb{N}$, equipped with the maximum norm $\|\cdot\|_\infty$. In the following, we denote by q the vector valued functions from $\mathcal{B}_k$ independently of the number of their components.

By $p \in \mathcal{B}_k$ we denote an approximation to the exact solution $x_\star$ of (3), and by $q \in \mathcal{B}_k$ an approximation to the exact solution $u_\star$ of the inherent ODE (20). As usually, to compute p and q, we set up the collocation equations augmented by the proper number of boundary conditions,

$$f(q'(t_{i,j}), p(t_{i,j}), t_{i,j}) = 0, \tag{29a}$$

$$D(t_{i,j})p(t_{i,j}) - q(t_{i,j}) = 0, \tag{29b}$$

$$B_0\, q(0) + B_1 q(1) = \beta, \tag{29c}$$

where $j = 1, \dots, k$ and $i = 0, \dots, N-1$. By inspection of the number of unknowns ($(k+1)N(n+m)$ polynomial coefficients) and equations ($Nk(n+m)$ collocation conditions, $(N-1)(n+m)$ continuity conditions for p and q, n boundary conditions) we see that further m conditions will be necessary to close the system for the numerical treatment. Clearly, these additional conditions have to be consistent with the original DAEs. Various choices are possible. In general, we may complete the above scheme by the $2m$ equations

$$f(D(1)w, D(1)^- q(1) + Q_0(1)w, 1) = 0, \quad p(1) = D(1)^- q(1) + Q_0(1)w, \tag{30}$$

which contain the additional variable $w \in \mathbb{R}^m$. This choice yields

$$w = \omega(q(1), 1), \quad p(1) = D(1)^- q(1) + Q_0(1)\omega(q(1), 1).$$

If the DAE (3a) is given with separated derivative free equations,

$$f_1((D(t)x(t))', x(t), t) = 0,$$

$$f_2(x(t), t) = 0,$$

where f_1 and f_2 have n and $m-n$ components, respectively, then we can augment the scheme (29) by

$$f_2(p(1), 1) = 0, \quad D(1)p(1) = q(1), \tag{31}$$

or by

$$f_2(p(0),0) = 0, \quad D(0)p(0) = q(0). \tag{32}$$

In order to show that the collocation scheme for the DAE (29) comprises exactly the same scheme applied to the inherent ODE, we introduce

$$u_{ij} := q(t_{ij}), \quad w_{ij} := D(t_{ij})^- q'(t_{ij}) + Q_0(t_{ij})p(t_{ij}), \tag{33}$$

and obtain from (29a)

$$f(D(t_{ij})w_{ij}, D(t_{ij})^- u_{ij} + Q_0(t_{ij})w_{ij}, t_{ij}) = 0. \tag{34}$$

By applying the decoupling function to (34), the relation

$$w_{ij} = \omega(u_{ij}, t_{ij}) = \omega(q(t_{ij}), t_{ij})$$

follows, and hence,

$$q'(t_{ij}) = D(t_{ij})\omega(q(t_{ij}), t_{ij}), \tag{35}$$

and

$$p(t_{ij}) = D(t_{ij})^- q(t_{ij}) + Q_0(t_{ij})\omega(q(t_{ij}), t_{ij}), \tag{36}$$

for all collocation points t_{ij}. The system (35) together with the boundary conditions (29c) form a classical collocation scheme for $q \in \mathcal{B}_k$. According to Theorem 3.1 in [18], there exists a unique collocation solution $q \in \mathcal{B}_k$ of the scheme (35) subject to (29c), under the assumptions that the underlying analytical problem is well-posed with sufficiently smooth data, and that the grid is sufficiently fine. Finally, $p \in \mathcal{B}_k$ is uniquely specified by the values of p at all collocation points, see (36), and one of the consistency conditions given by (30), (31) or (32).

The convergence analysis for the nonlinear collocation scheme (35) subject to appropriately posed boundary conditions has been given in [18]. Especially, the structure of the global error of the collocation solution at the collocation points has been described,

$$q(t_{ij}) - u_\star(t_{ij}) = \epsilon_u(t_{ij})h^k + r(t_{ij}) + O(h^{k+1}), \tag{37}$$

where ϵ_u is a smooth function with $\epsilon_u(0) \in \ker M(0)$, and $r = O(h^k)$ is in $\mathcal{B}_k$. Thus, we can conclude that $q(t_{ij}) - u_\star(t_{ij}) = O(h^k)$ holds.
It follows from the solution representation at the collocation points (19) and (36) that

$$\begin{aligned} p(t_{ij}) &= D(t_{ij})^- q(t_{ij}) + Q_0(t_{ij})\omega(q(t_{ij}), t_{ij}), \\ x_\star(t_{ij}) &= D(t_{ij})^- u_\star(t_{ij}) + Q_0(t_{ij})\omega(u_\star(t_{ij}), t_{ij}). \end{aligned}$$

Consequently,

$$p(t_{ij}) - x_\star(t_{ij}) = D(t_{ij})^-(q(t_{ij}) - u_\star(t_{ij})) + \int_0^1 Q_0(t_{ij})\omega_u(sq(t_{ij}) + (1-s)u_\star(t_{ij}), t_{ij})ds\,(q(t_{ij}) - u_\star(t_{ij})).$$

Taking into account the form of ω_u, we finally obtain the error representation

$$p(t_{ij}) - x_\star(t_{ij}) = \int_0^1 \left[I - Q_0(t_{ij})G_1^{-1}(P_{ij}(s))f_x(P_{ij}(s))\right]ds\, D(t_{ij})^-(q(t_{ij}) - u_\star(t_{ij})),$$

where

$$\begin{aligned} \big(P_{i,j}(s)\big) = \big(&D(t_{ij})\omega(sq(t_{ij}) + (1-s)u_\star(t_{ij}), t_{ij}), \\ &D(t_{ij})^-(sq(t_{ij}) + (1-s)u_\star(t_{ij})) + Q_0(t_{ij})\omega(sq(t_{ij}) + (1-s)u_\star(t_{ij}), t_{ij}),\, t_{ij}\big). \end{aligned}$$

