Analysis of a New Error Estimate for Collocation Methods Applied to Singular Boundary Value Problems

Winfried Auzinger
Othmar Koch
Ewa Weinmüller

to appear in SIAM J. Numer. Anal.

ANUM Preprint No. 13/02

TU WIEN

Technische
Universität
Wien

Vienna
University of
Technology

Institute for Applied Mathematics
and Numerical Analysis

ANALYSIS OF A NEW ERROR ESTIMATE FOR COLLOCATION METHODS APPLIED TO SINGULAR BOUNDARY VALUE PROBLEMS[*]

WINFRIED AUZINGER , OTHMAR KOCH , AND EWA WEINMÜLLER[†]

Abstract. We discuss an a-posteriori error estimate for the numerical solution of boundary value problems for nonlinear systems of ordinary differential equations with a singularity of the first kind. The estimate for the global error of an approximation obtained by collocation with piecewise polynomial functions is based on the defect correction principle. We prove that for collocation methods which are not superconvergent, the error estimate is asymptotically correct. As an essential prerequisite we derive convergence results for collocation methods applied to nonlinear singular problems.

Key words. Boundary value problems, singularity of the first kind, collocation methods, error estimate, defect correction, asymptotical correctness.

AMS subject classification. 65L05

1. Introduction.

In this paper, we discuss the numerical solution of singular boundary value problems of the form

$$z'(t) = \frac{M(t)}{t} z(t) + f(t, z(t)), \quad t \in (0, 1], \tag{1.1a}$$

$$B_a z(0) + B_b z(1) = \beta, \tag{1.1b}$$

$$z \in C[0, 1], \tag{1.1c}$$

where z is an n-dimensional real function, M is a smooth $n \times n$ matrix and f is an n-dimensional smooth function on a suitable domain. B_a and B_b are constant $r \times n$ matrices, with $r < n$. In §3 we will demonstrate that condition (1.1c) is equivalent to a set of $n - r$ linearly independent conditions $z(0)$ must satisfy. These boundary conditions are augmented by (1.1b) to yield an isolated solution z. In this paper, we restrict our attention to the class of singular boundary value problems which are equivalent to a wellposed singular initial value problem, where all boundary conditions are posed at $t = 0$. In this case, a shooting argument can be used to derive a representation of the solution convenient for our analysis. This implies certain restrictions on the spectrum of the matrix $M(0)$ which will be discussed in §3.

The search for an efficient numerical method to solve problems (1.1) is strongly motivated by numerous applications from physics, chemistry, mechanics, or ecology, see for example [15], [28]. Also, research activities in related fields, like the computation of connecting orbits in dynamical systems ([21]), or singular Sturm-Liouville problems ([6]), may benefit from techniques developed for problems of the form (1.1). The problem class discussed in this paper, where $M(0)$ has no eigenvalues with positive real parts, arises in applications from mechanics (buckling of spherical shells, [22], [25]), chemical reactor theory, cf. [11], or avalanche dynamics, see [18] and [19]. Moreover, Dirichlet problems for certain nonlinear elliptic equations lead to this problem class when certain symmetries are present, see [23]. The computation of self-similar solution profiles for the nonlinear Schrödinger equation is also essentially reduced to

[*] This project was supported by the Austrian Science Fund (FWF) under grant no. P-15072-MAT.

[†] All: Institute for Analysis and Scientific Computing, Vienna University of Technology, Austria.

this problem type, see [8]. However, our restriction on the spectrum of $M(0)$ excludes problems of the type

$$y''(t) + \frac{1}{t}y'(t) - \frac{1}{t^2}y(t) = f(t),$$

see [20], [22], from the treatment. The first order system resulting from the Euler transformation $z(t) = (y(t), ty'(t))$ does not belong to the class considered here.

To compute the numerical solution of (1.1), we use polynomial collocation at collocation points placed in the interior of every collocation interval. Collocation has been used in one of the best established standard codes for (regular) boundary value problems, COLSYS (COLNEW), see [1] and [2]. In COLSYS, (superconvergent) collocation at Gaussian points is used, cf. [7]. Our decision to use collocation was motivated by its advantageous convergence properties for (1.1), while in the presence of a singularity other high order methods show order reductions and become inefficient (see for example [14]). For linear problems (1.1) which can equivalently be posed as initial value problems, it was shown in [13] that the convergence order of collocation methods is at least equal to the *stage order* of the method. We will discuss the restrictions implied by the latter requirement in §3. For the general class (1.1), numerical evidence suggests that the convergence order is at least equal to the stage order for both the linear and the nonlinear case[1], cf. [5]. However, we cannot expect to observe superconvergence (cf. [7]) when collocation is applied to (1.1) in general. At most, a convergence order of $O(|\ln(h)|^{n_0-1}h^{m+1})$, for some positive integer n_0, holds for a method of stage order m, see [13]. Consequently, a restriction to collocation at an even number of equidistant points, which implies that the convergence order is at most $O(h^m)$, does not limit the method's accuracy significantly. We use these collocation nodes in practice, since it turns out that the error of the error estimate we propose in this paper is $O(|\ln(h)|^{n_0-1}h^{m+1})$. This means that the estimate is asymptotically correct when the order of the collocation method is not higher than the stage order.

Our main aim was to construct an efficient asymptotically correct error estimate for the global error of the numerical solution obtained by collocation. This estimate, introduced in [5], is based on the defect correction principle, which was first considered in [29] for the estimation of the global error of Runge-Kutta methods. In [29], the estimate for the error at the mesh points is obtained by applying the (high order) basic numerical scheme twice, to the original and to a suitably defined 'neighboring problem'. An extension of this idea proposed in [9], [24] avoids the second application of the high order scheme, using a cheap low order method instead. Again, this estimate is asymptotically correct at the mesh points only. A further modification proposed by the authors provides an error estimate which is asymptotically correct at both, the mesh and the collocation points. The analysis of this estimate in the context of nonlinear regular problems was given in [5]. It could be shown that for a collocation method of stage order $O(h^m)$, the error of the estimate (the difference between the global error and its estimate) is of order $O(h^{m+1})$. Numerical evidence suggests that this is also true for singular problems. In this paper, we will prove this assertion for the class of singular problems (1.1).

The collocation method and error estimate described in this paper were also implemented in the MATLAB code `sbvp` designed especially to solve singular boundary

[1] The analysis given in [26] for second order problems might provide tools to prove this assertion.

value problems. The error estimate yields a reliable basis for a mesh selection procedure which enables an efficient computation of the numerical solution. A description of the code and experimental evidence of its advantageous properties are given in [4].

The paper is organized as follows: The analytical properties of (1.1) which were discussed in detail in [12] are briefly recapitulated in §3. In §4.1, the results for collocation methods according to [13] are given. Using these results, we derive new, refined bounds for the errors of the numerical solution and its derivative, and extend these results to the nonlinear case. This analysis is carried out in §4.2. In §5 we use these estimates for the collocation solution in order to prove that our version of the error estimate is asymptotically correct for problem (1.1). Finally, in §6 we give a numerical example which illustrates the theory.

2. Preliminaries. Throughout the paper, the following notation is used. We denote by $\mathbb{R}^n$ the space of real vectors of dimension n and use $|\cdot|$,

$$|x| = |(x_1, x_2, \ldots, x_n)^T| := \max_{1\le i\le n} |x_i|,$$

to denote the maximum norm in $\mathbb{R}^n$. $C^p_n[0,1]$ is the space of real vector-valued functions which are p times continuously differentiable on $[0,1]$. For functions $y \in C^0_n[0,1]$ we define the maximum norm,

$$\|y\|_{[0,1]} := \max_{0\le t\le 1} |y(t)|,$$

or more generally for an interval $J \subseteq [0,1]$,

$$\|y\|_J := \max_{t\in J} |y(t)|.$$

$C^p_{n\times n}[0,1]$ is the space of real $n\times n$ matrices with columns in $C^p_n[0,1]$. For a matrix $A = (a_{ij})^n_{i,j=1}$, $A \in C^0_{n\times n}[0,1]$, $\|A\|_{[0,1]}$ is the induced norm,

$$\|A\|_{[0,1]} = \max_{0\le t\le 1} |A(t)| = \max_{0\le t\le 1} \left(\max_{1\le i\le n} \sum_{j=1}^n |a_{ij}(t)| \right).$$

Where there is no confusion, we will omit the subscripts n and $n\times n$ and denote $C[0,1] = C^0[0,1]$.

