Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Integration of stiff mechanical systems by Runge-Kutta methods

Ch. Lubich

Research Report No. 92-04
May 1992

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Submitted to ZAMP

Integration of stiff mechanical systems by Runge-Kutta methods

Ch. Lubich

Research Report No. 92-04
May 1992

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Submitted to ZAMP

Abstract

The numerical integration of stiff mechanical systems is studied in which a strong potential forces the motion to remain close to a manifold. The equations of motion are written as a singular singular perturbation problem with a small stiffness parameter ϵ. Smooth solutions of such systems are characterized, in distinction to highly oscillatory general solutions. Implicit Runge-Kutta methods using step sizes larger than ϵ are shown to approximate smooth solutions, and precise error estimates are derived. As $\epsilon \to 0$, Runge-Kutta solutions of the stiff system converge to Runge-Kutta solutions of the associated constrained system formulated as a differential-algebraic equation of index 3. Standard software for stiff initial-value problems does not work satisfactorily on the stiff systems considered here. The reasons of this failure are explained, and remedies are proposed.

Key words: stiff mechanical system, stiff ODE, singular singular perturbation problem, differential-algebraic equations, Runge-Kutta methods.

Subject Classification: 65L05, 70F20, 70H35.

This work was supported in part by the Austrian Science Foundation, grant P8443-PHY.

1. Introduction

In this article we study the numerical solution of the equations of motion of mechanical systems in which a strong potential forces the motion to remain close to a manifold. We defer the general formulation to Section 2, and consider in this introductory section a simple, yet instructive example: a plane stiff spring pendulum, consisting of a mass point suspended on a massless spring with Hooke's constant $1/\epsilon^2$, where ϵ is a small parameter. Assuming for simplicity unit mass, rest position of the spring at unit length, and unit gravity, the equations of motion in Cartesian coordinates (y_1, y_2) are given by

$$(1.1) \qquad \begin{aligned} \ddot{y}_1 &= -\frac{1}{\epsilon^2}\frac{y_1}{\sqrt{y_1^2+y_2^2}}\left(\sqrt{y_1^2+y_2^2}-1\right) \\ \ddot{y}_2 &= -\frac{1}{\epsilon^2}\frac{y_2}{\sqrt{y_1^2+y_2^2}}\left(\sqrt{y_1^2+y_2^2}-1\right)-1 \end{aligned} \qquad 0<\epsilon\ll 1 .$$

¿From the viewpoint of singular perturbation theory, this is a *singular* singularly perturbed problem [17]. Physical intuition as well as rigorous analysis, see [15], p.11ff., show that for initial values having bounded energy, the motion of system (1.1) is close to that of the associated constrained system. This is the fixed-length pendulum, whose equations of motion in Cartesian coordinates are most easily formulated as a differential-algebraic system of index 3 [2],[7]:

$$(1.2) \qquad \begin{aligned} \ddot{y}_1 &= -y_1\lambda \\ \ddot{y}_2 &= -y_2\lambda - 1 \\ 0 &= y_1^2+y_2^2-1 . \end{aligned}$$

An analogous property will be shown for the numerical solution of (1.1) by an implicit Runge-Kutta method satisfying mild stability conditions: As $\epsilon \to 0$, Runge-Kutta solutions of the stiff system (1.1) obtained with step sizes $h > \epsilon$ converge to Runge-Kutta solutions of the index-3 differential-algebraic system (1.2). It will be seen that this behavior is at the same time satisfactory and the cause of numerical difficulties.

Systems like (1.1) arise in the modeling of mechanical systems when stiff force elements, e.g. springs or elasticity in joints, are taken as such, rather than treating them as kinematic constraints. Such systems are also obtained when almost rigid bodies are modeled as elastic bodies. On the other hand, it has been a popular approach among mechanical engineers to replace constrained mechanical systems by stiff systems like (1.1), possibly with additional damping. This is done in hope for the bonus that should result from the numerical treatment of an ODE (albeit stiff) instead of a DAE. Work on such penalization methods in a numerical analysis context has been done in [14],[13],[5],[10].

Our interest here is in the numerical solution of stiff systems (1.1) and generalizations thereof, using step sizes larger than the stiffness parameter ϵ. We then obtain approximations of *smooth* solutions of (1.1), without following the high-frequency, small-amplitude oscillations (vibrations) present in general solutions of (1.1). We begin in Section 2 by describing the general framework of stiff mechanical systems considered in this paper, and characterize smooth solutions of such systems. In Section 3 we state the main results on Runge-Kutta methods applied to stiff mechanical systems. Theorem 3.1 deals with the case of starting values admitting a smooth solution according to the characterization of Section 2. Here the error of the Runge-Kutta approximation of the stiff system is shown to be essentially equal to the error of the Runge-Kutta approximation of the associated constrained system in index-3 differential-algebraic form, up to a perturbation of size ϵ^2 times some nonnegative power of the step size. For starting values near a smooth solution, but not on it, Theorem 3.2 shows that the corresponding Runge-Kutta solution tends rapidly towards one which started on a smooth solution. The proof of these results is spread over Sections 4 to 7, whose contents are described at the end of Section 3. In these sections, further properties of Runge-Kutta methods are also investigated.

Standard solvers for stiff ODEs do not work well when applied to stiff systems like (1.1). This failure is explained in Section 8, and remedies are proposed. It is seen, in particular, that the numerical solution of the stiff system is computationally at least as expensive as that of the associated constrained system.

We have not considered mechanical systems with strong damping in this paper. Such systems can, however, also be studied by the techniques used here, and similar results can be obtained. In particular, it can be shown that for strongly damped systems the numerical solution approaches that of a differential-algebraic system of index 2.

2. Smooth motion of stiff mechanical systems

When we consider a mechanical system where strong conservative forces penalize some directions of motion, we are led to study the second order differential equation with a small parameter ϵ,

$$M(y)\,\ddot{y} = f(y,\dot{y}) - \frac{1}{\epsilon^2}\nabla U(y)\ , \tag{2.1}$$

where the solution $y(t) \in \mathbf{R}^n$ is to be sought on an ϵ-independent bounded interval $[0,T]$. We have denoted $\nabla U = (\partial U/\partial y)^T$ the gradient of the potential $U : \mathbf{R}^n \to \mathbf{R}$. We assume throughout that $M : \mathbf{R}^n \to \mathbf{R}^{n\times n}$, $f : \mathbf{R}^{n+n} \to \mathbf{R}^n$, and U have sufficiently many bounded derivatives. Our further assumptions are as follows:

(2.2) $\qquad M(y)$ is symmetric and positive definite for every $y \in \mathbf{R}^n$.

U attains a (local) minimum on a d-dimensional manifold $\mathcal{U}$:

(2.3a) For some region $D \subset \mathbf{R}^n$,

$$\mathcal{U} = \{u \in D : U(u) = \min_{y \in D} U(y)\} = \{u \in D : \nabla U(u) = 0\}$$

In a neighborhood of $\mathcal{U}$, U is strongly convex along directions non-tangential to $\mathcal{U}$, i.e., there exists $\alpha > 0$ such that for $u \in \mathcal{U}$

$$v^T \nabla^2 U(u) \, v \geq \alpha \cdot v^T M(u) \, v \qquad \text{(2.3b)}$$

for all vectors v in the $M(u)$-orthogonal complement of the tangent space $T_u\mathcal{U}$.

We always let

$$m = n - d$$

be the number of independent constraints that locally describe the manifold $\mathcal{U}$.

A simple example of a potential U satisfying (2.3) with $\mathcal{U} = \mathbf{R}^d \times 0$ is given by

$$U(y) = \tfrac{1}{2}\|y^{\perp}\|^2 \qquad \text{for} \quad y = (y^{\|}, y^{\perp}) \in \mathbf{R}^n = \mathbf{R}^d \times \mathbf{R}^m ,$$

with the Euclidian norm $\|\cdot\|$. As the following lemma states, this is already the general situation in suitable local coordinates.

Lemma 2.1. *Let $u \in \mathcal{U}$ be given. There exists a change of local coordinates $y = y(z)$, with $y(0) = u$ and as often continuously differentiable as $\nabla^2 U$, such that*

$$U(y(z)) = \tfrac{1}{2}\|z^{\perp}\|^2 + Const. \qquad \textit{for} \quad z = (z^{\|}, z^{\perp}) \in \mathbf{R}^d \times \mathbf{R}^m \quad \textit{near} \quad 0 .$$

Proof. In a first step, we choose local coordinates $x = (x^{\|}, x^{\perp}) \in \mathbf{R}^d \times \mathbf{R}^m$ near 0, such that $y(0) = u$, and that $y(x) \in \mathcal{U}$ if and only if $x^{\perp} = 0$. In these coordinates, we then have for all $x^{\|}$ near 0 that $\dfrac{\partial U}{\partial x}(x^{\|}, 0) = 0$ by (2.3a), and $A(x^{\|}) := \dfrac{\partial^2 U}{(\partial x^{\perp})^2}(x^{\|}, 0)$ is positive definite by (2.3b). We now change coordinates

$$w^{\|} = x^{\|} , \quad w^{\perp} = \omega(x) x^{\perp}$$

where $\omega(x) \in \mathbf{R}$ is to be chosen such that we get

$$\tfrac{1}{2}(w^{\perp})^T A(w^{\|}) w^{\perp} = U(y(x)) - U(y(0)) .$$

Since the right-hand side equals

$$U(y(x^{\|}, x^{\perp})) - U(y(x^{\|}, 0)) = \tfrac{1}{2}(x^{\perp})^T A(x^{\|}) x^{\perp} + r(x)$$

with $r(x) = O(\|x^{\perp}\|^3)$, we have

$$\omega(x) = \sqrt{1 + \frac{2r(x)}{(x^{\perp})^T A(x^{\|}) x^{\perp}}} .$$

Let $A = LL^T$ be the Choleski decomposition of A. In the coordinates

$$z^{\|} = w^{\|} , \quad z^{\perp} = L^T(w^{\|}) \cdot w^{\perp}$$

the potential U then has the desired form. $\square$

Remark. Lemma 2.1 could also be obtained as a corollary of the Morse lemma, see [4] for a particularly transparent presentation, or [1]. The above proof, however, appears more elementary. □

In terms of the local coordinates z, the system (2.1) is of the form

$$\widehat{M}(z)\,\ddot{z} = \widehat{f}(z,\dot{z}) - \frac{1}{\epsilon^2}\begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix} z \tag{2.5}$$

with $\widehat{M}(z) = (\partial y/\partial z)^T M(y(z))(\partial y/\partial z)$. This system is again of the type (2.1)-(2.3). Since numerical methods are not invariant under this transformation, we shall nevertheless continue to consider the original system (2.1). The representation (2.5) will, however, be useful for deriving stability estimates, because z appears only linearly in the term divided by ϵ^2.

In the following it will be of interest to characterize *smooth* solutions of (2.1), that is, solutions with sufficiently many derivatives bounded independently of ϵ.

Theorem 2.2. (Smooth motion) *Assume (2.2)-(2.3). For every* $(y^0, \dot{y}^0)$ *in the tangent bundle of* $\mathcal{U}$ *there exists a pair* $(y^\epsilon, \dot{y}^\epsilon)$*, unique up to* $O(\epsilon^{2N})$ *for arbitrary* N*, with differences* $y^\epsilon - y^0$ *,* $\dot{y}^\epsilon - \dot{y}^0$ *of magnitude* $O(\epsilon^2)$ *and situated in the* $M(y^0)$*-orthogonal complement of the tangent space* $T_{y^0}\mathcal{U}$*, such that the solution of (2.1) with initial values* $(y^\epsilon, \dot{y}^\epsilon)$ *is smooth and of the form*

$$\begin{aligned} y(t) &= y^0(t) + \epsilon^2 y^1(t) + \cdots + \epsilon^{2N} y^N(t) + O(\epsilon^{2N+2}) \\ \dot{y}(t) &= \dot{y}^0(t) + \epsilon^2 \dot{y}^1(t) + \cdots + \epsilon^{2N} \dot{y}^N(t) + O(\epsilon^{2N+2}) \end{aligned} \tag{2.6}$$

with ϵ*-independent functions* $y^k(t)$ *. This solution exists on an* ϵ*-independent interval* $[0,T]$*. The collection of all such pairs* $(y^\epsilon, \dot{y}^\epsilon)$ *forms a* $2d$*-dimensional manifold* $\mathcal{M}^\epsilon$*. Solutions of (2.1) starting in* $\mathcal{M}^\epsilon$ *remain in* $\mathcal{M}^\epsilon$*, up to* $O(\epsilon^{2N})$ *on bounded time intervals.*

Proof. We will first construct a truncated expansion (2.6) such that its residual in (2.1) is small, and then conclude with the help of a stability estimate. It is clearly sufficient to prove the result for the special choice of coordinates in Lemma 2.1. This will be convenient for the derivation of the stability estimate, but for the construction of the expansion coefficients we prefer to work with the original coordinates in (2.1).

