Solving Algebraically Explicit DAEs with the MANPAK-Manifold-Algorithms

Werner C. Rheinboldt
Professor of Mathematics Emeritus, University of Pittsburgh

Abstract

Recently, the author introduced a package of algorithms, called MANPAK, for effective computations on implicitly defined submanifolds of \( \mathbb{R}^n \). Here algebraically explicit differential algebraic equations (DAEs) are considered; that is, DAEs in which either the algebraic equations and/or the algebraic variables are explicitly specified. Existence proofs for several types of such DAEs of index one, two, and three are given which directly suggest computational approaches. This is used in the development of solution algorithms for these DAEs, all of them intrinsically based on the MANPAK routines. Some numerical examples for the methods are included.

1 Introduction

In [14] a package of algorithms, called MANPAK, for computations on implicitly defined submanifolds of \( \mathbb{R}^n \) was presented. More specifically, we work with a sufficiently smooth mapping \( F: \mathbb{R}^n \to \mathbb{R}^m, \ d = n - m > 0 \), which is a submersion on its zero set \( M = F^{-1}(0) \), whence \( M \) is a \( d \)-dimensional submanifold of \( \mathbb{R}^n \). MANPAK includes algorithms for computing local parametrizations on such an implicitly defined manifold \( M \) and its tangent-bundle \( TM \), as well as other useful quantities on \( M \) including sensitivity measures and the second fundamental tensor.

In [14] several applications of the MANPAK algorithms were mentioned. Here we consider another such application, namely the solution of certain algebraically explicit differential algebraic equations (DAEs); that is, of DAEs in which either the algebraic
equations and/or the algebraic variables are explicitly specified. Algorithms are described for solving six types of DAEs of index one to three; all of them are based intrinsically on the MANPAK routines. These solvers have been implemented as a collection of FORTRAN-77 subroutines. As with all the MANPAK methods, the solvers are intended for application to small or medium-sized DAE problems, mainly because they involve many dense matrix computations.

Section 2 presents existence theories for algebraically explicit DAEs in a form exhibiting the algorithmic approach. Then, in section 3, the computational algorithms for these types of DAEs are developed. Finally, section 4 gives some numerical results obtained with the implementations.

## 2 Existence Results

We begin with autonomous DAEs of the form

\[
\begin{pmatrix}
F_1(u, u', w) \\
F_2(u)
\end{pmatrix} = 0
\tag{1}
\]

under the following condition:

**2.1. Assumption N2:** Let \(k_u, m_1 > 0, k_w, m_2 \geq 0\), be integers such that \(k_u + k_w = m_1 + m_2\), \(k_u \geq m_2\), and let \(E_u \subset \mathbb{R}^{k_u}\), \(E_w \subset \mathbb{R}^{k_w}\) be nonempty, open sets. Assume that

(i) \(F_1 : E_1 \to \mathbb{R}^{m_1}\) is \(C^1\) on \(E_1 = E_u \times \mathbb{R}^{k_u} \times E_w\);

(ii) \(F_2 : E_u \to \mathbb{R}^{m_2}\) is \(C^2\) on \(E_u\),

(iii) \(\text{rank } DF_2(u) = m_2\) on \(M_2 := \{u \in E_u : F_2(u) = 0\}\) whence \(M_2\) is a \(C^2\) submanifold of \(\mathbb{R}^{k_u}\) of dimension \(d = k_u - m_2\);

(iv) the \(n \times n\) matrix, \(n = k_u + k_w\),

\[
\begin{pmatrix}
D_pF_1(u, p, w) & D_uF_1(u, p, w) \\
D_F2(u) & 0
\end{pmatrix}
\tag{2}
\]

is nonsingular on the set \(K := \{(u, p, w) \in E_1 : F_1(u, p, w) = 0, (u, p) \in TM_2\}\) where \(TM_2\) is the tangent bundle of \(M_2\).

We are interested in a solution

\[
u : J \to E_u, \quad w : J \to E_w, \quad t \in J
\tag{3}
\]

of (N2) which is \(C^1\) on an open interval \(J\) containing the origin and satisfies the initial condition

\[
u(0) = u_0, \quad u'(0) = p_0, \quad w(0) = w_0.
\tag{4}
2 EXISTENCE RESULTS

For some existence results on such initial value problems we refer to [13], [8], and [9].

Here a proof is given that exhibits directly a computational approach.

Under Assumption N2 the index of (1) equals 2 if $k_\omega > 0$ and $m_2 > 0$, 1 if either $k_\omega = 0$ or $m_2 = 0$, and 0 if $k_\omega = 0$ and $m_2 = 0$. This is reflected in the label (N2) of the assumption where “N” stands for “nonlinear” and “2” indicates the (generic) index.

We shall assume first that $m_2 > 0$.

Given $(u_0, p_0, w_0) \in K$, let $(V_d, \varphi)$ be a $C^2$ local parametrization of $M_2$ near $u_0$; that is, $V_d \subset \mathbb{R}^d$ is an open neighborhood of the origin, and the $C^1$-mapping $\varphi: V_d \mapsto \mathbb{R}^{k_w}$, $\varphi(0) = u_0$

is a homeomorphism of $V_d$ onto its image and an immersion on $V_d$. Then $(V_d \times \mathbb{R}^d, (\varphi, D\varphi))$ is a $C^1$ local parametrization of $TM_2$ near $(u_0, p_0)$. In particular, we have $p_0 = D\varphi(0)z_0$ for some $z_0 \in \mathbb{R}^d$.