For the numerical analysis, we define meshes

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

and $h_i := \tau_{i+1} - \tau_i,\ i = 0, \ldots, N-1,\ \tau_0 = 0,\ \tau_N = 1$. On Δ, we define corresponding grid vectors

$$u_\Delta := (u_0, \ldots, u_N) \in \mathbb{R}^{(N+1)n}.$$

The norm on the space of grid vectors is given by

$$\|u_\Delta\|_\Delta := \max_{0\le k\le N} |u_k|.$$

For a continuous function $y \in C[0,1]$, we denote by R_Δ the pointwise projection onto the space of grid vectors,

$$R_\Delta(y) := (y(\tau_0), \ldots, y(\tau_N)).$$

For collocation, m points spaced at distances $h_i\delta_j,\ j = 1\ldots,m$, are inserted in each subinterval $J_i := [\tau_i, \tau_{i+1}]$. This yields the (fine) grid[2]

$$\Delta^m := \left\{ t_{i,j} :\ t_{i,j} = \tau_i + h_i \sum_{k=0}^{j} \delta_k,\ i = 0,\ldots,N-1,\ j = 0,\ldots,m+1 \right\}. \tag{2.1}$$

We restrict ourselves to grids where $\delta_1 > 0$ to avoid a special treatment of the singular point $t = 0$. For the analysis of collocation methods, we allow $\delta_{m+1} = 0$. In the discussion of the error estimate, we use the further restriction $\delta_{m+1} > 0$. This requirement is satisfied for equidistant collocation points which we use in practice (see §1), where

$$\delta_j := \frac{1}{m+1}, \quad j = 1,\ldots,m+1. \tag{2.2}$$

For a grid Δ^m, u_{Δ^m}, $\|\cdot\|_{\Delta^m}$ and R_{Δ^m} are defined accordingly.

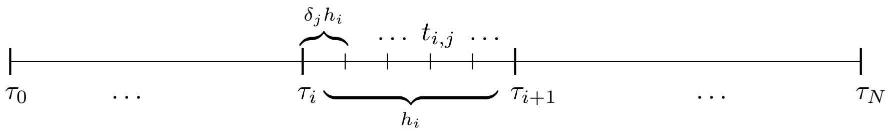

Fig. 2.1. *The computational grid*

3. Analytical results.

3. Analytical results. In this section we discuss the analytical properties of (1.1), cf. [12]. Here, we assume all eigenvalues of $M(0)$ to have nonpositive real parts. Moreover, the only eigenvalue of $M(0)$ on the imaginary axis is zero. These restrictions are necessary to ensure that we can use a shooting argument to derive a representation of the solution convenient for our theory[3].

First, we treat the linear case,

$$z'(t) = \frac{M(t)}{t} z(t) + f(t), \quad t \in (0,1], \tag{3.1a}$$

$$B_a z(0) + B_b z(1) = \beta, \tag{3.1b}$$

$$z \in C[0,1], \tag{3.1c}$$

where $B_a, B_b \in \mathbb{R}^{r\times n},\ r < n$, are constant matrices, and $\beta \in \mathbb{R}^r$ is a constant vector.

Throughout, we assume $M \in C^1[0,1]$. Consequently, we can rewrite $M(t)$ and obtain

$$M(t) = M(0) + tC(t) \tag{3.2}$$

with a continuous matrix $C(t)$.

Let X_0 be the kernel of $M(0)$ and let R be a projection onto X_0, where the rank of R is equal to r. We define

$$S := I_n - R,$$

[2] For convenience, we denote τ_i by $t_{i,0} \equiv t_{i-1,m+1},\ i = 1,\ldots,N-1$. Moreover, we define $\delta_0 := 0,\ \delta_{m+1} := (t_{i,m+1} - t_{i,m})/h_i$. Note that we choose the same distribution of collocation points in every subinterval J_i, and that $\sum_{j=0}^{m+1} \delta_j = 1$ holds for $i = 0,\ldots,N-1$.

[3] Note however that we do not use shooting when we actually compute the numerical solution.

where we denote by I_n the $n \times n$ identity matrix. The necessary and sufficient condition for z to be continuous on $[0,1]$ is

$$Sz(0) = 0.$$

This yields

$$z(0) = (S+R)z(0) = Rz(0),$$

and due to

$$M(0)z(0) = MRz(0) = 0$$

it follows that (3.1c) is equivalent to $z(0) \in \ker(M(0))$. These conditions are augmented by (3.1b) to yield a unique solution.

We denote by $\tilde{E}$ the $n \times r$ matrix consisting of a maximal set of linearly independent columns of R. Moreover, let $Z(t) = (Z_1(t), \dots, Z_r(t))$ be the fundamental solution matrix of the initial value problem

$$Z'(t) = \frac{M(t)}{t} Z(t), \quad t \in (0,1], \tag{3.3a}$$

$$Z(0) = \tilde{E}. \tag{3.3b}$$

The necessary and sufficient condition for problem (3.1) to have a unique solution is that the $r \times r$ matrix Q,

$$Q := B_a \tilde{E} + B_b Z(1) \tag{3.4}$$

is nonsingular. In this case, we can represent the solution z of (3.1) by

$$z(t) = \sum_{k=1}^{r} a_k Z_k(t) + \tilde{z}(t), \tag{3.5}$$

where $\tilde{z}$ is the solution of

$$\tilde{z}'(t) = \frac{M(t)}{t} \tilde{z}(t) + f(t), \quad t \in (0,1], \tag{3.6a}$$

$$\tilde{z}(0) = 0. \tag{3.6b}$$

The coefficients $a = (a_1, \dots, a_r)$ are uniquely determined by $Qa = \beta - B_b \tilde{z}(1)$.

For the solution of the linear problem (3.1), $z \in C^{k+1}[0,1]$ holds if $f \in C^k[0,1]$ and $M \in C^{k+1}[0,1]$.

Now we discuss the nonlinear problem[4]

$$z'(t) = \frac{M(t)}{t} z(t) + f(t, z(t)), \quad t \in (0,1], \tag{3.7a}$$

$$B_a z(0) + B_b z(1) = \beta, \tag{3.7b}$$

$$M(0)z(0) = 0. \tag{3.7c}$$

In order to formulate analogous smoothness properties for z, we make the following assumptions:

[4] Again, we assume that $M(0)$ has only eigenvalues with negative real parts or the eigenvalue 0.

1. $f : D_1 \to \mathbb{R}^n$ is a nonlinear mapping, where $D_1 \subseteq [0,1] \times \mathbb{R}^n$ is a suitable set.
2. Equation (3.7) has a solution $z \in C[0,1] \cap C^1(0,1]$. With this solution and a $\rho > 0$ we associate the closed balls
$$S_\rho(z(t)) := \{x \in \mathbb{R}^n : \ |z(t) - x| \le \rho\}$$
and the tube
$$T_\rho(z) := \{(t,x) : \ t \in [0,1], \ x \in S_\rho(z(t))\}.$$
3. $f(t,z)$ is continuously differentiable with respect to z, and $\frac{\partial f(t,z)}{\partial z}$ is continuous on $T_\rho(z)$.
4. The solution z is isolated. This means that
$$\begin{aligned} &u'(t) = \frac{M(t)}{t} u(t) + A(t)u(t), \quad t \in (0,1], \\ &B_a u(0) + B_b u(1) = 0, \\ &M(0)u(0) = 0, \end{aligned}$$
where
$$A(t) := \frac{\partial f}{\partial z}(t, z(t)),$$
has only the trivial solution.

Under these assumptions and for $f \in C^k(T_\rho(z)), \ M \in C^{k+1}[0,1]$, the solution z of (3.7) satisfies $z \in C^{k+1}[0,1]$.

For further details and proofs see [12].

4. Collocation methods. In this section, we derive new, refined error bounds for collocation methods applied to (1.1), relying on earlier results formulated in [13]. Moreover, we extend the convergence analysis to the nonlinear case. For reasons of simplicity, we restrict the discussion to equidistant meshes, $h_i = h, \ i = 0, \dots, N-1$, because the results from [13] are formulated for this situation. However, the results also hold for nonuniform meshes which have a limited variation in the stepsizes, see [13, §6].

Let us denote by B the Banach space of continuous, piecewise polynomial functions $q \in \mathbb{P}_m$ of degree $\le m, \ m \in \mathbb{N}$ (m is called the *stage order* of the method), equipped with the norm $\| \cdot \|_{[0,1]}$. As an approximation for the exact solution z of (1.1), we define an element of B which satisfies the differential equation (1.1a) at a finite number of points and which is subject to the same boundary conditions. Since we require the numerical solution to satisfy (1.1c), we introduce the space $B_1 \subset B$, such that $M(0)q(0) = 0, \ \forall\, q \in B_1$. Thus, we are seeking a function $p(t) = p_i(t), \ t \in J_i, \ i = 0, \dots, N-1$, in B_1 which satisfies

$$\text{(4.1a)} \quad p_i'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} p_i(t_{i,j}) + f(t_{i,j}, p_i(t_{i,j})), \quad i = 0, \dots, N-1, \ j = 1, \dots, m,$$
$$\text{(4.1b)} \quad B_a p(0) + B_b p(1) = \beta.$$

We consider collocation on general grids Δ^m as defined in §1, subject to the restriction $\delta_1 > 0$.

4.1. Earlier results. In [13], collocation methods for linear problems were studied. For the analysis of the nonlinear case in §4.2, bounds for the collocation solution $p \in B_1$ need to be specified. Here, the relevant preliminaries from [13] are recapitulated.

