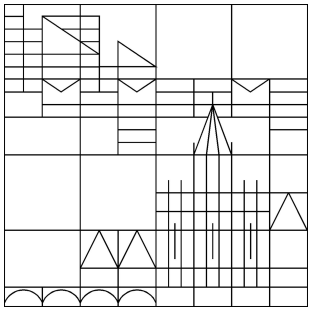

Universität Konstanz

Invariant manifolds in differential algebraic equations of index 3 and in their Runge-Kutta discretizations

Johannes Schropp

Konstanzer Schriften in Mathematik und Informatik

Nr. 226, März 2007

ISSN 1430-3558

Konstanzer Online-Publikations-System (KOPS)
URL: http://www.ub.uni-konstanz.de/kops/volltexte/2007/2386/
URN: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-23865

Universität Konstanz
Fach D 188, 78457 Konstanz, Germany
E-Mail: preprints@informatik.uni-konstanz.de
WWW: http://www.informatik.uni-konstanz.de/Schriften/

Invariant manifolds in differential algebraic equations of index 3 and in their Runge-Kutta discretizations

Johannes Schropp

FB Mathematik und Statistik, Universität Konstanz,

Universitätsstr. 10, D-78464 Konstanz, Germany

E-mail: johannes.schropp@uni-konstanz.de

Abstract

In the present paper we analyze the geometric properties of projected Runge-Kutta methods when applied to index 3 differential algebraic equations in Hessenberg form. These methods admit the integration of index 3 DAEs without any drift effects. We show that the phase portrait is well reproduced in its relationship between space and control variables.

Keywords. differential algebraic equations, Runge-Kutta methods, invariant manifolds, hidden constraints.

AMS classification. 34C05, 34C30, 65L05

1 Introduction

Differential-algebraic equations (DAEs) of index 3 arise naturally in many applications. Mechanical multibody systems in state formulation, electrical curcuits or, in general, second order ordinary differential equations (ODEs) with constraints provide as sources for wide classes of examples.

Higher index DAEs present numerical and analytical difficulties which do not occur in ODEs. Here we investigate the geometric properties of the projected index 3 Runge-Kutta methods. These numerical schemes are introduced in Schropp [14] as modified Runge-Kutta methods producing no drift effects when applied to index 3 DAEs. Of our particular interest is the relation between the discrete state and control variables as it has been worked out for the Runge-Kutta type methods applied to index 2 DAEs in Schropp [13]. To be more precise, the link between the state and control variables of an index 3 DAE in Hessenberg form is decribed by an algebraic relationship, that is, the phase space of the DAE is an

invariant submanifold in the state times control space. We show that the projected index 3 Runge-Kutta discretization scheme possesses an attractive, invariant submanifold too which is located nearby to the original one within the stage order minus one of the underlying Runge-Kutta method. Moreover, the attractivity is governd by the absolute value of the Runge-Kutta stability function at infinity.
Our main tools are embedding and invariant manifold techniques. We embed the original index 3 DAE into a DAE of the same index such that the corresponding index 0 ODE in the state variables admits a representation as dynamical system on the state space. Then we mimic that approach for the discrete Runge-Kutta type dynamics. An important feature of that approach is that it makes general ODE geometric phase portrait discretization results near equilibria, periodic orbits or general attracting sets (see, e.g., Beyn [2], [3], Garay [6] and Kloeden, Lorenz [11]) available for index 3 DAEs.

2 Projected Runge-Kutta methods

We consider the DAE

$$\begin{aligned} \dot{u} &= f(u,v), \quad u(0) = u_0, \\ \dot{v} &= k(u,v,\lambda), \quad v(0) = v_0, \\ 0 &= g(u), \quad \lambda(0) = \lambda_0, \end{aligned} \tag{2.1}$$

$u \in \mathbb{R}^N$, $v \in \mathbb{R}^M$ and $\lambda \in \mathbb{R}^l$ in Hessenberg form. Let C_b^ν denote the space of functions of class C^ν with bounded derivatives up to order ν. We make the following assumptions.

(A1) $f \in C_b^{\nu+1}(\mathbb{R}^{N+M}, \mathbb{R}^N)$, $k \in C_b^\nu(\mathbb{R}^{N+M+l}, \mathbb{R}^M)$, $g \in C_b^{\nu+2}(\mathbb{R}^N, \mathbb{R}^l)$ for ν sufficiently large.

(A2) There is a C_b^ν-function ψ_0 satisfying

$$\begin{aligned} Dg^2(u)(f(u,v), f(u,v)) + Dg(u)\frac{\partial f}{\partial u}(u,v)f(u,v) \\ +Dg(u)\frac{\partial f}{\partial v}(u,v)k(u,v,\psi_0(u,v)) &= 0 \end{aligned}$$

for $(u,v) \in D_\tau := \{(u,v) \in \mathbb{R}^{N+M} \mid \max(\| g(u) \|, \ \| Dg(u)f(u,v) \|) < \tau\}$, $\tau > 0$.

(A3) $Dg(u)\frac{\partial f}{\partial v}(u,v)\frac{\partial k}{\partial \lambda}(u,v,\psi_0(u,v))$ is invertible for $(u,v) \in D_\tau$ and the inverse has bounded norm.