Recall that $Q_0 G_1^{-1} f_x$ represents the canonical projector of the DAE. Evidently, in case that this projector function remains bounded, the error $p(t_{ij}) - x_\star(t_{ij})$ inherits the convergence order of $q(t_{ij}) - u_\star(t_{ij})$ which means that

$$p(t_{ij}) - x_\star(t_{ij}) = O(h^k). \tag{38}$$

This makes clear that the commutativity of the decoupling procedure and the collocation discretization, which is due to the numerically qualified DAE formulation (with $\mathcal{R}(D(t) = \mathbb{R}^n$, $t \in [0,1]$, cf. [9], [10]), ensures the expected convergence order at the collocation points.

4 Collocation Methods – Uniform Convergence

Beyond the convergence at the collocation points discussed in the previous section, we know from Theorem 3.1 in [18], that uniform convergence holds for the inherent explicit ODE,

$$\|q - u_\star\| = O(h^k), \quad \|q' - u_\star'\| = O(h^k). \tag{39}$$

Moreover, in case of an initial value problem we have

$$q(t) - u_\star(t) = tO(h^k). \tag{40}$$

We now consider the convergence behaviour of the whole numerical solution including algebraic components. We perform this error analysis relying on techniques presented in [19]. Here, we restrict ourselves to initial value problems. Let the exact solution and the related collocation polynomial be given as

$$\widehat{x}_\star(t) := \begin{pmatrix} x_\star(t) \\ u_\star(t) \end{pmatrix}, \quad \widehat{p}(t) := \begin{pmatrix} p(t) \\ q(t) \end{pmatrix},$$

respectively. Let us introduce the error function $\widehat{e}(t) := (e(t), e_u(t))^T$, $\widehat{e} \in \mathcal{B}_k$, by

$$\widehat{e}\,'(0) = 0, \quad \widehat{e}\,'(t_{i,j}) = \widehat{x}_\star'(t_{ij}) - \widehat{p}\,'(t_{ij}), \;\; j = 1, \dots, k, \;\; i = 0, \dots N-1,$$

which yields

$$\widehat{e}\,'(t) = \widehat{x}_\star'(t) - \widehat{p}\,'(t) + O(h^k), \quad \widehat{e}(t) = \widehat{x}_\star(t) - \widehat{p}(t) + t \begin{pmatrix} r(t) \\ s(t) \end{pmatrix},$$

where $r(t) = O(h^k)$ and $s(t) = O(h^k)$. From (27) and (29a), (29b) we know that

$$\begin{aligned} f(u_\star'(t_{ij}), x_\star(t_{ij}), t_{ij}) = 0, &\quad D(t_{ij})x_\star(t_{ij}) - u_\star(t_{ij}) = 0, \\ f(q'(t_{ij}), p(t_{ij}), t_{ij}) = 0, &\quad D(t_{ij})p(t_{ij}) - q(t_{ij}) = 0, \end{aligned}$$

leading to

$$\mathcal{A}(t_{ij})(q'(t_{ij}) - u_\star'(t_{ij})) + \mathcal{B}(t_{ij})(p(t_{ij}) - x_\star(t_{ij})) = 0, \tag{41a}$$

$$D(t_{ij})(p(t_{ij}) - x_\star(t_{ij})) - (q(t_{ij}) - u_\star(t_{ij})) = 0, \tag{41b}$$

with coefficients

$$\mathcal{A}(t) = \int_0^1 \tilde{\mathcal{A}}(t,\tau)d\tau, \quad \mathcal{B}(t) = \int_0^1 \tilde{\mathcal{B}}(t,\tau)d\tau,$$

$$\begin{aligned} \tilde{\mathcal{A}}(t,\tau) &= f_y(u_\star'(t) + \tau(q'(t) - u_\star'(t)), x_\star(t) + \tau(p(t) - x_\star(t)),\, t), \quad t, \tau \in [0,1] \\ \tilde{\mathcal{B}}(t,\tau) &= f_x(u_\star'(t) + \tau(q'(t) - u_\star'(t)), x_\star(t) + \tau(p(t) - x_\star(t)),\, t), \quad t, \tau \in [0,1]. \end{aligned}$$

The matrix functions $\mathcal{A}$ and $\mathcal{B}$ are continuous on $[0,1]$ because $x_\star, p, u_\star', q'$ are so.

From (38) and from the representations

$$e_u'(t_{ij}) = u_\star'(t_{ij}) - q'(t_{ij}), \tag{42a}$$
$$e_u(t_{ij}) = u_\star(t_{ij}) - q(t_{ij}) + t_{ij}s(t_{ij}), \tag{42b}$$
$$e(t_{ij}) = x_\star(t_{ij}) - p(t_{ij}) + t_{ij}r(t_{ij}), \tag{42c}$$

it follows that

$$\mathcal{A}(t_{ij})e_u'(t_{ij}) + \mathcal{B}(t_{ij})e(t_{ij}) = t_{ij}\mathcal{B}(t_{ij})r(t_{ij}), \tag{43a}$$
$$D(t_{ij})e(t_{ij}) - e_u(t_{ij}) = t_{ij}D(t_{ij})r(t_{ij}) - t_{ij}s(t_{ij}), \tag{43b}$$