Thus, we consider the solution $p \in B_1$ of

$$p'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} p(t_{i,j}) + f(t_{i,j}), \quad i = 0, \ldots, N-1, \ j = 1, \ldots, m, \tag{4.2a}$$

$$B_a p(0) + B_b p(1) = \beta. \tag{4.2b}$$

LEMMA 4.1. *For $\mu, \beta \in \{0,1\}$ and arbitrary constants $c_{i,j}$, there exists a unique $p \in B_1$ which satisfies*

$$p'(t_{i,j}) = \frac{M(0)}{t_{i,j}} p(t_{i,j}) + \frac{M(0)^\mu}{t_{i,j}^\beta} c_{i,j}, \quad i = 0, \ldots, N-1, \ j = 1, \ldots, m, \tag{4.3a}$$

$$p(0) = 0. \tag{4.3b}$$

Furthermore,

$$\|p\|_{J_i} \le \text{const.}\, \tau_{i+1}^{1-\beta} |\ln(h)|^{(\beta(n_0-\mu))_+} C_i, \quad i = 0, \ldots, N-1, \tag{4.4}$$

where n_0 is the dimension of the largest Jordan block of $M(0)$ corresponding to the eigenvalue 0*,*

$$(x)_+ := \begin{cases} x, & x \ge 0, \\ 0, & x < 0, \end{cases}$$

and

$$C_i := \max_{\substack{l = 0, \ldots, i \\ j = 1, \ldots, m}} |c_{l,j}|.$$

Proof. See [13, Lemma 4.4]. □

The following result is a slightly modified version of [13, Theorem 4.1].

THEOREM 4.2. *For $\mu, \beta \in \{0,1\}$, consider the problem*

$$p'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} p(t_{i,j}) + \frac{M(0)^\mu}{t_{i,j}^\beta} c_{i,j}, \quad i = 0, \ldots, N-1, \ j = 1, \ldots, m, \tag{4.5a}$$

$$p(0) = \delta \in \ker(M(0)). \tag{4.5b}$$

There exists a unique solution of (4.5) *when h is sufficiently small, and this solution satisfies*

$$\|p\|_{J_i} \le \text{const.}\, (|\delta| + \tau_{i+1}^{1-\beta} |\ln(h)|^{(\beta(n_0-\mu))_+} C_i), \quad i = 0, \ldots, i_0, \tag{4.6}$$

for a suitable $i_0 \le N - 1$.

Proof. In [13, Theorem 4.1], the estimate following [13, (4.15)] can be replaced by

$$\|p\|_{J_i} \le \kappa(\tau_{i+1} \|p\|_{[0,\tau_{i+1}]} + |\delta| + \tau_{i+1}^{1-\beta} |\ln(h)|^{(\beta(n_0-\mu))_+} C_i), \quad i = 0, \ldots, i_0, \tag{4.7}$$

if the results of [13, Lemma 4.4] are suitably applied. Substitution of the bound for p derived in [13, Theorem 4.1] into the right-hand side of (4.7) yields the result. □

Note that the existence of the solution of (4.5) and the estimate (4.6) are shown only on an interval $[0, b]$, where b is sufficiently small (but independent of h). Thus, we need to use classical theory for regular problems to ensure the existence of the solution on the whole interval. In the sequel, we treat the underlying singular problems only on the restricted interval, and apply classical results for collocation from [3], and the error estimate analysis for regular problems from [5], to complete the proofs.

4.2. New error bounds. First, we use Theorem 4.2 to derive bounds for the solution $p \in B_1$ of the general linear problem (4.2) and for its derivative p'. By the superposition principle, p can be written in the form

$$p(t) = \sum_{k=1}^{r} b_k P_k(t) + \tilde{p}(t), \tag{4.8}$$

analogous to (3.5) for the exact solution. Here, $P(t) = (P_1(1), \dots, P_r(t))$ is the $n \times r$ matrix solution of

$$P'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} P(t_{i,j}), \quad i = 0, \dots, N-1, \ j = 1, \dots, m, \tag{4.9a}$$

$$P(0) = \tilde{E}, \tag{4.9b}$$

whose columns are in B_1, and $\tilde{p}$ satisfies

$$\tilde{p}'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} \tilde{p}(t_{i,j}) + f(t_{i,j}), \quad i = 0, \dots, N-1, \ j = 1, \dots, m, \tag{4.10a}$$

$$\tilde{p}(0) = 0. \tag{4.10b}$$

It was shown in [13, Theorem 4.4] that the representation (4.8) is well defined. Of course, the coefficients b_k could be computed in principle from the boundary conditions as in the case of the analytical problem; the representation (4.8) is used only to describe the structure of the solution p and therefore we refrain from specifying b_k explicitly. Next, we derive convergence results for the quantities appearing in the representation (4.8) using arguments similar to those given in [13, Theorem 4.2].

Consider the solutions z and q of (3.1a) and (4.2a), respectively, subject to the initial conditions $z(0) = q(0) = \delta \in \ker(M(0))$. We define an error function $e \in B_1$ by

$$\begin{aligned} &e'(t_{i,j}) = z'(t_{i,j}) - q'(t_{i,j}), \quad i = 0, \dots, N-1, \ j = 1, \dots, m, \\ &e(0) = 0. \end{aligned}$$

From standard results for interpolation, see for example [10], we conclude that

$$e(t) = z(t) - q(t) + tO(h^m)$$

if z is sufficiently smooth, whence

$$e'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} e(t_{i,j}) + O(h^m), \quad i = 0, \dots, N-1, \ j = 1, \dots, m, \tag{4.11a}$$

$$e(0) = 0. \tag{4.11b}$$

Now, Theorem 4.2 yields

$$\|e\|_{J_i} \leq \tau_{i+1} O(h^m), \quad i = 0, \dots, i_0, \tag{4.12}$$

and consequently

$$\|z - q\|_{J_i} \le \tau_{i+1} O(h^m), \quad i = 0, \dots, i_0. \tag{4.13}$$

It follows from (4.11a) and (4.12) that $e'(t_{i,j}) = O(h^m)$, which implies

$$\|z' - q'\|_{[0,1]} = O(h^m). \tag{4.14}$$

Finally, we show that the residual of q w.r.t. (3.1a) has the same asymptotic quality. Since $q \in C[0,1]$ and q' has only a finite number of jump discontinuities in $[0,1]$, we can use the representations

$$q(t) = \delta + t \int_0^1 q'(st)\, ds, \tag{4.15a}$$

$$z(t) = \delta + t \int_0^1 z'(st)\, ds \tag{4.15b}$$

to conclude that

$$\begin{aligned} q'(t) - \frac{M(t)}{t} q(t) - f(t) &= q'(t) - z'(t) + \frac{M(t)}{t} t \int_0^1 (q'(st) - z'(st))\, ds \\ &= O(h^m), \quad t \in [0,1]. \end{aligned} \tag{4.16}$$

This means that the refined bounds (4.13), (4.14) and (4.16) hold for the fundamental modes P_k and the particular solution $\tilde{p}$ in (4.8). To show that these bounds also hold for the solution p of (4.2), we have to estimate the differences $|a_k - b_k|$ for $k = 1, \dots, r$. We substitute (3.5) and (4.8) into (3.1b) and obtain a system of linear equations for $a_k - b_k$. This system is nonsingular since Q from (3.4) is nonsingular and $P(1) = Z(1) + O(h^m)$. This implies

$$b_k = a_k + O(h^m), \quad k = 1, \dots, r, \tag{4.17}$$

see also [13, Theorem 4.5].

Consequently, the following result holds.

Theorem 4.3. *Consider the solution $p \in B_1$ of (4.2) as an approximation of the (sufficiently smooth[5]) solution z of (3.1). Then, for a sufficiently small stepsize h and a suitable $i_0 \le N - 1$ the following bounds hold:*

$$z(t) - p(t) = \tilde{E} O(h^m) + \tau_{i+1} O(h^m), \quad t \in J_i, \ i = 0, \dots, i_0, \tag{4.18a}$$

$$\|z' - p'\|_{[0,1]} = O(h^m), \tag{4.18b}$$

$$\left| p'(t) - \frac{M(t)}{t} p(t) - f(t) \right| = O(h^m), \quad t \in [0,1]. \tag{4.18c}$$

Proof. The result follows immediately on noting that $P(t)$ can also be written in a form given by (4.15a) and therefore,

$$\begin{aligned} z(t) - p(t) &= \sum_{k=1}^{r} a_k (Z_k(t) - P_k(t)) + \sum_{k=1}^{r} (a_k - b_k) P_k(t) + \tilde{z}(t) - \tilde{p}(t) \\ &= \tau_{i+1} O(h^m) + (\tilde{E} + t O(1)) O(h^m) + \tau_{i+1} O(h^m), \quad t \in J_i. \end{aligned}$$

[5] We require that $z \in C^{m+1}[0,1]$, which holds if $f \in C^m[0,1]$ and $M \in C^{m+1}[0,1]$.