(a) Since $\mathcal{U}$ is assumed to be a d-dimensional submanifold of $\mathbf{R}^n$, locally there exist constraint functions $g_1, \ldots, g_m : \mathbf{R}^n \to \mathbf{R}$ such that

$$\text{locally:} \quad u \in \mathcal{U} \quad \text{iff} \quad g_1(u) = \cdots = g_m(u) = 0 \ , \tag{2.7}$$

and such that the gradients $\nabla g_i(u)$ are linearly independent for $u \in \mathcal{U}$. Because of (2.3a), a vector v is in the tangent space $T_u\mathcal{U}$, if and only if

$$H(u)v = 0 \ , \tag{2.8}$$

with the Hessian $H = \nabla^2 U$. On the other hand, by (2.7) $v \in T_u\mathcal{U}$ if and only if

$$G(u)v = 0 \ , \tag{2.9}$$

where $G = \partial g/\partial y$ with $g = (g_1, \ldots, g_m)^T$ has linearly independent rows. (In what follows, G might be chosen also as any other $m \times n$-matrix with the same null-space as H.) H can then be written as

$$H = G^T K G \ , \tag{2.10}$$

with an invertible (even symmetric and positive definite) $m \times m$-matrix K. This is seen as follows: (2.9) is equivalent to

$$P(u)v = 0 \ ,$$

where P is the orthogonal projection $P = G^T(GG^T)^{-1}G$. Because of (2.8) we have (omitting the argument u)

$$v^T H w = v^T \, PHP \, w \qquad \text{for all} \quad v, w \in \mathbf{R}^n \ ,$$

and hence

$$H = PHP \ ,$$

which upon inserting the definition of P becomes (2.10). The $m \times m$-matrix K must be invertible, because H has rank m as a consequence of the equivalence of (2.8) and (2.9).

(b) The coefficients $y^k(t)$ in (2.6) will now be constructed recursively by comparing powers of ϵ in (2.1). The coefficient of ϵ^{-2} vanishes iff $\nabla U(y^0) = 0$, or equivalently, if

$$g(y^0) = 0 \ . \tag{2.11}$$

The coefficient of ϵ^0 vanishes iff

$$M(y^0)\,\ddot{y}^0 = f(y^0, \dot{y}^0) - H(y^0)y^1 \ .$$

Here, however, y^1 cannot yet be determined. Remembering (2.10), we introduce λ^0 by the condition

$$H(y^0)y^1 = G^T(y^0)\lambda^0 \ , \tag{2.12}$$

so that we get the equation

$$M(y^0)\,\ddot{y}^0 = f(y^0, \dot{y}^0) - G^T(y^0)\lambda^0 \ , \tag{2.13}$$

which together with (2.11) represents the Lagrange equations of motion of the mechanical system whose position is constrained to the manifold $\mathcal{U}$. This is a differential-algebraic system of index 3 (see, e.g., [2],[7]) which has a unique solution for every initial value $(y^0(0), \dot{y}^0(0))$ in the tangent bundle $T\mathcal{U}$.

(c) By (2.10) and because G has linearly independent rows, equation (2.12) is equivalent to

$$G(y^0)y^1 = K^{-1}(y^0)\lambda^0 \ . \tag{2.14}$$

The coefficient of ϵ^2 in (2.1) vanishes iff

$$M(y^0)\,\ddot{y}^1 = \phi^1(y^0,\dot{y}^0,\ddot{y}^0,y^1,\dot{y}^1) - G^T(y^0)\lambda^1 \ , \tag{2.15}$$

where

$$\begin{aligned}\phi^1(y^0,\dot{y}^0,\ddot{y}^0,y^1,\dot{y}^1) &= f_y(y^0,\dot{y}^0)y^1 + f_{\dot{y}}(y^0,\dot{y}^0)\dot{y}^1 \\ &\quad - \tfrac{1}{2}\nabla U_{yy}(y^0)\cdot(y^1,y^1) - \frac{\partial}{\partial y}\left(M(y)\ddot{y}^0\right)|_{y=y^0}\cdot y^1 \ ,\end{aligned}$$

and where the Lagrange multiplier λ^1 is introduced via

$$H(y^0)y^2 = G^T(y^0)\lambda^1. \tag{2.16}$$

If $y^0,\dot{y}^0,\ddot{y}^0$ are considered known, then (2.14), (2.15) is again an index-3 differential-algebraic system for $y^1,\dot{y}^1,\lambda^1$. Moreover, the initial values $y^1(0),\dot{y}^1(0)$ are determined uniquely if it is required that they both lie in the $M(y^0(0))$-orthogonal complement of the tangent space $T_{y^0(0)}\mathcal{U}$, i.e., in the range of $M^{-1}G^T(y^0(0))$.

Now (2.16) can be rewritten as

$$G(y^0)y^2 = K^{-1}(y^0)\lambda^1 \ ,$$

and comparison of the coefficients of ϵ^4 gives another index-3 differential-algebraic system for $y^2,\dot{y}^2,\lambda^2$. In this way we can continue to construct $y^k,\dot{y}^k$ such that the defect of the truncated expansion (2.6) inserted into (2.1) is of magnitude $O(\epsilon^{2N})$, for arbitrarily chosen N.

(d) To simplify matters, we assume from now on without loss of generality, that the equations are already formulated in the coordinates of Lemma 2.1, i.e., that

$$\nabla U(y) = \begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix} y \ .$$

In the previous parts of the proof we have constructed a truncated expansion

$$\eta(t) = y^0(t) + \epsilon^2 y^1(t) + \cdots + \epsilon^{2N} y^N(t) \tag{2.17}$$

such that the defect of (2.17) inserted into (2.1) is small:

$$M(\eta)\ddot{\eta} = f(\eta,\dot{\eta}) - \frac{1}{\epsilon^2}\begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix}\eta + O(\epsilon^{2N+2}) + \begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix} O(\epsilon^{2N}) \ . \tag{2.18}$$

We will show that every solution of (2.1), whose starting values satisfy

$$y(0) - \eta(0) = O(\epsilon^{2N+1}) \ , \quad \dot{y}(0) - \dot{\eta}(0) = O(\epsilon^{2N}) \ , \tag{2.19}$$

also satisfies

$$y(t) - \eta(t) = O(\epsilon^{2N}) \ , \quad \dot{y}(t) - \dot{\eta}(t) = O(\epsilon^{2N}) \ , \tag{2.20}$$

uniformly for all t on bounded intervals. With the abbreviations $M(t) = M(\eta(t))$ and $\Delta y = y - \eta$ we have by (2.1) and (2.18)

$$M(t)\Delta\ddot{y} = -\frac{1}{\epsilon^2}\begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}\Delta y + O(\Delta y) + O(\Delta\dot{y}) + O(\epsilon^{2N+2}) + \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} O(\epsilon^{2N}) \ , \tag{2.21}$$

as long as $\Delta y = O(\epsilon^2)$. With the transformation

$$\Delta x(t) = M^{1/2}(t)\Delta y(t) \ , \quad A(t) = M^{-1/2}(t)\begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} M^{-1/2}(t)$$

we have

$$\Delta\ddot{x} = -\frac{1}{\epsilon^2}A(t)\Delta x + O(\Delta x) + O(\Delta\dot{x}) + O(\epsilon^{2N+2}) + A(t)\cdot O(\epsilon^{2N}) \ .$$

Next we use a block diagonalization which separates the zero and non-zero eigenvalues of $A(t)$:

$$Q^T(t)A(t)Q(t) = \begin{pmatrix} 0 & 0 \\ 0 & B(t) \end{pmatrix} \ ,$$

with orthogonal matrix $Q(t)$ and with positive definite $m \times m$-matrix $B(t)$. Here, B and Q can be chosen as smooth functions, cf. Theorem II.5.11 in [11], p.115. (In contrast, the diagonalization of $B(t)$ may not be possible with a continuous (let alone twice continuously differentiable) transformation matrix, if eigenvalues of $B(t)$ coincide for some t. See Rellich's example in [11], p.110.) With the change of variables

$$Q^T\Delta x = \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}$$

we thus get

$$\begin{aligned} \Delta\ddot{u} &= O(\|\Delta u\| + \|\Delta\dot{u}\| + \|\Delta v\| + \|\Delta\dot{v}\|) + O(\epsilon^{2N+2}) \\ \Delta\ddot{v} &= -\tfrac{1}{\epsilon^2}B(t)\Delta v + O(\|\Delta u\| + \|\Delta\dot{u}\| + \|\Delta v\| + \|\Delta\dot{v}\|) + O(\epsilon^{2N}) \ . \end{aligned} \tag{2.22}$$

$B(t)$ has a smooth Choleski decomposition,

$$B(t) = L(t)L^T(t) \ .$$

Upon introducing $\Delta w = (\Delta v, \epsilon L^{-1} \Delta \dot{v})^T$, the second equation in (2.22) takes a first-order form with a skew-symmetric matrix divided by ϵ:

$$\Delta \dot{w} = \frac{1}{\epsilon} \begin{pmatrix} 0 & L(t) \\ -L^T(t) & 0 \end{pmatrix} \Delta w + O(\|\Delta w\| + \epsilon \|\Delta u\| + \epsilon \|\Delta \dot{u}\|) + O(\epsilon^{2N+1}) \ .$$

The usual energy estimate now gives

$$\begin{aligned} \|\Delta w\| \cdot \frac{d}{dt} \|\Delta w\| &= \tfrac{1}{2} \frac{d}{dt} \|\Delta w\|^2 = \Delta w^T \Delta \dot{w} \\ &= O\left(\|\Delta w\| \cdot (\|\Delta w\| + \epsilon \|\Delta u\| + \epsilon \|\Delta \dot{u}\| + O(\epsilon^{2N+1})) \right) \ , \end{aligned}$$

and using also $\Delta w(0) = O(\epsilon^{2N+1})$, Gronwall's inequality yields

$$\|\Delta w(t)\| \le C \epsilon \max_{0 \le \tau \le t} (\|\Delta u(\tau)\| + \|\Delta \dot{u}(\tau)\|) + O(\epsilon^{2N+1}) \ .$$

Reinserting this into the first equation of (2.22) then gives $\Delta u = O(\epsilon^{2N})$, $\Delta \dot{u} = O(\epsilon^{2N})$, and thus finally the desired bound (2.20).

(e) It remains to see that a solution y of (2.1) satisfying (2.20) is actually *smooth*. Subtracting (2.18) from (2.1) and using (2.20) gives us that

$$\ddot{y} - \ddot{\eta} = O(\epsilon^{2N-2}) \ . \tag{2.23}$$

Differentiating (2.1) and (2.18) once with respect to time and using once more (2.20) and also (2.23) shows that also

$$\dddot{y} - \dddot{\eta} = O(\epsilon^{2N-2}) \ . \tag{2.24}$$

Differentiating (2.1) and (2.18) further and using the previously obtained estimates of derivatives of y gives us subsequently

$$\begin{aligned} y^{(4)} - \eta^{(4)} &= O(\epsilon^{2N-4}) \ , \quad \dot{y}^{(4)} - \dot{\eta}^{(4)} = O(\epsilon^{2N-4}) \ , \\ y^{(6)} - \eta^{(6)} &= O(\epsilon^{2N-6}) \ , \quad \dot{y}^{(6)} - \dot{\eta}^{(6)} = O(\epsilon^{2N-6}) \ , \end{aligned}$$

and so on. This shows the smoothness of the constructed solution (2.6). □

For later reference, we recollect the differential-algebraic equations satisfied by the coefficients y^k in the ϵ^2-expansion (2.6) of a smooth solution of (2.1): $(y^0, \dot{y}^0)$ is obtained from

$$\begin{aligned} M(y^0)\, \ddot{y}^0 &= f(y^0, \dot{y}^0) - G(y^0)^T \lambda^0 \\ g(y^0) &= 0 \end{aligned} \tag{2.25.0}$$

with $G = \partial g / \partial y$ of full row rank. For $k \ge 1$ we have

$$\begin{aligned} M(y^0)\, \ddot{y}^k &= \phi^k(y^0, \dot{y}^0, \ddot{y}^0, \ldots, y^{k-1}, \dot{y}^{k-1}, \ddot{y}^{k-1}, y^k, \dot{y}^k) - G(y^0)^T \lambda^k \\ G(y^0)\, y^k &= K^{-1}(y^0)\, \lambda^{k-1} \end{aligned} \tag{2.25.k}$$

with ϕ^k linear in $\dot{y}^k$.

We note that (2.25.0) is a differential-algebraic system of index 3 (see, e.g., [2], [7], [8]). Similarly, if all functions with superscripts up to $k-1$ are considered to be known exactly, then (2.25.k) is again of index 3. The system (2.25.0-k) together is however of index $2k+3$. We thus get a sequence of differential-algebraic equations of index 3,5,7,... .

