IMA Journal of Numerical Analysis (2000) **20**, 561–599

On the numerical solution of involutive ordinary differential systems

JUKKA TUOMELA[†]
Department of Mathematics, University of Joensuu, PL 111, 80101 Joensuu, Finland

AND

TEIJO ARPONEN[‡]
Institute of Mathematics, Helsinki University of Technology, PL 1100, 02015 TKK, Finland

[Received 2 March 1998 and in revised form 26 June 1999]

We propose a method for the numerical solution of ordinary differential systems. The system is considered geometrically as a submanifold in a jet space. The solutions are then certain integral manifolds that can be computed numerically when the system has been transformed into involutive form.

Keywords: differential algebraic equations; overdetermined differential equations; formal theory of partial differential equations; jet spaces.

1. Introduction

1.1 *Historical background*

We shall present a new approach to the computation of solutions of ordinary differential systems which is based on jet space techniques. Jets were introduced in the early 1950s by Ehresmann (1951) and were almost immediately applied by Spencer, Kuranishi, Sternberg, Goldschmidt, Quillen and others to study arbitrary systems of PDEs (Spencer, 1969; Kuranishi, 1967; Goldschmidt, 1967). Apparently these results did not become well known to a larger audience because Vinogradov (1981) remarks (this is from his review of Pommaret (1978), written in 1981):

> Unfortunately, all these achievements have until now remained on the pages of difficult-to-read periodical literature (including the book of Kumpera & Spencer (1972), which is also difficult to read) and in fact remained inaccessible to those for whom they were written, namely, for specialists in differential equations.

Now there are several monographs on the subject of jets and PDEs (Krasilshchik *et al.*, 1986; Pommaret, 1978, 1983, 1988, 1994; Tarkhanov, 1995); see also the surveys (Dudnikov & Samborski, 1996; Alekseevskij *et al.*, 1991; Spencer, 1969) and the

[†]E-mail: jukka.tuomela@joensuu.fi
[‡]E-mail: teijo.arponen@hut.fi

collection of articles (Lychagin, 1995). Also, jets are used (among other techniques) in the analysis of the symmetries of differential equations (Olver, 1995) and in problems of nonholonomic dynamical systems (Vershik & Gershkovich, 1994). However, none of these works provides easy reading and require at least some background in differential geometry.

Now happily, though not unexpectedly, many of the complications disappear in the ODE case, so we do not have to introduce the whole framework below. In spite of that, we require many notions from differential geometry so this article cannot possibly be self-contained in this sense and we must refer to Spivak (1979), do Carmo (1992) and Saunders (1989) for more details and explanations on these basic matters. However, in Section 2 we have tried to explain some basic ideas with a minimum of technical details in order to motivate the introduction of the heavier machinery which is needed later on.

1.2 *Comparison to related work*

The main motivation of this article came from the theory of differential algebraic equations, but as the reader may remark, this term does not appear in the title, and in fact we shall not need this concept at all. The reason is simple: from our point of view there is no difference between ODEs and DAEs. Hence we shall usually use the term differential system to remind the reader that this covers all the cases. Perhaps this sounds a bit surprising since in the traditional approach to DAEs it is always stressed that ODEs and DAEs are basically different, see the well-known books by Brenan *et al.* (1989); Hairer & Wanner (1991); Hairer *et al.* (1989) and the survey article (März, 1992), where extensive references to the literature are given. Intuitively, this (apparent) paradox can be explained as follows. From the jet point of view the differential systems are certain submanifolds of jet spaces, and consequently one can speak of differential systems without actually writing down any particular equation. One might then say that the classical distinction between ODEs and DAEs is related to the representation of objects rather than to the objects themselves.

Since ODEs and DAEs are not distinguished, it is only natural that the notion of index is not needed either, see the above references for a discussion of this concept from the traditional point of view and (Le Vey, 1998; Seiler, 1999) from the jet point of view. In particular, the index is not related to any numerical difficulty of the problem when using our approach in the numerical solution.

Let us further mention two interesting consequences of our point of view. The first is that explicit methods can be used to solve general differential systems. In other words there is no intrinsic connection between DAEs and stiff systems. The second is that since the definition of the solution is not the usual one, some problems that are singular from the traditional point of view become regular, hence easily computable, in our framework. This aspect is discussed in detail in Tuomela (1997, 1998), so below we shall merely show some examples of this phenomenon.

In traditional methods one sometimes also talks about manifolds and projections to manifolds. Since both words are often used in the sequel, let us note that the manifolds in question are not the same as discussed below, nor are the projections the same.

Rabier, Rheinboldt and Reich have also studied DAEs from the differential geometric point of view, using the tangent bundle instead of jet spaces (Rheinboldt, 1984; Rabier & Rheinboldt, 1991, 1994; Reich, 1991). Some theorems and computations are seen to

be (equivalent to) special cases of the more general results of Goldschmidt (1967), see Korvola (1997) for details.

Recently there has been much interest in applying geometric ideas to the numerical solution of differential systems. Iserles (1996, 1997) discusses, in general, how qualitative features of the solutions should be taken into account when designing numerical methods. For example, nowadays Hamiltonian problems are usually treated with symplectic methods which preserve the underlying symplectic structure of the problem, see Sanz Serna & Calvo (1994) and the references therein. Symplecticity and jets appear also in Rangarajan (1996), but the use of jets there is quite different than in the present paper. In Seiler & Tucker (1995) and Seiler (1995) jets are used to study constrained dynamics, but the main motivation and results are not in the numerical analysis. There are also many papers dealing with flows on Lie groups and in general using Lie groups and Lie algebras in numerical methods. Since these works are not directly related to the present paper (except for the fact that differential geometric ideas play an essential role in both cases), we will simply refer the reader to the following articles for further details: (Crouch & Grossman, 1993; Iserles & Nørsett, 1999; Munthe-Kaas, 1995, 1999; Munthe-Kaas & Zanna, 1997; Zanna, 1998; Zanna & Munthe-Kaas, 1997).

1.3 *Outline of the article*

In Section 2 some motivating examples are given. By restricting our attention to simple equations, it is possible to explain some basic ideas relying on the geometrical intuition of the three-dimensional space. In Section 3 the jets are introduced and the important concept of involutivity is discussed. In Section 4 there are some computations illustrating the notions introduced in previous sections.

In Section 5 we use Riemannian geometry to analyse the local error of some one-step methods. The analysis shows that the orders of these methods in our context are the same as their classical orders. In Section 6 the actual numerical implementation is presented. It is seen that the subproblems arising in our algorithm can be treated with fairly standard techniques. In Section 7 the numerical examples are given. We have chosen some representative problems which have appeared frequently in the literature. Our methods work reliably, and the results show in particular that one can use explicit methods for these kind of problem, although in the literature only implicit methods are considered. We conclude with Section 8 where some suggestions for future work are indicated.

Essential parts of this article are based on the reports by Tuomela (1996) and Arponen & Tuomela (1996). That jet spaces are useful for analysing differential-algebraic equations was first suggested in Le Vey (1994) and Piirilä & Tuomela (1993) (independently).

2. Differential equation as a surface

Let us start by explaining the basic ideas in a simple setting where we can visualize the situation in three-dimensional space before formulating the general framework. A similar situation was considered in Arnold (1983) and Dara (1975) where the emphasis is on the analysis of singularities. We will also say a few words about singularities in the examples, but do not insist on this aspect of the problem. While discussing the examples we also

introduce some standard concepts of differential geometry. All maps and manifolds are assumed to be smooth, i.e. infinitely differentiable.

2.1 *Basic machinery*

Let us consider a single first-order differential equation

$$f(x, y, y_1) = 0 \tag{2.1}$$

where f is a smooth function and y_1 is the derivative of y.[†] Now, forgetting for the moment about derivatives, we can interprete (x, y, y_1) as coordinates of $\mathbb{R}^3$. Hence the zero set of (2.1) is a subset of $\mathbb{R}^3$, denoted by $\mathcal{R}_1 = f^{-1}(0) \subset \mathbb{R}^3$. If zero is a regular value of f, i.e. if the gradient of f never vanishes on $\mathcal{R}_1$, then $\mathcal{R}_1$ is a smooth two-dimensional manifold.

Now what are the solutions of (2.1)? Classically the solutions are defined as (smooth) curves $c : \mathbb{R} \to \mathbb{R}$ such that (2.1) is identically satisfied. Here we shall take another, more geometrical as well as more general, point of view. Having defined $\mathcal{R}_1$ one could perhaps consider defining solutions as certain curves $c : \mathbb{R} \to \mathcal{R}_1$. However, it is only the image of the curve which is important, its parametrization is irrelevant. Hence the solutions should be defined directly as certain one-dimensional submanifolds of $\mathcal{R}_1$ without any reference to parametrizations. So let S be a smooth one-dimensional submanifold of $\mathcal{R}_1$. When would it be meaningful to say that S is a solution of (2.1)?

To proceed let us introduce some terminology. If $p \in \mathcal{R}_1$, the tangent plane of $\mathcal{R}_1$ at p is denoted by $(T\mathcal{R}_1)_p$. Similarly TS_p is the line tangent to S at p, and it is evident that $TS_p \subset (T\mathcal{R}_1)_p$. Our next task is to add to this framework some extra structure in such a way that we can say that $\mathcal{R}_1$ is a differential equation. Consider $\mathbb{R}^3$ with coordinates (x, y, y_1); if y_1 is really the derivative of y with respect to x, then in infinitesimal notation we should have $\mathrm{d}y - y_1\,\mathrm{d}x = 0$. To make sense of this we must define what is meant by $\mathrm{d}y$ and $\mathrm{d}x$. Using Spivak's words (Spivak, 1979, vol. 1, p 153) symbols like $\mathrm{d}x$

> ... metamorphosed into functions, and it became clear that they must be distinguished from tangent vectors. Once this realization came, it was only a matter of making new definitions, which preserved the *old* notation, and waiting for everybody to catch up.

So $\mathrm{d}x$, $\mathrm{d}y$ and $\mathrm{d}y_1$ are linear functions on the tangent space $(T\mathbb{R}^3)_p$, hence elements of the dual space of $(T\mathbb{R}^3)_p$. This dual space is called the cotangent space and is denoted $(T^*\mathbb{R}^3)_p$. Let us then define $\alpha = \mathrm{d}y - y_1\,\mathrm{d}x$; at each point of $\mathbb{R}^3$, α is then a linear function on $(T\mathbb{R}^3)_p$. Now we can define a subspace of $(T\mathbb{R}^3)_p$ as follows

$$\mathcal{C}_p = \{v \in (T\mathbb{R}^3)_p \mid \alpha(v) = 0\}$$

All these $\mathcal{C}_p$ together are called a *contact distribution*, denoted $\mathcal{C}$. In general let us define:

DEFINITION 2.1 A distribution on a manifold M is a map which associates a certain subspace of the tangent space TM_p to each point of $p \in M$.

[†]Later y_1 is called a jet coordinate.

In the present case $\mathcal{C}$ is then a two-dimensional distribution on $\mathbb{R}^3$. One may think of distributions as generalizations of vector fields. Now a vector field may be interpreted as a differential equation, and hence the integral curves of the vector field are said to be solutions of the corresponding differential equation. A generalization of this idea leads to the notion of an integral manifold.

DEFINITION 2.2 Let S be a submanifold of M and $\mathcal{D}$ a distribution on M. S is an integral manifold of $\mathcal{D}$ if $TS_p = \mathcal{D}_p$ at each $p \in S$.[†]

Let us go back to our differential equation (2.1). Since $(T\mathcal{R}_1)_p$ can also be interpreted as a subspace of $(T\mathbb{R}^3)_p$, we can define an intersection $\mathcal{D}_p = (T\mathcal{R}_1)_p \cap \mathcal{C}_p$. Hence $\mathcal{D}$ is in general one dimensional. Of course at some points tangent and contact planes may coincide; we will discuss this event and its consequences in the examples below. Ignoring these exceptional points for the moment leads to

DEFINITION 2.3 Solutions of (2.1) are integral manifolds of $\mathcal{D}$.[‡]

Since one-dimensional distributions always have integral manifolds, see (Spivak, 1979, vol. 1, p 245), we get

THEOREM 2.1 Given $p \in \mathcal{R}_1$ such that $\mathcal{D}$ is one-dimensional in a neighbourhood of p, there exists a solution S of (2.1) such that $p \in S$.

Note that this kind of existence theorem is not valid for many dimensional distributions. For example, it is readily checked that the contact distribution does not admit any integral manifolds, even locally. Incidentally, this implies that tangent and contact planes cannot coincide in an open set of $\mathcal{R}_1$. Since we are interested only in one-dimensional distributions, i.e. ordinary differential equations, we will not consider these matters further and refer the reader to Spivak (1979, vol. 1, chapter 6) for more information.

2.2 *Comparison to classical solutions*

Consider again the problem (2.1) and let $p \in \mathcal{R}_1$ be a regular point, i.e. $\partial f/\partial y_1 \neq 0$ at p. Then by the implicit function theorem there is a map $\varphi : \mathbb{R}^2 \to \mathbb{R}$ such that in a neighbourhood of p, $\mathcal{R}_1$ can be represented by the equation $y_1 + \varphi(x, y) = 0$. The solutions of this equation can be defined in the usual way and one thinks of them either as curves $\mathbb{R} \to \mathbb{R}$ or as (one-dimensional) submanifolds of $\mathbb{R}^2$. Let us take the latter point of view. Let us define a projection $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ by $(x, y, y_1) \mapsto (x, y)$ and let S be a solution of (2.1) in the sense of Definition 2.3. Now if $p \in S$ is regular, then one can show that $\pi(S)$ is the classical solution in some neighbourhood of $\pi(p)$.

