NTNU
Norwegian University of
Science and Technology

Numerical Methods for Nonholonomic Mechanics

Sindre Kristensen Hilden

Master of Science in Physics and Mathematics
Submission date: June 2009
Supervisor: Elena Celledoni, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

A nonholonomic system is, roughly speaking, a mechanical system whose evolution is constrained depending on both its current position and velocity. Qualitative properties of such systems, as for example energy preservation and time reversibility, should be maintained under numerical discretization. The aim of this thesis is to understand the basic theoretical features of nonholonomically constrained systems and discuss which numerical methods are best suited for such problems. A comparison of different structure preserving methods will be considered.

Assignment given: 19. January 2009
Supervisor: Elena Celledoni, MATH

Abstract

We discuss nonholonomic systems in general and numerical methods for solving them. Two different approaches for obtaining numerical methods are considered; discretization of the Lagrange-d'Alembert equations on the one hand, and using the discrete Lagrange-d'Alembert principle to obtain nonholonomic integrators on the other. Among methods using the first approach, we focus on the super partitioned additive Runge-Kutta (SPARK) methods [14, 16, 17]. Among nonholonomic integrators, we focus on a reversible second order method by McLachlan and Perlmutter [22].

Through several numerical experiments the methods we present are compared by considering error-growth, conservation of energy, geometric properties of the solution and how well the constraints are satisfied. Of special interest is the comparison of the 2-stage SPARK Lobatto IIIA-B method and the nonholonomic integrator by McLachlan and Perlmutter, which both are reversible and of second order.

We observe a clear connection between energy-conservation and the geometric properties of the numerical solution. To preserve energy in long-time integrations is seen to be important in order to get solutions with the correct qualitative properties. Our results indicate that the nonholonomic integrator by McLachlan and Perlmutter sometimes conserves energy better than the 2-stage SPARK Lobatto IIIA-B method. In a recent work by Jay [19], however, the same two methods are compared and are found to conserve energy equally well in long-time integrations.

Preface

This thesis is written in the spring of 2009 at the Norwegian University of Science and Technology. It represents 20 weeks of work and completes the five years Master of Science program in Industrial Mathematics.

I would like to thank Laurent O. Jay for the plots we use for comparison in Section 4.5 and Robert McLachlan for the initial conditions used in the contact oscillator experiments in [22], which we use in Section 4.3 and 4.4. I would also like to thank my fellow student Martin Børter for proofreading. Many special thanks to supervisor Associate Professor Elena Celledoni for interesting discussions, help and feedback while working on this thesis.

Trondheim, June 12, 2009

Sindre Hilden

Contents

Introduction

A nonholonomic system is a mechanical system subject to nonholonomic constraints. Such constraints typically arise whenever there is sliding contact or rolling contact involved, e.g. if the system contains skates or wheels. One example of a nonholonomic system that we will study later, is a disk rolling on a horizontal plane without slipping (think of a coin rolling on a table).

In the literature nonholonomic mechanics is often studied from a geometric viewpoint, requiring the reader to be familiar with differential geometry. We aim to introduce nonholonomic mechanics at a more basic level. This thesis requires a general knowledge about linear algebra, partial differential equations and numerics in general, but it should be readable without any familiarity to manifolds.

The main objective of this thesis is to study numerical methods for solving nonholonomic systems. Different methods are presented and investigated through numerical experiments. Nonholonomic systems possess certain important properties. How well the numerical solutions preserve these properties plays a key role in determining the success of the methods.

Both unconstrained and constrained mechanical systems are discussed in Chapter 1. In particular, we study nonholonomic systems. The Lagrange-d'Alembert equations are derived. These equations are the governing equations for a nonholonomic system.

Chapter 2 formulates the Lagrange-d'Alembert equations as a differential-algebraic equation (DAE) and methods using this formulation are considered. A class of methods called super partitioned additive Runge-Kutta (SPARK) methods is the highlight of this chapter.

Another approach for integrating nonholonomic systems is considered in Chapter 3. Instead of discretizing the equations themselves, the derivation of the equations is discretized and the result is the discrete Lagrange-d'Alembert equations. Methods based on these equations are referred to as nonholonomic integrators and two such methods are presented.

Different numerical experiments are carried out in Chapter 4. The methods considered in the preceding chapters are put to the test and the properties of the different methods are analyzed and compared.

In Chapter 5 we summarize our work and draw some conclusions based on our results. Possibilities for further work are also discussed.

Chapter 1

Nonholonomic Mechanics

In this chapter, we first look at unconstrained mechanical systems. The movement of such systems is governed by Hamilton's equations. These equations are derived from Hamilton's principle. Next, we consider systems subject to constraints, and we in particular clearify what a nonholonomic system is. From the Lagrange-d'Alembert principle, the Lagrange-d'Alembert equations are derived. These equations govern the evolution of nonholonomic systems and solving them is the subject of the coming chapters.

1.1 Mechanical Systems

In this section, mechanical systems in general and how such systems are represented mathematically are discussed. An example is also introduced that will be used throughout this thesis to illustrate theory presented.

1.1.1 Generalized Coordinates

The *configuration* of a mechanical system is determined by the position and orientation in 3-space of all the physical parts of the system. At any given point in time, a mechanism has a particular configuration. The set of *all* physically possible configurations make up the *configuration space*. The configuration space is a manifold[1] and therefore also referred to as the *configuration manifold*. To describe a mechanical system more formally, a set of *generalized coordinates* are used, which is interpreted as the coordinates for the configuration space. These coordinates are minimal in the sense that no set of fewer coordinates are able to fully describe the system. The number of coordinate variables is called the number of *degrees of freedom* of the system and it is also the dimension of the configuration space. A set of generalized coordinates need not be Cartesian, but can be any set of coordinates describing the configuration of the system in a *unique* way.

Notation. The configuration space is denoted by Q and one particular configuration in this space by $q \in Q$. An element q is represented as a column vector of length n, where n is the dimension of Q. Components are denoted by superscript, so that the i'th component of q is denoted q^i. A mechanical system will change with time in a continuous way. At each time t, the current configuration of a system is denoted by $q(t) \in Q$. This way, q can

[1] If the reader is unfamiliar with manifolds and differential geometry, an introduction can be found in e.g. [24]. See also Appendix A.2. However, differential geometry is not of great importance in this thesis and it can easily be read without any knowledge about the subject.

be seen as a function $q\colon \mathbb{R} \to Q$ that will trace out a continuous curve in the configuration space Q.

1.1.2 Example: The Vertical Rolling Disk

To see a concrete example of a mechanical system, its configuration space and how a particular set of coordinates can be put on this space, the vertical rolling disk [1] is considered. This example consists of a disk rolling on a horizontal plane without slipping and without falling over. This is of course not a physically realistic situation as any real world disk would (eventually) fall over, but it is an illustrative example. The extension to the rolling disk that is allowed to fall over is not very difficult (see [1]).

If S^1 denotes the unit circle, the configuration space of this system is given by $Q = \mathbb{R}^2 \times S^1 \times S^1$. One possible choice of coordinates is given by $q = (x, y, \varphi, \theta)$, where (x, y) is the point of contact, φ is the orientation of the disk and θ is the rotation angle of the disk, as shown in Figure 1.1. This given choice of coordinates is not the only possibility.

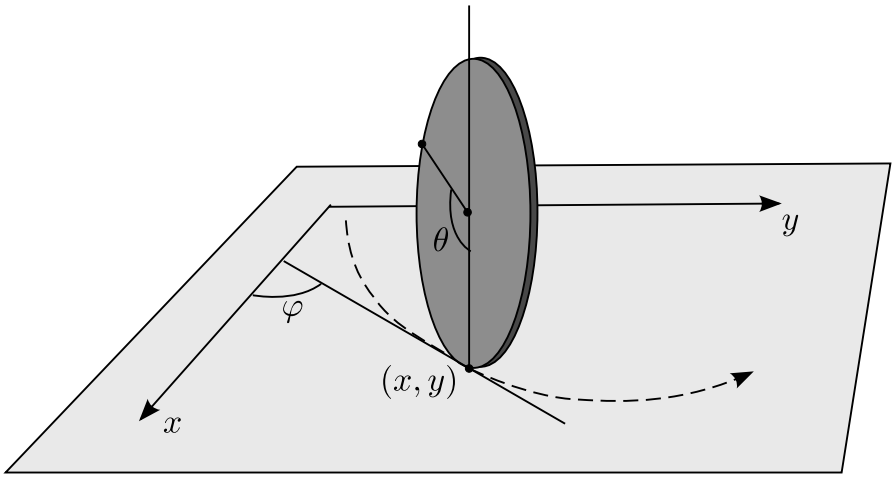

Figure 1.1: The vertical rolling disk with chosen coordinates.

Another choice could be to use polar coordinates to describe the point of contact instead of using Cartesian coordinates, or the orientation of the disk could have been given relative to the y-axis instead of the x-axis. The possibilities are many, but any choice would always consist of precisely four coordinates.

1.2 Hamilton's Principle

How an unconstrained mechanical system evolves in time is determined by Hamilton's principle. In this section, Hamilton's principle is first stated and then the Euler-Lagrange equations are derived from this principle.

Consider a mechanical system described by generalized coordinates q. Fix the time-interval $[0, T]$ and fix the initial and final states $q_0 = q(0)$ and $q_T = q(T)$. Consider continuous curves $q(t)$ in Q which connect these two points. Figure 1.2 shows one such continuous curve. Denote the time-derivative of q by $\dot{q}$ and let the kinetic energy of the system be given by $K = K(q, \dot{q})$ and the potential energy by $V = V(q)$. In a manifold

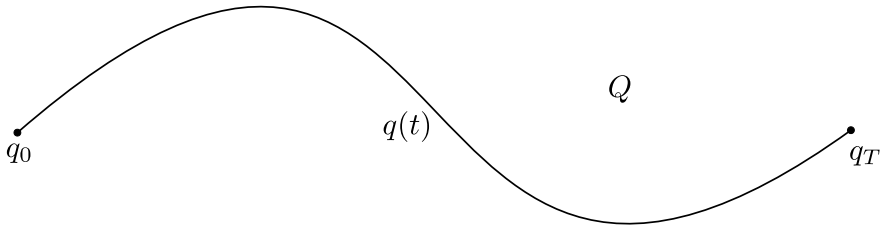

Figure 1.2: A curve $q(t)$ in Q with fixed endpoints q_0 and q_T.

setting, $\dot{q}$ is a tangent vector at the point q, i.e. $\dot{q} \in T_qQ$. Like q, also $\dot{q}$ is represented by a column vector of length n. The kinetic energy is often of the form

$$K(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q}, \tag{1.1}$$

where $M(q)$ is an $n \times n$ symmetric positive definite matrix called the *mass-matrix* [10]. Associated with such a system is a real-valued function $L\colon TQ \to \mathbb{R}$, called a *Lagrangian*, which is defined[2] by

$$L(q, \dot{q}) = K(q, \dot{q}) - V(q).$$

Given a Lagrangian $L(q, \dot{q})$, the *action functional* S is defined as

$$S(q(t)) = \int_0^T L(q(t), \dot{q}(t))\, dt. \tag{1.2}$$

This is also referred to as the *action integral*.

Definition 1.1 (Hamilton's Principle)**.** The true evolution $q(t)$ of a system described by the general coordinates q, with fixed endpoints $q_0 = q(0)$ and $q_T = q(T)$, is the curve $q(t)$ that extremizes the action functional $S(q(t))$ given by (1.2).

Said in a different way, among all possible continuous curves in Q connecting the two endpoints, the curve that represents the physically correct evolution of the system, is the curve that extremizes the action functional. To make Hamilton's principle a bit more precise, the approach taken in [1] is used. Let the *variation* of a curve $q(\cdot)$ with fixed endpoints be a smooth mapping

$$(t, \epsilon) \mapsto q(t, \epsilon), \quad 0 \le t \le T, \quad \epsilon \in (-\delta, \delta) \subset \mathbb{R}$$

such that $q(t, 0) = q(t)$ for $t \in [0, T]$. This concept of variations of the curve $q(t)$ is illustrated in Figure 1.3 on the following page. Let the *virtual displacement* $\delta q(t)$ corresponding to the variation of q be defined as

$$\delta q(t) \doteq \frac{\partial}{\partial \epsilon} q(t, \epsilon)\bigg|_{\epsilon=0}.$$

Since the endpoints of the curve q are fixed, it follows that $\delta q(0) = \delta q(T) = 0$. Hamilton's principle can then be restated in a precise way as follows. The true evolution $q(t)$ of a mechanical system satisfies

$$0 = \delta S \doteq \frac{d}{d\epsilon} \int_0^T L(q(t, \epsilon), \dot{q}(t, \epsilon))dt\bigg|_{\epsilon=0}, \tag{1.3}$$

[2]It is noted in [1] that $L(q, \dot{q}) = K(q, \dot{q}) - V(q)$ is the appropriate definition of the Lagrangian for most systems, but that for some systems, such as a particle in a magnetic field, a Lagrangian which is not of this form has to be chosen.

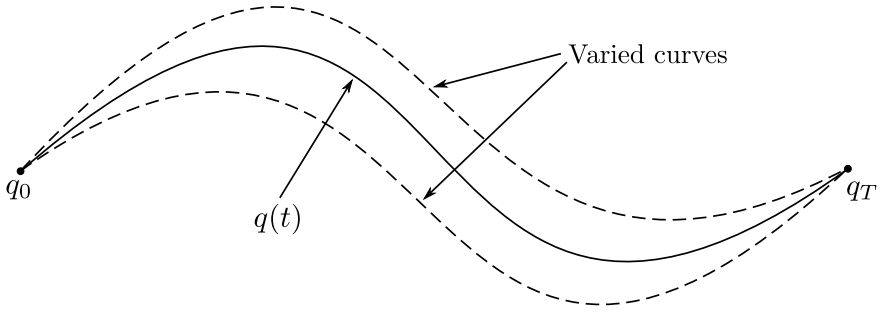

Figure 1.3: Variations of the curve $q(t)$.

for all variations of q.

Proposition 1.2. *Hamilton's principle for a curve $q(t)$, given in Definition 1.1, is equivalent to the condition that $q(t)$ satisfies the* Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = 0. \tag{1.4}$$

These equations are also referred to simply as the Lagrange equations*.*

Proof. Hamilton's principle is equivalent to (1.3). Using the chain rule yields

$$0 = \delta S = \int_0^T \left(\frac{\partial L}{\partial q} \cdot \frac{\partial q}{\partial \epsilon}\bigg|_{\epsilon=0} + \frac{\partial L}{\partial \dot{q}} \cdot \frac{\partial \dot{q}}{\partial \epsilon}\bigg|_{\epsilon=0} \right) dt,$$

and integrating by parts the integral involved in the second sum gives

$$0 = \delta S = \int_0^T \frac{\partial L}{\partial q} \cdot \delta q \; dt + \underbrace{\left[\frac{\partial L}{\partial \dot{q}} \cdot \delta q \right]_0^T}_{=0} - \int_0^T \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right) \cdot \delta q \; dt = \int_0^T \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right) \cdot \delta q \; dt$$

for all virtual displacements δq. The second term vanishes because $\delta q(0) = \delta q(T) = 0$. The dot $\cdot$ denotes the inner product. Since δq is arbitrary, the Euler-Lagrange equations (1.4) follow from the so-called fundamental lemma of the calculus of variations. □

1.2.1 Legendre Transformation and Hamilton's Equations

By a change of variables, it is possible to rewrite the Euler-Lagrange equations into a very symmetric form [10]. The result of this change of variables is Hamilton's equations.

If the Hessian matrix

$$\frac{\partial^2 L}{\partial \dot{q}^2} = \left(\frac{\partial^2 L}{\partial \dot{q}^i \partial \dot{q}^j} \right)_{i,j=1,\ldots,n}$$

is invertible, the Lagrangian L is called a *regular* Lagrangian [1]. For example, a Lagrangian of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$, where $M(q)$ is positive definite, is a regular Lagrangian [25]. Assume now that L is regular. With this assumption, a change

of variables from $(q, \dot{q})$ to the variables (q, p) can be made, where p is the momentum variables defined by

$$p = \frac{\partial L}{\partial \dot{q}}. \tag{1.5}$$

This change of variables is called the *Legendre transformation* [1, 10]. The *Hamiltonian* is a real-values function defined by

$$H(q, p) = p^T \dot{q} - L(q, \dot{q}), \tag{1.6}$$

where $\dot{q} = \dot{q}(q, p)$ is expressed as a function of q and p using the Legendre transform (1.5).

Proposition 1.3. *The Euler-Lagrange equations* (1.4) *are equivalent to Hamilton's Equations*

$$\dot{q} = \frac{\partial H}{\partial p}(q, p), \quad \dot{p} = -\frac{\partial H}{\partial q}(q, p). \tag{1.7}$$

Proof. This proof is found in [10]. Using the definition of the Hamiltonian (1.6), the Legendre transform (1.5) and the Euler-Lagrange equations (1.4),

$$\begin{aligned}
\frac{\partial H}{\partial p} &= \dot{q} + \left(\frac{\partial \dot{q}}{\partial p}\right)^T p - \left(\frac{\partial \dot{q}}{\partial p}\right)^T \frac{\partial L}{\partial \dot{q}} \overset{(1.5)}{=} \dot{q}, \\
\frac{\partial H}{\partial q} &= \left(\frac{\partial \dot{q}}{\partial q}\right)^T p - \frac{\partial L}{\partial q} - \left(\frac{\partial \dot{q}}{\partial q}\right)^T \frac{\partial L}{\partial \dot{q}} \overset{(1.5)}{=} -\frac{\partial L}{\partial q} \overset{(1.4)}{=} -\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \overset{(1.5)}{=} -\dot{p}.
\end{aligned}$$

So the Euler-Lagrange equations and Hamilton's equations are equivalent. □

In the particular case where the Lagrangian is of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$, where $M(q)$ is positive definite, the Legendre transform is given by

$$p = \frac{\partial L}{\partial \dot{q}} = M(q)\dot{q}$$

and thus

$$\dot{q} = M(q)^{-1} p.$$

Replacing $\dot{q}$ by $M(q)^{-1}p$ in the definition of the Hamiltonian (1.6) yields

$$\begin{aligned}
H(q, p) &= p^T M(q)^{-1} p - L(q, M(q)^{-1}p) = p^T M(q)^{-1} p - \frac{1}{2} p^T M(q)^{-1} p + V(q) \\
&= \frac{1}{2} p^T M(q)^{-1} p + V(q) = K(q, M(q)^{-1}p) + V(q) = K\big(q, \dot{q}(q, p)\big) + V(q).
\end{aligned}$$

Thus, the Hamiltonian is equal to the kinetic energy plus the potential energy, which is the *total energy* of the system.

1.2.2 Properties of Unconstrained Systems

The solution of Hamilton's equations (1.7) has some important properties. Some of the most important properties are described below.

Energy Conservation. Thinking of the Hamiltonian as a function of $(q, \dot{q})$ instead of (q, p), it can be written as $E(q, \dot{q})$ and referred to as *energy* [1]. Said in another way, the energy is defined as

$$E(q, \dot{q}) = \dot{q}^T \frac{\partial L}{\partial \dot{q}}(q, \dot{q}) - L(q, \dot{q}).$$

As seen above, in the particular case where $L = K - V = \frac{1}{2}\dot{q}^T M \dot{q} - V$, this becomes $E = K + V$, which makes this definition of energy seem like a natural choice. The energy is a conserved quantity since

$$\begin{aligned} \frac{d}{dt}E(q, \dot{q}) &= \left(\frac{\partial L}{\partial \dot{q}}\right)^T \frac{d\dot{q}}{dt} + \dot{q}^T \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right) - \dot{q}^T \frac{\partial L}{\partial q} - \left(\frac{\partial L}{\partial \dot{q}}\right)^T \frac{d\dot{q}}{dt} \\ &= \dot{q}^T \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q}\right) \overset{(1.4)}{=} 0. \end{aligned}$$

That is, for an unconstrained system, the total energy is constant in time.

Reversibility. The *flow* of a dynamical system is the mapping which associates an initial condition $y(0) = y_0$ to the solution $y(t)$ of the system at a later time t [10]. That is, the flow of a system is denoted by φ_t, it satisfies $\varphi_t(y_0) = y(t)$.

The flow φ_t of a dynamical system is *reversible* if

$$\varphi_t \circ \rho \circ \varphi_t = \rho$$

where $\rho(q, v) = (q, -v)$. This definition is illustrated in Figure 1.4. A reversible flow is also referred to as *time-reversible* or *symmetric*. The flow of Hamilton's equations (1.7) is reversible [10, V.1].

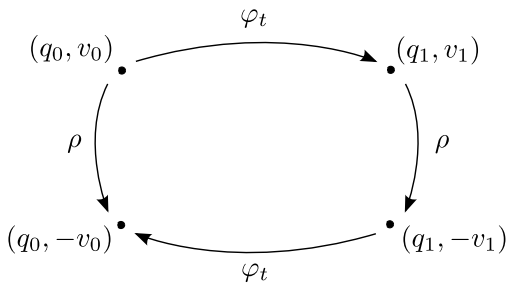

Figure 1.4: Illustration of a reversible (symmetric) flow φ_t.

Symplecticity. Let $J \in \mathbb{R}^{2n}$ be defined by

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

and let $U \subset \mathbb{R}^{2n}$. A differentiable map $g \colon U \to \mathbb{R}^{2n}$ is *symplectic* if the Jacobian matrix $g'(q, p)$ satisfies

$$g'(q, p)^T J g'(q, p) = J.$$

The flow of Hamilton's equations (1.7) is symplectic (see [10, VI.2] for a proof).

1.3 The Lagrange-d'Alembert Principle

First in this section, the definition of a holonomic and a nonholonomic system is given. As we are in particular interested in nonholonomic systems, this is what we focus on. Using the Lagrange-d'Alembert principle, the Lagrange-d'Alembert equations are derived, which can be seen as a generalization of the Euler-Lagrange equations to the nonholonomic case. At the end of this section, holonomic systems are also briefly discussed.

1.3.1 Holonomic and Nonholonomic Constraints

Mechanical systems subject to one or more constraints are now considered. A constraint is a condition imposed on the system, which limits its evolution or possible configurations. A constraint is either classified as holonomic or nonholonomic.

Consider a mechanical system described by generalized coordinates $q^j, j = 1, \ldots, n$, subject to $m < n$ constraints. Assume that the constraint equations are given by

$$\sum_{j=1}^{n} a_{ij}\dot{q}^j = 0, \quad i = 1, \ldots, m, \tag{1.8}$$

where the coefficients a_{ij} depend on q. That is, the constraint equations are assumed to be linear in the velocity variables $\dot{q}^j$. These constraints are called *holonomic* if it is possible to find m constraints on the position alone, which means it is possible to find m constraints

$$b_i(q) = 0, \quad i = 1, \ldots, m,$$

such that the time-derivative of these constraints equal (1.8), i.e.

$$\sum_{j=1}^{n} \frac{\partial b_i}{\partial q^j}\dot{q}^j = 0, \quad i = 1, \ldots, m.$$

If this is *not* possible, then the constraints (1.8) are *nonholonomic*. For example, the constraint that the distance between any two particles in a rigid body is fixed is a holonomic constraint, whereas the constraint of rolling without slipping (like in the vertical disk example) is a nonholonomic constraint. A mechanical system subject to holonomic constraints is referred to simply as a *holonomic system* and a system subject to nonholonomic constraints is referred to as a *nonholonomic system*.

By defining b as the column vector of length m, such that its i'th element of b is given by b_i, the holonomic constraints can be written simply as

$$b(q) = 0. \tag{1.9}$$

In a similar way, define A to be the $m \times n$ matrix such that its (i, j)-entry is given by a_{ij}. Then the nonholonomic constraints can be written as

$$A(q)\dot{q} = 0. \tag{1.10}$$

This concise notation is used hereafter.

1.3.2 The Lagrange-d'Alembert Principle

For unconstrained mechanical systems, Hamilton's principle determines how to find the physically correct evolution of the system. In a similar way for nonholonomic systems, the Lagrange-d'Alembert principle determines the true evolution.

Definition 1.4 (The Lagrange-d'Alembert Principle)**.** A curve $q(t)$ is an admissible evolution of the system if

$$\delta \int_0^T L(q(t), \dot{q}(t))dt = 0 \tag{1.11}$$

for all virtual displacements $\delta q(t)$, where $\delta q(0) = \delta q(T) = 0$, that satisfy the constraints for all times, i.e. $A(q(t))\delta q(t) = 0$ for all $t \in [0, T]$.

In a similar way that Hamilton's principle is shown to be equivalent to the Euler-Lagrange equations, the Lagrange-d'Alembert principle is shown to be equivalent to the Lagrange-d'Alembert equations, which can be seen as a generalization of the Euler-Lagrange equations to the nonholonomic case. Assuming that the constraint matrix $A(q)$ has full rang, the following proposition states the eqivalence.