With fixed $y \in V_d$ consider the nonlinear system of equations

$$H(y; z, w) \equiv F_1(\varphi(y), D\varphi(y)z, w) = 0,$$

in the unknown $z, w$, for which $H(0; z_0, w_0) = 0$. The Jacobian of $H$ is

$$DH(y; z, w) = (D_p F_1(\varphi(y), D\varphi(y)z, w) D\varphi(0) \quad D_w F_1(\varphi(y), D\varphi(y)z, w)).$$

and $DH(0; z_0, w_0)$ is nonsingular. In fact, if $(\xi_1, \xi_2) \in \mathbb{R}^d \times \mathbb{R}^{kw}$ is a null-vector of the matrix (7) then $(D\varphi(y)\xi_1, \xi_2)$ is a null-vector of (2) whence $D\varphi(y)\xi_1 = 0$, $\xi_2 = 0$, and therefore also $\xi_1 = 0$ because of the injectivity of $D\varphi(y)$. Thus, the implicit function theorem applies to the system (6). This proves that, after shrinking $V^d$ if needed, there exists an open neighborhood $U \subset \mathbb{R}^d \times E_w$ and a unique $C^1$-mapping

$$\Psi: V^d \mapsto U \subset \mathbb{R}^d \times \mathbb{R}^{kw}, \quad \Psi(y) = (\psi_1(y), \psi_2(y)) \in U, \forall y \in V^d,$$

such that, for any $y \in V^d$, the unique solution of (6) in $U$ is given by $(z, w) = \Psi(y)$.

Consider the local initial value problem

$$y' = \psi_1(y), \quad y(0) = 0,$$

which, by standard ODE-theory, has a unique $C^1$-solution $y: \mathcal{J} \mapsto V^d$ on an open interval $\mathcal{J}$ containing the origin. Then, by construction of $\psi$ and the definition of $\varphi$ we have

$$F_1(\varphi(y(t)), D\varphi(y(t))\psi_1(y(t)), \psi_2(y(t))) = 0,$$

$$F_2(\varphi(y(t))) = 0, \quad \forall t \in \mathcal{J},$$
which shows that $u(t) = \varphi(y(t)), w(t) = \psi_2(y(t))$, is a $C^1$ solution of (1). Evidently, this solution satisfies the initial conditions (4). In other words, we proved that for $(u_0, p_0, w_0) \in K$ there exists a local $C^1$ solution of the initial value problem.

Conversely, let (3) be any $C^1$-solution of (1) on an open interval $J$ containing the origin. Then we have $F_1(u(t), u'(t), w(t)) = 0$ and $F_2(u(t)) = 0$ for all $t \in J$. Hence by differentiation with respect to $t$ we see that $DF_2(u(t))u'(t) = 0$ for all $t \in J$ and therefore $(u(t), u'(t), w(t)) \in K$ for all $t \in J$. Thus, the above results apply at any point of this solution. Since (1) is autonomous, it suffices to work with the point $(u_0, p_0, w_0) = (u(0), u'(0), w(0))$. Then, with a local parametrization $(Y_d, \varphi)$ of $M_2$ at $u_0 \in M_2$ and $z_0 \in \mathbb{R}^d$ such that $D\varphi(0)z_0 = p_0$, the $C^1$ - mapping $\Psi$ of (8) is well defined on $Y_d$. For the local curve $y : J \mapsto Y_d, y = \varphi^{-1} \circ u$, we have $u(t) = \varphi(y(t)), u'(t) = D\varphi(y(t))y'(t)$, and $y(0) = 0$ which shows that $F_1(\varphi(y(t)), D\varphi(y(t))y'(t), w(t)) = 0$, for all $t \in J$. In view of the uniqueness of the solution of the equation (4), this implies that $\Psi(y) = (y'(t), w(t))$ and, hence, that the local curve is a solution of the initial value problem (9).

Altogether this proves the following result:

**2.2 Theorem.** Suppose that Assumption N2 holds for the DAE (1). Then, for any point $(u_0, p_0, w_0) \in K$ there exists a unique $C^1$-solution (2.2) of (1) on some open interval $J$, $t_0 \in J$ which satisfies the initial conditions (4).

As noted earlier, the index of (1) reduces to one if either $m_2 = 0$ or $k_w = 0$. Both cases are covered by the theorem. But, for $m_2 = 0$ the manifold $M_2$ equals the open set $E_u$ and the proof of Theorem 2.2 no longer suggests a very efficient computational approach. Accordingly, we give here a separate proof for this special case.

For $m_2 = 0$, we write (1) in the form

$$
\begin{pmatrix}
u' \\
F(u, p, w)
\end{pmatrix} = \begin{pmatrix} p \\
0
\end{pmatrix}
$$

where, in analogy to Assumption (1), we now use the condition:

**2.3. Assumption N1:** Let $k_u, k_w > 0, m = k_u + k_w$, be integers and $E_u \subset \mathbb{R}^{k_u}$, $E_w \subset \mathbb{R}^{k_w}$ nonempty, open sets. Assume that

(i) $F : E_1 \rightarrow \mathbb{R}^m$ is $C^2$ on $E = E_u \times E_u \times E_w$,  
(ii) the $m \times m$ matrix $\begin{pmatrix} D_pF(u, p, w) & D_wF(u, p, w) \end{pmatrix}$ is nonsingular for any $(u, p, w) \in M = F^{-1}(0)$, whence $M$ is a submanifold of $\mathbb{R}^n$, $n = 2 \ast k_u + k_w$, of dimension $d = k_u$.

Let $(Y_d, \varphi)$ be a $C^1$-local parametrization of $M$ near some point $(u_0, p_0, w_0) \in M$, where we use for $\varphi$ the component-notation

$$
\varphi : Y_d \mapsto \mathbb{R}^n \cong \mathbb{R}^{k_u} \times \mathbb{R}^{k_u} \times \mathbb{R}^{k_w}
$$

$$
\varphi(y) \equiv (\varphi_1(y), \varphi_2(y), \varphi_3(y)) \in M, \forall y \in Y_d, \varphi(0) = (u_0, p_0, w_0),
$$
For given \( y \in \mathcal{V}^d \) the \( d \times d \) matrix \( D\varphi_1(y) \) is nonsingular. In fact, suppose that \( D\varphi_1(y)\xi = 0 \) for some \( \xi \in \mathbb{R}^d \). Then

\[
0 = DF(\varphi(y))D\varphi(y)\xi = \begin{pmatrix} D_p F(\varphi(y)) & D_w F(\varphi(y)) \end{pmatrix} \begin{pmatrix} D\varphi_2(y)\xi \\ D\varphi_3(y)\xi \end{pmatrix}
\]

and, by condition (ii) of Assumption N1, the matrix on the right side is nonsingular whence \( D\varphi_2(y)\xi = 0 \), and \( D\varphi_3(y)\xi = 0 \). Thus altogether we have \( D\varphi(y)\xi = 0 \) which requires that \( \xi = 0 \).