Hence around regular points there is a bijective correspondence between solutions in the sense of Definition 2.3 and classical solutions. However, because the implicit function theorem is not constructive, the reduction to the standard form is not in general possible in practice. Note also that Definition 2.3 provides more smooth solutions than the classical

[†]Sometimes another terminology is used, for example in Dara (1975). If $\mathcal{D}$ has constant dimension on M, then it determines a foliation of M and each integral manifold is called a leaf of this foliation.

[‡]In Dara (1975) this way of defining solutions is called 'méthode de Lie'.

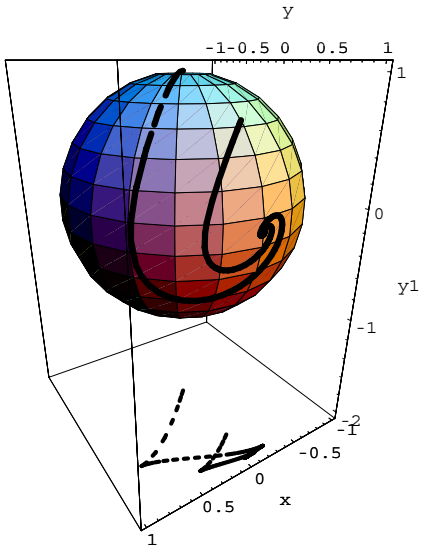

FIG. 1. Some solution curves of the equation (2.2) and their projections.

setting since tangent and contact planes need not (and in general do not) coincide even if $\partial f/\partial y_1 = 0$. We shall see shortly examples of this phenomenon.

REMARK 2.1 This property of having more smooth solutions than in the classical case is quite interesting from the numerical point of view. Indeed, it is possible to use general methods for a larger class of problems than the classical framework allows.

2.3 *Examples*

2.3.1 *Sphere.* Consider the problem

$$f(x, y, y_1) = (y_1)^2 + y^2 + x^2 - 1 = 0. \tag{2.2}$$

Hence $\mathcal{R}_1 = f^{-1}(0)$ is the unit sphere and at each point $p \in \mathcal{R}_1$, $\mathcal{D}_p$ is given by the nullspace of

$$A = \begin{pmatrix} -y_1 & 1 & 0 \\ x & y & y_1 \end{pmatrix}.$$

One observes that $\dim(N(A)) = 1$, except at $(0, \pm 1, 0)$. So only at these two points tangent and contact planes coincide; in the classical sense all the points of the equator are singular. In Fig. 1 we show some solution curves of (2.2) as well as their projections. When the solution crosses the equator, there is a cusp in the projected solution, which in convenient coordinates can be represented by the equation $u^2 = v^3$. The singularities at exceptional points $(0, \pm 1, 0)$ are called folded focuses (Arnold & Ilyashenko, 1988, p 34).

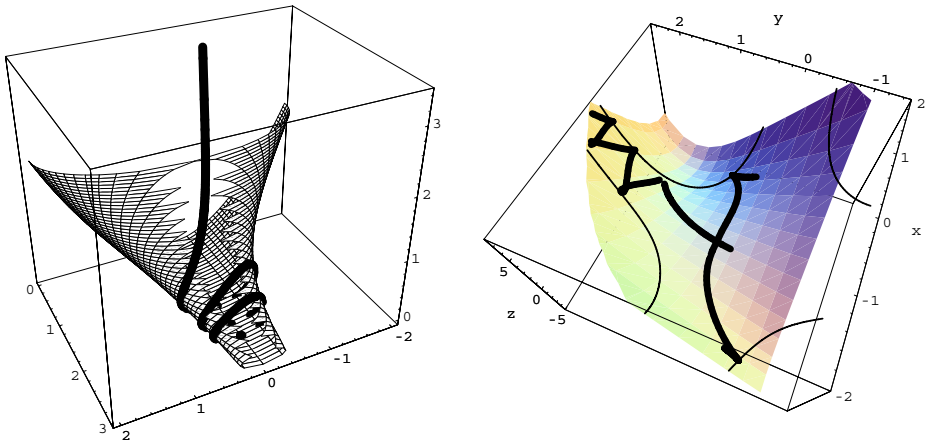

FIG. 2. A solution of equation (2.4) (left) and some asymptotic curves on surface (2.3) (right). Thin lines represent the sets of parabolic points.

2.3.2 *Asymptotic curves.* Let us consider the following problem of classical differential geometry. Let $N \subset \mathbb{R}^3$ be a surface defined by the equation

$$\tfrac{1}{4}(x^4 + y^4) - 3xy - z = 0. \tag{2.3}$$

We want to compute some asymptotic curves on this surface. There are several characterizations of these curves, see Spivak (1979, vol. 3) or any book on classical differential geometry. The one which is convenient for our purposes is the following: the tangent of an asymptotic curve makes the second fundamental form vanish. Asymptotic curves exist only in those parts of the surface where Gaussian curvature is negative or zero, in the present case this occurs in the region where $x^2y^2 \leqslant 1$. The curves can be computed by solving the differential equation

$$f(x, y, y_1) = y^2(y_1)^2 - 2y_1 + x^2 = 0. \tag{2.4}$$

The distribution is one-dimensional everywhere on $\mathcal{R}_1 = f^{-1}(0)$, except at $(1, -1, 1)$ and $(-1, 1, 1)$. Now the problem can be solved by computing the solution in $\mathcal{R}_1$, projecting it to $\mathbb{R}^2$ and finally lifting it to N. Note that the solutions of (2.4) are smooth in the sense of Definition 2.3, although the corresponding asymptotic curves on surface N have familiar cusps at $x^2y^2 = 1$, see Fig. 2. The singularities at exceptional points $(1, -1, 1)$ and $(-1, 1, 1)$ are called folded saddles, see Arnold & Ilyashenko (1988, p. 34) and Fig. 3 where there are some solutions and their projections around $(1, -1, 1)$.

2.3.3 *Distributions and their singularities.* Distributions are perhaps less intuitive and less familiar objects than vector fields, so let us see an example where distributions come directly from the problem. In fact the previous example is such a case. The condition defining the asymptotic directions clearly defines a subspace of the tangent space, hence a distribution. In this way we have arrived directly at distributions without passing by

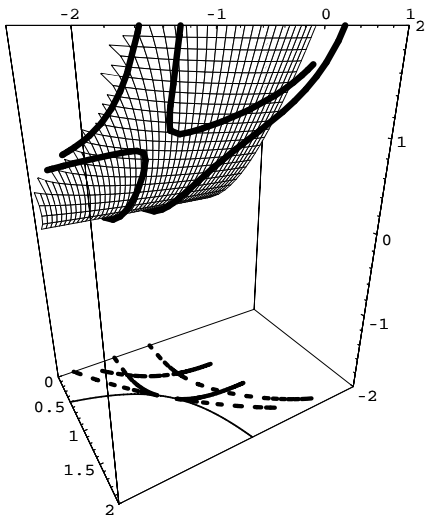

FIG. 3. Some solution curves of the equation (2.4) and their projections.

differential equations. In the previous example we have not used this distribution. Also the condition defining the lines of curvature of a surface leads directly to distributions. In this case the distribution is determined by the eigenspaces of the differential of the Gauss map.

Let us then show that singularities of distributions are more general than singularities, i.e. zeros, of vector fields. Consider any smooth surface $N \subset \mathbb{R}^3$ and let us suppose that distribution $\mathcal{D}$ is one-dimensional except at some isolated points which are called singular points of $\mathcal{D}$. Recall that around a regular point p there is a vector field V such that $0 \neq V_a \in \mathcal{D}_a$ for all a in some neighbourhood of p and conversely every such V spans a certain distribution. However, around a singular point such a vector field may not exist. Let us see an example of this phenomenon. Consider the problem

$$f(x, y, y_1) = x(y_1)^2 - 2yy_1 - x = 0. \tag{2.5}$$

It is easily checked that this is a regular problem, and in fact the classical solutions are seen to be a family of parabolas, see Fig. 4. Now clearly there is no vector field in the neighbourhood of the origin whose integral curves were the parabolas. Incidentally, the pattern of parabolas is the same as the pattern of the lines of curvature around an umbilic on an ellipsoid, see Spivak (1979, vol. 3, p 288). In other words the index of the singularity is the same, namely $1/2$ in both cases, see Spivak (1979, vol. 3, p 324). Note also that if one transforms the above problem into a classical problem using square roots, the correct solution is obtained only if one jumps from one branch of the square root to the other when crossing $x = 0$.

3. Differential systems and jet spaces

Lychagin (1995, p ix) describes the importance of jets in the analysis of differential equations as follows.

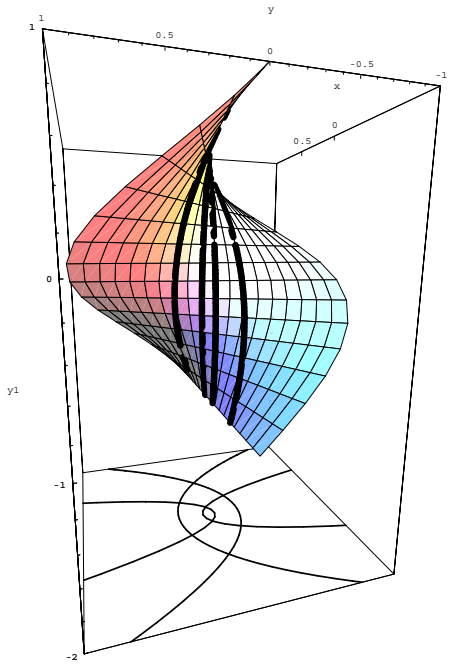

FIG. 4. Some solution curves of the equation (2.5) and their projections.

> Thus, whereas connections between linear differential equations and differential geometry were few and far between, as were the connections between analytic geometry (in the elementary sense of the term) and geometry in the sense of Felix Klein, in contrast the theory of nonlinear differential equations is clearly a geometric theory, based on the special geometry of jet spaces.

Hence to proceed we must introduce more ideas from differential geometry and, in particular, jets. For more details on jet geometry we refer to Saunders (1989). In all that follows various maps are required to be defined only in some convenient open set of the relevant manifold. For notational simplicity this set is not indicated. To further simplify the discussion we shall not define bundles in full generality. The version presented is, however, locally 'true', i.e. any bundle can locally be represented as described below.

3.1 *Jet bundles*

Let us consider two maps $y, z : \mathbb{R}^s \to \mathbb{R}^n$, let $\nu \in \mathbb{N}^s$ be a multi-index and $|\nu| = \nu_1 + \cdots + \nu_s$.

DEFINITION 3.1 The maps y and z are q-equivalent at x, if

$$\frac{\partial^{|\nu|} y}{\partial x^\nu}(x) = \frac{\partial^{|\nu|} z}{\partial x^\nu}(x)$$

for all $0 \leqslant |\nu| \leqslant q$.

The equivalence thus defined is clearly an equivalence relation so we can further define

DEFINITION 3.2 The q-jet of y at x, denoted by $j^q_x(y)$, is the equivalence class of the above equivalence relation. The q-jet of y, denoted by $j^q(y)$, is the map $x \mapsto j^q_x(y)$.

Intuitively the q-jet of y at x is its Taylor expansion up to order q. Since all these considerations are local, the previous definitions also make sense when $\mathbb{R}^s$ and $\mathbb{R}^n$ are replaced by manifolds. However, in the jet theory one usually prefers to talk about sections rather than maps. So let us introduce

DEFINITION 3.3 A (trivial) bundle is a triple $(\mathcal{E}, \pi, \mathcal{B})$ where $\mathcal{E} = \mathbb{R}^s \times \mathbb{R}^n$, $\mathcal{B} = \mathbb{R}^s$, $\pi : \mathcal{E} \to \mathcal{B}$ is a map defined by $\pi(x, y) = x$. $\mathcal{E}$ is called the total space, $\mathcal{B}$ the base space and π the projection. For each $p \in \mathcal{B}$, the set $\pi^{-1}(p)$ is called the fibre over p.

We shall usually refer to a bundle by the total space, although this might be confusing in some situations; indeed in jet theory there are frequently situations where the same total space is considered with different base spaces and projections. However, in our context this kind of confusion is not likely to occur.

DEFINITION 3.4 A section of a bundle $(\mathcal{E}, \pi, \mathcal{B})$ is a map $y : \mathcal{B} \to \mathcal{E}$ such that $\pi \circ y =$ identity.

Note that in our simplified context where $\mathcal{E} = \mathbb{R}^s \times \mathbb{R}^n$ a section of $\mathcal{E}$ is simply the graph of a map $\mathbb{R}^s \to \mathbb{R}^n$. Hence the definition of jets for maps carries immediately over to sections. Putting all this information together gives jet bundles.

DEFINITION 3.5 Let $(\mathcal{E}, \pi, \mathcal{B})$ be a bundle. The set of all q-jets of its sections is a bundle $(J_q(\mathcal{E}), \pi^q, \mathcal{B})$, the qth jet bundle of $(\mathcal{E}, \pi, \mathcal{B})$.