Proposition 1.5. *The Lagrange-d'Alembert principle given in Definition 1.4, together with the constraints* (1.10)*, is equivalent to the* Lagrange-d'Alembert equations of motion

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = A(q)^T\lambda, \tag{1.12a}$$

$$A(q)\dot{q} = 0. \tag{1.12b}$$

Proof. The constraints (1.12b) are simply the same as (1.10), so what needs to be shown is that the Lagrange-d'Alembert principle is equivalent to (1.12a). From the proof of Proposition 1.3,

$$0 = \delta S = \int_0^T \left(\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}\right) \cdot \delta q \, dt,$$

where now the virtual displacements δq^i are not independent, as they are in the unconstrained case. The above equation implies

$$\sum_{i=1}^{n} \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i}\right) \delta q^i = 0, \tag{1.13}$$

for all virtual displacements $\delta q(t)$ such that $A(q)\delta q = 0$ for each time $t \in [0, T]$. Now, for notational simplicity, let w be the column vector

$$w \doteq \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q}.$$

Equation (1.13) can then be written as $w^T\delta q = 0$ for all δq s.t. $A(q)\delta q = 0$. From the definition of kernel, we have that $A(q)\delta q = 0$ is equivalent to $\delta q \in \ker(A(q))$, and so

$$(1.13) \iff w \perp \ker\big(A(q)\big) \iff w \in \Big(\ker\big(A(q)\big)\Big)^{\perp} = \operatorname{im}\big(A(q)^T\big).$$

The last equality follows from the fundamental theorem of linear algebra. Then, since w is in the image of $A(q)^T$, there exists a $\lambda \in \mathbb{R}^m$ such that $w = A(q)^T\lambda$, which is exactly (1.12a). □

Remark. It is worth mentioning that the Lagrange-d'Alembert principle is not truly variational [1, 5]. The Lagrange-d'Alembert principle says to impose the constraints on the virtual displacements δq. The corresponding variational approach is to instead impose the constraints on the velocity vectors $\dot{q}$. This gives rise to the so-called *vakonomic mechanics* and the *variational nonholonomic equations*. These equations are in general *not* equivalent to the Lagrange-d'Alembert equations. Vakonomic mechanics is used as the setting for some optimization problems, whereas the Lagrange-d'Alembert principle should be used to obtain the correct dynamics for nonholonomic systems [1]. See [1, Chapter 1.3] and [5, Chapter 3.3] for more on the distinction between these two approaches.

1.3.3 Hamiltonian Formulation

As for unconstrained systems, if it again is assumed that the Hessian matrix $\partial^2 L/\partial \dot{q}^2$ is invertible, the Lagrange-d'Alembert equations can be rewritten using the Hamiltonian $H = p^T\dot{q} - L(q, \dot{q})$ and the Legendre transformation $p = \partial L/\partial \dot{q}$. Using the same approach as in the proof of Proposition 1.3, but now using the Lagrange-d'Alembert equations (1.12) instead of the Euler-Lagrange equations, it is found that

$$\begin{aligned}\frac{\partial H}{\partial p} &= \dot{q},\\ \frac{\partial H}{\partial q} &= -\frac{\partial L}{\partial q} \overset{(1.12)}{=} -\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} + A^T\lambda = -\dot{p} + A^T\lambda,\end{aligned}$$

so the Lagrange-d'Alembert equations are equivalent to

$$\dot{q} = \frac{\partial H}{\partial p}(q,p), \quad \dot{p} = -\frac{\partial H}{\partial q}(q,p) + A(q)^T\lambda.$$

The nonholonomic constraints can be written in terms of q and p as

$$A(q)\frac{\partial H}{\partial p}(q,p) = 0.$$

In the particular case where the Lagrangian is of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$, where $M(q)$ is positive definite, these equations become

$$\dot{q} = M(q)^{-1}p, \quad \dot{p} = -\frac{\partial H}{\partial q}(q,p) + A(q)^T\lambda = \frac{\partial L}{\partial q}(q, M(q)^{-1}p) + A(q)^T\lambda,$$

with constraints

$$A(q)M(q)^{-1}p = 0.$$

These equations are returned to in Section 2.2.2.

1.3.4 Properties of a Nonholonomic System

Three important properties of an unconstrained system are discussed in Section 1.2.2 — the energy of the system is conserved and the flow is reversible and symplectic. Nonholonomic systems also conserve energy and the flow is reversible. However, the flow is *not* symplectic [5].

When solving nonholonomic systems numerically, it is natural to search for methods that produce numerical solutions with these properties. In later chapters, when considering different numerical methods, how well these methods preserve energy and whether they are reversible or not play a key role in determining their success.

Energy conservation. As for unconstrained systems, the energy is defined as

$$E(q,\dot{q}) = \dot{q}^T \frac{\partial L}{\partial \dot{q}}(q,\dot{q}) - L(q,\dot{q}).$$

For a nonholonomic system with constraints that are linear in the velocity coordinates, i.e. the constraints can be written $A(q)\dot{q} = 0$, the energy of the system is conserved. This is seen by

$$\frac{d}{dt}E(q,\dot{q}) = \dot{q}^T \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} \right) \overset{(1.12a)}{=} \dot{q}^T A(q)^T \lambda = \Big(A(q)\dot{q} \Big)^T \lambda \overset{(1.12b)}{=} 0.$$

As before, if the Lagrangian is given by $L = \frac{1}{2}\dot{q}^T M \dot{q} - V$, the energy E equals the sum of the kintic energy and the potential energy.

Reversibility. Recall from Section 1.2.2 that the flow φ_t of a dynamical system is *reversible* [10] if

$$\rho \circ \varphi_t = \varphi_{-t} \circ \rho,$$

where $\rho(q,v) = (q,-v)$. If the Lagrangian is of the form $L = K - V$, where K is the kinetic energy and V is the potential energy, the flow of the Lagrange-d'Alembert equations is reversible [22].

1.3.5 Example: The Vertical Rolling Disk Revisited

The vertical rolling disk presented in Section 1.1.2 is now revisited to see an example of nonholonomic constraints and a Lagrangian and also to see what the Lagrange-d'Alembert equations look like in this case. Recall that this example consits of a vertical disk rolling on a horizontal plane without slipping and without falling over. The coordinates for this system are given by $q = \begin{bmatrix} x & y & \varphi & \theta \end{bmatrix}^T$. The system is depicted in Figure 1.1 on page 4.

The restriction that the disk cannot slip actually translates into two nonholonomic constraints. If the radius of the disk is R, then these constraints are

$$\begin{aligned} 0 &= \dot{x} - R(\cos\varphi)\dot{\theta}, \\ 0 &= \dot{y} - R(\sin\varphi)\dot{\theta}. \end{aligned}$$

Defining

$$A(q) = \begin{bmatrix} 1 & 0 & 0 & -R\cos\varphi \\ 0 & 1 & 0 & -R\sin\varphi \end{bmatrix},$$

these constraints can be written in matrix notation as $A(q)\dot{q} = 0$. As the disk cannot fall over, there is no gravity working and hence no potential energy. The Lagrangian is therefore given by the kinetic energy only and it reads

$$L(q,\dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) + \frac{1}{2}J\dot{\varphi}^2 + \frac{1}{2}I\dot{\theta}^2, \tag{1.14}$$

where m is the mass of the disk, J is the moment of inertia about an axis in the plane of the disk and I is the moment of inertia of the disk about the axis perpendicular to the

plane of the disk. Both axes go through the center of the disk. From (1.14), it is found that the mass matrix $M(q)$ is given by the diagonal matrix (zero elements are left out)

$$M(q) = \begin{bmatrix} m & & & \\ & m & & \\ & & J & \\ & & & I \end{bmatrix}. \tag{1.15}$$

This matrix is symmetric and positive definite as long as $m, J, I > 0$. If the Lagrangian multipliers are written as $\lambda = \begin{bmatrix} \lambda_1 & \lambda_2 \end{bmatrix}^T$, the Lagrange-d'Alembert equations (1.12) become

$$\begin{aligned} m\ddot{x} &= \lambda_1, \\ m\ddot{y} &= \lambda_2, \\ J\ddot{\varphi} &= 0, \\ I\ddot{\theta} &= -R(\cos\varphi)\lambda_1 - R(\sin\varphi)\lambda_2, \\ 0 &= \dot{x} - R(\cos\varphi)\dot{\theta}, \\ 0 &= \dot{y} - R(\sin\varphi)\dot{\theta}. \end{aligned} \tag{1.16}$$

These are the equations of motion, which describe the evolution of the system given consistent initial values. Consistent initial values means that the initial conditions have to satisfy the nonholonomic constraints and also the time-derivative of the constraint equations. This is discussed more in Section 2.1.

1.3.6 Holonomic systems

As we in this thesis only is concerned with solving nonholonomic systems, we will now give a brief overview of some ways to solve *holonomic* systems numerically. See the references given below for more on this subject.

Holonomic systems are mechanical systems subject to constraints on the position alone. Say we have as before $m < n$ constraints, where n is the number of coordinates. Mathematically, it is then possible to use these constraints to eliminate m of the coordinates of the system as we can use the constraints to solve for one coordinate at a time. Doing this leaves us with an unconstrained system of $n - m$ variables, which can be solved using the Euler-Lagrange equations.

However, even though a holonomic system is mathematically equivalent to an unconstrained system, there are many numerical methods that do not use this approach. Some of these methods are mentioned below.

SHAKE and RATTLE. The SHAKE method is proposed to solve the equations of motion for a holonomic system written in an Hamiltonian form. It is assumed that the Hamiltonian can be written as

$$H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q).$$

The SHAKE algorithm is a two-step method and an extension of the Störmer/Verlet scheme (also known as the leap-frog method).

A modification of the SHAKE algorithm, called RATTLE, has been proposed. This is a one-step method, which results in less round-off errors and the solution also lies on the correct manifold. The RATTLE algorithm can be extended to general Hamiltonians.

Both these methods are of second order and symmetric. They are very popular and widely used. See e.g. [10] for more on SHAKE and RATTLE.

SPARK Methods. In Section 2.5, a class of methods proposed in [14] called SPARK methods is discussed. There, these methods are only discussed in the context of nonholonomic systems, but the methods can handle holonomic systems just as well. They can also handle mixed holonomic and nonholonomic constraints.

Variational Integrators. In Section 3.2, a discretization of the Lagrange-d'Alembert principle is discussed and a discrete version of the Lagrange-d'Alembert equations is derived. These discrete equations are referred to as a variational integrator (or a nonholonomic integrator in the nonholonomic case). There is a choice on how to disctretize the Lagrangian and the constraints, and depending on this choice, different variational integrators with different properties are obtained. Variational integrators for holonomic systems are discussed in e.g. [21, 27].

Chapter 2

Numerical Integration of Index 2 DAEs

A differential-algebraic equation (DAE) is a class of differential equations. First, a more formal definition of differential-algebraic equations is given in Section 2.1, and then in Section 2.2, it is shown how the Lagrange-d'Alembert equations can be written as a DAE, both using a Lagrangian and a Hamiltonian formulation. In the rest of this chapter, numerical methods that solve the Lagrange-d'Alembert equations using these DAE formulations are considered.

2.1 Differential-Algebraic Equations (DAEs)

A description of differential-algebraic equations (or *DAEs* for short) is now given. In particular, what is called index 2 DAEs is considered. This is of special interest to us since the Lagrange-d'Alembert equations (1.12) are an example of an index 2 DAE (this is shown below). In the following, the discussion of DAEs in [9] is mainly used. See also [11] for more on DAEs.

A *differential-algebraic equations* (*DAE*) is in a general form an implicit differential equation

$$F(\dot{Y}, Y) = 0, \tag{2.1}$$

where F and Y are of the same dimensions. F is assumed to have sufficiently many bounded derivatives and $Y = Y(x)$, where x is the independent variable. The dot notation in (2.1) means differentiation with respect to x. Variables (i.e. elements of Y) for which derivatives are present in the equations are called *differential variables* and those for which no derivatives are present are called *algebraic variables*. This explains the name differential-algebraic equations. The non-autonomous system $F(\dot{Y}, Y, x) = 0$ can be written in the form (2.1) by adding the equation $\dot{x} = 1$.

It is common to classify different types of DAEs by the concept of index. The index of a DAE is a nonnegative integer which says something about the system of equations. In general, the higher the index is, the more difficulties in solving the DAE numerically one can expect [4]. There are actually several definitions of index. For simple problems, these definitions are identical, but for complicated systems, they can be different [4]. Index 0 DAEs are in all cases defined to be a system of ODEs.

One definition of index is the *differential index*, which is defined as (loosely speaking) the minimum number of times one has to differentiate equations in the DAE in order to

obtain a system of ODEs. This is an intuitively easy way of thinking of index. We will on the other hand use the *perturbation index* as our definition[1]. Actually, the definition of perturbation index is not given here (it can be found in [9]), but rather just the following result.

A system of the form

$$\dot{y} = f(y, z), \tag{2.2a}$$
$$0 = g(y), \tag{2.2b}$$

where

$$g_y(y) f_z(y, z) \text{ is invertible} \tag{2.3}$$

in a neighborhood of the solution is an index 2 DAE. The DAE form (2.2) is called *Hessenberg index-2 form* [4]. Here, $y \in \mathbb{R}^n$ are the differential variables and $z \in \mathbb{R}^m$ are the algebraic variables. The independent variable is $t \in \mathbb{R}$, such that $\dot{y} = dy/dt$. Also, g_y denotes the $m \times n$ Jacobian matrix, such that the (i, j)-entry of g_y is $\partial g^i / \partial y^j$. Similarly, f_z denotes the $n \times m$ Jacobian matrix of f with respect to z (see also Appendix A.1 for a definition of a Jabobian).

Differentiating (2.2b) once with respect to the independent variable t and substituting $\dot{y}$ yields

$$0 = g_y(y) f(y, z). \tag{2.4}$$

The equation (2.2a) together with these differentiated constraints (2.4) are, if (2.3) is satisfied, an index 1 DAE [9]. Differentiating (2.4) once more with respect to y yields

$$0 = g_{yy}(y)(\dot{y}, \dot{y}) + g_y(y) f_y(y, z)\dot{y} + g_y(y) f_z(y, z)\dot{z}, \tag{2.5}$$

where we by $g_{yy}(y)(\dot{y}, \dot{y})$ mean the vector

$$g_{yy}(y)(\dot{y}, \dot{y}) = \begin{bmatrix} \sum_{i,j=1}^{n} \dfrac{\partial^2 g^1}{\partial y^i \partial y^j} \dot{y}^i \dot{y}^j \\ \vdots \\ \sum_{i,j=1}^{n} \dfrac{\partial^2 g^m}{\partial y^i \partial y^j} \dot{y}^i \dot{y}^j \end{bmatrix} \in \mathbb{R}^m.$$

Because of the assumption (2.3), it is possible to solve for the derivate $\dot{z}$ in (2.5) to get a system of ordinary differential equations for y and z.

The initial values y_0, z_0 have to satisfy the relation (2.2b) and (2.4). If they do, they are referred to as *consistent initial values*. Given that consistent inital values are chosen, a solution of (2.2) exists and it is locally unique [10]. This follows mainly from known results about existence and uniqueness of ordinary differential equations (see [10, VII.1] for more details).

[1] It is stated in [9] that differential index $\leq$ perturbation index $\leq$ differential index $+ 1$.

2.2 The Lagrange-d'Alembert Equations as Index 2 DAE

As we are interested in solving the Lagrange-d'Alembert equations (1.12), we now show that these equations can be written in the form (2.2) and that they in this case also satisfy (2.3). Both the Lagrangian and the Hamiltonian formulation are considered. It is assumed that the Lagrangian $L\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is of the form

$$L(q,\dot{q}) = K(q,\dot{q}) - V(q) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q),$$

where $M(q) \in \mathbb{R}^{n\times n}$ is a symmetric positive definite (SPD) matrix. It is also assumed that the constraint matrix $A(q) \in \mathbb{R}^{m\times n}$ has full rank, i.e. all the m constraints are linearly independent.

2.2.1 Lagrangian Formulation

With the above assumptions, the Lagrange-d'Alembert equations (1.12) become the constraint equations $A(q)\dot{q} = 0$ and

$$\frac{d}{dt}\Big(M(q)\dot{q}\Big) - \frac{\partial L}{\partial q}(q,\dot{q}) = A(q)^T\lambda,$$

which can be written

$$M(q)\ddot{q} = \frac{\partial L}{\partial q}(q,\dot{q}) - \frac{dM(q)}{dt}\dot{q} + A(q)^T\lambda.$$

Let $v \doteq \dot{q}$, define the variable y by $y^T = \begin{bmatrix} q^T & v^T \end{bmatrix}$ and let z be the Lagrangian multipliers λ. This implies that $y \in \mathbb{R}^{2n}$, $z \in \mathbb{R}^m$, and the maps $f\colon \mathbb{R}^{2n}\times\mathbb{R}^m \to \mathbb{R}^{2n}$ and $g\colon \mathbb{R}^{2n} \to \mathbb{R}^m$ are given by

$$f(y,z) = \frac{d}{dt}y = \begin{bmatrix}\dot{q}\\ \dot{v}\end{bmatrix} = \begin{bmatrix} v \\ M^{-1}(q)\left(\frac{\partial L}{\partial q}(q,v) - \frac{dM(q)}{dt}v + A(q)^T z\right)\end{bmatrix}, \tag{2.6}$$

and

$$g(y) = A(q)v. \tag{2.7}$$

With this, the Lagrange-d'Alembert equations are written on the Hessenberg index-2 form (2.2). As stated in Section 2.1, with consistent initial values, these equations have a locally unique solution if (2.3) is satisfied, i.e. if $g_y(y)f_z(y,z)$ is invertible. Proposition 2.1 shows that this is the case, given the standing assumptions on the Lagrangian.

Proposition 2.1. *If $M(q)$ is positive definite and $A(q)$ has full rank, the m-by-m matrix $g_y(y)f_z(y,z)$ with $f(y,z)$ given by (2.6) and $g(y)$ given by (2.7) is invertible.*

Proof. First, $g_y \in \mathbb{R}^{m\times 2n}$ and $f_z \in \mathbb{R}^{2n\times m}$ are given by

$$g_y = \begin{bmatrix}\frac{\partial}{\partial q}\Big(A(q)v\Big) & A(q)\end{bmatrix} \quad \text{and} \quad f_z = \begin{bmatrix} 0 \\ M(q)^{-1}A^T(q)\end{bmatrix}.$$

So, leaving out the dependence of variables, this gives $g_y f_z = AM^{-1}A^T$. As M is SPD, a Cholesky decomposition can be done. We write $M = LL^T$, where L is invertible (and lower triangular). For notational simplicity, let $C = (L^{-1})^T$. Then

$$g_y f_z = AM^{-1}A^T = A(LL^T)^{-1}A^T = A(L^{-1})^T L^{-1}A^T = (AC)(AC)^T$$

As C is invertible and A is assumed to have full rank, $\text{rank}(AC) = \text{rank}(A) = m$. Using that for any matrix X, $\text{rank}(XX^T) = \text{rank}(X)$, this implies that $\text{rank}(g_y f_z) = \text{rank}((AC)(AC)^T) = \text{rank}(AC) = m$, and so $g_y f_z$ has full rank and is thus invertible. □

2.2.2 Hamiltonian Formulation

Now, the momentum coordinates $p = \frac{\partial L}{\partial \dot{q}} = M(q)\dot{q}$ are used in place of the velocity coordinates v. The Hamiltonian is given by

$$H(q,p) = \frac{1}{2}p^T M(q)^{-1}p + V(q),$$

and the Lagrange-d'Alembert equations become (see Section 1.3.3 for details)

$$\begin{aligned}
\dot{q} &= H_p(q,p),\\
\dot{p} &= -H_q(q,p) + A(q)^T\lambda,\\
0 &= A(q)H_p(q,p).
\end{aligned}$$

Let y be defined by $y^T = \begin{bmatrix} q^T & p^T \end{bmatrix}$ and let as above z be the Lagrangian multipliers λ. This implies that

$$f(y,z) = \frac{d}{dt}y = \begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} M(q)^{-1}p \\ -H_q(q,p) + A(q)^T z \end{bmatrix} = \begin{bmatrix} M(q)^{-1}p \\ L_q(q, M(q)^{-1}p) + A(q)^T z \end{bmatrix} \tag{2.9}$$

and

$$g(y) = A(q)M(q)^{-1}p. \tag{2.10}$$

The following proposition shows that (2.3) is satisfied, which implies existence and uniqueness as before.

Proposition 2.2. *If $M(q)$ is positive definite and $A(q)$ has full rank, the m-by-m matrix $g_y(y)f_z(y,z)$ with $f(y,z)$ given by* (2.9) *and $g(y)$ given by* (2.10) *is invertible.*

Proof. In a similar way as in the case of the Lagrangian formulation, it is found that we also now $g_y f_z = AM^{-1}A^T$. The rest of the proof is the same as the proof for Proposition 2.1. □

It is seen above that the equations of motion for a mechanical system with only nonholonomic constraints are an index 2 DAE. If both holonomic and nonholonomic constraints are present, the system can still be written as an index 2 DAE by introducing some extra Lagrangian multipliers. But, if *only* holonomic constraints are present, the system is an index 3 DAE [9].

2.3 Index Reduction

When solving differential algebraic equations numerically, a possible technique is to use *index reduction* [10, 11]. Instead of solving the DAE directly, some or all of the equations are differentiated to obtain a DAE of lower index and then this system is solved instead. We are interested in solving the Lagrange-d'Alembert equations, which can be written as a Hessenberg index-2 DAE (see Section 2.2).

It is seen in Section 2.1 that by differentiating the nonholonomic constraints once, the resulting equations are an index-1 DAE. For example, MATLAB's solvers `ode15s` and `ode23t` can solve DAEs of index 1 (in addition to ODEs), so one possible way of solving the Lagrange-d'Alembert equations is to differentiate the constraints once and then use one of these integrators.

By differentiating the constraints once more, a system of ODEs is obtained. This system can be solved using any regular ODE solver. In MATLAB, there are many functions for doing this, including the popular `ode45`.

This approach of differentiating the constraints and solving the DAE of a lower index has a drawback. The differentiated constraints are imposed on the system, but the original constraints are *not* imposed. The solution will therefore in general *not* satisfy the original constraints. That is, the solution will not be on the configuration manifold Q. This is not desirable as the solution will not be correct and it may even not be a physically possible solution.

2.3.1 Example: Index Reduction of the Vertical Rolling Disk

To illustrate the use of index reduction, we again return to the vertical rolling disk. This example has earlier been discussed in Section 1.1.2 and 1.3.5. Now, the nonholonomic constraints are differentiated to obtain both the underlying index-1 DAE and the underlying system of ODEs.

Recall once again that this example consits of a vertical disk rolling on a horizontal plane without slipping and without falling over. The coordinates for this system are $q = \begin{bmatrix} x & y & \varphi & \theta \end{bmatrix}^T$. The equations of motion for this problem are given by equation (1.16) on page 13.

Define the variable ξ as the column vector

$$\xi \doteq \begin{bmatrix} x & y & \varphi & \theta & v_x & v_y & v_\varphi & v_\theta & \lambda_1 & \lambda_2 \end{bmatrix}^T.$$

Differentiating the two constraint equations (1.16) yields the index-1 DAE

$$\underbrace{\begin{bmatrix} 1 & & & & & & & & & \\ & 1 & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 1 & & & & & & \\ & & & & m & & & & & \\ & & & & & m & & & & \\ & & & & & & J & & & \\ & & & & & & & I & & \\ & & & & & & & & 0 & \\ & & & & & & & & & 0 \end{bmatrix}}_{M_1(\xi)} \underbrace{\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\varphi} \\ \dot{\theta} \\ \dot{v}_x \\ \dot{v}_y \\ \dot{v}_\varphi \\ \dot{v}_\theta \\ \dot{\lambda}_1 \\ \dot{\lambda}_2 \end{bmatrix}}_{d\xi/dt} = \underbrace{\begin{bmatrix} v_x \\ v_y \\ v_\varphi \\ v_\theta \\ \lambda_1 \\ \lambda_2 \\ 0 \\ 0 \\ \lambda_1/m + R\, v_\varphi v_\theta \sin\varphi \\ \lambda_2/m - R\, v_\varphi v_\theta \cos\varphi \end{bmatrix}}_{f_1(\xi)}, \tag{2.11}$$

where $M_1(\xi)$ is referred to as the mass-matrix. Note that $M_1(\xi)$ is a singular matrix. Differentiating the two constraint equations once more results in the following system of ODEs

$$\frac{d\xi}{dt} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\varphi} \\ \dot{\theta} \\ \dot{v_x} \\ \dot{v_y} \\ \dot{v_\varphi} \\ \dot{v_\theta} \\ \dot{\lambda_1} \\ \dot{\lambda_2} \end{bmatrix} = \underbrace{\begin{bmatrix} v_x \\ v_y \\ v_\varphi \\ v_\theta \\ \lambda_1/m \\ \lambda_2/m \\ 0 \\ 0 \\ -mR\, v_\varphi^2\, v_\theta \cos\varphi \\ -mR\, v_\varphi^2\, v_\theta \sin\varphi \end{bmatrix}}_{f_2(\xi)}. \tag{2.12}$$

An initial condition has to be consistent. That is, it has to satisfy both the original constraints and the differentiated constraints. Let the initial condition be given by

$$\xi(0) = \xi_0 \doteq \begin{bmatrix} x_0 & y_0 & \varphi_0 & \theta_0 & v_{x,0} & v_{y,0} & v_{\varphi,0} & v_{\theta,0} & \lambda_{1,0} & \lambda_{2,0} \end{bmatrix}^T.$$

Denote the initial orientation $v_{\varphi,0}$ by ω and the initial rotation speed $v_{\theta,0}$ by Ω. The initial values are then consistent if

$$\begin{aligned} v_{x,0} &= R\Omega\cos\varphi_0, & \lambda_{1,0} &= -m\omega v_{y,0}, \\ v_{y,0} &= R\Omega\sin\varphi_0, & \lambda_{2,0} &= m\omega v_{x,0}. \end{aligned}$$

We choose to set $x_0 = y_0 = \varphi_0 = \theta_0 = 0$, $\omega = 2$ and $\Omega = 1$. The last initial values $v_{x,0}$, $v_{y,0}$, $\lambda_{1,0}$ and $\lambda_{2,0}$ are then determined by consistency. The different constants are chosen to be $m = J = I = 1$ and $R = 1/4$. Note that this implies $M(q) = I_4$, the 4×4 identity matrix.

