General Linear Methods for nonlinear DAEs in Circuit Simulation

Steffen Voigtmann

Humboldt University Berlin, Institute of Mathematics, Unter den Linden 6, D-10099 Berlin, Germany, steffen@math.hu-berlin.de

Summary. The Modified Nodal Analysis leads to differential algebraic equations with properly stated leading terms. In this article a special structure of the DAEs modelling electrical circuits is exploited in order to derive a new decoupling for nonlinear index-2 DAEs. This decoupling procedure leads to a solvability result and is also used to study general linear methods, a class of numerical schemes that covers both Runge-Kutta and linear multistep methods. Convergence for index-2 DAEs is proved.

1 Introduction

When simulating electrical circuits, one is confronted with solving differential algebraic equations (DAEs) of the form

$$A(t)\frac{d}{dt}d\big(x(t),t\big) + b\big(x(t),t\big) = 0, \qquad t \in \mathcal{I}. \tag{1}$$

In case of the charge oriented modified nodal analysis the vector d contains charges and fluxes while x represents all node potentials and currents of voltage defining elements like voltage sources and inductors. Typically the index of (1) does not exceed 2 [Est00, Theorem 3.1.3].

Common circuit simulators like Spice or Titan use the so-called charge oriented approach $A(t)\big(R(t)y(t)\big)' + b\big(x(t),t\big) = 0$, $y(t) - d(x(t),t) = 0$, where charges and fluxes are introduced as a new variable y. Notice that this enlarged system is of the form

$$A(t)\big(D(t)x(t)\big)' + b\big(x(t),t\big) = 0, \qquad t \in \mathcal{I}. \tag{2}$$

Solutions lie in the linear space $C^1_D(\mathcal{I},\mathbb{R}^m) := \big\{ z \in C(\mathcal{I},\mathbb{R}^m) \mid Dz \in C^1(\mathcal{I},\mathbb{R}^n) \big\}$.

Using the concept of the tractability index [Mär03] we study DAEs (2) having index $\mu \in \{1,2\}$. In section 2 we will exploit the specific structure of the MNA equations to derive a decoupling procedure for nonlinear index-2

DAEs. This will enable us to prove existence and uniqueness of solutions. In section 3 we study general linear methods for (2) and prove convergence.

We assume that $\mathcal{I}$ is a compact interval, $\mathcal{D} \subset \mathbb{R}^m$ a domain and that $A : \mathcal{I} \to L(\mathbb{R}^n, \mathbb{R}^m)$, $D : \mathcal{I} \to L(\mathbb{R}^m, \mathbb{R}^n)$ and $b : \mathcal{D}\times\mathcal{I} \to \mathbb{R}^m$ are continuous. Let b'_x exist and be continuous. Finally, the leading term of (2) is supposed to be properly stated, i.e. $\ker A(t) \oplus \operatorname{im} D(t) = \mathbb{R}^n$ for $t \in \mathcal{I}$ and there is a smooth projector function $R \in C^1\big(\mathcal{I}, L(\mathbb{R}^n, \mathbb{R}^n)\big)$ such that $\ker R(t) = \ker A(t)$, $\operatorname{im} R(t) = \operatorname{im} D(t)$ (see [HM04]).

For analysing (2) we introduce the following sequence of matrix functions and subspaces defined pointwise for $t \in \mathcal{I}$ and $x \in \mathcal{D}$.

$$\left.\begin{aligned}
&G_0(t) = A(t)D(t), \qquad B_0(x,t) = b'_x(x,t)\\
&N_0(t) = \ker G_0(t), \qquad S_0(x,t) = \big\{\, z \in \mathbb{R}^m \mid B_0(x,t)z \in \operatorname{im} G_0(t) \,\big\},\\
&Q_0(t) \text{ is a projector onto } N_0(t),\ P_0(t) = I - Q_0(t),\\
&G_1(x,t) = G_0(t) + B_0(x,t)Q_0(t),\\
&N_1(x,t) = \ker G_1(x,t),\\
&S_1(x,t) = \big\{\, z \in \mathbb{R}^m \mid B_0(x,t)z \in \operatorname{im} G_1(x,t) \,\big\}.
\end{aligned}\right\} \quad (3)$$