For example if $\mathcal{E} = \mathbb{R} \times \mathbb{R}$, then $\mathbb{R}^3$ in the previous section is denoted by $J_1(\mathcal{E})$. The projection π^1 is defined by $(x, y, y_1) \mapsto x$ and if y is a section of $\mathcal{E}$, then $j^1(y)$ is a section of $J_1(\mathcal{E})$ defined by $x \mapsto (x, y(x), y'(x))$. Note that not every section of $J_1(\mathcal{E})$ is a 1-jet of a section of $\mathcal{E}$. We are now ready to define a differential equation.

DEFINITION 3.6 Let $\mathcal{E}$ be a bundle. A (partial) differential system (or equation) of order q on $\mathcal{E}$ is a submanifold $\mathcal{R}_q$ of $J_q(\mathcal{E})$.

Since we are working with ODEs, the jet bundle $J_q(\mathcal{E})$ can be identified with Cartesian product $\mathbb{R} \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ where $\mathbb{R}^n$ appears $q + 1$ times. Hence we can write $J_q(\mathcal{E}) \simeq \mathbb{R}^{(q+1)n+1}$ and denote its coordinates by $(x, y, y_1, \ldots, y_q)$; these are called *jet coordinates*.

However, before restricting our attention to the ODE case, we add a few comments about the formal theory of PDEs and refer for more information to (Spencer, 1969; Goldschmidt, 1967; Krasilshchik *et al.*, 1986; Kuranishi, 1967; Pommaret, 1978, 1983, 1988, 1994; Tarkhanov, 1995; Dudnikov & Samborski, 1996; Alekseevskij *et al.*, 1991). The word formal is used for the following reason. When Riquier (1910) and others started studying fairly arbitrary systems of PDEs at the end of the 19th century, they had in mind a generalization of the Cauchy–Kovalevskaia theorem and this involved two steps: the first being to construct the solution as a *formal* power series. Now Riquier realized that given an arbitrary system, it was not possible to obtain the formal power series without first transforming the initial system to a 'canonical' form, or, in modern terms, a formally

integrable or involutive form. The second step is then to prove the convergence of the power series in the analytical case, or more generally that the solution exists in a stronger sense than merely as a formal power series. Note that these two steps are independent of each other.

Now, is it always possible to find the involutive form? Probably already in the beginning people working in this field thought that this must be the case for a 'generic' system, and the result was proved in full generality by Goldschmidt (1967), based on the work of Spencer. Stating precisely this theorem would require too much preparation, so we shall simply try to explain some ideas that are relevant from the point of view of the present article. Anyway let us formulate at least a slogan:

> Any system of PDEs can be transformed into an involutive form. Moreover this transformation is constructive.

In the ODE context, involutiveness says that if we have a system of differential equations of order at most q, then we cannot obtain new independent equations of order q by differentiating the given equations and eliminating higher-order derivatives. Later on we shall discuss the implications of these matters to numerical computations.

3.2 *Ordinary differential systems*

In Definition 3.6 we have defined a differential system without actually mentioning any equations. Now it is finally time to introduce the equations. Having identified $J_q(\mathcal{E})$ and $\mathbb{R}^{(q+1)n+1}$ we can define a submanifold of $J_q(\mathcal{E})$ as a zero set of some map $f : \mathbb{R}^{(q+1)n+1} \to \mathbb{R}^k$:

$$f(x, y, y_1, \ldots, y_q) = 0. \tag{3.1}$$

Hence we can set $\mathcal{R}_q = f^{-1}(0) \subset J_q(\mathcal{E})$. Now it is relatively straightforward to extend the framework of Section 2 to the general case. We denote the tangent (resp. cotangent) bundle of a manifold M by TM (resp. T^*M), and the tangent (resp. cotangent) space at $p \in M$ by TM_p (resp. T^*M_p). The sections of TM are vector fields and the sections of T^*M are one-forms. In the introductory examples α was a section of $T^*J_1(\mathbb{R} \times \mathbb{R})$ and now we must formulate this idea in the general case.

Let us then consider the system (3.1) and let us define the one-forms on $J_q(\mathcal{E})$ by

$$\alpha^i_j = \mathrm{d}y^i_{j-1} - y^i_j \,\mathrm{d}x \qquad i = 1, \ldots, n \qquad j = 1, \ldots, q. \tag{3.2}$$

With these forms one can formulate the appropriate generalization of the contact plane and contact distribution. Let $p \in J_q(\mathcal{E})$ and define

$$\mathcal{C}_p = \left\{ v_p \in (TJ_q(\mathcal{E}))_p \middle| \alpha^i_j(v_p) = 0 \right\}.$$

This defines a distribution $\mathcal{C}$ on $J_q(\mathcal{E})$, which is called a Cartan distribution, and it is easily checked that $\dim(\mathcal{C}) = n + 1$.[†] Let us see that this distribution captures the 'infinitesimal'

[†] The Cartan distribution defines a contact structure on $J_q(\mathcal{E})$ only in the case $n = q = 1$.

idea expressed in the notation of the one-forms. Let y (resp. z) be a section of $\mathcal{E}$ (resp. $J_q(\mathcal{E})$). Identifying sections and their images we may consider $j^q(y)$ and z as submanifolds of $J_q(\mathcal{E})$. Then we have the following lemma whose proof we omit.

LEMMA 3.1 For all $p \in j^q(y)$, $(Tj^q(y))_p \subset \mathcal{C}_p$ and conversely if for all $p \in z$, $(Tz)_p \subset \mathcal{C}_p$, then $z = j^q(y)$ for some y.

Let $\mathcal{R}_q = f^{-1}(0) \subset J_q(\mathcal{E})$ where f is given in (3.1) and, supposing that zero is a regular value of f, $\mathcal{R}_q$ is a smooth manifold. Now as in the simple examples in the beginning we can define a distribution $\mathcal{D}$ at each point of $p \in \mathcal{R}_q$ by

$$\mathcal{D}_p = (T\mathcal{R}_q)_p \cap \mathcal{C}_p. \tag{3.3}$$

Hence we can now define the solutions in a similar fashion to the scalar case.

DEFINITION 3.7 Supposing that $\mathcal{R}_q$ is involutive and $\mathcal{D}$ is one-dimensional, the solutions of (3.1) are integral manifolds of $\mathcal{D}$.

Now Theorem 2.1 generalizes as such to the present situation, so we conclude that there always exists a unique solution through $p \in \mathcal{R}_q$, if $\mathcal{D}$ is one-dimensional in some neighbourhood of p.

REMARK 3.1 Is $\mathcal{D}$ really one-dimensional? The dimensions of $\mathcal{C}_p$, $(T\mathcal{R}_q)_p$ and $(TJ_q(\mathcal{E}))_p$ are $n+1$, $qn+n-k+1$ and $qn+n+1$ which implies that $\mathcal{D}_p$ is at least $n-k+1$ dimensional. Consequently for $k < n$ (i.e. underdetermined systems) $\mathcal{D}$ cannot be one-dimensional. Note that in this case the solutions can still be defined with the help of sections, although evidently there is no unique solution through a given point. In some problems it may be interesting to analyse the space of solutions of underdetermined systems, and jet spaces are useful in this kind of analysis. However, such considerations are rather far removed from numerical analysis, so from now on we will always assume that the systems treated are not underdetermined.

For $k = n$ one would expect that $\mathcal{D}$ is 'generically' one-dimensional, and in fact this is always the case if f is in the standard form

$$f(x, y, y_1, \ldots, y_q) = y_q + \tilde{f}(x, y, y_1, \ldots, y_{q-1}).$$

More generally, the transversal intersection of $(T\mathcal{R}_q)_p$ and the Cartan distribution is one-dimensional.

Finally if $k > n$, then for f taken 'at random' one would expect that $\mathcal{D}$ is zero-dimensional. However, systems arising in practice have additional structure which 'allows' $\mathcal{D}$ to be one-dimensional.

REMARK 3.2 Why should the system be involutive? Consider the following systems in $J_1(\mathbb{R} \times \mathbb{R}^2) \simeq \mathbb{R}^5$

$$\mathcal{R}_1 : \begin{cases} y^1 + 1 = 0 \\ y_1^2 + 1 = 0 \end{cases} \qquad \mathcal{R}_1^{(1)} : \begin{cases} y^1 + 1 = 0 \\ y_1^1 = 0 \\ y_1^2 + 1 = 0. \end{cases}$$

$\mathcal{R}_1^{(1)}$ is the involutive form of $\mathcal{R}_1$. Note that the solution sets, in the classical sense, are the same for $\mathcal{R}_1$ and $\mathcal{R}_1^{(1)}$. Now the distribution on $\mathcal{R}_1$ is one-dimensional, *except* when $y_1^1 = 0$, and hence if the initial point is *not* on $\mathcal{R}_1^{(1)}$ we get a well defined integral manifold. However, this is not really what we want since the equation $y_1^1 = 0$ should also be satisfied. For the involutive system $\mathcal{R}_1^{(1)}$ one verifies that the corresponding distribution is everywhere one-dimensional. This is typical: noninvolutive systems can contain open sets where there are no solutions at all, and in spite of that the distribution can be one-dimensional in such a set. On the other hand, the dimension of the distribution of a noninvolutive system can be greater than one in an open set where there are unique solutions.

REMARK 3.3 Note that it is unnecessary to transform the system into a first-order system before the numerical solution. In fact it is better *not* to transform since the dimension of the ambient space is bigger for the first-order system. Let $q > 1$, $\mathcal{R}_q \subset J_q(\mathcal{E})$ and let the corresponding first-order system be $\mathcal{R}_1 \subset J_1(\mathbb{R} \times \mathbb{R}^{qn})$; then $\dim(J_1(\mathbb{R} \times \mathbb{R}^{qn})) = 2qn + 1 > (q + 1)n + 1 = \dim(J_q(\mathcal{E}))$. Of course $\dim(\mathcal{R}_q) = \dim(\mathcal{R}_1)$. For the same reason it does not pay to transform a nonautonomous system into autonomous form.

REMARK 3.4 It is interesting that when one computes the solution, one obtains automatically approximations to derivatives up to order q.

REMARK 3.5 Note that there is no natural way to extend the distribution $\mathcal{D}$ outside $\mathcal{R}_q$, even in the neighbourhood of $\mathcal{R}_q$. Such an extension would mean relaxing some of the 'tangent conditions' (the curve would not be tangent to $\mathcal{R}_q$) and/or 'Cartan conditions' (the jet coordinates would not correspond to derivatives). Evidently choosing this relaxation is quite arbitrary. Of course in some situations there can be an easily constructed extension, which might prove useful in the computations.

REMARK 3.6 Perhaps the reader feels that at this point one could make a distinction between ODEs and DAEs. A system $\mathcal{R}_q$ would be an ODE (strictly speaking a not underdetermined ODE) if its dimension is maximal, i.e. equal to $J_{q-1}(\mathcal{E})$, and the rest would be DAEs. However, this would not be intrinsic because the same manifold can be embedded in different jet spaces, and it would not be very informative to say that a system is sometimes an ODE and sometimes a DAE. In fact in the resolution of singularities there naturally arises a situation where one wants to consider the manifold in a higher-order jet space than initially given (Tuomela, 1998). Anyway we are mainly interested in what happens on the manifold itself and so it is hard to see how this kind of distinction which is based on the dimension of the *ambient* space could convey any useful information. Of course the term DAE can remain useful as a label for certain situations, but there is really no need to give a mathematical definition.

3.3 *Preliminary conclusions*

Having defined our basic framework we pause to make a few comments before moving on to the numerical part of the article. Since the basic objects of study, i.e. differential equations and solutions of differential equations, have been redefined, it is perhaps not so surprising that this approach opens up new points of view for many aspects of analysis of differential systems. First of all, on the purely conceptual level, it is quite nice to be able

to discuss such a large class of systems in a unified way; recall that no special property is required of the map f in (3.1).

The important concept which emerges from this framework is involutivity, and because of our slogan one might, from the theoretical point of view, restrict one's attention to involutive systems. However, in practice the systems are usually not involutive initially and hence they must first be transformed to an involutive form before actual numerical computations. Now this transformation usually involves some symbolic manipulation of the given system, and consequently can be rather time consuming. Of course the algorithms for symbolic manipulation are progressing all the time, so perhaps this is not such a severe problem. Moreover, it is not quite clear how much symbolic computation is really required: in some situations a mixture of numerical and symbolic techniques could be sufficient. Finally, situations arising in practice have a lot of structure so although some algorithms may be prohibitively slow, in general they could perhaps work quite well when this structure is taken into account. It is impossible to go into the details of the various algorithms here so we refer the interested reader to Boulier (1995); Carrà Ferro (1987); Macutan & Thomas (1998); Mansfield (1991); Thomas (1997) and the references therein.

However, the transformation to involutive form should also be seen as an analysis (of singularities) of the system. Recall that in our slogan we have suppressed the technical details which hide some sort of regularity assumptions. One may think that in general these conditions are satisfied in some open set of the relevant manifold. Now the set where these conditions fail indicates some sort of singularity of the system: the manifold is perhaps not smooth everywhere, or perhaps it can be factored to several subsystems. Obviously it is impossible to analyse the situation using only the original, i.e. non-involutive, system. Also it is quite interesting to locate the sets where the distribution fails to be one-dimensional, because this gives essential information about the solutions. Again these sets cannot be known in general, if the system is not involutive.