The bounds (4.18b) and (4.18c) are direct consequences of this representation. □

To prove the analogous convergence results for nonlinear problems, we use techniques developed in [17]. In order to show the existence of the solution and derive the error bounds, we rewrite the problem in an abstract Banach space setting and apply the Banach fixed point theorem. The arguments are similar to those given in the proof of [17, Theorem 3.6], but we cannot use this theorem directly, because some of the assumptions made there are violated and also, a refined error estimate is required. Therefore, we need to repeat the main steps of the proof.

We write the collocation problem as an operator equation

$$F(p) = 0, \tag{4.19}$$

where $F : B_1 \to B_2$ is defined by

$$F(p) = \begin{pmatrix} p'(t_{i,j}) - \frac{M(t_{i,j})}{t_{i,j}} p(t_{i,j}) - f(t_{i,j}, p(t_{i,j})), \quad i = 0, \dots, N-1,\ j = 1, \dots, m \\ B_a p(0) + B_b p(1) - \beta \end{pmatrix},$$

and B_1 and B_2 are Banach spaces,

$$B_1 = (\{q \in \mathbb{P}_m :\ M(0)q(0) = 0\}, \|\cdot\|_{[0,1]}), \quad B_2 = (\mathbb{R}^{Nmn+r}, |\cdot|).$$

For $p \in B_1$, the Fréchet derivative $DF(p) : B_1 \to B_2$ of F is given by

$$DF(p)q = \begin{pmatrix} q'(t_{i,j}) - \frac{M(t_{i,j})}{t_{i,j}} q(t_{i,j}) - D_2 f(t_{i,j}, p(t_{i,j}))q(t_{i,j}), \quad i = 0, \dots, N-1,\ j = 1, \dots, m \\ B_a q(0) + B_b q(1) \end{pmatrix},$$

where $D_2 f(t, z)$ is the Fréchet derivative of f with respect to z.

If $D_2 f$ is Lipschitz, then DF also satisfies a Lipschitz condition with the same constant,

$$\begin{aligned} |(DF(p_1) - DF(p_2))q| &= \left| \begin{pmatrix} (D_2 f(t_{i,j}, p_1(t_{i,j})) - D_2 f(t_{i,j}, p_2(t_{i,j})))q(t_{i,j}), \ \forall\, i, j \\ 0 \end{pmatrix} \right| \\ &\le L \|p_1 - p_2\|_{[0,1]} \|q\|_{[0,1]}. \end{aligned}$$

For the convergence proof, we require all assumptions from §3 to hold. In particular, this means that an isolated, smooth solution z of (1.1) exists. Using this function, we now construct an auxiliary element $p_{\mathrm{ref}} \in B_1$ for the proof of the existence of a solution p of (4.1). We require that p_{ref} satisfies

$$p_{\mathrm{ref}}'(t_{i,j}) = z'(t_{i,j}), \quad i = 0, \dots, N-1,\ j = 1, \dots, m, \tag{4.20a}$$

$$B_a p_{\mathrm{ref}}(0) + B_b p_{\mathrm{ref}}(1) = \beta. \tag{4.20b}$$

Since p_{ref}' is a piecewise polynomial of degree $\le m-1$, it is uniquely defined by the system (4.20a). Moreover,

$$\|z' - p_{\mathrm{ref}}'\|_{[0,1]} = O(h^m). \tag{4.21}$$

Representing p_{ref} by means of (4.15a) we conclude

$$z(t) - p_{\mathrm{ref}}(t) = \tilde{E}(r_1 - r_2) + tO(h^m), \quad r_1, r_2 \in \mathbb{R}^r.$$

Substitution into (4.20b) yields

$$(B_a + B_b)\tilde{E}(r_1 - r_2) = O(h^m).$$

For the further analysis, we assume that

$$\tilde{Q} := (B_a + B_b)\tilde{E} \quad \text{is nonsingular.} \tag{4.22}$$

This implies $r_1 - r_2 = O(h^m)$, and consequently,

$$z(t) - p_{\mathrm{ref}}(t) = \tilde{E}O(h^m) + tO(h^m). \tag{4.23}$$

Remark. Assumption (4.22) is quite natural. If we require that boundary value problems consisting of (1.1a) posed on intervals $(0, b]$, $0 < b \leq 1$, and boundary conditions $M(0)z(0) = 0$ and $B_a z(0) + B_b z(b) = \beta$, have unique, continuous solutions, then (4.22) follows. Moreover, we can interpret (4.20) as the (regular) collocation problem associated with the boundary value problem

$$\begin{aligned} &y'(t) = z'(t), \quad t \in (0,1], \\ &B_a y(0) + B_b y(1) = \beta, \\ &M(0)y(0) = 0. \end{aligned}$$

Obviously, $y(t) = z(t)$ is a solution of this reconstruction problem, and if we require the solution to be unique, then (4.22) must hold. Note that (4.22) always holds for problems with separated boundary conditions.

We now use (4.23) to derive the following relation:

$$\begin{aligned} F(p_{\mathrm{ref}}) &= \begin{pmatrix} p'_{\mathrm{ref}}(t_{i,j}) - \frac{M(t_{i,j})}{t_{i,j}} p_{\mathrm{ref}}(t_{i,j}) - f(t_{i,j}, p_{\mathrm{ref}}(t_{i,j})), \ \forall i,j \\ B_a p_{\mathrm{ref}}(0) + B_b p_{\mathrm{ref}}(1) - \beta \end{pmatrix} \\ &= \begin{pmatrix} p'_{\mathrm{ref}}(t_{i,j}) - z'(t_{i,j}) - \frac{M(t_{i,j})}{t_{i,j}} (p_{\mathrm{ref}}(t_{i,j}) - z(t_{i,j})) - \\ \qquad -f(t_{i,j}, p_{\mathrm{ref}}(t_{i,j})) + f(t_{i,j}, z(t_{i,j})), \ \forall i,j \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{M(0)}{t_{i,j}} (\tilde{E}O(h^m) + t_{i,j}O(h^m)) + O(h^m), \ \forall i,j \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} O(h^m) \\ 0 \end{pmatrix}. \end{aligned} \tag{4.24}$$

Finally, we give an estimate for $DF^{-1}(p_{\mathrm{ref}})$. Note that

$$q := DF^{-1}(p_{\mathrm{ref}}) \left(\begin{pmatrix} \gamma_{i,j}, \ \forall i,j \\ \tilde{\beta} \end{pmatrix} \right)$$

is the solution of the linear collocation problem

$$q'(t_{i,j}) = \frac{M(t_{i,j})}{t_{i,j}} q(t_{i,j}) + D_2 f(t_{i,j}, p_{\mathrm{ref}}(t_{i,j}))q(t_{i,j}) + \gamma_{i,j}, \ \forall i,j, \tag{4.25a}$$

$$B_a q(0) + B_b q(1) = \tilde{\beta}. \tag{4.25b}$$

Since for sufficiently small h, p_{ref} is in $T_\rho(z)$, this problem is well defined. Finally, from Theorem 4.2, we have

$$\|q\|_{J_i} \leq \mathrm{const.}\,(|\tilde{\beta}| + \tau_{i+1}\gamma_i), \tag{4.26}$$

where

$$\gamma_i = \max_{\substack{l=0,\dots,i\\ j=1,\dots,m}} |\gamma_{l,j}|.$$

With these preliminary results we can prove the next theorem.

Theorem 4.4. *Let z be an isolated, sufficiently smooth solution of* (1.1). *For sufficiently small h and $\rho > 0$, the nonlinear collocation scheme* (4.1) *has a unique solution p in the tube $T_\rho(z)$ around z. Moreover, the estimates* (4.18) *hold.*

Proof. We proceed in a manner similar to the proof of [17, Theorem 3.6]. Define a mapping $G : B_1 \to B_1$,

$$G(q) := q - DF^{-1}(p_{\mathrm{ref}})F(q). \tag{4.27}$$

Obviously, $F(p) = 0$ is equivalent to the fixed point equation $G(p) = p$. We use the Banach fixed point theorem to show that this equation has a unique solution in a suitably chosen closed ball

$$K := K(p_{\mathrm{ref}}, \rho_0) := \{q \in B_1 :\ \|q - p_{\mathrm{ref}}\|_{[0,1]} \le \rho_0\}.$$

To show that G is a contraction, we write

$$q := G(p_1) - G(p_2) = DF^{-1}(p_{\mathrm{ref}})(DF(p_{\mathrm{ref}}) - \widehat{DF}(p_1, p_2))(p_1 - p_2),$$

for $p_1, p_2 \in K$, where

$$\widehat{DF}(p_1, p_2) := \int_0^1 DF(\tau p_1 + (1-\tau)p_2)\, d\tau.$$

Consequently, q is the solution of the scheme (4.25), where

$$\begin{aligned}\left|\begin{pmatrix} \gamma_{i,j},\ \ \forall\, i,j \\ \tilde\beta \end{pmatrix}\right| &= \left|\int_0^1 (DF(p_{\mathrm{ref}}) - DF(\tau p_1 + (1-\tau)p_2))\, d\tau (p_1 - p_2)\right| \\ &\le L\rho_0 \|p_1 - p_2\|_{[0,1]},\end{aligned}$$

due to the Lipschitz condition DF satisfies. Thus, it follows from (4.26) that G is a contraction with constant $\tilde L < 1$ if ρ_0 is sufficiently small. To show that G maps K into itself, we estimate for $q \in K$,

$$\|p_{\mathrm{ref}} - G(q)\|_{[0,1]} \le \|p_{\mathrm{ref}} - G(p_{\mathrm{ref}})\|_{[0,1]} + \|G(p_{\mathrm{ref}}) - G(q)\|_{[0,1]},$$

where $p_{\mathrm{ref}} - G(p_{\mathrm{ref}}) = DF^{-1}(p_{\mathrm{ref}})F(p_{\mathrm{ref}})$ is the solution of (4.25) with $\gamma_{i,j} = O(h^m)$ and $\tilde\beta = 0$, cf. (4.24). Thus,

$$\|p_{\mathrm{ref}} - G(q)\|_{[0,1]} \le O(h^m) + \tilde L \rho_0 \le \rho_0 \tag{4.28}$$

provided that h is sufficiently small. The Banach fixed point theorem now implies that a solution $p \in B_1$ of (4.1) exists.

