BIT
1999, Vol. 39, No. 3, pp. 484–512

0006-3835/99/3903-0484 $12.00
© Swets & Zeitlinger

INDEXES AND SPECIAL DISCRETIZATION METHODS FOR LINEAR PARTIAL DIFFERENTIAL ALGEBRAIC EQUATIONS [*]

W. LUCHT, K. STREHMEL, and C. EICHLER-LIEBENOW

Institut für Numerische Mathematik, Fachbereich Mathematik und Informatik
Martin-Luther-Universität Halle–Wittenberg, Postfach, DE–06099 Halle, Germany
email: lucht@mail.mathematik.uni-halle.de

Abstract.

Linear partial differential algebraic equations (PDAEs) of the form $Au_t(t,x) + Bu_{xx}(t,x) + Cu(t,x) = f(t,x)$ are studied where at least one of the matrices A, $B \in \mathbb{R}^{n\times n}$ is singular. For these systems we introduce a uniform differential time index and a differential space index. We show that in contrast to problems with regular matrices A and B the initial conditions and/or boundary conditions for problems with singular matrices A and B have to fulfill certain consistency conditions. Furthermore, two numerical methods for solving PDAEs are considered. In two theorems it is shown that there is a strong dependence of the order of convergence on these indexes. We present examples for the calculation of the order of convergence and give results of numerical calculations for several aspects encountered in the numerical solution of PDAEs.

AMS subject classification: 35E99, 35M10, 65M06, 65M15, 65M20.

Key words: Differential algebraic equations, partial differential algebraic equations, coupled systems, indexes, consistency conditions, convergence of difference schemes, method of lines.

1 Introduction.

There are numerous mathematical models in science and engineering which lead to systems of equations of different type, e.g., the systems may consist of

- parabolic and elliptic differential equations or of
- parabolic, elliptic and ordinary differential equations or of
- elliptic and ordinary differential equations and algebraic equations.

There are many possibilities for the combination of equations of different character. In applications these systems must, in general, be supplemented by suitable initial conditions and boundary conditions. Under some assumptions such systems can be reduced by a Laplace transformation or by a Fourier analysis to a sequence of differential algebraic systems (DAEs). Hence the original system is also called a partial differential algebraic equation (PDAE).

We now give an example of such systems.

[*] Received March 1997. Revised January 1999. Communicated by Stig Skelboe.

EXAMPLE 1.1. The *modeling of a population dynamic* of n species in dependence of m ideally distributed food sources [13, 25]:

$$\begin{aligned}\frac{\partial u_j}{\partial t} &= D\,\Delta\, u_j \;+\; f_j(u,v), \qquad j=1,\dots,n,\\ \frac{\partial v_i}{\partial t} &= g_i(u,v), \qquad i=1,\dots,m\,,\end{aligned}$$

where $u=(u_1,\dots,u_n)^\top$ is the species density and $v=(v_1,\dots,v_m)^\top$ the food source density vector. The quantities $D>0$, f_j, g_i and suitable initial and boundary conditions are assumed to be given.

This system can also be interpreted as a model of a reaction-diffusion system. Here u_j may be seen as a concentration of a diffusing substance while the concentration v_i may be considered as a substance whose particles can not diffuse. n parabolic differential equations are coupled with m ordinary differential equations.

There are numerous applications in other scientific areas. Examples of PDAEs can be found in the field of Navier–Stokes equations [5, 14, 26], in chemical engineering [10, 13, 18, 20], in magneto-hydrodynamics [7, 21] and in the theory of elastic multibody systems [2, 22]. Some theoretical investigations for semilinear PDAEs are given in [15].

In this paper we investigate linear PDAEs with constant coefficients of the form[1]

$$A\,u_t(t,x)+B\,u_{xx}(t,x)+C\,u(t,x)=f(t,x), \quad (t,x)\in J\times\Omega \tag{1.1}$$

where $J=(0,t_e)$, $\Omega=(-l,l)$, $t_e>0$, $l>0$, and $A,B,C\in\mathbb{R}^{n\times n}$. Defining $\bar{J}=[0,t_e]$ for $t<\infty$, $\bar{J}=[0,\infty)$ for $t_e=\infty$ and $\bar{\Omega}=[-l,l]$, u and f are mappings $u,f:\bar{J}\times\bar{\Omega}\to\mathbb{R}^n$. We are interested in cases where at least one of the matrices A and B is singular. The two special cases $A=0$ or $B=0$ lead to ordinary differential equations or DAEs, which are not considered here. Therefore, in this paper we assume that none of the matrices A or B is the zero matrix.

Furthermore, for all $t\in\bar{J}$ we require boundary conditions (BCs, or boundary values, BVs) for the components u_j of u for all $j\in\mathfrak{M}_{BC}\subseteq\{1,2,\dots,n\}$ ($\mathfrak{M}_{BC}$ is specified in Section 2.1; see also the following example 1.2) for simplicity of the form

$$\mathrm{R_B}\ u_j(t,x):=u_j(t,\pm l)=0. \tag{1.2}$$

Beside these, for the initial conditions (ICs, or initial values, IVs) of the form

$$u(0,x)=g(x) \quad \text{for} \quad x\in\bar{\Omega} \tag{1.3}$$

we assume that the components g_j of g for all $j\in\mathfrak{M}_{IC}\subseteq\{1,2,\dots,n\}$ can be prescribed arbitrarily ($\mathfrak{M}_{IC}$ is defined in Section 2.2). For the components

[1] It may be that certain equations of the system are defined even on sets $\bar{J}\times\Omega$, $J\times\bar{\Omega}$ or $\bar{J}\times\bar{\Omega}$; see e.g., Examples 1.2 and 2.1 below.

u_k, $k \notin \mathfrak{M}_{BC}$, BCs of Dirichlet type are pertinent. In general, the BCs for u_k, $k \notin \mathfrak{M}_{BC}$, and the ICs for u_i, $i \notin \mathfrak{M}_{IC}$, must be determined with the help of the PDAE (1.1), the BCs for u_j, $j \in \mathfrak{M}_{BC}$, and the ICs for u_m, $m \in \mathfrak{M}_{IC}$ (of course, there may be cases which do not need all of these). Furthermore, we require the compatibility conditions (component wise)

$$\mathrm{R_B}\; g_i(x) = \mathrm{R_B}\, u_i(0,x), \qquad i \in \mathfrak{M}_{BC} \cap \mathfrak{M}_{IC}, \tag{1.4}$$

between the ICs and the BCs.

In contrast to problems with regular matrices A and B (e.g., parabolic problems) the IC (1.3) and/or the BC (1.2) for problems with singular matrices A and/or B have to fulfill certain supplementary conditions (so-called consistency conditions, see below). The following example illustrates this.

Example 1.2. Let $u = (u_1, u_2, u_3)^\top$ and the PDAE

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{=\,A} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix} + \underbrace{\begin{pmatrix} -1 & 0 & 0 \\ 0 & -b & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{=\,B} \begin{pmatrix} u_{1xx} \\ u_{2xx} \\ u_{3xx} \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & c_1 & 0 \\ 0 & 0 & c_2 \\ 0 & c_3 & 0 \end{pmatrix}}_{=\,C} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

$$u_j(t,-1) = u_j(t,1) = 0, \quad j \in \mathfrak{M}_{BC}, \qquad u_i(0,x) = g_i(x)\, \sin(\pi x), \quad i \in \mathfrak{M}_{IC}$$

be given for $t \in J$, $x \in \Omega = (-1,1)$ with $a, b > 0$, $c_i \neq 0$ and sufficiently smooth functions $f_i(t,x)$, $i = 1,2,3$. The term $\sin(\pi x)$ in the IC for u_i guarantees the compatibility (1.4) with the BCs for $i \in \mathfrak{M}_{BC}$. In the index set $\mathfrak{M}_{BC}$ we collect those indices of the components of u for which an arbitrary BC can be assigned ($\mathfrak{M}_{IC}$ is characterized in an analogous manner). Since we immediately get for the solution components u_2, u_3

$$u_2(t,x) = \frac{1}{c_3} f_3(t,x), \quad u_3(t,x) = \frac{1}{c_2}\Big(f_2(t,x) - \frac{a}{c_3} f_{3t}(t,x) + \frac{b}{c_3} f_{3xx}(t,x)\Big),$$

there is no free parameter left so that the IVs and BVs for the components u_2, u_3 are directly given by the IVs and BVs of the right hand functions and their derivatives. Therefore, we can not prescribe ICs and BCs for the components u_2, u_3, and hence it is $\mathfrak{M}_{BC} = \mathfrak{M}_{IC} = \{1\}$. The solution component $u_1(t,x)$ is the solution of the parabolic initial boundary value problem (IBVP)

$$\begin{aligned} u_{1t} &= u_{1xx} + f_1 - \frac{c_1}{c_3} f_3\,, \\ u_1(t,-1) = u_1(t,1) &= 0, \qquad u_1(0,x) = g_1(x) \sin(\pi x)\,, \end{aligned}$$

where we can assign $g_1(x)$ (the choice of the Dirichlet BC is also a free one).

This example clarifies that problems of the form (1.1) with singular matrices essentially differ from such with regular matrices A, B. It is analogous to the statement that ODEs are not DAEs [19].

This paper is organized as follows. In the next section we introduce the basic notions of the differential spatial index by means of a Laplace transform (2.1) and the uniform differential time index by a Fourier transform (2.2), respectively. These indexes are used to characterize the PDAE in several contexts. Of importance for the treatment of PDAEs is the fact that, as illustrated in the foregoing example, in general, any ICs and BCs cannot be prescribed for all components of the solution vector u. To overcome this difficulty we define in Section 2 the fundamental index sets $\mathfrak{M}_{IC}$ and $\mathfrak{M}_{BC}$. If we know these two sets, then we know for which components of the solution vector u of the PDAE ICs and BCs can be assigned.

In the third section we investigate two numerical schemes for solving IBVPs of linear PDAEs by means of the method of lines (MOL). The full discretization of the PDAE is considered in Section 3.2. In Sections 3.2.1 and 3.2.2 the BTCS scheme is investigated in detail. Of special interest are the influence of the differential spatial and the uniform differential time index on the convergence of the numerical solution to the exact solution. We derive order relations for the convergence by means of an algebraic characterization of the indexes. The Crank–Nicolson scheme is the subject of Section 3.2.3.

Numerical examples using the BTCS scheme are presented in Section 4. First we determine numerically order relations for convergence. Hereby the theoretical results of Sections 3.2.1 and 3.2.2 are confirmed. In a second example we study numerically the influence of inconsistent BVs. A third example deals with the problem of an index jump. Although, the PDAE does not have an index jump the MOL–DAE may have such a jump. This may be origin of an instability (as discussed in Section 4) and of large errors in the numerical solution.

2 Indexes of the PDAE.

As in the case of DAEs which can be characterized conveniently by a differential index (cf. [8, 11]) it is useful to introduce indexes also for PDAEs. These describe special properties of the systems (with respect to both the analytical solution and the numerical treatment). For PDAEs it is suitable to distinguish between a differential spatial index and a uniform differential time index which are defined in this section by means of the Laplace transform and a Fourier analysis, respectively.

We note that indexes for PDAEs have been defined already by other authors; see [5, 6, 18]. Our index definitions given below are partly the same as the corresponding ones presented in the references. However, in detail ours differ from those. The reason for using slightly different index definitions is that we can prove with these some convergence theorems when solving PDAEs by certain discretization methods (see Section 3).

In the sequel the following assumptions are used:

(I) The IBVP (1.1)–(1.4) has one and only one solution where a function u is a solution of the problem, if it is sufficiently smooth, uniquely determined by its IVs and BVs and if it solves the PDAE pointwise.

(II) Each component of the vectors u, u_t and f satisfy a growth condition of the form

$$|y(t,x)| \leq M\, e^{\alpha\, t}\,, \quad \alpha \geq 0,\ t \geq 0$$

(M and α are independent of t and x).

(III) The matrix pencil $(B, \xi A + C)$, $\mathrm{Re}(\xi) > \alpha$, is regular.

(IV) The matrix pencil $(A, \mu_k B + C)$ is regular for all k. μ_k is an eigenvalue of the operator $\frac{\partial^2}{\partial x^2}$ with prescribed BCs (1.2).

(V) The right hand vector $f(t,x)$ and the initial vector $g(x)$ are sufficiently smooth.