This leads to yet another aspect, namely genericity. When analysing singularities of the solutions for example, it is well known that certain singularities are typical while others disappear when the system is perturbed a little (Arnold *et al.*, 1985). Of course this kind of analysis is important in applications, where usually models and the coefficients are only known approximatively. Without pursuing this matter further let us just mention one interesting fact. Now the points where the distribution fails to be one-dimensional are evidently the generalizations of the singularities (i.e. zeros) of vector fields. It is known that generically singularities of vector fields appear only at isolated points. This is not true for distributions. In fact one expects that generically the singularities of distributions appear in the sets of codimension two, see Arnold *et al.* (1985); Tuomela (1997) and Tuomela (1998) for more information.

In applications one sometimes encounters situations where f is not smooth and the transformation to involutive form may fail because f cannot be differentiated a sufficient number of times. In this situation it is not clear how useful this geometric framework is in general. However, doing the transformation *as if* f were smooth might still reveal some important information about the nature of singularities of the solutions.

Finally there is the numerical aspect which is the main topic of the present article. Below we will show how to use jet spaces in numerical computations. Here also the new point of view leads to different conclusions than the classical approach. This will be discussed in more detail below.

4. Examples and simplifications

Since a lot of problems arising in practice are of the type of mechanical systems with constraints it seems worthwhile to discuss explictly some problems in this class. Moreover, the computations that follow are also useful as illustrations of the concepts and definitions introduced above.

4.1 *Pendulum*

Consider the simple pendulum which in classical notation is given by

$$\begin{cases} x'' + \lambda x = 0 \\ y'' + \lambda y + 1 = 0 \\ x^2 + y^2 - 1 = 0. \end{cases} \tag{4.1}$$

Here x and y are the Cartesian coordinates and λ is the tension in the string (or Lagrange multiplier). Let us introduce the jet coordinates:

$$x \longleftrightarrow y^1 \qquad y \longleftrightarrow y^2 \qquad \lambda \longleftrightarrow y^3.$$

The system is not involutive, but differentiating and eliminating four times produces the following involutive form:[†]

$$\begin{cases} y_2^1 + y^1 y^3 = 0 \\ y_2^2 + y^2 y^3 + 1 = 0 \\ y_2^3 + 3y_2^2 = 0 \\ y^1 y_1^1 + y^2 y_1^2 = 0 \\ (y_1^1)^2 + (y_1^2)^2 - y^2 - y^3 = 0 \\ 3y_1^2 + y_1^3 = 0 \\ (y^1)^2 + (y^2)^2 - 1 = 0. \end{cases} \tag{4.2}$$

Let $\mathcal{E} = \mathbb{R} \times \mathbb{R}^3$ and denote the above equations by $f(x, y, y_1, y_2) = 0$ and $\mathcal{R}_2 = f^{-1}(0) \subset J_2(\mathcal{E})$. Evidently $\dim(\mathcal{R}_2) = 3$. Note that the original system and the involutive system have the same set of solutions in the classical sense, but Definition 3.7 makes sense only for the involutive form. Now the relevant distribution can be computed from

[†]We have to differentiate four times, although the system is said to be of (differential) index 3.

the nullspace on the following matrix

$$A = \begin{pmatrix} -y_1^1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -y_1^2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -y_1^3 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -y_2^1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -y_2^2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -y_2^3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & y^3 & 0 & y^1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & y^3 & y^2 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & y_1^1 & y_1^2 & 0 & y^1 & y^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 2y_1^1 & 2y_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & y^1 & y^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.3}$$

It is easily checked that $\dim(N(A)) = 1$ everywhere on $\mathcal{R}_2$, though A is usually of full rank outside $\mathcal{R}_2$. Hence the problem is regular and through every initial point there exists a unique solution. Of course in this simple problem one sees immediately that one could drop the last four rows of the matrix in actual computations. However, in general one must use the full matrix and hence treat an overdetermined system of linear equations.

The problem is quite simple and one could use the above formulation in the numerical computations. However, the number of equations can be reduced, or more geometrically, the manifold and the distribution can be projected to a lower-dimensional ambient space. Since this projection can also be done for quite a large class of problems (see below), let us see how this can be achieved.

Recall that if one has a system of the form $y_1 + \varphi(x, y) = 0$, this always induces a distribution or vector field on $\mathcal{E}$, because in this case $\mathcal{R}_1$ is simply the graph of $-\varphi$. Hence the projection $(x, y, y_1) \mapsto (x, y)$ induces a diffeomorphism between $\mathcal{R}_1$ and $\mathcal{E}$, and the differential of the projection transports the distribution from $\mathcal{R}_1$ to $\mathcal{E}$. Note that the graph cannot be vertical, so that the distribution remains one dimensional.

Now looking at the system (4.2) we see that it is 'graph like' with respect to y_2, so in this case we could project the system from $J_2(\mathcal{E})$ to $J_1(\mathcal{E})$ with the natural map π_1^2 : $(x, y, y_1, y_2) \mapsto (x, y, y_1)$. This gives

$$\begin{cases} y^1 y_1^1 + y^2 y_1^2 = 0 \\ (y_1^1)^2 + (y_1^2)^2 - y^2 - y^3 = 0 \\ 3y_1^2 + y_1^3 = 0 \\ (y^1)^2 + (y^2)^2 - 1 = 0 \end{cases} \tag{4.4}$$

and the distribution is given by the nullspace of

$$\tilde{A} = \begin{pmatrix} -y_1^1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -y_1^2 & 0 & 1 & 0 & 0 & 0 & 0 \\ -y_1^3 & 0 & 0 & 1 & 0 & 0 & 0 \\ y^1 y^3 & 0 & 0 & 0 & 1 & 0 & 0 \\ y^2 y^3 + 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ -3y^2 y^3 - 3 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & y_1^1 & y_1^2 & 0 & y^1 & y^2 & 0 \\ 0 & 0 & -1 & -1 & 2y_1^1 & 2y_1^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & y^1 & y^2 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Denote the system (4.4) by $\mathcal{R}_1 = \tilde{f}^{-1}(0) \subset J_1(\mathcal{E})$. Again it is immediate that $\dim(\mathcal{R}_1) = 3$, $\dim(N(\tilde{A})) = 1$ on $\mathcal{R}_1$, and that $\pi_1^2 : \mathcal{R}_2 \to \mathcal{R}_1$ is a diffeomorphism. Hence there is a bijective correspondence between the solutions of (4.2) and (4.4).

However, one can do another reduction. Recall that λ in the original equations (4.1) was 'only' a Lagrange multiplier, so one suspects that perhaps it is not necessary to treat it and other variables in the same way. Let us try to get a formulation where we can use the special features of λ in the problem. So let us introduce

$$x \longleftrightarrow y^1 \qquad y \longleftrightarrow y^2.$$

Hence here $y = (y^1, y^2)$ and $\mathcal{E} = \mathbb{R} \times \mathbb{R}^2$. Now differentiating twice the constraint in (4.1), we get

$$\begin{cases} y_2^1 + y^1 \lambda = 0 \\ y_2^2 + y^2 \lambda + 1 = 0 \\ \langle y, y_1 \rangle = 0 \\ |y_1|^2 + \langle y, y_2 \rangle = 0 \\ |y|^2 - 1 = 0. \end{cases} \tag{4.5}$$

Consequently y_2 and λ can be solved from the equations

$$\begin{pmatrix} 1 & 0 & y^1 \\ 0 & 1 & y^2 \\ y^1 & y^2 & 0 \end{pmatrix} \begin{pmatrix} y_2^1 \\ y_2^2 \\ \lambda \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ |y_1|^2 \end{pmatrix} = 0. \tag{4.6}$$

Note that the matrix cannot be singular since y^1 and y^2 cannot be both zero. Now we can regard y_2 and λ simply as parameters which have to be (numerically!) computed, and this leads to a simple problem in $J_1(\mathcal{E})$. First let us consider the system (4.5) in $J_2(\mathcal{E})$ and denote the manifold by $\mathcal{R}_2$. Now of course $\mathcal{R}_2$ depends on λ, but obviously $\mathcal{R}_1 = \pi_1^2(\mathcal{R}_2)$ does not; in fact $\mathcal{R}_1$ is given by the system

$$\begin{cases} |y|^2 - 1 = 0 \\ \langle y, y_1 \rangle = 0. \end{cases}$$

Hence we can solve the problem in $\mathcal{R}_1 \subset J_1(\mathcal{E})$, provided we can compute the distribution on $\mathcal{R}_1$. This is done with (4.6). The projected distribution is the nullspace of

$$A = \begin{pmatrix} -y_1^1 & 1 & 0 & 0 & 0 \\ -y_1^2 & 0 & 1 & 0 & 0 \\ -y_2^1 & 0 & 0 & 1 & 0 \\ -y_2^2 & 0 & 0 & 0 & 1 \\ 0 & y_1^1 & y_1^2 & y^1 & y^2 \\ 0 & y^1 & y^2 & 0 & 0 \end{pmatrix}$$

where y_2^1 and y_2^2 are computed from (4.6). Again it is seen that the last two rows can be dropped, and we can immediately write down a vector which spans the nullspace: $(1, y_1^1, y_1^2, y_2^1, y_2^2)$, or more briefly and conveniently $(1, y_1, y_2)$. Hence the final form of the problem is:

$$\begin{cases} |y|^2 - 1 = 0 \\ \langle y, y_1 \rangle = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = \big(1, y_1, y_2\big) \\ \begin{pmatrix} 1 & 0 & y^1 \\ 0 & 1 & y^2 \\ y^1 & y^2 & 0 \end{pmatrix} \begin{pmatrix} y_2^1 \\ y_2^2 \\ \lambda \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ |y_1|^2 \end{pmatrix} = 0. \end{cases} \tag{4.7}$$

In the next section we generalize the above reduction argument.

4.2 *Mechanical systems with holonomic constraints*

Let $f : \mathbb{R}^m \to \mathbb{R}^k$ be a smooth map. The first (resp. second) differential of f, denoted by $\mathrm{d}f$ (resp. $\mathrm{d}^2 f$), is a map $\mathbb{R}^m \to L(\mathbb{R}^m, \mathbb{R}^k)$ (resp. $\mathbb{R}^m \to L(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^k)$) where $L(\mathbb{R}^m, \mathbb{R}^k)$ (resp. $L(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^k)$) denotes the space of linear (resp. bilinear) maps. The value of $\mathrm{d}f$ (resp. $\mathrm{d}^2 f$) at p is denoted by $\mathrm{d}f_p$ (resp. $\mathrm{d}^2 f_p$).

The following class of systems occurs often in applications

$$\begin{cases} B(x, y, y_1)y_2 + f(x, y, y_1) + (\mathrm{d}g)^t \lambda = 0 \\ g(y) = 0 \end{cases} \tag{4.8}$$

where B is invertible, dg has a full rank and λ is the Lagrange multiplier. Note that the pendulum system (4.1) is of this form and that the rank conditions need to be satisfied only on the zero set of g. Now proceeding as in the pendulum case we differentiate twice the constraint and get the system

$$\begin{pmatrix} B & (\mathrm{d}g)^t \\ \mathrm{d}g & 0 \end{pmatrix} \begin{pmatrix} y_2 \\ \lambda \end{pmatrix} + \begin{pmatrix} f \\ \mathrm{d}^2 g\,(y_1, y_1) \end{pmatrix} = 0.$$

This linear system has a solution under the present hypothesis. Hence projecting the system

from $J_2(\mathcal{E})$ to $J_1(\mathcal{E})$ leads to

$$\begin{cases} g(y) = 0 \\ \mathrm{d}gy_1 = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = \left(1, y_1, y_2\right) \\ \begin{pmatrix} B & (\mathrm{d}g)^t \\ \mathrm{d}g & 0 \end{pmatrix} \begin{pmatrix} y_2 \\ \lambda \end{pmatrix} + \begin{pmatrix} f \\ \mathrm{d}^2 g(y_1, y_1) \end{pmatrix} = 0. \end{cases} \tag{4.9}$$

Note that the overdetermined character of the system has disappeared and that the solution of the linear system can be computed numerically.

4.3 *On the index*

Before discussing the actual numerical methods, let us say a few words about the index. There are many definitions of the index, see Brenan *et al.* (1989); Hairer & Wanner (1991); Hairer *et al.* (1989); März (1992), and this notion (or these notions) has been used to characterize the numerical difficulty of the problem. It is unnecessary to analyse here more closely various definitions since we can recommend to interested readers the articles by Le Vey (1998) and Seiler (1999) where they are discussed from the jet point of view, also in the context of PDEs. So we simply make some remarks on the numerical implications of this matter.

One of the definitions, the differential index, is rather close in spirit to the following definition: the index is the number of differentiation and elimination steps needed to transform the given system into an involutive system. Hence one may interpret the index as measuring the lack of information in the original system, and so it is quite natural that higher-index problems are numerically more difficult, *if one uses the original system in the numerical computations*. We use the involutive form so we do not 'see' any index numerically. We might conclude that the index is a useful concept when using certain algorithms to deal with certain representations of differential equations. Hence it is quite natural that when other representations and algorithms are used the index may cease to be necessary or useful.