Let $Q_1(x,t)$ be a projector function onto N_1 and $P_1(x,t) = I - Q_1(x,t)$. Finally denote with $D^-(t)$ the generalised reflexive inverse of $D(t)$ defined by

$$DD^-D = D, \quad D^-DD^- = D^-, \quad D^-D = P_0, \quad DD^- = R.$$

Definition 1. (see [Mär03]) *The DAE* (2) *with a properly stated leading term is regular with tractability index* $\mu \in \{1, 2\}$ *on* $\mathcal{D}\times\mathcal{I}$ *if there is a sequence* (3) *such that for* $(x,t) \in \mathcal{D}\times\mathcal{I}$

1. G_i *has constant rank* $r_i < m$ *for* $0 \le i < \mu$,
2. Q_i *is continuous for* $i = 1, \dots, \mu$, $Q_1(x,t)Q_0(t) = 0$ *and* DN_1, DS_1 *are spanned by continously differentiable basis functions,*
3. $N_{\mu-1} \cap S_{\mu-1} = \{0\}$.

For index-2 DAEs $G_2(x,t) = G_1(x,t) + B_0(x,t)P_0(t)Q_1(x,t)$ remains nonsingular on $\mathcal{D}\times\mathcal{I}$ and we have $N_1(x,t) \oplus S_1(x,t) = \mathbb{R}^m$. In the following we will adopt the convention to choose Q_1 to be the canonical projector onto N_1 along S_1. Due to $N_0 \subset S_1$ the property $Q_1Q_0 = 0$ is then always given.

The space $N_0(t) \cap S_0(x,t) = \operatorname{im} Q_0(t)Q_1(x,t)$ (see [Est00]) is of vital importance as it describes the so-called index-2 components, i.e. the particular part of the solution that can be calculated only by performing an inherent differentiation process. In [Est00] it is shown that the circuit's layout determines this subspace. Thus it is independent of x. We choose a projector function T onto $N_0(t) \cap S_0(x,t)$ that depends on t only. Note that $U = I - T$ satisfies $\ker U(t) = \operatorname{im} Q_0(t)Q_1(x,t)$. T can be chosen such that $TP_0 = 0$ and $P_0T = 0$ for $t \in \mathcal{I}$. Then the following properties are valid: $Q_0T = T = TQ_0$, $Q_1T = 0$, $P_0U = P_0 = UP_0$, $Q_1UQ_0 = 0$.

2 Decoupling nonlinear index-2 equations

Given the conditions stated in [Est00, Corollary 3.2.8] it is well known that for the charge oriented modified nodal analysis the index-2 components Tx enter the equations only in a linear way, i.e. (2) has the structure

$$A(t)\big(D(t)x(t)\big)' + b\big(U(t)x(t),t\big) + \mathfrak{B}(t)T(t)x(t) = 0. \tag{4}$$

This particular form of the DAEs arising in circuit simulation makes it possible to develop a new decoupling procedure for index-2 DAEs. For a given solution $x(\cdot)$ of (4) denote $x_0 = x(t_0)$ and introduce the new variable

$$w = \bar{P}_1 D^-(Dx)' + (Q_0 + \bar{Q}_1)x \tag{5}$$

where $\bar{P}_1(t) = P_1(x_0,t)$ and $\bar{Q}_1(t) = Q_1(x_0,t)$. Here and in the sequel t arguments are generally omitted for better readability. Notice that

$$\bar{Q}_1 w = \bar{Q}_1 x, \qquad Q_0 w = -Q_0\bar{Q}_1 D^-(Dx)' + Q_0 x + Q_0\bar{Q}_1 x,$$
$$D\bar{P}_1 w = D\bar{P}_1 D^-(Dx)'.$$

From $G_1(x_0,\cdot)\bar{Q}_1 = 0$ we infer $A(Dx)' + \mathfrak{B}Tx = (AD+\mathfrak{B}T)w$ and, denoting $u = D\bar{P}_1 x$, we find

$$x = P_0\bar{P}_1 x + P_0\bar{Q}_1 x + Q_0 x = D^-u + (P_0\bar{Q}_1 + Q_0\bar{P}_1)w + Q_0\bar{Q}_1 D^-(Dx)'.$$

The component $Ux = D^-u + (P_0\bar{Q}_1 + UQ_0)w$ can be written in terms of u and w such that (4) is equivalent to

$$F(u,w,t) := \big(AD + \mathfrak{B}T\big)w + b\big(D^-u + (P_0\bar{Q}_1 + UQ_0)w,\cdot\big) = 0. \tag{6}$$