subject to

$$e(0) = 0, \quad e_u(0) = 0. \tag{44}$$

Next, we take a closer look at the coefficients $\mathcal{A}$ and $\mathcal{B}$ and show that $\mathcal{A}(t_{ij}) \in \mathbb{R}^{m\times n}$ has full column rank n and $\mathcal{G}_1(t_{ij}) := \mathcal{A}(t_{ij})D(t_{ij}) + \mathcal{B}(t_{ij})Q_0(t_{ij})$ is nonsingular provided that the stepsize h is sufficiently small. Rewrite

$$\begin{aligned}
\tilde{\mathcal{A}}(t,\tau) &= f_y(u_\star'(t), x_\star(t), t) \\
&+ \int_0^1 \{f_{yy}(u_\star'(t) + (1-\xi)\tau(q'(t) - u_\star'(t)), x_\star(t) + (1-\xi)\tau(p(t) - x_\star(t)), t)\tau(q'(t) - u_\star'(t)) \\
&+ f_{yx}(u_\star'(t) + (1-\xi)\tau(q'(t) - u_\star'(t)), x_\star(t) + (1-\xi)\tau(p(t) - x_\star(t)), t)\tau(p(t) - x_\star(t))\}\, d\xi \\
&= A_\star(t) + \int_0^1 \{\cdots\}\, d\xi, \\
\tilde{\mathcal{B}}(t,\tau) &= f_x(u_\star'(t), x_\star(t), t) \\
&+ \int_0^1 \{f_{xy}(u_\star'(t) + (1-\xi)\tau(q'(t) - u_\star'(t)), x_\star(t) + (1-\xi)\tau(p(t) - x_\star(t)), t)\tau(q'(t) - u_\star'(t)) \\
&+ f_{xx}(u_\star'(t) + (1-\xi)\tau(q'(t) - u_\star'(t)), x_\star(t) + (1-\xi)\tau(p(t) - x_\star(t)), t)\tau(p(t) - x_\star(t))\}\, d\xi \\
&= B_\star(t) + \int_0^1 \{\cdots\}\, d\xi.
\end{aligned}$$

Taking into account the convergence results formulated in (38) and (39) we obtain

$$\begin{aligned}
\mathcal{A}(t_{ij}) &= A_\star(t_{ij}) + \int_0^1\int_0^1 \{f_{yy}(\cdots, t_{ij})\tau(\underbrace{q'(t_{ij}) - u_\star'(t_{ij})}_{O(h^k)}) + f_{yx}(\cdots, t_{ij})\tau(\underbrace{p(t_{ij}) - x_\star(t_{ij})}_{O(h^k)})\}\, d\xi d\tau \\
&= A_\star(t_{ij}) + O(h^k),
\end{aligned}$$

analogously

$$\mathcal{B}(t_{ij}) = B_\star(t_{ij}) + O(h^k),$$

and hence

$$\mathcal{G}_1(t_{ij}) = G_{\star 1}(t_{ij}) + O(h^k).$$

We now suppose the grid to be sufficiently fine so that the terms $O(h^k)$ are small and therefore $\mathcal{A}(t_{ij})$ inherits from $A_\star(t_{ij})$ the full rank n, while $\mathcal{G}_1(t_{ij})$ inherits the invertibility of $G_{\star 1}(t_{ij})$.

Let us now introduce

$$\mathcal{M}(t_{ij}) := -t_{ij} D(t_{ij}) \mathcal{G}_1(t_{ij})^{-1} \mathcal{B}(t_{ij}) D(t_{ij})^-.$$

We multiply (43a) by $D(t_{ij})\mathcal{G}(t_{ij})^{-1}$ and by $Q_0(t_{ij})\mathcal{G}(t_{ij})^{-1}$ and also (43b) by $D(t_{ij})^-$. This implies

$$\begin{aligned}
&e_u'(t_{ij}) = \frac{1}{t_{ij}} \mathcal{M}(t_{ij}) e_u(t_{ij}) - \mathcal{M}(t_{ij}) s(t_{ij}),\\
&Q_0(t_{ij})\mathcal{G}_1(t_{ij})^{-1}\mathcal{B}(t_{ij})P_0(t_{ij})e(t_{ij}) + Q_0(t_{ij})e(t_{ij}) = t_{ij} Q_0(t_{ij})\mathcal{G}_1(t_{ij})^{-1}\mathcal{B}(t_{ij}) r(t_{ij}),\\
&P_0(t_{ij})e(t_{ij}) - D(t_{ij})^- e_u(t_{ij}) = t_{ij} P_0(t_{ij}) r(t_{ij}) - t_{ij} D(t_{ij})^- s(t_{ij}),
\end{aligned}$$

respectively. Thus, the values of $e(t_{ij})$ for the interpolation process from which we can recover the error function $e(t)$ read $e(0) = 0$ and

$$\begin{aligned}
e(t_{ij}) &= (P_0(t_{ij}) + Q_0(t_{ij}))e(t_{ij})\\
&= (I - Q_0(t_{ij})\mathcal{G}_1(t_{ij})^{-1}\mathcal{B}(t_{ij}))(D(t_{ij})^- e_u(t_{ij}) + t_{ij}P_0(t_{ij})r(t_{ij}) - t_{ij}D(t_{ij})^- s(t_{ij}))\\
&\quad + t_{ij}Q_0(t_{ij})\mathcal{G}_1(t_{ij})^{-1}\mathcal{B}(t_{ij})r(t_{ij}).
\end{aligned}$$

It can be seen that $e_u(t_{ij}) = t_{ij}O(h^k)$. This follows from $e_u(t) = u_\star(t) - q(t) + ts(t)$ and (40). Consequently, the error function is determined by the values $e(0) = 0$ and

$$e(t_{ij}) = (I - Q_0(t_{ij})\mathcal{G}_1(t_{ij})^{-1}\mathcal{B}(t_{ij}))t_{ij}O(h^k) + t_{ij}Q_0(t_{ij})\mathcal{G}_1(t_{ij})^{-1}\mathcal{B}(t_{ij})O(h^k).$$