5. Computation of one-dimensional integral manifolds

5.1 *Preliminary remarks*

We have seen above that the computation of the solutions of a differential system reduces to the following:

> Given a manifold M, a point $p \in M$ and a one-dimensional distribution $\mathcal{D}$ on M, compute the integral manifold of $\mathcal{D}$ through p.

By Theorem 2.1 there exists a unique solution to this problem. Now in practice M is given as a zero set of some map as in (3.1), but usually $\mathcal{D}$ is not explicitly known. However, given a point $p \in M$ we can numerically evaluate $\mathcal{D}_p$ and this is sufficient in applications. The

standard ODE codes work in the same way: it is not necessary to know a formula for the vector field, it is enough to be able to evaluate it at a given point.

Recall that around a regular point p one can find a vector field V such that $0 \neq V_a \in \mathcal{D}_a$ for all a in some neighbourhood of p and we can also normalize by requiring $|V_a| = 1$. Such a V is called a vector field associated to $\mathcal{D}$. Note that V may not exist globally, as our example in Section 2 showed. Finally let us remark that V is not needed in actual computations, but it is useful in the analysis of numerical methods.

Let us also remark that there is no reason why the problem should always be stiff, so there is no need to restrict one's attention to implicit methods. Of course our formulation includes all the classical stiff systems: simply take M to be some Euclidean space and V to be a vector field which defines a stiff system. However, the point is that overdetermined systems (or DAEs) are not intrinsically stiff.

To proceed in our construction and analysis of numerical methods to the above problem we must pause to introduce some terminology.

5.2 *Some notions of Riemannian geometry*

For more details on Riemannian geometry we refer to Spivak (1979, vol. 2 and vol. 4) or do Carmo (1992). Let M be a smooth submanifold of $\mathbb{R}^m$. Then its normal space at p is denoted by NM_p. In these circumstances NM_p is the orthogonal complement of TM_p in $(T\mathbb{R}^m)_p$. Let us give M the Riemannian metric induced by the standard metric in $\mathbb{R}^m$ and denote by ∇ the (unique) symmetric connection which is compatible with this metric.

Let X and Y be vector fields on M; denoting their extensions to $\mathbb{R}^m$ by the same letter we have

$$\mathrm{d}Y(X_p) = \nabla_{X_p} Y + S(X_p, Y_p) \tag{5.1}$$

where $\mathrm{d}Y(X_p)$ is the standard directional derivative, $\nabla_{X_p} Y$ is the covariant derivative and S is the second fundamental tensor. Note that $\left(\nabla_{X_p} Y\right)_p \in TM_p$ and that at each point p, S is a symmetric bilinear map $TM_p \times TM_p \to NM_p$.

Note that the covariant derivative is intrinsic to M while the second fundamental tensor depends on how M is embedded in $\mathbb{R}^m$. Let V be a vector field on M and c be an integral curve of V. If $|V_p| = 1$ for all p, i.e. if c is parametrized by arclength, then $|\nabla_{V_p} V|$ is called the geodesic curvature and $|S(V_p, V_p)|$ is called the normal curvature of c at p.

5.3 *Analysis of local error for some methods*

Let $k < m$ and $f : \mathbb{R}^m \to \mathbb{R}^k$ be a smooth map and let zero be a regular value of f. We think of f as in (3.1), but for the present discussion it is convenient to regard $J_q(\mathcal{E})$ simply as a big Euclidean space $\mathbb{R}^m$ where $m = (q+1)n + 1$. Hence $M := f^{-1}(0) \subset \mathbb{R}^m$ is a smooth manifold of dimension $m - k$. Let us give M the Riemannian metric induced by the standard metric in $\mathbb{R}^m$. When it is convenient, the objects defined on M will also be considered as defined on $\mathbb{R}^m$ without writing explicitly the inclusion map. Let $\mathcal{D}$ be a

smooth one-dimensional distribution on M and let V be the vector field associated with it. We would like to compute some integral manifolds of $\mathcal{D}$.†

Let $p \in M$ be the current point and $c : \mathbb{R} \to M$ be the integral curve of the associated vector field with $c(0) = p$ and $c'(0) = V_p$. The arclength parameter is denoted by s. Let us compute an approximation of $c(h)$ for small h. The simplest possibility is to use Euler's method in the following form.

1. Choose the step-size h and take a 'Euler step' along the tangent space; this gives a point $p + V_p h$.
2. Since $p + V_p h \notin M$ in general we project it orthogonally back to M. This projection is well defined for h sufficiently small by the tubular neighbourhood theorem (Spivak, 1979, vol. 1, p 466).

Hence an approximation q of $c(h)$ is obtained by solving

$$\begin{cases} q + (\mathrm{d}f_q)^t \mu = p + V_p h \\ f(q) = 0 \end{cases} \tag{5.2}$$

where $\mu \in \mathbb{R}^k$. It can easily be verified that this system has a solution for small h. To get error estimates we compute the first few terms in the expansions

$$q = q^0 + q^1 h + q^2 h^2 + \cdots$$
$$\mu = \mu^0 + \mu^1 h + \mu^2 h^2 + \cdots.$$

This leads to

PROPOSITION 5.1

$$q = p + V_p h - (\mathrm{d}f_p)^t \mu^2 h^2 + O(h^3)$$
$$\mu = \tfrac{1}{2} (\mathrm{d}f_p (\mathrm{d}f_p)^t)^{-1} \mathrm{d}^2 f_p(V_p, V_p) h^2 + O(h^3).$$

Proof. Obviously $q^0 = p$ and $\mu^0 = 0$. For linear terms we get the equations

$$q^1 + (\mathrm{d}f_p)^t \mu^1 = V_p$$
$$\mathrm{d}f_p q^1 = 0.$$

This implies that $q^1 = V_p$ and $\mu^1 = 0$. Then the quadratic terms give

$$q^2 + (\mathrm{d}f_p)^t \mu^2 = 0$$
$$2\mathrm{d}f_p q^2 + \mathrm{d}^2 f_p(V_p, V_p) = 0$$

from which the result follows. □

PROPOSITION 5.2

$$q^2 = \tfrac{1}{2} S(V_p, V_p).$$

†In the limiting case where $m = k + 1$ and hence the integral manifold is M itself, this problem has been studied quite extensively, see Allgower & Georg (1990).

Proof. Let $c : \mathbb{R} \to M$ be any curve (parametrized by arclength) such that $c(0) = p$ and $c'(0) = V_p$. Recall that $(T\mathbb{R}^m)_p = TM_p \oplus NM_p$ and denote by π the orthogonal projection $\pi : \mathbb{R}^m \to NM_p$. Now it is known that $\pi(c''(0)) = S(V_p, V_p)$, see Spivak (1979, vol. 3, p 5). On the other hand differentiating twice the identity $f \circ c = 0$ we obtain

$$\mathrm{d}f_p c''(0) + \mathrm{d}^2 f_p(V_p, V_p) = 0.$$

Now $c''(0) = v + \pi(c''(0)) = v + (\mathrm{d}f_p)^t a$ for some $v \in TM_p$ and some $a \in \mathbb{R}^k$. Since $\mathrm{d}f_p v = 0$ we can use the above equation to compute a which then gives the required result. □

Let us then compare the approximation to the exact value. Using (5.1) we obtain

$$c(h) = p + V_p h + \tfrac{1}{2}((\nabla_{V_p} V)_p + S(V_p, V_p))h^2 + O(h^3).$$

Subtracting the expansions we get immediately

COROLLARY 5.1

$$c(h) - q = \tfrac{1}{2}(\nabla_{V_p} V)_p h^2 + O(h^3).$$

Hence our simple method (5.2) is of first order (i.e. the local error is $O(h^2)$). Note that the main error term is rather natural: it reduces to the standard directional derivative of the classical case when M is a Euclidean space.

Let us then consider some simple implicit methods. One can formulate the implicit Euler method in two ways.

$$\begin{cases} p + (\mathrm{d}f_p)^t \mu = q - V_q h \\ f(q) = 0 \end{cases} \qquad \begin{cases} q + (\mathrm{d}f_q)^t \mu = p + V_q h \\ f(q) = 0. \end{cases} \tag{5.3}$$

In either case we obtain

PROPOSITION 5.3 Let q solve either of the systems (5.3); then

$$c(h) - q = -\tfrac{1}{2}(\nabla_{V_p} V)_p h^2 + O(h^3).$$

Proof. Expanding as before $q = p + q^1 h + q^2 h^2 + \cdots$, we immediately get $q^1 = V_p$ and $\mu^1 = 0$. Expanding the second system in (5.3) we get

$$\begin{aligned} q^2 + (\mathrm{d}f_p)^t \mu^2 &= \mathrm{d}V V_p \\ \mathrm{d}f_p q^2 + \tfrac{1}{2}\mathrm{d}^2 f_p(V_p, V_p) &= 0. \end{aligned}$$

Hence $(\mathrm{d}f_p(\mathrm{d}f_p)^t)\mu^2 = \frac{1}{2}\mathrm{d}f_p S(V_p, V_p)$ and $q^2 = \nabla_{V_p} V + \frac{1}{2}S(V_p, V_p)$. Expanding the first system produces the same q^2, but μ^2 changes sign. □

The implicit mid-point rule can be formulated as follows.

$$\begin{cases} q + (\mathrm{d}f_q)^t \mu = p + V_r h \\ f(q) = 0 \\ r + (\mathrm{d}f_r)^t \nu = \frac{1}{2}(p + q) \\ f(r) = 0. \end{cases} \tag{5.4}$$

PROPOSITION 5.4 Let q solve the above system; then

$$c(h) - q = O(h^3).$$

Proof. Evidently $q^0 = r^0 = p$, $q^1 = V_p$, $r^1 = V_p/2$, $\mu^0 = \mu^1 = \nu^0 = \nu^1 = 0$. For second-order terms we get

$$q^2 + (\mathrm{d}f_p)^t \mu^2 = \tfrac{1}{2}\,\mathrm{d}VV$$
$$\mathrm{d}f q^2 + \tfrac{1}{2}\,\mathrm{d}^2 f(V, V) = 0.$$

Hence $\mu^2 = 0$ and $q^2 = \frac{1}{2}\,\mathrm{d}VV$ from which the result follows. □

We have seen that it is rather straightforward to formulate low-order schemes in the present context. Moreover their error terms are what one expects by the classical theory. The analysis of higher-order schemes is more involved and will be treated elsewhere, see Tuomela & Arponen (to appear). Note that the methods (5.2)–(5.4) are one-step methods whose local errors are of order 2, 2 and 3. By standard theorems we can then conclude that the global errors for sufficiently small h are 1, 1 and 2.

6. Numerical implementation

We have actually implemented the explicit Euler method (5.2) and the midpoint rule (5.4). Implicit Euler (5.3) is only used for step size control as explained below. There are three basic tasks in our methods.

1. Given a point $p \in M$, compute the distribution $\mathcal{D}_p$.
2. Step size control.
3. Solving the nonlinear system (5.2), (5.3) or (5.4).

All these subproblems reduce to fairly standard numerical problems so we will say just a few words about them. All computations were done with *Mathematica* (Wolfram, 1991). Of course this is not as efficient as using standard languages like Fortran or C. However, at this stage we found *Mathematica* to be more convenient than the traditional languages, so the implementation in Fortran or C has to await future versions of our code.

6.1 *Computing the distribution*

As we have seen in the examples, the distribution defined in (3.3) can easily be represented as a nullspace of some matrix A, as in (4.3). Let us analyse more closely the structure of this matrix. Consider the system (3.1); then A is given by

$$A = \begin{pmatrix} -v & I_{nq} & 0_{nq\times n} \\ w & A_1 & A_2 \end{pmatrix}$$

where $w = \partial f/\partial x$, $A_2 = \partial f/\partial y_q$, A_1 contains the partial derivatives with respect to $y, y_1, \ldots, y_{q-1}$ and finally $v \in \mathbb{R}^{nq\times 1}$ contains the vectors $y_1, y_2, \ldots, y_q$. Let us further set $B = (-v, I_{nq}, 0_{nq\times n})$ and let $z \in \mathbb{R}^{(q+1)n+1}$ be a solution of $Az = 0$. From the equations $Bz = 0$ we immediately get $z^2, \ldots, z^{nq+1}$ if we know z^1. Hence the computation

of the nullspace of A can be reduced to the computation of the nullspace of the following $k \times (n+1)$ matrix:

$$C = \begin{pmatrix} w + A_1 v & A_2 \end{pmatrix}. \tag{6.1}$$

Note that the dimensions of C are independent of q. Now this problem is (mildly) ill-posed numerically since in general $k > n$ and hence because of the round-off errors the nullspace would be trivial. To circumvent this difficulty, we use the standard singular value decomposition

$$C = U \Sigma V^t$$

where U and V are orthogonal and Σ contains the singular values in descending order in its diagonal. The last singular value should be very close to zero, and hence the last column of V should give a good approximation of the required direction. Of course we must check that there is only one singular value which is much smaller than the others; if this is not the case, this would indicate some singularity of the solution. We did not encounter any problems of this sort in the examples below.

Note also that for mechanical systems with holonomic constraints, i.e. problems of the form (4.8), the distribution can be computed from the regular linear system and the SVD is not needed at all.