Consider now the local initial value problem

\[
D\varphi_1(y)y' = \varphi_2(y), \ y(0) = 0,
\]

which by ODE theory has a unique \( C^1 \)-solution \( y : \mathcal{J} \mapsto \mathcal{V}^d \) on some open interval \( \mathcal{J} \) containing the origin. Then, by construction, we have

\[
0 = F(\varphi(y)) = F(\varphi_1(y(t)), D\varphi_1(y(t))y'(t), \varphi_3(y(t))), \quad \forall t \in \mathcal{J}
\]

which shows that \( u(t) = \varphi_1(y(t)), w(t) = \varphi_3(y(t)) \), define a \( C^1 \) solution of (N1). Moreover, because of \( \varphi(0) = (u_0, p_0, w_0) \), the initial conditions (4) hold.

Conversely let (3) be any \( C^1 \)-solution of (N1) on an open interval \( \mathcal{J} \) containing the origin. Then, clearly, \( (u(t), u'(t), w(t)) \in M \) for all \( t \in \mathcal{J} \) and the earlier considerations apply at any point of this solution. As before, it suffices to work only with the point \( (u_0, p_0, w_0) = (u(0), u'(0), w(0)) \). Let \( (\mathcal{V}^d, \varphi) \) be a \( C^1 \) local parametrization of \( M \) at this point and consider the local curve \( y : \mathcal{J} \mapsto \mathcal{V}^d, y = \varphi_1^{-1} \circ u \). Then \( u(t) = \varphi(y(t)) \) and \( u'(t) = \varphi_2(y(t)) \) imply that \( D\varphi(y(t))y'(t) = \varphi_2(y(t)) \) for \( t \in \mathcal{J} \) and from \( \varphi(0) = u_0 = u(0) = \varphi(y(0)) \) it follows that \( y(0) = 0 \). Hence, \( y \) is a solution of the initial value problem (12).

With this we obtained the following result:

**2.4 Theorem.** Suppose that Assumption N1 holds for the DAE (10). Then for any point \( (u_0, p_0, w_0) \in M \) there exists a unique \( C^1 \)-solution (11) of (10) on some open interval \( \mathcal{J}, 0 \in \mathcal{J} \), which satisfies the initial conditions (4).

These results extend also to second order, algebraically explicit DAEs which, in general, have index three. As illustration, we consider here only the special case of quasilinear systems of the form

\[
\begin{pmatrix} A(u, u')u'' + B(u, u')w \\ F(u) \end{pmatrix} = \begin{pmatrix} G(u, u') \\ 0 \end{pmatrix}
\]

under the following condition:
2.5. Assumption Q3: Let $k_u, m_1 > 0$, $k_w, m_2 \geq 0$, be integers such that $k_u + k_w = m_1 + m_2$, $k_u \geq m_2$, and $E_u \subset \mathbb{R}^{k_u}$, $E_w \subset \mathbb{R}^{k_w}$ nonempty, open sets. Assume that

(i) $A : E_1 \mapsto L(\mathbb{R}^{k_u}, \mathbb{R}^{m_1})$, and $B : E_1 \mapsto L(\mathbb{R}^{k_w}, \mathbb{R}^{m_1})$, $G : E_1 \mapsto \mathbb{R}^{m_1}$ are $C^1$ on $E_1 = E_u \times \mathbb{R}^{k_u}$;
(ii) $F : E_2 \mapsto \mathbb{R}^{m_2}$ is $C^3$ on $E_2 = E_u$;
(iii) $\text{rank } DF(u) = m_2$ on $M = F^{-1}(0)$ whence $M$ is a $C^3$-submanifold of $\mathbb{R}^{k_u}$ of dimension $d = k_u - m_2$;
(iv) the $n \times n$ matrix, $n = k_u + k_w$,
\[
\begin{pmatrix}
A(u,p) & B(u,p) \\
DF(u) & 0
\end{pmatrix},
\]

is nonsingular for any $(u,p) \in TM$.

Given $(u_0, p_0) \in TM$ let $(\mathcal{V}^d, \varphi)$ be a $C^3$-local parametrization of $M$ near $u_0$, and $(\mathcal{V}^d \times \mathbb{R}^d, (\varphi, D\varphi))$ the induced $C^2$-local parametrization of $TM$ near $(u_0, p_0)$. As before, there is some $z_0 \in \mathbb{R}^d$ such that $p_0 = D\varphi(0)z_0$.