Note that the crucial terms in the above expression are $\mathcal{G}_1(t_{ij})^{-1}$ for t_{ij} close to zero. Due to the basic assumption (9) we know that $t_{ij}G_{\star 1}(t_{ij})^{-1}$ is uniformly bounded (for arbitrary grids). When this also holds for the perturbed expression $t_{ij}\mathcal{G}_1(t_{ij})^{-1}$ then the interpolation leads to $e(t) = O(h^k)$ and we arrive at the uniform convergence result for the global error of the collocation solution to the DAE system,

$$x_\star(t) - p(t) = O(h^k). \tag{45}$$

In the above consideration, the uniform boundedness of the expression $t_{ij}\mathcal{G}_1(t_{ij})^{-1}$ plays an important role. It may happen that the error term in $\mathcal{G}_1(t_{ij}) = G_{\star 1}(t_{ij}) + O(h^k)$ disappears, and $\mathcal{G}_1(t_{ij})$ equals $G_{\star 1}(t_{ij})$, such that $t_{ij}\mathcal{G}_1(t_{ij})^{-1}$ inherits the uniform boundedness from $t_{ij}G_{\star 1}(t_{ij})^{-1}$. In particular, this situation is given for the special quasi-linear DAEs

$$A(t)(D(t)x(t))' + B(t)x(t) + \mathcal{S}(D(t)x(t), t) = 0,$$

i.e. for

$$f(y, x, t) = A(t)y + B(t)x + \underbrace{\mathcal{S}(D(t)x, t)}_{=:h(x,t)}, \quad y \in \mathbb{R}^n, x \in \mathcal{D}, t \in [0, 1].$$

Namely, then we have

$$\begin{aligned}
&f_y(y, x, t) = A(t), \quad f_x(y, x, t) = B(t) + h_x(x, t),\\
&f_{yy} = 0, \quad f_{yx} = f_{xy} = 0, \quad f_{xx}(y, x, t) = h_{xx}(x, t),
\end{aligned}$$

and further

$$h(x, t) = h(P_0(t)x, t),$$

such that $P_0(t)w = 0$ implies $h_{xx}(x,t)w = 0$. This yields

$$\mathcal{M}(t_{ij}) = A_\star(t_{ij}) = A(t_{ij}),$$

$$\mathcal{L}(t_{ij}) = B_\star(t_{ij}) + \int_0^1\int_0^1 h_{xx}(x_\star(t_{ij}) + (1-\xi)\tau(p(t_{ij}) - x_\star(t_{ij})), t_{ij}) \cdot \tau(p(t_{ij}) - x_\star(t_{ij}))d\xi d\tau,$$

$$\mathcal{L}(t_{ij})Q_0(t_{ij}) = B_\star(t_{ij})Q_0(t_{ij}),$$

and finally

$$\begin{aligned}\mathcal{G}_1(t_{ij}) &= \mathcal{M}(t_{ij})D(t_{ij}) + \mathcal{L}(t_{ij})Q_0(t_{ij}) \\ &= A_\star(t_{ij})D(t_{ij}) + B_\star(t_{ij})Q_0(t_{ij}) = G_{\star 1}(t_{ij}).\end{aligned}$$

Let us also mention that $t_{ij}\mathcal{G}_1(t_{ij})^{-1}$ remains uniformly bounded for the problem class which we discuss in the next section.

5 Quasi-Linear DAEs

The aim of this section is to derive sufficient conditions in terms of the original problem data to ensure the special structure of the inherent ODE (21) for a quasi-linear DAE of the form

$$A(t)(D(t)x(t))' + B(t)x(t) + h(x(t),t) = 0. \tag{46}$$

As for the general case we assume $\mathcal{R}(D(t)) = \mathbb{R}^n$, $t \in [0,1]$ and $\ker A(t) = \{0\}$, $t \in (0,1]$. As before, we use the orthoprojectors $Q_0(t)$ and $P_0(t) = I - Q_0(t)$ onto $\ker D(t)$ and $\ker D(t)^\perp$, respectively. We also make the following assumptions:

(i) Let the matrix $\tilde{G}_1(t) := A(t)D(t) + B(t)Q_0(t)$ be nonsingular for all $t \in (0,1]$.

(ii) Let $t\tilde{G}_1(t)^{-1}$ have a continuous extension on $[0,1]$.

(iii) Let $\tilde{G}_1(t)^{-1}h(x,t)$ have a smooth extension on $\mathcal{D} \times [0,1]$.

(iv) Let $Q_0(t)\tilde{G}_1(t)^{-1}B(t)$ be bounded for $t \to 0$.

(v) Let the function h have the structure

$$(I - A(t)A(t)^-)(h(x,t) - h(P_0(t)x,t)) = 0.$$

The properties (i)–(iv) have already been useful in the analysis of linear problems, cf. [19]. The structural assumption (v) is satisfied in case that at least $h(x(t),t) = \bar{h}(D(t)x(t),t)$. In Section 2 we further assumed that

A1 $G_1(y,x,t) := A(t)D(t) + (B(t) + h_x(x,t))Q_0(t)$ is nonsingular for $t > 0$,

A2 $tG_1(y,x,t)^{-1}$ has a continuous extension for $t \to 0$,

A3 the decoupling function has the form, see (21),

$$D(t)\omega(u,t) = \frac{1}{t}M(t)u + g(u,t)$$

hold.