6.2 *Step size*

In the explicit Euler case we used the standard technique: take one big step of size $2h$ achieving point b and two smaller ones of size h, thus achieving points a_1 and a_2. Then we computed $|a_2 - b|$ in the ambient space and accepted the step if the difference was smaller than a given tolerance. If the step was rejected we halved the step-size because then 'old' a_1 could be used as a 'new' b.

In the case of the midpoint rule it was convenient to use the auxiliary point r in (5.4) which has to be computed anyway. One can easily check that the local error of r is of order 2. Hence we can take an implicit Euler step from r with the step $h/2$ and compare this value to the value q obtained by the midpoint rule.

6.3 *Projection*

We used Newton's method to solve the systems (5.2)–(5.4). The values $q := p$ and $\mu := 0$ were used as initial guesses. The Jacobian was computed symbolically, and it was evaluated at every third step during the iteration of one point. However, usually two iterations were enough. In test problems we typically used $10^{-2} \cdots 10^{-8}$ as an error tolerance for the Newton iteration. For example, in the Hénon–Heiles case the error tolerance of the iteration was 0·01 and the distance of the point from M was of the order of 10^{-5}.

REMARK 6.1 In a forthcoming article (Tuomela & Arponen, to appear) we study higher-order methods with more sophisticated step-size controls than above. Also, expansions used in analysing the error can be used to obtain a better initial guess for the Newton iteration. This is also explained in Tuomela & Arponen (to appear).

This leaves the following important problems which must be addressed in future versions of our code.

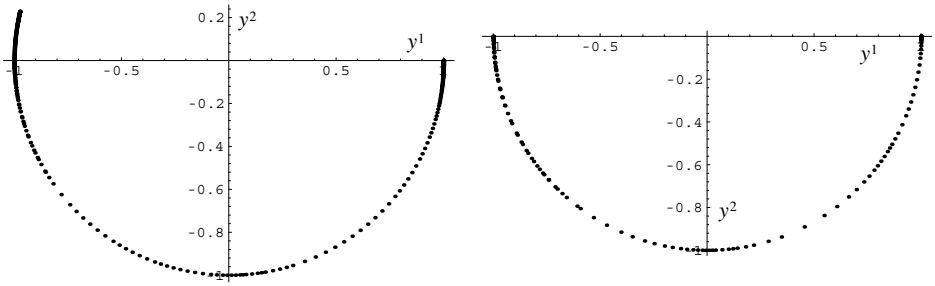

FIG. 5. A solution of system (4.7) and the system (7.1), $0 \leqslant x \leqslant 5{\cdot}1$.

- Using the full singular value decomposition to compute the nullspace of the matrix C in (6.1) is rather excessive, and consequently this step could be done more efficiently. Also C normally only changes slightly during a single step so maybe it would be possible to exploit this in the computations.
- Instead of carrying out the full Newton iteration, one could try pseudoinverses which would result in a smaller system to be solved. Since we are close to the manifold, one would expect that this kind of iteration would also converge quite rapidly. However, the projection would no longer be orthogonal, and it is not clear how this would affect the accuracy.
- The symbolic computation of the various required derivatives is rather time consuming. Fortunately this is not necessary: automatic differentiation permits a very accurate and efficient numerical evaluation of the required quantities. However, discussion of these techniques is outside the scope of the present article and we refer to Berz *et al.* (1996) for more information on this subject.

7. Numerical examples

7.1 *Pendulum*

The pendulum system is of the form (4.8) with $y \in \mathbb{R}^2$, $\lambda \in \mathbb{R}$, $g(y) = (|y|^2 - 1)/2$, $f = (0, 1)$ and $B = I$. The resulting simplified system was given in (4.7). In Fig. 5 we show a solution of (4.7) with quite a large tolerance and with initial value $(x, y, y_1) = (0, 1, 0, 0, 0)$. The solution is not very satisfactory since the energy deviates quite rapidly from the correct value and in particular the computed solution goes above the y^1-axis although the correct solution would always stay below it. This is not surprising because our method is of first order, so one expects that the energy is also only 'first-order correct'. However, one can easily impose conservation of energy. It is sufficient to add the energy

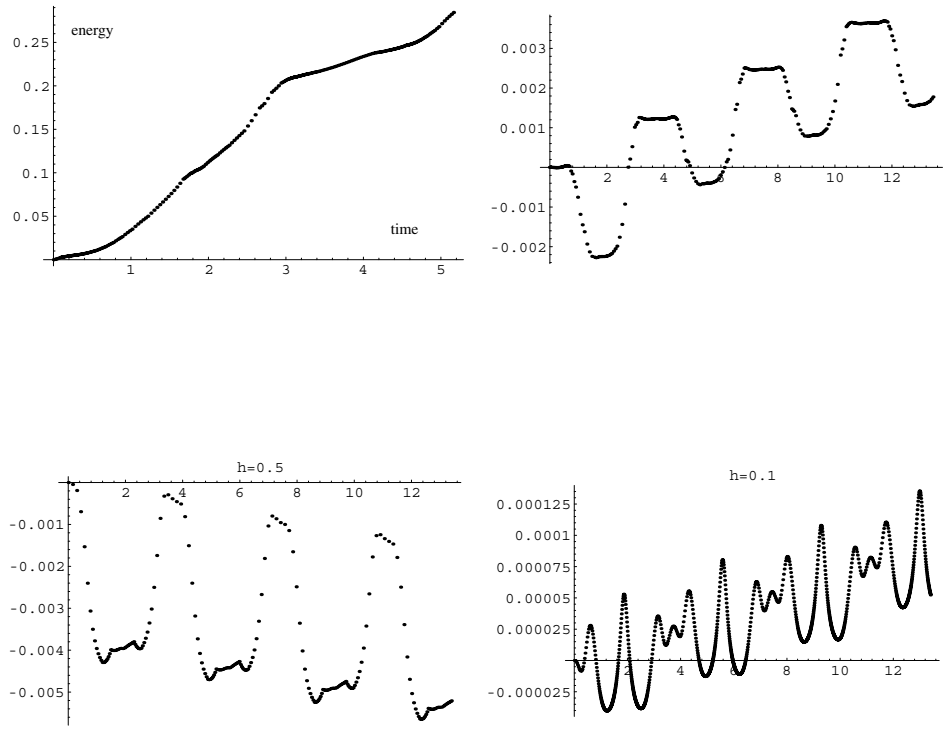

FIG. 6. Evolution of the energy: top left, the Euler method; top right, the midpoint method using variable stepsize, and the midpoint method using step sizes $h = 0{\cdot}5$ and $h = 0{\cdot}1$.

equation to system (4.7) which yields

$$\begin{cases} |y|^2 - 1 = 0 \\ \langle y, y_1 \rangle = 0 \\ \frac{1}{2}|y_1|^2 + y^2 - a = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = (1, y_1, y_2) \\ \begin{pmatrix} 1 & 0 & y^1 \\ 0 & 1 & y^2 \\ y^1 & y^2 & 0 \end{pmatrix} \begin{pmatrix} y_2^1 \\ y_2^2 \\ \lambda \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ |y_1|^2 \end{pmatrix} = 0 \end{cases} \tag{7.1}$$

where a is the constant energy. Note that the distribution restricts to the appropriate submanifold. In Fig. 5 there is also a solution of (7.1) with the same initial values and same tolerance that were used without the conservation of energy.

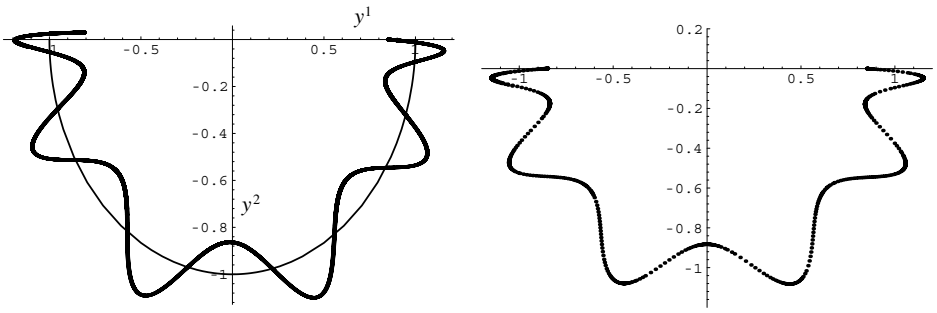

FIG. 7. A solution of system (7.2), $0 \leqslant x \leqslant 1{\cdot}7$, $\varepsilon = 0{\cdot}1$. The Euler method is shown on the left and the midpoint method on the right.

In Fig. 6 we show the evolution of the energy in various cases. It can be seen that with the midpoint method the energy deviates quite slowly from the correct value. Recall that the classical midpoint method is symplectic, so it is perhaps not so surprising that our version of it behaves well with respect to conservation of energy. Note that in the case of variable stepsize, the stepsizes were between 0·05 and 1, so the case $h = 0{\cdot}5$ represents a kind of average stepsize. Anyway let us reiterate that adding the energy constraint does not make the system more difficult to solve.

7.2 *Stiff pendulum*

Let us then consider the stiff pendulum (Hairer *et al.*, 1989, p 119) whose involutive form is

$$\begin{cases} y_2^1 + y^1 y^3 = 0 \\ y_2^2 + y^2 y^3 + 1 = 0 \\ ((y^1)^2 + (y^2)^2)(\varepsilon^2 y^3 - 1)^2 - 1 = 0 \\ \varepsilon^2 y_1^3 + (y^1 y_1^1 + y^2 y_1^2)(\varepsilon^2 y^3 - 1)^3 = 0 \\ (\varepsilon^2 y^3 - 1)\varepsilon^2 y_2^3 - 3\varepsilon^4 (y_1^3)^2 + ((y_1^1)^2 + (y_1^2)^2 - y^2)(\varepsilon^2 y^3 - 1)^4 - y^3(\varepsilon^2 y^3 - 1)^2 = 0. \end{cases}$$

Now y^3 cannot be solved without introducing square roots, and we want to avoid this. However, we can still do the projection $J_2 \to J_1$. In addition, y_1^3 is 'graph-like', so one can eliminate it also. The simplified system can be written as

$$\begin{cases} ((y^1)^2 + (y^2)^2)(\varepsilon^2 y^3 - 1)^2 - 1 = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = (1, y_1^1, y_1^2, -(y^1 y_1^1 + y^2 y_1^2)(\varepsilon^2 y^3 - 1)^3/\varepsilon^2, -y^1 y^3, -y^2 y^3 - 1). \end{cases} \tag{7.2}$$

Here V is given in symbolic form just for the purposes of illustration. This form is not needed, but if a symbolic form is available, it could be given to the programme. This would somewhat speed up the computation.

Let us take $\varepsilon = 0{\cdot}1$ and compute the solution in the interval $0 \leqslant x \leqslant 1{\cdot}7$ using the initial value $(x, y^1, y^2, y^3, y_1^1, y_1^2) = (0, 0{\cdot}85, 0, -17{\cdot}65, 0, 0)$. In Fig.7 we show the results obtained with the Euler method and the midpoint method. In the former case we needed about 4700 points and in the latter about 500 points. Then in Fig. 8 we computed the solution using the midpoint method with $\varepsilon = 0{\cdot}01$ and the initial value $(x, y^1, y^2, y^3, y_1^1, y_1^2) = (0, 1, 0, -17{\cdot}65, 0{\cdot}016, 0)$. Of course, even in the implicit case the stepsize has to be quite small, because of the rapid oscillations.

7.3 *Discharge pressure control problem*

Let us consider the discharge pressure flow problem given in Hairer *et al.* (1989, p 116), whose involutive form is

$$\begin{cases} 20y_1^1 + y^1 - y^2 = 0 \\ 75y_1^2 + 5y_1^3 + y^3 - c_1 = 0 \\ 20y_1^4 + y^5 - f = 0 \\ c_2 y^4 (y^5)^2 - c_3 y^4 y^5 + c_4 y^4 - y^3 = 0 \\ (y^1)^2 (y^4)^2 - c_5 (y^1)^2 + c_6 f^2 = 0 \\ (40c_2 y^4 y^5 - 20c_3 y^4) y_1^5 - 20y_1^3 - c_2 (y^5)^3 + g_1 (y^5)^2 - g_2 y^5 + c_4 f = 0 \\ y^1 y^2 (y^4)^2 - (y^1)^2 y^4 y^5 + f (y^1)^2 y^4 - c_5 y^1 y^2 + c_6 g_3 = 0 \\ 20y^1 ((y^4)^2 - c_5) y_1^2 - 20(y^1)^2 y^4 y_1^5 + (y^2)^2 (y^4)^2 + (y^1)^2 y^4 y^5 \\ -4y^1 y^2 y^4 y^5 + (y^1)^2 (y^5)^2 - f (y^1)^2 y^4 - 2f (y^1)^2 y^5 + 4f y^1 y^2 y^4 \\ +20f' (y^1)^2 y^4 + f^2 (y^1)^2 - c_5 (y^2)^2 + c_6 g_4 = 0 \end{cases} \tag{7.3}$$

where

$$\begin{array}{lll} c_1 = 99{\cdot}1 & c_2 = 0{\cdot}001 & c_3 = 0{\cdot}075 \\ c_4 = 3{\cdot}35 & c_5 = 2458{\cdot}18 & c_6 = 1/1{\cdot}44 \end{array}$$

and

$$\begin{array}{ll} f = 5\tanh(x - 10) + 15 & g_1 = c_2 f + c_3 \\ g_2 = c_3 f + c_4 & g_3 = f^2 + 20ff' \\ g_4 = f^2 + 60ff' + 400(f')^2 + 400ff''. & \end{array}$$

The values of the parameters are those given in Hairer *et al.* (1989, p 116). We computed the solution with initial value

$$\begin{aligned} (x, y, y_1) = (&0, 0{\cdot}25, 0{\cdot}25, 99{\cdot}1, 36{\cdot}7, 9{\cdot}998, \\ &0, -8{\cdot}45 \times 10^{-7}, 1{\cdot}27 \times 10^{-5}, 8{\cdot}5 \times 10^{-5}, 1{\cdot}1 \times 10^{-4}). \end{aligned} \tag{7.4}$$

The solution curves obtained with the Euler scheme are shown in Fig. 9 and those obtained with the midpoint scheme are shown in Fig. 10. In Hairer *et al.* (1989) only y^5 (output) is given and the solution obtained is qualitatively the same as here.