We introduce the $C^1$-mapping
\[
\hat{G} : \mathcal{V}^d \times \mathbb{R}^d \mapsto \mathcal{V}^{m_1}
\]
\[
\hat{G}(y, z) = G(\varphi(y), D\varphi(y)z) - A(\varphi(y), D\varphi(y)z)D^2\varphi(y)(z, z), \forall (y, z) \in \mathcal{V}^d \times \mathbb{R}^d,
\]

and the $C^2$ mapping
\[
\Gamma : \mathcal{V}^d \times \mathbb{R}^d \mapsto L(\mathbb{R}^{m_1})
\]
\[
\Gamma(y, z) = (A(\varphi(y), D\varphi(y)z)D\varphi(y)B(\varphi(y), D\varphi(y)z)), \forall (y, z) \in \mathcal{V}^d \times \mathbb{R}^d.
\]

Then, the same proof as for the matrix (7) shows that the matrix $\Gamma(y, z)$ is nonsingular for any fixed $(y, z) \in \mathcal{V}^d \times \mathbb{R}^d$ Let $\pi_d$ and $\pi_w$ denote the canonical projections of the space $\mathbb{R}^d \times \mathbb{R}^{k_w}$ onto its first and second component, respectively. We consider the local initial value problem
\[
y' = z, \quad z' = \pi_d \Gamma(y, z)^{-1}\hat{G}(y, z), \quad y(0) = 0, z(0) = z_0.
\]

Since the right side is of class $C^1$, ODE theory guarantees that (17) has a unique $C^1$-solution $y : \mathcal{J} \mapsto \mathcal{V}^d$ on some open interval $\mathcal{J}$ containing the origin. Set
\[
u(t) = \varphi(y(t)), \quad w(t) = \pi_w \Gamma(y, z)^{-1}\hat{G}(y(t), z(t)), \quad \forall t \in \mathcal{J}
\]

Then $F(u(t)) = 0$ for $t \in \mathcal{J}$, and, by differentiation we find that for all $t \in \mathcal{J}$
\[
u'(t) = D\varphi(y(t))z(t), \quad \nu''(t) = D\varphi(y(t))z'(t) + D^2\varphi(y(t))(z(t), z(t)).
\]
The second equation of (17) implies that
\[ A(u(t), u'(t))D\varphi(y(t))z'(t) + B(u(t), u'(t))w(t) = G(u(t), u'(t))z(t) - A(u(t), u'(t))D^2\varphi(y(t))(z(t), z(t)) \]
which, together with (19), shows that (18) is a $C^1$-solution of (13) for which
\[ u(t_0) = u_0, \ u'(t_0) = p_0. \] (20)
Note that the second parts of equations (18) and (19) enforce initial values for $w(0)$ and $u''(0)$ which therefore cannot be prescribed.

Conversely, let (3) be any $C^2$-solution of (13) on an open interval $J$ containing the origin. Then $F(u(t)) = 0$ and $DF(u(t))u'(t) = 0$; that is, $(u(t), u'(t)) \in TM$ for all $t \in J$. Thus, the previous consideration apply at any point of this solution and it suffices again to work with the point $(u_0, p_0, w_0) = (u(0), u'(0), w(0))$ where $p_0 = D\varphi(0)z_0$ for some $z_0 \in \mathbb{R}^d$.

Then with a local parametrization $(V^d, \varphi)$ of $M$ at $u_0 \in M$, the mappings (15) and (16) are well-defined and the matrix $\Gamma(y, z)$ is nonsingular for any fixed $(y, z) \in V^d \times \mathbb{R}^d$.

For the $C^2$ local curve $y : J \mapsto V^d, y = \varphi^{-1} \circ u$, it follows that
\[ \varphi(y(t)) = u(t), \ D\varphi(y(t))y'(t) = u'(t), \ D^2\varphi(y(t))y''(t) + D^2\varphi(y(t))(y'(t), y'(t)) = u''(t). \]
and $y(0) = 0$ and $y'(0) = z_0$. Moreover, by substituting these representations for $u$ and its derivatives into the first equation of (13) and using (15) and (16), we find that
\[ \Gamma(y(t), y'(t)) \begin{pmatrix} y''(t) \\ w(t) \end{pmatrix} = \hat{G}(y(t), y'(t)), \ \forall t \in J, \]
whence $y$ and $z = y'$ is a solution of (17).

This proves the following result:

**2.6 Theorem.** Suppose that Assumption Q3 holds for the DAE (13). Then for any point $(u_0, p_0) \in E_1$ such that $(u_0, p_0) \in TM$ there exist a unique $C^2$-solution (2.2) of (13) on some open interval $J \subset E_t, t_0 \in J$ which satisfies the initial condition (20).

### 3 Computational Algorithms

All algorithms in this section (except one) are written for non-autonomous versions of the DAEs in Section 2. Our existence results are easily extended to these non-autonomous cases by adding, as usual, the equation $t' = 0$. The details should not be required but some comments will be provided with the algorithms.
The existence proofs suggest that, in each case, we should solve the corresponding local ODE to compute the solution of the (global) DAE. Thus the process always begins with the construction of some local parametrization of the manifold constraining the problem, followed by the application of a standard ODE solver to the resulting local system. When the computed points appear to leave the domain of validity of the local coordinate system, a new local parametrization is generated and the process is continued by applying the ODE solver to the new local ODE.

For the algorithms discussed here, a new reverse-communication version of the explicit Runge-Kutta method DOPRI5 of [5] was chosen as the ODE solver. Other integrators are readily applicable in this setting. For simplicity, the codes establish a new local parametrization at each accepted point; that is, after each successful RK-step. Once all stages and the next approximate solution of the local ODE have been computed, the RK-routine performs the standard error computation. If the RK-step is accepted then from the computed results a new approximate solution of the DAE is determined and the process is repeated with this new global point until the required termination time has been reached. If the RK-step was rejected, then, as usual, the stepsize is reduced and, from the previous approximate global solution, the step is repeated with the estimated smaller stepsize. A stepsize reduction may also have to be enforced if, during an RK-step the local parametrization evaluation fails to converge which indicates that the local vector, at that stage, fails to belong to the domain of the local parametrization.

We note that there is no fundamental obstacle to retaining the local parametrization for several successful RK-steps.

There should be no need to detail the overall framework of the codes involving the use of the ODE solver. Instead we focus on the principal aspect of the different algorithms, namely the evaluation of the right side of the local DAEs. All these algorithms are called in two distinct modes to distinguish whether or not the local parametrization has to be established at the particular point.