Next, we show that properties A1–A3 follow from (i)–(v):

We first note that (i), (iii) and (v) imply

$$\begin{aligned} Q_0(t)\tilde{G}_1(t)^{-1}h(x,t) &= Q_0(t)\tilde{G}_1(t)^{-1}[\underbrace{A(t)}_{\tilde{G}_1(t)P_0(t)} A(t)^- + I - A(t)A(t)^-]h(x,t) \\ &= Q_0(t)\tilde{G}_1(t)^{-1}[I - A(t)A(t)^-]h(P_0(t)x,t) = Q_0(t)\tilde{G}_1(t)^{-1}h(P_0(t)x,t), \end{aligned} \quad (47)$$

and additionally

$$Q_0(t)\tilde{G}_1(t)^{-1}h_x(x,t) = Q_0(t)\tilde{G}_1(t)^{-1}h_x(P_0(t)x,t)P_0(t).$$

Thus

$$Q_0(t)\tilde{G}_1(t)^{-1}h_x(x,t)Q_0(t) = 0.$$

Now, we show A1. For $t > 0$ we have

$$\begin{aligned} G_1(y,x,t) &= \tilde{G}_1(t) + h_x(x,t)Q_0(t) = \tilde{G}_1(t)(I + \tilde{G}_1(t)^{-1}h_x(x,t)Q_0(t)) \\ &= \tilde{G}_1(t)(I + P_0(t)\tilde{G}_1(t)^{-1}h_x(x,t)Q_0(t)). \end{aligned}$$

Since $I + P_0(t)\tilde{G}_1(t)^{-1}h_x(x,t)Q_0(t)$ has the inverse $I - P_0(t)\tilde{G}_1(t)^{-1}h_x(x,t)Q_0(t)$, the matrix $G_1(y,x,t)$ is nonsingular for $t > 0$ together with $\tilde{G}_1(t)$.

A2 is now a simple consequence of the representation

$$tG_1(y,x,t)^{-1} = (I - P_0(t)\tilde{G}_1(t)^{-1}h_x(x,t)Q_0(t))t\tilde{G}_1(t)^{-1}.$$

As a valuable byproduct we can see that the canonical projector has a continuous extension for $t \to 0$, since

$$\begin{aligned} Q_{can}(y,x,t) &:= Q_0(t)G_1(y,x,t)^{-1}(B(t) + h_x(x,t)) = Q_0(t)\tilde{G}_1(t)^{-1}(B(t) + h_x(x,t)) \\ &= Q_0(t)\tilde{G}_1(t)^{-1}B(t) + Q_0(t)\tilde{G}_1(t)^{-1}h_x(x,t). \end{aligned}$$

Now we turn to A3. The equation defining the decoupling function $\omega(u,t)$ reads:

$$A(t)D(t)w + B(t)(D(t)^-u + Q_0(t)w) + h(D(t)^-u + Q_0(t)w,t) = 0.$$

For $t > 0$ this is equivalent to

$$w = -\tilde{G}_1(t)^{-1}B(t)D(t)^-u - \tilde{G}_1(t)^{-1}h(D(t)^-u + Q_0(t)w,t).$$

Taking (47) into account we find

$$Q_0(t)w = -Q_0(t)\tilde{G}_1(t)^{-1}B(t)D(t)^-u - Q_0(t)\tilde{G}_1(t)^{-1}h(D(t)^-u,t).$$

This means that the function $\omega(u,t)$ satisfies

$$Q_0(t)\omega(u,t) = -Q_0(t)\tilde{G}_1(t)^{-1}B(t)D(t)^-u - Q_0(t)\tilde{G}_1(t)^{-1}h(D(t)^-u,t).$$

The right hand side of the above identity has a continuous extension for $t \to 0$. We now insert the last expression into the identity

$$D(t)\omega(u,t) = \underbrace{-D(t)\tilde{G}_1(t)^{-1}B(t)D(t)^-}_{=:\frac{1}{t}M(t)} u \underbrace{-D(t)\tilde{G}_1(t)^{-1}h(D(t)^-u + Q_0(t)\omega(u,t),t)}_{=:g(u,t)}$$

and see that A3 follows, with $M(t)$ and $g(u,t)$ continuous for $t \to 0$.
Finally, we show the uniform boundedness of the expression $t_{ij}\mathcal{G}_1(t_{ij})^{-1}$, which is important for the uniform convergence considered in the previous section. Derive (cf. the definitions in Section 4)

$$\begin{aligned}
\mathcal{G}_1(t) &:= \int_0^1 (\tilde{\mathcal{M}}(t,\tau)D(t) + \tilde{\mathcal{L}}(t,\tau)Q_0(t))d\tau \\
&= \underbrace{A(t)D(t) + B(t)Q_0(t)}_{\tilde{G}_1(t)} + \int_0^1 h_x(x_\star(t) + \tau(p(t) - x_\star(t)), t)Q_0(t)d\tau \\
&= \tilde{G}_1(t)\left\{ I + \int_0^1 \tilde{G}_1(t)^{-1}h_x(\cdots, t)Q_0(t)d\tau \right\} \\
&= \tilde{G}_1(t)\left\{ I + \int_0^1 P_0(t)\tilde{G}_1(t)^{-1}h_x(\cdots, t)Q_0(t)d\tau \right\} \\
&= \tilde{G}_1(t)\left\{ I + P_0(t)\int_0^1 \tilde{G}_1(t)^{-1}h_x(\cdots, t)d\tau Q_0(t) \right\} \\
&=: \tilde{G}_1(t)\mathcal{C}(t).
\end{aligned}$$

The matrix function $\mathcal{C}(t)$ is nonsingular, and it has the inverse $\mathcal{C}(t)^{-1} = I - P_0(t)\int_0^1 \tilde{G}_1(t)^{-1}h_x(\cdots, t)d\tau Q_0(t)$, which is bounded on $[0,1]$. As a consequence, $t_{ij}\mathcal{G}_1(t_{ij})^{-1} = \mathcal{C}(t_{ij})^{-1}t_{ij}\tilde{G}_1(t_{ij})^{-1}$ is uniformly bounded since $t_{ij}\tilde{G}_1(t_{ij})^{-1}$ is so.