The index-1 DAE (2.11) is solved using `ode15s` and the system of ODEs (2.12) using `ode45`. We want to see how well the original constraints of the system are satisfied. Recall that the vertical rolling disk is subject to two nonholonomic constraints. If the radius of the disk is R, then these constraints are

$$\begin{aligned} 0 &= \dot{x} - R(\cos\varphi)\dot{\theta}, \\ 0 &= \dot{y} - R(\sin\varphi)\dot{\theta}. \end{aligned} \tag{2.13}$$

For both integrators `ode15s` and `ode45`, the tolerance values are set to `RelTol` $= 10^{-6}$ and `AbsTol` $= 10^{-9}$. Integrating from $t_0 = 0$ to $t_{\text{end}} = 100$, the value of the constraints, i.e. the values on the right hand side of (2.13), are plotted in Figure 2.1 on the next page. For both solutions, a clear drift in the value of the constraints is observed. The same behavior is seen for any tolerance. This illustrates one negative side of solving the systems with differentiated constraints instead of the original problem. Long time integration using this technique is not possible as the solution will drift off the correct manifold. The rest of this chapter is devoted to other approaches for solving the Lagrange-d'Alembert equations.

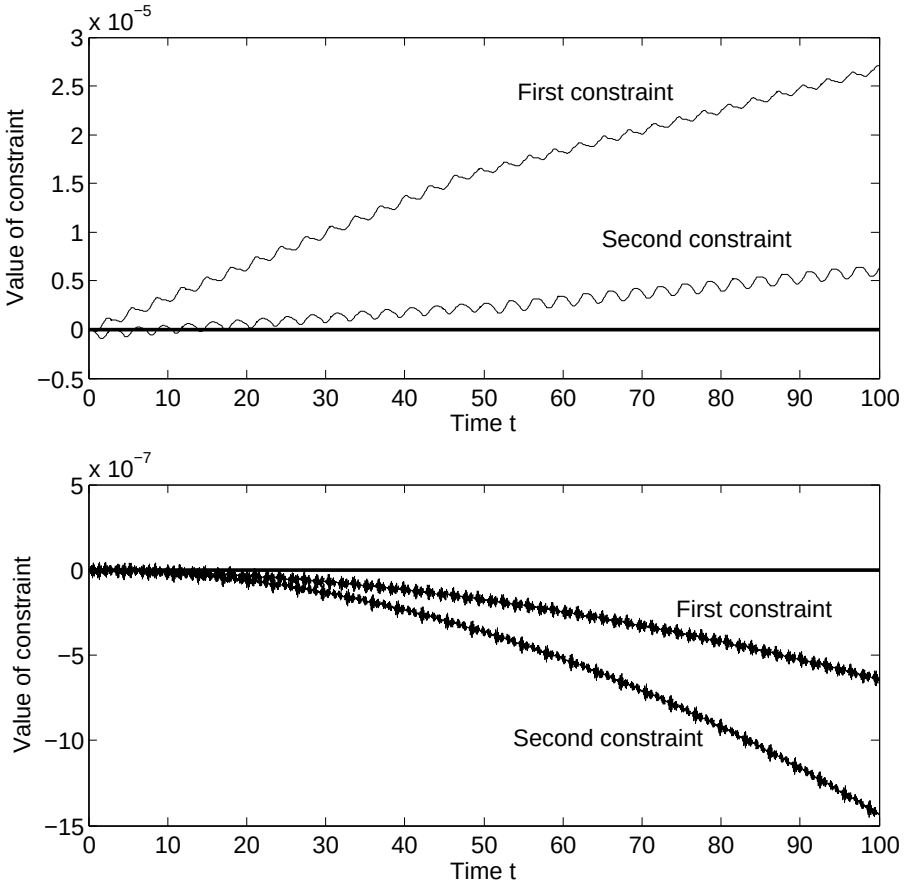

Figure 2.1: The error in the two constraint equations (2.13). (Top) Solving the index-1 DAE (2.11) using `ode15s`. (Bottom) Solving the system of ODEs (2.12) using `ode45`. The bold line in each plot is the exact value of zero. A clear drift in the constraints is observed for both solutions.

2.4 Runge-Kutta Methods Applied to Index 2 DAEs

Consider an s-stage Runge-Kutta method, defined by the Butcher-tableau

$$\begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array}$$

Define as is customary the column vector of weights $b \doteq (b_i)_{i=1,\ldots,s}$, the column vector of the nodes $c \doteq (c_i)_{i=1,\ldots,s}$ and the RK matrix of coefficients $A \doteq (a_{ij})_{i,j=1,\ldots,s}$.

Assuming that the initial values y_0 and z_0 are consistent with (2.2), i.e. they satisfy

$$g(t_0, y_0) = 0 \quad \text{and} \quad g_y(t_0, y_0) f(t_0, y_0, z_0) = 0, \tag{2.14}$$

the standard definition of an RK-method applied to (2.2) is as follows [9]. Starting from y_0 at t_0, the numerical solution y_1 at $t_1 = t_0 + h$ is given by

$$y_1 = y_0 + h\sum_{i=1}^{s} b_i f(T_i, Y_i, Z_i), \tag{2.15a}$$

where $T_i \doteq t_0 + c_i h$ and the s internal stages Y_i, Z_i, $i = 1, \ldots, s$, are the solution of the system of nonlinear equations

$$Y_i = y_0 + h\sum_{j=1}^{s} a_{ij} f(T_i, Y_i, Z_i), \quad i = 1, ..., s, \tag{2.15b}$$

$$0 = g(T_i, Y_i), \qquad i = 1, ..., s. \tag{2.15c}$$

The assumption that the initial conditions satisfy (2.14) ensures that the system (2.15) has a solution and that it is locally unique [9].

An important part of implementing these methods is the solution of the non-linear system of equations (2.15) arising at each time step. How this is done using the simplified Newton method is discussed in [9, Chapter 7]. We use the simplified Newton method when solving a very similar non-linear system of equations that arise in SPARK methods in Section 2.5.1.

Runge-Kutta methods that satisfy $b_i = a_{si}$ for $i = 1, \ldots, s$, are called *stiffly accurate* RK methods. Methods that do not satisfy this property are referred to as *nonstiffly accurate*. For stiffly accurate RK methods, $Y_s = y_1$, and so $g(t_1, y_1) = 0$ is automatically satisfied [9, 18]. This is a desirable property as this means the constraints are satisfied at the solution y_1 and not only at the internal stages.

Convergence rates for some classical RK-methods based on Gauss, Radau and Lobatto quadrature given[2] in [9, 12] are listed in Table 2.1. The stiffly accurate methods Radau IIA, Lobatto IIIA and Lobatto IIIC preserve their high order of convergence, whereas the nonstiffly accurate methods Gauss and Radau IA do not [18]. In the next section, one way to reestablish this so called superconvergence for nonstiffly accurate methods is considered.

Table 2.1: Order of convergence for the method (2.15) based on different RK-methods.

Method	Stages	Convergence
Gauss	s { odd / even	{ $s+1$ / s
Radau IA	s	s
Radau IIA	s	$2s-1$
Lobatto IIIA	s	$2s-2$
Lobatto IIIC	s	$2s-2$

[2]The convergence rates were proven in [9] with the exception of the result for Lobatto IIIA methods, which were only proven for $s = 2, 3$ and conjectured for larger s. This conjecture was then later proven in [12].

2.5 SPARK Methods for Index 2 DAEs

As noted in the last section, when using nonstiffly accurate RK methods (i.e. methods that do not satisfy $b_i = a_{si}$ for all i) to solve index 2 DAEs, the methods do in general not have a high order of convergence. The reason for this loss of superconvergence is thought to be the fact that nonstiffly accurate RK methods do not automatically satisfy $0 = g(t_1, y_1)$ [18]. One would ideally like both $0 = g(T_i, Y_i)$ for $i = 1, \ldots, s$, and $0 = g(t_1, y_1)$ to be satisfied also for nonstiffly accurate methods. This is however not possible as there are only sm algebraic variables Z_i for $(s+1)m$ constraints.

Some methods had been proposed to deal with the problem discussed above before the method considered next in this section was proposed (see [10, 11, 18] and references therein). One is projection methods, where the solution y_1 is projected onto the constraints in an additional step. Another is called partitioned RK methods, which require introduction of additional internal stages. In this section, another method is considered, first introduced in [14], which does not require any projection or the introduction of additional internal stages. This method is called *super partitioned additive Runge-Kutta (SPARK) methods* [14, 16, 17].

SPARK methods can be applied to a broad class of DAEs, but only its application to Hessenberg index 2 DAEs is discussed here, i.e. its application to the system

$$\dot{y} = f(t, y, z), \tag{2.16a}$$
$$0 = g(t, y). \tag{2.16b}$$

Consider a decomposition of the right hand side $f(t, y, z)$ of (2.16a) which reads

$$f(t, y, z) = \sum_{m=1}^{M} f_m(t, y, z). \tag{2.17}$$

The reason for this decomposition is that $f(t, y, z)$ may represent different forces in a mechanical system. The functions f_m are supposed to have distinct properties which may be beneficial to treat in different ways numerically. The number of different classes of right hand side terms is usually small, e.g. $M = 5$. It is assumed that f_1 is independent of z, i.e.

$$f_1(t, y, z) = f_1(t, y).$$

The main idea of the SPARK methods is then as follows. To ensure that the constraints are satisfied at the solution y_1, the equation $0 = g(t_1, y_1)$ is added. Then, in order to not have an overdetermined system, the equations $0 = g(T_i, Y_i)$ for $i = 1, \ldots, s$, are not used, but instead $s - 1$ well-chosen linear combinations of $g(T_i, Y_i)$ are enforced to be equal to zero. The difficulty then lies in finding the coefficients of these linear combinations such that the method has highest possible order.

Some definitions and assumptions need to be made before giving the actual definition of SPARK methods. In the following, the i'th column of the $s \times s$ identity matrix is denoted by e_i and the zero column vector of length s is denoted by 0_s. It is assumed that $s \geq 2$. As before, the RK weight vector is denoted by b and the RK node vector by c. The SPARK methods use the RK coefficient matrices of M distinct RK methods. These are

denoted by A_m for $m = 1, \ldots, M$. The following assumptions are made.

$$e_1^T A_1 = 0_s^T, \tag{2.18a}$$
$$e_s^T A_1 = b^T, \tag{2.18b}$$
$$A_1 A_m = \begin{bmatrix} 0_s^T \\ N \end{bmatrix} \quad \text{for } m = 2, \ldots, M, \tag{2.18c}$$
$$\begin{bmatrix} N \\ b^T \end{bmatrix} \text{ is invertible,} \tag{2.18d}$$
$$e_s^T A_3 = b^T. \tag{2.18e}$$

These assumptions are for example satisfied by the s-stage Lobatto SPARK families with $M = 5$ and A_1 through A_5 being the RK matrices of the Lobatto IIIA-B-C-C*-D methods, respectively [17]. Note that these methods are not the only choice for the SPARK coefficients. We will on the other hand only consider the coefficients of the Lobatto III family.

Now, let $\widetilde{A}_1$ be the $(s-1) \times s$ sub-matrix of A_1 given by the relation

$$A_1 = \begin{bmatrix} 0_s^T \\ \widetilde{A}_1 \end{bmatrix},$$

and let the $s \times (s+1)$ matrix Q be defined by

$$Q \doteq L \begin{bmatrix} \widetilde{A}_1 & 0_{s-1} \\ 0_s^T & 1 \end{bmatrix}, \tag{2.19}$$

where the $s \times s$ matrix L is any invertible matrix. One choice for L is given by

$$L = \begin{bmatrix} \widetilde{A}_1 \\ e_s^T \end{bmatrix}^{-1}.$$

With this, the definition of SPARK methods are given as stated in [17].

Definition 2.3 (SPARK method applied to index 2 DAEs)**.** One step of an s-stage *super partitioned additive Runge-Kutta (SPARK) method* applied with stepsize h to the system of index 2 DAEs (2.16) with decomposition (2.17) and initial values y_0, z_0 and t_0 reads

$$0 = Y_i - y_0 - h \sum_{j=1}^{s} \sum_{m=1}^{M} a_{ij,m} f_m(T_j, Y_j, Z_j), \qquad \text{for } i = 1, \ldots, s, \tag{2.20a}$$
$$0 = \sum_{j=1}^{s} q_{ij} g(T_j, Y_j) + q_{i,s+1} g(t_1, y_1), \qquad \text{for } i = 1, \ldots, s, \tag{2.20b}$$
$$0 = y_1 - y_0 - h \sum_{j=1}^{s} b_j f(T_j, Y_j, Z_j), \tag{2.20c}$$

where the coefficients q_{ij} are those of the matrix Q satisfying (2.19). This implies the equation $0 = g(t_1, y_1)$ is satisfied.

Existence and uniqueness for the SPARK methods are proven in [17]. We only state the result.

Theorem 2.4 (Existence and uniqueness)**.** *Assume that the initial conditions are consistent, i.e.* (y_0, z_0) *satisfy* (2.14) *and that* $g_y(y)f_z(y,z)$ *exists and is invertible. Assume also that the SPARK coefficients satisfy* (2.18) *and that* $\sum_{j=1}^{s} a_{ij,m} = c_i$ *for* $i = 1, \ldots, s$ *hold for* $m = 1$ *and* $m = 3$*. Then there exists for* $h \leq h_0$ *a locally unique solution for the system* (2.20)*.*

With some additional assumptions on the SPARK coefficients, the order of the local and global error are shown in [17]. What is of main interest to us, is what this implies for the Lobatto III SPARK methods. The following result is also proven in [17].

Theorem 2.5 (Global error and symmetry)**.** *The global error of the s-stage Lobatto IIIA-B-C-C*-D SPARK method satisfies*

$$y_k - y(t_k) = \mathcal{O}(h^{2s-2}).$$

Also, if $f_3 \equiv 0$ *and* $f_4 \equiv 0$*, then these methods are symmetric (reversible).*

The Butcher tableaux for the 2-stage and 3-stage Lobatto IIIA-B-C-C*-D methods are listed in Appendix B.

2.5.1 Nonlinear Equations and the Simplified Newton Method

A major difficulty in using SPARK methods lies in solving the system of non-linear equations (2.20). A classical algorithm for solving a set of nonlinear equations $F(X) = 0$ is the *simplified Newton method* [9, 15]. In the Newton method, the exact Jacobian $F'(X_k)$ is computed at each iteration. In the simplified Newton method on the other hand, the Jacobian $F'(X_0)$ computed at the initial value is reused. This method may require more iterations than the Newton method to converge, but the evaluation of the Jacobian only has to be done once, resulting in a lower computational cost. Given an initial guess X_0, sufficiently close to the locally unique zero X^*, the simplified Newton method is given by Algorithm 2.1 below.

Algorithm 2.1 Simplified Newton. Solves $F(X) = 0$.

Input: Jacobian $F'(X_0)$, initial guess X_0
Output: X, an approximation to X^* where $F(X^*) = 0$

1: Set $k = 0$
2: **while** not convergent **do**
3: Solve $F'(X_0)\Delta X_k = -F(X_k)$ for ΔX_k
4: Set $X_{k+1} = X_k + \Delta X_k$
5: Set $k = k + 1$
6: **end while**
7: $X = X_k$

A stopping criterion for the simplified Newton method often used in practice is given in [15]. Let

$$\theta_k \doteq \frac{||\Delta X_k||}{||\Delta X_{k-1}||}$$

for $k \geq 1$ and let $\widehat{\theta}_1 \doteq \theta_1$ and $\widehat{\theta}_k \doteq \sqrt{\widehat{\theta}_{k-1}\theta_k}$ for $k \geq 2$. Then this stopping criterion is given by

$$\eta_k ||\Delta X_k|| \leq \kappa_1 TOL, \quad \text{where} \quad \eta_k \doteq \frac{\widehat{\theta}_k}{1-\widehat{\theta}_k}.$$

Here, TOL is an error tolerance and κ_1 is a security factor such as $\kappa_1 = 0.03$.

How to apply the simplified Newton method to the set of non-linear SPARK equations (2.20) is now considered. Define the column vector X of length $d = s(n+m)+n$ as

$$X = \begin{bmatrix} Y_1^T & \cdots & Y_s^T & Z_1^T & \cdots & Z_s^T & y_1^T \end{bmatrix}^T. \tag{2.21}$$

Let $F\colon \mathbb{R}^d \to \mathbb{R}^d$ be a map such that $F(X)$ is the right hand side of (2.20). Also, let $\widehat{Q} \in \mathbb{R}^{s\times s}$ and $q_{s+1} \in \mathbb{R}^s$ be such that

$$Q = \begin{bmatrix} \widehat{Q} & q_{s+1} \end{bmatrix}.$$

For the Jacobian needed in the simplified Newton method, the Jacobian $F'(X)$ evaluated at $Y_i = y_0$, $Z_i = z_0$ for $i = 1, \ldots, s$, and $y_1 = y_0$ is used. Denote this Jacobian by J. It is given by

$$J = \begin{bmatrix} I_{sn} - h\sum_{m=1}^{M} A_m \otimes f_{my} & -h\sum_{m=1}^{M} A_m \otimes f_{mz} & 0 \\ \widehat{Q} \otimes g_y & 0 & q_{s+1} \otimes g_y \\ -hb^T \otimes f_y & -hb^T \otimes f_z & I_n \end{bmatrix},$$

where the notation $f_{my} = \frac{\partial}{\partial y} f_m$ and similarly $f_{mz} = \frac{\partial}{\partial z} f_m$ is used. I_n denotes the $n \times n$ identity matrix. All derivatives appearing in J are evaluated at (t_0, y_0, z_0). The sign $\otimes$ denotes the Kronecker product.

The following initial guess is used for the simplified Newton iterations

$$\begin{aligned} Y_i &= y_0 + c_i h f(y_0, z_0), && \text{for } i = 1, \ldots, s, \\ Z_i &= z_0, && \text{for } i = 1, \ldots, s, \\ y_1 &= y_0 + h f(y_0, z_0). \end{aligned} \tag{2.22}$$

This is the initial guess given in [9] for Runge-Kutta methods applied to DAEs. It is also stated in [9] that if one uses the natural initial guess $Y_i = y_0$, $Z_i = z_0$ for $i = 1, \ldots, s$, and $y_1 = y_0$, then the simplified Newton iterations may diverge. We have experienced that this may happen for the SPARK equations (2.20) as well, so using the starting values (2.22) is found to be crucial.

The derivatives f_{my} for $m = 1, \ldots, M$ and g_y in J must either be computed exactly beforehand, or they need to be estimated. One can estimate derivatives using a finite difference approximation (see e.g. [23, Chapter 8.2.2] and [8]). In MATLAB, the function `numjac` uses finite differences to compute approximate Jacobians.

Further details about how to solve the nonlinear SPARK equations can be found in [15, 16], where the use of a preconditioner is also discussed. We have chosen not to look into the use of a preconditioner as our focus first of all is on the qualitative properties of the integrators. For a more efficient implementation, the use of a precoditioner should be considered.

2.5.2 Numerical Confirmation of the Order of Convergence

To confirm the convergence rates given by Theorem 2.5 numerically, a test problem given in [17] is solved using our implementation of the s-stage Lobatto IIIA-B-C-C*-D SPARK methods. The test problem is given by

$$\frac{d}{dt}y = \frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = f_1(y_1, y_2) + \sum_{m=2}^{5} f_m(y_1, y_2, z_1), \quad 0 = g(y_1, y_2), \tag{2.23}$$

where

$$f_1(y_1, y_2) = \begin{bmatrix} y_2 - 2y_1^2 y_2 \\ -y_1^2 \end{bmatrix}, \qquad f_2(y_1, y_2, z_1) = \begin{bmatrix} y_1 y_2^2 z_1^2 \\ e^{-t} z_1 - y_1 \end{bmatrix},$$
$$f_3(y_1, y_2, z_1) = \begin{bmatrix} -y_2^2 z_1 \\ -3y_2^2 z_1 \end{bmatrix}, \qquad f_4(y_1, y_2, z_1) = \begin{bmatrix} 2y_1 y_2^2 - 2e^{-2t} y_1 y_2 \\ z_1 \end{bmatrix},$$
$$f_5(y_1, y_2, z_1) = \begin{bmatrix} 2y_2^2 z_1^2 \\ y_1^2 y_2^2 \end{bmatrix}, \qquad g(y_1, y_2) = y_1^2 y_2 - 1.$$

The initial conditions are given by $t_0 = 0$, $y_1(0) = 1$ and $y_2(0) = 1$. With these initial conditions, the exact solution of the system is given by $y_1(t) = e^t$, $y_2(t) = e^{-2t}$ and $z_1(t) = e^{2t}$. Let $E_1(h)$ be the global error at $t = 1$ using stepsize h, i.e.

$$E_1(h) \doteq \|y_K - y(1)\|_2,$$

where we by y_K mean the numerical solution after $K = 1/h$ steps. If the error $E_1(h)$ is of order p, then $E_1(h) = Ch^p + \mathcal{O}(h^{p+1})$. For $h \ll 1$, we can therefore assume $E_1(h) = Ch^p$. Taking the logarithm on both sides yields $\log(E_1) = p\log(h) + \log C$, so plotting $\log(E_1)$ against $\log(h)$ should give a straight line with slope p. This is done in Figure 2.2 on the next page for different values of the stepsize h and for both $s = 2$ and $s = 3$. Theorem 2.5 predicts that the 2-stage method should have a convergence rate of 2 and the 3-stage a convergence rate of 4. Using least squares to fit a straight line to the data, we find that $p = 1.97$ for $s = 2$ and $p = 4.12$ for $s = 3$. This confirms the predicted values of Theorem 2.5.

2.5.3 SPARK Methods Used on the Lagrange-d'Alembert Equations

The SPARK Lobatto IIIA-B-C-C*-D methods for general Hessenberg index-2 DAEs are presented and discussed in Section 2.5. For these methods, a decomposition of the right hand side

$$f(t, y, z) = \sum_{m=1}^{5} f_m(t, y, z)$$

is assumed, but it is not yet discussed how such a decomposition should be made when solving the Lagrange-d'Alembert equations.

We will use a decomposition suggested for holonomic systems in [13]. There, a symplectic partitioned Runge-Kutta method is given that uses a Lobatto IIIA-B pair to solve the equations of motion written in Hamiltonian form. The Lobatto IIIA coefficients are used on the equations for q and the Lobatto IIIB coefficients on the equations for p. We hope that since this decomposition has good properties in the holonomic case, it will do well also for nonholonomic systems.

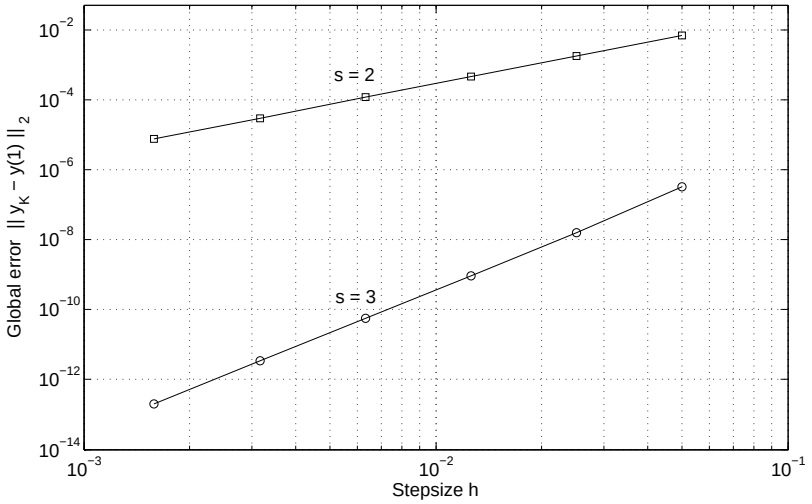

Figure 2.2: Global error of the s-stage Lobatto IIIA-B-C-C*-D SPARK methods ($s = 2, 3$) applied to the test problem (2.23).