2.1 Laplace transform and the spatial index.

Let now $t_e = \infty$ and $y : [0,\infty) \to \mathbb{R}$ be continuous with (growth condition)

$$|y(t)| \leq M\, e^{\alpha\, t}\,, \quad t \in [0,\infty), \quad 0 < M < \infty, \quad \alpha \geq 0.$$

Define (as usual) the Laplace transform of $y(t)$ by

$$y_\xi := \int_0^\infty e^{-t\,\xi}\, y(t)\, dt\,, \quad \mathrm{Re}(\xi) > \alpha.$$

Assumption (II) implies that equation (1.1) can be transformed into

$$B\, u_\xi''(x) + (\xi A + C)\ u_\xi(x) = f_\xi(x) + A\, g(x)\,, \quad \mathrm{Re}(\xi) > \alpha\,, \tag{2.1}$$

where $g(x)$ is the initial vector (1.3). If B is a singular matrix, then (2.1) is a DAE depending on the parameter ξ. Therefore, BCs can be prescribed in general only for certain components of u_ξ. To characterize these we introduce $\mathfrak{M}_{BC}^{(\xi)} \subseteq \{1, \ldots, n\}$ as the set of indices of components of u_ξ for which BCs can be prescribed arbitrarily.

In order to define a spatial index we need the Kronecker normal form of the DAE (2.1). Assumption (III) guarantees that there are nonsingular matrices $P_{L,\xi}, Q_{L,\xi} \in \mathbb{C}^{n\times n}$ such that

$$P_{L,\xi}\, B\, Q_{L,\xi} = \begin{pmatrix} I_{m_1} & 0 \\ 0 & N_{L,\xi} \end{pmatrix}, \qquad P_{L,\xi}\ (\xi A + C)\ Q_{L,\xi}\ = \begin{pmatrix} R_{L,\xi} & 0 \\ 0 & I_{m_2} \end{pmatrix} \tag{2.2}$$

where $R_{L,\xi} \in \mathbb{C}^{m_1\times m_1}$, and $N_{L,\xi} \in \mathbb{R}^{m_2\times m_2}$ is a nilpotent Jordan chain matrix with $m_1 + m_2 = n$. I_k is the unit matrix of order k. The Riesz index (or nilpotency) of $N_{L,\xi}$ is denoted by $\nu_{L,\xi}$ (i.e. $N_{L,\xi}^{\nu_{L,\xi}} = 0$, $N_{L,\xi}^{\nu_{L,\xi}-1} \neq 0$). With these relations (2.1) can be transformed into the decoupled system of equations

$$v_\xi''(x) + R_{L,\xi}\, v_\xi(x) = r_{\xi,1}(x), \tag{2.3}$$

$$N_{L,\xi}\, w_\xi''(x) + w_\xi(x) = r_{\xi,2}(x)\ . \tag{2.4}$$

Here, we introduced the definitions $(v_\xi^\top(x), w_\xi^\top(x))^\top := Q_{L,\xi}^{-1} u_\xi(x)$ and $(r_{\xi,1}^\top(x), r_{\xi,2}^\top(x))^\top := P_{L,\xi}(f_\xi(x) + Ag(x))$. (2.3) shows that we have to solve a system of m_1 differential equations of second order. In general, we have to require for each component of v_ξ a BC. Under assumption (I) this BVP has an unique solution. Thus, if we transform $(v_\xi^\top, w_\xi^\top)^\top$ back to u_ξ we see that $\mathfrak{M}_{BC}^{(\xi)}$ is a set with m_1 elements. Equation (2.4) is equivalent to the algebraic equation

(2.5)
$$w_\xi(x) = r_{\xi,2}(x) - N_{L,\xi} r''_{\xi,2}(x) + \cdots (-1)^{\nu_{L,\xi}-1} N_{L,\xi}^{\nu_{L,\xi}-1} r_{\xi,2}^{(2\nu_{L,\xi}-2)}(x).$$

From this equation we obtain formally for each component of the vector $w_\xi \in \mathbb{C}^{m_2}$ an expression in terms of the given vector $r_{\xi,2}$ and its derivatives with respect to x up to the order $2\nu_{L,\xi} - 2$. This means that any BCs cannot be prescribed for the components of w_ξ.

The matrices $P_{L,\xi}, Q_{L,\xi}, R_{L,\xi}, N_{L,\xi}$ (and $m_1, m_2, \nu_{L,\xi}$) depend on the parameter ξ. For which components of $u_\xi(x)$ (and finally of $u(t,x)$) we can prescribe a BC depends on the matrix $Q_{L,\xi}$. In this paper we will assume that there is a real number $\alpha^* \geq \alpha$ such that the index set $\mathfrak{M}_{BC}^{(\xi)}$ is independent of the Laplace parameter ξ, provided $\mathrm{Re}(\xi) \geq \alpha^*$. Taking into account that the inverse Laplace transformation of u_ξ must be performed we conclude from this assumption that the index set $\mathfrak{M}_{BC}$ (which is needed after all) is then simply given by $\mathfrak{M}_{BC} = \mathfrak{M}_{BC}^{(\xi)}$.

From (2.5) we get a hint for a definition of a space index analogous to the definition of the differential index for DAEs. One differentiation of this equation with respect to x yields

(2.6)
$$w'_\xi(x) = r'_{\xi,2}(x) - N_{L,\xi} r'''_{\xi,2}(x) + \cdots (-1)^{\nu_{L,\xi}-1} N_{L,\xi}^{\nu_{L,\xi}-1} r_{\xi,2}^{(2\nu_{L,\xi}-1)}(x) \ .$$

Thus, in order to get an explicit differential equation for $w_\xi(x)$ one needs $2\nu_{L,\xi}-1$ differentiations with respect to x. The equivalence of the transformed PDAE (2.1) and the decoupled system (2.3), (2.4) with the differential equation (2.6) for $w_\xi(x)$ suggests to define the differential spatial index $\nu_{d,x}$ by

Definition 2.1. *Let $\alpha^* \in \mathbb{R}^+$ be a number with $\alpha^* \geq \alpha$, such that for all $\xi \in \mathbb{C}$ with* $\mathrm{Re}(\xi) \geq \alpha^*$

1. *the matrix pencil $(B, \xi A + C)$ is regular,*
2. *$\mathfrak{M}_{BC}^{(\xi)}$ is independent of ξ, i.e. $\mathfrak{M}_{BC}^{(\xi)} = \mathfrak{M}_{BC}$,*
3. *the nilpotency of $N_{L,\xi}$ is $\nu_L \geq 1$.*

Then $\nu_{d,x} := 2\nu_L - 1$ is called the differential spatial index of the PDAE (1.1). *If $\nu_L = 0$ then the differential spatial index of PDAE is defined to be zero.*

The index $\nu_{d,x}$ shows directly which differentiation properties with respect to the variable x the function $r_{\xi,2}(x)$ (and hence $f(t,x)$, $g(x)$) must have at least in order to convert the Laplace transformed PDAE (2.1) into an explicit system of differential equations (2.3), (2.6).

Remark 2.1. Since n is small in many applications (such as $n = 2, 3$ or 4) the nilpotency ν_L of N_L can be determined easily. In the case that N_L consists of only one block, it is $\nu_L = m_2$.

Sometimes the notion of consistent BVs is used:

Definition 2.2. *Assume* $\mathfrak{M}_{BC}^{(\xi)} = \mathfrak{M}_{BC}$, $\mathrm{Re}(\xi) \geq \alpha^*$. *Then a given BV of* u_j *for* $j \notin \mathfrak{M}_{BC}$ *is called consistent if its Laplace transform satisfies* $u_{\xi j}(\pm l) = \big(Q_{L,\xi}(v_\xi^\top(\pm l), w_\xi^\top(\pm l))^\top\big)_j$.

In the following example we calculate consistent BVs.

Example 2.1. Let $n = 2$, $u = (u_1, u_2)^\top$ and

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u_t + \begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix} u_{xx} + \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} u \;=\; f, \qquad \beta < 0. \tag{2.7}$$

This PDAE has differential spatial index $\nu_{d,x} = 1$. Because A is regular it is $\mathfrak{M}_{IC} = \{1, 2\}$ (cf. Section 2.2). We use

$$P_{L,\xi} = \begin{pmatrix} \frac{1}{\xi} & \frac{1}{\xi-\beta} \\ 0 & \frac{1}{\xi-\beta} \end{pmatrix} \text{ and } Q_{L,\xi} = \begin{pmatrix} -\frac{\xi^2}{\xi-\beta} & -1 \\ \frac{\beta\,\xi}{\xi-\beta} & 1 \end{pmatrix}$$

with $\mathrm{Re}(\xi) > 0$ to perform the Kronecker transformation of the Laplace transformed PDAE (2.1) of this example. The new equation is

$$\begin{pmatrix} v_\xi'' \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{\xi^2}{\beta-\xi} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_\xi \\ w_\xi \end{pmatrix} = P_{L,\xi}(f_\xi + A\,g)$$

with

$$v_\xi = -\frac{1}{\xi}\,(u_{1\xi} + u_{2\xi})\,, \qquad w_\xi = \frac{1}{\xi-\beta}\,(\beta\, u_{1\xi} + \xi\, u_{2\xi})\,.$$

From these formulas it is seen immediately that $\mathfrak{M}_{BC}^{(\xi)}$ is defined not uniquely. We can choose for $\mathrm{Re}(\xi) \geq \alpha^*$ either $\mathfrak{M}_{BC}^{(\xi)} = \{1\}$ or $\mathfrak{M}_{BC}^{(\xi)} = \{2\}$. Using $\mathfrak{M}_{BC}^{(\xi)} = \{1\}$ we infer from the original equation (2.7) that the BVs for u_2 can be calculated by solving the ODE

$$\frac{d}{dt} u_2(t, \pm l) = f_2(t, \pm l) - \beta u_1(t, \pm l) \tag{2.8}$$

with given IVs $u_2(0, \pm l)$. After solving this simple ordinary initial value problem we have consistent BV for u_2.

2.2 Fourier transform and the time index.

Beside the spatial index defined in the previous section a time index for PDAEs can be introduced too. We multiply the PDAE (1.1) with a scalar function

$\phi_k(x)$ and integrate with respect to x on the interval $[-l, l]$. The functions $\phi_k(x)$, $k = 1, 2, \ldots$, are chosen as orthonormal eigenfunctions of the operator $\frac{\partial^2}{\partial x^2}$ with eigenvalue μ_k. $\phi_k(x)$ fulfills the same BC as $u_j(t,x)$, $j \in \mathfrak{M}_{BC}$, i.e. the homogeneous condition (1.2). Defining the kth Fourier coefficient of a vector valued function $\chi(t,x)$ with respect to ϕ_k, as usual, by

$$\hat{\chi}_k(t) = \frac{1}{l}\int_{-l}^{l} \chi(t,x)\phi_k(x)\,dx$$

we get

$$A\hat{u}_k'(t) + (\mu_k\, B + C)\,\hat{u}_k(t) = \hat{f}_k(t) + B\rho_k(t) \;=:\; \bar{f}_k(t) \tag{2.9}$$

with $\rho_k(t) = (\rho_{k1}(t), \ldots, \rho_{kn}(t))^\top$ and

$$\rho_{ki}(t) = 0 \text{ for } i \in \mathfrak{M}_{BC}, \quad \rho_{kj}(t) = \tfrac{1}{l}\Big[\phi_k'(x)u_j(t,x) - \phi_k(x)u_{x,j}(t,x)\Big]_{x=-l}^{x=l}$$

for $j \notin \mathfrak{M}_{BC}$, which results from partial integration of the term $\int_{-l}^{l} u_{xx}(t,x)\phi_k(x)\,dx$.

If the matrix A is singular (as it may be here), (2.9) is a DAE depending on the parameter μ_k which can be solved uniquely with suitable ICs under the assumptions (IV) and (V). Analogous to the case of the Laplace transform in Section 2.1 assumption (IV) implies that there exist regular matrices $P_{F,k}, Q_{F,k}$ such that

(2.10)
$$P_{F,k}\, A\, Q_{F,k} = \begin{pmatrix} I_{n_1} & 0 \\ 0 & N_{F,k} \end{pmatrix}, \qquad P_{F,k}\; (\mu_k B + C)\; Q_{F,k} \;=\; \begin{pmatrix} R_{F,k} & 0 \\ 0 & I_{n_2} \end{pmatrix}$$

with $R_{F,k} \in \mathbb{R}^{n_1 \times n_1}$. $N_{F,k} \in \mathbb{R}^{n_2 \times n_2}$ is again a nilpotent Jordan chain matrix with Riesz index $\nu_{F,k}$. Of course $n_1 + n_2 = n$. These relations imply that (2.9) is equivalent to the decoupled system of equations

$$y_k'(t) + R_{F,k}\, y_k(t) = s_{k,1}(t), \tag{2.11}$$
$$N_{F,k}\, z_k'(t) + z_k(t) = s_{k,2}(t) \tag{2.12}$$

where $(y_k^\top(t), z_k^\top(t))^\top := Q_{F,k}^{-1}\hat{u}_k(t)$, $(s_{k,1}^\top(t), s_{k,2}^\top(t))^\top := P_{F,k}\bar{f}_k(t)$. The solution of equation (2.12) can be written as

$$z_k(t) = \sum_{i=0}^{\nu_{F,k}-1} (-N_{F,k})^i\, s_{k,2}^{(i)}(t)\ . \tag{2.13}$$

This representation shows that, in general, we cannot prescribe any IV $z_k(0)$ because this equation implies that the initial values of z_k are given by the function on the right. Furthermore, we also see that the right hand side function f of the original PDAE has for indexes $\nu_{F,k} \geq 2$ to fulfill strong requirements concerning its differentiation properties.