In particular, problem (2.1) is of index 3 and consistent initial values (u_0, v_0, λ_0) for (2.1) must satisfy the relations

$$g(u_0) = 0,$$

$$
\begin{aligned}
Dg(u_0)f(u_0,v_0) &= 0, \qquad (2.2)\\
Dg^2(u_0)(f(u_0,v_0),f(u_0,v_0)) + Dg(u_0)\frac{\partial f}{\partial u}(u_0,v_0)f(u_0,v_0) & \\
+Dg(u_0)\frac{\partial f}{\partial v}(u_0,v_0)k(u_0,v_0,\lambda_0) &= 0.
\end{aligned}
$$

Additionally, (A3) implies that $Dg(u)$ and $Dg(u)\frac{\partial f}{\partial v}(u,v)$ are of full rank. Thus the solutions of the equations $g(u)=0$, $Dg(u)f(u,v)=0$ define the $(N+M-2l)$-dimensional submanifold

$$
S_0 := \{(u,v)\in\mathbb{R}^{N+M} \mid g(u)=0,\ Dg(u)f(u,v)=0\} \qquad (2.3)
$$

of $\mathbb{R}^{N+M}$ and the underlying index 0 ODE reads

$$
\begin{aligned}
\dot{u} &= f(u,v), \quad u(0)=u_0,\\
\dot{v} &= k(u,v,\psi_0(u,v)), \quad v(0)=v_0, \quad (u_0,v_0)\in S_0 \qquad (2.4)
\end{aligned}
$$

(for an illustration of S_0 and the dynamics on it, see Hairer, Wanner [8], p.458). Throughout that paper we assume S_0 to be nonempty. We denote the solution flow of (2.4) with $(\bar{u}(t,u_0,v_0),\bar{v}(t,u_0,v_0))$, $(u_0,v_0)\in S_0$. Then, (A2) implies the flow

$$
\begin{aligned}
\varphi_t(u_0,v_0) &= (\bar{u}(t,u_0,v_0),\bar{v}(t,u_0,v_0),\bar{\lambda}(t,u_0,v_0)),\\
\bar{\lambda}(t,u_0,v_0) &= \psi_0(\bar{u}(t,u_0,v_0),\bar{v}(t,u_0,v_0)), \quad (u_0,v_0)\in S_0
\end{aligned}
$$

for equation (2.1). This means that the manifold

$$
M_0 = \{(u,v,\lambda)\in\mathbb{R}^{N+M+l} \mid (u,v)\in S_0,\ \lambda=\psi_0(u,v)\}
$$

is the phase space of the solution flow of (2.1).

We are interested in the qualitative, geometric features of s-stage one-step Runge-Kutta type methods with Butcher tableau

$$
\begin{array}{c|c} c & A \\ \hline & b^T \end{array}, \quad A=(a_{ij})_{1\le i,j\le s}\in\mathbb{R}^{s,s},\ b,c\in\mathbb{R}^s \qquad (2.5)
$$

and constant step size Δt when applied to (2.1). It is well known that numerical schemes applied to (2.1) in general produces drift terms, i.e., the numerical solution does not satisfy the index 1 and 2 constraints $g(u)=0$ and $Dg(u)f(u,v)=0$. We are interested in Runge-Kutta type methods whose iterates retain these constraints exactly. To that purpose we combine classical Runge-Kutta methods with projection techniques. This approach was first proposed by Ascher and Petzold [1] for index 2 DAEs in Hessenberg form. A generalization to the index 3 case was given by Schropp [14].
For the Butcher tableau of the underlying Runge-Kutta method we impose the conditions:

(B1) The Runge-Kutta matrix A is invertible.

(B2) $R(\infty) = 1 - b^T A^{-1}\mathbb{I}$, $\mathbb{I} = (1, \ldots, 1)$ satisfies $|R(\infty)| < 1$.

(B3) The method is of classical order p and possesses stage order q with $p \geq q + 1$ and $q \geq 2$.

A definition of the stage order can be found in Hairer and Wanner [8], p. 226. Applied to equation (2.1) a projected Runge-Kutta method with step size Δt is a combination of a Runge-Kutta step with a projection. We denote the projected Runge-Kutta iterates at time $t_n = n\Delta t$ shortly by (u_n, v_n, λ_n) or more precisely

$$\Phi_{n\Delta t}^{\Delta t}(u_0, v_0, \lambda_0) \;=\; (\hat{u}_{\Delta t}, \hat{v}_{\Delta t}, \hat{\lambda}_{\Delta t})(n\Delta t, u_0, v_0, \lambda_0),$$

if the dependence of the initial value (u_0, v_0, λ_0) and the step size Δt is emphasized. The Runge-Kutta step has the form

$$\begin{aligned} u_{R,n+1} &= u_n + \Delta t(b^T \otimes I)\bar{f}(U^n, V^n), \\ v_{R,n+1} &= v_n + \Delta t(b^T \otimes I)\bar{k}(U^n, V^n, \Lambda^n), \\ \lambda_{n+1} &= (1 - b^T A^{-1}\mathbb{I})\lambda_n + (b^T A^{-1} \otimes I)\Lambda^n \end{aligned} \tag{2.6}$$

where $U^n = (U_1^n, \ldots, U_s^n) \in \mathbb{R}^{Ns}$, $V^n = (V_1^n, \ldots, V_s^n) \in \mathbb{R}^{Ms}$, $\Lambda^n = (\Lambda_1^n, \ldots, \Lambda_s^n) \in \mathbb{R}^{ls}$ denote the solution of the algebraic system

$$\begin{aligned} U - (\mathbb{I} \otimes u_n) &= \Delta t(A \otimes I)\bar{f}(U, V), \\ V - (\mathbb{I} \otimes v_n) &= \Delta t(A \otimes I)\bar{k}(U, V, \Lambda), \\ 0 &= \bar{g}(U). \end{aligned} \tag{2.7}$$

Here the functions $\bar{f}, \bar{k}, \bar{g}$ stand for $\bar{f}(U^n, \Lambda^n) = (f(U_1^n, \Lambda_1^n), \ldots, f(U_s^n, \Lambda_s^n))$, $\bar{k}(U^n, V^n, \Lambda^n) = (k(U_1^n, V_1^n, \Lambda_1^n), \ldots, k(U_s^n, V_s^n, \Lambda_s^n))$, $\bar{g}(U^n) = (g(U_1^n), \ldots, g(U_s^n))$.
Finally, the projection step

$$\begin{aligned} u_{n+1} &= u_{R,n+1} + \frac{\partial f}{\partial v}(u_{n+1}, v_{n+1})\frac{\partial k}{\partial \lambda}(u_{n+1}, v_{n+1}, \lambda_{n+1})\mu_1, \\ v_{n+1} &= v_{R,n+1} + \frac{\partial k}{\partial \lambda}(u_{n+1}, v_{n+1}, \lambda_{n+1})\mu_2, \\ 0 &= g(u_{n+1}), \\ 0 &= Dg(u_{n+1})f(u_{n+1}, v_{n+1}) \end{aligned} \tag{2.8}$$

determines u_{n+1} and v_{n+1}. In (2.8) the variables μ_i, $i = 1, 2$ are needed for the projection only. An obvious advantage of the projection step is that, by construction, the numerical iterates (u_n, v_n) live in the continuous state space S_0.