We now prove the convergence results (4.18). From

$$\begin{aligned}\|p_{\mathrm{ref}} - p\|_{J_i} = \|p_{\mathrm{ref}} - G(p)\|_{J_i} &\le \|p_{\mathrm{ref}} - G(p_{\mathrm{ref}})\|_{J_i} + \|G(p_{\mathrm{ref}}) - G(p)\|_{J_i} \\ &\le \tau_{i+1} O(h^m) + \tilde L \|p_{\mathrm{ref}} - p\|_{J_i}\end{aligned}$$

we have $\|p_{\mathrm{ref}} - p\|_{J_i} \le \tau_{i+1} O(h^m)$, which together with (4.23) yields

$$\begin{aligned} z(t) - p(t) &= z(t) - p_{\mathrm{ref}}(t) + p_{\mathrm{ref}}(t) - p(t) \\ &= \tilde{E} O(h^m) + t O(h^m) + \tau_{i+1} O(h^m), \quad t \in J_i. \end{aligned} \tag{4.29}$$

Consequently, (4.18a) follows. Next, we choose a piecewise polynomial function $e \in B_1$ satisfying $e'(t_{i,j}) = z'(t_{i,j}) - p'(t_{i,j})$. Therefore, $e'(t) = z'(t) - p'(t) + O(h^m)$. Moreover, (4.29) implies

$$\begin{aligned} e'(t_{i,j}) &= z'(t_{i,j}) - p'(t_{i,j}) \\ &= \frac{M(t_{i,j})}{t_{i,j}} (z(t_{i,j}) - p(t_{i,j})) + f(t_{i,j}, z(t_{i,j})) - f(t_{i,j}, p(t_{i,j})) \\ &= O(h^m), \quad i = 0, \ldots, N-1, \ j = 1, \ldots, m. \end{aligned}$$

Thus $e'(t) = O(h^m) = z'(t) - p'(t) + O(h^m)$ and (4.18b) follows. Finally, (4.18c) is shown by using (4.18b), (4.29), and the Lipschitz condition for f in

$$\begin{aligned} p'(t) &- \frac{M(t)}{t} p'(t) - f(t, p(t)) \\ &= p'(t) - z'(t) + \frac{M(t)}{t} (p(t) - z(t)) - f(t, p(t)) + f(t, z(t)) \\ &= O(h^m), \quad t \in [0,1]. \qquad \Box \end{aligned}$$

Under the previous assumptions we can also show that Newton's method converges quadratically when it is applied to compute the collocation solution p, provided that the starting approximation $p^{[0]}$ is chosen sufficiently close to p_{ref}.

THEOREM 4.5. *Let all assumptions of Theorem* 4.4 *hold. Newton's method converges quadratically to the solution* $p \in K(p_{\mathrm{ref}}, \rho_0)$ *of* (4.1) *if the starting iterate* $p^{[0]}$ *is chosen in a ball* $K(p_{\mathrm{ref}}, \rho_1)$, $\rho_1 \le \rho_0$, *provided that* ρ_0, ρ_1 *and the stepsize* h *are sufficiently small.*

Proof. The proof is analogous to that of [17, Theorem 3.7], taking into account the modifications made earlier in the proof of Theorem 4.4.

We write[6]

$$DF(q) = DF(p_{\mathrm{ref}})(I + DF^{-1}(p_{\mathrm{ref}})(DF(q) - DF(p_{\mathrm{ref}})))$$

for $q \in K(p_{\mathrm{ref}}, \rho_0)$ and use the bound for $DF^{-1}(p_{\mathrm{ref}})$, the Lipschitz condition for DF and the Banach lemma to show that $DF^{-1}(q)$ is bounded if ρ_0 is sufficiently small,

$$\|DF^{-1}(q)\|_{[0,1]} \le K_{\rho_0}, \tag{4.30}$$

where K_{ρ_0} is a constant depending on ρ_0. Furthermore, let $p^{[0]}$ in $K(p_{\mathrm{ref}}, \rho_1)$, then

$$\begin{aligned} p^{[1]} - p^{[0]} &= -DF^{-1}(p^{[0]}) F(p^{[0]}) \\ &= -DF^{-1}(p^{[0]}) F(p_{\mathrm{ref}}) + DF^{-1}(p^{[0]})(\widehat{DF}(p_{\mathrm{ref}}, p^{[0]})(p_{\mathrm{ref}} - p^{[0]})) \end{aligned}$$

[6] I is the identical mapping on the space of operators mapping $B_1 \to B_2$, that is, $I : DF(p_{\mathrm{ref}}) \mapsto DF(p_{\mathrm{ref}})$.

holds, with $\widehat{DF}(p_1,p_2)$ specified in Theorem 4.4. Using the Lipschitz condition for DF we obtain

$$\begin{aligned}
&\|DF^{-1}(p^{[0]})\widehat{DF}(p_{\mathrm{ref}},p^{[0]})(p_{\mathrm{ref}}-p^{[0]})\|_{[0,1]} \\
&\qquad = \|p_{\mathrm{ref}}-p^{[0]}+DF^{-1}(p^{[0]})(\widehat{DF}(p_{\mathrm{ref}},p^{[0]})-DF(p^{[0]}))(p_{\mathrm{ref}}-p^{[0]})\|_{[0,1]} \\
&\qquad \le \left(1+\frac{L\rho_1}{2}K_{\rho_0}\right)\rho_1 =: C\rho_1.
\end{aligned}$$

Finally, we conclude

$$\|p^{[1]}-p^{[0]}\|_{[0,1]} \le K_{\rho_0}O(h^m)+C\rho_1.$$

Consider a ball $K(p^{[0]},r)$. For a sufficiently small ρ_1 it is possible to choose the radius $r \le \rho_0$ in such a way that $K(p^{[0]},r) \subseteq K(p_{\mathrm{ref}},\rho_0)$. Moreover, let

$$\|DF^{-1}(p^{[0]})(DF(q_1)-DF(q_2))\|_{[0,1]} \le \omega(\|q_1-q_2\|_{[0,1]}),\ \ \forall\, q_1,q_2 \in K(p^{[0]},r),$$

and choose r such that the condition $\omega(r)=2K_{\rho_0}Lr \le 1/2$ holds. Consequently,

$$\|p^{[1]}-p^{[0]}\|_{[0,1]} \le K_{\rho_0}O(h^m)+C\rho_1 \le (1-2\omega(r))r,$$

provided that ρ_1 and h are sufficiently small, cf. [16, (6c)]. This implies that the assumptions of [16, Theorem 1] are satisfied and the quadratic convergence of Newton's method in $K(p^{[0]},r)$ follows. □

5. The error estimate. In this section, we analyze an error estimate based on the defect correction principle for the numerical solution p on the collocation grid Δ^m. For reasons explained in §1, it is sufficient for practical purposes to consider equidistant collocation, cf. (2.2), where we choose m even. However, the argument is valid on any collocation grid with $t_{i,m} < t_{i,m+1},\ i=0,\dots,N-1$.