Denote by $\mathfrak{M}_{IC}^{(k)} \subseteq \{1, \dots, n\}$ the set of indices of components of $\hat{u}_k$ whose IVs can be prescribed arbitrarily. Then we always assume in the context of a Fourier analysis of u that $\mathfrak{M}_{IC}^{(k)}$ is independent of $k \in \mathbb{N}_+$, i.e. $\mathfrak{M}_{IC}^{(k)} = \mathfrak{M}_{IC}$. We need this assumption because there are PDAEs which do not have this property (see Example 2.5).

After these preliminaries we can define the following index.

Definition 2.3. *Assume for* $k = 1, 2, \dots$ *that*

1. *the matrix pencil* $(A, \mu_k B + C)$ *is regular,*
2. $\mathfrak{M}_{IC}^{(k)}$ *is independent of* k*, i.e.* $\mathfrak{M}_{IC}^{(k)} = \mathfrak{M}_{IC}$,
3. *the nilpotency of* $N_{F,k}$ *is* $\nu_{F,k} = \nu_F$.

Then the PDAE (1.1) *is said to have uniform differential time index* $\nu_{d,t} = \nu_F$.

If the matrices A, B, C have a special structure it can happen that assumption (2) of this definition is not satisfied. Then it may be that the pencil $(A, \mu_k B + C)$ does not have the same index for all k and/or that $N_{F,k}$ is not the same for all k $(N_{F,k} \neq N_F)$. This can only occur for a finite number of different k. The first case is called an *index jump* already mentioned in [6, 18] and discussed shortly below. These two cases are the subject of the Examples 2.4 and 2.5, respectively. Problems like these differ qualitatively from those with an uniform time index such that it is advantageous to assign them to an other class of PDAEs.

An index quite similar to $\nu_{d,t}$ has been defined in [6, 18]. However, the authors of these works do not rule out some difficult special cases as we do here. Especially, their index definitions do not distinguish between a PDAE with or without an index jump. We prefer the above definition because we then can get some further results (convergence theorems in the context of special numerical methods; see Section 3).

Example 2.2. For the PDAE of Example 1.2 it is easily shown that $\nu_L = 2$, i.e. $\nu_{d,x} = 3$, $\nu_{d,t} = 2$.

The analogue of Definition 2.2 is given by

Definition 2.4. *Assume* $\mathfrak{M}_{IC}^{(k)} = \mathfrak{M}_{IC}$, $k = 1, 2, \dots$. *Then a given IV of* u_j *for* $j \notin \mathfrak{M}_{IC}$ *is called consistent if its finite Fourier transform satisfies* $\hat{u}_{kj}(0) = \big(Q_{F,k}(y_k^\top(0), z_k^\top(0))^\top\big)_j$. The following example illustrates the determination of the sets $\mathfrak{M}_{BC}$ and $\mathfrak{M}_{IC}$.

Example 2.3. Let $A = \left(\begin{smallmatrix}1 & 1\\ 0 & 0\end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix}-1 & 0\\ -1 & 0\end{smallmatrix}\right)$, $C = \left(\begin{smallmatrix}c_1 & c_2\\ -1 & 0\end{smallmatrix}\right)$, $c_1, c_2 > 0$. Then we get $\nu_{d,t} = 1$ $(n_1 = n_2 = 1)$, $\nu_{d,x} = 1$ $(m_1 = m_2 = 1)$ from the Kronecker transformation using the matrices

$$P_{F,k} = \begin{pmatrix} \frac{1}{c_2} & \frac{c_1 - c_2 - \mu_k}{c_2(\mu_k - 1)} \\ 0 & \frac{1}{\mu_k - 1} \end{pmatrix}, \qquad Q_{F,k} = \begin{pmatrix} 0 & -1 \\ c_2 & 1 \end{pmatrix},$$

$$P_{L,\xi} = \begin{pmatrix} 0 & \frac{1}{\xi + c_2} \\ \frac{1}{\xi + c_2} & -\frac{1}{\xi + c_2} \end{pmatrix}, \qquad Q_{L,\xi} = \begin{pmatrix} -(\xi + c_2) & 0 \\ \xi + c_1 - 1 & 1 \end{pmatrix}.$$

This implies the transformations

(2.14)
$$\begin{pmatrix}\hat{u}_{k1}\\ \hat{u}_{k2}\end{pmatrix} = \begin{pmatrix}-z_k\\ c_2 y_k + z_k\end{pmatrix}, \qquad \begin{pmatrix}u_{\xi 1}\\ u_{\xi 2}\end{pmatrix} = \begin{pmatrix}-(\xi + c_2)v_\xi\\ (\xi + c_1 - 1)v_\xi + w_\xi\end{pmatrix}.$$

We know from Section 2 that, in general, the initial value for z_k and the boundary values for w_ξ can not be prescribed arbitrarily. Equation (2.14) fixes the component $u_1(t,x)$ to be that one for which we can not prescribe an IC, i.e. $\mathfrak{M}_{IC} = \mathfrak{M}_{IC}^{(k)} = \{2\}$. On the other hand since a BC for v_ξ can be assigned, we conclude from (2.14) that $\mathfrak{M}_{BC} = \mathfrak{M}_{BC}^{(\xi)} = \{1\}$. Thus a BC for $u_{\xi 2}$ (or $u_2(t,x)$) can not be prescribed because the BC for $u_{\xi 2}$ must be computed in the manner shown in (2.14). On the other hand for $\mathrm{Re}(\xi) > 1 - c_1$ it is also possible to set $\mathfrak{M}_{BC} = \mathfrak{M}_{BC}^{(\xi)} = \{2\}$. Then $u_{\xi 1} = -\frac{\xi + c_2}{\xi + c_1 - 1}(u_{\xi 2} - w_\xi)$. The set $\mathfrak{M}_{BC}$ is not determined uniquely. In order to illustrate the case of an index jump and a change of the sets $\mathfrak{M}_{IC}^{(k)}$ with respect to k we consider the following examples. They also motivate the term *uniform* in Definition 2.3.

Example 2.4. Let $n = 2$, A be singular, B be regular and homogeneous Dirichlet BCs $u(t,-l) = u(t,l) = 0$ for $t \geq 0$ be given. For B regular it holds that $\mathfrak{M}_{BC} = \{1,2\}$ and $\rho_k(t) = 0$ in (2.9). Let

$$A = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \qquad B = \begin{pmatrix}b_1 & b_2\\ b_3 & b_4\end{pmatrix}, \qquad C = \begin{pmatrix}c_1 & c_2\\ c_3 & c_4\end{pmatrix}.$$

The matrix pencil $(A, \mu_k B + C)$ is regular if either $\{b_4\,\mu_k = -c_4$ and $(b_2\,\mu_k + c_2)(b_3\,\mu_k + c_3) \neq 0\}$ or $\{b_4\,\mu_k \neq -c_4\}$. Let ν be such that the matrix $\nu A + (\mu_k B + C)$ is regular. Then the Riesz index of the matrix pencil is 1 for $b_4\,\mu_k \neq -c_4$ and 2 for $b_4\,\mu_k = -c_4$ (cf. [3, 9]). We get

- uniform time index 1 if $b_4\,\mu_k \neq -c_4$ for all $k = 1, 2, \ldots$,
- uniform time index 2 if $b_4 = c_4 = 0$,
- index jump if $\mu_k = -c_4/b_4$ for one k.

Let for simplicity the matrix $\mu_k B + C$ be regular for all $k > 0$, i.e. $D_k := \det(\mu_k B + C) \neq 0$. Using the Jordan normal form of $(\mu_k B + C)^{-1} A$ we obtain the following two cases:

1. For $b_4\,\mu_k \neq -c_4$, i.e. if the pencil $(A, \mu_k B + C)$ has the Riesz index 1, we get from the solution $\hat{u}_k = (\hat{u}_{k1}, \hat{u}_{k2})^\top$ for $t = 0$ with $\hat{f}_k = (\hat{f}_{k1}, \hat{f}_{k2})^\top$

$$\hat{u}_k(0) = \begin{pmatrix}\hat{u}_{k1}(0)\\ \hat{u}_{k2}(0)\end{pmatrix} = \hat{u}_{k1}(0)\begin{pmatrix}1\\ -\frac{b_3\,\mu_k + c_3}{b_4\,\mu_k + c_4}\end{pmatrix} + \begin{pmatrix}0\\ \frac{1}{b_4\,\mu_k + c_4}\end{pmatrix}\hat{f}_{k2}(0).$$

That means we have in this case a consistency condition between the initial values and the right hand side function $\hat{f}_{k2}(0) = (b_3\,\mu_k + c_3)\hat{u}_{k1}(0) + (b_4\,\mu_k + c_4)\hat{u}_{k2}(0)$. If we prescribe $\hat{u}_{k1}(0)$, then $\hat{u}_{k2}(0)$ is uniquely determined by $\hat{u}_{k1}(0)$ and $\hat{f}_{k2}(0)$ and cannot be prescribed arbitrarily.

2. From the solution of (2.9) for $b_4\,\mu_k = -c_4$ (for Riesz index 2) we get for the initial values the consistency conditions

$$\begin{aligned}\hat{u}_{k1}(0) &= \frac{\hat{f}_{k2}(0)}{b_3\,\mu_k + c_3},\\ \hat{u}_{k2}(0) &= \frac{(b_3\,\mu_k + c_3)\hat{f}_{k1}(0) - (b_1\,\mu_k + c_1)\hat{f}_{k2}(0) - \hat{f}'_{k2}(0)}{(b_2\,\mu_k + c_2)(b_3\,\mu_k + c_3)}\end{aligned}$$

which means that, in general, we can prescribe neither $\hat{u}_{k1}(0)$ nor $\hat{u}_{k2}(0)$.

In this example we can interpret the solution $\hat{u}_k(t)$ of (2.9) formally as the kth Fourier coefficient of $u(t,x)$, i.e. $u(t,x) = \sum_{k=1}^{\infty} \hat{u}_k(t)\phi_k(x)$. For uniform differential time index 1 of the PDAE (1.1) all DAEs (2.9), $k = 1, 2, \ldots$, have differential index 1. That means that for all these initial value problems we can assign $\hat{u}_{k1}(0)$, and we can calculate a consistent $\hat{u}_{k2}(0)$ uniquely from $\hat{u}_{k1}(0)$ and the given right hand functions. Prescribing initial values (1.3) for the first component of $g(x)$ (i.e. $\mathfrak{M}_{IC} = \{1\}$), we get consistency between the initial values and the right hand side functions. Analogously we cannot prescribe any initial condition in the case of an uniform time index 2 (i.e. $\mathfrak{M}_{IC} = \emptyset$) because it is completely determined by the right hand side.

In the second case, the condition $\mathfrak{M}_{IC}^{(k)} = \mathfrak{M}_{IC}$ does not hold. Hence, from the Fourier analysis we cannot obtain consistent IVs (1.3) for u in the case of an index jump. In the example it is $\mu_j = -\frac{c_4}{b_4}$ for one $j > 0$, and the jth Fourier coefficient of $u_1(0,x)$ is given by the right hand side function. All other Fourier coefficients $\hat{u}_{k1}(0)$, $k \neq j$, are free parameters.