In the case of (1) the algorithm is applied to the DAE

\[
\begin{pmatrix}
F_1(u, u', w, t) \\
F_2(u, t)
\end{pmatrix} = 0
\]  

(21)

where the equation \(t' = 0\) is not explicitly stated. Note that now \(M_2 = F_2^{-1}(0)\) is a submanifold of \(\mathbb{R}^n\), \(n = k_u + 1\), of dimension \(d = n - m_2\). Accordingly, for the local parametrization \((\mathcal{V}^d, \varphi)\) at any \((u_0, t_0) \in M_2\) the component notation

\[
\varphi : \mathcal{V}^d \mapsto \mathbb{R}^{k_u} \times \mathbb{R}^1, \quad \varphi(y) \equiv (\varphi_1(y), \varphi_2(y)) \in \mathcal{E}_u \times \mathcal{E}_t, \quad \forall y \in \mathcal{V}^d, \quad (u_0, t_0) = \varphi(0)
\]  

(22)
is introduced. The evaluation of the right side of the local ODE then requires the solution of the nonlinear system

\[ H(y, z, w) \equiv \left( \begin{array}{c} F_1(\varphi_1(y), D\varphi_1(y)z, w, \varphi_2(y)) \\ D\varphi_2(y)z - 1 \end{array} \right) = 0. \]  \hspace{1cm} (23)

With the solution function \( \Psi \) of (8) we have here

\[ D\varphi_2(y)\psi_1(y) = 1 \quad \text{for all} \quad y \in \mathbb{V} \]

which implies that now the relation between the local and global solution is

\[ u(t) = \varphi(y(t - t_0)), \quad w(t) = \psi_2(y(t - t_0)). \]

For the solution of (23) a standard chord-Newton method is applied with the Jacobian of \( H \) as iteration matrix typically evaluated at the time of the establishment of the local parametrization; that is, with a matrix of the form

\[
\begin{pmatrix}
(D_p F_1(u, p, w, t) D\varphi_1(0) & 0) \\
(D w F_1(u, p, w, t)) & 0
\end{pmatrix}
\]  \hspace{1cm} (24)

Then the algorithm for the evaluation of the right side of the local ODE has the form:

**DYN2 Input:** mode

If mode = ‘init’ then

**Input:** Global point \((u, p, w, t)\);

Evaluate \( DF(u, t) \);

With GNBAS compute the basis of the local parametrization \( \varphi \) at \((u, t)\);

Set the local point \( y = (0, 0) \in \mathbb{R}^{k_u} \times \mathbb{R}^1 \);

Set up and factor the iteration matrix (24) at the point \((u, p, w, t)\);

Use DGPHI to evaluate \( D\varphi(y) = (D\varphi_1(y), D\varphi_2(y)) \)

Set the start vector \((\zeta, w)\) with \( \zeta = D\varphi_1(y)^{\top} p \);

Set the parametrization time \( t_c = t \);

Else

**Input:** Local point \( y \) in the parametrization \( \varphi \) at time \( t_c \), last vector \((\zeta_\ell, w_\ell)\);

With GPHI evaluate the global point \((u, t) = \varphi(y)\)

If GPHI fails to converge then force an RK-step reduction

Use DGPHI to evaluate \( D\varphi(y) = (D\varphi_1(y), D\varphi_2(y)) \)

Set the start vector \((\zeta, w) = (\zeta_\ell, w_\ell)\);

End If;

Use the chord-Newton method with the current iteration matrix to solve the nonlinear system \( H(0, \zeta, w) = 0 \) of (6) for \( \zeta, w \);

If the method diverges Then error return

Set \( p = D\varphi_1(y)(\zeta) \);

Output: \( y' := (\zeta, 1), (u, p, t), (\zeta_\ell, w_\ell) := (\zeta, w) \).
Since $M_2$ is here a submanifold of $\mathbb{R}^{k_u} \times \mathbb{R}^1$, where the one-dimensional space represents the $t$ variable, the local parametrization is constructed by means of the algorithm GN-BAS of [14] which preserves the $t$ variable. Correspondingly, for the computation of the derivative $D\varphi(0)$ of the local parametrization the algorithm DGPHI of [14] is applied.

If, in the 'init'-mode, we were assured that $(u_0, p_0, w_0, t_0) \in K$ then we would expect that $H(0, z, w_0) = 0$ for $z = [D\varphi(0)^\top D\varphi(0)]^{-1}D\varphi(0)^\top p_0$. For the algorithm it was found to be more advantageous to enforce the validity of the equation explicitly by applying the chord Newton method to (24) starting with $(\zeta, w) = D\varphi(0)^\top (p, 1)$ and $w = w_0$. This usually converges in a step or two and is no more costly than the indicated direct evaluation of $z$. Moreover, no convergence indicates here that the point is too far away from $K$ to be corrected onto that manifold. The algorithm for the solution of initial value problems for the nonautonomous version (21) of (1) has been implemented as a subroutine suite called DAEN2.

The process simplifies considerably in several special cases of (1). In particular, in the quasilinear case

\[
\begin{pmatrix}
A(u, t)u' + B(u, t)w \\
F_2(u, t)
\end{pmatrix} = \begin{pmatrix}
G(u, t) \\
0
\end{pmatrix}
\]  

the chord Newton iteration is no longer needed. Moreover, only the initial condition $(u_0, t_0) \in M_2$ has to be given and no initial vectors $p_0$ and $w_0$ are needed. For each local vector $y \in V_d$ arising during the RK-step, we have to set up and solve the linear system

\[
\begin{pmatrix}
(A(\varphi(y)) D\varphi_1(y) & 0 \\
D\varphi_2(y) & 0
\end{pmatrix} \begin{pmatrix}
z \\
w
\end{pmatrix} = G(\varphi(y))
\]  

(26)

to obtain $z = \psi_1(y)$ and $w = \psi_2(y)$. All other parts of the algorithm remain the same. The algorithm for solving initial value problems for the special case (21) of (1) has been implemented as a subroutine suite called DAEQ2.