3. Runge-Kutta methods for stiff mechanical systems: statement of main results

A Runge-Kutta method applied to a second-order differential (or differential-algebraic) equation, such as (2.1) or (2.25),

$$F(y,\dot y,\ddot y)=0\ , \tag{3.1}$$

yields approximations $(y_n,\dot y_n)$ to the solution values $(y,\dot y)$ at gridpoints t_n recursively via

$$y_{n+1}=y_n+h\sum_{j=1}^{s} b_j\dot Y_{nj}\ ,\quad \dot y_{n+1}=\dot y_n+h\sum_{j=1}^{s} b_j\ddot Y_{nj}\ , \tag{3.2a}$$

with internal stages (for $i=1,\dots,s$)

$$Y_{ni}=y_n+h\sum_{j=1}^{s} a_{ij}\dot Y_{nj}\ ,\quad \dot Y_{ni}=\dot y_n+h\sum_{j=1}^{s} a_{ij}\ddot Y_{nj}\ , \tag{3.2b}$$

which satisfy an equation of the form of (3.1),

$$F(Y_{ni},\dot Y_{ni},\ddot Y_{ni})=0\ . \tag{3.2c}$$

Here a_{ij} and b_j are the coefficients that determine the Runge-Kutta method. It will be convenient for the presentation to take a constant step size h. It should be noted, however, that our results do not depend on such an assumption.

Under appropriate stability conditions, it will be seen that the numerical solution $(y_n,\dot y_n)$ of the stiff mechanical system (2.1) converges for $\epsilon\to 0$ to a Runge-Kutta solution $(y_n^0,\dot y_n^0)$ of the associated constrained system in the index-3 formulation (2.25.0). These stability conditions (only) involve the *stability function* for the linear test equation $\dot y=\lambda y$ (see, e.g., [8]):

$$R(z)=1+zb^T(I-zA)^{-1}\mathbb{1}\ , \tag{3.3}$$

where $b^T=(b_1,\dots,b_s)$, $A=(a_{ij})_{i,j=1}^s$, and $\mathbb{1}=(1,\dots,1)^T$. We assume the following:

(3.4) A is invertible, and $R(\infty)=1-b^TA^{-1}\mathbb{1}$ has absolute value strictly smaller than 1.

(3.5) A has no eigenvalues on the imaginary axis, and
$$|R(i\omega)|<1 \quad \text{for all} \quad \omega\in\mathbf{R}\ ,\ \omega\neq 0\ .$$

Condition (3.5) can be dropped in all our results if we restrict our attention to $\epsilon\ll h$.

The Runge-Kutta method has *stage order* $q \geq 1$, if

$$\sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{c_i^k}{k} \qquad \text{for} \quad k = 1, \dots, q \qquad \text{and all } i \ . \tag{3.6}$$

(Here c_i is defined by (3.6) with $k = 1$.)
We assume throughout that the *order* p of the method when applied to nonstiff ordinary differential equations satisfies $p \geq q$.

Our first main result concerns numerical solutions which start on the manifold $\mathcal{M}^\epsilon$ of smooth motion, as characterized by Theorem 2.2. It will be convenient in the following to distinguish such solutions notationally by adding a superscript ϵ.

Theorem 3.1. *Let the Runge-Kutta method have stage order q and satisfy the stability conditions (3.4)-(3.5). Suppose that the starting value $(y_0^\epsilon, \dot y_0^\epsilon)$ is on the manifold $\mathcal{M}^\epsilon$ of Theorem 2.2, i.e., that the exact solution $(y^\epsilon(t), \dot y^\epsilon(t))$ of (2.1) with initial value $(y_0^\epsilon, \dot y_0^\epsilon)$ is smooth. For $0 < \epsilon \leq h \leq h_0$, with h_0 sufficiently small but independent of ϵ, there then exists a unique Runge-Kutta solution (3.2) of the stiff equation (2.1), whose error satisfies*

$$\begin{aligned} y_n^\epsilon - y^\epsilon(t_n) &= y_n^0 - y^0(t_n) + O(\epsilon^2 h^{q-2}) \\ \dot y_n^\epsilon - \dot y^\epsilon(t_n) &= \dot y_n^0 - \dot y^0(t_n) + O(\epsilon^2 h^{q-2}) \ , \end{aligned} \tag{3.7}$$

uniformly for $0 \leq t_n \leq T$. Here y_n^0 and $y^0(t)$ denote the Runge-Kutta and exact solution, respectively, of the index-3 differential-algebraic equation (2.25.0), where the starting value $(y_0^0, \dot y_0^0)$ is the coefficient of ϵ^0 in the ϵ^2-expansion of $(y_0^\epsilon, \dot y_0^\epsilon)$.

Since we have $y^\epsilon(t) = y^0(t) + O(\epsilon^2)$ and $\dot y^\epsilon(t) = \dot y^0(t) + O(\epsilon^2)$ by Theorem 2.2, we get in particular

$$y_n^\epsilon = y_n^0 + O(\epsilon^2) \ , \quad \dot y_n^\epsilon = \dot y_n^0 + O(\epsilon^2) \ ,$$

if $q \geq 2$. Theorem 3.1, however, gives a far sharper estimate. The errors $y_n^0 - y^0(t_n)$ and $\dot y_n^0 - \dot y^0(t_n)$ for the index-3 problem have been studied in [7] and [9]. They are at most $O(h^q)$, a sharper bound is restated in Theorem 4.1 below.

Since it is quite exceptional that the starting values lie on the manifold $\mathcal{M}^\epsilon$ of smooth motion, it is of interest to know how numerical solutions behave for starting values which are not in $\mathcal{M}^\epsilon$ but sufficiently close. The following theorem shows that such a Runge-Kutta solution rapidly approaches one which started in $\mathcal{M}^\epsilon$.

Theorem 3.2. *Let the Runge-Kutta method of stage order q satisfy the stability conditions (3.4)-(3.5). If the starting value $(y_0, \dot y_0)$ satisfies $\nabla U(y_0) = O(h^2)$ and $\nabla^2 U(y_0) \dot y_0 = O(h)$, then there exists $(y_0^\epsilon, \dot y_0^\epsilon) \in \mathcal{M}^\epsilon$ such that the corresponding Runge-Kutta solutions satisfy for $0 < \epsilon \leq h \leq h_0$ and $0 \leq t_n \leq T$*

$$\|y_n - y_n^\epsilon\| + \|\dot y_n - \dot y_n^\epsilon\| \leq \begin{cases} C \cdot (h\rho^n + \epsilon^{2k}) & \text{for} \quad q = 2k \\ C \cdot (h\rho^n + h\epsilon^{2k}) & \text{for} \quad q = 2k+1 \end{cases} \tag{3.8}$$

with $\rho < 1$. If $\epsilon \ll h$, then ρ can be chosen as any fixed number larger than $|R(\infty)|$.

Remarks. The condition $\nabla U(y_0) = O(h^2)$ may seem rather restrictive at first sight. It should however be noted that the potential energy in the mechanical system (2.1) is $\frac{1}{\epsilon^2}U(y) + O(1)$.
The *damping* in (3.8) is implied by conditions (3.4) and (3.5). In contrast, analytical solutions of (2.1) *oscillate* about $\mathcal{M}^\epsilon$. Note, however, that it is not at all obvious *a priori* that the damping conditions (3.4),(3.5) for the linear test equation $\ddot{y} = -\omega^2 y$ are sufficient for damping in the nonlinear problem (2.1).

Theorems 3.1 and 3.2 will be proved in the course of the following sections. To prove Theorem 3.1, we will take a route similar to [6]: In the ϵ^2-expansion of the numerical solution,

$$y_n^\epsilon = y_n^0 + \epsilon^2 y_n^1 + \epsilon^4 y_n^2 + \dots , \qquad \dot{y}_n^\epsilon = \dot{y}_n^0 + \epsilon^2 \dot{y}_n^1 + \epsilon^4 \dot{y}_n^2 + \dots , \tag{3.9}$$

the coefficients $(y_n^k, \dot{y}_n^k)$ are the Runge-Kutta solution of the differential-algebraic system (2.25.0-k). This comes as a direct consequence of the fact that (3.2c) is of the same form as the underlying differential equation, and of the linearity of the Runge-Kutta relations (3.2a,b). To obtain estimates of $y_n^\epsilon - y^\epsilon(t_n)$, we therefore study in Section 4 the errors $y_n^k - y^k(t_n)$ for the differential-algebraic systems (2.25) of index 3,5,7, ... The remainder terms in (3.9) will be bounded in Section 6, using the tools developed in Section 5. In that section we study existence and uniqueness of the numerical solution, and influence of perturbations and error propagation in the Runge-Kutta scheme applied to the stiff problem (2.1). Theorem 3.2 will be proved in Section 7, using the results of previous sections and an invariant manifold theorem of Kirchgraber et al. [12],[16].

4. Runge-Kutta approximation of differential-algebraic equations associated with the stiff problem

In this section we give bounds for the errors $y_n^k - y^k(t_n)$ and $\dot{y}_n^k - \dot{y}^k(t_n)$ of the Runge-Kutta method applied to the sequence of differential-algebraic equations (2.25). We begin with $k = 0$.

Theorem 4.1. [7],[9] (Error estimates for the differential-algebraic system of index 3) *Let the Runge-Kutta method have stage order $q \geq 1$ and satisfy (3.4). Then the error of the method applied to the differential-algebraic system (2.25.0) of index 3 satisfies*

$$y_n^0 - y^0(t_n) = O(h^q) , \qquad \dot{y}_n^0 - \dot{y}^0(t_n) = O(h^q) , \tag{4.1}$$

uniformly for $0 \leq t_n \leq T$. The error bound for y_n^0 can be improved for collocation methods of order $p > q$ with nodes $0 < c_1 < \cdots < c_s = 1$ (in particular, for Radau IIA methods, where $p = 2s - 1$):

$$y_n^0 - y^0(t_n) = O(h^p) . \tag{4.2}$$

Remarks. The bound (4.2), which was only conjectured in [7], p.86, has recently been proved by Jay [9]. The error bounds (4.1) were obtained in Theorem 6.4 of [7], p.78. There the result was formulated under the assumption $q \geq 2$ (thus excluding, for example, the backward Euler method and diagonally implicit RK methods). This condition is not needed here because equation (2.25.0) is linear in λ^0, cf. the discussion on p.74 of [7].

For the higher-index equations (2.25.0-k) with $k \geq 1$ we have the following result.

Theorem 4.2. (Error estimates for the differential-algebraic system of index $2k+3$) *Let $k \geq 1$. Let the Runge-Kutta method have stage order $q \geq 2k$ and satisfy (3.4). Then the error of the method applied to the differential-algebraic system (2.25) satisfies*

$$y_n^k - y^k(t_n) = O(h^{q-2k}) , \qquad \dot{y}_n^k - \dot{y}^k(t_n) = O(h^{q-2k}) , \tag{4.3}$$

uniformly for $0 \leq t_n \leq T$.

Proof. The proof relies on results and techniques from [7], Ch. 6. We consider first the case $k = 1$. The result for general k can then be obtained in the same way by an induction argument.

(a) We begin by studying the *local error* in y^1 and $\dot{y}^1$, i.e., the difference of the exact solution $(y^1(t_{n+1}), \dot{y}^1(t_{n+1}))$ to the Runge-Kutta solution which starts at $(y_n^0, \dot{y}_n^0, y^1(t_n), \dot{y}^1(t_n))$. It is convenient to first consider $n = 0$. The result after one Runge-Kutta step is obtained via (cf. (3.2))

$$y_1^1 = y_0^1 + h \sum_{j=1}^{s} b_j \dot{Y}_j^1 , \quad \dot{y}_1^1 = \dot{y}_0^1 + h \sum_{j=1}^{s} b_j \ddot{Y}_j^1 , \tag{4.4a}$$

with internal stages

$$Y_i^1 = y_0^1 + h \sum_{j=1}^{s} a_{ij} \dot{Y}_j^1 , \quad \dot{Y}_i^1 = \dot{y}_0^1 + h \sum_{j=1}^{s} a_{ij} \ddot{Y}_j^1 , \tag{4.4b}$$

related by (see (2.25.1))

$$\begin{aligned} M(Y_i^0)\ddot{Y}_i^1 &= \phi^1(Y_i^0, \dot{Y}_i^0, \ddot{Y}_i^0, Y_i^1, \dot{Y}_i^1) - G(Y_i^0)^T \Lambda_i^1 \\ G(Y_i^0) Y_i^1 &= K^{-1}(Y_i^0) \Lambda_i^0 . \end{aligned} \tag{4.4c}$$

We will show that

$$\begin{aligned} y_1^1 - y^1(h) &= O(h^{q-1}) , \qquad & P_0 \cdot (y_1^1 - y^1(h)) &= O(h^q) , \\ \dot{y}_1^1 - \dot{y}^1(h) &= O(h^{q-2}) , \qquad & P_0 \cdot (\dot{y}_1^1 - \dot{y}^1(h)) &= O(h^{q-1}) , \end{aligned} \tag{4.5}$$

where $P_0 = I - (M^{-1}G^T(KGM^{-1}G^T)^{-1}KG)(y_0^0)$ is a projection.