Example 2.5. Let $n = 4$ and the matrices A, B, C be given by

$$A = \begin{pmatrix} 0&1&0&0\\ 0&0&0&0\\ 0&0&1&0\\ 0&0&0&0 \end{pmatrix},\ B = \begin{pmatrix} -1&0&0&0\\ 0&-1&0&0\\ 0&0&-1&0\\ 0&0&0&-1 \end{pmatrix},\ C = \begin{pmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&1\\ 0&0&1&c \end{pmatrix}.$$

The matrix pencil $(A, \mu_k B + C)$ has Riesz index 2 for all $k \in \mathbb{N}_+$. Let $c = \mu_j$ for a fixed $j \in \mathbb{N}_+$. The Kronecker transformations (2.10) of the pencils $(A, \mu_j\,B + C)$ and $(A, \mu_k\,B + C)$ lead to

$$P_{F,j}A \underbrace{\begin{pmatrix} 1&0&0&0\\ 0&1-c&0&0\\ 0&0&0&1\\ 0&0&1&0 \end{pmatrix}}_{Q_{F,j}} = \underbrace{\begin{pmatrix} 0&1&0&0\\ 0&0&0&0\\ 0&0&0&1\\ 0&0&0&0 \end{pmatrix}}_{N_{F,j}}, \qquad P_{F,k}A \underbrace{\begin{pmatrix} 0&1&1&1\\ 0&0&1-\mu_k&0\\ \mu_k-c&0&0&0\\ 1&0&1&1 \end{pmatrix}}_{Q_{F,k}} = \left(\begin{array}{c|c} 1 & \begin{matrix}0&0&0\end{matrix}\\ \hline \begin{matrix}0\\0\\0\end{matrix} & \underbrace{\begin{matrix}0&1&0\\0&0&0\\0&0&0\end{matrix}}_{N_{F,k}} \end{array}\right),$$

i.e. the matrix $N_{F,j} \in \mathbb{R}^{4\times4}$ is different from the matrices $N_{F,k} \in \mathbb{R}^{3\times3}$, $k \neq j$. It follows from (2.13) that $u_j(0)$ is determined by the right hand side function but we get from $u_k(t) = Q_{F,k}(y_k^\top, z_k^\top)^\top$ that for all $k \neq j$ any value of the third component of $u_k(0)$ can be chosen arbitrarily, i.e. $\mathfrak{M}_{IC}^{(k)} = \{3\} \neq \mathfrak{M}_{IC}^{(j)} = \emptyset$ $\forall k \neq j$. Thus by means of a Fourier analysis we cannot define $\mathfrak{M}_{IC}$ for the original PDAE defined by the matrices A, B, C considered in this example. For details and further investigations (including examples) see [16].

3 Two discretization methods,

In this section we consider two numerical discretization schemes for solving PDAEs of the type (1.1) with BVs (1.2) and IVs (1.3) under the general assumptions (I)–(V) of Section 2 and under the further condition that $\nu_{d,x}$ and $\nu_{d,t}$ are defined. The discretization methods used here can be derived along the ideas of the well-known method of lines (MOL). We suppose that the BVs for the components u_j, $j \notin \mathfrak{M}_{BC}$, which may be needed in the discretization scheme can be calculated from the PDAE (see Example 2.1) and possibly from the prescribed BVs and IVs. More details of the investigations in this section are given in [17].

3.1 Space discretization.

For the space discretization of the considered PDAE-problem we approximate u_{xx} by its second order difference quotient $u_{xx}(t,x_k) \approx (u(t,x_{k+1}) - 2u(t,x_k) + u(t,x_{k-1}))/h^2$, $k = 1,\ldots,N$, on the uniform grid $\Omega_h = \{x_k : x_k = -l + kh,\ k = 1,\ldots,N,\ h = \frac{2l}{N+1}\}$. Then we obtain the semi-discretized equation (using the Kronecker product)

$$(I_N \otimes A)\, U'(t) + \left(\frac{1}{h^2} P \otimes B + I_N \otimes C\right) U(t) = F(t) - r(t) \tag{3.1}$$

where $U(t) = (u_1^\top(t),\ldots,u_N^\top(t))^\top \in \mathbb{R}^{nN}$ with $u_k(t) \approx u(t,x_k)$, $k = 1,\ldots,N$, and the $nN-$dimensional vectors $r(t)$, $F(t)$ are given by

$$r(t) = \left(\tfrac{1}{h^2} I_N \otimes B\right)(u^\top(t,-l), 0, \ldots, 0, u^\top(t,l))^\top, \quad F(t) = (f_1^\top(t),\ldots,f_N^\top(t))^\top,$$

$f_k(t) = f(t,x_k)$. I_N is the $(N \times N)$-unit matrix, and the matrix P is defined by

$$P = \begin{pmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & & \ddots & \\ & & 1 & -2 \end{pmatrix} \in \mathbb{R}^{N\times N}\,. \tag{3.2}$$

Equation (3.1) is (according to (1.3)) completed by the consistent initial vector

$$U(0) = \left(\tilde{g}^\top(x_1),\ldots,\tilde{g}^\top(x_N)\right)^\top \in \mathbb{R}^{nN} \tag{3.3}$$

where the difference $\tilde{g} - g$ goes (component wise) to zero for $h \to 0$. If A is singular, then (3.1) is for fixed h a DAE which may be handled by standard methods [12]. Its differential index is identical (at least for small h) with the uniform differential time index $\nu_{d,t}$. This is easily proved by the expansion

$$U_\Phi(t) = \sum_{k=1}^{N} \Phi_k \otimes w_k(t), \quad \Phi_k \in \mathbb{R}^N, \quad w_k \in \mathbb{R}^n, \tag{3.4}$$

where Φ_k is an eigenvector of the matrix $\frac{1}{h^2}P$ ([24]), i.e.,

$$\tfrac{1}{h^2} P\, \Phi_k = \lambda_k \Phi_k, \qquad k = 1,\ldots,N, \tag{3.5}$$

and U_Φ denotes the projection of $U \in \mathbb{R}^{nN}$ in the space of vectors corresponding for $N \to \infty$ to $n-$dimensional vector functions with zero BCs. The eigenvalues of $\frac{1}{h^2}P$ are given by $\lambda_k = -\frac{4}{h^2}\sin^2\left(\frac{k\pi}{2(N+1)}\right)$ with $h = \frac{2l}{N+1}$. These eigenvalues converge for $k \in \{1,\dots,N\}$ in the limit $h \to 0$ towards the eigenvalues $\mu_k = -\left(\frac{k\pi}{2l}\right)^2$ with corresponding eigenfunctions $\sin\left(\frac{k\pi}{2l}(x+l)\right)$ of the operator $\frac{\partial^2}{\partial x^2}$ with homogeneous Dirichlet BCs on the interval $(-l, l)$. Furthermore the set $\{\Phi_k,\, k = 1,\dots,N\}$ is assumed to be orthonormal.

In the following $\|\cdot\|$ denotes a suitable norm on the space $\mathbb{R}^{nN}$, e.g., the discrete L_2-norm defined as usual: If $w = (w_1,\dots,w_{nN})^\top \in \mathbb{R}^{nN}$ and $h = \frac{2l}{N+1}$, then $\|w\| := \left(h\sum_{i=1}^{nN} w_i^2\right)^{1/2}$.

Let $U_h(t)$ be the restriction of $u(t,x)$ to the spatial grid Ω_h. Of interest is the space truncation error of (3.1) defined by

$$\alpha_h(t) = (I_N \otimes A)U_h'(t) + \left(\frac{1}{h^2}P \otimes B + I_N \otimes C\right)U_h(t) - F(t) + r(t). \tag{3.6}$$

With a simple Taylor expansion (recall that $U_h(t)$ is smooth by assumption) we obtain

$$\left(\frac{1}{h^2}P \otimes B\right)U_h(t) + r(t) = (I_N \otimes B)U_{xx,h}(t) + \alpha_h(t) \tag{3.7}$$

with $\alpha_h(t) = h^2(I_N \otimes B)\gamma_h(t)$, where $\gamma_h(t) = (\gamma_1,\dots,\gamma_{nN})^\top$. The components γ_i of γ_h are bounded under the assumption that $\max\limits_{-l\le x\le l}\left\|\frac{\partial^4}{\partial x^4}u(t,x)\right\|_n$ is bounded for $t \in I^* := [0,t^*]$, $t^* > 0$. ($\|\cdot\|_n$ denotes the Euclidean norm on the space $\mathbb{R}^n$.) Then the estimate $\|\alpha_h(t)\| = \mathcal{O}(h^2)$, $t \in I^*$, follows. That means the discretization in space is consistent of second order. The space truncation error will enter the analysis below.

3.2 Time discretization and convergence of the full discretization.

For the time integration of the DAE (3.1) we use the implicit Euler and the trapezoidal rule. The result is the BTCS and the Crank–Nicolson scheme for the PDAE, respectively. For these two discretization methods we examine the behavior of the full local and global error associated with the exact solution of the IBVP (1.1)–(1.4). For ease of representation we restrict ourselves to a constant stepsize τ.

3.2.1 The BTCS scheme.

This scheme for the PDAE (1.1) is given by

$$G(\tau,h^2)\,U_{m+1} = \left(\tfrac{1}{\tau}I_N \otimes A\right)U_m + F(t_{m+1}) - r(t_{m+1}), \quad U_0 = U(0), \tag{3.8}$$

$$G(\tau,h^2) = \tfrac{1}{\tau}I_N \otimes A + \tfrac{1}{h^2}P \otimes B + I_N \otimes C. \tag{3.9}$$

Of course, if (3.8) is to be used for the calculation of U_{m+1} then it should have an unique solution (G regular).

LEMMA 3.1. *Suppose the matrices*

$$G_k(\tau, h^2) = \frac{1}{\tau} A + \lambda_k \, B + C, \quad k = 1, \dots, N, \tag{3.10}$$

are regular for $0 < \tau \le \tau_0$, $0 < h \le h_0$ (τ, h *fixed*). *Then the matrix* $G(\tau, h^2)$ *is regular and there is an unique solution* U_{m+1} *of* (3.8).

PROOF. Let $\Phi := \sqrt{h}(\Phi_1, \dots, \Phi_N)$ be the matrix of the eigenvectors of the matrix $\frac{1}{h^2}P$ and $\Lambda := \operatorname{diag}(\lambda_1, \dots, \lambda_N)$ (see (3.5)). Then it is $\Phi = \Phi^{-1}$ and we can write

$$\begin{aligned} G &= (\Phi \otimes I_n) \left(\frac{1}{\tau} I_N \otimes A + \Lambda \otimes B + I_N \otimes C \right) (\Phi \otimes I_n) \\ &= (\Phi \otimes I_n) \begin{pmatrix} G_1 & & \\ & \ddots & \\ & & G_N \end{pmatrix} (\Phi \otimes I_n). \end{aligned}$$

Thus if the low order matrices $G_k(\tau, h^2)$, $k = 1, \dots, N,$ are regular, then $G(\tau, h^2)$ is regular. □

This lemma is important for the introduction of the proper quality (defined below) of the matrix $G^{-1}(\tau, h^2)$. This is seen from

LEMMA 3.2. *Let* $G(\tau, h^2)$ *be regular and* $\nu_{d,t}$ *and* $\nu_{d,x}$ *be the uniform differential time index and the differential space index of the PDAE* (1.1), *respectively. Then the following relations for the regular matrix* $G^{-1}(\tau, h^2)$ *hold for* $i, j = 1, \dots, nN$

(i) *for* $\nu_{d,t} \ge 0$ *and fixed* $h > 0$: $\lim_{\tau \to 0} \tau^{\nu_{d,t}} \left(G^{-1}(\tau, h^2) \right)_{i,j} = 0,$

(ii) *for* $\nu_{d,t} \ge 0$ *and fixed* $h > 0$: $\lim_{\tau \to 0} \tau^{\nu_{d,t}} \left(G^{-1}(\tau, h^2) \frac{1}{\tau} (I_N \otimes A) \right)_{i,j} = 0,$

(iii) *for* $\nu_{d,x} \ge 0$ *and fixed* $\tau > 0$: $\lim_{h \to 0} h^{\nu_{d,x}} \left(G^{-1}(\tau, h^2) \right)_{i,j} = 0,$

(iv) *for* $\nu_{d,x} \ge 1$ *and fixed* $\tau > 0$: $\lim_{h \to 0} h^{\nu_{d,x}} \left(G^{-1}(\tau, h^2) \frac{1}{h^2} (I_N \otimes B) \right)_{i,j} = 0.$

PROOF. First we consider the matrix pencil $(I_N \otimes A, \frac{1}{h^2} P \otimes B + I_N \otimes C)$ which is regular. Thus we may perform a Kronecker transformation of this pencil with regular matrices P_h, $Q_h \in \mathbb{R}^{nN \times nN}$, i.e.

$$I_N \otimes A = P_h^{-1} \begin{pmatrix} I_1 & 0 \\ 0 & \bar{N} \end{pmatrix} Q_h^{-1}, \qquad \frac{1}{h^2} P \otimes B + I_N \otimes C = P_h^{-1} \begin{pmatrix} R_h & 0 \\ 0 & I_2 \end{pmatrix} Q_h^{-1},$$

where $\bar{N}$ is a nilpotent matrix provided the matrix A is singular, and $I_1, R_h \in \mathbb{R}^{N_1 \times N_1}$, I_2, $\bar{N} \in \mathbb{R}^{N_2 \times N_2}$, $N_1 + N_2 = nN$ (I_m is an unit matrix). Thus G^{-1} can be written

$$G^{-1}(\tau, h^2) = Q_h \begin{pmatrix} \tau (I_1 + \tau R_h)^{-1} & 0 \\ 0 & (I_2 + \frac{1}{\tau} \bar{N})^{-1} \end{pmatrix} P_h.$$