A Runge-Kutta method satisfying $a_{sj} = b_j$, $j = 1, \ldots, s$ is called stiffly accurate. Stiffly accurate Runge-Kutta solutions satisfy the index 1 constraint $g(u) = 0$ (see, e.g., Hairer,

Lubich and Roche [7]). In this case we obtain $u_{n+1} = u_{R,n+1},\ \mu_1 = 0$ in (2.8) and the projection step can be reduced to

$$\begin{aligned} v_{n+1} &= v_{R,n+1} + \frac{\partial k}{\partial \lambda}(u_{n+1}, v_{n+1}, \lambda_{n+1})\mu_2, \\ 0 &= Dg(u_{n+1}) f(u_{n+1}, v_{n+1}). \end{aligned} \tag{2.9}$$

Assuming (A1)-(A3) for the DAE (2.1) and (B1)-(B3) for the underlying Runge-Kutta scheme convergence order q is shown for the $u-$ and $v-$component and order $q-1$ for the λ-component (see Th. 2.1, Schropp [14]). For a special subclass of Runge-Kutta methods, the so called stiffly accurate collocation methods with the projection (2.9), convergence of order p in the $u-$ and $v-$variables is shown (see Jay [9, 10] for details).

In the continuous case the algebraic relation $\lambda = \psi_0(u, v)$ determines the link between the state variables (u, v) and the control variable λ. In the discrete dynamics the following characterization holds.

Theorem 2.1 *Consider the DAE (2.1) and assume (A1)-(A3). Moreover, let (u_n, v_n, λ_n) denote the sequences generated with a projected Runge-Kutta method satisfying (B1)-(B3), when applied to (2.1) with consistent initial values (u_0, v_0, λ_0).*
Then there exists a positive constant Δt_0 such that for $0 < \Delta t < \Delta t_0$ the iterates (u_n, v_n, λ_n) exist for $n \in \mathbb{N}$. Moreover, for $\Delta t \in]0, \Delta t_0]$ there is a C_b^ν-function $\psi_{0,\Delta t} : S_0 \to \mathbb{R}^l$, $S_0 = \{(u, v) \in \mathbb{R}^{N+M} \mid g(u) = 0,\ Dg(u)f(u, v) = 0\}$ satisfying the following assertions.

i) The set $M_{0,\Delta t} = \{(u, v, \lambda) \in D_\tau \times \mathbb{R}^l \mid g(u) = 0,\ Dg(u)f(u, v) = 0,\ \lambda = \psi_{0,\Delta t}(u, v)\}$ is invariant for the projected Runge-Kutta map (2.6)-(2.8).

ii) The manifold $M_{0,\Delta t}$ is uniformly attractive with the constant $\chi_{\Delta t} = |\, R(\infty)\,| + O(\Delta t^q)$, that is,

$$\| \lambda_{n+1} - \psi_{0,\Delta t}(u_{n+1}, v_{n+1}) \| \ \leq\ \chi_{\Delta t} \, \| \lambda_n - \psi_{0,\Delta t}(u_n, v_n) \|$$

for every discrete evolution (u_n, v_n, λ_n) starting sufficiently close to $M_{0,\Delta t}$.

iii) For every initial value (u_0, v_0, λ_0) with $\| \lambda_0 - \psi_{0,\Delta t}(u_0, v_0) \|$ sufficiently small there is $(\tilde{u}_0, \tilde{v}_0, \tilde{\lambda}_0) \in M_{0,\Delta t}$ and $c, \hat{c} > 0$ such that the corresponding evolutions (u_n, v_n, λ_n) and $(\tilde{u}_n, \tilde{v}_n, \tilde{\lambda}_n)$ satisfy

$$\begin{aligned} \| (u_i, v_i) - (\tilde{u}_i, \tilde{v}_i) \| &\leq c\chi_{\Delta t}^i \, \| \lambda_0 - \psi_{0,\Delta t}(u_0, v_0) \|, \quad i \in \mathbb{N}, \\ \| \lambda_i - \tilde{\lambda}_i \| &\leq \hat{c}\chi_{\Delta t}^i \, \| \lambda_0 - \psi_{0,\Delta t}(u_0, v_0) \|, \quad i \in \mathbb{N}. \end{aligned}$$

iv) $\| \psi_0(u, v) - \psi_{0,\Delta t}(u, v) \| \leq C\Delta t^{q-1}$ for $(u, v) \in S_0$.

Theorem 2.1 confirms that the set $M_{0,\Delta t}$ can be viewed as a discrete analogue of the continuous phase space M_0. But in general the discrete numerical scheme is initialized at a consistent initial value $(u_0, v_0, \lambda_0) \in M_0$. Thus we have to discuss the attractivity properties of $M_{0,\Delta t}$ in more detail.
A numerical scheme is infinitely attractive, if $\chi_{\Delta t} = 0$. This implies $(u_n, v_n, \lambda_n) \in M_{0,\Delta t}$ for $n \geq 1$. Numerical methods with $R(\infty) = 0$ are highly attractive. In this case we have $\chi_{\Delta t} = O(\Delta t^q)$ and, thus, the iterates (u_n, v_n, λ_n), $n \geq 1$ for small step sizes Δt are indistinguishable from their in phase counterparts $(\tilde{u}_n, \tilde{v}_n, \tilde{\lambda}_n)$ on $M_{0,\Delta t}$. The relation $R(\infty) = 0$ is valid, i.e., for stiffly accurate Runge-Kutta methods.
Finally, an application of the implicit function theorem to the index 3 constraint equation (2.2) in combination with condition iv) shows that the numerical iterates (u_n, v_n, λ_n) satisfy the index 3 constraint (last equation in formula (2.2)) up to a deviation of order $O(\Delta t^{q-1})$ for $n \in \mathbb{N}$.