Assume the Lagrangian is of the form $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$, where $M(q)$ is positive definite. Recall from Section 1.3.3 that the Hamiltonian formulation of the Lagrange-d'Alembert equations in the nonholonomic case then is given by

$$\dot{q} = \frac{\partial H}{\partial p} = M^{-1}p, \tag{2.24a}$$

$$\dot{p} = -\frac{\partial H}{\partial q} + A^T\lambda = \frac{\partial L}{\partial q} + A^T\lambda. \tag{2.24b}$$

Motivated by the use of a Lobatto IIIA-B pair for holonomic systems in [13], we will use Lobatto IIIA coefficients on (2.24a) and Lobatto IIIB coefficients on (2.24b). This means that the decomposition of $f(t, y, z)$ we will use for the SPARK Lobatto IIIA-B-C-C*-D methods is given by

$$f_1(t, y) = \begin{bmatrix} M^{-1}p \\ 0 \end{bmatrix}, \quad f_2(t, y, z) = \begin{bmatrix} 0 \\ L_q + A^T z \end{bmatrix}, \quad f_3 \equiv f_4 \equiv f_5 \equiv 0. \tag{2.25}$$

Note that f_1 is independent of z, which is assumed to be the case for SPARK methods. Also note that since $f_3 \equiv f_4 \equiv 0$, Theorem 2.5 states that these methods are symmetric.

When we later use the s-stage SPARK Lobatto IIIA-B-C-C*-D method with decomposition (2.25) to solve the Lagrange-d'Alembert equations, it will simply be referred to as the *s-stage SPARK Lobatto IIIA-B method*.

2.6 Specialized Runge-Kutta Methods for Index 2 DAEs

A method very similar to the SPARK methods is proposed in [18]. These methods are called *specialized Runge-Kutta methods for index 2 differential-algebraic equations* or *SRK-DAE2* for short. The main idea of these methods is the same as for the SPARK methods. The difference lies for the most part in that the SRK-DAE2 methods do not consider

a decomposition of the right hand side $f(y, z)$. For certain cases though, the SPARK methods and SRK-DAE2 methods can be shown to be equivalent [18].

The definition of SRK-DAE2 methods as stated in [18] is now given. For these methods, the RK coefficient matrix A is in assumed to be invertible.

Definition 2.6 (SRK-DAE2 methods applied to index 2 DAEs)**.** Assuming the initial values y_0 and z_0 satisfy (2.14), one step of an s-stage specialized Runge-Kutta method applied to (2.2) (also known as an SRK-DAE2 method) with stepsize h is given by y_1 where y_1 and the s internal stages Y_i, Z_i for $i = 1, ..., s$ are the solution of the system of nonlinear equations

$$0 = Y_i - y_0 - h\sum_{j=1}^{s} a_{ij} f(Y_j, Z_j), \quad i = 1, ..., s, \tag{2.26a}$$

$$0 = \sum_{j=1}^{s} \omega_{ij} g(Y_j) + \omega_{i,s+1} g(y_1), \qquad i = 1, ..., s, \tag{2.26b}$$

$$0 = y_1 - y_0 - h\sum_{j=1}^{s} b_j f(Y_j, Z_j). \tag{2.26c}$$

Let the matrix $\Omega \in \mathbb{R}^{s\times(s+1)}$ be such that (i, j)-entry of Ω is the coefficient ω_{ij} in (2.26b). Let $0_s^T = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^s$ be the zero vector of length s and $C = \operatorname{diag}(c_1, \ldots, c_s) \in \mathbb{R}^{s\times s}$ be a diagonal matrix with the nodes of the RK method on its diagonal. One choice of the matrix of weight Ω is

$$\widetilde{\Omega}_0 \doteq \begin{bmatrix} 0_s^T & 1 \\ b^T & 0 \\ b^T C & 0 \\ \vdots & \vdots \\ b^T C^{s-2} & 0 \end{bmatrix}. \tag{2.27}$$

Existence, uniqueness and convergence rates for SRK-DAE2 methods are proven in [18], given certain conditions on the chosen RK method. These results are not given in general here, only what they imply for the Gauss and Radau IA methods. This is summarized in the following theorem.

Theorem 2.7. *Assume that the initial conditions are consistent, i.e. they satisfy* (2.14)*, and that $g_y(y)f_z(y, z)$ exists and is invertible. Then, for a Gauss or Radau IA SRK-DAE2 method there exists for $h \leq h_0$ a locally unique solution to the equations* (2.26)*. The global error of the s-stage Gauss SRK-DAE2 method with matrix Ω as in* (2.27) *satisfies*

$$y_n - y(t_n) = \mathcal{O}(h^{2s}).$$

and the global error of the s-stage Radau IA SRK-DAE2 method with matrix Ω as in (2.27) *satisfies*

$$y_n - y(t_n) = \mathcal{O}(h^{2s-1}).$$

Also, the Gauss SRK-DAE2 methods are symmetric.

From this theorem, it is seen that using the SRK-DAE2 method, the nonstiffly Gauss and Radau IA methods regain their natural high order of convergence. Recall from Section 2.4 that these methods do not preserve their natural high order of convergence when used

in the method (2.15). Also, note that the convergence rate of Radau IA SRK-DAE2 methods is one order higher than for projected Radau IA methods [18].

When it comes to solving the set of nonlinear equations (2.26), we use the simplified Newton method in the same way as for the SPARK methods (see Section 2.5.1). We solve $F(X) = 0$, where X is as before given by (2.21) and the function F now is given by the right hand side of (2.26). Let $\widehat{\Omega} \in \mathbb{R}^{s\times s}$ and $q_{s+1} \in \mathbb{R}^s$ be such that

$$\Omega = \begin{bmatrix} \widehat{\Omega} & \omega_{s+1} \end{bmatrix}.$$

The Jacobian needed in the simplified Newton method is then given by

$$J = \begin{bmatrix} I_{sn} - hA \otimes f_y & -hA \otimes f_z & 0 \\ \widehat{\Omega} \otimes g_y & 0 & \omega_{s+1} \otimes g_y \\ -hb^T \otimes f_y & -hb^T \otimes f_z & I_n \end{bmatrix}.$$

The Butcher tableaux for the 2-stage and 3-stage Gauss and Radau IA methods are listed in Appendix B.

2.6.1 Numerical Confirmation of the Order of Convergence

As for the Lobatto IIIA-B-C-C*-D SPARK methods, we now want to check numerically the convergence rates given in Theorem 2.7 for both Gauss and Radau IA SRK-DAE2 methods. Consider the following test problem given in [18]

$$\frac{d}{dt}y = \frac{d}{dt}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_1 y_2^2 z_1^2 \\ y_1^2 y_2^2 - 3y_2^2 z_1 \end{bmatrix}, \quad 0 = y_1^2 y_2 - 1, \tag{2.28}$$

with initial conditions $y_1(0) = 1$ and $y_2(0) = 1$ at $t_0 = 0$. The exact solution of this problem is given by $y_1(t) = e^t$, $y_1(t) = e^{-2t}$ and $z_1(t) = e^{2t}$.

For the 2-stage and 3-stage Gauss methods, the global error at $t = 1$ is plotted on a logarithmic scale in Figure 2.3 on the facing page for different values of the stepsize h. Using least squares to fit a straight line to these points, it is found that the 2-stage method has a slope of 4.06 and the 3-stage method a slope of 6.07. This confirms the convergence rates predicted by Theorem 2.7, which were 4 and 6, respectively.

The global error at $t = 1$ for the 2-stage and 3-stage Radau IA methods is plotted on a logarithmic scale in Figure 2.4 on the next page. The convergence rates predicted by Theorem 2.7 were 3 and 5 for the 2-stage and 3-stage method, respectively. Using least squares, we find that the 2-stage method has a slope of 2.97 and the 3-stage method a slope of 5.09. Thus, also in this case the predicted values are confirmed by the numerical example.

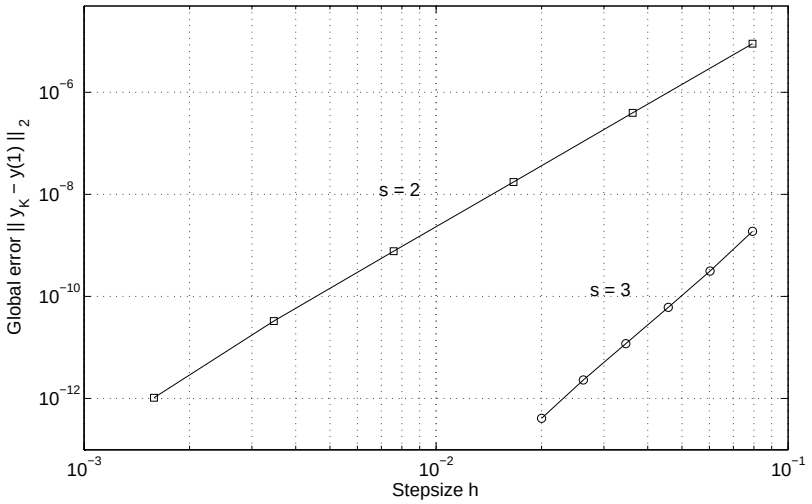

Figure 2.3: Global error of the 2-stage and the 3-stage Gauss SRK-DAE2 methods applied to the test problem (2.28).

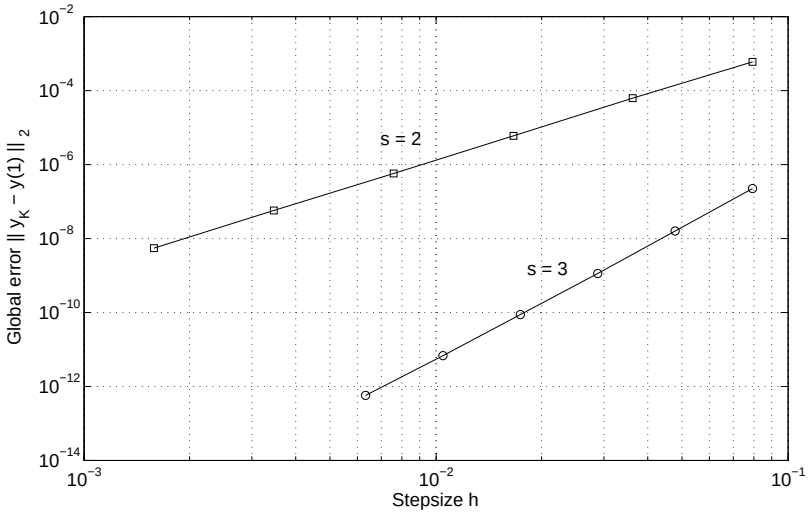

Figure 2.4: Global error of the 2-stage and the 3-stage Radau IA SRK-DAE2 methods applied to the test problem (2.28).

Chapter 3

Nonholonomic Integrators

In the previous chapter, it is seen how the Lagrange-d'Alembert equations can be discretized and solved using the DAE formulation of the equations. In this chapter, a different approach is considered. Instead of discretizing the equations of motion, a discrete version of the Lagrange-d'Alembert principle is used. This discrete principle leads to the discrete Lagrange-d'Alembert equations using a derivation analagous to the derivation used in the continuous case. Before considering nonholonomic systems however, the unconstrained case is considered and the discrete Euler-Lagrange equations are derived.

3.1 Discrete Euler-Lagrange Equations

In the continuous case, given a Lagrangian $L(q, \dot{q})$, the action integral is defined as

$$S(q(t)) = \int_0^T L(q(t), \dot{q}(t))\, dt.$$

Hamilton's principle then states that the curve $q(t)$ that extremizes this action integral is the curve that describes the correct physical evolution of the system. Then, by solving $\delta S(q(t)) = 0$, the Euler-Lagrange equations can be derived (see Section 1.2). Now, in the discrete case, a discrete analogue to this derivation is performed, and the *discrete* Euler-Lagrange equations are derived. Ideas from [21, 27] are mainly used in the following.

A continuous curve $q(t)$ between two fixed endpoints is replaced by a discrete curve $[q]$, consisting of points $q_0, \ldots, q_K$, where the endpoints q_0 and q_K are fixed. Such a curve is illustrated in Figure 3.1. Fix a stepsize h and think of two adjacent points q_k and q_{k+1} as being h apart in time. Each point q_k is in this way thought of as the state of the discrete system at time kh and so q_k can be viewed as an approximation to $q(kh)$.

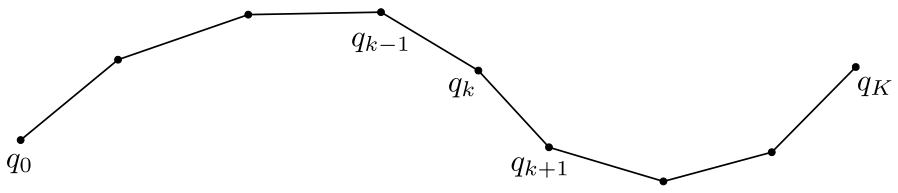

Figure 3.1: Illustration of a discrete curve $[q]$ with fixed endpoints q_0 and q_K.

Introduce a discrete Lagrangian $L_d(q_k, q_{k+1}, h)$, which depend on the stepsize h. This discrete Lagrangian is thought of as an approximation of the action integral along the

curve segment between q_k and q_{k+1}. The action integral $S(q(t))$ of the continuous curve is then replaced by the *action sum* $S_d([q])$ of the discrete curve, defined by

$$S_d([q]) = \sum_{k=0}^{K-1} L_d(q_k, q_{k+1}, h).$$

Analagous to the continuous case, consider the variations of this action sum with the endpoints fixed. Such discrete variations are illustrated in Figure 3.2.

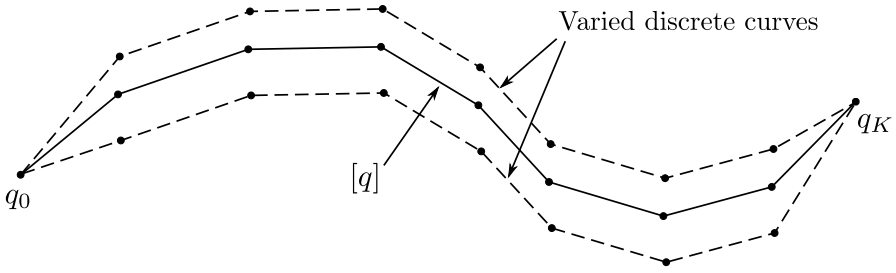

Figure 3.2: Illustration of variations of a discrete curve.

Among all discrete variations, the curve $[q]$ that extremizes the action sum is of interest. In the following, the shorter notation $L_d(q_k, q_{k+1})$ is used instead of $L_d(q_k, q_{k+1}, h)$, and it is understood that L_d still depends on h. Take the variation of $S_d([q])$ and set it equal to zero. This reads

$$\delta S_d([q]) = \delta \sum_{k=0}^{K-1} L_d(q_k, q_{k+1}) = \sum_{k=0}^{K-1} \Big(D_1 L_d(q_k, q_{k+1}) \cdot \delta q_k + D_2 L_d(q_k, q_{k+1}) \cdot \delta q_{k+1} \Big)$$

$$= D_1 L_d(q_0, q_1) \cdot \delta q_0 + \sum_{k=1}^{K-1} \Big(D_2 L_d(q_{k-1}, q_k) \cdot \delta q_k + D_1 L_d(q_k, q_{k+1}) \cdot \delta q_k \Big) + D_2 L_d(q_{K-1}, q_K) \cdot \delta q_K$$

$$= \sum_{k=1}^{K-1} \Big(D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) \Big) \cdot \delta q_k = 0,$$

for all variations δq_k. It is used that $\delta q_0 = \delta q_N = 0$, since there is no variation at the endpoints as they are fixed. $D_1 L_d(q_k, q_{k+1})$ means differentiation of $L_d(q_k, q_{k+1})$ with respect to the first argument q_k and D_2 is similarly differentiation with respect to the second argument. Since $\delta S_d([q]) = 0$ holds for all variations δq_k, the following set of equations is obtained.

Definition 3.1 (Discrete Euler-Lagrange equations)**.** The *discrete Euler-Lagrange (DEL) equations* are

$$D_2 L_d(q_{k-1}, q_k) + D_1 L_d(q_k, q_{k+1}) = 0, \tag{3.1}$$

for $k = 1, \ldots, K-1$.

These equations define a *two-step method*, which means that the solution at the two previous points in time are needed in order to compute the current solution.

In a similar fashion as flow is defined for continuous systems (see Section 1.3.4), the discrete flow is now defined as follows.

Definition 3.2 (Discrete flow)**.** The the *discrete flow* (or the *discrete flow map*) of a two-step method with stepsize h is the map F such that $F(h)\colon (q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$.

The discrete flow will depend on the choice of L_d and to emphasize this dependence, the discrete flow is sometimes denoted $F_{L_d}(h)$.

With some restrictions on the discrete Lagrangian L_d, the discrete flow $F_{L_d}(h)$ of the discrete Euler-Lagrange equations (3.1) is a well defined map (see [5, Chapter 7.2] for details). Thus, given an initial condition (q_0, q_1), one can use the DEL equations to compute the sequence $\{q_k\}_{k=0}^{K}$ in a recursive way.

Variational Integrators. The discrete Euler-Lagrange equations are also referred to as a *variational integrator*. With different choices of the discrete Lagrangian L_d, different integrators are obtained. We will not dwell on such methods for unconstrained systems as we are more interested in (nonholonomically) constrained systems. In the next section, variational integrators in the constrained case are derived. When dealing with nonholonomic systems, variational integrators are referred to as nonholonomic integrators. For more on variational integrators in the unconstrained case, see [10, VI.6].

3.2 Discrete Lagrange-d'Alembert Equations

Nonholonomic systems are now discussed. As a discrete analogue to the derivation of the Lagrange-d'Alembert equations, a *discrete* version of the Lagrange-d'Alembert principle is used to derive the *discrete* Lagrange-d'Alembert equations. This approach is used also for holonomic systems (see e.g. [21, 27]), but we will focus here on the nonholonomic case [6, 22].

As in the case of unconstrained systems, consider a discrete curve $[q]$ consisting of points $q_0, \dots, q_K$, where the endpoints are fixed. Given a discrete Lagrangian $L_d(q_k, q_{k+1}, h)$, the action sum is given by

$$S_d([q]) = \sum_{k=0}^{K-1} L_d(q_k, q_{k+1}, h).$$

Of interest is the curve that extremizes this action sum. In the continuous case, the variations δq have to satisfy the constraints for all times, i.e. $A(q(t))\delta q(t) = 0$ for all $t \in [0, T]$. In an analogous way, the variations δq_k now have to satisfy the constraints for all discrete times, so that $A(q_k)\delta q_k = 0$ for $k = 0, \dots, K$.

Also, the continuous curve $q(t)$ has to satisfy the constraint equations itself, such that $g(q, \dot{q}) \doteq A(q)\dot{q} = 0$ for all $t \in [0, T]$. As the curve $[q]$ is now considered, which is not continuous, it is not possible to differentiate it and hence not directly imposed these constraints. Instead, some *discrete constraints* g_d are found such that (q_k, q_{k+1}) satisfy these discrete constraints if $g_d(q_k, q_{k+1}) = 0$.

The above procedure leads to the discrete Lagrange-d'Alembert equations [6], which can be seen as a generalization of the discrete Euler-Lagrange equations (3.1) to the nonholonomic case.

Definition 3.3 (Discrete Lagrange-d'Alembert equations)**.** Given a discrete Lagrangian $L_d(q_k, q_{k+1})$ and discrete constraints $g_d(q_k, q_{k+1})$, the *discrete Lagrange-d'Alembert (DLA)*

equations are given by

$$D_1 L_d(q_k, q_{k+1}) + D_2 L_d(q_{k-1}, q_k) = A(q_k)^T \lambda, \tag{3.2a}$$

$$g_d(q_k, q_{k+1}) = 0, \tag{3.2b}$$

for $k = 1, \ldots, K-1$. The discrete Lagrange-d'Alembert equations are also referred to as a *nonholonomic integrator*.

Like the discrete Euler-Lagrange equations, these equations define a two-step method, so a discrete flow $F_{L_d}(h)\colon (q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$ can be associated with (3.2). The following result is stated in [6] and gives conditions on L_d and g_d for this discrete flow to be well defined.

Theorem 3.4. *The discrete flow $F_{L_d}(h)\colon (q_{k-1}, q_k) \mapsto (q_k, q_{k+1})$ defined by the discrete Lagrange-d'Alembert equations* (3.2) *is well defined for sufficiently small timesteps h if the matrix*

$$\begin{bmatrix} D_1 D_2 L_d(q_k, q_k) & A(q_k)^T \\ D_2 g_d(q_k, q_k) & 0 \end{bmatrix}$$

is invertible for each $q_k \in Q$.

3.3 Nonholonomic Integrators

A nonholonomic integrator is another name for the discrete Lagrange-d'Alembert equations. Using different discrete Lagrangians L_d and discrete constraints g_d, different methods with different properties are obtained. One important property is the order the method.

Definition 3.5. A nonholonomic integrator has order r if $q_k = q(kh) + \mathcal{O}(h^{r+1})$, where $q(t)$ is the exact solution of the Lagrange-d'Alembert equations.

Recall that if the Lagrangian is given by $L = K - V$, where K is the kinetic energy and V is the potential energy, then the flow of the solution of a continuous nonholonomic system is reversible (see Section 1.3.4). This is a property which it is preferable that the discrete solution also possesses. In a similar way reversibility is defined in the continuous case, discrete reversibility is now defined as follows.

Definition 3.6 (Discrete reversibility)**.** Let R_d be such that $R_d(q_k, q_{k+1}) = (q_{k+1}, q_k)$ and let F_{L_d} be the discrete flow determined by the discrete Lagrange-d'Alembert equations. F_{L_d} is said to be *discrete (time) reversible* if

$$F_{L_d} \circ R_d \circ F_{L_d} = R_d.$$

This is a natural discrete analog of continuous reversibility, where R_d can be seen as a discrete version of ρ used in the continuous case. This definition is illustrated in Figure 3.3.

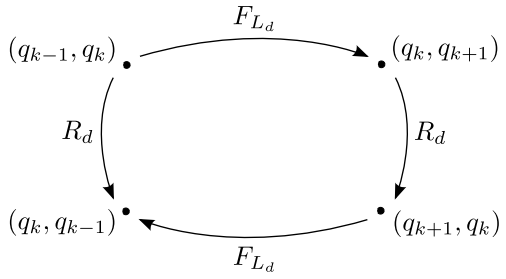

Figure 3.3: Illustration of a discrete reversible flow F_{L_d}.

Two second order nonholonomic integrators are now be presented. The first is given by Cortés and Martínez [6] and the other by McLachlan and Perlmutter [22].

3.3.1 Integrator by Cortés and Martínez

This is a second order nonholonomic integrator presented in [6]. It is given by (3.2) with the discrete Lagrangian

$$L_d(q_k, q_{k+1}) = L\left(\frac{q_k + q_{k+1}}{2}, \frac{q_{k+1} - q_k}{h}\right),$$

and the discrete constraints

$$g_d(q_k, q_{k+1}) = g\left(\frac{q_k + q_{k+1}}{2}, \frac{q_{k+1} - q_k}{h}\right).$$

In the case where the continuous Lagrangian is given by $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T\dot{q} - V(q)$, this method can be written as follows. Given values (q_{k-1}, q_k), let

$$v_k = \frac{q_k - q_{k-1}}{h} \quad \text{and} \quad q_{k-1/2} = q_{k-1} + \frac{1}{2}hv_k.$$

This method can then be written as

$$\begin{aligned} q_{k+1/2} &= q_k + \frac{1}{2}hv_{k+1}, \\ v_{k+1} &= v_k + h\left(-\frac{1}{2}\Big(V_q(q_{k-1/2}) + V_q(q_{k+1/2})\Big) + A(q_k)^T\lambda_k\right), \\ A(q_{k+1/2})v_{k+1/2} &= 0, \\ q_{k+1} &= q_k + hv_{k+1}. \end{aligned} \tag{3.3}$$

Note that the contstraints are satisfied at the intermediate step $q_{k+1/2}$, but not at the next solution q_{k+1}, which is a drawback of this method.

3.3.2 Integrator by McLachlan and Perlmutter

This is a second order reversible nonholonomic integrator presented in [22]. It is composition method, constructed by composing a specific first order method F with its adjoint F^*. To explain what this means, the definition of an adjoint method is given first [10, 27].