The statements (i) and (ii) now follow easily from the transformation of $I_N \otimes A$ and from the identity $(I_2 + \frac{1}{\tau}\bar{N})^{-1} = \sum_{l=0}^{\nu_{d,t}-1}(-\frac{1}{\tau}\bar{N})^l$. In the same manner we show with a Kronecker transformation of the matrix pencil $\left(\frac{1}{h^2}P \otimes B, I_N \otimes \left(\frac{1}{\tau}A + C\right)\right)$ the statements (iii) and (iv). □

Therefore, alternatively (cf. Definitions 2.1 and 2.3) we can characterize for the BTCS scheme the differential space index $\nu_{d,x}$ and the uniform differential time index $\nu_{d,t}$ of the PDAE (1.1) by the following

Corollary 3.3. *The differential space index $\nu_{d,x}$ and the uniform differential time index $\nu_{d,t}$ are for the BTCS scheme the smallest integers such that property* (iii) *and property* (i) *in Lemma* 3.2 *hold, respectively.*

Definition 3.1. *The matrices $\tau^{\nu_{d,t}}G^{-1}(\tau,h^2)$ satisfying property* (i) *and $h^{\nu_{d,x}}G^{-1}(\tau,h^2)$ satisfying property* (iii) *in Lemma* 3.2 *are called τ-proper and h-proper, respectively.*

From property (ii) and property (iv) we see that for $\nu_{d,t}, \nu_{d,x} \geq 1$ the matrices

$$\tau^{\nu_{d,t}}G^{-1}(\tau,h^2)\tfrac{1}{\tau}(I_N \otimes A) \quad \text{and} \quad h^{\nu_{d,x}}G^{-1}(\tau,h^2)\tfrac{1}{h^2}(I_N \otimes B)$$

have the same proper quality as the matrices $\tau^{\nu_{d,t}}G^{-1}(\tau,h^2)$ and $h^{\nu_{d,x}}G^{-1}(\tau,h^2)$, respectively.

Remark 3.1. The characterization of the indexes of the PDAE for the BTCS scheme by means of the proper quality is of a purely algebraic nature. In contrast, its definitions given earlier in Definitions 2.1 and 2.3 use the differential properties of the algebraic solution components obtained after a Kronecker transformation.

Remark 3.2. The algebraic x-index for the regular matrix $R(s,z) = (sA + z^2B + C)^{-1}$ (s and z some parameters) defined by Campbell and Marszalek [5] is equal to the differential space index $\nu_{d,x}$. These authors also introduced the proper quality of the matrix $R(s,z) = (sA+z^2B+C)^{-1}$ for $|s| \to \infty$ or $z^2 \to \infty$.

Now let us introduce the *full truncation error*

$$le_{m+1} = U_h(t_{m+1}) - \hat{U}_{m+1} \tag{3.11}$$

with $\hat{U}_{m+1} = G^{-1}(\tau,h^2)\left(\left(\frac{1}{\tau}I_N \otimes A\right)U_h(t_m) + \tilde{F}(t_{m+1})\right)$. We emphasize that this error is defined with respect to the true PDAE solution $U_h(t)$ and not for the true DAE solution $U(t)$. We introduce the following:

Definition 3.2. *The BTCS scheme* (3.8) *is consistent with the PDAE* (1.1) *if the full truncation error le_{m+1} satisfies*

$$\tfrac{1}{\tau}\|le_{m+1}\| \to 0 \quad \textit{for} \quad \tau,\, h \to 0, \quad m = 0,1,\ldots. \tag{3.12}$$

Definition 3.3. *The BTCS scheme is said to be accurate of order (p,q), $p,q \geq 1$, if*

$$\frac{1}{\tau}\|le_{m+1}\| = \mathcal{O}(h^p) + \mathcal{O}(\tau^q) \quad \textit{for} \quad \tau,\, h \to 0. \tag{3.13}$$

It is easy to see that if the BTCS scheme is accurate of order (p, q), then it is consistent. If the relations (3.12) and (3.13) hold under a time space restriction of the form (3.18) we say that the BTCS scheme is conditionally consistent and conditionally accurate of order (p, q), respectively. With equation (3.6) we obtain

$$\begin{aligned} le_{m+1} = G^{-1}(\tau, h^2)\Big\{(\tfrac{1}{\tau} I_N \otimes A)\Big[U_h(t_{m+1}) - U_h(t_m)\Big] \\ - (I_N \otimes A)U_h'(t_{m+1}) + \alpha_h(t_{m+1})\Big\} \end{aligned}$$

and from this follows with a Taylor expansion of $U_h(t_{m+1})$ and $U_h'(t_{m+1})$ in t_m

(3.14)
$$\begin{aligned} le_{m+1} = \tau G^{-1}(\tau, h^2)(I_N \otimes A)\Big[\tfrac{1}{2}U_h''(t_m + \zeta\tau) - U_h''(t_m + \overline{\zeta}\tau)\Big] \\ + h^2 G^{-1}(\tau, h^2)(I_N \otimes B)\gamma_h(t_{m+1}) \end{aligned}$$

where $\zeta, \overline{\zeta} \in (0, 1)$ and may be different for each component of U_h. Therefore, for the full truncation error we have for $t \in I^*$ (cf. the foregoing section) the estimate

(3.15)
$$\|le_{m+1}\| \leq C_0 \left(\tau\|G^{-1}(\tau, h^2)(I_N \otimes A)\| \, + \, h^2\|G^{-1}(\tau, h^2)(I_N \otimes B)\|\right)$$

where C_0 is a positive constant independent of τ and h. The terms on the right hand side of this inequality can be estimated easily with Lemma 3.2 because it implies that the following relations hold for $\tau, h \to 0$ with constants κ_i (independent of τ, h)

(3.16)
$$\begin{aligned} \tau^{\nu_{d,t}-1}\|G^{-1}(\tau, h^2)\| &\leq \kappa_1, &&\text{for } \nu_{d,t} \geq 0, \\ h^{\nu_{d,x}-1}\|G^{-1}(\tau, h^2)\| &\leq \kappa_2, &&\text{for } \nu_{d,x} \geq 0, \\ \tau^{\nu_{d,t}-2}\|G^{-1}(\tau, h^2)(I_N \otimes A)\| &\leq \kappa_3, &&\text{for } \nu_{d,t} \geq 1, \\ h^{\nu_{d,x}-3}\|G^{-1}(\tau, h^2)(I_N \otimes B)\| &\leq \kappa_4, &&\text{for } \nu_{d,x} \geq 1 \end{aligned}$$

whereby possibly τ and h cannot go independent towards zero. This means it may be that one (or none) of the following conditions must be satisfied:

(3.17)
$$c_0 \leq \frac{\tau}{h^2} \quad \text{or} \quad \frac{\tau}{h^2} \leq c_1 \quad \text{or} \quad c_0 \leq \frac{\tau}{h^2} \leq c_1, \qquad c_0, c_1 \in \mathbb{R}^+ .$$

The estimate (3.15) and the asymptotic relations (3.16) can now be used to state (in the sense of Definition 3.3) an accuracy result:

THEOREM 3.4. *Let for a given PDAE* (1.1) *the asymptotic relations* (3.16) *be valid* (*possibly under one of the conditions* (3.18)). *Then the BTCS scheme is for* $t \in I^*$ (*conditionally*) *accurate of order* $(2, 1)$ *for*

- *a uniform time index* $\nu_{d,t} = 0$ *and a space index* $\nu_{d,x} \geq 0$,
- $\nu_{d,t} \geq 0$ *and* $\nu_{d,x} = 0$,
- $\nu_{d,t} = 1$ *and* $\nu_{d,x} = 1$.

Proof. Let K be a constant which is independent of τ and h and possibly different in different expressions. First we consider the case $\nu_{d,t} = 0$, $\nu_{d,x} \geq 0$. Using in (3.16) the first inequality we can estimate the right hand side of inequality (3.15) further by (the arguments of G are suppressed)

$$K \left(\tau^2 \|\tau^{-1} G^{-1}\| + \tau h^2 \|\tau^{-1} G^{-1}\|\right) = \mathcal{O}(\tau^2) + \mathcal{O}(\tau h^2).$$

According to Definition 3.3 we see that the BTCS scheme is accurate of order (2,1) for these indexes.
In the same way we obtain in the case $\nu_{d,t} \geq 0$, $\nu_{d,x} = 0$ by means of the second inequality in (3.16) for the right hand side of (3.15) the upper bound

$$K \left(\tau h^2 \|h^{-2} G^{-1}\| + h^4 \|h^{-2} G^{-1}\|\right) = \mathcal{O}(\tau h^2) + \mathcal{O}(h^4).$$

If we now take into account the condition $\tau/h^2 \leq c_1$, then the right hand side reduces to $\mathcal{O}(\tau^2) + \mathcal{O}(\tau h^2)$. Using again Definition 3.3 we obtain the second statement of the theorem.

The third statement can be proved by using the last two inequalities of (3.16). We find for the right of (3.15) the upper bound

$$K \left(\tau^{3-\nu_{d,t}} \|\tau^{\nu_{d,t}-2} G^{-1}(I_N \otimes A)\| + h^{5-\nu_{d,x}} \|h^{\nu_{d,x}-3} G^{-1}(I_N \otimes B)\|\right).$$

Inserting $\nu_{d,t} = \nu_{d,x} = 1$ into this bound it reduces to an expression of the form $\mathcal{O}(\tau^2) + \mathcal{O}(h^4)$ which is equivalent to $\mathcal{O}(\tau^2) + \mathcal{O}(\tau h^2)$ because $0 < c_0 \leq \frac{\tau}{h^2}$. This proves the third statement of the theorem. □

Now we introduce the *full discretization error* at time level t_{m+1}

$$\varepsilon_h(t_{m+1}) = U_h(t_{m+1}) - U_{m+1}$$

and:

Definition 3.4. *If the order relation*

$$\|\varepsilon_h(t_{m+1})\| = \mathcal{O}(h^p) + \mathcal{O}(\tau^q) \quad \textit{for} \quad \tau,\, h \to 0,\; m = 0, 1, \ldots, \tag{3.18}$$

is valid, then the BTCS scheme is called convergent of order (p, q). *If this relation holds under a time space condition* (3.18), *then the BTCS scheme is said to be conditionally convergent of order* (p, q).

With (3.11) for the truncation error we obtain the recursion

$$\varepsilon_h(t_{m+1}) = G^{-1}(\tau, h^2) \left(\tfrac{1}{\tau} I_N \otimes A\right) \varepsilon_h(t_m) + le_{m+1}, \tag{3.19}$$

and by induction we get

(3.20)
$$\varepsilon_h(t_{m+1}) = \Big(G^{-1}(\tau,h^2)\tfrac{1}{\tau}I_N \otimes A\Big)^{m+1}\varepsilon_h(0) + \sum_{i=0}^{m}\Big(G^{-1}(\tau,h^2)\tfrac{1}{\tau}I_N \otimes A\Big)^{i} le_{m+1-i}.$$

It follows that

(3.21)
$$\begin{aligned}\|\varepsilon_h(t_{m+1})\| \le{}& \Big\|\big(G^{-1}(\tau,h^2)\tfrac{1}{\tau}I_N \otimes A\big)^{m+1}\Big\|\,\|\varepsilon_h(0)\|\\ &+ \max_{i=0}^{m}\|le_{i+1}\|\sum_{i=0}^{m}\Big\|\big(G^{-1}(\tau,h^2)\tfrac{1}{\tau}I_N \otimes A\big)^{i}\Big\|\end{aligned}$$

We require for $0 < m\,\tau \le t^*$

$$\sup_{i\in\mathbb{N}}\Big\{\|\,\big(G^{-1}(\tau,h^2)\tfrac{1}{\tau}I_N \otimes A\big)^{i}\,\|\Big\} \;<\; \infty. \tag{3.22}$$

where τ, h may be restricted according to (3.18).