Lemma 1. *Let* (2) *be a regular DAE with index* $\mu \in \{1,2\}$. *Let* $y_0 \in \operatorname{im} D(t_0)$, $(x_0,t_0) \in \mathcal{D}\times\mathcal{I}$ *be given such that* $A(t_0)y_0 + b\big(U(t_0)x_0,t_0\big) + \mathfrak{B}(t_0)T(t_0)x_0 = 0$. *Denote*

$$u_0 = D(t_0)\bar{P}_1(t_0)x_0, \qquad w_0 = \bar{P}_1(t_0)D^-(t_0)y_0 + \big(Q_0+\bar{Q}_1\big)(t_0)x_0$$

and consider F *to be defined on a neighbourhood* $\mathcal{N}_0 \subset \mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}$ *of* (u_0,w_0,t_0). *Then there is a neighbourhood* $\mathcal{N}_1 \subset \mathbb{R}^n\times\mathbb{R}$ *of* (u_0,t_0) *and a continuous mapping* $\mathtt{w} : \mathcal{N}_1 \to \mathbb{R}^m$ *such that* $F\big(u,\mathtt{w}(u,t),t\big) = 0 \quad \forall\ (u,t) \in \mathcal{N}_1$.

Proof. Due to (6) we have $F(u_0,w_0,t_0) = 0$ and

$$F'_w(u,w,\cdot) = AD + \mathfrak{B}T + b'_x\big(D^-u + (P_0\bar{Q}_1 + UQ_0)w,\cdot\big)\,(P_0\bar{Q}_1 + UQ_0)$$

implies that $F'_w(u_0,w_0,t_0) = G_2(x_0,t_0)$ is nonsingular. Thus the assertion follows from the implicit function theorem. □

Notice that the mapping $\mathtt{w}$ from the previous lemma is defined only locally around (u_0, t_0). For simplicity we assume that the interval $\mathcal{I}$ is sufficiently small such that $\mathtt{w}$ is defined for all $t \in \mathcal{I}$.

We arrive at the following representation of the solution:

$$x = D^- u + (Q_0 \bar{P}_1 + P_0 \bar{Q}_1)\mathtt{w}(u,\cdot) + Q_0 \bar{Q}_1 D^- \big(u + D\bar{Q}_1 \mathtt{w}(u,\cdot)\big)'. \qquad (7)$$

The component $u = D\bar{P}_1 x$ satisfies the ordinary differential equation

$$D\bar{P}_1 D^- u' = D\bar{P}_1 \mathtt{w}(u,\cdot) - D\bar{P}_1 D^- \big(D\bar{Q}_1 \mathtt{w}(u,\cdot)\big)'. \qquad (8)$$

As for linear DAEs this equation will be called the inherent regular ODE. Since, by the index-2 condition, $D\bar{P}_1 D^- \in C^1(\mathcal{I}, \mathbb{R}^n)$, we may rewrite (8) as

$$u' = (D\bar{P}_1 D^-)' u + D\bar{P}_1 \mathtt{w}(u,\cdot) + (D\bar{P}_1 D^-)' D\bar{Q}_1 \mathtt{w}(u,\cdot). \qquad (9)$$

Similar to [HM04] we will now study (9) without assuming the existence of a solution.

Theorem 1. *Let the assumptions of Lemma 1 be given. Then*

1. $\operatorname{im} D\bar{P}_1 D^-$ *is a (time-varying) invariant subspace of the inherent ODE* (9), *i.e.* $u(t_0) \in \operatorname{im}\big(D\bar{P}_1 D^-\big)(t_0)$ *implies* $u(t) \in \operatorname{im}\big(D\bar{P}_1 D^-\big)(t)\ \forall\ t \in \mathcal{I}$.
2. *If the subspaces* $\operatorname{im} D\bar{P}_1 D^-$ *and* $\operatorname{im} D\bar{Q}_1 D^-$ *are constant, then* (9) *simplifies to* $u' = D\bar{P}_1 \mathtt{w}(u,\cdot)$, $u(t_0) \in \operatorname{im}\big(D\bar{P}_1 D^-\big)(t_0)$.