6 Numerical Experiments

In this section we consider a nonlinear system of the form

$$f((D(t)x(t))', x(t), t) = A(t)(D(t)x(t))' + b(x(t), t) = \begin{cases} tx_1'(t) + b_1(x(t), t) = 0, \\ b_2(x(t), t) = 0, \end{cases}, \quad t \in [0,1],$$

where $n = 2$, $m = 4$, and

$$f(y, x, t) = \underbrace{\begin{pmatrix} tI \\ 0 \end{pmatrix}}_{A(t)} y + \underbrace{Bx + \tilde{B}(x)x + \beta(t)}_{b(x,t)},$$

$$D(t) = (I \quad 0), \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} -11 & -18 & 3 & -1 \\ 12 & 19 & -2 & 1 \\ 1 & 1 & 1 & 0 \\ 2 & 3 & 0 & \frac{1}{5} \end{pmatrix},$$

$$\tilde{B} = \begin{pmatrix} \tilde{B}_{11}(x) & \tilde{B}_{12}(x) \\ \tilde{B}_{21}(x_1) & \tilde{B}_{22}(x) \end{pmatrix} = \begin{pmatrix} \sin(x_{12}) & 0 & e^{-x_{11}} & 0 \\ 0 & \cos(x_{22}) & 0 & \sin(x_{11} + x_{21}) \\ x_{12}^3 & 0 & x_{11} & 0 \\ 0 & x_{11}x_{12} & 0 & x_{12}^2 \end{pmatrix}.$$

The inhomogeneity $\beta(t)$ is chosen in such a way that

$$x_\star(t) = \begin{pmatrix} t^2\sin(t) \\ te^t \\ t\cos(t) \\ \sin(t) \end{pmatrix}$$

is the exact solution and $\beta(0) = 0$ holds. Moreover, $y = Dx' \in \mathbb{R}^2$ and

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{pmatrix} \in \mathcal{D} := \{x \in \mathbb{R}^4 : x_{11} > -1\}, \quad Q_0(t) = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

The above problem satisfies the following assumptions: $G_1(y,x,t)$ is nonsingular for $t > 0$, $tG_1(y,x,t)^{-1}$ remains bounded for $t \to 0$, the canonical projector

$$Q_{can}(y,x,t) := Q_0(t)G_1(y,x,t)^{-1}b_x(x,t),$$

is continuous on $\mathbb{R}^2 \times \mathcal{D} \times [0,1]$, and the nonlinear inherent ODE is singular with a singularity of the first kind,

$$D(t)\omega(u,t) = \frac{1}{t}Mu + g(u,t),$$

where

$$M = -B_{11} + B_{12}B_{22}^{-1}B_{21} = \begin{pmatrix} 4 & 6 \\ -4 & -6 \end{pmatrix},$$

and $g(u,t)$ is smooth. Clearly, $u = x_1$. We now augment the DAE system by the necessary number of boundary conditions given by either

$$2x_{11}(0) + 3x_{12}(0) = 0, \tag{48a}$$
$$x_{11}(1) + x_{12}(1) = \sin(1) + \mathrm{e}, \tag{48b}$$

or

$$2x_{11}(0) + 3x_{12}(0) = 0, \tag{49a}$$
$$x_{11}(0) + x_{12}(0) = 0, \tag{49b}$$

and the consistency condition

$$b_2(x(0),0) = 0.$$

A careful analysis of the matrices $\mathcal{G}(t_{ij})$, cf. Section 4 on uniform convergence, shows that $\mathcal{G}(t_{ij})$ are nonsingular for all sufficiently fine grids, and $t_{ij}\mathcal{G}(t_{ij})^{-1}$ are uniformly bounded.

In the following tables, we have collected the results of our numerical experiments. All calculations have been carried out with MATLAB. In Table 1 we report on the global error of the solution x and its differential x_1 and algebraic x_2 components. In the upper part of the table we illustrate the asymptotical properties of the differential components x_1 with error at points τ_i defined as $\max_{\tau_i}\{\max\{|x_{11}(\tau_i)|, |x_{12}(\tau_i)|\}\}$. Similarly, the global error at the collocation points is given by $\max_{t_{ij}}\{\max\{|x_{11}(t_{ij})|, |x_{12}(t_{ij})|\}\}$. For the algebraic components the above quantities are specified in an analogous way. The order and the error constant are computed out of two consecutive steps in the usual fashion. We can see that the observed order of convergence is $k+1$ (since $k = 1$ is odd).
In the lower part of the tabular we report on the asymptotical behavior of the whole solution vector. The left column serves as an illustration for the estimate given in (38). In order to illustrate (45), in the right

column the maximal global error of x is calculated by considering its values at 1000 uniformly spaced points in the interval $[0, 1]$. In both cases, we also observe the convergence of order $k + 1$.