Our estimate was introduced in [5], where it was shown to be asymptotically correct for regular problems. The numerical solution p obtained by collocation is used to define a 'neighboring problem' to (1.1). The original and neighboring problems are solved by the backward Euler method at the points $t_{i,j},\ i=0,\dots,N-1,\ j=1,\dots,m+1$. This yields the grid vectors[7] $\xi_{i,j}$ and $\pi_{i,j}$ as the solutions of the following schemes, subject to boundary conditions (1.1b) and (1.1c),

$$\text{(5.1a)}\qquad \frac{\xi_{i,j}-\xi_{i,j-1}}{t_{i,j}-t_{i,j-1}} = \frac{M(t_{i,j})}{t_{i,j}}\xi_{i,j}+f(t_{i,j},\xi_{i,j}),\quad \text{and}$$

$$\text{(5.1b)}\qquad \frac{\pi_{i,j}-\pi_{i,j-1}}{t_{i,j}-t_{i,j-1}} = \frac{M(t_{i,j})}{t_{i,j}}\pi_{i,j}+f(t_{i,j},\pi_{i,j})+\bar{d}_{i,j},$$

where $\bar{d}_{i,j}$ is a defect term defined by

$$\text{(5.2)}\quad \bar{d}_{i,j} := \frac{p(t_{i,j})-p(t_{i,j-1})}{t_{i,j}-t_{i,j-1}} - \sum_{k=1}^{m+1}\alpha_{j,k}\left(\frac{M(t_{i,k})}{t_{i,k}}p(t_{i,k})+f(t_{i,k},p(t_{i,k}))\right).$$

Here, the coefficients $\alpha_{j,k}$ are chosen in such a way that the quadrature rules given by

$$\frac{1}{t_{i,j}-t_{i,j-1}}\int_{t_{i,j-1}}^{t_{i,j}}\varphi(\tau)\,d\tau \approx \sum_{k=1}^{m+1}\alpha_{j,k}\varphi(t_{i,k})$$

[7] Here and in Theorem 5.1, we assume throughout $i=0,\dots,N-1,\ j=1,\dots,m+1$.

have precision $m+1$.

In the next theorem, we show that the difference $\xi_{\Delta^m} - \pi_{\Delta^m}$ is an asymptotically correct estimate for the global error of the collocation solution, $R_{\Delta^m}(z) - R_{\Delta^m}(p)$.

THEOREM 5.1. *Assume that the singular boundary value problem* (1.1) *has an isolated (sufficiently smooth*[8]*) solution z. Then, provided that h is sufficiently small, the following estimate holds:*

$$\|(R_{\Delta^m}(z) - R_{\Delta^m}(p)) - (\xi_{\Delta^m} - \pi_{\Delta^m})\|_{\Delta^m} = O(|\ln(h)|^{n_0-1}h^{m+1}), \tag{5.3}$$

with n_0 specified in Lemma 4.1.

Proof. The general idea of the proof is similar to that for regular problems. In particular, the smooth nonlinear part in the right-hand side of (1.1a) can be treated analogously. Therefore, we give a general outline of the proof here, and discuss those aspects which are crucial for the singular case. For further technical details we refer the reader to [5].

Let

$$\varepsilon_{\Delta^m} := \xi_{\Delta^m} - R_{\Delta^m}(z), \quad \bar{\varepsilon}_{\Delta^m} := \pi_{\Delta^m} - R_{\Delta^m}(p), \tag{5.4}$$

then the quantity to be estimated is

$$\tilde{\varepsilon}_{\Delta^m} := (R_{\Delta^m}(p) - R_{\Delta^m}(z)) - (\pi_{\Delta^m} - \xi_{\Delta^m}) = \varepsilon_{\Delta^m} - \bar{\varepsilon}_{\Delta^m}. \tag{5.5}$$

Here, ε_{Δ^m}, the error of the backward Euler scheme applied to the original problem, satisfies

$$\begin{aligned}\frac{\varepsilon_{i,j} - \varepsilon_{i,j-1}}{t_{i,j} - t_{i,j-1}} &= \frac{M(t_{i,j})}{t_{i,j}}\xi_{i,j} + f(t_{i,j}, \xi_{i,j}) - \frac{z(t_{i,j}) - z(t_{i,j-1})}{t_{i,j} - t_{i,j-1}} \\ &= \frac{M(t_{i,j})}{t_{i,j}}\xi_{i,j} + f(t_{i,j}, \xi_{i,j}) - \\ &\quad - \sum_{k=1}^{m+1} \alpha_{j,k}\left(\frac{M(t_{i,k})}{t_{i,k}} z(t_{i,k}) + f(t_{i,k}, z(t_{i,k}))\right) + O(h^{m+1}),\end{aligned} \tag{5.6}$$

since the $\alpha_{j,k}$ define quadrature rules of precision $O(h^{m+1})$. Moreover, $\bar{\varepsilon}_{\Delta^m}$ satisfies

$$\begin{aligned}\frac{\bar{\varepsilon}_{i,j} - \bar{\varepsilon}_{i,j-1}}{t_{i,j} - t_{i,j-1}} &= \frac{M(t_{i,j})}{t_{i,j}}\pi_{i,j} + f(t_{i,j}, \pi_{i,j}) + \bar{d}_{i,j} - \frac{p(t_{i,j}) - p(t_{i,j-1})}{t_{i,j} - t_{i,j-1}} \\ &= \frac{M(t_{i,j})}{t_{i,j}}\pi_{i,j} + f(t_{i,j}, \pi_{i,j}) - \\ &\quad - \sum_{k=1}^{m+1} \alpha_{j,k}\left(\frac{M(t_{i,k})}{t_{i,k}} p(t_{i,k}) + f(t_{i,k}, p(t_{i,k}))\right).\end{aligned} \tag{5.7}$$

Both (5.6) and (5.7) hold for $i = 0, \dots, N-1,\ j = 1, \dots, m+1$, and ε_{Δ^m} as well as $\bar{\varepsilon}_{\Delta^m}$ satisfy homogeneous boundary conditions.

In order to proceed, we use Taylor's Theorem to conclude that

$$\begin{aligned} f(t_{i,j}, \xi_{i,j}) - f(t_{i,j}, z(t_{i,j})) &= \int_0^1 D_2 f(t_{i,j}, z(t_{i,j}) + \tau(\xi_{i,j} - z(t_{i,j})))\, d\tau \cdot \varepsilon_{i,j} \\ &=: A(t_{i,j})\varepsilon_{i,j},\end{aligned} \tag{5.8}$$

[8] In fact, we require $z \in C^{m+2}[0,1]$.

and analogously,

$$f(t_{i,j},\pi_{i,j}) - f(t_{i,j},p(t_{i,j})) =: \bar{A}(t_{i,j})\bar{\varepsilon}_{i,j}. \tag{5.9}$$

Next, we note that due to (4.18c),

$$\begin{aligned}\bar{d}_{i,j} &= \frac{p(t_{i,j}) - p(t_{i,j-1})}{t_{i,j} - t_{i,j-1}} - \sum_{k=1}^{m+1} \alpha_{j,k}\left(\frac{M(t_{i,k})}{t_{i,k}}p(t_{i,k}) + f(t_{i,k},p(t_{i,k}))\right)\\ &= \frac{1}{t_{i,j}-t_{i,j-1}}\int_{t_{i,j-1}}^{t_{i,j}} p'(\tau)\,d\tau - \sum_{k=1}^{m+1}\alpha_{j,k}p'(t_{i,k}) +\\ &\quad + \alpha_{j,m+1}\left(p'(t_{i,m+1}) - \frac{M(t_{i,m+1})}{t_{i,m+1}}p(t_{i,m+1}) - f(t_{i,m+1},p(t_{i,m+1}))\right)\\ &= O(h^m). \end{aligned} \tag{5.10}$$

From this we conclude that $\xi_{i,j} = \pi_{i,j} + O(h^m)$ using the following argument:

The backward Euler schemes (5.1a) and (5.1b) can be written as collocation methods with $m = 1$ and the collocating condition posed at the right endpoint of each interval $[t_{i,j-1}, t_{i,j}]$. Thus, we discuss the collocation solutions $\xi(t)$, $\pi(t)$ of two singular boundary value problems whose right-hand sides differ by a term $O(h^m)$. This term can be assumed to be smooth, if a suitable interpolant g of $\bar{d}_{i,j}$ is used. More precisely, $\xi(t)$ is an approximation to the solution z of (1.1), and $\pi(t)$ is an approximation to the solution of

$$\begin{aligned}&z'_{\mathrm{def}}(t) = \frac{M(t)}{t}z_{\mathrm{def}}(t) + f(t,z_{\mathrm{def}}(t)) + g(t), \quad t \in (0,1],\\ &B_a z_{\mathrm{def}}(0) + B_b z_{\mathrm{def}}(1) = \beta,\\ &M(0)z_{\mathrm{def}}(0) = 0.\end{aligned}$$

For (1.1), we make the assumption that the analytical problem is stable in the sense that

$$\|z - z_{\mathrm{def}}\|_{[0,1]} \le \text{const.}\, \|g\|_{[0,1]} = O(h^m)$$

holds. For results on this type of stability analysis, see [27].

As in §4 we can prove that (locally) unique solutions $\xi(t)$ and $\pi(t)$ of (5.1a) and (5.1b) exist in a neighborhood of z and z_{def}, respectively.

Subtracting (5.1b) from (5.1a) and using Taylor expansion about $\pi(t_{i,j})$, we can show that $q(t) := \xi(t) - \pi(t)$ satisfies the linear scheme

$$\frac{q(t_{i,j}) - q(t_{i,j-1})}{t_{i,j} - t_{i,j-1}} = \frac{M(t_{i,j}) + t_{i,j}B(t_{i,j})}{t_{i,j}}q(t_{i,j}) + O(h^m),$$

with a suitable, bounded matrix B and homogeneous boundary conditions. Since this is equivalent to a collocation scheme, we may use [13, Theorem 4.4] with $\gamma = \delta = 0$ (or alternatively a combination of stability results from §4), to conclude that $\|R_{\Delta^m}(q)\|_{\Delta^m} = \|\xi_{\Delta^m} - \pi_{\Delta^m}\|_{\Delta^m} = O(h^m)$.