Proof. Similar to [HM04, Theorem 2.2]. Replace R by $D\bar{P}_1 D^-$.

Theorem 2. *Let the assumptions of Lemma 1 be given. Assume that the mapping* $t \mapsto D(t)Q_1(t)(x_0,t)\mathtt{w}(u(t),t)$ *is* C^1, *where* u *is the solution of the inherent regular ODE* (9) *with initial value* $u(t_0) = DP_1(x_0,t_0)x_0$. *Then there is a unique solution* $x \in C^1_D(\mathcal{I}, \mathbb{R}^m)$ *of the initial value problem*

$$A(t)\big(Dx\big)'(t) + b\big((Ux)(t),t\big) + \big(\mathfrak{B}T\big)(t)x(t) = 0, \quad DP_1(x_0,t_0)\big(x_0 - x(t_0)\big) = 0.$$

Proof. From Lemma 1 we get the mapping $\mathtt{w}(u,t)$ and thus the solution u of the inherent regular ODE (9). Due to Theorem 1 $u(t) \in \operatorname{im} D(t)P_1(x_0,t)D^-(t)$ holds for all t where u is defined. Then the mapping x as defined in (7) is a solution since

$$\begin{aligned} A(Dx)' + b(Ux,\cdot) + \mathfrak{B}Tx &= A(Dx)' + b(Ux,\cdot) + \mathfrak{B}Tx - F\big(u, \mathtt{w}(u,\cdot),\cdot\big) \\ &= (AD + \mathfrak{B}T)\bar{Q}_1 D^-(Dx)' + AD\bar{P}_1 D^-(Dx)' - AD\bar{P}_1 \mathtt{w}(u,\cdot) = 0. \qquad \square \end{aligned}$$

Remark 1. If (2) was an index-1 DAE, then $\bar{P}_1 = I$, $\bar{Q}_1 = 0$ and all results can be reinterpreted also for index-1 equations. In particular, (5) reduces to $w = D^-(Dx)' + Q_0 x$. This is exactly the mapping studied in [HM04] and the decoupling procedure presented here generalises [HM04].

3 Numerical Integration by General Linear Methods

For the numerical solution of index-2 equations (4) we investigate general linear methods (GLM). This class of methods seems to be very attractive for circuit simulation since there are methods with diagonally implicit structure that have high stage order and a stability behaviour similar to fully implicit Runge-Kutta methods. The diagonally implicit structure yields a very efficient implementation. Examples of such methods are given in [But03, Wri03]. Also, recall that both Runge-Kutta and linear multistep methods can be cast into general linear form.

A GLM is given by a partitioned matrix $M = \left[\begin{array}{c|c} \mathcal{A} & \mathcal{U} \\ \hline \mathcal{B} & \mathcal{V} \end{array}\right] \in L(\mathbb{R}^{s+r}, \mathbb{R}^{s+r})$. We will always assume that $\mathcal{A}$ is nonsingular. The discretisation of the DAE (4) using the general linear method M reads

$$A_{ni}[DX]'_{ni} + b(U_{ni}X_{ni}, t_{ni}) + \mathfrak{B}_{ni}T_{ni}X_{ni} = 0, \quad i = 1, \ldots, s, \tag{10a}$$

$$[DX]_n = h(\mathcal{A} \otimes I_m)[DX]'_n + (\mathcal{U} \otimes I_m)[Dx]^{[n-1]}, \tag{10b}$$

$$[Dx]^{[n]} = h(\mathcal{B} \otimes I_m)[DX]'_n + (\mathcal{V} \otimes I_m)[Dx]^{[n-1]}. \tag{10c}$$

For better readability we will drop the Kronecker products in the sequel. As in the case of linear multistep methods r pieces of information $[Dx]_k^{[n-1]} \in \mathbb{R}^m$, $k = 1, \ldots, r$, are passed on from step to step. These quantities represent some approximations to $D(t)x(t)$ or it's derivative. See [But03] for more details. Observe that only information about the exact solution's D component is carried on. Thus errors in the null-space of D are not propagated.