REMARK 3.3. *If the matrices A and B are regular, then this requirement is the Lax stability condition.* Assuming $\|\varepsilon_h(0)\| = \mathcal{O}(h^2)$, we get for $t \in I^*$ the estimate

$$\|\varepsilon_h(t_{m+1})\| \le \overline{K}(m+1)\max_{i=0}^{m}\|le_{i+1}\| \;\le\; K\max_{i=0}^{m}\frac{\|le_{i+1}\|}{\tau}$$

where $\overline{K}$ and K are positive constants independent of τ and h for $\tau,\ h\ \to 0$ under the condition $\tau\, m = t \in I^*$ (t fixed). With Theorem 3.4 we obtain immediately

THEOREM 3.5. *Suppose the assumptions of Theorem* 3.4 *and the estimate* (3.23) *are valid. Then the BTCS scheme is* (*conditionally*) *convergent with respect to $\|\cdot\|$ of order* $(2,1)$ *for $t \in I^*$ and for*

- $\nu_{d,t} = 0$ *and* $\nu_{d,x} \ge 0,$
- $\nu_{d,t} \ge 0$ *and* $\nu_{d,x} = 0,$
- $\nu_{d,t} = \nu_{d,x} = 1.$

3.2.2 Finite Fourier representation of the full discretization error.

Beside the full discretization error $\varepsilon_h(t_{m+1})$ we define a generalized Fourier component of this error in terms of the low order $(n \times n)$–matrices $G_k^{-1}\frac{A}{\tau}$ and $G_k^{-1}B$. This Fourier error component is (as shown below) convenient for an investigation of necessary conditions for the convergence of the numerical scheme considered. In the following we use in the vector space $\mathbb{R}^N$ the discrete L_2-norm corresponding to the inner product

$$\langle x, y\rangle = hx^\top y = \sum_{i=1}^{N} h\, x_i\, y_i\,, \quad x, y \in \mathbb{R}^N, \quad h = \tfrac{2l}{N+1}.$$

The eigenvectors Φ_k (cf. (3.5)) are supposed to be normed in the norm induced by this scalar product.

LEMMA 3.6. *Let for fixed $\tau, h > 0$*

(i) G_k^{-1} *be regular for $k = 1, \dots, N$;*

(ii) $u(t,x)$ (*and hence $U_h(t)$ and $\gamma_h(t)$*) *be sufficiently smooth for $t \in I^*$, $x \in [-l, l]$;*

(iii) $\varepsilon_h(t_j)$ *be represented as a finite generalized Fourier series with respect to Φ_k*

$$\varepsilon_h(t_j) = \sum_{k=1}^{N} \Phi_k \otimes e_{h,k}^j, \qquad e_{h,k}^j \in \mathbb{R}^n, \quad j = 0, 1, \dots. \tag{3.23}$$

Then $e_{h,k}^{m+1}$ can be estimated for $t \in I^*$ by

(3.24)
$$\|e_{h,k}^{m+1}\|_n \le \left\| \left(G_k^{-1} \tfrac{A}{\tau}\right)^{m+1} \right\|_n \|e_{h,k}^0\|_n + \omega \sum_{i=0}^{m} \left\| \left(G_k^{-1} \tfrac{A}{\tau}\right)^{i} \right\|_n \left[h^2 \|G_k^{-1} B\|_n + \tau^2 \left\|G_k^{-1} \tfrac{A}{\tau}\right\|_n \right]$$

where the positive coefficient ω is independent of τ and h.

NOTE. When M is an $(n \times n)-$matrix, then $\|M\|_n$ is the spectral norm compatible with the Euclidean vector norm.

PROOF. We rewrite (3.20) in the form

$$G(\tau, h^2)\, \varepsilon_h(t_{m+1}) = \frac{1}{\tau}(I_N \otimes A)\, \varepsilon_h(t_m) + G(\tau, h^2)\, le_{m+1}$$

and multiply it from the left with $h(\Phi_{k'}^{\top} \otimes I_n)$, $k' \in \{1, \dots, N\}$. Using now (3.14), one gets with the Fourier representation indicated in assumption (iii)

$$G_k\, e_{h,k}^{m+1} = \tfrac{A}{\tau}\, e_{h,k}^m + h^2 (h\, \Phi_k^{\top} \otimes B) \gamma_h(t_{m+1}) + \tau (h\, \Phi_k^{\top} \otimes A) \left[\tfrac{1}{2} U_h''(t_m + \zeta\tau) - U_h''(t_m + \overline{\zeta}\tau) \right].$$

Combining this formula with assumptions (i) and (ii) the assertion of the lemma follows by taking norm estimates. □

The Fourier error components $e_{h,k}^{m+1}$ can be used to derive a necessary condition for the convergence of the numerical scheme.

LEMMA 3.7. *Suppose that the assumptions of Lemma 3.6 are valid for $0 < \tau \le \tau_0$, $0 < h \le h_0$ and that $\|\varepsilon_h(t_{m+1})\|$ goes to zero for τ, $h \to 0$. Then the relation $\|e_{h,k}^{m+1}\|_n \to 0$ for τ, $h \to 0$ holds for $t \in I^*$. Furthermore the convergence order (with respect to τ and h) of $\|e_{h,k}^{m+1}\|_n$ is the same as the convergence order of $\|\varepsilon_h(t_{m+1})\|$.*

PROOF. We multiply (3.24) from the left by $h(\Phi_{k'}^\top \otimes I_n)$ and use the orthogonality of the eigenvectors Φ_k, $\Phi_{k'}$ for $k \neq k'$ to get

$$e_{h,k'}^{m+1} = h(\Phi_{k'}^\top \otimes I_n)\varepsilon_h(t_{m+1}).$$

From this formula the statement of the lemma follows by taking norms. □

In the following example it is assumed that the initial error component $\|e_{h,k}^0\|_n$ in inequality (3.25) is $\mathcal{O}(h^2)$.

EXAMPLE 3.1. Let $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} b_1 & b_2 \\ 0 & 0 \end{smallmatrix}\right)$ with $b_1 < 0$, $C = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. The uniform differential time index and the differential spatial index are easily shown to be $\nu_{d,t} = 1$ and $\nu_{d,x} = 1$, respectively. $\|G_k^{-1}\frac{A}{\tau}\|_n$ is of order $\mathcal{O}(1)$ for $\tau, h \to 0$, while $\|G_k^{-1}B\|_n$ is of order $\mathcal{O}(\tau)$ as well for $\frac{\tau}{h^2} \geq c_0$ as for $\frac{\tau}{h^2} \leq c_1$ (cf. the inequalities in (3.18)). Therefore, no restrictions for τ/h^2 are needed in this example. Of special interest is the norm $\left\|G_k^{-1}\frac{A}{\tau}\right\|_n = 1/|1 + \tau(\lambda_k b_1 + 1)|$ which must be summed over in the estimate of the right hand side of inequality (3.25). Because $\|G_k^{-1}\frac{A}{\tau}\|_n \leq 1$ for τ, $h \to 0$ and $m\tau = t \in I^*$ (t fixed), we find the order relation of the sum

$$\sum_{i=0}^{m} \left\| \left(G_k^{-1}\frac{A}{\tau} \right)^i \right\|_n \leq m + 1 = \mathcal{O}\left(\frac{1}{\tau}\right)$$

for $\tau, h \to 0$. Inserting this into inequality (3.25) and taking $\|e_{h,k}^0\|_n = \mathcal{O}(h^2)$ into account we get the result that the BTCS scheme for this example gives convergence of the error component $e_{h,k}^{m+1}$ of the order $\|e_{h,k}^{m+1}\|_n = \mathcal{O}(\tau) + \mathcal{O}(h^2)$.

3.2.3 The Crank–Nicolson scheme.

This discretization scheme can be generated if the trapezoidal rule (in the following expressed by the lower index T) is applied to (3.1). Using the grid Ω_h defined in Section 3 the analogue to (3.8) for the Crank–Nicolson scheme is

(3.25)
$$G_T(\tau, h^2)U_{T,m+1} = \Big[\frac{1}{\tau}I_N \otimes A - H_h\Big]U_{T,m} + \frac{1}{2}\left(\tilde{F}(t_{m+1}) + \tilde{F}(t_m)\right)$$

with $G_T(\tau, h^2) = \frac{1}{\tau}I_N \otimes A + H_h$, $H_h = \frac{1}{2}\left(\frac{1}{h^2}P \otimes B + I_N \otimes C\right)$ and $\tilde{F}(t) = F(t) - r(t)$. We consider again the full truncation error $le_{T,m+1} = U_h(t_{m+1}) - \hat{U}_{T,m+1}$ where now

$$\hat{U}_{T,m+1} = G_T^{-1}(\tau, h^2)\Big[\tfrac{1}{\tau}I_N \otimes A - H_h\Big]U_h(t_m) + \tfrac{1}{2}G_T^{-1}(\tau, h^2)\Big[\tilde{F}(t_{m+1}) + \tilde{F}(t_m)\Big]. \tag{3.26}$$

In the same manner as in Section 3.2.1 we get for $le_{T,m+1}$ the representation

(3.27)
$$le_{T,m+1} = G_T^{-1}(\tau, h^2)\Big\{\left(\tfrac{1}{\tau}I_N \otimes A\right)[U_h(t_{m+1}) - U_h(t_m) - \tfrac{\tau}{2}\left(U_h'(t_{m+1}) - U_h'(t_m)\right)] + \tfrac{1}{2}\left(\alpha_h(t_{m+1}) + \alpha_h(t_m)\right)\Big\},$$

with the estimate

$$\|le_{T,m+1}\| \leq C_1 \left(\tau^3 \left\|G_T^{-1}(\tau,h^2)\tfrac{1}{\tau}(I_N \otimes A)\right\| + h^2 \|G_T^{-1}(\tau,h^2)(I_N \otimes B)\|\right), \tag{3.28}$$

where C_1 is positive and independent of τ and h. This is the analogue of inequality (3.15) for the BTCS scheme.

Taking into account that the proper quality is valid also for the matrix G_T^{-1} (cf. Lemma 3.2) we obtain from the last estimate the following accuracy result (analogous to Theorem 3.4 for the BTCS scheme):

THEOREM 3.8. *Let for a given PDAE* (1.1) *the asymptotic relations* (3.16) (*where* G *must be replaced by* G_T) *be valid* (*possibly under one of the conditions* (3.18)). *Then the Crank–Nicolson scheme is for* $t \in I^*$ (*conditionally*) *accurate of order* $(2,2)$ *for*

- $\nu_{d,t} = 0$ *and* $\nu_{d,x} \geq 0$,
- $\nu_{d,t} \geq 0$ *and* $\nu_{d,x} = 0$,
- $\nu_{d,t} = \nu_{d,x} = 1$.

It is (*conditionally*) *accurate of order* $(2,1)$ *for*

- $\nu_{d,t} = 2$ *and* $\nu_{d,x} = 1$.

In order to obtain a convergence result also for the Crank–Nicolson scheme (like Theorem 3.5 for the BTCS scheme) we consider the full discretization error $\varepsilon_{T,h}(t_{m+1}) = U_h(t_{m+1}) - U_{T,m+1}$ under the assumption that the matrix $G_T(\tau,h^2)$ is regular. Because

$$U_{T,m+1} = G_T^{-1}(\tau,h^2)\Big(\left[\tfrac{1}{\tau} I_N \otimes A - H_h\right] U_{T,m} + \tfrac{1}{2}\left[\tilde{F}(t_{m+1}) + \tilde{F}(t_m)\right] \Big)$$

(this follows from (3.26)) and $U_h(t_{m+1}) = le_{T,m+1} + \hat{U}_{T,m+1}$ (see (3.11)) we can write

$$\varepsilon_{T,h}(t_{m+1}) = G_T^{-1}(\tau,h^2)\left(\tfrac{1}{\tau} I_N \otimes A - H_h\right)\varepsilon_{T,h}(t_m) + le_{T,m+1} \tag{3.29}$$

which is analogous to (3.20). Thus we also get the analogue of inequality (3.22) which reads for the Crank–Nicolson method

$$\begin{aligned}\|\varepsilon_{T,h}(t_{m+1})\| \leq{}& \left\|\left(G_T^{-1}(\tau,h^2)\left[\tfrac{1}{\tau} I_N \otimes A - H_h\right]\right)^{m+1}\right\| \cdot \|\varepsilon_{T,h}(0)\| \\ &+ \max_{0\leq j\leq m} \|le_{j+1}\| \sum_{i=0}^{m} \left\|\left(G_T^{-1}(\tau,h^2)\left[\tfrac{1}{\tau} I_N \otimes A - H_h\right]\right)^i\right\|.\end{aligned} \tag{3.30}$$

Requiring that for $0 < m\tau \leq t^*$

$$\sup_{i\in\mathbb{N}_+} \left\{\left\|\left(G_T^{-1}(\tau,h^2)\left[\tfrac{1}{\tau} I_N \otimes A + H_h\right]\right)^i\right\|\right\} < \infty \tag{3.31}$$

(τ, h may be restricted; see (3.18)) we obtain the following convergence theorem:

THEOREM 3.9. *Suppose the assumptions of Theorem* 3.8 *and the estimate* (3.32) *are valid. Then the Crank–Nicolson scheme is for $t \in I^*$* (*conditionally*) *convergent of order* $(2, 2)$ *for*

- $\nu_{d,t} = 0$ *and* $\nu_{d,x} \geq 0$,
- $\nu_{d,t} \geq 0$ *and* $\nu_{d,x} = 0$,
- $\nu_{d,t} = \nu_{d,x} = 1$.