In Table 2 the analogous results are given for $k = 2$ and equidistant points in the upper and Gaussian points in the lower table. For the equidistant points, we observe the order of convergence $k = 2$, as predicted by the analysis. Note that due to the singularity, we so not observe the superconvergence order $2k$, not even for the differential components x_1 at the mesh points τ_i, cf. [12].

uniform mesh		differential components x_1 at points τ_i			differential components x_1 at points $t_{i,j}$		
N	h	error	order	const.	error	order	const.
10	1.00e-01	2.721e-02			1.630e-02		
20	5.00e-02	6.850e-03	2.0	2.657e+00	5.270e-03	1.6	6.946e-01
40	2.50e-02	1.717e-03	2.0	2.707e+00	1.567e-03	1.7	9.949e-01
80	1.25e-02	4.297e-04	2.0	2.736e+00	4.344e-04	1.9	1.448e+00
160	6.25e-03	1.074e-04	2.0	2.746e+00	1.149e-04	1.9	1.944e+00
320	3.13e-03	2.686e-05	2.0	2.749e+00	2.960e-05	2.0	2.367e+00
uniform mesh		algebraic components x_2 at points τ_i			algebraic components x_2 at points t_{ij}		
N	h	error	order	const.	error	order	const.
10	1.00e-01	1.440e-01			7.898e-02		
20	5.00e-02	4.164e-02	1.8	8.869e+00	2.822e-02	1.5	2.410e+00
40	2.50e-02	1.089e-02	1.9	1.372e+01	8.784e-03	1.7	4.380e+00
80	1.25e-02	2.755e-03	2.0	1.633e+01	2.486e-03	1.8	7.264e+00
160	6.25e-03	6.909e-04	2.0	1.729e+01	6.642e-04	1.9	1.045e+01
320	3.13e-03	1.729e-04	2.0	1.759e+01	1.719e-04	2.0	1.321e+01
uniform mesh		solution x at points t_{ij}			solution x at 1000 uniform points		
N	h	error	order	const.	error	order	const.
10	1.00e-01	7.898e-02			1.440e-01		
20	5.00e-02	2.822e-02	1.5	2.410e+00	4.164e-02	1.8	8.869e+00
40	2.50e-02	8.784e-03	1.7	4.380e+00	1.089e-02	1.9	1.372e+01
80	1.25e-02	2.486e-03	1.8	7.264e+00	2.755e-03	2.0	1.633e+01
160	6.25e-03	6.642e-04	1.9	1.045e+01	6.909e-04	2.0	1.729e+01
320	3.13e-03	1.719e-04	2.0	1.321e+01	1.729e-04	2.0	1.759e+01

Table 1: Numerical experiments with midpoint rule, collocation at one point $k = 1$.

uniform mesh		differential components x_1 at points τ_i			differential components x_1 at points $t_{i,j}$		
N	h	error	order	const.	error	order	const.
10	1.00e-01	2.172e-03			1.965e-03		
20	5.00e-02	5.252e-04	2.0	2.425e-01	5.004e-04	2.0	1.849e-01
40	2.50e-02	1.295e-04	2.0	2.231e-01	1.264e-04	2.0	1.913e-01
80	1.25e-02	3.216e-05	2.0	2.143e-01	3.179e-05	2.0	1.963e-01
160	6.25e-03	8.017e-06	2.0	2.099e-01	7.970e-06	2.0	1.998e-01
320	3.13e-03	2.001e-06	2.0	2.075e-01	1.995e-06	2.0	2.019e-01
uniform mesh		algebraic components x_2 at points τ_i			algebraic components x_2 at points τ_i		
N	h	error	order	const.	error	order	const.
10	1.00e-01	2.143e-02			1.808e-03		
20	5.00e-02	5.427e-03	2.0	2.055e+00	3.663e-04	2.3	3.631e-01
40	2.50e-02	1.361e-03	2.0	2.142e+00	8.308e-05	2.1	2.232e-01
80	1.25e-02	3.405e-04	2.0	2.169e+00	2.004e-05	2.1	1.607e-01
160	6.25e-03	8.515e-05	2.0	2.177e+00	4.935e-06	2.0	1.413e-01
320	3.13e-03	2.129e-05	2.0	2.179e+00	1.224e-06	2.0	1.339e-01
uniform mesh		solution x at points t_{ij}			solution x at 1000 uniform points		
N	h	error	order	const.	error	order	const.
10	1.00e-01	1.965e-03			2.143e-02		
20	5.00e-02	5.004e-04	2.0	1.849e-01	5.427e-03	2.0	2.055e+00
40	2.50e-02	1.264e-04	2.0	1.913e-01	1.361e-03	2.0	2.142e+00
80	1.25e-02	3.179e-05	2.0	1.963e-01	3.405e-04	2.0	2.169e+00
160	6.25e-03	7.970e-06	2.0	1.998e-01	8.515e-05	2.0	2.177e+00
320	3.13e-03	1.995e-06	2.0	2.019e-01	2.129e-05	2.0	2.179e+00

uniform mesh		differential components x_1 at points τ_i			differential components x_1 at points $t_{i,j}$		
N	h	error	order	const.	error	order	const.
10	1.00e-01	1.088e-04			1.617e-04		
20	5.00e-02	1.288e-05	3.1	1.305e-01	2.070e-05	3.0	1.494e-01
40	2.50e-02	1.567e-06	3.0	1.160e-01	2.639e-06	3.0	1.521e-01
80	1.25e-02	1.930e-07	3.0	1.085e-01	3.339e-07	3.0	1.582e-01
160	6.25e-03	2.394e-08	3.0	1.039e-01	4.202e-08	3.0	1.637e-01
320	3.13e-03	2.980e-09	3.0	1.010e-01	5.272e-09	3.0	1.676e-01
uniform mesh		algebraic components x_2 at points τ_i			algebraic components x_2 at points t_{ij}		
N	h	error	order	const.	error	order	const.
10	1.00e-01	1.632e-02			1.206e-03		
20	5.00e-02	4.121e-03	2.0	1.579e+00	1.492e-04	3.0	1.249e+00
40	2.50e-02	1.033e-03	2.0	1.631e+00	1.858e-05	3.0	1.213e+00
80	1.25e-02	2.584e-04	2.0	1.647e+00	2.321e-06	3.0	1.196e+00
160	6.25e-03	6.461e-05	2.0	1.652e+00	2.900e-07	3.0	1.190e+00
320	3.13e-03	1.615e-05	2.0	1.653e+00	3.625e-08	3.0	1.188e+00
uniform mesh		solution x at points t_{ij}			solution x at 1000 uniform points		
N	h	error	order	const.	error	order	const.
10	1.00e-01	1.206e-03			1.632e-02		
20	5.00e-02	1.492e-04	3.0	1.249e+00	4.121e-03	2.0	1.579e+00
40	2.50e-02	1.858e-05	3.0	1.213e+00	1.033e-03	2.0	1.631e+00
80	1.25e-02	2.321e-06	3.0	1.196e+00	2.584e-04	2.0	1.647e+00
160	6.25e-03	2.900e-07	3.0	1.190e+00	6.461e-05	2.0	1.652e+00
320	3.13e-03	3.625e-08	3.0	1.188e+00	1.615e-05	2.0	1.653e+00

Table 2: Numerical experiment for collocation with two collocation points $k = 2$; equidistant (top), and Gaussian points (bottom).