Similar to Runge-Kutta methods internal stages $X_{ni} \in \mathbb{R}^m$, $i = 1, \ldots, s$, are calculated at intermediate time points $t_{ni} = t_{n-1} + c_i h$ within every step. In (10) we wrote $A_{ni} = A(t_{ni})$ and used similar notations for $\mathfrak{B}$, T and U. For compactness of notation we introduced $X_n = \big(X_{n1}^T \cdots X_{ns}^T\big)^T \in \mathbb{R}^{ms}$ and similarly $[DX]_{ni} = D_{ni}X_{ni}$. The initial input vector $[Dx]^{[0]}$ can be calculated by generalised Runge-Kutta methods [But03].

From [HMT03] it is well known that one should investigate numerically qualified DAEs in order to get good numerical results. We will therefore restrict attention to DAEs where the subspaces $\operatorname{im} D\bar{P}_1 D^-$ and $\operatorname{im} D\bar{Q}_1 D^-$ are constant. Recall from Theorem 1 that the inherent regular ODE (9) now reads

$$u' = D\bar{P}_1 \mathtt{w}(u, \cdot), \qquad\qquad u(t_0) \in \operatorname{im}\big(D\bar{P}_1 D^-\big)(t_0). \tag{11}$$

We want to apply the decoupling procedure to the discretised problem (10). Therefore we need to split the vector $[Dx]^{[n-1]}$ into it's $D\bar{P}_1$ and $D\bar{Q}_1$ parts. If $[Dx]^{[0]}$ was calculated by a generalised Runge-Kutta method, then

$$[Dx]_k^{[0]} = \mathbf{u}_k^{[0]} + \mathbf{v}_k^{[0]} \in \operatorname{im}(D\bar{P}_1)(t_0) \oplus \operatorname{im}(D\bar{Q}_1)(t_0), \quad k = 1, \ldots, r. \tag{12}$$

Splitting the stages $\mathbf{U}_{ni} = D_{ni}\bar{P}_{1,ni}X_{ni}$, $\mathbf{V}_{ni} = D_{ni}\bar{Q}_{1,ni}X_{ni}$ and defining $\mathbf{U}'_n$, $\mathbf{V}'_n$ by $\mathbf{U}_n = h\mathcal{A}\mathbf{U}'_n + \mathcal{U}\mathbf{u}^{[n-1]}$ and $\mathbf{V}_n = h\mathcal{A}\mathbf{V}'_n + \mathcal{U}\mathbf{v}^{[n-1]}$, respectively, we find that (12) is given for every step via

$$\mathbf{u}^{[n]} = \mathcal{B}\mathcal{A}^{-1}\mathbf{U}_n + \mathcal{M}_\infty \mathbf{u}^{[n-1]}, \qquad \mathbf{v}^{[n]} = \mathcal{B}\mathcal{A}^{-1}\mathbf{V}_n + \mathcal{M}_\infty \mathbf{v}^{[n-1]}.$$

Notice that $\mathcal{M}_\infty = \mathcal{V} - \mathcal{B}\mathcal{A}^{-1}\mathcal{U}$ is the methods stability matrix $\mathcal{M}(z)$ evaluated at infinity. This matrix plays a role similar to $R(\infty) = 1 - b^T\mathcal{A}^{-1}\mathbb{1}$ for Runge-Kutta methods.

As in (5) we define $\mathbf{W}_{ni} = P_{1,ni}D_{ni}^{-}[DX]'_{ni} + \big(Q_{0,ni} + \bar{Q}_{1,ni}\big)X_{ni}$ such that

$$X_{ni} = D_{ni}^{-}\mathbf{U}_{ni} + \big(Q_{0,ni}\bar{P}_{1,ni} + P_{0,ni}\bar{Q}_{1,ni}\big)\mathbf{W}_{ni} + Q_{0,ni}\bar{Q}_{1,ni}D_{ni}^{-}[DX]'_{ni}. \tag{13}$$

From (10a) it follows that $F\big(\mathbf{U}_{ni}, \mathbf{W}_{ni}, t_{ni}\big) = 0$. Thus $\mathbf{W}_{ni} = \mathtt{w}(\mathbf{U}_{ni}, t)$ is given by the mapping $\mathtt{w}$ from Lemma 1. Here we have to assume that the stepsize h is small enough to guarantee that $(\mathbf{U}_{ni}, t_{ni})$ remains in the neighbourhood $\mathcal{N}_1$ of (u_0, t_0) where $\mathtt{w}$ is defined.