To this end we look at the defect when the exact solution $(y^1, \dot{y}^1)$ is inserted into (4.4). By (3.6), the defect in (4.4a,b) is not larger than $O(h^{q+1})$. In the second equation of (4.4c) we have the defect (after multiplying the equation with $K(Y_i^0)$)

$$(KG)(Y_i^0)\,y^1(c_ih) - \Lambda_i^0 = \big((KG)(y^0(c_ih)) - \lambda^0(c_ih)\big) \\ + \big((KG)(Y_i^0) - (KG)(y^0(c_ih))\big)\,y^1(c_ih) + \big(\lambda^0(c_ih) - \Lambda_i^0\big)\ .$$

This is $O(h^{q-1})$ because of (2.25.1) and the following error estimates for the internal stages in (2.25.0) which are shown in [7], p.77:

$$(4.6)\quad Y_i^0 - y^0(c_ih) = O(h^q)\ , \qquad \dot{Y}_i^0 - \dot{y}^0(c_ih) = O(h^q)\ , \qquad \Lambda_i^0 - \lambda^0(c_ih) = O(h^{q-1})\ .$$

By (2.25.0) and the corresponding equation (3.2c) these estimates immediately imply

$$(4.7)\qquad \ddot{Y}_i^0 - \ddot{y}^0(c_ih) = O(h^{q-1})\ ,$$

and so we get a defect of size $O(h^{q-1})$ also in the first equation of (4.4c). We now apply Theorem 6.2 of [7], p.75, with $q-2$, KG, $M^{-1}G^T$ in the roles of q, g_y, k_u there. Since the problem (2.25.1) is linear in $\dot{y}^1$ and λ^1 and the constraint is linear in y^1, it is sufficient to assume $q \geq 2$ (instead of $q \geq 4$). We thus get

$$(4.8)\quad Y_i^1 - y^1(c_ih) = O(h^{q-1})\ , \qquad \dot{Y}_i^1 - \dot{y}^1(c_ih) = O(h^{q-2})\ , \qquad \Lambda_i^1 - \lambda^1(c_ih) = O(h^{q-3})\ .$$

As in the proof of Lemma 6.3 of [7], p.77, we then get the desired bound (4.5) for the local error, for $n = 0$. For step numbers $n \geq 1$ the estimation of the local error is the same as above when one uses that the global errors of the internal stages Y_{ni}^0, $\dot{Y}_{ni}^0$, Λ_{ni}^0 are still bounded as in (4.6). This fact is seen from the proof of Theorem 6.4 in [7].

(b) The desired bounds (4.3) of the *global error* now follow with the proof of Theorem 6.4 of [7], which also gives the bounds

$$Y_{ni}^1 - y^1(t_n + c_ih) = O(h^{q-2})\ , \quad \dot{Y}_{ni}^1 - \dot{y}^1(t_n + c_ih) = O(h^{q-2})\ , \quad \Lambda_{ni}^1 - \lambda^1(t_n + c_ih) = O(h^{q-3})$$

to be used for the proof of $k = 2$. □

5. Runge-Kutta discretization of the stiff problem: basic properties

In this section we study existence and uniqueness of a Runge-Kutta solution of the stiff equations of motion (2.1), the influence of perturbations in the scheme, and error propagation. Our analysis relies in an essential way on various coordinate transforms, like that of Lemma 2.1 and those used in part (d) of the proof of Theorem 2.2. The Runge-Kutta method is not invariant under these transformations, but their use is, broadly speaking, that powers of ϵ can be gained in exchange for powers of h in the local estimates, and that "fast" and "slow" solution components are separated in the analysis of error propagation. These appear to be the principal rules in the game, in addition to some basic techniques of [6].

One step of the Runge-Kutta method (3.2) applied to the stiff problem (2.1) is reformulated as

$$y_1 = y_0 + h\dot{y}_0 + h^2 \sum_{j,k=1}^{s} b_j a_{jk} \ddot{Y}_k \ , \qquad \dot{y}_1 = \dot{y}_0 + h \sum_{j=1}^{s} b_j \ddot{Y}_j \ , \tag{5.1a}$$

with internal stages

$$Y_i = y_0 + c_i h \, \dot{y}_0 + h^2 \sum_{j,k=1}^{s} a_{ij} a_{jk} \ddot{Y}_k \ , \qquad \dot{Y}_i = \dot{y}_0 + h \sum_{j=1}^{s} a_{ij} \ddot{Y}_j \tag{5.1b}$$

satisfying

$$M(Y_i)\, \ddot{Y}_i = f(Y_i, \dot{Y}_i) - \frac{1}{\epsilon^2} \nabla U(Y_i) \ . \tag{5.1c}$$

Lemma 5.1. (Existence and local uniqueness) *Suppose that the Runge-Kutta matrix $A = (a_{ij})$ is invertible and has no eigenvalues on the imaginary axis. If $\nabla U(y_0) = O(h^2)$, $\nabla^2 U(y_0)\, \dot{y}_0 = O(h)$, then the scheme (5.1) has for $0 < \epsilon \le h$ a unique solution, with $\ddot{Y}_i = O(1)$ for $i = 1, \dots, s$. This holds for $h \le h_0$, where h_0 is sufficiently small but independent of ϵ.*

Proof. (a) We use the coordinate transform $y = y(z)$ of Lemma 2.1 and denote the inverse transform by $z = z(y)$. We set $z_0 = z(y_0)$,

$$Z_i = z(Y_i) \ , \quad \dot{Z}_i = \left(\frac{\partial z}{\partial y} \right)(y_0)\, \dot{Y}_i \ , \tag{5.2a}$$

and further $\ddot{Z}_i = \left(\frac{\partial z}{\partial y} \right)(y_0)\, \ddot{Y}_i$, i.e.,

$$\ddot{Y}_i = \left(\frac{\partial y}{\partial z} \right)(z_0)\, \ddot{Z}_i \ . \tag{5.2b}$$

In these variables (5.1c) becomes

$$\widetilde{M}(Z_i)\, \ddot{Z}_i = \widetilde{f}(Z_i, \dot{Z}_i) - \frac{1}{\epsilon^2} \begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix} Z_i \tag{5.3}$$

with $\widetilde{M}(z) = \left(\frac{\partial y}{\partial z} \right)^T (z)\, M(y(z)) \left(\frac{\partial y}{\partial z} \right)(z_0)$ and $\widetilde{f}(z, \dot{z}) = \left(\frac{\partial y}{\partial z} \right)^T (z)\, f(y(z), \left(\frac{\partial y}{\partial z} \right)(z_0) \dot{z})$. We now consider all unknowns as functions of $\ddot{Z}_i$, according to the scheme

$$\ddot{Z}_i \xrightarrow[(5.2b)]{} \ddot{Y}_i \xrightarrow[(5.1b)]{} Y_i, \dot{Y}_i \xrightarrow[(5.2a)]{} Z_i, \dot{Z}_i \ . \tag{5.4}$$

(b) Similarly to [7],[6], we consider $\ddot{Z}_i$ as a function of τ in the homotopy

(5.5)
$$\widetilde{M}(Z_i)\,\ddot{Z}_i = \widetilde{f}(Z_i,\dot{Z}_i) - \frac{1}{\epsilon^2}\begin{pmatrix}0 & 0\\ 0 & I_m\end{pmatrix} Z_i + (\tau-1)\left(\widetilde{f}(Z_i^{(0)},\dot{z}_0) - \frac{1}{\epsilon^2}\begin{pmatrix}0 & 0\\ 0 & I_m\end{pmatrix} Z_i^{(0)}\right) ,$$

where we have denoted $Z_i^{(0)} = z_0 + c_i h \dot{z}_0$. For $\tau = 0$ the solution is $\ddot{Z}_i|_{\tau=0} = 0$. We will show that a unique bounded solution of (5.5) exists up to $\tau = 1$. When we differentiate (5.5) with respect to τ, we obtain with (5.4) and the notation $\ddot{Z} = (\ddot{Z}_i)_{i=1}^s$, $\widetilde{F}^{(0)} = (\widetilde{f}(Z_i^{(0)},\dot{z}_0))_{i=1}^s$, $Z^{(0)} = (Z_i^{(0)})_{i=1}^s$ the differential equation

$$\left(I_s \otimes \widetilde{M}(z_0) + O(h) + O(h\|\ddot{Z}\|) + \frac{h^2}{\epsilon^2}A^2 \otimes \begin{pmatrix}0 & 0\\ 0 & I_m + O(h) + O(h^2\|\ddot{Z}\|)\end{pmatrix}\right)\cdot \frac{d\ddot{Z}}{d\tau}$$
$$= \widetilde{F}^{(0)} - \frac{1}{\epsilon^2} I_s \otimes \begin{pmatrix}0 & 0\\ 0 & I_m\end{pmatrix}\cdot Z^{(0)} .$$

Multiplying the lower block by ϵ^2/h^2, we get a right-hand side which is $O(1)$ by our assumption on the starting values. With the block notation

$$\widetilde{M}(z_0) = \begin{pmatrix} M_{11} & M_{12}\\ M_{21} & M_{22}\end{pmatrix}$$

we have

$$\begin{pmatrix} I_s \otimes M_{11} + O(h) & I_s \otimes M_{12} + O(h)\\ \frac{\epsilon^2}{h^2}\cdot I_s \otimes (M_{21} + O(h)) & \frac{\epsilon^2}{h^2}\cdot I_s \otimes M_{22} + A^2 \otimes I_m + O(h)\end{pmatrix}\frac{d\ddot{Z}}{d\tau} = O(1)$$

as long as $\|\ddot{Z}\| = O(1)$. The assumption on the eigenvalues of A implies that the above matrix has a uniformly bounded inverse for $h \le h_0$. It follows that the solution $\ddot{Z}$ exists uniquely and remains $O(1)$ up to $\tau = 1$. □

In addition to (5.1), we now consider a perturbed system

(5.6a)
$$\widehat{y}_1 = \widehat{y}_0 + h\dot{\widehat{y}}_0 + h^2 \sum_{j,k=1}^{s} b_j a_{jk} \widehat{\ddot{Y}}_k , \qquad \dot{\widehat{y}}_1 = \dot{\widehat{y}}_0 + h \sum_{j=1}^{s} b_j \widehat{\ddot{Y}}_j ,$$

with internal stages

(5.6b)
$$\widehat{Y}_i = \widehat{y}_0 + c_i h\,\dot{\widehat{y}}_0 + h^2 \sum_{j,k=1}^{s} a_{ij} a_{jk} \widehat{\ddot{Y}}_k , \qquad \dot{\widehat{Y}}_i = \dot{\widehat{y}}_0 + h \sum_{j=1}^{s} a_{ij} \widehat{\ddot{Y}}_j$$

satisfying

(5.6c)
$$M(\widehat{Y}_i)\,\widehat{\ddot{Y}}_i = f(\widehat{Y}_i, \dot{\widehat{Y}}_i) - \frac{1}{\epsilon^2}\nabla U(\widehat{Y}_i) + d_i .$$

We denote the differences to the solution of (5.1) by $\Delta y_0 = y_0 - \widehat{y}_0$, $\Delta Y_i = Y_i - \widehat{Y}_i$, etc. When we study the effect of the perturbations Δy_0, $\Delta \dot{y}_0$, and d_i on the Runge-Kutta solution, we have to take recourse to the variable transforms used in part (d) of the proof of Theorem 2.2.

Lemma 5.2. (Influence of perturbations) *In addition to the conditions of Lemma 5.1 we suppose that also* $\nabla U(\widehat{y}_0) = O(h^2)$, $\nabla^2 U(\widehat{y}_0)\,\widehat{\dot{y}_0} = O(h)$, *and further* $\Delta y_0 = O(h)$, $\Delta\dot{y}_0 = O(h)$. *Then there exist invertible matrices* $T = T(\widehat{y}_0)$ *and* $S_i = S(\widehat{y}_0, \widehat{Y}_i)$, *such that the transformed variables*

$$\begin{pmatrix} \Delta\ddot{U}_i \\ \Delta\ddot{V}_i \end{pmatrix} = T\Delta\ddot{Y}_i\ , \qquad \begin{pmatrix} \Delta u_0 \\ \Delta v_0 \end{pmatrix} = T\Delta y_0\ , \qquad \begin{pmatrix} \Delta\dot{u}_0 \\ \Delta\dot{v}_0 \end{pmatrix} = T\Delta\dot{y}_0\ , \qquad \begin{pmatrix} \delta_i \\ \theta_i \end{pmatrix} = S_i d_i$$

satisfy the bounds

$$\begin{aligned} \|\Delta\ddot{U}_i\| &\le C\cdot\left(\frac{1}{h}\|\Delta v_0\| + \|\Delta u_0\| + \|\Delta\dot{v}_0\| + \|\Delta\dot{u}_0\| + \delta + h\frac{\epsilon^2}{h^2}\theta\right) \\ \|\Delta\ddot{V}_i\| &\le C\cdot\left(\frac{1}{h^2}\|\Delta v_0\| + \frac{1}{h}\|\Delta u_0\| + \frac{1}{h}\|\Delta\dot{v}_0\| + \|\Delta\dot{u}_0\| + h\delta + \frac{\epsilon^2}{h^2}\theta\right)\ . \end{aligned}$$

Here $\delta = \max_i \|\delta_i\|$, $\theta = \max_i \|\theta_i\|$, *and* C *is a constant independent of* $0 < \epsilon \le h \le h_0$.