In the special index-one case (10) of (1) our proof already indicated the considerable simplifications in the algorithm. The algorithm is applied to

\[
\begin{pmatrix}
u' \\
F(u, p, w, t)
\end{pmatrix} = \begin{pmatrix}
p \\
0
\end{pmatrix}
\]  

(27)

where, again, the equation $t' = 0$ is subsumed. In this case, $M_1$ is a $d = k_u + 1$-dimensional submanifold of $\mathbb{R}^n$, $(n = 2k_u + k_w + 1)$, and, analogous to (22), we write the local parametrization in the component form

$\varphi : V_d \rightarrow \mathbb{R}^n$, $\varphi(y) = (\varphi_1(y), \varphi_2(y), \varphi_3(y), \varphi_4(y)) \in \mathbb{R}^{k_u} \times \mathbb{R}^{k_u} \times \mathbb{R}^{k_u} \times \mathbb{R}^1$, $\forall y \in V_d$.  

(28)

Then the local ODE becomes

\[
\begin{pmatrix}
D\varphi_1(y) \\
D\varphi_4(y)
\end{pmatrix} y' = \begin{pmatrix}
\varphi_2(y) \\
1
\end{pmatrix}
\]  

(29)
and the evaluation of its right side has the following form:

**DYN1:** **Input:** mode
   
   If mode = ‘init’ then
      
      **Input:** Global point \((u, p, w, t)\);
      
      Evaluate \(DF(u, p, w, t)\);
      
      With GNBAS compute the basis of \(\varphi\) at \((u, p, w, t)\);
      
      Set the local point \(y = (0, 0) \in \mathbb{R}^k \times \mathbb{R}^1\);
      
      Parametrization time \(t_c = t\);
   
   Else
      
      **Input:** Local point \(y\) in the parametrization \(\varphi\) at time \(t_c\);
      
      With GPHI evaluate the global point \((u, p, w, t) = (\varphi_1(y), \varphi_2(y), \varphi_3(y), \varphi_4(y))\);
      
      If GPHI fails to converge then force an RK-step reduction
   
   End If;
   
   With DGPHI evaluate \(D\varphi(y) = (D\varphi_1(y), D\varphi_2(y), D\varphi_3(y), D\varphi_4(y))\);
   
   Solve the linear system \(D\varphi_1(y)\zeta = \varphi_2(y)\) for \(\zeta\);
   
   **Output:** \(y' := (\zeta, 1), (u, p, t)\).

The algorithm for the solution of initial value problems for the nonautonomous version (27) of (10) has been implemented as subroutine suite called DAEN1.

In analogy to (25) we may admit also a specialization of (10) to quasilinear form. Such index-one systems, albeit in the autonomous form

\[
\begin{pmatrix}
A(u)u' \\
F_2(u)
\end{pmatrix} = \begin{pmatrix} G(u) \\ 0 \end{pmatrix}
\]  

(30)

have been considered earlier in [11] in connection with the computation of impasse points (see also [10]). In this case the reduced system has the form

\[
A(\varphi(y))D\varphi(y)y' = G(\varphi(y))
\]  

(31)

and it is obvious how to change the above algorithm DYN1 for this case. The autonomous form was retained here to allow for a direct integration of the solution algorithm with the methods of [11]. In our setting the impasse points are those points where the reduced ODE has a standard singularity, in the sense that \(\hat{A}(y) = A(\varphi(y))D\varphi(y)\) satisfies

\[
\dim \ker \hat{A}(y) = 1, \quad D\hat{A}(y)(z, z) \notin \text{rge} \ \hat{A}(y), \forall z \in \ker \ \hat{A}(y).
\]

The combined algorithm has been implemented in the form of a suite of subroutines called DAESQ1 where “S” stands for “singular” point.
In the case of (13) the algorithm is applied to the nonautonomous version
\[
\begin{pmatrix}
A(u, u', t)u'' + B(u, u', t)w \\
F(u, t)
\end{pmatrix} = \begin{pmatrix}
G(u, u', t) \\
0
\end{pmatrix}
\] (32)

Here \( M \) is a submanifold of \( \mathbb{R}^n, n = k_u + 1 \), of dimension \( d = n - m_2 \) and we use the component notation (22) for the local parametrizations. The algorithm for the evaluation of the right side of the local ODE now has the form:

**DYQ3**  
**Input:** mode  
If mode = ‘init’ then  
- **Input:** Global point \((u, p, t)\);  
  Evaluate \(DF(u, t)\);  
  With GNBAS compute basis matrix \(U_c\) of \(\varphi\) at \((u, t)\);  
  Set the local point \(y = (0, 0)\); \(z = U_c^\top p_c\)  
  Set the parametrization time \(t_c = t\);  
Else  
- **Input:** Local point \(y, z\) in the local parametrization \(\varphi\) at time \(t_c\)  
  With GPHI evaluate the global point \((u, t) = (\varphi_1(y), \varphi_2(y))\)  
  If GPHI fails to converge then force an RK-step reduction  
**End If:**  
- Use DGPHI to evaluate the derivative \(D\varphi(y) = (D\varphi_1(y), D\varphi_2(y))\);  
  Set \(p = D\varphi_1(y)z\);  
  Evaluate \(q = D^2\varphi(y)(z, z)\);  
  Evaluate \(G = G(u, p, t) - A(u, p, t)q\);  
  Evaluate \(\Gamma = (A(u, p, t)D\varphi(y)B(u, p, t))\);  
  Solve the linear system \(\Gamma(\zeta, w)^\top = \hat{G}\) for \(\zeta\) and \(w\);  
**Output:** \(y' := z; z' := \zeta; (u, p, w, t)\).

The algorithm for the solution of initial value problems for the nonautonomous version (25) of (13) has been implemented as subroutine suite called DAEQ3.