References

[1] P. Bailey, W. Everitt, and A. Zettl. Computing eigenvalues of singular Sturm-Liouville problems. *Results Math.*, 20:391–423, 1991.

[2] K. Balla and R. März. A unified approach to linear differential algebraic equations and their adjoint equations. *J. Anal. Appl.*, 21(3):783–802, 2002.

[3] C.J. Budd and R. Kuske. Localised periodic pattern for the non-symmetric generalized Swift-Hohenberg equations. *Physica D*, 208:73–95, 2005.

[4] C.J. Budd and J.F. Williams. Parabolic Monge-Ampère methods for blow-up problems in several spatial dimensions. *J. Phys. A*, 39:5425–5463, 2006.

[5] M. Drmota, R. Scheidl, H. Troger, and E.B. Weinmüller. On the imperfection sensitivity of complete spherical shells. *Comp. Mech.*, 2:63–74, 1987.

[6] G.B. Froment and K.B. Bischoff. *Chemical reactor analysis and design.* John Wiley & Sons Inc., New York, 1990.

[7] S. Golub. Measures of restrictions in inward foreign direct investment in OECD countries. *OECD Economics Dept. WP Nr. 357.*

[8] E. Helpman, M.J. Melitz, and S.R. Yeaple. Export versus FDI with heterogeneous firms. *Amer. Econ. Rev.*, 94(1):300–316, 2004.

[9] I. Higueras and R. März. Differential algebraic equations with properly stated leading terms. *Comp. Math. Appl.*, 48:215–235, 2004.

[10] I. Higueras, R. März, and C. Tischendorf. Stability preserving integration of index-1 DAEs. *Appl. Numer. Math*, 45:175–200, 2003.

[11] F.R. de Hoog and R. Weiss. Difference methods for boundary value problems with a singularity of the first kind. *SIAM J. Numer. Anal.*, 13:775–813, 1976.

[12] F.R. de Hoog and R. Weiss. Collocation methods for singular boundary value problems. *SIAM J. Numer. Anal.*, 15:198–217, 1978.

[13] T. Kapitula. Existence and stability of singular heteroclinic orbits for the Ginzburg-Landau equation. *Nonlinearity*, 9:669–685, 1996.

[14] B. Karabay. Foreign direct investment and host country policies: A rationale for using ownership restrictions. Technical Report, University of Virginia, WP, 2005.

[15] H. Keller. Approximation methods for nonlinear problems with application to two-point boundary value problems. *Math. Comp.*, 29:464–474, 1975.

[16] G. Kitzhofer, O. Koch, and E.B. Weinmüller. Kollokationsverfahren für singuläre Randwertprobleme zweiter Ordnung in impliziter Form. Techn. Rep. ANUM Preprint Nr. 9/04, Inst. for Anal. and Sci. Comput., Vienna Univ. of Technology, Austria, 2004. Available at `http://www.math.tuwien.ac.at/~inst115/preprints.htm`.

[17] G. Kitzhofer, O. Koch, G. Pulverer, Ch. Simon, and E.B. Weinmüller. Numerical treatment of singular BVPs: The new Matlab code `bvpsuite`. Accepted for *JNAIAM, J. Numer. Anal. Indust. Appl. Math.*

[18] O. Koch. Asymptotically correct error estimation for collocation methods applied to singular boundary value problems. *Numer. Math.*, 101:143–164, 2005.

[19] O. Koch, R. März, D. Praetorius, and E.B. Weinmüller. Collocation methods for linear index 1 DAEs with a singularity if the first kind. *Math. Comp.*, 79:281-304, 2010.

[20] O. Koch and E.B. Weinmüller. Analytical and numerical treatment of a singular initial value problem in avalanche modeling. *Appl. Math. Comput.*, 148(2):561–570, 2004.

[21] R. März. Differential algebraic equations anew. *Appl. Numer. Math.*, 42:315–335, 2002.

[22] R. März and R. Riaza. On linear differential-algebraic equations with properly stated leading terms: A-critical points. *Math. Comp. Model. Dyn. Sys.*, 13:291–314, 2004.

[23] D.M. McClung and A.I. Mears. Dry-flowing avalanche run-up and run-out. *J. Glaciol.*, 41(138):359–369, 1995.

[24] G. Moore. Geometric methods for computing invariant manifolds. *Appl. Numer. Math.*, 17:319–331, 1995.

[25] V.V. Ranade. *Computational flow modeling for chemical engineering.* Academic Press, San Diego, 2002.

[26] K. Sundmacher and U. Hoffmann. Multicomponent mass and energy transport on different length scales in a packed reactive distillation column for heterogeneously catalyzed fuel ether production. *Chem. Eng. Sci.*, 49:4443–4464, 1994.

[27] C.Y. Yeh, A.B. Chen, D.M. Nicholson, and W.H. Butler. Full-potential Korringa-Kohn-Rostoker band theory applied to the Mathieu potential. *Phys. Rev. B*, 42(17):10976–10982, 1990.