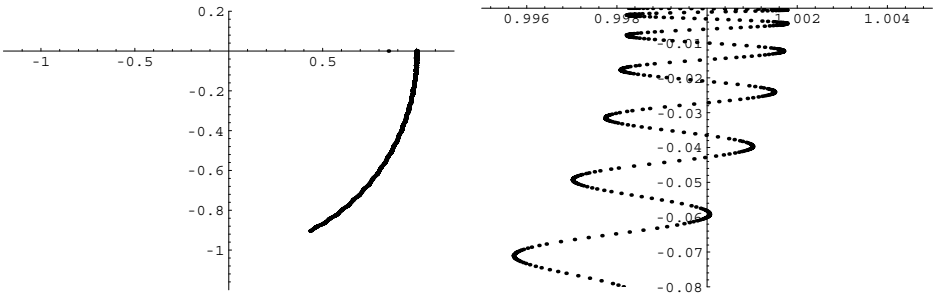

FIG. 8. A solution of system (7.2), $\varepsilon = 0{\cdot}01$, using the midpoint method. On the right we have a magnified view of part of the image on the left.

7.4 *Hamiltonian systems*

Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ and let $M = H^{-1}(0)$; then the Hamiltonian system on $\mathbb{R} \times M$ is given by the vector field

$$V = \left(1, \frac{\partial H}{\partial y^{n+1}}, \ldots, \frac{\partial H}{\partial y^{2n}}, -\frac{\partial H}{\partial y^{1}}, \ldots, -\frac{\partial H}{\partial y^{n}}\right)$$

which defines a distribution on $\mathbb{R} \times M$. If the given system has more invariants (or integrals) then we restrict V to an appropriate submanifold of $\mathbb{R} \times M$. Now in many cases (in particular in the examples that follow), the momenta and first-order jets are identified, so it is convenient to identify $\mathbb{R} \times \mathbb{R}^{2n}$ and $J_1(\mathbb{R} \times \mathbb{R}^n)$. Hence the system is $\mathcal{R}_1 = \mathbb{R} \times M$. Note that V induces a symplectic flow on $\mathbb{R}^{2n}$. However, the flow restricted to M or $\mathcal{R}_1$ is not symplectic.

Consider the following Kepler problem (Sanz Serna & Calvo, 1994, p 6)

$$\begin{cases} \frac{1}{2}|y_1|^2 - 1/|y| - a = 0 \\ y^1 y_1^2 - y^2 y_1^1 - b = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = (1, y_1^1, y_1^2, -y^1/|y|^3, -y^2/|y|^3) \end{cases} \tag{7.5}$$

where a and b are constants (energy and angular momentum). The analytic solution is periodic with period 2π. In Fig. 11 we show two solutions with the same initial values, $(x, y, y_1) = (0, 0{\cdot}5, 0, 0, 1{\cdot}73)$; on the left is the Euler method and on the right the midpoint method. The tolerances were chosen such that qualitatively the solutions were similar. The solutions were computed for $0 \leqslant x \leqslant 100$, and with the Euler method about 8000 points was needed while for the midpoint method about 900 points was sufficient.

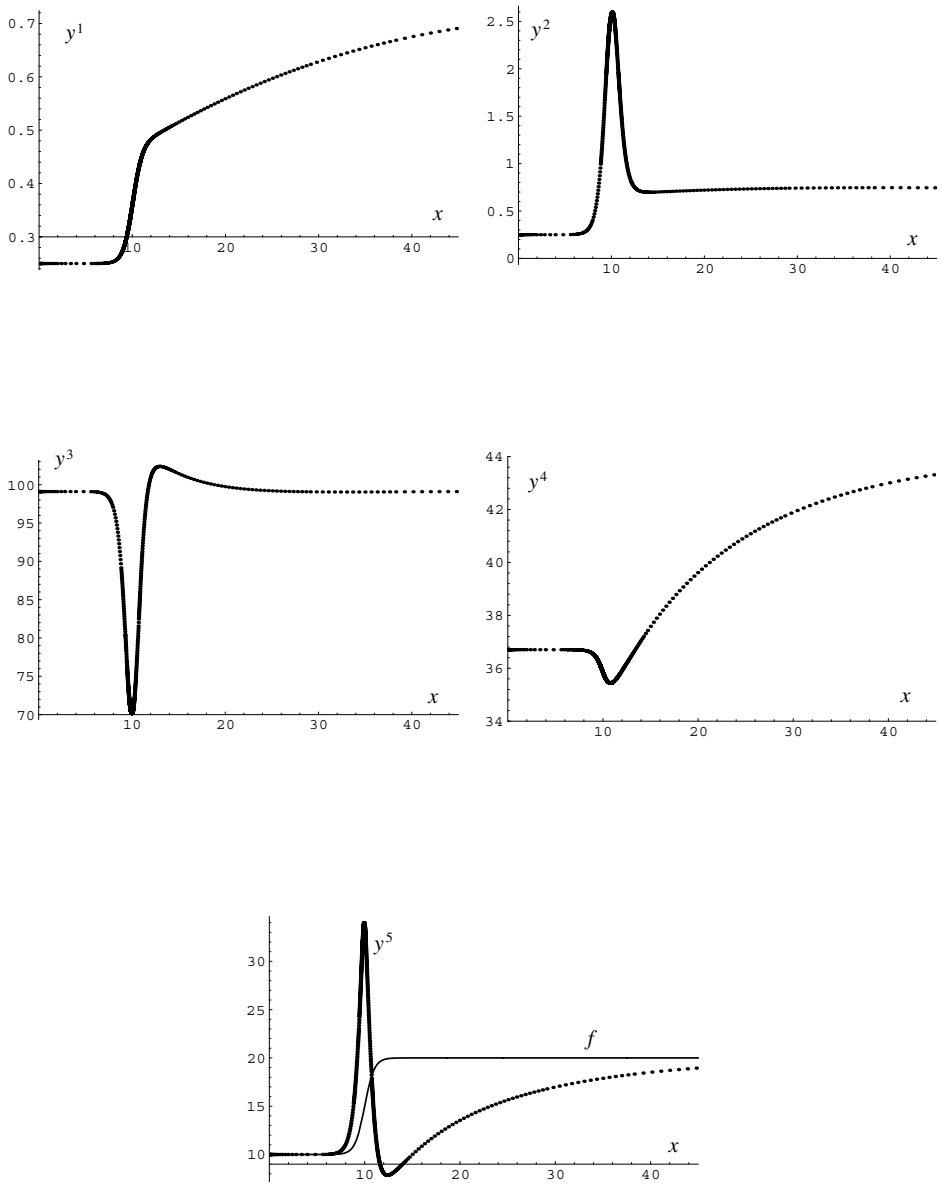

FIG. 9. A solution of system (7.3) with $f = 5\tanh(x - 10) + 15$ using the explicit Euler method.

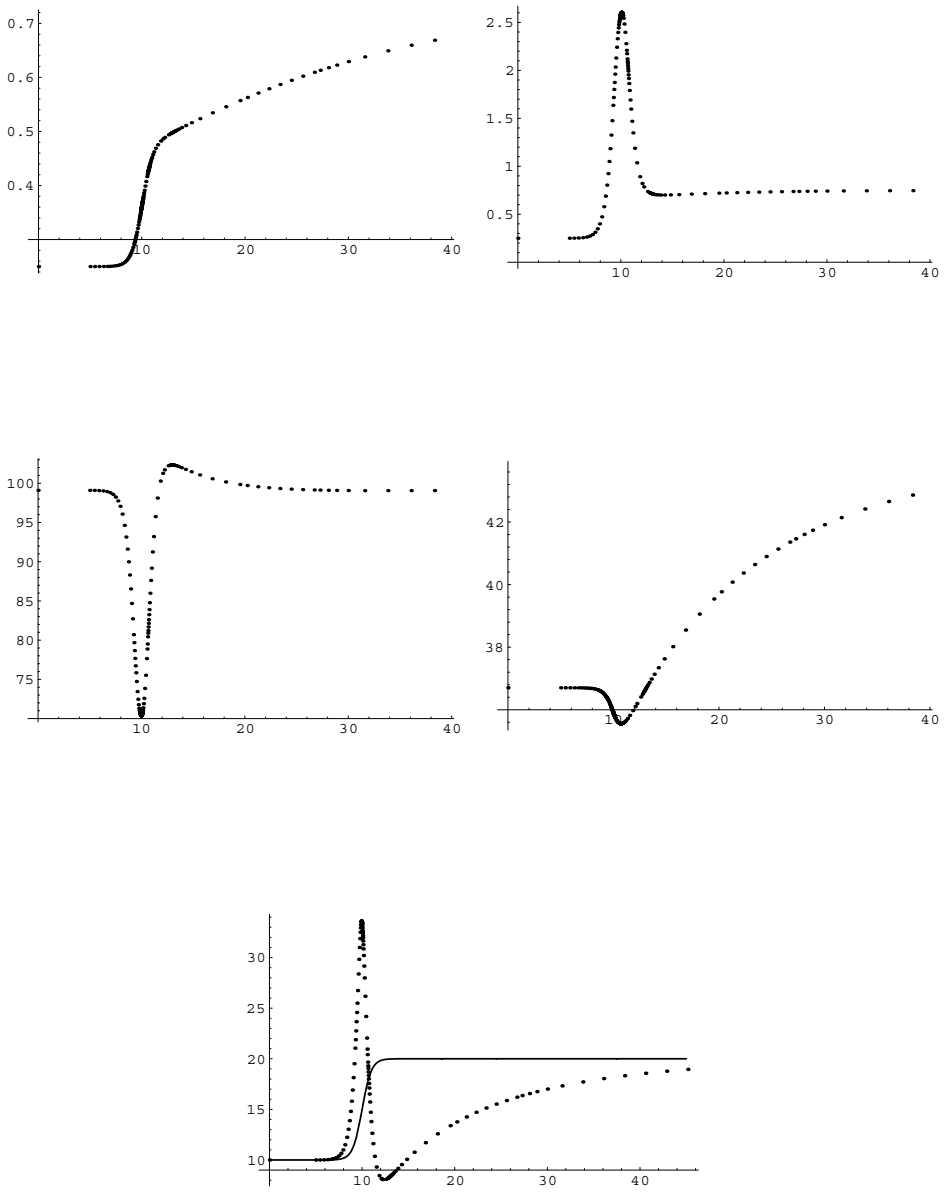

FIG. 10. A solution of system (7.3) with $f = 5\tanh(x - 10) + 15$ using the midpoint method.

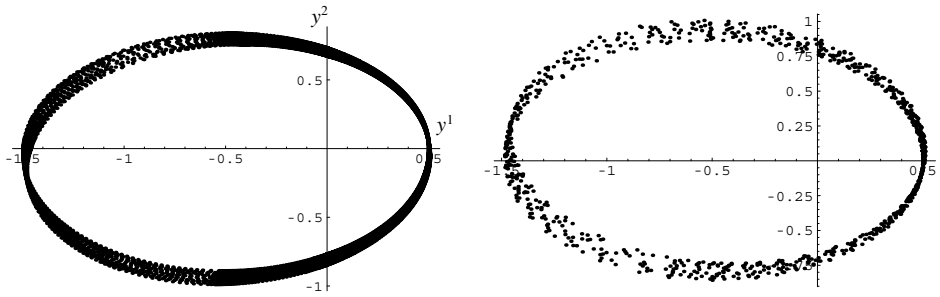

FIG. 11. A solution of the Kepler problem (7.5) using the Euler (left) and midpoint (right) methods.

Next we take a Hénon–Heiles system (Sanz Serna & Calvo, 1994, p 13)

$$\begin{cases} \frac{1}{2}|y_1|^2 + \frac{1}{2}|y|^2 + (y^1)^2 y^2 - \frac{1}{3}(y^2)^3 - a = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = (1, y_1^1, y_1^2, y^1 + 2y^1 y^2, y^2 + (y^1)^2 - (y^2)^2) \end{cases} \tag{7.6}$$

where a is again the (constant) energy. For this system we took as an initial value $(x, y, y_1) = (0, 0{\cdot}12, 0{\cdot}12, 0{\cdot}12, 0{\cdot}12)$, which yields quasiperiodic solutions. We computed the solution using both methods, for $0 \leqslant x \leqslant 303$. In Fig. 12 we have plotted the values of the solution in the (y^2, y_1^2) plane when the hyperplane $y^1 = 0$ was crossed. On the left we show the solution with the Euler method (101 points) and on the right the solution with the midpoint method (141 points). With the Euler method, 5000 points had to be computed and with the midpoint method 2500. The plots should yield closed curves rather than the spirals that can be seen in the figures. However, the results are quite satisfactory, taking into account the fact that tolerances were not particularly strict and the methods are of low order.