3 Embedding techniques for index 3 DAEs

In this section we give the first part of the proof of Theorem 2.1. Theorem 2.1 is the index 3 analogue of Theorem 2.1 in Schropp [13] where the corresponding statement was shown for index 2 DAEs. We will prove Theorem 2.1 generalizing the ideas displayed there. The main technical tool was the embedding of the index 0 form of the original DAE in an open neighbourhood of S_0 in the state space.
Assuming (A1)-(A3), an embedding of (2.4) into D_{τ_0}, $\tau_0 \in]0, \tau]$ sufficiently small can be established as follows. Let

$$C := \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \in \mathbb{R}^{2l,2l},\ C_{ij} \in \mathbb{R}^{l,l},\ 1 \leq i, j \leq 2$$

and consider the problem

$$\begin{aligned} \dot{u} &= f(u, v), \quad u(0) = u_0, \\ \dot{w} &= -(C_{11} w + C_{12} z), \quad w(0) = w_0, \\ \dot{v} &= k(u, v, \lambda), \quad v(0) = v_0 \\ \dot{z} &= -(C_{21} w + C_{22} z), \quad z(0) = z_0, \\ 0 &= g(u) - w, \quad \lambda(0) = \lambda_0. \end{aligned} \tag{3.1}$$

We suppose $\mu_2(-C) \leq -\eta$, $\eta > 0$ and C_{12} invertible (e.g., set $C_{11} = C_{22} = 2I$, $C_{12} = C_{21} = -I$). Here $\mu_2(C)$ stands for the logarithmic norm of a matrix C (see Dekker, Verwer [5], p. 27 for definition). Introducing the variables $\tilde{u} = (u, w)$, $\tilde{v} = (v, z)$, $\tilde{\lambda} = \lambda$ and the functions

$$\tilde{f}(\tilde{u}, \tilde{v}) = \begin{pmatrix} f(u, v) \\ -(C_{11} w + C_{12} z) \end{pmatrix},$$

$$\begin{aligned}\tilde{k}(\tilde{u},\tilde{v},\tilde{\lambda}) &= \begin{pmatrix} k(u,v,\lambda) \\ -(C_{21}w + C_{22}z) \end{pmatrix} \\ \tilde{g}(\tilde{u}) &= g(u) - w\end{aligned}$$

equation (3.1) takes the formal Hessenberg index 3 form (cf. (2.1))

$$\begin{aligned}\dot{\tilde{u}} &= \tilde{f}(\tilde{u},\tilde{v}), \\ \dot{\tilde{v}} &= \tilde{k}(\tilde{u},\tilde{v},\tilde{\lambda}), \\ 0 &= \tilde{g}(\tilde{u}).\end{aligned} \tag{3.2}$$

Our first aim here to show is that (A1)-(A3) for (2.1) imply the following assertions for the DAE (3.1).

(A1') $f \in C_b^{\nu+1}(\mathbb{R}^{N+M},\mathbb{R}^N)$, $k \in C_b^{\nu}(\mathbb{R}^{N+M+l},\mathbb{R}^M)$, $g \in C_b^{\nu+2}(\mathbb{R}^N,\mathbb{R}^l)$, $C \in C_b^{\nu}(\mathbb{R}^N,\mathbb{R}^{2l,2l})$ for ν sufficiently large.

(A2') There is a C_b^{ν}-function ψ_e satisfying

$$\begin{aligned}Dg^2(u)(f(u,v),f(u,v)) + Dg(u)\frac{\partial f}{\partial u}(u,v)f(u,v) \\ +Dg(u)\frac{\partial f}{\partial v}(u,v)k(u,v,\psi_e(u,v,w,z)) \\ -C_{11}(C_{11}w + C_{12}z) - C_{12}(C_{21}w + C_{22}z) &= 0\end{aligned}$$

for $(u,v) \in D_{\tau_0}$, $\|\,(w,z)\,\| < \tau_1$.

(A3') $Dg(u)\frac{\partial f}{\partial v}(u,v)\frac{\partial k}{\partial \lambda}(u,v,\psi_e(u,v,w,z))$ is invertible for $(u,v) \in D_{\tau_0}$, $\|\,(w,z)\,\|_2 < \tau_1$ and the inverse has bounded norm.

(A1')-(A3') correspond to (A1)-(A3) with the exception that the vectorfield on the right-hand side of the DAE (3.1) is not bounded.
We construct the function ψ_e by applying Lemma 4.2, Beyn, Schropp [4] to the equation

$$\begin{aligned}F_{u,w,v,z}(\zeta) &:= Dg^2(u)(f(u,v),f(u,v)) + Dg(u)\frac{\partial f}{\partial u}(u,v)f(u,v) \\ &\quad +Dg(u)\frac{\partial f}{\partial v}(u,v)k(u,v,\psi_0(u,v)+\zeta) \\ &\quad -C_{11}(C_{11}w + C_{12}z) - C_{12}(C_{21}w + C_{22}z) \\ &= 0.\end{aligned} \tag{3.3}$$

Lemma 4.2 in Schropp [13] ensures exactly one solution $\hat{\zeta}_{u,w,v,z}$ of the equation (3.3) in $B_{r_0}(0)$, that is, $\psi_e(u,w,v,z) := \psi_0(u,v) + \hat{\zeta}_{u,w,v,z}$ satisfies the equation in (A2') for $(u,v) \in D_{\tau_0}$, $\|\,(w,z)\,\|_2 \leq \tau_1$, $\tau_1 > 0$ sufficiently small and an implicit function argument guarantees the

smoothness of ψ_e. The reader may notice that we have $\psi_0(u,v) = \psi_e(u,0,v,0)$ by uniqueness.
Then, an application of the Banach lemma with $Dg(u)\frac{\partial f}{\partial v}(u,v)\frac{\partial k}{\partial \lambda}(u,v,\psi_0(u,v))$ and the perturbation $Dg(u)\frac{\partial f}{\partial v}(u,v)\frac{\partial k}{\partial \lambda}(u,v,\psi_e(u,w,v,z))$ shows that the perturbed matrix is invertible, the inverse possesses a bounded norm and (A1')-(A3') are verified.
(A1')-(A3') imply that equation (3.1) is of index 3. Consistent initial values must satisfy