Proof. As in the preceding proof, we make the term divided by ϵ^2 linear by using the transformed variables Z_i and their perturbed counterparts $\widehat{Z}_i$ which are defined in the same way. We introduce

$$\begin{aligned} &\Delta z_0 = \frac{\partial z}{\partial y}(\widehat{y}_0)\Delta y_0\ , \quad \Delta\dot{z}_0 = \frac{\partial z}{\partial y}(\widehat{y}_0)\Delta\dot{y}_0\ , \quad \Delta\ddot{Z}_i = \frac{\partial z}{\partial y}(\widehat{y}_0)\Delta\ddot{Y}_i\ , \\ &\Delta\dot{Z}_i = \frac{\partial z}{\partial y}(\widehat{y}_0)\Delta\dot{Y}_i = \Delta\dot{z}_0 + h\sum_{j=1}^{s} a_{ij}\Delta\ddot{Z}_j\ , \end{aligned} \tag{5.7}$$

but we have to insist on setting

$$\begin{aligned} \Delta Z_i &= z(Y_i) - z(\widehat{Y}_i) \\ &= \Delta z_0 + c_i h\Delta\dot{z}_0 + h^2\sum_{j,k=1}^{s} a_{ij}a_{jk}\Delta\ddot{Z}_k + O(h\|\Delta z_0\| + h^2\|\Delta\dot{z}_0\| + h^3\|\Delta\ddot{Z}\|)\ . \end{aligned} \tag{5.8}$$

With (5.7), (5.8) and the *a priori* estimate $\ddot{Z}_i = O(1)$ of Lemma 5.1 we then get from (5.3) and from the analogously transformed formula (5.6c) the equation

$$\widetilde{M}(\widehat{z}_0)\Delta\ddot{Z}_i + \frac{1}{\epsilon^2}\begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix}\Delta Z_i = O(h\|\Delta\ddot{Z}\|) + O(\|\Delta z_0\| + \|\Delta\dot{z}_0\|) + \frac{\partial y}{\partial z}(\widehat{Z}_i)^T d_i\ . \tag{5.9}$$

As in part (d) of the proof of Theorem 2.2, we block-diagonalize

$$\widetilde{M}^{-1/2}(\widehat{z}_0)\begin{pmatrix} 0 & 0 \\ 0 & I_m \end{pmatrix}\widetilde{M}^{-1/2}(\widehat{z}_0) = Q\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}Q^T \tag{5.10}$$

with symmetric positive definite B. For the transformed variables

$$\begin{pmatrix} \Delta\ddot{U}_i \\ \Delta\ddot{V}_i \end{pmatrix} = Q^T \widetilde{M}^{1/2}(\widehat{z}_0)\cdot \Delta\ddot{Z}_i = T(\widehat{z}_0)\cdot\Delta\ddot{Y}_i \qquad \text{with} \quad T = Q^T\widetilde{M}^{1/2}\frac{\partial z}{\partial y} \ ,$$

$$\begin{pmatrix} \Delta\dot{U}_i \\ \Delta\dot{V}_i \end{pmatrix} = T(\widehat{z}_0)\cdot\Delta\dot{Y}_i \qquad \begin{pmatrix} \Delta U_i \\ \Delta V_i \end{pmatrix} = Q^T\widetilde{M}^{1/2}(\widehat{z}_0)\cdot\Delta Z_i \ ,$$

$$\begin{pmatrix} \delta_i \\ \theta_i \end{pmatrix} = S_i d_i \qquad \text{with} \quad S_i = S(\widehat{z}_0,\widehat{Z}_i) = Q^T\widetilde{M}^{-1/2}(\widehat{z}_0)\frac{\partial y}{\partial z}(\widehat{Z}_i)^T \ ,$$

we obtain from (5.9) the following system:

$$\begin{pmatrix} \Delta\ddot{U}_i \\ \Delta\ddot{V}_i + \frac{1}{\epsilon^2}B\cdot\Delta V_i \end{pmatrix} = O(h\|\Delta\ddot{U}\| + h\|\Delta\ddot{V}\|) + O(\|\Delta u_0\| + \|\Delta v_0\| + \|\Delta\dot{u}_0\| + \|\Delta\dot{v}_0\|) + \begin{pmatrix} \delta_i \\ \theta_i \end{pmatrix} . \tag{5.11}$$

¿From (5.8) we have

$$\Delta V_i = \Delta v_0 + c_i h\Delta\dot{v}_0 + h^2\sum_{j,k=1}^{s} a_{ij}a_{jk}\Delta\ddot{V}_k + O(h(\|\Delta u_0\| + \|\Delta v_0\|) + h^2(\|\Delta\dot{u}_0\| + \|\Delta\dot{v}_0\|)) + O(h^3(\|\Delta\ddot{U}\| + \|\Delta\ddot{V}\|)) \ .$$

We insert this into the second equation of (5.11) and multiply by ϵ^2/h^2 to obtain

$$\left(\frac{\epsilon^2}{h^2}I_s\otimes I_m + A^2\otimes B\right)\Delta\ddot{V} = O(h\|\Delta\ddot{U}\| + h\|\Delta\ddot{V}\|) + O(\frac{1}{h^2}\|\Delta v_0\| + \frac{1}{h}\|\Delta\ddot{v}_0\|) + O(\frac{1}{h}\|\Delta u_0\| + \|\Delta\dot{u}_0\|) + \frac{\epsilon^2}{h^2}\Theta \ ,$$

where $\Delta\ddot{V} = (\Delta\ddot{V}_i)_{i=1}^s$, and $\Theta = (\theta_i)_{i=1}^s$. The matrix on the left-hand side has a uniformly bounded inverse for $0 < \epsilon \leq h$. We thus obtain the bounds stated in the theorem. □

With the bounds of Lemma 5.2 at hand, it is now an easy task to estimate the errors in the numerical solution after one step.

Lemma 5.3. (Error after one step) *Under the assumptions of Lemma 5.2 we have with* $\alpha = \|\Delta u_0\| + \|\Delta v_0\| + h\cdot(\|\Delta\dot{u}_0\| + \|\Delta\dot{v}_0\|)$ *the relations*

$$\begin{pmatrix} \Delta u_1 \\ h\Delta\dot{u}_1 \end{pmatrix} = \begin{pmatrix} \Delta u_0 + h\Delta\dot{u}_0 \\ h\Delta\dot{u}_0 \end{pmatrix} + O(h\alpha + h^2\delta + h\epsilon^2\theta)$$

$$\begin{pmatrix} \Delta v_1 \\ \epsilon L^{-1}\Delta\dot{v}_1 \end{pmatrix} = R\left(\frac{h}{\epsilon}\begin{pmatrix} 0 & L \\ -L^T & 0 \end{pmatrix}\right)\cdot\begin{pmatrix} \Delta v_0 \\ \epsilon L^{-1}\Delta\dot{v}_0 \end{pmatrix} + \begin{pmatrix} O(h\alpha + h^3\delta + \epsilon^2\theta) \\ O(\epsilon\alpha + \epsilon h^2\delta + \epsilon^3/h\cdot\theta) \end{pmatrix}$$

where $R(z)$ is the stability function (3.3), and L is the Choleski factor of the positive definite matrix $B = LL^T$ in (5.10).

Proof. The relations for Δu_1 and $h\Delta\dot{u}_1$ follow directly from (5.6), (5.7) and the bounds of Lemma 5.2, used in (5.1a).
The second equation of (5.11), with $\|\Delta\ddot{U}\|$ and $\|\Delta\ddot{V}\|$ on the right-hand side bounded by Lemma 5.2, reads

$$\Delta\ddot{V}_i + \frac{1}{\epsilon^2}B\cdot\Delta V_i = O(\frac{1}{h}\|\Delta v_0\| + \|\Delta u_0\| + \|\Delta\dot{v}_0\| + \|\Delta\dot{u}_0\| + h\delta + \theta)\ . \tag{5.12}$$

In analogy to the proof of Theorem 2.2, we rewrite this as a first-order system with a skew-symmetric matrix:

$$\begin{pmatrix} \Delta\dot{V}_i \\ \epsilon L^{-1}\Delta\ddot{V}_i \end{pmatrix} = \frac{1}{\epsilon}\begin{pmatrix} 0 & L \\ -L^T & 0 \end{pmatrix}\cdot\begin{pmatrix} \Delta V_i \\ \epsilon L^{-1}\Delta\dot{V}_i \end{pmatrix} + \begin{pmatrix} 0 \\ O(\epsilon\beta) \end{pmatrix} \tag{5.13}$$

where β is the expression on the right-hand side of (5.12). ¿From (5.7), (5.8) and Lemma 5.2 we get the relations

$$\begin{aligned} \Delta V_i =& \Delta v_0 + h\sum_{j=1}^{s} a_{ij}\Delta\dot{V}_j + O(h\alpha + h^3\delta + \epsilon^2\theta) \\ \epsilon L^{-1}\Delta\dot{V}_i =& \epsilon L^{-1}\Delta\dot{v}_0 + h\sum_{j=1}^{s} a_{ij}\epsilon L^{-1}\Delta\ddot{V}_j + O(\epsilon\alpha + \epsilon h^2\delta + \epsilon^3/h\cdot\theta)\ . \end{aligned} \tag{5.14}$$

Similarly we have

$$\begin{aligned} \Delta v_1 &= \Delta v_0 + h\sum_{j=1}^{s} b_j\Delta\dot{V}_j + O(h\alpha + h^3\delta + \epsilon^2\theta) \\ \epsilon L^{-1}\Delta\dot{v}_1 &= \epsilon L^{-1}\Delta\dot{v}_0 + h\sum_{j=1}^{s} b_j\epsilon L^{-1}\Delta\ddot{V}_j + O(\epsilon\alpha + \epsilon h^2\delta + \epsilon^3/h\cdot\theta)\ . \end{aligned} \tag{5.15}$$

If we ignore the $O(\dots)$ terms, then equations (5.13)-(5.15) are just the Runge-Kutta equations for the linear differential equation

$$\dot{w} = \frac{1}{\epsilon}\begin{pmatrix} 0 & L \\ -L^T & 0 \end{pmatrix} w\ ,$$

whose Runge-Kutta solution is

$$w_1 = R\left(\frac{h}{\epsilon}\begin{pmatrix} 0 & L \\ -L^T & 0 \end{pmatrix}\right) w_0\ .$$

Taking the perturbation terms in (5.13)-(5.15) into account leads to the statement of the lemma. □

Next we consider error propagation over many steps of the Runge-Kutta method applied to (2.1) and its perturbed variant. We denote by $(y_n, \dot y_n)$ and $(\widehat y_n, \widehat{\dot y}_n)$ the corresponding numerical solutions. We assume

$$
\begin{aligned}
&\nabla U(\widehat y_n) = O(h^2)\ , \quad \nabla^2 U(\widehat y_n)\widehat{\dot y}_n = O(h) \qquad \text{for } \ 0 \le nh \le T\ , \\
&\widehat y_{n+1} - \widehat y_n = O(h)\ ,
\end{aligned} \tag{5.16}
$$

and for the starting value of the unperturbed Runge-Kutta scheme we assume

$$
\begin{aligned}
&\nabla U(y_0) = O(h^2)\ , \quad \nabla^2 U(y_0)\dot y_0 = O(h)\ , \\
&y_0 - \widehat y_0 = O(h)\ , \quad \dot y_0 - \widehat{\dot y}_0 = O(h)\ .
\end{aligned} \tag{5.17}
$$

We denote $\Delta y_n = y_n - \widehat y_n$, $\Delta \dot y_n = \dot y_n - \widehat{\dot y}_n$, and the transformed variables

$$
\begin{pmatrix} \Delta u_n \\ \Delta v_n \end{pmatrix} = T(\widehat y_n)\Delta y_n\ , \qquad \begin{pmatrix} \Delta \dot u_n \\ \Delta \dot v_n \end{pmatrix} = T(\widehat y_n)\Delta \dot y_n\ , \tag{5.18}
$$

with the transformation matrix $T(\widehat y_n)$ of Lemma 5.2. Similarly, we let

$$
\begin{pmatrix} \delta_{ni} \\ \theta_{ni} \end{pmatrix} = S(\widehat y_n, \widehat Y_{ni}) d_{ni} \qquad (i = 1, \ldots, s) \tag{5.19}
$$

where d_{ni} is the perturbation in the n-th step of (5.6c). We denote by δ and θ the corresponding bounds

$$
\|\delta_{ni}\| \le \delta\ , \quad \|\theta_{ni}\| \le \theta \qquad \text{for } \ 0 \le nh \le T\ , \quad \text{and all } i\ , \tag{5.20a}
$$

of which we assume

$$
\delta = O(h)\ , \qquad \epsilon^2\theta = O(h^2)\ . \tag{5.20b}
$$

Then we have the following result:

Lemma 5.4. (Error propagation) *Let the Runge-Kutta method satisfy the stability conditions (3.4) and (3.5). Under conditions (5.16)-(5.20), the Runge-Kutta solution $(y_n, \dot y_n)$ exists for $0 \le nh \le T$, $0 < \epsilon \le h \le h_0$, and it satisfies*

$$
\begin{aligned}
\|\Delta u_n\| + \|\Delta \dot u_n\| &\le C \cdot \big(\|\Delta u_0\| + \|\Delta \dot u_0\| + \|\Delta v_0\| + h\|\Delta \dot v_0\| + \delta + \epsilon^2\theta/h\big) \\
\|\Delta v_n\| + h\|\Delta \dot v_n\| &\le C \cdot \big(h(\|\Delta u_0\| + \|\Delta \dot u_0\|) + (\rho^n + h)(\|\Delta v_0\| + h\|\Delta \dot v_0\|) + h\delta + \epsilon^2\theta\big)\ ,
\end{aligned}
$$

where $0 < \rho < 1$. If $\epsilon \ll h$, then ρ can be chosen as any fixed number larger than $|R(\infty)|$.