Definition 3.7 (Adjoint method)**.** For a method with discrete flow $F(h)$ the *adjoint method* is defined by the discrete flow $F^*(-h)$ where $F^*(h) \circ F(-h) = \mathrm{Id}$. Said in another way, $F^*(h) \circ F(-h) \colon (q_k, v_k) \mapsto (q_k, v_k)$.

Begin by considering the first order method defined by the discrete Lagrangian

$$L_d(q_k, q_{k+1}) = L\left(q_k, \frac{q_{k+1} - q_k}{h}\right),$$

and the discrete constraints

$$g_d(q_k, q_{k+1}) = g\left(q_k, \frac{q_{k+1} - q_k}{h}\right).$$

Assuming the Lagrangian is given by $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T\dot{q} - V(q)$, the Lagrange-d'Alembert equations (3.2) gives the first order nonholonomic integrator

$$\begin{aligned} q_{k+1} &= q_k + hv_k, \\ v_{k+1} &= v_k + h\Big(- V_q(q_{k+1}) + A(q_{k+1})^T\lambda_{k+1}\Big), \\ A(q_{k+1})v_{k+1} &= 0. \end{aligned}$$

Denote the discrete flow of this method by $F_{L_d}(h)$. Its adjoint method $F^*_{L_d}(-h)$ is given by

$$\begin{aligned} v_{k+1} &= v_k + h\Big(- V_q(q_k) + A(q_k)^T\lambda_k\Big), \\ q_{k+1} &= q_k + hv_{k+1}, \\ A(q_{k+1})v_{k+1} &= 0. \end{aligned}$$

The second order method which is to be derived is then found by the composition $F^*_{L_d}(-h/2) \circ F_{L_d}(h/2)$ of these two methods. In this composition, the two velocity updates are merged to obtain

$$\begin{aligned} q_{k+1/2} &= q_k + \frac{1}{2}hv_k, \\ A(q_{k+1/2})v_{k+1/2} &= 0, \\ v_{k+1} &= v_k + h\Big(- V_q(q_{k+1/2}) + A(q_{k+1/2})^T\Big(\frac{\lambda_1 + \lambda_2}{2}\Big)\Big), \\ q_{k+1} &= q_{k+1/2} + \frac{1}{2}hv_{k+1}, \\ A(q_{k+1})v_{k+1} &= 0. \end{aligned}$$

The intermediate velocity state $v_{k+1/2}$ is only involved in $A(q_{k+1/2})v_{k+1/2} = 0$, the first constraint. Therefore, this variable and constraint equations are both dropped, and the two Lagrangian multipliers λ_1 and λ_2 are merged. This leads to our second order method

$$\begin{aligned} q_{k+1/2} &= q_k + \frac{1}{2}hv_k, \\ v_{k+1} &= v_k + h\Big(- V_q(q_{k+1/2}) + A(q_{k+1/2})^T\lambda_k\Big), \\ q_{k+1} &= q_{k+1/2} + \frac{1}{2}hv_{k+1}, \\ A(q_{k+1})v_{k+1} &= 0. \end{aligned} \tag{3.4}$$

It is shown in [22] that this second order method also satisfies the discrete Lagrange-d'Alembert equations (3.2). If we use the discrete Lagrangian

$$\widehat{L}_d(q_k, q_{k+1}) = L\left(q_k, \frac{q_{k+1} - q_k}{h}\right), \tag{3.5}$$

and the discrete constraints

$$\widehat{g}_d(q_k, q_{k+1}) = g\left(\frac{q_k + q_{k+1}}{2}, \frac{q_{k+1} - q_k}{h}\right), \tag{3.6}$$

we obtain the method (3.4) using the DLA equations and a change of variables $\widehat{q}_k = (q_k + q_{k-1})/2$ and $\widehat{v}_k = (q_k - q_{k-1})/h$. This is easily verified.

3.3.3 The Integrator by McLachlan and Perlmutter as a SPARK Method

We have made the observation that mathematically, the second order nonholonomic integrator (3.4) by McLachlan and Perlmutter can be seen as a special case of a 2-stage Lobatto III SPARK method discussed in Section 2.5. This is interesting to observe as these two methods are derived using two very different approaches.

Consider a nonholonomic system with Lagrangian given by $L(q, \dot{q})$ and constraints $A(q)\dot{q} = 0$. We write the Lagrange-d'Alembert equations (1.12) as a Hessenberg index 2 DAE (2.16), where $y = \begin{bmatrix} q^T & \dot{q}^T \end{bmatrix}^T$. The right hand side $f(y, z)$ is decomposed as $f(y, z) = \sum_{m=1}^{5} f_m(y, z)$ where

$$f_1 = \begin{bmatrix} 0 \\ -V_q(q) + A(q)^T z \end{bmatrix}, \quad f_2 = \begin{bmatrix} v \\ 0 \end{bmatrix}, \quad f_3 \equiv f_4 \equiv f_5 \equiv 0, \tag{3.7}$$

and the constraint function is $g(y) = A(q)v$. We then apply the 2-stage Lobatto IIIA-B-C-C*-D SPARK method (2.20) using this decomposition. The SPARK equations (2.20) become

$$Q_1 = q_0 + \frac{1}{2}hV_1, \tag{3.8a}$$
$$V_1 = v_0, \tag{3.8b}$$
$$Q_2 = q_0 + \frac{1}{2}hV_1, \tag{3.8c}$$
$$V_2 = v_0 + \frac{1}{2}h\Big(\big(-V_q(Q_1) + A(Q_1)^T Z_1\big) + \big(-V_q(Q_2) + A(Q_2)^T Z_2\big)\Big), \tag{3.8d}$$
$$0 = \frac{1}{2}\Big(A(Q_1)V_1 + A(Q_2)V_2\Big), \tag{3.8e}$$
$$0 = A(q_1)v_1, \tag{3.8f}$$
$$q_1 = q_0 + \frac{1}{2}hV_1 + \frac{1}{2}hV_2, \tag{3.8g}$$
$$v_1 = v_0 + \frac{1}{2}h\Big(\big(-V_q(Q_1) + A(Q_1)^T Z_1\big) + \big(-V_q(Q_2) + A(Q_2)^T Z_2\big)\Big). \tag{3.8h}$$

Equations (3.8a), (3.8b) and (3.8c) give $Q_1 = Q_2 = q_0 + (h/2)v_0 \doteq q_{1/2}$ and equations (3.8d) and (3.8h) imply $V_2 = v_1$. Let $\lambda = (Z_1 + Z_2)/2$ and $v_{1/2} = (v_0 + v_1)/2$. Then (3.8)

can be rewritten as

$$\begin{aligned} q_{1/2} &= q_0 + \frac{1}{2}hv_0, \\ 0 &= A(q_{1/2})v_{1/2}, \\ 0 &= A(q_1)v_1, \\ q_1 &= q_{1/2} + \frac{1}{2}hv_1, \\ v_1 &= v_0 + h\Big(- V_q(q_{1/2}) + A(q_{1/2})^T\lambda\Big). \end{aligned}$$

Like in the derivation of the method by McLachlan and Perlmutter, the intermediate velocity state $v_{1/2}$ is now dropped as it only appears in the first constraint $0 = A(q_{1/2})v_{1/2}$. The resulting set of equations is then exactly the same as (3.4).

Note. In this case $f_1(y, z)$ is dependent of z, which is assumed *not* to be the case for the SPARK methods. This means that the results about existence, uniqueness and convergence rates given for the Lobatto III SPARK methods in Section 2.5 do not apply. We also observe in practice that our implementation of SPARK methods fails when using this decomposition of $f(y, z)$.

The SPARK Lobatto III methods are used on the Lagrange-d'Alembert equations in Section 2.5.3 on page 27. The decomposition used there is the same as (3.7), if only $f_1(y, z)$ and $f_2(y, z)$ are swapped.

That the nonholonomic integrator (3.4) by McLachlan and Perlmutter can be seen as a 2-stage SPARK Lobatto IIIB-A method is also noted in [19].

3.3.4 Numerical Confirmation of the Order of Convergence

The two numerical integrators (3.3) and (3.4) are both of order two. We would like to confirm this numerically. As a test problem, the equations of motion for the vertical rolling disk example studied in earlier chapters are used.

Recall from Section 1.3.5 that the coordinates for this problem is $q = \begin{bmatrix} x & y & \varphi & \theta \end{bmatrix}^T$ and the Lagrangian is given by $L(q, \dot{q}) = \dot{q}^T M(q)\dot{q}$, where and $M(q) = \operatorname{diag}(m, m, J, I)$. The two nonholonomic integrators (3.3) and (3.4) assume a Lagrangian of the form $L(q, \dot{q}) = \dot{q}^T\dot{q}$, so we therefore choose $m = J = I = 1$. The constraint matrix for the vertical rolling disk is given by

$$A(q) = \begin{bmatrix} 1 & 0 & 0 & -R\cos\varphi \\ 0 & 1 & 0 & -R\sin\varphi \end{bmatrix},$$

and $V_q = \partial V/\partial q$ used in both the nonholonomic integrators is simply the zero vector, as there is no potential energy.

We choose the same initial conditions as we did in Section 2.3.1. That is,

$$\begin{aligned} q_0 &= \begin{bmatrix} x_0 & y_0 & \varphi_0 & \theta_0 \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T, \\ v_0 &= \begin{bmatrix} v_{x,0} & v_{y,0} & \omega & \Omega \end{bmatrix}^T = \begin{bmatrix} 1/4 & 0 & 2 & 1 \end{bmatrix}^T, \\ \lambda_0 &= \begin{bmatrix} \lambda_{1,0} & \lambda_{2,0} \end{bmatrix}^T = \begin{bmatrix} 0 & 1/2 \end{bmatrix}^T. \end{aligned}$$

These are consistent initial values. The nonlinear equations in each of the two methods are solved using the built in function `fsolve` in MATLAB.

The equations of motion for the vertical rolling disk (1.16) actually has an analytic solution. Assuming consistent initial values, it is shown in [1] that the analytic solution of the equations of motion is

$$q(t) = \begin{bmatrix} x(t) \\ y(t) \\ \varphi(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} (\Omega R/\omega)\Big(\sin(\omega t + \varphi_0) - \sin(\varphi_0)\Big) + x_0 \\ -(\Omega R/\omega)\Big(\cos(\omega t + \varphi_0) - \cos(\varphi_0)\Big) + y_0 \\ \omega t + \varphi_0 \\ \Omega t + \theta_0 \end{bmatrix}.$$

Let $E_1(h)$ be the global error at $t = 1$ using stepsize h, i.e.

$$E_1(h) \doteq \|q_K - q(1)\|_2,$$

where we by q_K mean the numerical solution after $K = 1/h$ steps. In Figure 3.4, the global error $E_1(h)$ is plotted for different values of the stepsize h on a logarithmic scale for both methods (3.3) and (3.4). Using least squares to fit straight lines to the data, we find that the slope for (3.3) is 1.99 and the slope for (3.4) is 2.00. As explained in Section 2.5.2, this confirmes the predicted order of two for both methods.

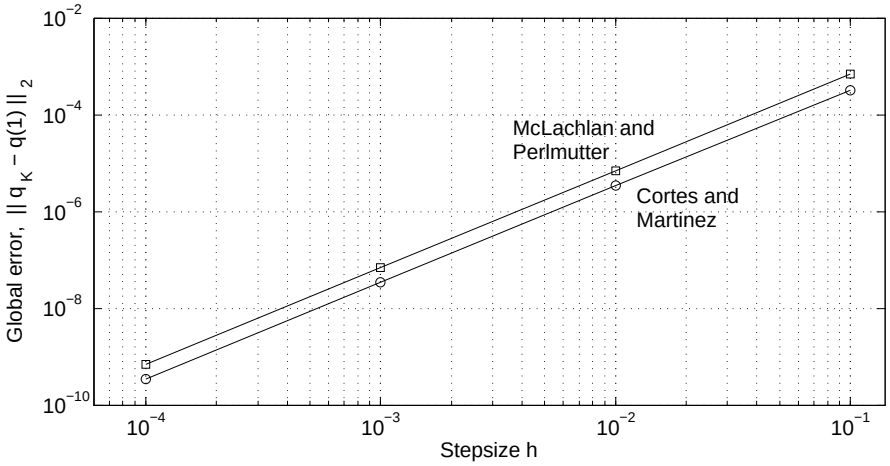

Figure 3.4: The global error at $t = 1$ for the vertical rolling disk using integrator (3.3) by Cortés and Martínez and integrator (3.4) by McLachlan and Perlmutter. For both methods, this plot confirmes a convergence rate of 2.

Chapter 4

Numerical Experiments

Note. Close to deadline for this thesis, we became aware of a recent work by Jay [19], where the 2-stage SPARK Lobatto IIIA-B method is compared to the nonholonomic integrator (3.4) by McLachlan and Perlmutter. The energy conservation for these two methods are seen to be equally good in a long-time integration done in [19]. In this chapter, we sometimes observe that the 2-stage SPARK Lobatto IIIA-B method has a drift in the energy. One explenation for the difference between our results and those in [19] might be that we do not solve the nonlinear SPARK equations accurately enough. As we did not become aware of this until at the very end, our results stand as they were. However, to argue that our implementation is correct, we compare our own plots to plots obtained from Laurent O. Jay through private communications. This is done in Section 4.5. We also discuss this matter further in the conclusion in Chapter 5.

4.1 The Vertical Rolling Disk

The vertical rolling disk example has followed us through this thesis. It is discussed in Section 1.1.2, 1.3.5, 2.3.1 and 3.3.4. Now, we wish to compare the error of the solution for the nonholonomic integrator (3.4) by McLachlan and Perlmutter with the error for the 2-stage SPARK Lobatto IIIA-B method. These methods are both reversible and of order 2. We reuse the vertical rolling disk example one last time because this problem has an analytical solution.

Recall that we are considering a vertical disk rolling on a plane without slipping and without falling over. The coordinates for the system are given by $q = \begin{bmatrix} x & y & \varphi & \theta \end{bmatrix}^T$. The Lagrangian and the constraints can be found in Section 1.3.5 on page 12. The same values for the constants and the same initial condition as in Section 3.3.4 are chosen. This implies that the Lagrangian is given simply by $L(q, \dot{q}) = \frac{1}{2}\dot{q}^T\dot{q}$.

Integrating over just a short period of time reveals a big difference in the two solutions. The error in all four components integrating from $t_0 = 0$ to $t_{\text{end}} = 50$ using a constant stepsize of $h = 0.1$ are shown for the nonholonomic integrator (3.4) in Figure 4.1 and for the SPARK Lobatto IIIA-B method in Figure 4.2. We observe that the SPARK Lobatto IIIA-B solution has a linear growth in the error of the θ-component. Even though a beginning linear growth is observed for the nonholonomic integrator as well, the slope is much less steep than for the SPARK Lobatto IIIA-B solution. This same behavior is seen using different stepsizes and also integrating much longer in time.

To investigate this linear growth further, we run a longer simulation from $t_0 = 0$ to

$t_{\text{end}} = 1000$ using both methods. The error of q in 2-norm for this simulation is shown in Figure 4.3 on the next page. The same plot using a logarithmic scale on the vertical axis is shown in Figure 4.4. The linear growth for the SPARK Lobatto IIIA-B method is clearly seen both using stepsize $h = 0.1$ and $h = 0.05$. There is no sign of any growth in the solution by the nonholonomic integrator. The error in energy for the same simulation is shown in Figure 4.5. The SPARK Lobatto IIIA-B solutions have a clear drift in the energy-error.

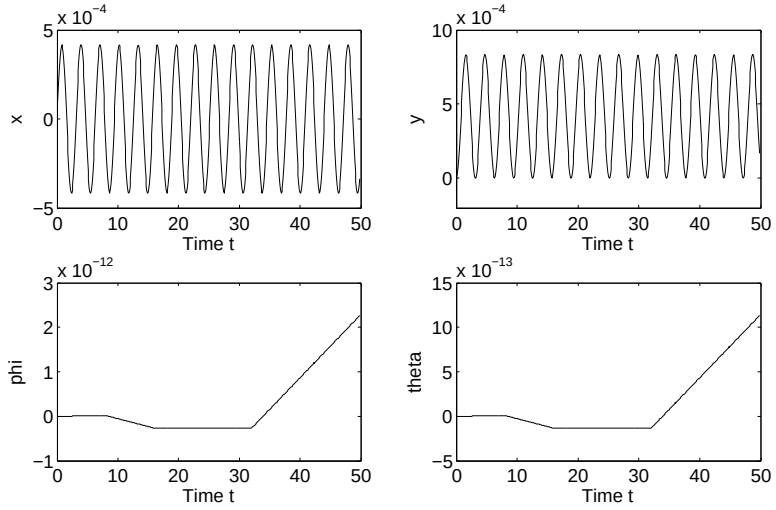

Figure 4.1: Error in each of the four components for the vertical rolling disk using the nonholonomic integrator (3.4) by McLachlan and Perlmutter with a constant stepsize $h = 0.1$.

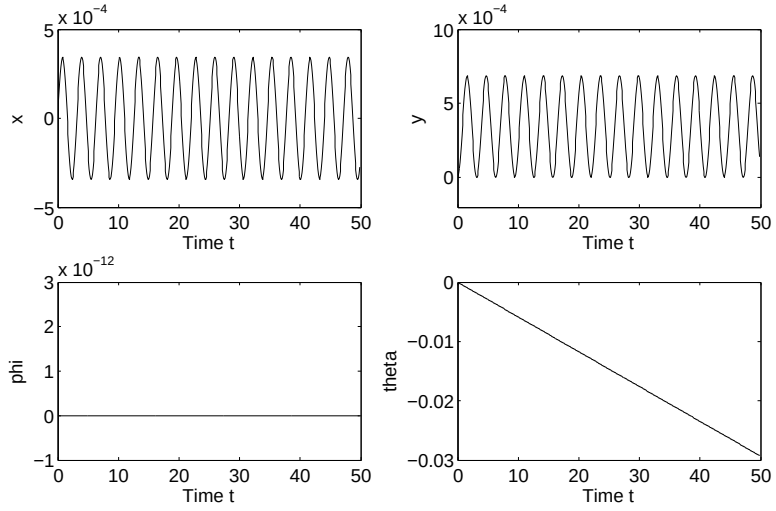

Figure 4.2: Error in each of the four components for the vertical rolling disk using the 2-stage SPARK Lobatto IIIA-B method with a constant stepsize $h = 0.1$. Note the linear growth in the θ-component. The error in φ is exactly zero.

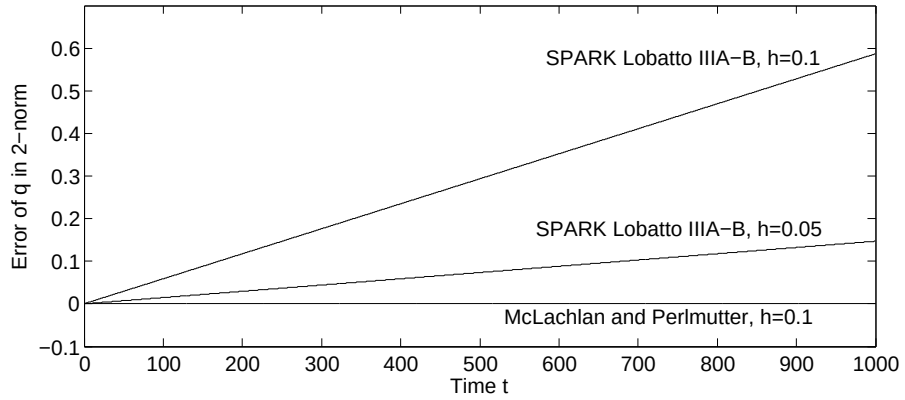

Figure 4.3: The error of q in 2-norm for the vertical rolling disk using the 2-stage SPARK Lobatto IIIA-B method and the nonholonomic integrator (3.4) by McLachlan and Perlmutter. A linear growth in the error for the SPARK Lobatto IIIA-B solutions is observed.

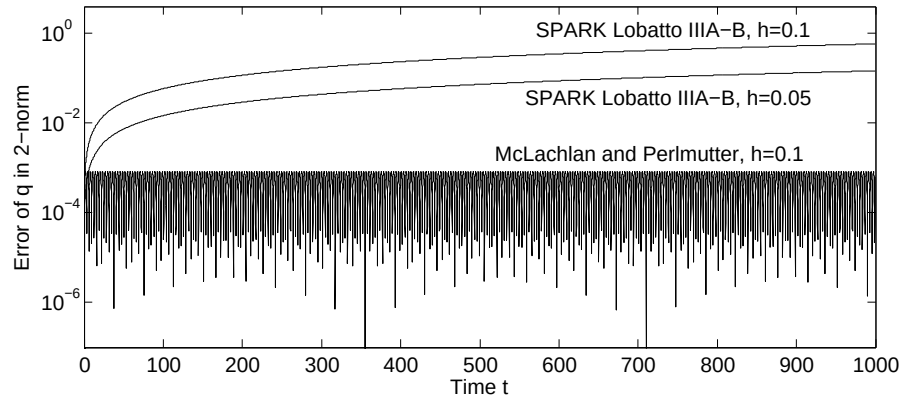

Figure 4.4: The same plot as in Figure 4.3 using a logarithmic scale on the vertical axis. The error for the nonholonomic integrator stays bounded below about 10^{-3}.

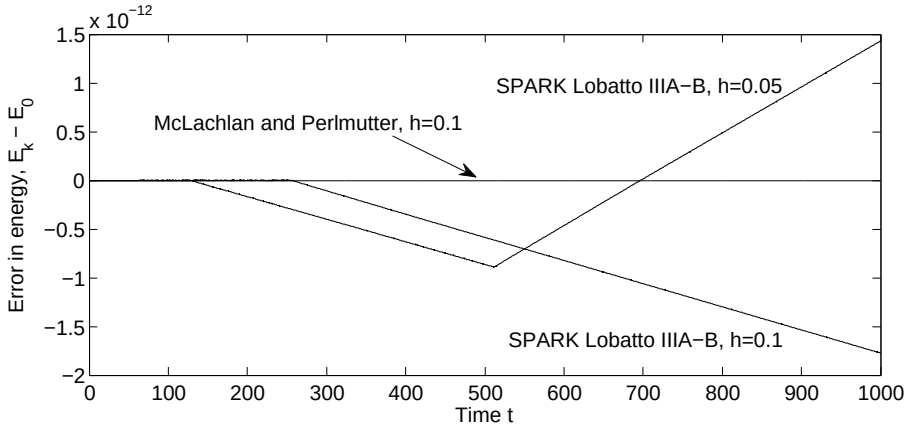

Figure 4.5: The error in energy for the same simulation as shown in Figure 4.3 and in Figure 4.4. The SPARK Lobatto IIIA-B solutions have a clear drift in the energy-error.

4.2 The Chaplygin Sleigh

In this experiment, we consider the SPARK and SRK methods, as well as MATLABs `ode15s`. We also generalize the nonholonomic integrator (3.4) by McLachlan and Perlmutter so it can be used in this case where the mass matrix is not equal to the identity matrix. The methods are compared, and we focus on how well these methods conserve energy, and also how well the nonholonomic constraint of this example is satisfied.

The Chaplygin sleigh is a nonholonomic system discussed, among other places, in [1, 7]. The sleigh is a rigid body supported by a knife edge and two other points as shown in Figure 4.6. The knife edge cannot move in the direction perpendicular to its edge, whereas the other two points of contact are free to move in any direction without friction. We consider the case where the sleigh is on an inclined plane. This is the same situation as described in [7].

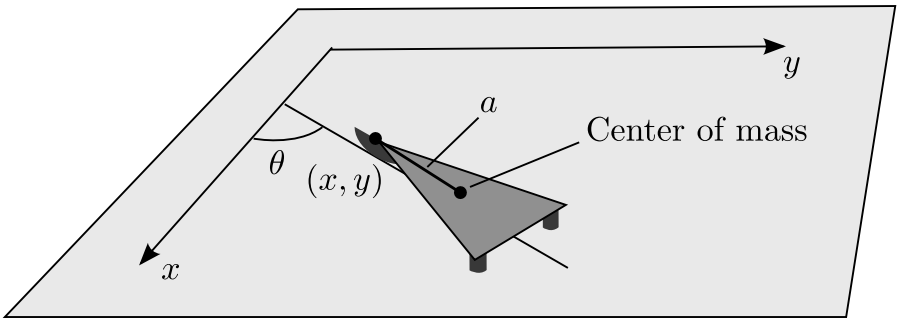

Figure 4.6: The Chaplygin sleigh with chosen coordinates.