$$\begin{aligned} g(u_0) - w_0 &= 0, \\ Dg(u_0)f(u_0,v_0) + C_{11}w_0 + C_{12}z_0 &= 0, \\ Dg^2(u_0)(f(u_0,v_0), f(u_0,v_0)) + Dg(u_0)\frac{\partial f}{\partial u}(u_0,v_0)f(u_0,v_0) & \\ + Dg(u_0)\frac{\partial f}{\partial v}(u_0,v_0)k(u_0,v_0,\psi_e(u_0,w_0,v_0,z_0)) & \\ - C_{11}(C_{11}w_0 + C_{12}z_0) - C_{12}(C_{21}w_0 + C_{22}z_0) &= 0. \end{aligned}$$

The solution flow of (3.1) has the form

$$\varphi_e(t,u_0,v_0) = (\bar{u},\bar{w},\bar{v},\bar{z},\bar{\lambda})(t,u_0,v_0), \quad (u_0,v_0) \in D_{\tau_0}$$

with the relations

$$\begin{aligned} \bar{w}(t,u_0,v_0) &= g(\bar{u}(t,u_0,v_0)), \\ \bar{z}(t,u_0,v_0) &= h((\bar{u},\bar{v})(t,u_0,v_0)), \\ h(u,v) &= -(C_{12})^{-1}(Dg(u)f(u,v) + C_{11}g(u)), \\ \bar{\lambda}(t,u_0,v_0) &= \psi_e((\bar{u},\bar{w},\bar{v},\bar{z})(t,u_0,v_0)). \end{aligned}$$

Moreover,

$$\begin{aligned} M_e := \{&(u,w,v,z,\lambda) \in D_{\tau_0} \times \mathbb{R}^{3l} \mid g(u) - w = 0, \\ &Dg(u)f(u,v) + C_{11}w + C_{12}z = 0,\ \lambda = \psi_e(u,w,v,z)\} \end{aligned}$$

is the phase space of equation (3.1). Applying the theory of logarithmic norms (see, e.g., Dekker, Verwer [5], Th. 1.5.2) we obtain that

$$\| (\bar{w},\bar{z})(t,u_0,v_0) \|_2 \leq \| (\bar{w},\bar{z})(0,u_0,v_0) \|_2 \ \exp(-\eta t) \tag{3.4}$$

is valid for the (w,z)-component of every solution of (3.1). In particular, in the case $w(0) = w_0 = 0$, $z(0) = z_0 = 0$ problem (3.1) reduces to (2.1). After eliminating the (w,z)-variables with the setting $\hat{\psi}_e(u,v) = \psi_e(u,g(u),v,h(u,v))$ the (u,v)-component of the underlying index 0 ODE of (3.1) reads

$$\begin{aligned} \dot{u} &= f(u,v), \quad u(0) = u_0, \\ \dot{v} &= k(u,v,\hat{\psi}_e(u,v)), \quad v(0) = v_0, \quad (u_0,v_0) \in D_{\tau_0} \subset \mathbb{R}^{N+M} \text{ open.} \end{aligned} \tag{3.5}$$

Next we summarize the qualitative properties of the solutions of (3.1).

Lemma 3.1 *Consider equation (3.1) on the phase space M_e, and let (A1)-(A3) hold. Then every solution of (3.1) with consistent initial values $(u_0, w_0, v_0, z_0, \lambda_0)$ exists for all $t \geq 0$. Moreover, $M_{0,e} = \{(u, w, v, z, \lambda),\ (u, v) \in D_{\tau_0},\ (w, z, \lambda) \in \mathbb{R}^{3l} \mid\ g(u) = w = 0,\ h(u, v) = z = 0,\ \lambda = \hat{\psi}_e(u, v)\}$ is an invariant and globally attractive subset of the phase space M_e.*

Proof: The proof of Lemma 3.2 is a direct consequence of (3.4) and the fact that f, g, k, C are C_b^ν-functions.

We are interested in the behaviour of s-stage projected Runge-Kutta methods when applied to the embedded DAE (3.1). The reader should keep in mind that (3.1) is an index 3 DAE in Hessenberg form (see (3.2)). The numerical scheme has the form

$$\begin{aligned}
u_{R,n+1} &= u_n + \Delta t(b^T \otimes I)\bar{f}(U^n, V^n),\\
w_{n+1} &= w_n - \Delta t(b^T \otimes I)((C_{11} \otimes I)W^n + (C_{12} \otimes I)Z^n),\\
v_{R,n+1} &= v_n + \Delta t(b^T \otimes I)\bar{k}(U^n, V^n, \Lambda^n),\\
z_{n+1} &= z_n - \Delta t(b^T \otimes I)((C_{21} \otimes I)W^n + (C_{22} \otimes I)Z^n),\\
\lambda_{n+1} &= (1 - b^T A^{-1}\mathbb{I})\lambda_n + (b^T A^{-1} \otimes I)\Lambda^n
\end{aligned} \tag{3.6}$$

where $U^n \in \mathbb{R}^{Ns}$, $W^n \in \mathbb{R}^{ls}$, $V^n \in \mathbb{R}^{Ms}$, $Z^n \in \mathbb{R}^{ls}$, $\Lambda^n \in \mathbb{R}^{ls}$ denote the solution of the algebraic system

$$\begin{aligned}
U - (\mathbb{I} \otimes u_n) &= \Delta t(A \otimes I)\bar{f}(U, V),\\
W - (\mathbb{I} \otimes w_n) &= -\Delta t(A \otimes I)((C_{11} \otimes I)W + (C_{12} \otimes I)Z),\\
V - (\mathbb{I} \otimes v_n) &= \Delta t(A \otimes I)\bar{k}(U, V, \Lambda),\\
Z - (\mathbb{I} \otimes z_n) &= -\Delta t(A \otimes I)((C_{21} \otimes I)W + (C_{22} \otimes I)Z),\\
0 &= \bar{g}(U) - W.
\end{aligned} \tag{3.7}$$