Proof. (a) Let us suppose for the moment that a solution $(y_n, \dot{y}_n)$ exists for $0 \le nh \le T$, which satisfies (5.17) for all n. This can be verified by an induction argument, as it will be seen that the conditions of Lemma 5.1 remain satisfied uniformly over the whole integration interval. For brevity we denote

$$\mu_n = \|(\Delta u_n, \Delta \dot{u}_n)^T\| ,$$

and

$$\begin{aligned} \nu_n &= \|(\Delta v_n, \epsilon L^{-1}(\widehat{y}_n)\cdot \Delta \dot{v}_n)^T\| && \text{for} \quad \epsilon \le h \le K\epsilon \\ \nu_n &= \|(\Delta v_n, h\Delta \dot{v}_n)^T\| && \text{for} \quad h > K\epsilon \end{aligned}$$

where K is a sufficiently large constant. We will show that for suitably chosen (h- and ϵ-independent) norms we have

$$\begin{pmatrix} \mu_{n+1} \\ \nu_{n+1} \end{pmatrix} \le \begin{pmatrix} 1+O(h) & O(1) \\ O(h) & \rho + O(h) \end{pmatrix} \begin{pmatrix} \mu_n \\ \nu_n \end{pmatrix} + \begin{pmatrix} O(h\delta + \epsilon^2\theta) \\ O(h^3\delta + \epsilon^2\theta) \end{pmatrix} \tag{5.21}$$

with a $\rho < 1$. This is seen from Lemma 5.3 as follows. For ϵ of the same magnitude as h, this holds with the Euclidian norm in (5.21), because by (3.5)

$$\|R\left(\frac{h}{\epsilon}\begin{pmatrix} 0 & L \\ -L^T & 0 \end{pmatrix}\right)\| = \max_{j=1,\dots,m} |R(i\frac{h}{\epsilon}\omega_j)| \le \rho < 1$$

where $\omega_j^2 > 0$ are the eigenvalues of $B = LL^T$. On the other hand, for small ϵ/h a calculation gives

$$\begin{pmatrix} I_m & 0 \\ 0 & \frac{h}{\epsilon}L \end{pmatrix} R\left(\frac{h}{\epsilon}\begin{pmatrix} 0 & L \\ -L^T & 0 \end{pmatrix}\right)\begin{pmatrix} I_m & 0 \\ 0 & \frac{\epsilon}{h}L^{-1} \end{pmatrix} = \begin{pmatrix} R(\infty) & 0 \\ R'(\infty) & R(\infty) \end{pmatrix} \otimes I_m + O((\epsilon/h)^2) ,$$

where $R(z) = R(\infty) + R'(\infty)z^{-1} + O(z^{-2})$ as $z \to \infty$. Since $|R(\infty)| < 1$, we can choose a norm such that the induced operator norm of the matrix on the right-hand side is strictly smaller than 1. This gives us (5.21).

(b) To bound μ_n and ν_n, we transform the matrix in (5.21) to diagonal form, so that

$$\begin{aligned} \begin{pmatrix} \mu_n \\ \nu_n \end{pmatrix} \le & X^{-1}\begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} X \begin{pmatrix} \mu_0 \\ \nu_0 \end{pmatrix} + \\ & \sum_{j=1}^{n} X^{-1}\begin{pmatrix} \lambda_1^{n-j} & 0 \\ 0 & \lambda_2^{n-j} \end{pmatrix} X \begin{pmatrix} O(h\delta + \epsilon^2\theta) \\ O(h^3\delta + \epsilon^2\theta) \end{pmatrix} \end{aligned}$$

with $\lambda_1 = 1 + O(h)$, $\lambda_2 = \rho + O(h)$, and transformation matrix

$$X = \begin{pmatrix} 1 & O(1) \\ O(h) & 1 \end{pmatrix} .$$

Direct computation now yields the estimates stated in the lemma. □

6. Estimation of the remainder term in the ϵ^2-expansion of the numerical solution

In this section we will prove the following result, which together with Theorems 2.2 and 4.2 will imply Theorem 3.1.

Theorem 6.1. (Asymptotic ϵ^2-expansion of the numerical solution) *Let the Runge-Kutta method have stage order q and satisfy the stability conditions (3.4)-(3.5). Suppose that the starting value $(y_0^\epsilon, \dot{y}_0^\epsilon)$ is on the manifold $\mathcal{M}^\epsilon$ of Theorem 2.1, i.e., that the exact solution of (2.1) with initial value $(y_0^\epsilon, \dot{y}_0^\epsilon)$ is smooth. For $0 < \epsilon \le h \le h_0$, with h_0 sufficiently small but independent of ϵ, there then exists a unique Runge-Kutta solution (3.2) of the stiff equation (2.1), which is of the form*

$$\begin{aligned} y_n^\epsilon &= y_n^0 + \epsilon^2 y_n^1 + \cdots + \epsilon^{2k} y_n^k + r_n \\ \dot{y}_n^\epsilon &= \dot{y}_n^0 + \epsilon^2 \dot{y}_n^1 + \cdots + \epsilon^{2k} \dot{y}_n^k + \dot{r}_n \end{aligned} \tag{6.1}$$

Here $y_n^l, \dot{y}_n^l$ $(l = 0, \ldots, k)$ denote the Runge-Kutta solution of the differential-algebraic equation (2.25), where the starting value $(y_0^l, \dot{y}_0^l)$ is the coefficient of ϵ^{2l} in the ϵ^2-expansion of $(y_0^\epsilon, \dot{y}_0^\epsilon)$. The remainder terms are bounded by

$$\begin{aligned} \|r_n\| + \|\dot{r}_n\| &= O(\epsilon^{2k}) && \textit{if } q = 2k \ , \\ &= O(h\epsilon^{2k}) && \textit{if } q = 2k+1 \ . \end{aligned} \tag{6.2}$$

This result will follow from Lemma 5.4, once we have studied the defect obtained upon inserting the truncated expansions (with $k = [q/2]$)

$$\begin{aligned} \widehat{y}_n &= y_n^0 + \epsilon^2 y_n^1 + \cdots + \epsilon^{2k} y_n^k \\ \widehat{\dot{y}}_n &= \dot{y}_n^0 + \epsilon^2 \dot{y}_n^1 + \cdots + \epsilon^{2k} \dot{y}_n^k \end{aligned} \tag{6.3}$$

and the truncated expansions of the internal stages into the formulas (3.2) defining $y_n, \dot{y}_n$. By linearity, there is no defect in the Runge-Kutta relations (3.2a,b), and the defect d_{ni} in (3.2c),

$$M(\widehat{Y}_{ni})\widehat{\ddot{Y}}_{ni} = f(\widehat{Y}_{ni}, \widehat{\dot{Y}}_{ni}) - \frac{1}{\epsilon^2}\nabla U(\widehat{Y}_{ni}) + d_{ni} \ , \tag{6.4}$$

is of the special form given in the following lemma.

Lemma 6.2. (Defect of truncated expansions) *With $k = [q/2]$, we have in (6.4)*

$$S(Y_{ni}^0) d_{ni} = \epsilon^{2k} \begin{pmatrix} 0 \\ B(Y_{ni}^0)\Lambda_{ni}^k \end{pmatrix} + O(\epsilon^{2k+2}) \ . \tag{6.5}$$

Here the matrices are $S(Y_{ni}^0) = S(\widehat{y}_n, \widehat{Y}_{ni}) + O(h)$ and $B(Y_{ni}^0) = B(\widehat{y}_n) + O(h)$, with $S(\widehat{y}_n, \widehat{Y}_{ni})$ and $B(\widehat{y}_n)$ defined as in the proof of Lemma 5.2. Λ_{ni}^k is the internal stage approximation of $\lambda^k(t_n + c_i h)$ in (2.25.k), which satisfies

$$\begin{aligned} \Lambda_{ni}^k &= O(h^{-1}) && \textit{if } q = 2k \ , \\ &= O(1) && \textit{if } q = 2k+1 \ . \end{aligned} \tag{6.6}$$

Proof. Since the Runge-Kutta relations (3.2a,b) are linear with ϵ-independent coefficients, since (3.2c) is of the same form as the corresponding differential or differential-algebraic equations (2.1) or (2.25), and since $\widehat{Y}_{ni}, \widehat{\dot{Y}}_{ni} = O(1)$ by Theorem 4.2, the same construction as in parts (b) and (c) of the proof of Theorem 2.2 gives

$$d_{ni} = \epsilon^{2k} G^T(Y_{ni}^0)\Lambda_{ni}^k + O(\epsilon^{2k+2}) \ .$$

Here $G^T(y)$ may be any matrix of full column rank which has the same range as $H(y) = \nabla^2 U(y)$. The representation (6.5) is obtained with a special choice of G^T: The positive definite matrix B has been constructed in the proofs of Theorem 2.2 and Lemma 5.2 such that

$$Q^T \widetilde{M}^{-1/2} \left(\frac{\partial y}{\partial z}\right)^T H \left(\frac{\partial y}{\partial z}\right) \widetilde{M}^{-1/2} Q = \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} ,$$

where $\widetilde{M} = \left(\frac{\partial y}{\partial z}\right)^T M \left(\frac{\partial y}{\partial z}\right)$. It follows that

$$G^T = \left(\frac{\partial z}{\partial y}\right)^T \widetilde{M}^{1/2} Q \begin{pmatrix} 0 \\ B \end{pmatrix} \equiv S^{-1} \begin{pmatrix} 0 \\ B \end{pmatrix}$$

is an admissible choice. This gives (6.5). Finally, the bounds (6.6) for Λ_{ni}^k are given explicitly for $k = 0$ and 1 in the proof of Theorem 4.2, and follow with the indicated induction argument also for higher k. □

Proof of Theorem 6.1. We apply Lemma 5.4 with $\widehat{y}_n, \widehat{\dot{y}}_n$ of (6.3). We have

$$y_0^\epsilon - \widehat{y}_0 = O(\epsilon^{2k+2}) \ , \quad \dot{y}_0^\epsilon - \widehat{\dot{y}}_0 = O(\epsilon^{2k+2}) \ ,$$

and Lemma 6.2 yields

$$\begin{aligned} \delta &= O(\epsilon^{2k}) \ , \quad \theta = O(\epsilon^{2k}/h) \qquad \text{if } q = 2k \ , \\ \delta &= O(h\epsilon^{2k}) \ , \quad \theta = O(\epsilon^{2k}) \qquad \text{if } q = 2k+1 \ . \end{aligned}$$

Lemma 5.4 now gives the stated bounds for $r_n = y_n^\epsilon - \widehat{y}_n$ and $\dot{r}_n = \dot{y}_n^\epsilon - \widehat{\dot{y}}_n$. □

Proof of Theorem 3.1. Combining Theorem 6.1 with Theorem 2.2, we have for $q = 2k+1$

$$\begin{aligned} y_n^\epsilon - y^\epsilon(t_n) &= \left(y_n^0 - y^0(t_n)\right) + \epsilon^2 \left(y_n^1 - y^1(t_n)\right) + \cdots + \epsilon^{2k} \left(y_n^k - y^k(t_n)\right) + O(h\epsilon^{2k}) \\ \dot{y}_n^\epsilon - \dot{y}^\epsilon(t_n) &= \left(\dot{y}_n^0 - \dot{y}^0(t_n)\right) + \epsilon^2 \left(\dot{y}_n^1 - \dot{y}^1(t_n)\right) + \cdots + \epsilon^{2k} \left(\dot{y}_n^k - \dot{y}^k(t_n)\right) + O(h\epsilon^{2k}) \ . \end{aligned}$$

By Theorem 4.2, the position error terms $\epsilon^{2l}\left(y_n^l - y^l(t_n)\right)$ $(l = 1, \ldots, k)$ and the corresponding velocity error terms are all bounded by $O(\epsilon^2 h^{q-2})$ for $\epsilon \le h$. This gives Theorem 3.1 for odd stage order q. The case of even q is treated in the same way. □

7. Starting values near, but not on the manifold of smooth motion

In this section we prove Theorem 3.2. We will use the results of Sections 5 and 6, and a variant of an invariant manifold theorem due to Kirchgraber, Nipp and Stoffer [12],[16]. There, one considers a recurrence relation

$$\begin{aligned} \xi_{n+1} &= A\xi_n + \varphi(\xi_n, \eta_n) \\ \eta_{n+1} &= \psi(\xi_n, \eta_n) \end{aligned} \qquad n \geq 0 \; , \tag{7.1}$$

where A is an invertible matrix, and $\begin{pmatrix} \varphi \\ \psi \end{pmatrix} : \mathbf{R}^{n_\xi} \times \mathbf{R}^{n_\eta} \to \mathbf{R}^{n_\xi} \times \mathbf{R}^{n_\eta}$ is Lipschitz continuous. We denote the Lipschitz constant of φ with respect to ξ by $L_{\xi\xi}$, and define $L_{\xi\eta}$, $L_{\eta\xi}$, $L_{\eta\eta}$ analogously.