The coordinates of the knife edge are used. The point (x, y) is the point of contact for the blade and θ is the angle of the blade relative to the x-axis. Let m be the mass, I the moment of inertia about the center of mass and let a be the distance from the blade's point of contact to the center of mass (see Figure 4.6). Let $q = \begin{bmatrix} x & y & \theta \end{bmatrix}^T$. The Lagrangian for this system is given by

$$L(q, \dot{q}) = K(q, \dot{q}) - V(q),$$

where the kinetic energy K and the potential energy V are given by

$$K(q, \dot{q}) = \frac{1}{2}\left(m\left(\dot{x}^2 + \dot{y}^2\right) + \left(I + ma^2\right)\dot{\theta}^2 + 2ma\dot{\theta}\left(\dot{y}\cos\theta - \dot{x}\sin\theta\right)\right)$$

and

$$V(q) = mg\left(y + a\sin\theta\right).$$

The potential energy is due to the inclination of the plane. From the restriction on the movement of the blade, we get the nonholonomic constraint

$$\dot{y}\cos\theta - \dot{x}\sin\theta = 0, \tag{4.1}$$

so the constraint matrix is $A(q) = \begin{bmatrix} -\sin\theta & \cos\theta & 0 \end{bmatrix}$. The Lagrangian can be written in the form

$$L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$$

where the mass matrix $M(q)$ is given by

$$M(q) = \begin{bmatrix} m & 0 & -ma\sin\theta \\ 0 & m & ma\cos\theta \\ -ma\sin\theta & ma\cos\theta & I + ma^2 \end{bmatrix}. \tag{4.2}$$

This mass matrix is symmetric and it has determinant $Im^2 \neq 0$, so it is invertible. In earlier discussions, the assumption that the mass matrix is positive definite is made several times. We show that this also is the case here, so that earlier results apply. A symmetric matrix is positive definite if all eigenvalues are positive [3]. As $M(q)$ is symmetric, it has only real eigenvalues. Using Descartes' Sign Rule [26], we find that all real eigenvalues must be positive. This implies that $M(q)$ is a symmetric positive definite (SPD) matrix. The inverse of $M(q)$ is given by

$$M(q)^{-1} = \frac{1}{I}\begin{bmatrix} I/m + a^2\sin^2\theta & -a^2\sin\theta\cos\theta & a\sin\theta \\ -a^2\sin\theta\cos\theta & I/m + a^2\cos^2\theta & -a\cos\theta \\ a\sin\theta & -a\cos\theta & 1 \end{bmatrix}.$$

4.2.1 The LDA Equations as an Index-1 DAE

To solve this problem using `ode15s`, the underlying index-1 DAE needs to be obtained. Let ξ be defined as

$$\xi = \begin{bmatrix} x & y & \theta & v_x & v_y & v_\theta & \lambda \end{bmatrix}^T.$$

Differentiating the constraint (4.1) once then yields the index-1 DAE

$$\underbrace{\begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & m & & & -ma\sin\theta \\ & & & & m & & ma\cos\theta \\ & & & -ma\sin\theta & ma\cos\theta & I+ma^2 & \\ & & & -\sin\theta & \cos\theta & & \end{bmatrix}}_{M_1(\xi)} \underbrace{\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{v}_x \\ \dot{v}_y \\ \dot{v}_\theta \\ \dot{\lambda} \end{bmatrix}}_{d\xi/dt} = \underbrace{\begin{bmatrix} v_x \\ v_y \\ v_\theta \\ mav_\theta^2\cos\theta - \lambda\sin\theta \\ -mg + mav_\theta^2\sin\theta + \lambda\cos\theta \\ -mga\cos\theta \\ v_x v_\theta\cos\theta + v_y v_\theta\sin\theta \end{bmatrix}}_{f_1(\xi)}.$$

Only non-zero elements of $M_1(\xi)$ are shown. Note that the last column of $M_1(\xi)$ is only zeros.

4.2.2 Using the Nonholonomic Integrator by McLachlan and Perlmutter

The nonholonomic integrator (3.4) by McLachlan and Perlmutter is only presented in [22] for systems where $L = \frac{1}{2}\dot{q}^T\dot{q} - V(q)$, i.e. systems where the mass matrix $M(q)$ equals the identity matrix. For the Chaplygin Sleigh, the mass matrix is given by (4.2) and is not equal to the identity matrix except for the special case where $a = 0$ and $m = I = 1$. We want to generalize this integrator so it can be used in this experiment anyway.

It is noted at the end of Section 3.3.2 that the integrator (3.4) is given by the discrete Lagrange-d'Alembert equations (3.2) using the discrete Lagrangian (3.5) and discrete constraints (3.6) and also a change of variables $\widehat{q}_k = (q_k + q_{k-1})/2$ and $\widehat{v}_k = (q_k - q_{k-1})/h$.

Our approach is then simply to do the exact same thing only using instead the continuous Lagrangian $L = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q)$ where $M(q)$ is not necessarily equal to the identity

matrix. The resulting method cannot be written in a simple way in general. We therefore only derive this method for the Chaplygin Sleigh example, where the mass matrix is given by (4.2). The resulting method is as follows.

Let $q_k = \begin{bmatrix} x_k & y_k & \theta_k \end{bmatrix}^T$ and $v_k = \begin{bmatrix} v_k^x & v_k^y & v_k^\theta \end{bmatrix}^T$ for $k = 0, \cdots, K$. Then, given q_0 and v_0, let $q_{-1/2} = q_0 - \frac{1}{2}hv_0$. Then, q_1 and v_1 are found by solving the non-linear equations

$$q_{1/2} = q_0 + \frac{1}{2}hv_0, \tag{4.3a}$$

$$\begin{aligned} &\begin{bmatrix} mv_1^x - mav_1^\theta \sin\theta_{1/2} \\ mv_1^y + mav_1^\theta \cos\theta_{1/2} \\ (I+ma^2)v_1^\theta - ma\sin\theta_{1/2}\big(v_1^x - hv_1^y v_1^\theta\big) - ma\cos\theta_{1/2}\big(-v_1^y - hv_1^x v_1^\theta\big) \end{bmatrix} \\ &= \begin{bmatrix} mv_0^x - mav_0^\theta \sin\theta_{-1/2} \\ mv_0^y + mav_0^\theta \cos\theta_{-1/2} \\ (I+ma^2)v_0^\theta - mav_0^x \sin\theta_{-1/2} + mav_0^y \cos\theta_{-1/2} \end{bmatrix} \\ &\quad + h\Big(-V_q(q_{1/2}) + A(q_{1/2})^T\lambda\Big), \end{aligned} \tag{4.3b}$$

$$q_1 = q_{1/2} + \frac{1}{2}hv_1, \tag{4.3c}$$

$$A(q_1)v_1 = 0. \tag{4.3d}$$

Note that when $a = 0$ and $m = I = 1$, the mass matrix $M(q)$ reduces to I_3 and (4.3) reduces like it should to the integrator (3.4).

4.2.3 Numerical Experiment

Let the energy for the numerical solution at time-step k be given by

$$E_k \doteq \frac{1}{2}v_k^T M(q_k)v_k + V(q_k).$$

The energy of the continuous system is conserved (see Section 1.3.4 on page 11) and we wish to see how well the different numerical methods preserve the energy in their solution. We will do this by considering the absolute error in energy at each time step. That is, we consider $|E_0 - E_k|$ for $k = 1, \ldots, K$.

We choose the constants to be given by $g = 9.8$, $m = 0.001$, $a = 0.04$ and $I = 0.01$. The small values for m, a and I are chosen because we have experienced that bigger values cause the Jacobian matrix J used in the simplified Newton iterations to be very ill-conditioned after some time, resulting in great round-off errors.

An initial condition has to satisfy the constraints (4.1) and also the differentiated constraints, which can be written as

$$\left(\frac{a^2}{I} + \frac{1}{m}\right)\lambda = g\cos\theta + v_x v_\theta \cos\theta + v_y v_\theta \sin\theta. \tag{4.4}$$

As the initial value, we choose $x_0 = 1$, $y_0 = 0$, $\theta_0 = 0.2$, $v_{x,0} = 0$ and $v_{\theta,0} = 0$. To get consistency, the value of $v_{y,0}$ is then given by (4.1) and λ_0 is given by (4.4). We integrate from $t_0 = 0$ to $t_{\text{end}} = 30$.

The index-1 DAE $M_1(\xi)\dot{\xi} = f_1(\xi)$ is solved using MATLABs `ode15s`. Four different tolerance levels are used. The results are summarized in Table 4.2.3. The absolute error

in energy for each tolerance level is plotted in Figure 4.7 on the next page. The absolute error in the constraint $|A(q_k)v_k|$ for the same simulations is plotted in Figure 4.8. Both plots use a logarithmic scale on the vertical axis.

The original index-2 DAE is solved using the nonholonomic integrator (3.4) by McLachlan and Perlmutter as well as the 2- and 3-stage SPARK Lobatto IIIA-B methods and the 2- and 3-stage SRK method with both Radau IA and Gauss coefficients. The absolute error in energy for all these seven methods using a constant stepsize of $h = 0.1$ is shown in Figure 4.9 on page 51 using logarithmic values on the vertical axis. The same plot only using stepsize $h = 0.01$ is shown in Figure 4.10. We observe that the energy-error for the nonholonomic integrator and the 2-stage SPARK Lobatto IIIA-B method is about the same for both stepsizes.

As seen in the vertical rolling disk example, the constraint is not exactly satisfied for the `ode15s` solution. This is seen in Figure 4.8 to be the case also here. For all the SPARK and SRK methods however, the constraint is satisfied close to machine precision, so it is not plotted.

We compare the energy-error of the `ode15s` solution when the average stepsize is 0.1304 with the other solutions when the stepsize is $h = 0.1$. The error for `ode15s` is between the error for the nonholonomic integrator, the 2-stage SPARK Lobatto IIIA-B method and the 2-stage SRK Radau IA method. When the average stepsize for `ode15s` is 0.0163, the error is about the same as for the 2-stage SRK Radau IA method with stepsize $h = 0.01$. The 2-stage SRK Gauss method and all the 3-stage methods conserve energy better than `ode15s` for similar step-sizes.

The lowest error in energy we obtain using `ode15s` is of order 10^{-10}. Trying to reduce the tolerance level further, results in an error message. Both the SPARK and SRK methods are close to machine precision in Figure 4.10. Especially for the two SRK methods, the solution seems to be dominated by round-off errors.

Table 4.1: Average stepsize, absolute energy-error and absolute constraint error using `ode15s` on the Chaplygin Sleigh problem. Four different tolerance levels are used.

RelTol	AbsTol	Avg. h	$\max_k \lvert E_0 - E_k \rvert$	$\max_k \lvert A(q_k)v_k \rvert$
10^{-5}	10^{-8}	0.1304	$2.516 \cdot 10^{-5}$	$8.256 \cdot 10^{-4}$
10^{-7}	10^{-10}	0.0638	$8.325 \cdot 10^{-7}$	$1.320 \cdot 10^{-5}$
10^{-9}	10^{-12}	0.0303	$1.398 \cdot 10^{-8}$	$1.838 \cdot 10^{-7}$
10^{-11}	10^{-14}	0.0163	$4.267 \cdot 10^{-10}$	$8.647 \cdot 10^{-9}$

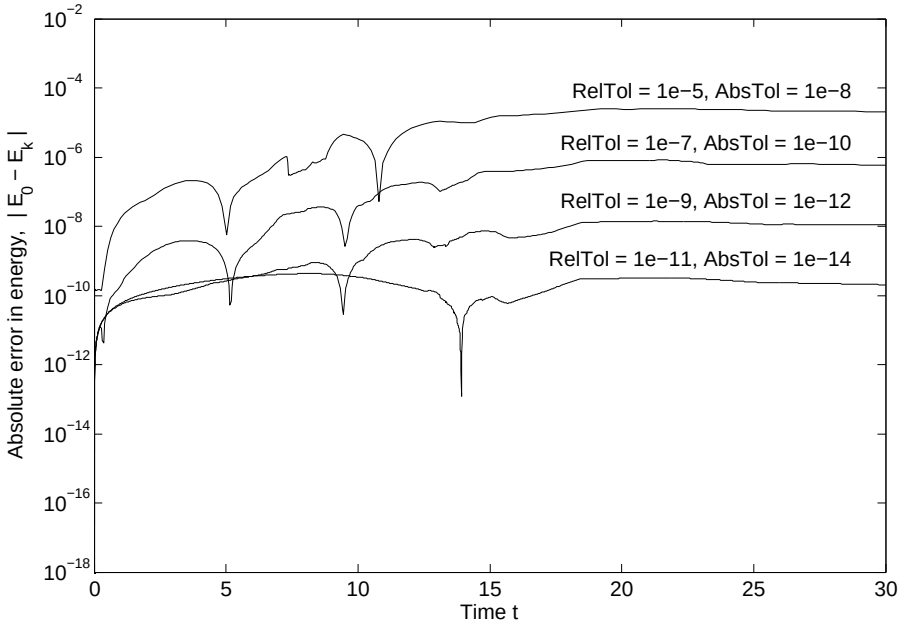

Figure 4.7: Absolute value of the energy error $|E_0 - E_k|$ for the Chaplygin Sleigh problem using `ode15s`. The different tolerance levels used are shown. The average stepsize for each tolerance level is, from the top and down, 0.1304, 0.0638, 0.0303 and 0.0163.

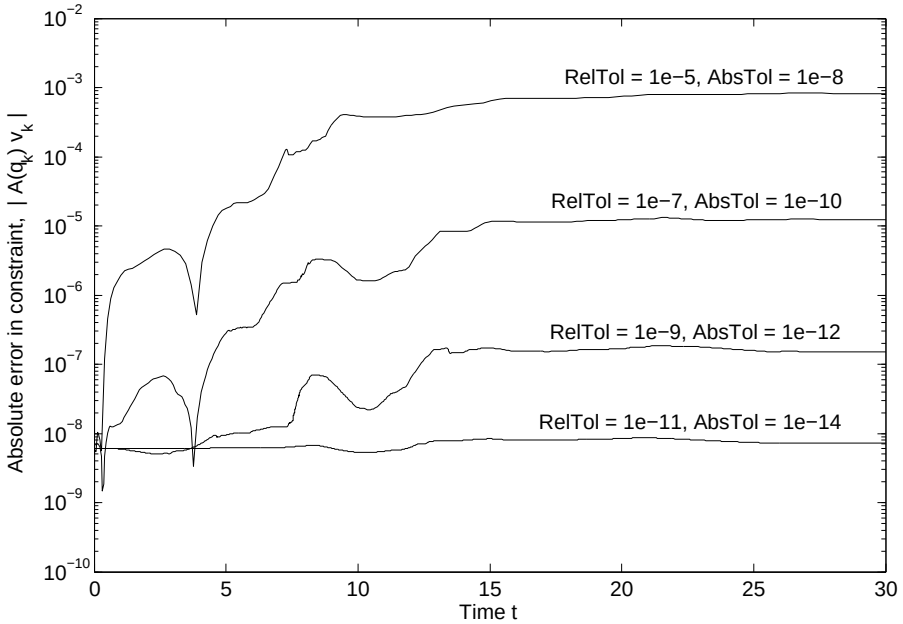

Figure 4.8: Absolute error in the constraint $|A(q_k)v_k|$ for the Chaplygin Sleigh problem using `ode15s`. The different tolerence levels used are shown. The average stepsize for each tolerence level is, from the top and down, 0.1304, 0.0638, 0.0303 and 0.0163.

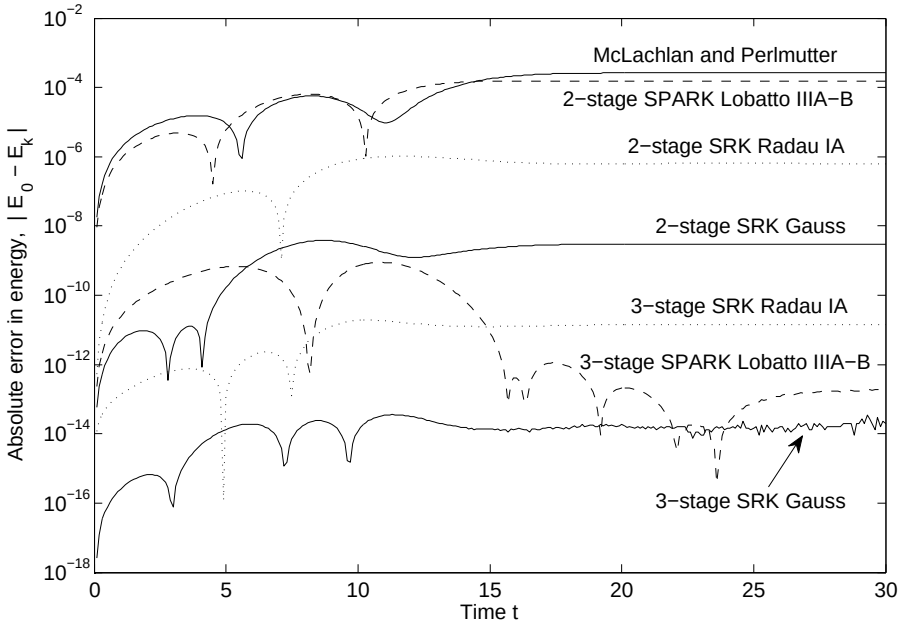

Figure 4.9: Absolute value of the energy error $|E_0 - E_k|$ for the Chaplygin Sleigh problem using the nonholonomic integrator (4.3) by McLachlan and Perlmutter and the 2- and 3-stage SPARK Lobatto IIIA-B, SRK Radau IA and SRK Gauss methods. A constant stepsize of $h = 0.1$ is used for all methods.

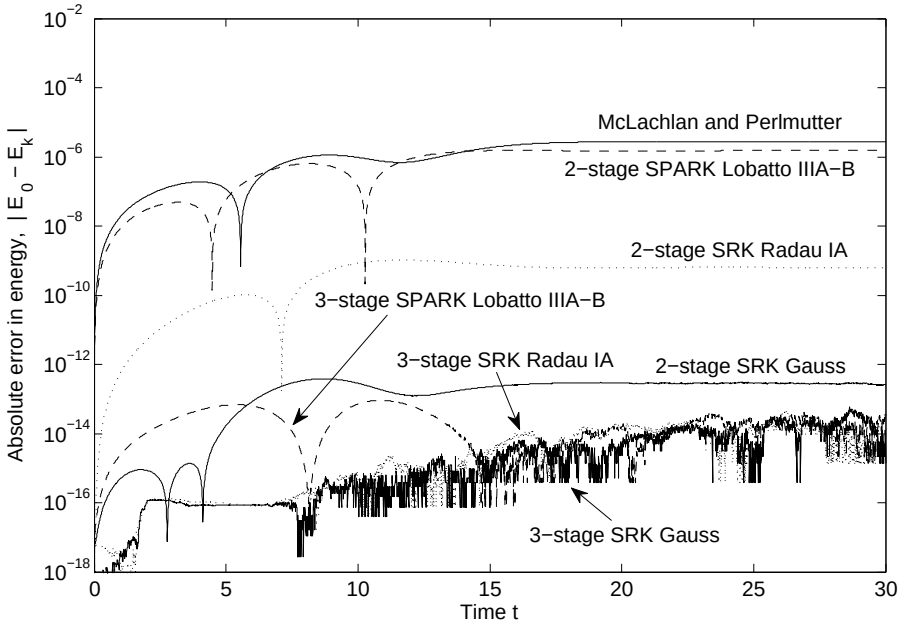

Figure 4.10: The same plot as in Figure 4.9 only now using a constant stepsize of $h = 0.01$ with all methods.

4.3 The Contact Oscillator

In this section, we consider the contact oscillator as presented in [22, Section 5.1]. We want to reproduce the results obtained by McLachlan and Perlmutter in [22] using their nonholonomic integrator (3.4). We will also do the same experiments using the 2-stage SPARK Lobatto IIIA-B method with the intention of comparing their solutions. These integrators are both of order two and they are both reversible. We investigate how well the integrators preserve both geometric properties of the solution and the energy. In the next section, we consider a nonlinear perturbation of the contact oscillator and do a similar analysis of the same two integrators.

Let the general coordinates for the system be $q = \begin{bmatrix} x & y & z \end{bmatrix}^T$. The Lagrangian for the contact oscillator is

$$L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q) = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - \frac{1}{2}\left(x^2 + y^2 + z^2\right),$$

which implies that the mass matrix is $M(q) = I_3$. The nonholonomic constraint is given by $\dot{x} + y\dot{z} = 0$, so the constraint matrix is $A(q) = \begin{bmatrix} 1 & 0 & y \end{bmatrix}$. From the Lagrange-d'Alembert equations (1.12), we find that the equations of motion for this system is given by

$$\begin{aligned} \ddot{x} + x &= \lambda, \\ \ddot{y} + y &= 0, \\ \ddot{z} + z &= \lambda y, \\ \dot{x} + y\dot{z} &= 0. \end{aligned} \tag{4.5}$$

We will have a closer look at these equations and properties of the solution. Before doing so, a few consepts need to be explained.

The *phase space* of a mechanical system is the space consisting of all possible values of the position q and momentum p. A plot of the solution of the system in the phase space or in a subspace of the phase space is referred to as a *phase portrait*. An *orbit* is a continuous curve in the phase space, such that a solution with initial condition on the orbit will stay on the orbit for all times.

4.3.1 Properties of the Exact Solution

Following the calculations done in [22], we will show that all the orbits are quasiperiodic with at most two frequencies. That is, we will show that for any solution, either all the variables are periodic with the same frequency, or all the variables are periodic, each with one of two possible frequencies.

The y-coordinate is unconstrained and hence independent of the other coordinates. Integrating the separable equation $\ddot{y} + y = 0$ twice yields $y(t) = c_1 \sin t + c_2 \cos t$. Using that $b\cos(t + \varphi) = (b\cos\varphi)\cos t + (-b\sin\varphi)\sin t \doteq c_1 \sin t + c_2 \cos t = y(t)$, we see that we can choose the origin of time such that $y(t) = b\cos t$ for some $b \in \mathbb{R}$.

By differentiating the constraint and using the equations of motion, we can explicitly solve for the Lagrangian multiplier λ and we find that

$$\lambda = \frac{x + yz - \dot{y}\dot{z}}{1 + y^2} = \frac{x + zb\cos t + \dot{z}b\sin t}{1 + b^2\cos^2 t}. \tag{4.6}$$

Introducing the notation $v_x = \dot{x}$ and $v_z = \dot{z}$, the equations of motion (4.5) give the four equations

$$\begin{aligned} \dot{x} &= v_x, & \dot{v}_x &= \lambda - x, \\ \dot{z} &= v_z, & \dot{v}_z &= \lambda b \cos t - z, \end{aligned}$$

where now λ is given by (4.6) and linear in the unknowns x, z and $\dot{z}$. Thus these equations form a system of nonautonomous linear ODEs. From the constraint, we have that $\dot{x} = -v_z a \sin t$, which means we can eliminate v_x from the above equations. This leaves us with a system of three ODEs for x, z and v_z. With the new variable $\widetilde{v}_z \doteq (1 + b^2 \cos^2 t)^{1/2} v_z$, these three equations can be written as

$$\begin{bmatrix} \dot{x} \\ \dot{z} \\ \dot{\widetilde{v}}_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & \alpha(t) \\ 0 & 0 & \beta(t) \\ -\alpha(t) & -\beta(t) & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ \widetilde{v}_z \end{bmatrix} \doteq A(t) \begin{bmatrix} x \\ z \\ \widetilde{v}_z \end{bmatrix}, \tag{4.7}$$

where

$$\alpha(t) = \frac{-b \cos t}{\sqrt{1 + b^2 \cos^2 t}} \quad \text{and} \quad \beta(t) = \frac{1}{\sqrt{1 + b^2 \cos^2 t}}.$$

Let now $\xi(t) = \begin{bmatrix} x & z & \tilde{v}_z \end{bmatrix}^T$ and for each b, let $Y(t)$ be such that $\xi(t) = Y(t)\xi(0)$ and $Y(0) = I$. Since $A(t)$ is skew-symmetric, it is known that $Y(t) \in \mathrm{SO}(3)$, where $\mathrm{SO}(3)$ is the special orthogonal group in 3 dimensions. That is, $Y(t)$ is orthogonal with determinant $+1$. Since $A(t)$ also is 2π-periodic, we must have that $Y(t + 2\pi) = Y(t)$. This means that there is an orthogonal matrix $\Omega(b) \in \mathrm{SO}(3)$ independent of time such that the time-2π flow of (4.7) is given by $\begin{bmatrix} x & z & \tilde{v}_z \end{bmatrix}^T \mapsto \Omega(b) \begin{bmatrix} x & z & \tilde{v}_z \end{bmatrix}^T$. As all orthogonal 3×3 matrices are, this is a rotation about some axis, where now both the angle of rotation and the axis depend on b.

The orbits of (4.7) are in [22] classified into three cases. We will explain each case a bit more detailed.