It is convergent of order $(2, 1)$ *for*

- $\nu_{d,t} = 2$ *and* $\nu_{d,x} = 1$ *under a condition* $0 < c_0 \leq \tau/h^2$.

Furthermore, as for the BTCS method a finite Fourier representation of the full discretization error $\varepsilon_{T,h}(t_{m+1})$ can be given also for the Crank–Nicolson scheme. It is based on the eigenvectors $\Phi_{T,k}$ of the matrix $\frac{1}{2h^2}P$ (for P see (3.2)): $\frac{1}{2h^2}P\Phi_{T,k} = \lambda_{T,k}\Phi_{T,k}$, $k = 1, \ldots, N$. For completeness we give the analogue of Lemma 3.6:

LEMMA 3.10. *Let for fixed τ, $h > 0$*

(i) $G_{T,k}^{-1} = \left(\frac{1}{\tau}A + \lambda_{T,k}B + \frac{1}{2}C\right)^{-1}$ *be regular for $k = 1, \ldots, N$;*

(ii) $u(t, x)$ *(and hence $U_h(t)$ and $\gamma_h(t)$) be sufficiently smooth for $t \in I^*$, $x \in [-l, l]$;*

(iii) $\varepsilon_{T,h}(t_j)$ *be represented as a finite Fourier series with respect to $\Phi_{T,k}$,*

$$\varepsilon_{T,h}(t_j) = \sum_{k=1}^{N} \Phi_{T,k} \otimes e_{T,h,k}^{j}, \quad j = 0, 1, \ldots, \quad e_{T,h,k}^{j} \in \mathbb{R}^n .$$

Then $e_{T,h,k}^{m+1}$ can be estimated for $t \in I^*$ by

$$\begin{aligned} \|e_{T,h,k}^{m+1}\|_n \leq{}& \left\| \left(G_{T,k}^{-1} \left[\tfrac{1}{\tau}A - \lambda_{T,k}B - \tfrac{1}{2}C \right] \right)^m \right\|_n \cdot \|e_{T,h,k}^{0}\|_n \\ &+ \omega_T \sum_{i=0}^{m} \left\| \left(G_{T,k}^{-1} \left[\tfrac{1}{\tau}A - \lambda_{T,k}B - \tfrac{1}{2}C \right] \right)^i \right\|_n \\ &\qquad \cdot \left[h^2 \left\| G_{T,k}^{-1} B \right\|_n + \tau^3 \left\| G_{T,k}^{-1} \tfrac{1}{\tau} A \right\|_n \right] \end{aligned}$$

where ω_T is a positive constant independent of τ and h.

PROOF. We multiply (3.30) from the left by G_T and apply (3.28) to express $G_T(\tau, h^2)le_{T,m+1}$ (after a Taylor expansion; cf. Section 3.2.1) in terms of A and B. The resulting equation is multiplied from the left by $h\left(\Phi_{T,k'}^{\top} \otimes I_n\right)$. Using the orthogonality of the system $\{\Phi_{T,k}\}_1^N$, the definition of the scalar product (cf. Section 3.2.2) and assumption (i) we obtain the statement of the lemma by taking norms. □

We note that the analogue of Lemma 3.7 holds also for the Crank–Nicolson scheme ($\varepsilon_h(t_{m+1})$ must be replaced by $\varepsilon_{T,h}(t_{m+1})$ and so on). Lemma 3.10 is useful especially because the norms of only low order $(n \times n)$-matrices must be analyzed (provided n is low as is often the case in applications).

4 Numerical examples.

In the following some results of numerical test calculations with the BTCS scheme are presented. For specific examples we confirm the convergence results derived in Sections 3.2.1 and 3.2.2, and we use this scheme also to demonstrate that the consistency of the BCs is essential for an accurate solution of the PDAE (1.1). Examples with inconsistent ICs and BCs can be found also in [4] (an extended version of [6]) and in [18]. Using again the BTCS scheme it is shown that if an index-jump-free PDAE is semidiscretized not properly, then this semidiscretized version may have an index jump. This can cause large errors in the corresponding numerical solution.

In the examples considered the inhomogeneity $f(t,x)$ of the PDAE (1.1) is chosen such that an exact solution is known. Further we take $l = 1$ and BCs (1.2) and ICs (1.3) needed for numerical calculations are then taken from the exact solution. Therefore the BCs and ICs can (if necessary) assumed to be consistent.

4.1 Order relations.

First we note that the full discretization error ε_h defined in Section 3 can be considered to be a function of τ and h (see, e.g., (3.21)). We therefore write here $\varepsilon_{h,\tau}$.

Example 4.1. In order to show the convergence order clearly, we consider example 3.1 with $b_1 = b_2 = -1$. Let $f(t,x)$ be such that

$$u_I(t,x) = \big(x(x^2-1)\,e^{-t},\ (x^2-1)\,e^{10\,t}\big)^\top \in \mathbb{R}^2$$

is the exact solution. Because the components of $u_I(t,x)$ are polynomials in x of degree ≤ 3, there will be no spatial error by discretization of u_{xx} with standard second order finite differences, i.e. $\alpha_h(t) = 0$ (cf. (3.7)). Comparing $\varepsilon_{h,\tau}$ with $\varepsilon_{h,\frac{\tau}{2}}$ under, e.g., the condition $\tau\, m = t = 1$ we can determine the numerical order of convergence $p_{num,\tau} = \log_2 \|\varepsilon_{h,\tau}\| - \log_2 \|\varepsilon_{h,\frac{\tau}{2}}\|$ with respect to τ. Analogously if we choose $f(t,x)$ such that

$$u_{II}(t,x) = \big(x^6(x^2-1)\,t,\ x^4(x^2-1)\,t\big)^\top \in \mathbb{R}^2,$$

is the exact solution, then the second order derivatives in time of $U_h(t)$ and $F(t)$ in (3.13) vanish identically, and the error is only an error in space. Its convergence order can be determined numerically by means of the expression

$p_{num,h} = \log_2 \|\varepsilon_{h,\tau}\| - \log_2 \|\varepsilon_{\frac{h}{2},\tau}\|$ with respect to h. Table 4.1 shows (in agreement with the investigation in the foregoing subsections) that the τ-order is 1, and the h-order is 2.

Table 4.1: Numerical order of convergence for example 4.1.

	$p_{num,\tau}$ using $u_I(t,x)$						$p_{num,h}$ using $u_{II}(t,x)$					
$0.1/\tau$	2^2	2^3	2^4	2^5	2^6	2^7	2^2	2^3	2^4	2^5	2^6	2^7
$0.1/h$												
1	1.01	1.01	1.00	1.00	1.00	1.00	1.92	1.92	1.92	1.92	1.92	1.92
2^1	1.01	1.01	1.00	1.00	1.00	1.00	1.98	1.98	1.98	1.98	1.98	1.98
2^2	1.01	1.01	1.00	1.00	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00
2^3	1.01	1.01	1.00	1.00	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00

EXAMPLE 4.2. As the next example we consider the PDAE (1.1) generated by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Its indexes are $\nu_{d,t} = 2, \nu_{d,x} = 3$. Although Theorem 3.4 cannot be applied, the BTCS scheme may be practically useful. If we choose

$$\begin{aligned} u_I(t,x) &= \big(x(x^2-1)e^{-t},\ (x^2-1)e^{t},\ (x+1)(x^2-1)e^{10\,t}\big)^\top \in \mathbb{R}^3 \\ u_{II}(t,x) &= \big(x^6(x^2-1)\,t,\ x^4(x^2-1)\,t,\ (x^4-1)\,t\big)^\top \in \mathbb{R}^3 \end{aligned}$$

we get numerical convergence orders for $m\tau = 1$ as given in Table 4.1, i.e. the convergence is of order $\mathcal{O}(\tau) + \mathcal{O}(h^2)$. This shows that the assumptions of Theorem 3.4 can only be sufficient.

Table 4.2: Numerical order of convergence for example 4.2.

	$p_{num,\tau}$ using $u_I(t,x)$						$p_{num,h}$ using $u_{II}(t,x)$					
$0.1/\tau$	2^2	2^3	2^4	2^5	2^6	2^7	2^2	2^3	2^4	2^5	2^6	2^7
$0.1/h$												
1	0.99	0.99	1.00	1.00	1.00	1.00	1.77	1.77	1.77	1.77	1.77	1.77
2^2	0.99	0.99	1.00	1.00	1.00	1.00	1.95	1.95	1.95	1.95	1.95	1.95
2^4	0.99	0.99	1.00	1.00	1.00	1.00	1.99	1.99	1.99	1.99	1.99	1.99
2^6	1.03	1.01	1.00	1.00	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00

4.2 Consistency of BCs.

The following example shows that it is important to know those components of u for which boundary conditions can be prescribed arbitrarily. For this we consider the example

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} u_t + \begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix} u_{xx} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} u = f \tag{4.1}$$

whose right hand side function $f = (f_1, f_2)^\top$ is chosen such that the exact solution is given by

$$u(t,x) = \begin{pmatrix} v(t,x) \\ w(t,x) \end{pmatrix} := \begin{pmatrix} (x^2-1)(x^2+4)x\cos(\pi\, t) \\ (x^2-1)(x^5-2x^2+5)e^{-t} \end{pmatrix}.$$

Since the matrix A is regular the corresponding MOL-system (3.1) is an ordinary differential equation which can be treated without problems. The set $\mathfrak{M}_{BC}$ can be chosen as $\mathfrak{M}_{BC} = \{1\}$ (another possibility could be $\mathfrak{M}_{BC} = \{2\}$).

In order to illustrate the difference between consistent and inconsistent BCs for the second component of the solution vector, we compare the numerical solution $u_h(t,x)$ with another numerical solution $\tilde{u}_h(t,x)$ where $\tilde{v}_h(t,\pm 1) = v(t,\pm 1)$ and $\tilde{w}_h(t,\pm 1) = w(t,\pm 1) + e^{2t} - 1$, i.e. $\tilde{w}_h(t,\pm 1)$ is an inconsistent BC because f vanishes at the boundary. The ICs $u_h(0,x)$ and $\tilde{u}_h(0,x)$ are the same (i.e. equal to $u(0,x)$) such that the compatibility condition (1.4) is guaranteed. Comparing Figures 4.1 and 4.2 we see that the solutions for consistent and inconsistent BCs are very different. Of course, this is expected because two different problems were solved. Especially the second component $\tilde{w}_h(t,x)$ of the numerical solution calculated with an inconsistent BC has in part very large gradients near the boundary points ±1.

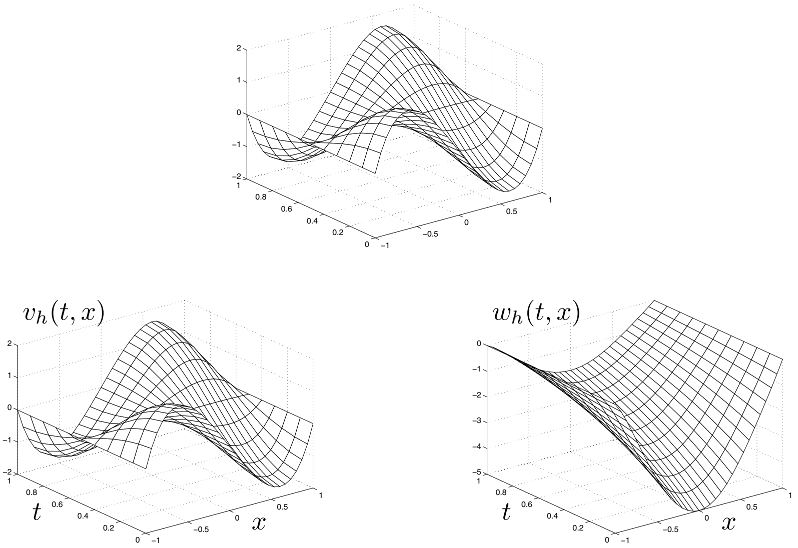

Figure 4.1: Example: PDAE 4.1, consistent BC, $h = \frac{1}{15}$, $\tau = 0.1$.