Theorem 7.1. [16],[12] (Invariant manifold theorem) *For (7.1), suppose that*

$$\|A^{-1}\| \leq \alpha \; , \qquad L_{\xi\xi} + L_{\eta\eta} + 2\sqrt{L_{\xi\eta} L_{\eta\xi}} < 1/\alpha \; , \qquad \alpha \geq 1 \; . \tag{7.2}$$

Then the following holds:

(i) There exists a Lipschitz continuous function $s : \mathbf{R}^{n_\xi} \to \mathbf{R}^{n_\eta}$ *such that*

$$\eta_0 = s(\xi_0) \qquad \textit{implies} \qquad \eta_n = s(\xi_n) \quad \textit{for all } n. \tag{7.3}$$

Hence, $\mathcal{S} = \{(\xi, s(\xi)) : \xi \in \mathbf{R}^{n_\xi}\}$ *is an invariant manifold for (7.1).*

(ii) The invariant manifold is attractive: With $\rho = L_{\eta\eta} + \sqrt{L_{\xi\eta} L_{\eta\xi}} < 1$*, solutions of (7.1) satisfy*

$$\|\eta_n - s(\xi_n)\| \leq \rho^n \cdot \|\eta_0 - s(\xi_0)\| \; , \qquad n \geq 0 \; , \tag{7.4}$$

for all starting values (ξ_0, η_0).

(iii) For every (ξ_0, η_0)*, there exists* $(\xi_0^*, \eta_0^*) \in \mathcal{S}$*, such that the corresponding solutions of (7.1) converge geometrically to each other:*

$$\|(\xi_n, \eta_n) - (\xi_n^*, \eta_n^*)\| \leq C\rho^n \cdot \|(\xi_0, \eta_0) - (\xi_0^*, \eta_0^*)\| \; , \tag{7.5}$$

and

$$\|(\xi_0, \eta_0) - (\xi_0^*, \eta_0^*)\| \leq C' \cdot \|\eta_0 - s(\xi_0)\| \; . \tag{7.6}$$

The constants C *and* C' *depend only on the quantities in (7.2).*

We have now all ingredients ready for the proof of Theorem 3.2.

Proof of Theorem 3.2. (a) Let a starting value $(y_0, \dot y_0)$ with

$$\nabla U(y_0) = O(h^2) \ , \quad \nabla^2 U(y_0)\dot y_0 = O(h) \tag{7.7}$$

be given. To distinguish it from other starting values needed in the course of the proof, we denote it henceforth by $(\overline{y}_0, \overline{\dot y}_0)$. By (7.7), $(\overline{y}_0, \overline{\dot y}_0)$ is $O(h)$-close to the manifold $\mathcal{M}^\epsilon$ of smooth motion, and we can thus choose an $(y(0), \dot y(0)) \in \mathcal{M}^\epsilon$ which is $O(h)$-close to $(\overline{y}_0, \overline{\dot y}_0)$. We take this as initial value of a smooth solution $(y(t), \dot y(t))$ of (2.1).
We will transform variables in all Runge-Kutta solutions $(y_n, \dot y_n)$ of (2.1) with starting values $(y_0, \dot y_0)$ which satisfy (7.7) and are $O(h)$-close to $(\overline{y}_0, \overline{\dot y}_0)$. We then know from Lemma 5.4 and Theorem 3.1 that for such solutions

$$y_n - y(t_n) = O(h) \ , \quad \dot y_n - \dot y(t_n) = O(h) \qquad \text{uniformly for } \ 0 \le t_n \le T \ . \tag{7.8}$$

Like in Lemma 5.2, we now introduce transformed variables

$$\begin{pmatrix} u_n \\ v_n \end{pmatrix} = T(y(t_n)) \cdot y_n \ , \qquad \begin{pmatrix} \dot u_n \\ \dot v_n \end{pmatrix} = T(y(t_n)) \cdot \dot y_n \ , \tag{7.9}$$

where $T(y) = (Q^T \widetilde{M}^{1/2} \frac{\partial z}{\partial y})(y)$ as in Lemma 5.2. In these variables, the Runge-Kutta method for (2.1) can be expressed as a recursion of the form (7.1): By Lemma 5.3 (and (7.8)), and by part (a) of the proof of Lemma 5.4, we have (7.1) and (7.2) with

$$\begin{aligned}
\xi_n &= (t_n, u_n, h\dot u_n)^T \\
\eta_n &= \begin{cases} (v_n, \epsilon L^{-1}(y(t_n)) \cdot \dot v_n)^T & \text{for } \ \epsilon \le h \le K\epsilon \ , \\ (v_n, h\dot v_n)^T & \text{for } \ h > K\epsilon \ . \end{cases} \\
A &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & I_d & I_d \\ 0 & 0 & I_d \end{pmatrix} \\
L_{\xi\xi} &= O(h) \ , \quad L_{\xi\eta} = O(h) \ , \quad L_{\eta\xi} = O(h) \ , \quad L_{\eta\eta} = \rho_0 + O(h) \ ,
\end{aligned}$$

where ρ_0 is the ρ in Lemma 5.4.
At first, this only holds *locally* along $y(t)$, $0 \le t \le T$, but outside we may modify and extend φ and ψ such that (7.2) holds with global Lipschitz constants. We can therefore apply Theorem 7.1, which gives us the following in the original variables $(y, \dot y)$: There exist $2d$-dimensional manifolds $M_h^\epsilon(t)$, $0 \le t \le T$, such that

$$(y_0, \dot y_0) \in M_h^\epsilon(0) \qquad \text{implies} \qquad (y_n, \dot y_n) \in M_h^\epsilon(t_n) \quad \text{for all } \ n \ . \tag{7.10}$$

For every $(y_0, \dot y_0)$ satisfying (7.7) and $O(h)$-close to $(\overline{y}_0, \overline{\dot y}_0)$, there exists $(y_0^*, \dot y_0^*) \in M_h^\epsilon(0)$, such that the corresponding Runge-Kutta solutions satisfy

$$(y_n, \dot y_n) = (y_n^*, \dot y_n^*) + O(\rho^n h) \ , \tag{7.11}$$

where $\rho = \rho_0 + O(h)$. The last statement follows from (7.5) and (7.6), and holds in particular for $(\overline{y}_0, \overline{\dot y}_0)$ itself.

(b) On the other hand, Theorem 6.1 gives us an "almost-invariant" manifold: Let $\mathcal{M}_{h,n}^{\epsilon}$ denote the set of all $(\widehat{y}_n, \widehat{\dot{y}}_n)$ of the form of a truncated ϵ^2-expansion (6.3), where the coefficients $(y_n^l, \dot{y}_n^l)$ are the Runge-Kutta solution of the differential-algebraic system (2.25), and where the starting value is a truncated ϵ^2-expansion $(\widehat{y}_0, \widehat{\dot{y}}_0) \in \mathcal{M}^{\epsilon} + O(\epsilon^{2k+2})$. The set $\mathcal{M}_{h,n}^{\epsilon}$ is then a $2d$-dimensional manifold, which by Theorem 3.1 (and Lemma 5.4 for the $O(\epsilon^{2k+2})$-perturbation off $\mathcal{M}_h^{\epsilon}$) is $O(h^q)$-close to $\mathcal{M}^{\epsilon}$.
Let now $(y_0^{\epsilon}, \dot{y}_0^{\epsilon}) \in \mathcal{M}^{\epsilon}$. Then we have

$$(y_n^{\epsilon}, \dot{y}_n^{\epsilon}) \in \mathcal{M}_{h,n}^{\epsilon} + O(\epsilon^{2k})$$

by Theorem 6.1 (for $q = 2k$, and with an additional factor h before ϵ^{2k} for $q = 2k+1$), and

$$(y_n^{\epsilon}, \dot{y}_n^{\epsilon}) \in M_h^{\epsilon}(t_n) + O(\rho^n h)$$

as a consequence of (7.11). Since this holds uniformly for $0 \le t_n \le T$ and for all $(y_0^{\epsilon}, \dot{y}_0^{\epsilon}) \in \mathcal{M}^{\epsilon}$ (in an $O(h)$-neighborhood of $(\overline{y}_0, \overline{\dot{y}}_0)$), we see that the two manifolds $\mathcal{M}_{h,n}^{\epsilon}$ and $M_h^{\epsilon}(t_n)$ are $O(\rho^n h + \epsilon^{2k})$-close. As they have the same dimension $2d$, we can conclude conversely: For every $(y_0^*, \dot{y}_0^*) \in M_h^{\epsilon}(0)$, there exists $(y_0^{\epsilon}, \dot{y}_0^{\epsilon}) \in \mathcal{M}^{\epsilon} = \mathcal{M}_{h,0}^{\epsilon} + O(\epsilon^{2k+2})$ such that the corresponding Runge-Kutta solutions satisfy

$$(y_n^*, \dot{y}_n^*) = (y_n^{\epsilon}, \dot{y}_n^{\epsilon}) + O(\rho^n h + \epsilon^{2k}) \; . \tag{7.12}$$

Theorem 3.2 now follows by combining (7.11) and (7.12). □

8. Computational aspects

Standard software for stiff initial value problems does not work satisfactorily when applied to stiff mechanical systems (2.1). There are two reasons for this failure:

(i) The numerical treatment of the stiff problem (2.1) is very close to that of the associated constrained problem in the index-3 form (2.25.0). This can cause a breakdown of the standard error estimation and step size control.

(ii) The modified Newton iterations usually employed for the solution of the nonlinear system of equations in each Runge-Kutta step only converge under a severe step size restriction $h = O(\epsilon^{2/3})$, except when the potential U is quadratic.

We will see in this section that these difficulties can be overcome by suitable modifications.

Concerning (i), we remark that the usual step size control works well when the velocities are multiplied by the step size h in the error estimation, which should thus be based on $\|\Delta y\| + h \cdot \|\Delta \dot{y}\|$. Compare [7], p.104. An incurable consequence of (i) remains the order reduction as implied by Theorem 3.1. This appears tolerable, however, for Runge-Kutta methods with higher stage order, e.g., Radau IIA methods.

We now turn to problem (ii). The following lemma gives sufficient conditions for the convergence of the usual modified Newton iterations, and the proof and numerical experiments indicate that they are actually necessary, unless the potential is quadratic.

Lemma 8.1. (Convergence of modified Newton iterations under step size restriction) *Consider the solution of the nonlinear system of equations (5.1c) for* $\ddot{Y}_i$ $(i = 1, \dots s)$*, with* Y_i *and* $\dot{Y}_i$ *inserted from (5.1b). Suppose that the starting values satisfy* $\nabla U(y_0) = O(h^2)$, $\nabla^2 U(y_0)\dot{y}_0 = O(h)$*, and let the iteration start with* $\ddot{Y}_i^{(0)} = O(1)$*. Then the modified Newton iteration with an approximate Jacobian*

$$J_0 = I_s \otimes M(y_0) + \frac{h^2}{\epsilon^2} A^2 \otimes \nabla^2 U(y_0) + O(h) \tag{8.1}$$

converges if $h \le c\epsilon^{2/3}$*. With the approximate Jacobian*

$$J_1 = I_s \otimes M(y_0) + \frac{h^2}{\epsilon^2} \left(\sum_{j=1}^{s} a_{ij} a_{jk} \nabla^2 U(y_0 + c_i h\, \dot{y}_0) \right)_{i,k=1}^{s} + O(h) \tag{8.2}$$

the iteration converges if $h \le c\epsilon^{1/2}$*. In both cases* c *is a sufficiently small constant.*

Proof. Denoting

$$F(\ddot{Y}) = \left(M(Y_i)\ddot{Y}_i - f(Y_i, \dot{Y}_i) + \frac{1}{\epsilon^2} \nabla U(Y_i) \right)_{i=1}^{s} ,$$

with Y_i, $\dot{Y}_i$ inserted from (5.1b), the modified Newton iteration is the fixed-point iteration

$$\ddot{Y}^{(k+1)} = \Phi(\ddot{Y}^{(k)}) , \qquad \text{with} \quad \Phi(\ddot{Y}) = \ddot{Y} - J^{-1} F(\ddot{Y}) .$$

For both $J = J_0$ and $J = J_1$ we have $\|J^{-1}\| = O(1)$, and since $\nabla^2 U(y_0 + c_i h\, \dot{y}_0 + h^2 \sum_{j,k} a_{ij} a_{jk} \ddot{Y}_k) = \nabla^2 U(y_0 + c_i h\, \dot{y}_0) + O(h^2) = \nabla^2 U(y_0) + O(h)$ for $\ddot{Y} = O(1)$, we get

$$\|\Phi'(\ddot{Y})\| = \begin{cases} O(h^3/\epsilon^2) & \text{for} \quad J = J_0 , \\ O(h^4/\epsilon^2) & \text{for} \quad J = J_1 . \end{cases}$$