Since $\varepsilon_{i,j} = O(h)$ and $\bar{\varepsilon}_{i,j} = O(h)$, we may finally write (see [5])

$$\bar{A}(t_{i,j})\bar{\varepsilon}_{i,j} = A(t_{i,j})\bar{\varepsilon}_{i,j} + (\bar{A}(t_{i,j}) - A(t_{i,j}))\bar{\varepsilon}_{i,j} = A(t_{i,j})\bar{\varepsilon}_{i,j} + O(h^{m+1}).$$

Now we use (5.8), (5.9) to rewrite (5.6), (5.7) and obtain

$$\begin{aligned}\frac{\varepsilon_{i,j}-\varepsilon_{i,j-1}}{t_{i,j}-t_{i,j-1}} &= \frac{M(t_{i,j})}{t_{i,j}}\varepsilon_{i,j} + A(t_{i,j})\varepsilon_{i,j} + \frac{M(t_{i,j})}{t_{i,j}} z(t_{i,j}) + f(t_{i,j}, z(t_{i,j})) - \\ &\quad - \sum_{k=1}^{m+1} \alpha_{j,k}\left(\frac{M(t_{i,k})}{t_{i,k}} z(t_{i,k}) + f(t_{i,k}, z(t_{i,k}))\right) + O(h^{m+1}),\end{aligned} \tag{5.11}$$

and

$$\begin{aligned}\frac{\bar\varepsilon_{i,j}-\bar\varepsilon_{i,j-1}}{t_{i,j}-t_{i,j-1}} &= \frac{M(t_{i,j})}{t_{i,j}}\bar\varepsilon_{i,j} + A(t_{i,j})\bar\varepsilon_{i,j} + \frac{M(t_{i,j})}{t_{i,j}} p(t_{i,j}) + f(t_{i,j}, p(t_{i,j})) - \\ &\quad - \sum_{k=1}^{m+1} \alpha_{j,k}\left(\frac{M(t_{i,k})}{t_{i,k}} p(t_{i,k}) + f(t_{i,k}, p(t_{i,k}))\right) + O(h^{m+1}).\end{aligned} \tag{5.12}$$

Systems (5.11) and (5.12) are a pair of 'parallel' backward Euler schemes, with related inhomogeneous terms. Let us use the shorthand notation $\phi(t) := f(t,p(t)) - f(t,z(t))$. It can be shown that for the difference in the smooth parts of the inhomogeneous terms the estimate

$$\begin{aligned}|\phi(t_{i,j}) - \sum_{k=1}^{m+1} \alpha_{j,k}\phi(t_{i,k})| &\le \text{const.}\, h_i \|\phi'\|_{J_i} \\ &\le \text{const.}\, h_i (\|z-p\|_{J_i} + \|z'-p'\|_{J_i}) \le O(h^{m+1})\end{aligned} \tag{5.13}$$

holds. To see this we use Taylor expansion of $\phi(t_{i,k})$ about $t_{i,j}$ and the fact that $\sum_{k=1}^{m+1} \alpha_{j,k} = 1$ for all j, see [5]. The estimate finally follows from Theorem 4.4.

In the next step, we derive a representation for the difference in the singular terms occurring in the inhomogeneous parts of the schemes (5.11) and (5.12). With $\epsilon(t) := z(t) - p(t)$ and with $\sigma := t_{i,j} + \tau(t_{i,k} - t_{i,j})$, we rewrite

$$\begin{aligned}&\frac{M(t_{i,j})}{t_{i,j}}\epsilon(t_{i,j}) - \sum_{k=1}^{m+1}\alpha_{j,k}\frac{M(t_{i,k})}{t_{i,k}}\epsilon(t_{i,k}) \\ &= \frac{M(t_{i,j})}{t_{i,j}}\epsilon(t_{i,j}) - \sum_{k=1}^{m+1}\alpha_{j,k}\Bigg(\frac{M(t_{i,j})}{t_{i,j}}\epsilon(t_{i,j}) + \\ &\quad + \int_0^1 \frac{d}{d\sigma}\left(\frac{M(\sigma)}{\sigma}\epsilon(\sigma)\right) d\tau (t_{i,k} - t_{i,j})\Bigg) \\ &= \sum_{k=1}^{m+1}\alpha_{j,k}(t_{i,j} - t_{i,k}) \int_0^1 \Bigg(\frac{M(0)}{\sigma}\epsilon'(\sigma) - \\ &\quad - \frac{M(0)}{\sigma^2}\epsilon(\sigma) + C'(\sigma)\epsilon(\sigma) + C(\sigma)\epsilon'(\sigma)\Bigg) d\tau \\ &= \frac{M(0)}{t_{i,j}} O(h^{m+1}) + O(h^{m+1}),\end{aligned} \tag{5.14}$$

on noting that

$$\frac{1}{\sigma} \le \frac{m}{t_{i,j}},\ k = 1, \dots, m+1,\ j = 1, \dots, m,\ \tau \in [0,1],$$

and using the results of Theorem 4.4.

Altogether, we have shown that the error of the error estimate $\tilde{\varepsilon}_{\Delta^m}$, cf. (5.5), satisfies a linear Euler difference scheme

$$\text{(5.15a)}\quad \frac{\tilde{\varepsilon}_{i,j}-\tilde{\varepsilon}_{i,j-1}}{t_{i,j}-t_{i,j-1}} = \frac{M(t_{i,j})}{t_{i,j}}\tilde{\varepsilon}_{i,j} + A(t_{i,j})\tilde{\varepsilon}_{i,j} + \frac{M(0)}{t_{i,j}}O(h^{m+1}) + O(h^{m+1}),\ \ \forall\, i,j,$$

$$\text{(5.15b)}\quad B_a\tilde{\varepsilon}_{0,0} + B_b\tilde{\varepsilon}_{N-1,m+1} = 0,$$

$$\text{(5.15c)}\quad M(0)\tilde{\varepsilon}_{0,0} = 0.$$

This scheme can also be interpreted as a collocation scheme with $m=1$ where the only collocation point is the right endpoint of every collocation interval. To estimate the solution of (5.15) we use a representation according to (4.8) for $\tilde{\varepsilon}_{\Delta^m}$. Then we apply Theorem 4.2 to derive bounds for the quantities occurring in (4.8), and conclude that altogether the estimate (5.3) holds for the solution of (5.15). □

Remark. Obviously, the arguments used to prove the last theorem are valid for any choice of collocation nodes. The only necessary restriction is $t_{i,m+1} > t_{i,m}$. However, if we consider superconvergent schemes, the error estimate is no longer asymptotically correct, because the basic collocation solution has a higher convergence order in that case. Therefore we restrict ourselves to an even number of equidistant collocation points. This restriction is not severe, since in the case of singular problems, the highest convergence order that can generally be expected at the mesh points τ_i is $O(|\ln(h)|^{n_0-1}h^{m+1})$, see [13].

Finally, we would like to mention an alternative variant of our error estimate closely related to the so-called “Version B” of defect correction according to Stetter [24]. If instead of (5.1) we solve

$$\text{(5.16)}\qquad \frac{\zeta_{i,j}-\zeta_{i,j-1}}{t_{i,j}-t_{i,j-1}} = \frac{M(t_{i,j})}{t_{i,j}}\zeta_{i,j} + D(t_{i,j})\zeta_{i,j} - \bar{d}_{i,j},$$

where

$$D(t_{i,j}) := D_2 f(t_{i,j}, p(t_{i,j})),$$

then $\zeta_{i,j}$ is an asymptotically correct error estimate. To see this, we note that the difference between this error estimate and the estimate analyzed earlier in this paper,

$$x_{i,j} := (\xi_{i,j} - \pi_{i,j}) - \zeta_{i,j}$$

satisfies

$$\begin{aligned}\frac{x_{i,j}-x_{i,j-1}}{t_{i,j}-t_{i,j-1}} &= \frac{M(t_{i,j})}{t_{i,j}}x_{i,j} + f(t_{i,j},\xi_{i,j}) - f(t_{i,j},\pi_{i,j}) - D(t_{i,j})\zeta_{i,j}\\ &= \frac{M(t_{i,j})}{t_{i,j}}x_{i,j} + O(h^{m+1}).\end{aligned}$$

Consequently, the error of this error estimate has a bound analogous to (5.3). Note that for linear problems, this alternative error estimate coincides with the variant discussed earlier in the paper. For nonlinear problems, the practical usability and numerical stability of the new estimate still has to be carefully assessed.