Finally, the projection step for the embedded DAE reads

$$\begin{aligned}
u_{n+1} &= u_{R,n+1} + \frac{\partial f}{\partial v}(u_{n+1}, v_{n+1})\frac{\partial k}{\partial \lambda}(u_{n+1}, v_{n+1}, \lambda_{n+1})\mu_1,\\
v_{n+1} &= v_{R,n+1} + \frac{\partial k}{\partial \lambda}(u_{n+1}, v_{n+1}, \lambda_{n+1})\mu_2,\\
0 &= g(u_{n+1}) - w_{n+1},\\
0 &= Dg(u_{n+1})f(u_{n+1}, v_{n+1}) + C_{11}w_{n+1} + C_{12}z_{n+1}.
\end{aligned} \tag{3.8}$$

Now we present existence results for the projected Runge-Kutta schemes applied to the embedded DAE (3.1) in the concept of vectornorms (see, e.g., Beyn, Schropp [4], section 4 for a definition).

Lemma 3.2 *Let the assumptions of Theorem 2.1 hold and* $(u_0, w_0, v_0, z_0, \lambda_0)$ *be a consistent initial value for the DAE (3.1).*
Then for $0 < \Delta t \leq \Delta t_0$, $\Delta t_0 > 0$ *sufficiently small the projected Runge-Kutta iterates* $(u_n, w_n, v_n, z_n, \lambda_n)$ *exist for* $n \in \mathbb{N}$. *For the stages* (U, W, V, Z, Λ) *of the projected Runge-Kutta dynamics we have with* $s_0(u, w, v, z) := (\mathbb{I} \otimes u, \mathbb{I} \otimes w, \mathbb{I} \otimes v, \mathbb{I} \otimes z, \mathbb{I} \otimes \psi_e(u, w, v, z))$, $(u, v) \in D_{\tau_0}$, $\| (w, z) \| < \tau_1$ *the inequality*

$$| (U, W, V, Z, \Lambda) - s_0(u, w, v, z) | \quad \leq \quad O(\Delta t) \, (1, 1, 1, 1, 1) \tag{3.9}$$

Moreover, the functions $\mu_i = \mu_i(\Delta t, u, w, v, z, \lambda)$, $i = 1, 2$ *from the projection step (2.8) satisfy*

$$\| \mu_i \| \quad = \quad O(\Delta t^q) \tag{3.10}$$

for (u, w, v, z) *satisfying* $g(u) - w = 0$, $Dg(u) f(u, v) + C_{11} w + C_{12} z = 0$ *and* $\| \lambda - \psi_e(u, w, v, z) \|$ *sufficiently small.*

Proof: The DAE (3.1) can be written in Hessenberg index 3 form (see (3.2)) and satisfies (A1')-(A3'). A detailed inspection of the proof of Lemma 3.2 in Schropp [14] shows that that it works with (A1')-(A3') instead of (A1)-(A3) too. Thus, Lemma 3.2 is a direct consequence of Lemma 3.2 in Schropp [14].

4 Embedded index 3 systems under discretization

In this section we complete the proof of Theorem 2.1. We show the assertions i)-iv) of Theorem 2.1 simultaneously by applying Theorem 5 of Nipp, Stoffer [12] on the discrete Runge-Kutta dynamics of the DAE (3.1). First, we follow the lines of Schropp [13], section 4. We use the formulae (3.8), (3.10) and Lemma 3.2 to eliminate the variables w, z and obtain

$$\begin{aligned}
u_{n+1} &= u_n + \Delta t [(b^T \otimes I) \bar{f}(U(\Delta t, u_n, v_n), V(\Delta t, u_n, v_n)) \\
&\quad + \Delta t^{q-1} \hat{f}_1(\Delta t, u_n, v_n, \lambda_n)] \\
v_{n+1} &= v_n + \Delta t [(b^T \otimes I) \bar{k}(U(\Delta t, u_n, v_n), V(\Delta t, u_n, v_n), \Lambda(\Delta t, u_n, v_n)) \\
&\quad + \Delta t^{q-1} \hat{f}_2(\Delta t, u_n, v_n, \lambda_n)] \\
\lambda_{n+1} &= R(\infty) \lambda_n + (b^T A^{-1} \otimes I) \Lambda(\Delta t, u_n, v_n)
\end{aligned} \tag{4.1}$$

with smooth and bounded functions $\hat{f}_i$, $i = 1, 2$.
Introducing $\eta_n := \lambda_n - \hat{\psi}_e(u_n, v_n) = \lambda_n - \psi_e(u_n, g(u_n), v_n, h(u_n, v_n))$ and rewriting (4.1) yields

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} \quad = \quad \begin{pmatrix} u_n \\ v_n \end{pmatrix} + \Delta t (b^T \otimes I) \begin{pmatrix} \bar{f}((U, V)(\Delta t, u_n, v_n)) \\ \bar{k}((U, V, \Lambda)(\Delta t, u_n, v_n)) \end{pmatrix}$$

$$
\begin{aligned}
& +\Delta t^q \begin{pmatrix} \hat{f}_1(\Delta t, u_n, v_n, \eta_n + \hat{\psi}_e(u_n, v_n)) \\ \hat{f}_2(\Delta t, u_n, v_n, \eta_n + \hat{\psi}_e(u_n, v_n)) \end{pmatrix} \\
=: \; & (u_n, v_n) + G_1(\Delta t, u_n, v_n, \eta_n), \qquad (4.2) \\
\eta_{n+1} \; = \; & R(\infty)\eta_n + (b^T A^{-1} \otimes I)(\Lambda(\Delta t, u_n, v_n) - \mathbb{I} \otimes \hat{\psi}_e(u_n, v_n)) \\
& +\hat{\psi}_e(u_n, v_n) - \hat{\psi}_e((u_n, v_n) + G_1(\Delta t, u_n, v_n, \eta_n)) \\
=: \; & R(\infty)\eta_n + \beta(\Delta t, u_n, v_n, \eta_n) \; =: \; G_2(\Delta t, u_n, v_n, \eta_n).
\end{aligned}
$$