Evidently, the Euler-Lagrange equations
\[
\begin{pmatrix}
K(u, t)u'' + D_u F(u, t)^\top w \\
F(u, t)
\end{pmatrix} = \begin{pmatrix}
G(u, u', t) \\
0
\end{pmatrix}
\] (33)
arising as models of multibody systems, are a special case of (26). In that case, the condition (iv) of Assumption Q3 is equivalent with the usual assumption that the mass matrix \(K\) is definite on \(T_{u,t} M\). In [12] a different algorithm for the numerical solution of (33) was given which uses the second fundamental tensor of the manifold \(M\). This algorithm has also been added to the present collection of DAE solvers under the name DAEUL3.
4 Computational Examples

As described in the previous sections, the package of solvers for algebraically explicit DAEs consists of the following suites of routines:

DAEN1 for nonlinear, index-1 DAEs (21) of order one
DAESQ1 for quasilinear, index-1 DAEs (30) of order one with singular points
DAEN2 for nonlinear, index-2 DAEs (21) of order one
DAEQ2 for quasilinear, index-2 DAEs (25) of order one
DAEQ3 for quasilinear, index-3 DAEs (32) of order two
DAEUL3 for Euler-Lagrange equations (33).

FORTRAN 77 versions of the codes are available.

We give here some numerical examples.

The following index-two problem was given in [7]

\begin{align*}
  u_1' + \sin(\arccos u_1) - \frac{1}{u_1} + w^2 + 1 &= 0 \\
  u_2' + w &= 0 \\
  u_2 - \log u_1 &= 0
\end{align*}

(34)

For \( u(0) = (1, 0), w(0) = 0 \), the exact solution is \( u_1^*(t) = \cos t, u_2^* = \log \cos t, w^*(t) = \tan t \). In [7] the system was integrated from \( t = 0.5 \) to \( t = 1.5 \) using accurate starting values and, among others, a BDF-solver of order 2 with fixed stepsize \( 10^{-5} \) and a tolerance of \( 10^{-8} \) for the Newton iteration. The resulting global errors at \( t = 1.5 \) were

\(|u_1^* - u_1^*| = 0.560(-9), |u_2^* - u_2^*| = 0.791(-8), |w - w^*| = 0.112(-6)|

When DAEN2 was applied to (34) with accurate starting values and a relative tolerance \( RTOL = 10^{-8} \), the terminal point at \( t = 1.5 \) was reached in 38 RK-steps (no rejections) and the final absolute, global errors were

\(|u_1 - u_1^*| = 2.734(-10), |u_2 - u_2^*| = 3.115(-9), |w - w^*| = 5.476(-8)|

More illuminating are the relative, global errors

\(|u_1 - u_1^*|/|u_1| = 3.865(-9), |u_2 - u_2^*|/|u_2^*| = 1.176(-9), |w - w^*|/|w^*| = 3.884(-9)\),
which show no deterioration in the error of the algebraic variable.

The second example is the trajectory-prescribed-path control problem of [3] which is discussed in detail in [4]. We refer to the latter reference for the equations of this index two problem. It involves six differential variables, \( u_1, \ldots, u_6 \), two algebraic variables \( w_1, w_2 \), both of which occur nonlinearly. The system consists of six differential and two algebraic equations. At \( t = 0 \) the initial conditions

\[
\mathbf{u}(0) = \frac{1}{4} (0.0, 100000.0, 0.0, 12000.0, -1.0, \pi)^	op
\]

were imposed where \( u_1, u_3, u_5, u_6 \) are in radians and \( u_2, u_4 \) in feet. The remaining initial values

\[
\begin{align*}
\mathbf{w}(0) &= (0.046650383, -0.91122917(-3))^	op \\
\mathbf{u}'(0) &= (0.40394369(-3), -209.42888, 0.40394369(-3), -34.978260, 0.0, 0.0)^	op
\end{align*}
\]

where computed by DAEN1 and are in agreement with those given in [4]. With a relative tolerance \( RTOL = 10^{-8} \) the system was integrated to \( t = 300 \). The final point was reached in 44 RK-steps, including 3 rejections, and, as Table 4.1 shows, agrees very closely with the reference values cited in [4].

<table>
<thead>
<tr>
<th></th>
<th>DAEN2</th>
<th>BCP89</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>0.0727991598693</td>
<td>0.0727991600</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>14200.8114923</td>
<td>14200.8114</td>
</tr>
<tr>
<td>( u_3 )</td>
<td>0.0406923052169</td>
<td>0.0406923053</td>
</tr>
<tr>
<td>( u_4 )</td>
<td>1433.29213943</td>
<td>1433.29213</td>
</tr>
<tr>
<td>( u_5 )</td>
<td>-0.174532925199</td>
<td>-0.174532925</td>
</tr>
<tr>
<td>( u_6 )</td>
<td>2.35619449019</td>
<td>2.35619449</td>
</tr>
<tr>
<td>( w_1 )</td>
<td>0.124888510970</td>
<td>0.124888511</td>
</tr>
<tr>
<td>( w_1 )</td>
<td>0.460375007767</td>
<td>0.460375012</td>
</tr>
</tbody>
</table>

Table 4.1

As an example for DAEQ2 we use the index-two DAE proposed in [1]

\[
\begin{align*}
\mathbf{u}' + a(t-2)w &= \left[a + 1/(t-2)\right]u_1 + [(3-t)/(2-t)]e^t \\
\mathbf{u}_2' - (a-1)w &= [(a-1)/(t-2)]u_1 + u_2 - 2e^t \\
0 &= (2+t)u_1 + (t^2 - 4)u_2 - (t^2 + t - 2)e^t.
\end{align*}
\] (35)

For \( a = 50 \) and the initial conditions \( \mathbf{u}(0) = (1, 1) \), the exact solution is \( u_1(t) = u_2(t) = \exp t, w = (\exp t)/(t-2) \). With a relative tolerance \( RTOL = 10^{-8} \), DAEQ2 reached \( t = 1.0 \) in 202 RK-steps (no rejections) and the maximum norm of the final, absolute
error, was $0.4(-9)$. When run with DAEN2, (35) happens to be one of the few problems where it does make a difference what updating strategy is used for the iteration matrix in the chord-Newton process. When the iteration matrix is retained throughout an entire RK-step, then DAEN2 required 328 RK-steps (including 97 rejections) to reach $t = 1.0$. On the other hand, as expected, the performance of DAEN2 was identical to that of DAEQ2 when the iteration matrix is updated for each chord-Newton iteration.