Case (i). If $b = 0$, then $\alpha(t) = 0$ and $\beta(t) = 1$ for all t. This makes $A(t)$ a constant matrix, and we can easily solve (4.7) since A is diagonalizable. We find that

$$Y(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{bmatrix},$$

which is a 2π-periodic rotation about the x-axis. That $b = 0$ implies $v_z(t) = \widetilde{v}_z(t)$ and $y(t) = \dot{y}(t) = 0$ for all t. From considering $Y(t)$, we see that $x(t) = x(0) \doteq x_0$ and $v_x(t) = 0$. The orbits in this case are thus circles in the z-$\widetilde{v}_z$ plane at $x = x_0$, centered at $z = v_z = 0$ and with period 2π. In addition to x_0, the radius of these circles need to be determined, so we have two degrees of freedom. We use x_0 and the initial energy E_0 as parameters. Note that even though the orbit is fully determined by the choice of x_0 and E_0, we are still left with the choice of where on the orbit we set the origin of time.

Case (ii). If $b \neq 0$ and (x, z, v_z) happens to lie on the axis of rotation of $\Omega(b)$, then $\Omega(b)$ will leave (x, z, v_z) unchanged and hence these coordinates are 2π-periodic. This implies that v_x must be 2π-periodic as well, and we also know that both $y(t) = b \cos t$ and $\dot{y}(t) = -b \sin t$ are. The orbits are therefore 2π-periodic. The orbits are determined by two parameters. The axis of rotation of $\Omega(b)$ is determined by the value of b and where on this axis (x, z, v_z) lies can be determined by the initial energy E_0.

Case (iii). If $b \neq 0$ and (x, z, v_z) does not lie on the axis of rotation of $\Omega(b)$, we will have a 2π-periodicity in y and $\dot{y}$ and the other coordinates will be periodic with period $2\pi/\gamma$, where γ is the angle of rotation of $\Omega(b)$. In addition to the two degrees of freedom we had in case (ii), we now also have the distance of (x, z, v_z) from the axis of rotation of $\Omega(b)$. We choose to use z_0 as our third parameter.

We have shown that all the orbits are quasiperiodic with at most two frequencies. Specifically, we found that the orbits are 2π-periodic in case (i) and (ii) above, and quasiperiodic with periods 2π and $2\pi/\gamma$ in case (iii).

4.3.2 Properties of the Numerical Solutions

An analysis of how the nonholonomic integrator (3.4) behaves when applied to this problem is carried out in [22]. We first do this same analysis below and then we make a few comments about the 2-stage SPARK Lobatto IIIA-B method.

McLachlan and Perlmutter. As the discrete variables $(y, v_y)_k$ are unconstrained, it can be seen from (3.4) that $(y, v_y)_k \mapsto M(h)(y, v_y)_k$, where

$$M(h) = \begin{bmatrix} 1 - h^2/2 & h(4 - h^2)/4 \\ -h & 1 - h^2/2 \end{bmatrix}.$$

If we assume $h < 2$, $M(h)$ has eigenvalues

$$\lambda = \frac{1}{2}\left(2 - h^2\right) \pm i\frac{1}{2}h\sqrt{4 - h^2} = |\lambda|\, e^{i \arg(\lambda)},$$

where $|\lambda| = 1$ and

$$\arg(\lambda) = 2 \arctan\left(\frac{\mathrm{Im}(\lambda)}{|\lambda| + \mathrm{Re}(\lambda)}\right) = \pm 2 \arcsin\left(\frac{h}{2}\right) \doteq \pm\theta.$$

As $M(h)$ is diagonalizable, we have $M(h) = S\Lambda S^{-1}$, where S depends on h and $\Lambda = \mathrm{diag}(e^{i\theta}, e^{-i\theta})$. So if we choose $\theta = 2\pi/N$ for some integer N, then $\Lambda = \mathrm{diag}(e^{i2\pi/N}, e^{-i2\pi/N})$ and so $M^N = S\Lambda^N S^{-1} = SIS^{-1} = I$. This means that if we choose $h = 2\sin(\pi/N)$ for some positive integer N, then $(y, v_y)_{k+N} = M^N(y, v_y)_k = (y, v_y)_k$, so the numerical solution $(y, v_y)_k$ is periodic with period N.

SPARK Lobatto IIIA-B. We have not done a similar analytical analysis of the SPARK Lobatto IIIA-B method, but we have found through experiments that also the 2-stage SPARK Lobatto IIIA-B method is periodic with period N using the same stepsize h as the integrator by McLachlan and Perlmutter (3.4). That is, using stepsize $h = 2\sin(\pi/N)$.

4.3.3 Numerical Experiment

To see if the integrators can quantitively preserve the geometry explained in Section 4.3.1 above, we redo the example done in [22]. Different initial values are used that each defines a solution on an orbit of type (iii), i.e. the solution is quasiperiodic. We are then interested to see whether the numerical solutions preserve these orbits, and we compare

the nonholonomic integrator (3.4) with the 2-stage SPARK Lobatto IIIA-B method. For the SPARK Lobatto IIIA-B method, we use the decomposition described in Section 2.5.3 on page 27.

For case (iii) above, we chose b, E_0 and z_0 as parameters to determine the orbit. For an initial condition, we need one parameter more to determine where on the orbit we start at $t = 0$, and for this, $v_{z,0}$ is used. A consistent initial value using these four parameters is then given by

$$q_0 = \begin{bmatrix} \left(2E_0 - 2v_{z,0}^2 - z_0^2 - 1\right)^{1/2} \\ b \\ z_0 \end{bmatrix}, \quad v_0 = \begin{bmatrix} -bv_{z,0} \\ 0 \\ v_{z,0} \end{bmatrix}, \quad \lambda_0 = \frac{x_0 + bz_0}{1 + b^2}. \tag{4.8}$$

We wish to reproduce the same plot for the contact oscillator as in [22, Figure 1]. The initial data for this plot were kindly given to us by Robert McLachlan. Ten different initial values are used. They are given by (4.8), where

$$\begin{aligned} b = 1, \quad E_0 = 1.5, \quad v_{z,0} = 0, \quad \text{and} \\ z_0 = (-0.9 + 0.2k)\sqrt{2}, \quad \text{for } \ k = 0, 1, \ldots, 9. \end{aligned} \tag{4.9}$$

We define a Poincaré section[1] by $v_y = 0$, $\dot{v}_y < 0$. As v_y is 2π-periodic, the Poincaré map is the 2π-flow of the system. From our earlier analysis, we know that the 2π-flow of the solution in (x, z, v_z)-space is a rotation about some fixed axis (determined by the value of b). Therefore, when plotting this Poincaré section, we expect the solution to contain a set of circles. It is given in [22] that the Poincaré section is in fact a sphere in (x, z, v_z)-space.

In Figure 4.11, we plot the numerical solution on this section[2] for $x > 0$, which then shows one half of the sphere. 8000 iterations of the Poincaré map with 40 timesteps per iteration is shown. The solution is plotted for both the nonholonomic integrator (3.4) and the 2-stage SPARK Lobatto IIIA-B method. In both plots, we clearly recognize the shape of the circles we expect to find. All orbits look qualitatively correct for the nonholonomic integrator (3.4). For the 2-stage SPARK Lobatto IIIA-B method, three of the orbits seem to drift. Numbering the orbits from left to right, these three orbits are numbered 7, 8 and 9. In Figure 4.12, we plot the energy for all ten orbits in the SPARK Lobatto IIIA-B solution. We see that the same three orbits have a clear drift in energy, whereas for all the other orbits, the energy stays bounded for the whole simulation. It seems to be a connection between conservation of energy and the geometric correctness of the solution. This is investigated further in the next experiment.

[1] The Poincaré section is in [22] defined by $v_y = 0$, $\dot{v}_y > 0$ (note the difference in the inequality sign). If we use this definition, our plots in Figure 4.11 on the following page become flipped about $z = 0$. This might be a simple sign error in our implementation, but as the geometry seem to be correct, we do not consider this important.

[2] Explained in words — every time the numerical solution has $v_y = 0$ and the value of v_y is decreasing, we plot a small dot if $x > 0$.

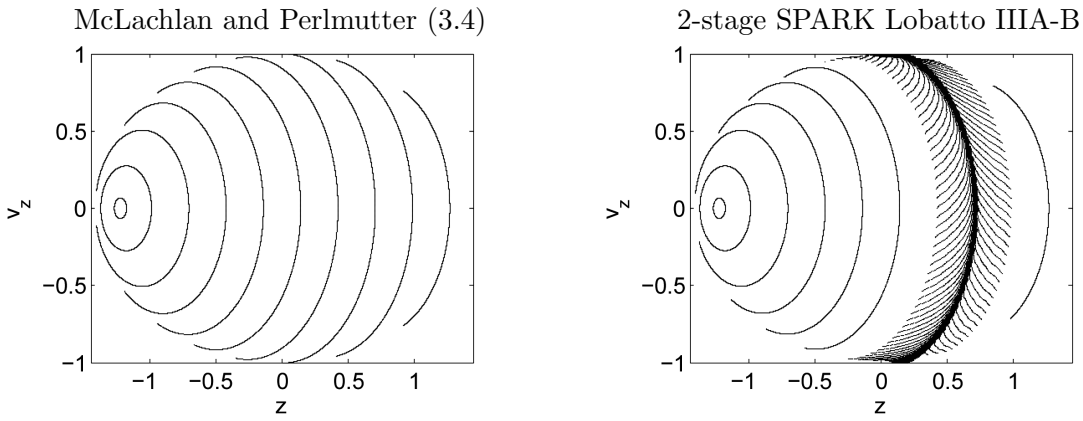

Figure 4.11: The Poincaré section of the contact oscillator defined by $v_y = 0$, $\dot{v}_y < 0$ for $x > 0$. To the left using the nonholonomic integrator (3.4) and to the right using the 2-stage SPARK Lobatto IIIA-B method. 8000 iterations of the Poincaré map are shown with 40 timesteps per iteration using the 10 different initial values given by (4.8) and (4.9).

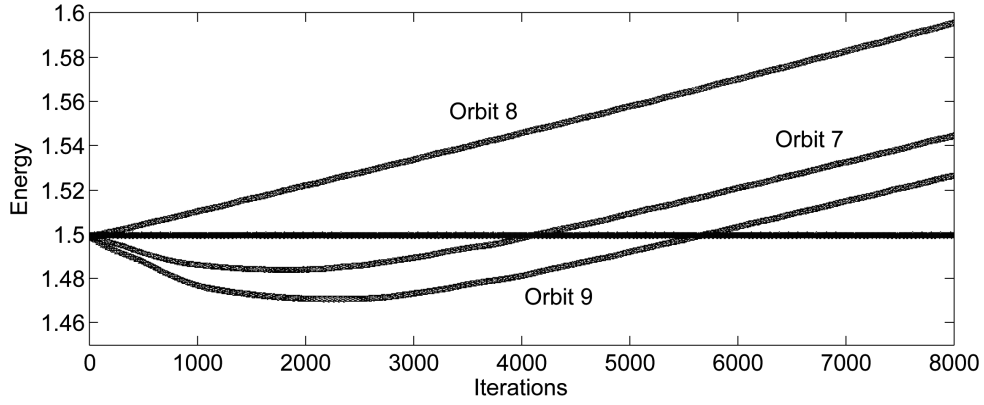

Figure 4.12: The energy of the SPARK Lobatto IIIA-B solution for each of the 10 initial conditions shown in Figure 4.11. We number the orbits in Figure 4.11 from left to right. The energy corresponding to orbits 7, 8 and 9 all have a clear drift. The other 7 orbits are also plotted. They all stay bounded and close to the initial energy level of 1.5.

4.4 The Perturbed Contact Oscillator

We consider a nonlinear perturbation of the contact oscillator. This example is, like the contact oscillator was, presented in [22].

Using coordinates $q = \begin{bmatrix} x & y & z \end{bmatrix}^T$, this system is defined by the Lagrangian

$$L(q, \dot{q}) = \frac{1}{2}\dot{q}^T M(q)\dot{q} - V(q) = \frac{1}{2}\dot{q}^T\dot{q} - \frac{1}{2}\Big(q^T q + \varepsilon x^2 z^2\Big)$$

and constraint matrix is still $A(q) = \begin{bmatrix} 1 & 0 & y \end{bmatrix}$. Note that this system reduces to the contact oscillator for $\varepsilon = 0$. As $M(q) = I_3$, the momentum p is simply given by $p = \partial L/\partial\dot{q} = \dot{q}$.

The Lagrange-d'Alembert equations (1.12) give

$$\begin{aligned} \ddot{x} + x &= \lambda - \varepsilon x z^2, \\ \ddot{y} + y &= 0, \\ \ddot{z} + z &= y\lambda - \varepsilon x^2 z, \\ \dot{x} + y\dot{z} &= 0. \end{aligned} \tag{4.10}$$

Note that the y-coordinate is still unconstrained as in the unperturbed case. As before, we choose the origin of time such that $y(t) = b\cos t$.

Differentiating the constraint equation $\dot{x} + y\dot{z} = 0$ with respect to the time t, using (4.10) and solving for λ yields

$$\lambda = \frac{x + yz - \dot{y}\dot{z} + \varepsilon\Big(xz^2 + x^2 yz\Big)}{1 + y^2}. \tag{4.11}$$

Denote the velocity $\dot{q}$ by $v = \begin{bmatrix} v_x & v_y & v_z \end{bmatrix}^T$. We continue to use b, z_0, E_0 and $v_{z,0}$ as parameters for an initial condition. With these parameters, a consistent initial condition is given by

$$q_0 = \begin{bmatrix} x_0 \\ b \\ z_0 \end{bmatrix}, \quad v_0 = \begin{bmatrix} -bv_{z,0} \\ 0 \\ v_{z,0} \end{bmatrix}, \tag{4.12}$$

where x_0 is given by

$$x_0 = \sqrt{\frac{2E_0 - 2v_{z,0}^2 - z_0^2 - 1}{1 + \varepsilon z_0^2}}. \tag{4.13}$$

The initial value for λ is found by using (4.11).

Like we did for the unperturbed contact oscillator, we define a Poincaré section by $v_y = 0$, $\dot{v}_y < 0$ and the Poincaré map is still the 2π-flow of the system since y is 2π-periodic as before. We will look at the solutions from both the nonholonomic integrator (3.4) and the 2-stage SPARK Lobatto IIIA-B method.

In Figure 4.13 on the next page, the Poincaré section for $x > 0$ with $\varepsilon = 0.3$ is plotted. The solutions using both of the two integrators are shown. The same initial values are used as for the contact oscillator, which were given by (4.9). 8000 iterations of the Poincaré map for each of the 10 initial values are shown using 40 timesteps per iteration. Three of the orbits in the SPARK Lobatto IIIA-B solution seem to break up

and behave chaotic, whereas the same orbits in the solution by the nonholonomic integrator (3.4) are quasiperiodic and not chaotic.

In Figure 4.14 on the facing page, the same plots are made using $\varepsilon = 0.8$ and z_0 values given by

$$z_0 = -1.2,\ -1,\ -0.75,\ -0.55,\ -0.4,\ -0.1,\ 0.93,\ 1.07.$$

The rest of the initial values are the same as before and given by (4.9). These values are the same as were used in [22, Figure 1]. 8000 iterations of the Poincaré map are shown using 40 timesteps per iteration. In the solution by nonholonomic integrator (3.4), there is one quasiperiodic orbit in between the chaotic orbits. In the SPARK Lobatto IIIA-B solution on the other hand, this orbit is also chaotic.

We do a longer simulation of the first chaotic orbit from the left in Figure 4.14 (corresponding to initial value $z_0 = -0.55$). 50000 iterations of the Poincaré map, still using 40 timesteps per iteration, are shown in Figure 4.15 on the next page for both integrators. For the nonholonomic integrator (3.4), the Poincaré map seems to visit different parts of the chaotic orbit evenly throughout the simulation. For the SPARK Lobatto IIIA-B method however, the Poincaré map visits different parts of the chaotic orbit at first. Then, after about 2500 iterations, it stays close to the two circles seen as darker areas in Figure 4.15 for the rest of the simulation.

The energy for the same simulation is shown in Figure 4.16 on page 60. The energy using the integrator (3.4) stays bounded, but the energy for the SPARK Lobatto IIIA-B solution drifts away from the initial value of 1.5 and setteles at a lower energy level. Note that the energy starts to level off after about 2500 iterations, which is when the solution in Figure 4.15 starts to stay close to the two circles. To have a closer look at what happens with the SPARK Lobatto IIIA-B solution for this simulation, the energy for the first 1000 iterations is shown in Figure 4.17 on page 60. The energy seems to stay bounded for the first 500 iterations, before it suddenly drifts away. Like we did in the unperturbed contact oscillator example, we once again see a clear connection between the geometry of the solution and energy-conservation.

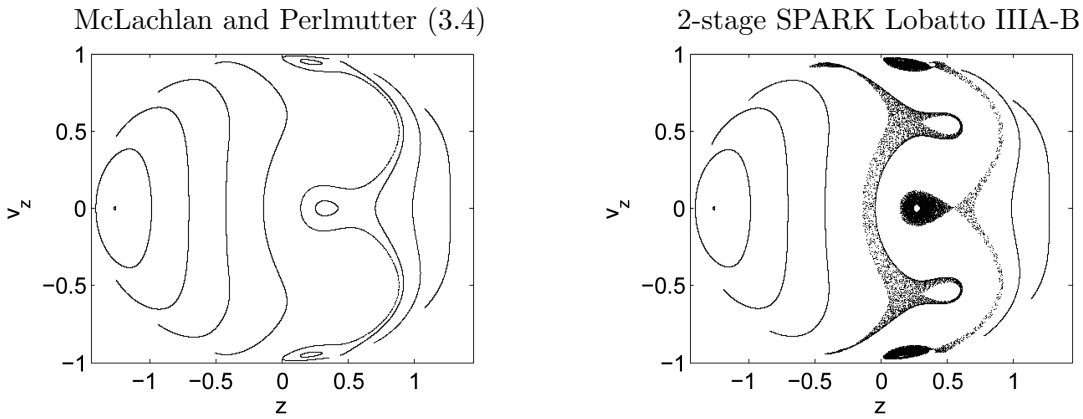

Figure 4.13: The Poincaré map defined by $v_y = 0$, $\dot{v}_y < 0$ of the perturbed contact oscillator with $\varepsilon = 0.3$. To the left using the nonholonomic integrator (3.4) and to the right using the 2-stage SPARK Lobatto IIIA-B method. 8000 iterations of the Poincaré map are taken, with 40 timesteps per iteration for each of 10 initial values. Only values for which $x > 0$ are shown.

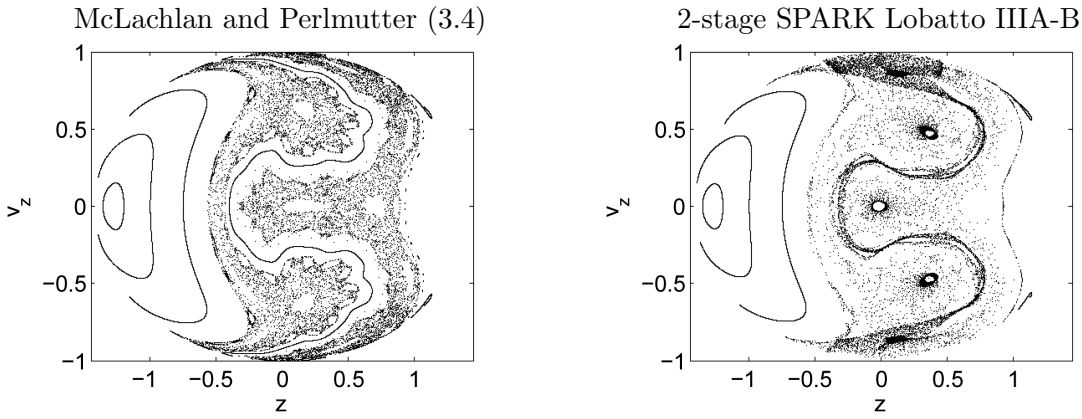

Figure 4.14: The Poincaré map defined by $v_y = 0$, $\dot{v}_y < 0$ of the perturbed contact oscillator with $\varepsilon = 0.8$. To the left using the nonholonomic integrator (3.4) and to the right using the 2-stage SPARK Lobatto IIIA-B method. 8000 iterations of the Poincaré map are taken, with 40 timesteps per iteration for each of 8 initial values. Only values for which $x > 0$ are shown.

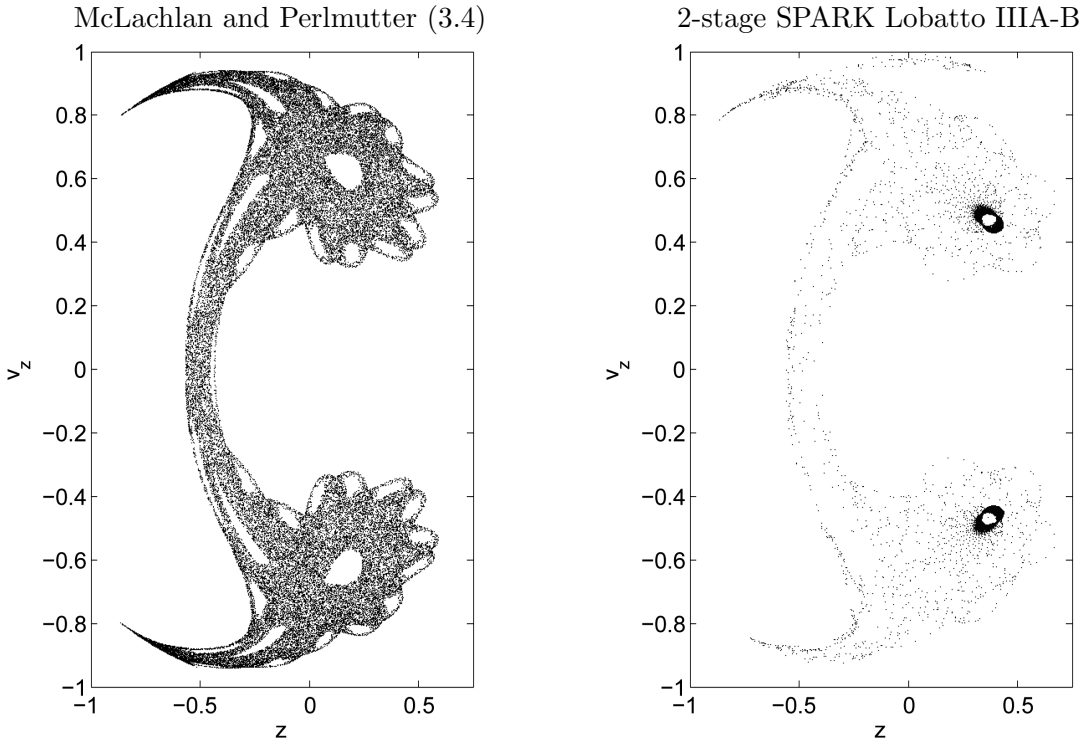

Figure 4.15: 50,000 iterations of the first chaotic orbit from the left in Figure 4.14 (corresponding to initial value $z_0 = -0.55$). To the left using the nonholonomic integrator (3.4) and to the right using the 2-stage SPARK Lobatto IIIA-B method. The energy for this simulation is shown in Figure 4.16 on the next page.

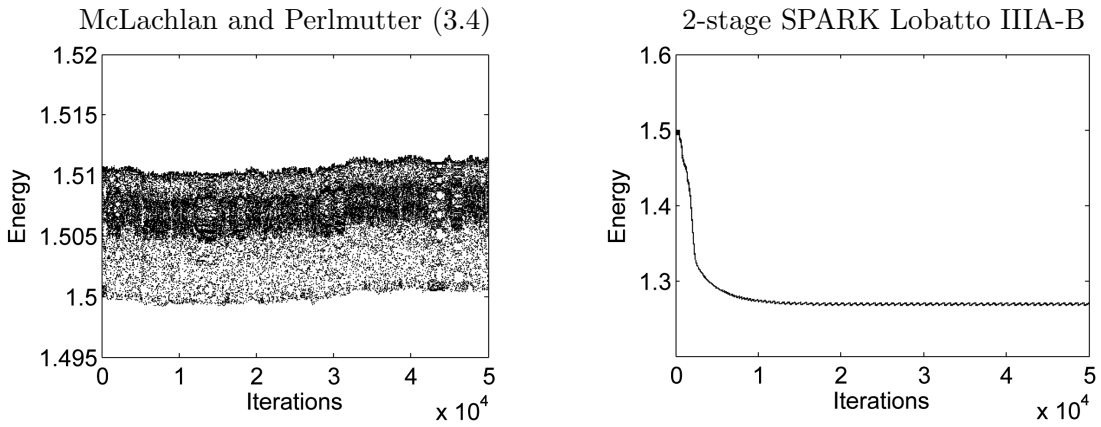

Figure 4.16: Energy for the same simulation as in Figure 4.15 on the previous page. To the left using the nonholonomic integrator (3.4) and to the right using the 2-stage SPARK Lobatto IIIA-B method. The initial energy is 1.5. The energy for the integrator (3.4) stays well bounded, whereas the energy for the SPARK Lobatto IIIA-B method drifts away from the initial value.