The functions G_1, G_2 are lipschitzian with the constants

$$
\begin{aligned}
L_{G_1,(u,v)} \; &= \; O(\Delta t), \;\; L_{G_1,\eta} \; = \; O(\Delta t^q), \\
L_{G_2,(u,v)} \; &= \; O(1), \;\; L_{G_2,\eta} \; = |\, R(\infty)\,| + O(\Delta t^q) < 1. \qquad (4.3)
\end{aligned}
$$

Obviously, for a fixed number $r \in \mathbb{N}$ the conditions

$$
\begin{aligned}
2\sqrt{L_{G_1,\eta} L_{G_2,(u,v)}} \; &< \; 1 - L_{G_1,(u,v)} - L_{G_2,\eta}, \\
L_{G_2,\eta} + L_{G_1,\eta}\alpha \; &< \; (1 - L_{G_1,(u,v)} - L_{G_1,\eta}\alpha)^r
\end{aligned}
$$

with

$$
\alpha \; := \; \frac{2L_{G_2,(u,v)}}{1 - L_{G_1,(u,v)} - L_{G_2,\eta} + \sqrt{(1 - L_{G_1,(u,v)} - L_{G_2,\eta})^2 - 4L_{G_1,\eta} L_{G_2,(u,v)}}}
$$

are satisfied for $\Delta t > 0$ sufficiently small. Now, Theorem 5 of Nipp, Stoffer [12] guarantees the existence of a C_b^ν-function $\eta_{\Delta} t$ which defines the discrete invariant manifold by $\eta = \eta_{\Delta t}(u, v)$. In the (u, v, λ)-coordinates we obtain with $\psi_{e,\Delta t}(u, v) := \hat{\psi}_e(u, v) + \eta_{\Delta t}(u, v)$, $(u, v) \in D_{\tau_0}$ for $0 < \Delta t \leq \Delta t_0$, $\Delta t_0 > 0$ sufficiently small the following result:

i) The set $M_{e,\Delta t} = \{(u, v, \lambda) \in D_{\tau_0} \times \mathbb{R}^l \mid \lambda = \psi_{e,\Delta t}(u, v)\}$ is invariant for the projected Runge-Kutta map (4.1).

ii) The manifold $M_{e,\Delta t}$ is uniformly attractive with attractivity constant $\chi_{\Delta t} = |\, R(\infty)\,| + O(\Delta t^q)$.

iii) For every initial value (u_0, v_0, λ_0) with $\| \lambda_0 - \psi_{e,\Delta t}(u_0, v_0) \|$ sufficiently small there is $(\tilde{u}_0, \tilde{v}_0, \tilde{\lambda}_0) \in M_{e,\Delta t}$ and $c, \hat{c} > 0$ such that the corresponding evolutions (u_n, v_n, λ_n) and $(\tilde{u}_n, \tilde{v}_n, \tilde{\lambda}_n)$ satisfy

$$
\begin{aligned}
\| (u_i, v_i) - (\tilde{u}_i, \tilde{v}_i) \| \; &\leq \; c\chi_{\Delta t}^i \, \| \lambda_0 - \psi_{e,\Delta t}(u_0, v_0) \|, \;\; i \in \mathbb{N}, \\
\| \lambda_i - \tilde{\lambda}_i \| \; &\leq \; \hat{c}\chi_{\Delta t}^i \, \| \lambda_0 - \psi_{e,\Delta t}(u_0, v_0 \| \;\; i \in \mathbb{N}.
\end{aligned}
$$

iv) $\| \hat{\psi}_e(u, v) - \psi_{e,\Delta t}(u, v) \| \leq C \sup\{\| \beta(\Delta t, u, v, \hat{\psi}_e(u, v) - \hat{\psi}_e(u, v)) \| \mid (\Delta t, u, v) \in]0, \Delta t_0] \times D_{\tau_0}\}$.

Here the reader may notice that to apply Theorem 5 of Nipp, Stoffer [12] formally we have to enlarge the domain of G_1, G_2 for $(u,v) \in \mathbb{R}^{N+M}$ as C_b^ν-maps which satisfy the lipschitz conditions (4.3).
Reduced to the invariant manifold $M_{e,\Delta t}$ the (u,v)-component of a projected Runge-Kutta method reads

$$\begin{aligned} u_{n+1} &= u_n + \Delta t[(b^T \otimes I)\bar{f}(U(\Delta t, u_n, v_n), V(\Delta t, u_n, v_n)) \\ &\quad +\Delta t^{q-1}\hat{f}_1(\Delta t, u_n, v_n, \psi_{e,\Delta t}(u_n, v_n))], \\ v_{n+1} &= v_n + \Delta t[(b^T \otimes I)\bar{k}(U(\Delta t, u_n, v_n), V(\Delta t, u_n, v_n), \Lambda(\Delta t, u_n, v_n)) \\ &\quad +\Delta t^{q-1}\hat{f}_2(\Delta t, u_n, v_n, \psi_{e,\Delta t}(u_n, v_n))]. \end{aligned} \tag{4.4}$$

Obviously, the iteration scheme (4.4) can be regarded as a $(q-1)$-th order one-step method applied to the index 0 problem in the (u,v)-variables

$$\begin{aligned} \dot{u} &= f(u,v),\ u(0) = u_0, \\ \dot{v} &= k(u,v,\hat{\psi}_e(u,v)),\ v(0) = v_0 \end{aligned} \tag{4.5}$$

of the embedded equation (3.1). Next we estimate the distance between $\hat{\psi}(u,v)$ and $\psi_{0,\Delta t}(u,v)$. With β from (4.2) we set

$$\tilde{\beta}_n := \beta(\Delta t, u_n, v_n, \psi_{e,\Delta t}(u_n, v_n) - \hat{\psi}_e(u_n, v_n)).$$

According to iv) we have to show $\tilde{\beta}_n = O(\Delta t^{q-1})$.