As a first example for DAEN1 we used the simple problem

$$u^2 + u'^2 - 1 = 0, \quad 2uu' - w = 0. \quad (36)$$

For the initial conditions $u(0) = 0$, $u'(0) = 1$, $w(0) = 0$ it has the exact solution $u(t) = \sin t$, $w(t) = \sin 2t$ which represents a horizontal figure eight in the plane and hence is not a submanifold of $\mathbb{R}^2$. Note that here the matrix $(D_p F(u, w) \ D_w F(u, w))$ of Assumption N2 is nonsingular on $M_1$ except at the point $(u, p, w) = (1, 0, 0)$ which turns out to be a removable singularity of $M_1$. In all runs, DAEN1 stepped over this point when it was reached for $t = (2k + 1)\pi/2$, $k = 0, 1, \ldots$. In fact, the absolute error, in the maximum norm, remained about of the same order of magnitude as the chosen relative tolerance $RTOL = 10^{-8}$. On the other hand, DASSL (see [4]) always terminated just before $t = \pi/2$.

A second example is the batch reactor model given in [2]

$$\begin{align*}
u_1' + k_2 u_2 w_2 &= 0 \\
u_2' + k_1 u_2 u_6 - k_{-1} w_4 + k_2 u_2 w_2 &= 0 \\
u_3' - k_3 u_2 w_2 - k_3 u_4 u_6 + k_{-3} w_3 &= 0 \\
u_4' + k_1 u_4 u_6 - k_{-3} w_3 &= 0 \\
u_5' - k_1 u_2 u_6 + k_{-1} w_4 &= 0 \\
u_6' + k_1 u_2 u_6 + k_3 u_4 u_6 - k_{-1} w_4 - k_{-3} w_3 &= 0 \\
u_6 - w_1 + w_2 + w_3 + w_4 - a &= 0 \\
w_2 - (K_2 u_1)/(K_2 + w_1) &= 0 \\
w_3 - (K_3 u_3)/(K_3 + w_1) &= 0 \\
w_4 - (K_1 u_5)/(K_1 + w_1) &= 0,
\end{align*}$$

where

$$\begin{align*}
k_1 &= 21.893, k_{-1} = 2.14(+9), k_2 = 32.318, k_3 = 21.893, k_{-3} = 1.07(+9), \\
K_1 &= 7.65(-18), K_2 = 4.03(-11), K_3 = 5.32(-18), a = 0.0131
\end{align*}$$

The initial conditions are

$$\begin{align*}
u(0) &= (1.5776, 8.32, 0.0, 0.0, 0.0, 0.0, 0.0)^	op, \\
w(0) &= (0.79735161(-5), 0.79735161(-5), 0.0, 0.0)^	op,
\end{align*}$$
and the values of the derivatives were determined by the differential equations. For this problem, DASSL did not start while DAEN1 reached $t = 1.0$ [hr] with 229 RK-steps (including 30 rejections). Problems of this type suggest the need for introducing also an implicit RK-solver into these DAE-codes.

Examples for use of DAESQ1 were given in [11] and for the application of DAEUL3 we refer to [12] and the comparative study [15]. DAEQ3 also runs for these Euler-Lagrange examples. Instead, we consider here only the following example for DAEQ3 which is not of Euler-Lagrange type

$$
\begin{align*}
    u_1'' + u_1w &= (1 + \sin t)e^t \\
    u_2'' + u_2w &= 2/(1+t)^2 + [1/(1+t)]\sin t \\
    0 &= e^t/(1+t)^3 - 2e^t/(1+t)^2 - e^t/(1+t)/(1 + t) - u_1u_2^3 + 3u_1u_2^2 - u_2e^u_2 \\
\end{align*}
$$

For the initial conditions $u(0) = (1, 1)^\top$, $u'(0) = (1, -1)^\top$ the problem has the solution $u_1^* = \exp t$, $u_2^* = 1/(1+t)$, $w^* = \sin t$. But in this case the matrix (16) of Assumption N3 has the form

$$
\begin{pmatrix}
    1 & 0 \\
    0 & 1 \\
    -u_2^3 + 3u_2^2 & -3u_1u_2^2 + 6u_1u_2 - u_2e^u_2 - e^u_2
\end{pmatrix}
$$

and, on the exact solution, this matrix becomes singular at $t^* \approx 0.03756275$. When started at $t = 0.0$, DAEQ3 produces a solution for which, near $t^*$, the algebraic variable tends to $-\infty$ while the other variables remain close to their exact values. As an illustration, Table 4.2 shows both the solution obtained by the code at $t = 0.037562749184$ and the corresponding exact solution.

<table>
<thead>
<tr>
<th></th>
<th>DAEQ3</th>
<th>Exact Sol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>1.0382771509</td>
<td>1.0382771461</td>
</tr>
<tr>
<td>$u_2$</td>
<td>0.96379713463</td>
<td>0.96379713014</td>
</tr>
<tr>
<td>$w$</td>
<td>-25014116.226</td>
<td>0.037553916550</td>
</tr>
</tbody>
</table>

Table 4.2

The same behavior is seen when the code is run backwards from $t = 0.05$ (with exact starting values), except that now, as expected, $w(t)$ tends to $+\infty$. On the other hand, when run forward from $t = 0.05$ the computed solution approximates the exact solution with an absolute error of the same magnitude as the given tolerance.

Singularities of quasilinear, second order ODEs have been studied in [6]. A discussion of the relation between these results and the observed singularity for (4.4) cannot be given here. It is noteworthy that the singularity is not apparent in the given DAE.
REFERENCES

References