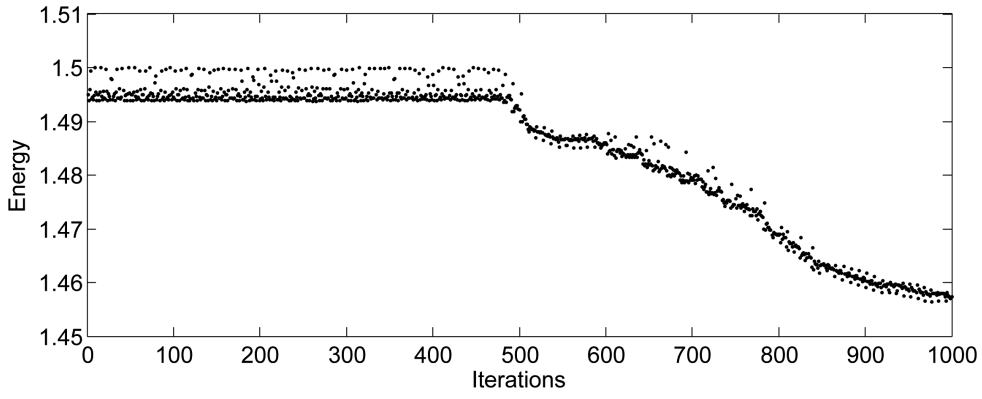

Figure 4.17: Close up of the right plot in Figure 4.16. The energy for the first 1000 iterations of the Poincaré map using the SPARK Lobatto IIIA-B method is shown. The initial energy is 1.5. The energy stays bounded until about 500 iterations before it drifts away.

4.5 Comparison with Results by Laurent O. Jay

To argue that our implementation of SPARK methods is correct, we compare plots obtained from Laurent O. Jay through private communications with our own plots. The experiment we consider is the Chaplygin sleigh, which we also discuss in Section 4.2. As the details are discussed there, we only give the constants and initials conditions chosen here.

The constants are chosen as $m = 1$, $a = 1$, $g = 0.1$ and $I = 1$. The initial condition is given by $q_0 = (1, 0, 0)^T$, $v_0 = (0, 0, 0)^T$ and $\lambda_0 = 1/20$. A stepsize of $h = 0.12$ is used. The Chaplygin sleigh experiment with these values was originally carried out in [7].

The value for each of the three components using the 4-stage SPARK Lobatto IIIA-B method is shown in Figure 4.18. The error in the energy for the 2-, 3- and 4-stage SPARK Lobatto IIIA-B methods is shown in Figure 4.19.

The plot showing the error in the components is indistinguishable from the same plot created by Jay. Considering the plots showing the energy-error, no difference is seen in the plots for the 2- and 3-stage SPARK Lobatto IIIA-B methods. For the 4-stage SPARK Lobatto IIIA-B method, the plots are very similar, but not identical. Only a slight difference in the irregularities is seen. As these irregularities are caused by round-off errors, this is not surprising and one cannot expect identical plots.

These results indicate that our implementation is correct. Our other experiments sometimes show a poor energy conservation for the 2-stage SPARK Lobatto IIIA-B method. The results in [19] on the other hand show very good energy conservation for this method. Because of this, we do not exclude that there might be room for improvement in our implementation. One explanation might be that we do not solve the nonlinear SPARK equations accurately enough. We discuss this matter further in the conclusion in Chapter 5.

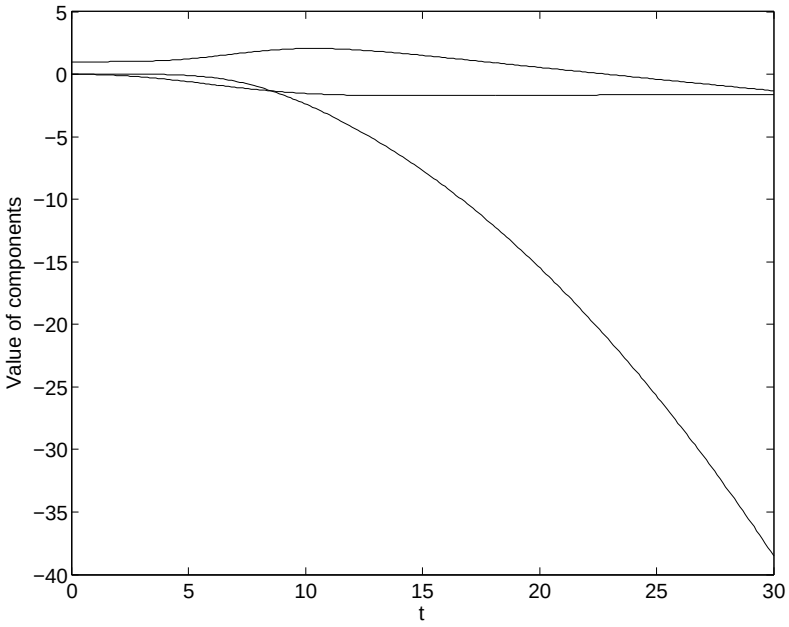

Figure 4.18: The value of each of the three components using the 4-stage SPARK Lobatto IIIA-B method.

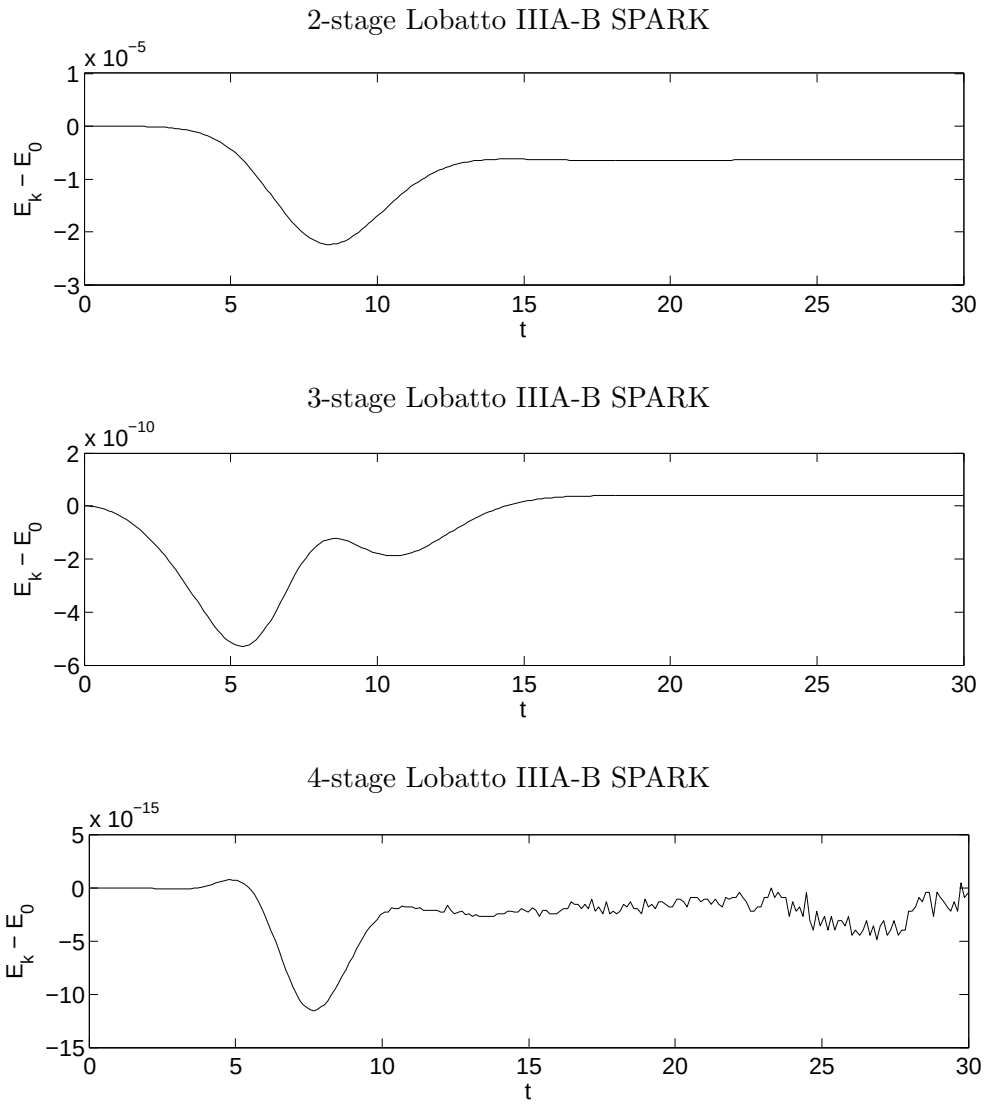

Figure 4.19: The error in energy for the Chaplygin sleigh example.

Chapter 5

Conclusion

Summary and Conclusion

Two main types of integrators for nonholonomic systems are studied — SPARK and SRK methods on the one hand and nonholonomic integrators on the other. Through numerical experiments, both types of methods are seen to perform well compared to standard integrators such as MATLABs `ode15s`. The nonholonomic constraints are satisfied at the solution and energy is in general well conserved.

In the vertical rolling disk experiment, we compare the error in the solution for the 2-stage SPARK Lobatto IIIA-B method and the nonholonomic integrator (3.4). We see that while the error for the nonholonomic integrator behaves well and stays bounded, a linear growth in the error for the SPARK Lobatto IIIA-B solutions is observed. The SPARK Lobatto IIIA-B solutions also have a drift in the energy-error, while no drift is observed for the nonholonomic integrator.

In the Chaplygin Sleigh experiment the SPARK and SRK methods show good conservation of energy in short time simulations. We see that increasing the order of the method and lowering the size of the timestep both reduce the energy-error as one would expect. Compared to MATLABs `ode15s`, the higher order methods perform very well.

In the contact oscillator experiments we compare the 2-stage SPARK Lobatto IIIA-B method with the nonholonomic integrator (3.4). These two methods are both reversible and of order 2. We focus in these experiments on observing the geometry of the solution and also on long-time energy conservation. In some cases, a clear drift in the energy for the SPARK Lobatto IIIA-B solution is observed and this is believed to cause the geometry of the solution to sometimes be incorrect. This shows that energy conservation is crucial and can be a good measure of how well a method performs. The nonholonomic integrator (3.4) shows very good energy conservation in these experiments.

In a recent work by Jay [19] the 2-stage SPARK Lobatto IIIA-B method is compared to the nonholonomic integrator (3.4) by McLachlan and Perlmutter. In a numerical experiment in [19] (the McLachlan and Perlmutter's particles) the energy-conservation for long-time integration is seen to be equally good for the two methods. Even though we do not perform the same experiment, our results several times show a better conservation of energy for the nonholonomic integrator. In the Chaplygin sleigh example the energy error for the nonholonomic integrator (3.4) and the 2-stage SPARK Lobatto IIIA-B method is about the same. But, in the contact oscillator example, where we integrate much longer in time, the energy in some cases drift off in the 2-stage SPARK Lobatto IIIA-B solution, but not in the nonholonomic integrator solution. Also, in the vertical rolling disk experiment,

the energy in the 2-stage SPARK Lobatto IIIA-B solution has a drift.

Because of the difference in our results and those in [19], we do not exclude that there might be room for improvement in our implementation. One explanation for this might be that we do not solve the nonlinear SPARK equations accurately enough. To argue that our implementation still is correct, we compare our own results to results by Laurent O. Jay in Section 4.5. We observe that the energy-conservation is as good as identical for our results and those done by Jay. How to solve the nonlinear SPARK equations is discussed in [15, 16] and also recently in [20]. These references should be studied in greater detail. As we did not become aware of the results in [19] before at the very end, we have not had time to study this more. We will leave it for further work.

We also note that in resent work done in [19], a new discrete principle is defined based on the discrete Lagrange-d'Alembert principle for forced Lagrangian systems and it is shown that a large class of SPARK methods satisfies this principle. This means that the two very different approaches for obtaining numerical methods we consider in this thesis may be closer related than what we have thought.

Further Work

There are many possibilities for further work. If we had had more time to work on this thesis, we would have looked into the following.

- Repeat the numerical experiment in [19] where the 2-stage Lobatto IIIA-B SPARK method is compared to the nonholonomic integrator (3.4) by McLachlan and Perlmutter. See if the same results are obtained. If not, study in greater detail how to solve the nonlinear SPARK equations. See [15, 16, 20] for more on this. This would have been the first issue we would have done.

- Consider systems with mixed holonomic and nonholonomic constraints. For example can SPARK methods be considered for this purpose as these methods can handle such mixed constraints (see e.g. [14]).

- We have not focused on efficient implementations, but as long-time integrations often are of interest, efficiency could be looked into more. For the SPARK methods, the use of a preconditioner, discussed in [15, 16], could be considered. If writing in MATLAB, implementing parts of the code as Fortran or C sub-routines by the use of mex-files can create huge cuts in computational time.

Bibliography

[1] A.M. Bloch, J. Baillieul, P. Crouch, and J. Marsden. *Nonholonomic Mechanics and Control*. Springer New York, 2003.

[2] John C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley-Blackwell, 2nd edition, 2008.

[3] Jr. C. H. Edwards and David E. Penny. *Elementary Linear Algebra*. Prentice Hall, Inc., 1988.

[4] S. L. Campbell, V. Hoang Linh, and L. R. Petzold. Differential-algebraic equations. *Scholarpedia*, 3(8):2849, 2008. Revision nr. 44943.

[5] J. Cortés. Geometric, control and numerical aspects of nonholonomic systems. *Lecture Notes in Mathematics*, 1793:219, 2002.

[6] J. Cortés and S. Martínez. Non-holonomic integrators. *Nonlinearity 14*, pages 1365–1392, 2001.

[7] Dag Frohde Evensberget. Numerical simulation of nonholonomic dynamics. Master's thesis, Norwegian University of Science and Technology, July 2006.

[8] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization*. Academic Press, 1984.

[9] Ernst Hairer, Christian Lubich, and Michel Roche. *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*, volume 1409 of *Lecture Notes in Mathematics*. Springer-Verlag, 1989.

[10] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration*. Springer New York, 2nd edition, 2006.

[11] Ernst Hairer and Gerhard Wanner. *Solving Ordinary Differential Equations: Stiff and Differential-Algebraic Problems*. Springer New York, 1991.

[12] Laurent O. Jay. Convergence of a class of Runge-Kutta methods for differential-algebraic systems of index 2. *BIT*, 33:137–150, 1993.

[13] Laurent O. Jay. Symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems. *SIAM J. Numer. Anal.*, 33:368–387, 1996.

[14] Laurent O. Jay. Structure preservation for constrained dynamics with super partitioned additive runge-kutta methods. *SIAM J. Sci. Comput*, 20:416–446, 1998.

[15] Laurent O. Jay. Inexact simplified Newton iterations for implicit Runge-Kutta methods. *SIAM J. Numer. Anal.*, 38:1369–1388, 2000.

[16] Laurent O. Jay. Iterative solution of nonlinear equations for SPARK methods applied to DAEs. *Numer. Alg*, 31:171–191, 2002.

[17] Laurent O. Jay. Solution of index 2 implicit differential-algebraic equations by Lobatto Runge-Kutta methods. *BIT Numerical Mathematics*, 43:93–106, 2003.

[18] Laurent O. Jay. Specialized Runge-Kutta methods for index 2 differential-algebraic equations. *Mathematics of Computation*, 75:641–654, 2006.

[19] Laurent O. Jay. Lagrange-d'Alembert SPARK integrators for nonholonomic Lagrangian systems. Submitted for publication. `http://www.math.uiowa.edu/~ljay/publications.html`, June 2009.

[20] Laurent O. Jay. On modified Newton iterations for SPARK methods applied to constrained systems in mechanics. Conference proceedings paper. Submitted. `http://www.math.uiowa.edu/~ljay/publications.html`, June 2009.

[21] J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numerica*, pages 357–514, 2001.

[22] R. McLachlan and M. Perlmutter. Integrators for nonholonomic mechanical systems. *J. Nonlinear Sci.*, 16:283–328, 2006.

[23] Lawrence F. Shampine. *Numerical Solution of Ordinary Differential Equations*. Chapman & Hall, 1994.

[24] Loring W. Tu. *An Introduction To Manifolds*. Springer New York, 2008.

[25] A.J. van der Schaft and B.M. Maschke. On the Hamiltonian formulation of nonholonomic mechanical systems. *Reports on mathematical physics*, 34(2):225–233, 1994.

[26] Eric W. Weisstein. Descartes' sign rule. `http://mathworld.wolfram.com/DescartesSignRule.html`, April 2009.

[27] M. West. *Variational Integrators*. PhD thesis, California Institute of Technology, 2004.

Appendix A

Mathematical Background

This thesis assumes a basic knowledge about linear algebra, partial differential equations and numerics in general. As vector notation is used throughout, a brief review of vector differentiation is given in Section A.1. Also, some definitions in differential geometry are given in Section A.2.

A.1 Vector Differentiation

Let x be a column vector $x = \begin{bmatrix} x^1 & \dots & x^n \end{bmatrix}^T \in \mathbb{R}^n$. We define the vector differentiation operator as the column vector[1]

$$\frac{d}{dx} = \begin{bmatrix} \frac{\partial}{\partial x^1} & \cdots & \frac{\partial}{\partial x^n} \end{bmatrix}^T.$$

Some properties of this operator are given without proof. First,

$$\frac{d}{dx}\left(x^T x\right) = 2x.$$

If $b \in \mathbb{R}^n$ is a column vector independent of x, then

$$\frac{d}{dx}\left(b^T x\right) = \frac{d}{dx}\left(x^T b\right) = b.$$

Let A be an m-by-n matrix and let S be a symmetric n-by-n matrix, both independent of x. Then

$$\frac{d}{dx}\left(Ax\right) = A,$$

and

$$\frac{d}{dx}\left(x^T S x\right) = 2Sx.$$

Let $f = (f^1, \dots, f^m) \colon \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function. The matrix of partial derivatives, called the *Jacobian matrix of* f, is defined as

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f^1}{\partial x^1} & \cdots & \frac{\partial f^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial f^m}{\partial x^1} & \cdots & \frac{\partial f^m}{\partial x^n} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

[1] Some authors choose to define the vector differentiation operator as a row vector.

The notation $f_x = \partial f / \partial x$ will also be used.

A.2 Differential Geometry

This section does not give a review of differential geometry. A few definitions that are used in the thesis are simply stated. Some basic consepts of manifolds are used. However, differential geometry is not of great importance throughout this thesis and it can easily be read without any knowledge about the subject. Readers not familiar with differential geometry are referred to [24].

Definition A.1 (Tangent space)**.** Let Q be a manifold. For each point $q \in Q$, the *tangent space* T_qQ of Q is the vector space of all the tangent vectors at q.

Definition A.2 (Tangent bundle)**.** The *tangent bundle* TQ of a manifold Q is the disjoint union of all the tangent spaces of Q, i.e.

$$TQ = \coprod_{q \in Q} T_qQ = \bigcup_{q \in Q} \{q\} \times T_qQ.$$

The tangent bundle TQ is a smooth manifold.

Definition A.3 (Cotangent space)**.** The *cotangent space* of a smooth manifold Q at a point $q \in Q$ is denoted by T_q^*Q and is the dual space of the tangent space T_qQ. An element of T_q^*Q is called a *covector* at q.

This means that an element $p \in T_q^*Q$ is a linear map $p \colon T_qQ \to \mathbb{R}$.

Definition A.4 (Cotangent bundle)**.** The *cotangent bundle* T^*Q of a manifold Q is the disjoint union of the cotangent spaces at all the points of Q,

$$T^*Q = \coprod_{q \in Q} T_q^*Q = \bigcup_{q \in Q} \{q\} \times T_q^*Q.$$

The cotangent bundle T^*Q is a smooth manifold.

Appendix B

Butcher Tableaux for Selected Runge-Kutta Methods

We give here the Butcher tableaux for selected Runge-Kutta methods [2, 10, 14]. Written in parentheses are the number of stages s and the order p of the method.

Gauss ($s = 2$, $p = 4$)

$$\begin{array}{c|cc} \frac{1}{2}-\frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4}-\frac{\sqrt{3}}{6} \\ \frac{1}{2}+\frac{\sqrt{3}}{6} & \frac{1}{4}+\frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Gauss ($s = 3$, $p = 6$)

$$\begin{array}{c|ccc} \frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{5}{36} & \frac{2}{9}-\frac{\sqrt{15}}{15} & \frac{5}{36}-\frac{\sqrt{15}}{30} \\ \frac{1}{2} & \frac{5}{36}+\frac{\sqrt{15}}{24} & \frac{2}{9} & \frac{5}{36}-\frac{\sqrt{15}}{24} \\ \frac{1}{2}+\frac{\sqrt{15}}{10} & \frac{5}{36}+\frac{\sqrt{15}}{30} & \frac{2}{9}+\frac{\sqrt{15}}{15} & \frac{5}{36} \\ \hline & \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \end{array}$$

Radau IA ($s = 2$, $p = 3$)

$$\begin{array}{c|cc} 0 & \frac{1}{4} & -\frac{1}{4} \\ \frac{2}{3} & \frac{1}{4} & \frac{5}{12} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}$$

Radau IA ($s = 3$, $p = 5$)

$$\begin{array}{c|ccc} 0 & \frac{1}{9} & \frac{-1-\sqrt{6}}{18} & \frac{-1+\sqrt{6}}{18} \\ \frac{6-\sqrt{6}}{10} & \frac{1}{9} & \frac{88+7\sqrt{6}}{360} & \frac{88-43\sqrt{6}}{360} \\ \frac{6+\sqrt{6}}{10} & \frac{1}{9} & \frac{88+43\sqrt{6}}{360} & \frac{88-7\sqrt{6}}{360} \\ \hline & \frac{1}{9} & \frac{16+\sqrt{6}}{36} & \frac{16-\sqrt{6}}{36} \end{array}$$

Lobatto IIIA ($s = 2$, $p = 2$)

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Lobatto IIIB ($s = 2$, $p = 2$)

$$\begin{array}{c|cc} 0 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Lobatto IIIC ($s = 2$, $p = 2$)

$$\begin{array}{c|cc} 0 & \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Lobatto IIIC* ($s = 2$, $p = 2$)

0	0	0
1	1	0
	$\frac{1}{2}$	$\frac{1}{2}$

Lobatto IIID ($s = 2$, $p = 2$)

0	$\frac{1}{4}$	$-\frac{1}{4}$
1	$\frac{3}{4}$	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$

Lobatto IIIA ($s = 3$, $p = 4$)

0	0	0	0
$\frac{1}{2}$	$\frac{5}{24}$	$\frac{1}{3}$	$-\frac{1}{24}$
1	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Lobatto IIIB ($s = 3$, $p = 4$)

0	$\frac{1}{6}$	$-\frac{1}{6}$	0
$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	0
1	$\frac{1}{6}$	$\frac{5}{6}$	0
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Lobatto IIIC ($s = 3$, $p = 4$)

0	$\frac{1}{6}$	$-\frac{1}{3}$	$\frac{1}{6}$
$\frac{1}{2}$	$\frac{1}{6}$	$\frac{5}{12}$	$-\frac{1}{12}$
1	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Lobatto IIIC* ($s = 3$, $p = 4$)

0	0	0	0
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0
1	0	1	0
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Lobatto IIID ($s = 3$, $p = 4$)

0	$\frac{1}{12}$	$-\frac{1}{6}$	$\frac{1}{12}$
$\frac{1}{2}$	$\frac{5}{24}$	$\frac{1}{3}$	$-\frac{1}{24}$
1	$\frac{1}{12}$	$\frac{5}{6}$	$\frac{1}{12}$
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

Lobatto IIIA ($s = 4$, $p = 6$)

0	0	0	0	0
$\frac{5-\sqrt{5}}{10}$	$\frac{11+\sqrt{5}}{120}$	$\frac{25-\sqrt{5}}{120}$	$\frac{25-13\sqrt{5}}{120}$	$\frac{-1+\sqrt{5}}{120}$
$\frac{5+\sqrt{5}}{10}$	$\frac{11-\sqrt{5}}{120}$	$\frac{25+13\sqrt{5}}{120}$	$\frac{25+\sqrt{5}}{120}$	$\frac{-1-\sqrt{5}}{120}$
1	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$
	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$

Lobatto IIIB ($s = 4$, $p = 6$)

0	$\frac{1}{12}$	$\frac{-1-\sqrt{5}}{24}$	$\frac{-1+\sqrt{5}}{24}$	0
$\frac{5-\sqrt{5}}{10}$	$\frac{1}{12}$	$\frac{25+\sqrt{5}}{120}$	$\frac{25-13\sqrt{5}}{120}$	0
$\frac{5+\sqrt{5}}{10}$	$\frac{1}{12}$	$\frac{25+13\sqrt{5}}{120}$	$\frac{25-\sqrt{5}}{120}$	0
1	$\frac{1}{12}$	$\frac{11-\sqrt{5}}{24}$	$\frac{11+\sqrt{5}}{24}$	0
	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$