Using the definition of β in (4.2) and the discrete Δt-time map $(\hat{u}, \hat{v}, \hat{\lambda})(\Delta t, u, v, \lambda)$ we obtain with the local error

$$\begin{aligned} locerr_{u,v} &:= (\bar{u}, \bar{v})(\Delta t, u, v) - (\hat{u}, \hat{v})(\Delta t, u, v, \hat{\psi}_e(u,v)) = O(\Delta t^q), \\ locerr_\lambda &:= \bar{\lambda}(\Delta t, u, v) - \hat{\lambda}(\Delta t, u, v, \hat{\psi}_e(u,v)) = O(\Delta t^{q-1}) \end{aligned}$$

the estimation

$$\begin{aligned} \tilde{\beta}_n &= (b^T A^{-1} \otimes I)(\Lambda(\Delta t, u_n, v_n) - \mathbb{I} \otimes \hat{\psi}_e(u_n, v_n)) + \hat{\psi}_e(u_n, v_n) \\ &\quad -\hat{\psi}_e((\hat{u}, \hat{v})(\Delta t, u_n, v_n, \psi_{e,\Delta t}(u_n, v_n)) \\ &= -locerr_{\lambda_n} + \hat{\psi}_e((\bar{u}, \bar{v})(\Delta t, u_n, v_n)) - \hat{\psi}_e((\hat{u}, \hat{v})(\Delta t, u_n, v_n, \hat{\psi}_e(u_n, v_n))) \\ &\quad +\hat{\psi}_e((\hat{u}, \hat{v})(\Delta t, u_n, v_n, \hat{\psi}_e(u_n, v_n))) \\ &\quad -\hat{\psi}_e((\hat{u}, \hat{v})(\Delta t, u_n, v_n, \psi_{e,\Delta t}(u_n, v_n))) \\ &= -locerr_{\lambda_n} + O(1)(locerr_{u_n,v_n}) + O(\Delta t)(\hat{\psi}_e(u_n, v_n) - \psi_{e,\Delta t}(u_n, v_n)) \\ &= O(\Delta t^{q-1}) + O(\Delta t) * (\hat{\psi}_e(u_n, v_n) - \psi_{e,\Delta t}(u_n, v_n)). \end{aligned} \tag{4.6}$$

Combining (4.6) with property iv) this leads to

$$(1 + O(\Delta t)) * \| \hat{\psi}_e(u_n, v_n) - \psi_{e,\Delta t}(u_n, v_n) \| = O(\Delta t^{q-1}) \tag{4.7}$$

Since $(I + O(\Delta t))^{-1} = I + O(\Delta t)$ holds, formulae (4.6), (4.7) imply the desired estimation

$$\| \hat{\psi}_e(u_n, v_n) - \psi_{e,\Delta t}(u_n, v_n) \| \quad = \quad O(\Delta t^{q-1}), \quad (u_n, v_n) \in D_{\tau_0}.$$

Now, we complete the proof of Theorem 2.1. Since the projected Runge-Kutta methods applied to the original DAE (2.1) can be regarded as the same method applied to the embedded DAE (3.1) with $w(0) = w_0 = 0$ and $z(0) = z_0 = 0$ we can draw back the derived results. We restrict (4.4) to the invariant set S_0 and define $\psi_{0,\Delta t} := \psi_{e,\Delta t} \mid_{S_0}$ by $\psi_{0,\Delta t}(u, v) = \psi_{e,\Delta t}(u, v)$, $(u, v) \in S_0$ as well as $M_{0,\Delta t} = \{(u, v, \lambda) \in D_{\tau_0} \times \mathbb{R}^l \quad | \quad g(u) = 0,\ Dg(u)f(u, v) = 0,\ \lambda = \psi_{0,\Delta t}(u, v)\}$.
This finishes the proof of Theorem 2.1.

References

[1] Ascher, U., Petzold, L.R., *Projected Runge-Kutta methods for Differential-Algebraic Equations*, SIAM J. of Numer. Anal., **28** (1991), 1097–1120.

[2] Beyn, W.-J., *On invariant closed curves for one-step methods*, Numer. Math., **51** (1987), 103–122.

[3] Beyn, W.-J., *On the numerical approximation of phase portraits near stationary points*, SIAM J. Numer. Anal., **24** (1987), 1095–1113.

[4] Beyn, W.-J., Schropp, J., *Runge-Kutta discretizations of singularly perturbed gradient equations*, BIT Numerical Mathematics, **40** (2000), pp. 415–433.

[5] Dekker, K., Verwer, J.G., *Stability of Runge-Kutta methods for stiff nonlinear differential equations* (1984), CWI Monographs, North-Holland.

[6] Garay, B., *Discretization and some qualitative properties of ordinary differential equations about equilibria*, Acta Math. Com. Univ., **62** (1993), 249–275.

[7] Hairer, E., Lubich, Ch., Roche, M., *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*, Lecture Notes in Mathematics, 1409 (1989), Springer.

[8] Hairer, E., Wanner, G., *Solving Ordinary Differential Equations II*, second edition, Springer, Heidelberg, (1996).

[9] Jay, L., *Collocation methods for differential-algebraic equations of index 3*, Numer. Math., **65** (1993), 407–421.

[10] Jay, L., *Convergence of Runge-Kutta methods for differential-algebraic equations of index 3*, Appl. Num. Math., **17** (1995), 97–118.

[11] Kloeden, P., Lorenz, J., *Stable attracting sets in dynamical systems and in their one-step discretizations*, SIAM J. Numer. Anal., **23** (1986), 986–993.

[12] Nipp, K., Stoffer, D., *Attractive invariant manifolds for maps*, SAM Research report, No. 92-11 (1992), ETH Zürich.

[13] Schropp, J., *Geometric properties of Runge-Kutta discretizations for index 2 Differential Algebraic Systems*, SIAM J. of Numer. Anal., **40** (2002), p.872-890.

[14] Schropp, J., *Projected Runge-Kutta methods for differential algebraic equations of index 3*, Konstanzer Schriften in Mathematik und Informatik Nr. 191, Universität Konstanz.