The Banach fixed-point theorem then gives the result. □

In the proof of Lemma 8.1, one sees that no convergence problems would occur if the Jacobian could use values Y_i^0 in the argument of $\nabla^2 U$ which are $O(\epsilon^2)$-close to the solution of (5.1c). Such values can be obtained by first solving the Runge-Kutta equations for the constrained problem (2.25.0), see [7], p.95 for a study of the corresponding modified Newton iteration. Then we use the obtained values Y_i^0 in modified Newton iterations for (5.1) with Jacobian

$$J_2 = I_s \otimes M(y_0) + \frac{h^2}{\epsilon^2} \left(\sum_{j=1}^{s} a_{ij} a_{jk} \nabla^2 U(Y_i^0) \right)_{i,k=1}^{s} + O(h) . \tag{8.3}$$

A potential difficulty with all the iteration matrices in (8.1)-(8.3) is that their condition numbers grow without bound as $\epsilon/h \to 0$. Moreover, in (8.2) and (8.3) we have lost the tensor product structure of the Jacobian in (8.1) which leads to significant savings in the linear algebra, see [3], and [8], p.131f. This can be preserved in the following approach, which appears particularly effective if the rank of $\nabla^2 U$, m, is considerably smaller than the dimension of the system, n. The starting point is Hairer's reformulation ([7], p.120) of the equations of the stiff pendulum (1.1),

$$\begin{aligned} \ddot{y}_1 &= -y_1\lambda \\ \ddot{y}_2 &= -y_2\lambda - 1 \\ \epsilon^2\lambda &= \frac{\sqrt{y_1^2+y_2^2}-1}{\sqrt{y_1^2+y_2^2}}\ . \end{aligned} \tag{8.4}$$

Formally setting $\epsilon = 0$ in this system of equations yields the differential-algebraic equations of the fixed-length pendulum in index-3 form. When a Runge-Kutta method is applied to (8.4), the numerical solution is still the same as in (1.1), but now the modified Newton method converges without a step size restriction as in Lemma 8.1. The reformulation (8.4) depends on an explicit representation $\nabla U(y) = G^T(y)\phi(y)$, where $G^T(y) = (y_1, y_2)^T$ is a full-rank matrix having the same range as $\nabla^2 U(y)$. Unfortunately, such a representation is not possible in general. However, along a smooth solution we "almost" have that $\nabla U(y)$ is in the range of $\nabla^2 U(y)$ (up to a perturbation of magnitude $O(\epsilon^2)$). This observation is at the heart of the following iteration: Beginning with

$$r_i^{(0)} = 0\ , \qquad i = 1,\ldots,s\ , \tag{8.5}$$

we solve iteratively for $k \ge 0$ the nonlinear systems in $\ddot{Y}_i^{(k)}$ and $\Lambda_i^{(k)}$ (with $Y_i^{(k)}$, $\dot{Y}_i^{(k)}$ related to $\ddot{Y}_i^{(k)}$ by (5.1b))

$$\begin{aligned} M(Y_i^{(k)})\ddot{Y}_i^{(k)} &= f(Y_i^{(k)}, \dot{Y}_i^{(k)}) - G^T(Y_i^{(k)})\Lambda_i^{(k)} + r_i^{(k)} \\ \frac{\epsilon^2}{h^2}\Lambda_i^{(k)} &= \frac{1}{h^2}\left(G^T(Y_i^{(k)})\right)^{-} \cdot \nabla U(Y_i^{(k)})\ , \end{aligned} \tag{8.6}$$

and then set

$$r_i^{(k+1)} = G^T(Y_i^{(k)})\Lambda_i^{(k)} - \frac{1}{\epsilon^2}\nabla U(Y_i^{(k)})\ . \tag{8.7}$$

Here $G^T \in \mathbf{R}^{n\times m}$ is a full-rank matrix having the same range as $\nabla^2 U \in \mathbf{R}^{n\times n}$, e.g., a selection of m linearly independent columns of $\nabla^2 U$. $(G^T)^-$ denotes a left inverse of G^T, i.e., $(G^T)^- \cdot G^T = I_m$. For example, $(G^T)^-$ can be obtained from the LU decompositon of an invertible $m \times m$-submatrix of G^T: $(G^T)^- = (E^{-1}, 0)$ if $G^T = \begin{pmatrix} E \\ F \end{pmatrix}$ with invertible E. This is computationally inexpensive if $m \ll n$.

The nonlinear system (8.6) is solved by inner iterations with a modified Newton method with the approximate Jacobian

$$(8.8)\qquad \begin{pmatrix} I_s \otimes (M(y_0) + O(h)) & I_s \otimes G^T(y_0) \\ A^2 \otimes \left(G^T(y_0)\right)^- \cdot \nabla^2 U(y_0) & -\frac{\epsilon^2}{h^2} I_s \otimes I_m \end{pmatrix} ,$$

which will be seen to have a uniformly bounded inverse as $\epsilon/h \to 0$. The convergence properties of these iterations are summarized in the following.

Lemma 8.2. (Convergence of iterations (8.5)-(8.8)) *Suppose that the starting values satisfy* $\nabla U(y_0) = O(h^2)$, $\nabla^2 U(y_0)\dot{y}_0 = O(h)$. *Then the errors of the outer iteration (8.5)-(8.7) satisfy for* $\epsilon \le ch$

$$\ddot{Y}_i^{(k)} - \ddot{Y}_i = O(\epsilon^2 h^{2k}) , \qquad k = 0, 1, 2, \dots ,$$

where $\ddot{Y}_i$ *is the solution of (5.1). The inner iteration with matrix (8.8) converges with rate* $O(h)$.

Remark. The first outer iteration thus already gives

$$Y_i^{(0)} - Y_i = O(\epsilon^2 h^2) , \quad \dot{Y}_i^{(0)} - \dot{Y}_i = O(\epsilon^2 h) .$$

If a further outer iteration is required, then care should be taken that $Y_i^{(0)}$ has been computed sufficiently accurately to account for the perturbation sensitivity in the computation of $r_i^{(1)}$ in (8.7). Formula (8.7) will usually be used only for $k = 0$ (or at the utmost $k = 1$), and unlike (8.3) there is no iteration with ill-conditioned matrices.

Proof. For the sake of brevity, we only outline the main points of the proof.
(a) In the study of the inner iteration with (8.8), the essential observation is that by (2.10), $H = \nabla^2 U$ can be written as $H = G^T K G$ with invertible K, and hence $(G^T)^- H = KG$. Since $\begin{pmatrix} M & G^T \\ G & 0 \end{pmatrix}$ is invertible, this implies that (8.8) has a uniformly bounded inverse. It then follows without difficulty that the modified Newton iteration for (8.6) converges with rate $O(h)$.

(b) The convergence analysis of the outer iteration relies on a bound of the difference between the solution of (8.6) and that of a perturbed system with additional defect $\delta = O(h)$ in the first equation of (8.6). This difference can be shown to be bounded by

$$(8.9)\qquad \Delta Y_i = O(h^2\delta) , \quad \Delta\dot{Y}_i = O(h\delta) , \quad \Delta\ddot{Y}_i + O(\delta) , \quad \Delta\Lambda_i = O(\delta) .$$

We now consider the defect when the exact values $\ddot{Y}_i$ are inserted into (8.6) with $k = 0$. With $\Lambda_i = \frac{1}{\epsilon^2} G^T(Y_i)^- \cdot \nabla U(Y_i)$, the defect is

$$G^T(Y_i)\Lambda_i - \frac{1}{\epsilon^2}\nabla U(Y_i) = \frac{1}{\epsilon^2}\left(G^T(Y_i) \cdot G^T(Y_i)^- - I_n\right) \nabla U(Y_i) .$$

By Lemma 5.1, we have $\ddot{Y}_i = O(1)$, and hence $\nabla U(Y_i) = O(\epsilon^2)$ by (5.1c). So there exists $Y_i^0 = Y_i + O(\epsilon^2)$ with $\nabla U(Y_i^0) = 0$, and we have

$$\nabla U(Y_i) = H(Y_i) \cdot (Y_i - Y_i^0) + O(\epsilon^4) \ .$$

Since by (2.10),

$$\big(G^T(G^T)^- - I_n\big) \cdot H = 0 \ , \tag{8.10}$$

the defect is $O(\epsilon^2)$, and (8.9) gives the result for $k = 0$. To proceed further one then shows with the help of (8.9) and (8.10)

$$G^T(Y_i^{(0)})\Lambda_i^{(0)} - \frac{1}{\epsilon^2}\nabla U(Y_i^{(0)}) = G^T(Y_i)\Lambda_i - \frac{1}{\epsilon^2}\nabla U(Y_i) + O(\epsilon^2 h^2) \ .$$

For $k = 1$, we can then use (8.9) with $\delta = O(\epsilon^2 h^2)$, and so on. □

Acknowledgement. This article is an outgrowth of earlier joint work with E. Hairer and M. Roche [6],[7]. I am grateful for their support during the early stages of this work and for their continuing interest. I thank K. Nipp and D. Stoffer for helpful discussions on Theorem 7.1.

References

[1] R. Abraham, J.E. Marsden, *Foundations of Mechanics.* Benjamin/Cumming Publ. Comp., Reading, Mass., 1978.

[2] K.E. Brenan, S.L. Campbell, L.R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations.* North-Holland, New York, 1989.

[3] J.C. Butcher, *On the implementation of implicit Runge-Kutta methods.* BIT 16 (1976), 237-240.

[4] M. Crouzeix, J. Rappaz, *On Numerical Approximation in Bifurcation Theory.* Masson, Paris, and Springer-Verlag, Berlin, 1990.

[5] E. Eich, M. Hanke, *Regularization methods for constrained mechanical multibody systems.* Preprint, Humboldt-Universität zu Berlin, 1991.

[6] E. Hairer, Ch. Lubich, M. Roche, *Error of Runge-Kutta methods for stiff problems studied via differential algebraic equations.* BIT 28 (1988), 678-700.

[7] E. Hairer, Ch. Lubich, M. Roche, *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods.* Springer LNM 1409, 1989.

[8] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems.* Springer-Verlag, Berlin, 1991.

[9] L. Jay, *Collocation methods for differential-algebraic equations of index 3.* Preprint, Section de Math., Univ. de Genève, 1992.

[10] L. Kalachev, R.E. O'Malley, Jr., *Regularization of nonlinear differential-algebraic equations.* Report 92-4, Dept. Appl. Math., Univ. of Washington, 1992.

[11] T. Kato, *Perturbation Theory for Linear Operators.* 2nd Ed., Grundlehren math. Wiss. 132, Springer-Verlag, Berlin, 1980.

[12] U. Kirchgraber, F. Lasagni, K. Nipp, D. Stoffer, *On the application of invariant manifold theory, in particular to numerical analysis.* Intern. Ser. Num. Math., Vol. 97 (1991), 189-197.

[13] M. Knorrenschild, *Regularisierung von differentiell-algebraischen Systemen – Theoretische und numerische Aspekte.* Doctoral dissertation, RWTH Aachen, 1988.

[14] P. Lötstedt, *On a penalty function method for the simulation of mechanical systems subject to constraints.* Report TRITA-NA-7919, Royal Inst. of Technology, Stockholm, 1979.

[15] D. Morgenstern, I. Szabó, *Vorlesungen über theoretische Mechanik.* Grundlehren math. Wiss. 112, Springer-Verlag, Berlin, 1961.

[16] K. Nipp, D. Stoffer, *Attractive invariant manifolds for maps: Existence, smoothness, and continuous dependence on the map.* Report 92-11, SAM, ETH Zürich, 1992.

[17] R.E. O'Malley, Jr., *Singular Perturbation Methods for Ordinary Differential Equations.* Springer-Verlag, New York, 1991.

Research Reports

No.	Authors	Title
92-04	Ch. Lubich	Integration of stiff mechanical systems by Runge-Kutta methods
92-03	M. Fey, R. Jeltsch, S. Müller	Stagnation point analysis
92-02	C. W. Schulz-Rinne, J. P. Collins, H. M. Glaz	Numerical Solution of the Riemann Problem for Two-Dimensional Gas Dynamics
92-01	R. J. LeVeque, K. M. Shyue	Shock Tracking Based on High Resolution Wave Propagation Methods
91-10	M. Fey, R. Jeltsch	Influence of numerical diffusion in high temperature flow
91-09	R. J. LeVeque, R. Walder	Grid Alignment Effects and Rotated Methods for Computing Complex Flows in Astrophysics
91-08	Ch. Lubich, R. Schneider	Time discretization of parabolic boundary integral equations
91-07	M. Pirovino	On the Definition of Nonlinear Stability for Numerical Methods
91-06	Ch. Lubich, A. Ostermann	Runge-Kutta Methods for Parabolic Equations and Convolution Quadrature
91-05	C. W. Schulz-Rinne	Classification of the Riemann Problem for Two-Dimensional Gas Dynamics
91-04	R. Jeltsch, J. H. Smit	Accuracy Barriers of Three Time Level Difference Schemes for Hyperbolic Equations
91-03	I. Vecchi	Concentration-cancellation and Hardy spaces
91-02	R. Jeltsch, B. Pohl	Waveform Relaxation with Overlapping Splittings