6. Numerical examples. To illustrate the theory, we first consider the following nonlinear problem:

$$\text{(6.1a)}\quad z'(t) = \frac{1}{t}\begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} z(t) + t\begin{pmatrix} 0 \\ -\frac{2(t^2+2)+8}{(t^2+2)^2} z_1^2(t) + \frac{8t^2}{(t^2+2)^2} z_1^3(t) \end{pmatrix},$$

$$\text{(6.1b)}\quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} z(0) + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} z(1) = \begin{pmatrix} 0 \\ 1/\ln(3) \end{pmatrix}.$$

Its exact solution is

$$z(t) = \left(\frac{1}{\ln(t^2+2)}, -\frac{2t^2}{(t^2+2)\ln^2(t^2+2)} \right)^T.$$

The computations were carried out with the subroutines from our MATLAB code `sbvp` (cf. [4]) on fixed, equidistant grids. For the purpose of determining the empirical convergence orders the mesh adaptation strategy was disabled. The tests were performed in IEEE double precision with EPS $\approx 1.11 \cdot 10^{-16}$. In Table 6.1, we give the exact global errors $\mathrm{err}_{\mathrm{coll}}$ of the collocation solutions for the respective mesh width h, and the convergence orders p_{coll} computed from the errors for two consecutive stepsizes. Moreover, the errors of the error estimate with respect to the exact global errors, $\mathrm{err}_{\mathrm{est}}$, are recorded, together with associated empirical convergence orders p_{est}. In accordance with the theoretical results from §§4–5, convergence orders $O(h^4)$ for collocation and $O(h^5)$ for the error estimate are observed. This illustrates the asymptotical correctness of the error estimate analyzed in this paper. Test runs given in [4] demonstrate that this error estimate can be used as a dependable basis for a mesh adaptation algorithm, providing an efficient, high precision numerical solver.

Table 6.1
Convergence orders of collocation and error estimate for (6.1)

h	$\mathrm{err}_{\mathrm{coll}}$	p_{coll}	$\mathrm{err}_{\mathrm{est}}$	p_{est}
2^{-2}	1.5763e−04		2.2232e−05	
2^{-3}	9.5865e−06	4.04	6.5978e−07	5.07
2^{-4}	5.9574e−07	4.01	1.7873e−08	5.21
2^{-5}	3.7189e−08	4.00	5.1077e−10	5.13
2^{-6}	2.3237e−09	4.00	1.5205e−11	5.07
2^{-7}	1.4522e−10	4.00	4.6274e−13	5.04
2^{-8}	9.0772e−12	4.00	1.4655e−14	4.98

Finally, we demonstrate the favorable performance of our error estimate for a practically relevant example from applications. The following boundary value problem is a model from the theory of shallow spherical shells, see [22], [25]. The transformation of the original two-dimensional system of second order to the first order form, yields

$$\text{(6.2)}\quad z'(t) = \frac{1}{t}\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix} z(t) + t\begin{pmatrix} 0 \\ 0 \\ z_2(t)(-\mu^2 + z_1(t)) - 2\gamma \\ z_1(t)(\mu^2 - \frac{1}{2}z_1(t)) \end{pmatrix},$$

where the eigenvalues of $M(0)$ are $\lambda = 0,\ 0,\ -2,\ -2$. The boundary conditions read $z_3(0) = z_4(0) = z_1(1) = 0$, $z_4(1) + 2/3z_2(1) = 0$, and the parameters are chosen as $\mu = 9$, $\gamma = 6000$. We solve (6.2) using our code `sbvp` ([4]) equipped with the error estimate from §5 and adaptive mesh selection routine. The numerical solution

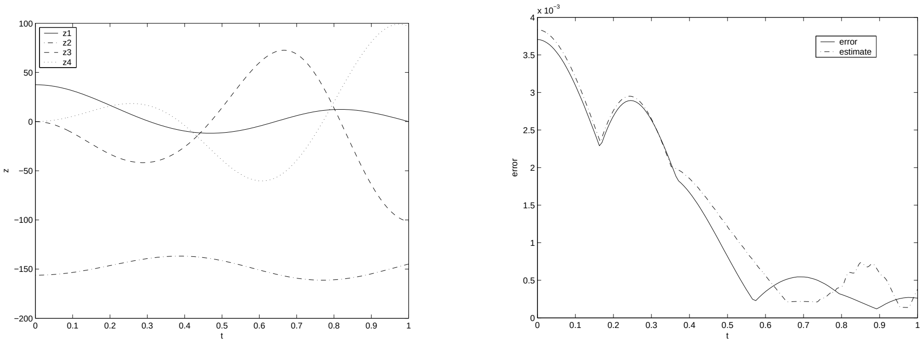

Fig. 6.1. *Solution, global error and error estimate for (6.2).*

satisfies a mixed tolerance requirement with absolute and relative tolerance equal to 10^{-4} at a mesh containing 124 mesh points, where the variation in the mesh width is just below 2. In Figure 6.1 four components of the numerical solution are given and the estimate of the global error on the final mesh is compared with the error of the collocation solution. In order to calculate the error of the collocation solution we used a reference solution computed with tolerances $5 \cdot 10^{-6}$. The maximum of the error estimate is 0.0038367, and the maximum of the error with respect to the reference solution is 0.003706. For the most part of the integration interval, the estimate slightly overestimates the "true" error.

Acknowledgment. We wish to thank the referees for their valuable suggestions, in particular for pointing out the alternative variant (5.16).

REFERENCES

[1] U. Ascher, J. Christiansen, and R. Russell, *A collocation solver for mixed order systems of boundary values problems*, Math. Comp., 33 (1978), pp. 659–679.

[2] ———, *Collocation software for boundary value ODEs*, ACM Transactions on Mathematical Software, 7 (1981), pp. 209–222.

[3] U. Ascher, R. Mattheij, and R. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[4] W. Auzinger, G. Kneisl, O. Koch, and E. Weinmüller, *A collocation code for boundary value problems in ordinary differential equations*, Numer. Algorithms, 33 (2003), pp. 27–39.

[5] W. Auzinger, O. Koch, and E. Weinmüller, *Efficient collocation schemes for singular boundary value problems*, Numer. Algorithms, 31 (2002), pp. 5–25.

[6] P. Bailey, W. Everitt, and A. Zettl, *Computing eigenvalues of singular Sturm-Liouville problems*, Results in Mathematics, 20 (1991), pp. 391–423.

[7] C. d. Boor and B. Swartz, *Collocation at Gaussian points*, SIAM J. Numer. Anal., 10 (1973), pp. 582–606.

[8] C. J. Budd, O. Koch, and E. Weinmüller, *Self-similar blow-up in nonlinear PDEs.* In preparation.

[9] R. Frank and C. Überhuber, *Iterated Defect Correction for differential equations, Part I: Theoretical results*, Computing, 20 (1978), pp. 207–228.

[10] F. B. Hildebrand, *Introduction to Numerical Analysis*, McGraw-Hill, New York, 2^{nd} ed., 1974.

[11] V. Hlavacek, M. Marek, and M. Kubicek, *Modelling of chemical reactors*, Chem. Eng. Sci., 23 (68), pp. 1083–1097.

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

[13] ———, *Collocation methods for singular boundary value problems*, SIAM J. Numer. Anal., 15 (1978), pp. 198–217.

[14] ———, *The application of Runge-Kutta schemes to singular initial value problems*, Math. Comp., 44 (1985), pp. 93–103.

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

[16] H. Keller, *Newton's method under mild differentiability conditions*, J. Comput. System Sci., 4 (1970), pp. 15–28.

[17] ———, *Approximation methods for nonlinear problems with application to two-point boundary value problems*, Math. Comp., 29 (1975), pp. 464–474.

[18] O. Koch and E. Weinmüller, *Analytical and numerical treatment of a singular initial value problem in avalanche modeling*, Appl. Math. Comput., 148 (2003), pp. 561–570.

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

[20] H. Meissner and P. Tholfsen, *Cylindrically symmetric solutions of the Ginzburg-Landau equation*, Phys. Rev., 169 (1968), pp. 413–416.

[21] G. Moore, *Geometric methods for computing invariant manifolds*, Appl. Num. Math., 17 (1995), pp. 319–331.

[22] P. Rentrop, *Eine Taylorreihenmethode zur numerischen Lösung von Zwei-Punkt Randwertproblemen mit Anwendung auf singuläre Probleme der nichtlinearen Schalentheorie*. TUM-MATH-7733. Technische Universität München 1977.

[23] R. Russell and L. Shampine, *Numerical methods for singular boundary value problems*, SIAM J. Numer. Anal., 12 (1975), pp. 13–36.

[24] H. J. Stetter, *The defect correction principle and discretization methods*, Numer. Math., 29 (1978), pp. 425–443.

[25] H. J. Weinitschke, *On the stability problem of shallow spherical shells*, JMP, 38 (1958), pp. 209–235.

[26] E. Weinmüller, *Collocation for singular boundary value problems of second order*, SIAM J. Numer. Anal., 23 (1986), pp. 1062–1095.

[27] ———, *Stability of singular boundary value problems and their discretization by finite differences*, SIAM J. Numer. Anal., 26 (1989), pp. 180–213.

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

[29] P. Zadunaisky, *On the estimation of errors propagated in the numerical integration of ODEs*, Numer. Math., 27 (1976), pp. 21–39.