Theorem 3. *Let M be a stiffly accurate general linear method with nonsingular $\mathcal{A}$. Assume that $\mathcal{V}$ is power bounded and that the spectral radius of $\mathcal{M}_\infty = \mathcal{V} - \mathcal{B}\mathcal{A}^{-1}\mathcal{U}$ is less than 1. Then M is convergent for numerically qualified DAEs (2) with index $\mu \in \{1, 2\}$.*

If M has order p and stage order q for ordinary differential equations, then the order of convergence is (at least) q.

Proof. Since $\mathbf{U}'_{ni} = D_{ni}\bar{P}_{1,ni}D_{ni}^{-}[DX]'_{ni} = D_{ni}P_{1,ni}\mathtt{w}(\mathbf{U}_{ni}, t_{ni})$ holds for numerically qualified DAEs, the decoupling procedure shows that (10) is equivalent to the split system

$$\left.\begin{aligned} \mathbf{U}'_{ni} &= D_{ni}\bar{P}_{1,ni}\mathtt{w}(\mathbf{U}_{ni}, t_{ni}), & \mathbf{V}_{ni} &= D_{ni}\bar{Q}_{1,ni}\mathtt{w}(\mathbf{U}_{ni}, t_{ni}), \\ \mathbf{U}_n &= h\mathcal{A}\,\mathbf{U}'_n + \mathcal{U}\,\mathbf{u}^{[n-1]}, & \mathbf{V}_n &= h\mathcal{A}\,\mathbf{V}'_n + \mathcal{U}\,\mathbf{v}^{[n-1]}, \\ \mathbf{u}^{[n]} &= h\mathcal{B}\,\mathbf{U}'_n + \mathcal{V}\,\mathbf{u}^{[n-1]}, & \mathbf{v}^{[n]} &= h\mathcal{B}\,\mathbf{V}'_n + \mathcal{V}\,\mathbf{v}^{[n-1]}. \end{aligned}\right\} \tag{14}$$

The left hand block of equations is exactly the numerical scheme resulting from applying M directly to the inherent regular ODE (11). Thus ODE theory for general linear methods [But03] yields

$$\mathbf{U}_{ni} = u(t_{n-1} + c_i h) + \mathcal{O}(h^{\tilde{q}+1}), \qquad \mathbf{U}'_{ni} = u'(t_{n-1} + c_i h) + \mathcal{O}(h^{\tilde{q}+1}),$$

where we denoted $\tilde{q} = \min(p-1, q)$. Let u be the inherent ODE's exact solution and introduce $v(t) = D(t)\bar{Q}_1(t)\mathtt{w}\big(u(t), t\big)$. Then

$$\|\mathbf{V}_{ni} - v(t_{ni})\| \le \textstyle\int_0^1 \|\mathtt{w}'_u\big(\tau\mathbf{U}_{ni} + (1-\tau)u(t_{ni}), t_{ni}\big)\|\mathrm{d}\tau\, \|\mathbf{U}_{ni} - u(t_{ni})\|$$

and thus $\mathbf{V}_{ni} = v(t_{n-1} + c_i h) + \mathcal{O}(h^{\tilde{q}+1})$. Denoting exact input quantities by $\hat{\mathbf{v}}^{[n]}$ and using techniques from [HLR89, Theorem 3.1] we obtain the recursion

$$\Delta\mathbf{v}^{[n]} = \mathcal{M}_\infty^n \Delta\mathbf{v}^{[0]} + \sum_{i=1}^{n} \mathcal{M}_\infty^{n-i}\delta_i$$

where $\Delta\mathbf{v}^{[n]} = \mathbf{v}^{[n]} - \hat{\mathbf{v}}^{[n]}$ and $\delta_{ij} = \mathcal{B}\mathcal{A}^{-1}\big(\mathbf{V}_{ij} - v(t_{i-1} + c_j h)\big) = \mathcal{O}(h^{\tilde{q}+1})$. Given that the spectral radius of $\mathcal{M}_\infty$ is less than 1 and $\Delta\mathbf{v}^{[0]} = \mathcal{O}(h^{\tilde{q}+1})$ we find $\Delta\mathbf{v}^{[n]} = \mathcal{O}(h^{\tilde{q}+1})$ and, consequently, $\mathbf{V}'_{ni} = v'(t_{n-1} + c_i h) + \mathcal{O}(h^{\tilde{q}})$.