4.3 The problem of an index jump in the MOL-DAE

Suppose the matrices A, B, C are defined by

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 1 \\ 1 & c \end{pmatrix}.$$

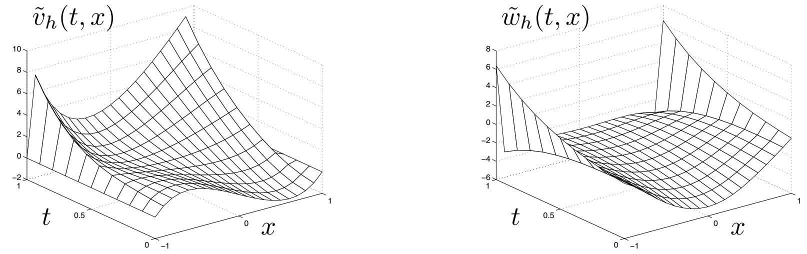

Figure 4.2: Example: PDAE (4.1), inconsistent BC, $h = \frac{1}{15}$, $\tau = 0.1$.

Let further $\mu_k := -\left(\frac{k\pi}{2l}\right)^2$, $k = 1, 2, \ldots$, be the kth eigenvalue mentioned in assumption (IV) of Section 2 (cf. also Section 2.2).

It is easy to see that the matrix pencil $(A, \mu_k B + C)$ has Riesz index 1 for $c \neq \mu_k$ and Riesz index 2 for $c = \mu_{\bar{k}}$ for some $\bar{k}$, i.e. the PDAE has uniform differential time index $\nu_{d,t} = 1$ if $c \neq \mu_k$ $\forall k \in \mathbb{N}_+$ and an index jump if $c = \mu_{\bar{k}}$ for one $\bar{k} \in \mathbb{N}_+$. Of course it may be that the MOL-DAE (3.1) has an index jump for some h and special matrices A, B, C even if the PDAE has an uniform differential time index. In this example this would happen if $c = \lambda_{\bar{k}}$ $(c \neq \mu_k \,\forall\, k)$ for a certain $h > 0$ (this can be shown by a linear transformation of the MOL-DAE using the Kronecker product).

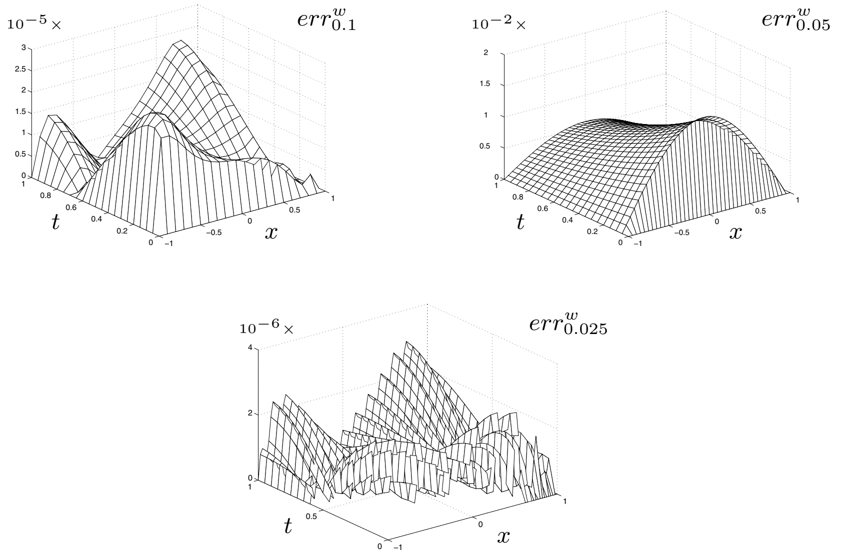

Figure 4.3: $|w_{num}(t,x) - w(t,x)|$, $h = 0.1, 0.05, 0.025$, $\tau = 0.05$.

Let $f(t,x)$ be chosen such that the exact solution of the PDAE is

$$u(t,x) = \begin{pmatrix} v(t,x) \\ w(t,x) \end{pmatrix} := \begin{pmatrix} x^5\,(x^2 - l^2)\cos(\pi t) \\ x^2\,(x^2 - l^2)\,e^{-t} \end{pmatrix}.$$

and define $c := \lambda_1(h) = -\frac{4}{h^2}\sin^2\left(\frac{\pi\, h}{4\, l}\right)$ for $h = h_0 = 0.05$ ($N = 39$). (The numerical value of $\lambda_1(h_0)$ is -2.466133.) Figure 4.3 shows the absolute error err_h^w of the w-component between the numerical solution and the exact solution in each space and time point.

We see that the error for $h = h_0$ is greater than for $h \neq h_0$. Effects like these can also occur in other more complicated examples. We note that because $\lambda_k(h) = \mu_k + \mathcal{O}(h^2)$ for $h \to 0$ there is no index jump of the MOL-DAE for sufficiently small h. At the whole this example demonstrates that we must be careful in the numerical treatment of PDAEs.

We note that the error shown in Figure 4.3 can be explained also from another point of view. For this we insert the expansion (3.4) into the MOL-equation (3.1) to get the DAE for the time dependent coefficient vector w_k

$$Aw_k'(t) + [\lambda_k(h)\, B + C]w_k(t) = \tilde{f}_k(t), \quad k = 1, \ldots, N,$$

where $\tilde{f}_k(t)$ is the generalized Fourier coefficient of $\tilde{F}(t) = F(t) - r(t)$ (analogous to (3.4)). In the example considered here the DAE for $w_k = (w_{k1}, w_{k2})^\top$ can be written

$$\begin{aligned} w_{k1}' + [1 - \lambda_k(h)\,]w_{k1} + w_{k2} &= f_{k1}, \\ w_{k1} + [c - \lambda_k(h)\,]w_{k2} &= f_{k2} \end{aligned}$$

which gives for $w_{k1}(t)$ the simple ODE

$$w_{k1}' + \sigma_k(h)w_{k1} = f_{k1} - \frac{f_{k2}}{c - \lambda_k(h)}$$

where for short $\sigma_k(h) = 1 - \lambda_k(h) + 1/[\lambda_k(h) - c]$. For our example we choose $k = 1$. The stability of w_1 and the numerical method considered here depend on the sign of $\sigma_1(h)$ which is $\sigma_1(h) > 0$ for $h > h_0$ and $\sigma_1(h) < 0$ for $\underline{h}_0 \leq h < h_0$ where $\underline{h}_0$ is some step size near h_0. Furthermore, it holds that

$$\lim_{h \to h_0} |\sigma_1(h)| = \infty.$$

From these relations we get the result that the numerical method is unstable for $\underline{h}_0 \leq h < h_0$. This is reflected in the last figure for the case $h = 0.025$. A similar phenomenon is considered also in [1, 23].

5 Conclusion.

After a definition of two differential indexes ($\nu_{d,t}$, $\nu_{d,x}$) of the PDAE (where we took into account earlier definitions due to Campbell and coworkers) and after a discussion of the problem of consistent IVs and BVs we considered the numerical solution of IBVPs for PDAEs of the type (1.1) by means of two discretization methods (BTCS and Crank–Nicolson scheme). Especially we investigated the convergence of these methods for the case that the time and space step sizes tend to zero. Compared with the convergence of these schemes for regular systems

(1.1) (i.e. when both matrices A and B are regular, in this case $\nu_{d,t} = \nu_{d,x} = 0$) we found as a main result a strong dependence (strong assumptions) of the convergence on the two indexes; see the theorems in Sections 3.2.1 and 3.2.3. Numerical experiments confirm the theoretical results.

What is the reason for the strong requirements for convergence? Why do the schemes not converge (this is also observed in most numerical computations) especially in the case when none of the two indexes is zero and at least one of them is higher? Taking into account the fact that classical numerical methods for ordinary differential equations have trouble with DAEs, it is also expected that numerical methods for partial differential equations may be inadequate for PDAEs. The convergence theorems indeed show that this is true. The main cause for the strong convergence assumptions is that the PDAE of the form (1.1) may be a system of equations of different type. Furthermore, each of them may be require different types of ICs and/or BCs and some of them must be consistent too. And it may be, e.g., that one or more parabolic equations are in the system and their solution components are coupled by algebraic equations being also in the system. These may force the parabolic components to be on a manifold (represented by the algebraic equations, these are some side conditions). Since all these cases must be contained in a general convergence theorem it is expected that it cannot be so nice as in the regular case ($\nu_{d,t} = \nu_{d,x} = 0$).

Although the numerical methods considered in this paper are simple ones the solution of IBVPs for PDAEs seems to be difficult. In such problems it is often useful to take into account special properties (if any) of the system (1.1). Sometimes such special properties become clearer by a transformation of the original dependent variables (the components of u) to new variables.

Acknowledgement.

The authors are indebted to the referees for helpful comments leading to an improved version of the paper.

REFERENCES

1. M. Arnold, *A note on the uniform perturbation index*, Heft 52, Rostocker Math. Kolloquium, 1998.
2. M. Arnold and B. Simeon, *The simulation of pantograph and catenary: A PDAE approach*, Preprint Nr. 1990, Fachbereich Mathematik, TU Darmstadt, 1998.
3. K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*, North-Holland, Amsterdam, 1989.
4. S. L. Campbell and W. Marszalek, *ODE/DAE integrators and MOL problems*, ICIAM 95 Minisymposium on MOL, 1995.
5. S. L. Campbell and W. Marszalek, *The index of an infinite dimensional implicit system*, Math. Modelling Syst., 1:1 (1996), pp. 1–25.
6. S. L. Campbell and W. Marszalek, *ODE/DAE integrators and MOL problems*, ZAMM, 76:S1 (1996), pp. 251–254.
7. S. L. Campbell and W. Marszalek, *DAEs arising from traveling wave solutions of PDEs*, J. Comput. Appl. Math., 82:1–2 (1997), pp. 41–58.

8. C. W. Gear and L. R. Petzold, *ODE methods for the solution of differential/algebraic systems*, SIAM J. Numer. Anal., 21:4 (1984), pp. 716–728.
9. E. Griepentrog and R. März, *Differential-algebraic Equations and their Numerical Treatment*, Teubner-Texte zur Mathematik, Band 88, Leipzig, 1986.
10. P. Grindrod, *The Theory and Applications of Reaction-diffusion Equations*, Clarendon Press, Oxford, 1996.
11. E. Hairer, Ch. Lubich, and M. Roche, *The Numerical Solution of Differential-algebraic Systems by Runge–Kutta Methods*, Vol. 1409, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1989.
12. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II*, Springer-Verlag, Berlin, 1996.
13. A. W. Leung, *Systems of Nonlinear Partial Differential Equations*, Kluwer Academic Publishers, Dordrecht, 1989.
14. Ping Lin, *A sequential regularization method for time-dependent incompressible Navier–Stokes equations*, SIAM J. Numer. Anal., 34:3 (1997), pp. 1051–1071.
15. W. Lucht and K. Strehmel, *Discretization based indices for semilinear partial differential algebraic equations*, Appl. Numer. Math., 28 (1998), pp. 371–386.
16. W. Lucht, K. Strehmel, and C. Eichler-Liebenow, *Linear partial differential algebraic equations, Part I: Indexes, consistent boundary/initial conditions*, Report 17, Fachbereich Mathematik und Informatik, Martin-Luther-Universität, Halle, 1997.
17. W. Lucht, K. Strehmel, and C. Eichler-Liebenow, *Linear partial differential algebraic equations, Part II: Numerical solution*, Report 18, Martin-Luther-Universität Halle, Fachbereich Mathematik und Informatik, 1997.
18. W. Marszalek, *Analysis of partial differential algebraic equations*, PhD thesis, North Carolina State University, Raleigh, NC, 1997.
19. L. R. Petzold, *Differential-algebraic equations are not ODE's*, SIAM J. Sci. Stat. Comput., 3 (1982), pp. 367–384.
20. K. G. Pipilis, *Higher order moving finite elements method for systems described by partial differential-algebraic equations*, PhD thesis, Dept. of Chemical Engineering, Imperial College of Science, Technology and Medicine, London, 1990.
21. M. Sezgin, *Magnetohydrodynamic flow in a rectangular channel*, Internat. J. Numer. Methods Fluids, 7 (1987), pp. 697–718.
22. B. Simeon, *Modelling a flexible slider crank mechanism by a mixed system of DAEs and PDEs*, Math. Modelling Syst., 2(1):1–18, 1996.
23. G. Söderlind, *Remarks on the stability of high-index DAEs with respect to parametric perturbations*, Computing, 49 (1992), pp. 303–314.
24. J. W. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, Springer-Verlag, New York, 1995.
25. W. Walter, *Differential and Integral Inequalities*, Springer-Verlag, New York, 1970.
26. J. Weickert, *Navier–Stokes equations as a differential-algebraic system*, Preprint SFB 393/96-08, Technische Universität Chemnitz–Zwickau, 1996.