By assumption the general linear method has stiff accuracy, i.e. the numerical result $x_n = X_{ns}$ coincides with the last stage. Thus we can use (7), (13) to find a bound for the global error

$$\begin{aligned} \|x_n - x(t_n)\| \leq \quad & C_1\|\mathbf{U}_{ns} - u(t_n)\| + C_2\|\mathtt{w}\big(\mathbf{U}_{ns}, t_n\big) - \mathtt{w}\big(u(t_n), t_n\big)\| \\ & + C_3\big(\|\mathbf{U}'_{ns} - u'(t_n)\| + \|\mathbf{V}'_{ns} - v'(t_n)\|\big) = \mathcal{O}(h^{\min(p-1,q)}) \qquad \square \end{aligned}$$

If $p > q$, Theorem 3 predicts order q behaviour for the global error. This agrees with the results in [HLR89]. However, since the proof above is given for general linear methods it not only covers Runge-Kutta methods but also linear multistep methods and even more general methods such as those studied in [Wri03]. In particular, for general linear methods $p = q$ is possible even for diagonally implicit methods. Also the BDF methods have the same property. With a global error of order $\mathcal{O}(h^p)$ they indeed have a higher order than predicted by Theorem 3. From [BCP96] it is known that the k step BDF methods exhibit the true order of convergence for index-2 DAEs only after $k + 1$ steps.

For general linear methods a similar statement holds. For completeness we formulate this result in the following remark.

Remark 2. Let the assumptions of theorem 3 hold. Assume that, in addition, $p = q \geq 2$ and $\mathcal{M}_\infty^{k_0} = 0$. Then M is convergent for (2) with order p after $k_0 + 1$ steps.

Notice that (14) is the general linear method's discretisation of a Hessenberg index-1 DAE

$$u'(t) = f\big(u(t), t\big), \qquad\qquad v(t) = g\big(u(t), t\big) \tag{15}$$

using the direct approach as in [HLR89]. In order to prove the statement of remark 2 one needs to show that $\mathbf{V}'$ is calculated with order $p = q$ when M is applied to (15).

To reach this goal a careful analysis of numerical methods for fully implicit index-1 DAEs can be performed using the language of B-series for differential algebraic equations. In [Kvæ90] Kværnø studied the case of Runge-Kutta methods, but her approach has to be generalised for the much larger class of general linear methods.

The technical effort introducing elementary differentials and B-series for fully implicit index-1 DAEs is far too much to be presented here. We will therefore skip the proof which will be given in [Voi04].

References

[BCP96] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*, Classics in Applied Mathematics, vol. 14, Society for Industrial and Applied Mathematics (SIAM), 1996.

[But03] J. C. Butcher, *Numerical methods for ordinary differential equations*, John Wiley & Sons Ltd., Chichester, 2003.

[Est00] D. Estévez Schwarz, *Consistent initialization for index-2 differential algebraic equations and it's application to circuit simulation*, Ph.D. thesis, Humboldt Universität zu Berlin, 2000.

[HLR89] E. Hairer, C. Lubich, and M. Roche, *The numerical solution of differential-algebraic systems by Runge-Kutta methods*, Lecture Notes in Mathematics, vol. 1409, Springer-Verlag, Berlin, 1989.

[HM04] I. Higueras and R. März, *Differential algebraic equations with properly stated leading terms*, Comp. & Math. with Appl. **48** (2004), no. 1-2, 215–235.

[HMT03] I. Higueras, R. März, and C. Tischendorf, *Stability preserving integration of index-2 DAEs*, Appl. Numer. Math. **45** (2003), no. 2-3, 201–229.

[Kvæ90] A. Kværnø, *Runge-Kutta methods applied to fully implicit differential-algebraic equations of index* 1, Math. Comp. **54** (1990), no. 190, 583–625.

[Mär03] R. März, *Differential algebraic systems with properly stated leading term and MNA equations*, Modeling, simulation, and optimization of integrated circuits (Oberwolfach, 2001), Internat. Ser. Numer. Math., vol. 146, Birkhäuser, Basel, 2003, pp. 135–151.

[Voi04] S. Voigtmann, *General linear methods for nonlinear DAEs with properly stated leading terms*, in preparation, 2004.

[Wri03] W. Wright, *General linear methods with inherent Runge-Kutta stability*, Ph.D. thesis, The University of Auckland, New Zealand, 2003.