7.5 *Mechanical system*

As a final example we consider the system given in Leimkuhler & Skeel (1994). It is of the form (4.8) with $y \in \mathbb{R}^{12}$, $\lambda \in \mathbb{R}^5$, $B = I_{12}$, $f = \nabla F$, where $F = \frac{5}{2}\langle y, Ky\rangle$ and K and g

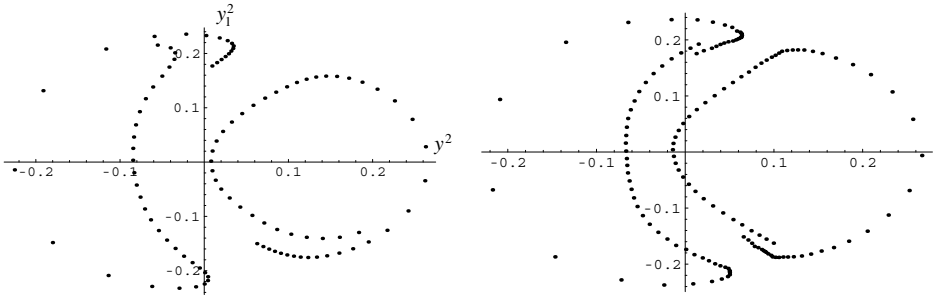

FIG. 12. The Poincaré sections for the Hénon–Heiles system (7.6); the Euler method is shown on the left and the midpoint method on the right.

are given by

$$K = \begin{pmatrix} I_2 & 0 & -I_2 & 0 & 0 & 0 \\ 0 & I_2 & 0 & -I_2 & 0 & 0 \\ -I_2 & 0 & 2I_2 & 0 & -I_2 & 0 \\ 0 & -I_2 & 0 & 2I_2 & 0 & -I_2 \\ 0 & 0 & -I_2 & 0 & I_2 & 0 \\ 0 & 0 & 0 & -I_2 & 0 & I_2 \end{pmatrix}$$

$$g(y) = \frac{1}{2}\begin{pmatrix} (y^1 - y^3)^2 + (y^2 - y^4)^2 - 1 \\ (y^3 - y^5)^2 + (y^4 - y^6)^2 - 1 \\ (y^5 - y^7)^2 + (y^6 - y^8)^2 - 1 \\ (y^7 - y^9)^2 + (y^8 - y^{10})^2 - 1 \\ (y^9 - y^{11})^2 + (y^{10} - y^{12})^2 - 1 \end{pmatrix}.$$

Variables $y^1, \ldots, y^{12}$ are coordinates of six points in the plane. In this case the system (4.9) can be written as

$$\mathcal{R}_1 : \quad \begin{cases} g(y) = 0 \\ \mathrm{d}g y_1 = 0 \\ \mathcal{D} = \operatorname{span}(V) \\ V = \left(1, y_1, -(\mathrm{d}g)^t \lambda - 5Ky\right) \\ \mathrm{d}g(\mathrm{d}g)^t \lambda + 5\,\mathrm{d}g K y - \mathrm{d}^2 g(y_1, y_1) = 0. \end{cases} \tag{7.7}$$

So the final problem is in $J_1(\mathbb{R} \times \mathbb{R}^{12}) \simeq \mathbb{R}^{25}$ and $\dim(\mathcal{R}_1) = 15$. It can readily be verified that treating Lagrange multipliers like other variables and transforming the system into involutive form produces a 15-dimensional system in 52-dimensional ambient space.

The energy of the system is

$$E = \tfrac{1}{2}|y_1|^2 + \tfrac{5}{2}\langle y, Ky\rangle.$$

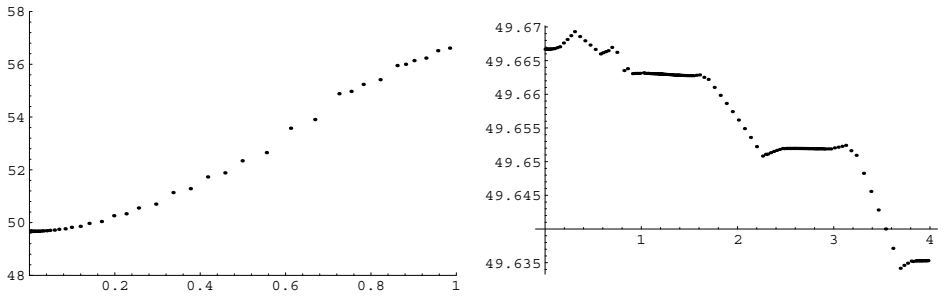

FIG. 13. Evolution of the energy using the Euler (left) and midpoint (right) methods.

In Fig. 13 we show the evolution of the energy for the Euler and midpoint methods, and as in the pendulum case, the midpoint method remains quite close to the correct manifold. In Fig. 14 we show the time evolution of the configuration of the system (7.7) with initial value

$$(x, y, y_1) = (0, 0, 0, 0{\cdot}5, 0{\cdot}866, 1, 0, 1{\cdot}5, 0{\cdot}866, 2, 0, 2{\cdot}5, 0{\cdot}866,\\ 1, -5{\cdot}77, -1, -4{\cdot}62, 1, -3{\cdot}464, -1, -2{\cdot}31, 1, -1{\cdot}155, -1, 0).$$

For small values of x, the solutions with both methods are very similar so we have plotted only one figure. For bigger values of x, the energy in the Euler case starts to grow quite fast, so the results also quickly become quite different in the two cases.

Now adding the energy equation gives the system

$$\begin{cases} g(y) = 0 \\ \mathrm{d}g y_1 = 0 \\ \frac{1}{2}|y_1|^2 + \frac{5}{2}\langle y, Ky\rangle - a = 0 \\ \mathcal{D} = \mathrm{span}(V) \\ V = (1, y_1, -(\mathrm{d}g)^t\lambda - 5Ky) \\ \mathrm{d}g(\mathrm{d}g)^t\lambda + 5\mathrm{d}gKy - \mathrm{d}^2g(y_1, y_1) = 0 \end{cases} \tag{7.8}$$

where a is the constant energy. The solution of system (7.8) is given in Fig. 15 where the configurations are given at the same time instants as in Fig. 14. The same initial values and same tolerance were used as for system (7.7). Again there was practically no difference between the solutions obtained with the Euler and midpoint methods, so only one plot is given. However, comparing the results obtained with conservation of energy to those obtained without conservation of energy, it is seen that the results may become quite different. This is surprising because the midpoint method remains rather close to the energy surface even without explicit conservation of energy. The same phenomenon was also observed with other initial values.

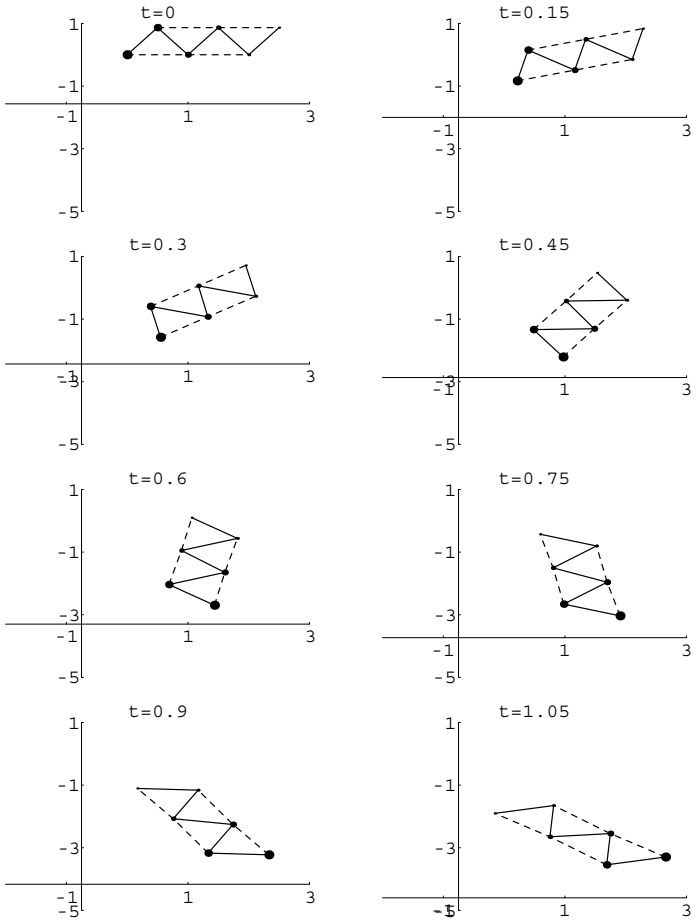

FIG. 14. Evolution of the configuration of the system (7.7) without conservation of energy.

8. Conclusion

We have introduced a new framework to treat arbitrary systems of ODEs and DAEs. The main notion which emerges is the involutivity of the system, and we have seen that using involutive form in the computations (and analysis) leads to rather different conclusions than the traditional treatment of such systems. To sum up, let us list a few of them:

- We have not defined DAEs. This is simply because geometrically there is no difference between ODEs and DAEs.
- It is possible to obtain reliable results even with low-order explicit methods. Hence DAEs are not intrinsically stiff.
- The concept of index is not needed in our approach. Hence the index is not related to the numerical difficulty of solving an *involutive* system.

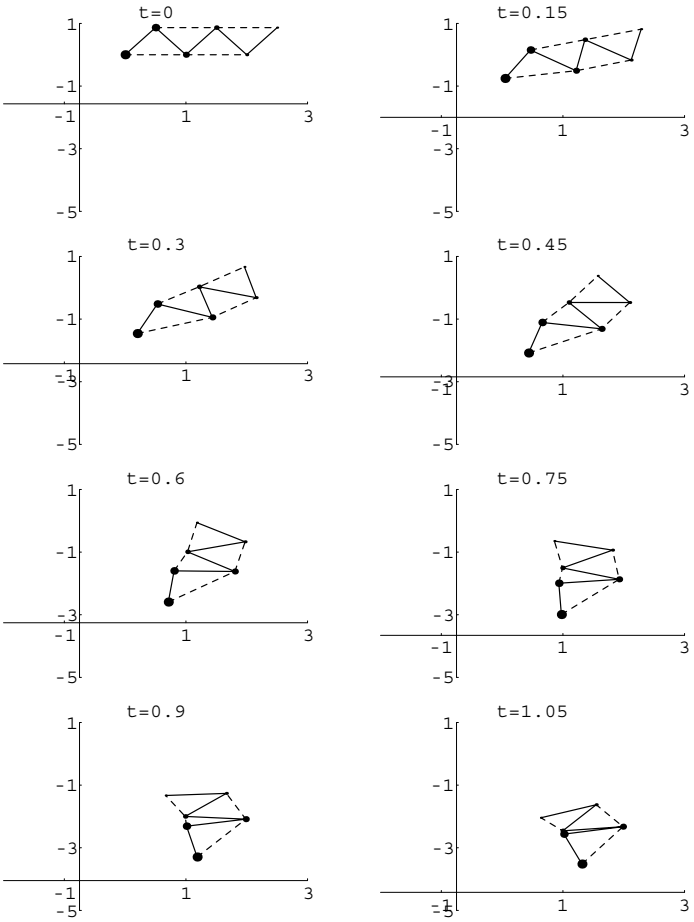

FIG. 15. Evolution of the configuration of the system (7.8) with conservation of energy.

- Because the notion of the solution is more general than the traditional one we obtain more smooth solutions and hence encounter less singularities than in the traditional setting.

This work can be extended in many directions. We have already completed a second article where higher-order Runge–Kutta-type methods are analysed and implemented in our context (Tuomela & Arponen, to appear). The most immediate task after this is to study how the subproblems discussed in Section 6 could be implemented as efficiently as possible. An interesting aspect is also the question of genericity of the system and analysis of the system with respect to perturbations. This in turn is closely related to the analysis of singularities of the system and its solutions.

REFERENCES

ALEKSEEVSKIJ, D. V., VINOGRADOV, A. M., & LYCHAGIN, V. V. 1991 Basic ideas and concepts of differential geometry. *Geometry I (Encyclopaedia of Mathematical Sciences 28)*. (R. V. Gamkrelidze ed). Berlin: Springer.

ALLGOWER, E. & GEORG, K. 1990 *Numerical Continuation Methods (Springer Series in Computational Mathematics 13)*, Berlin: Springer.

ARNOLD, V. 1983 *Geometrical Methods in the Theory of Ordinary Differential Equations (Grundlehren 250)*. Berlin: Springer.

ARNOLD, V. I., GUSSEIN-ZADE, S. M., & VARCHENKO, A. N. 1985 *Singularities of Differentiable Maps I*. Basel: Birkhäuser.

ARNOLD, V. I. & ILYASHENKO, YU. S. 1988 Ordinary differential equations. *Dynamical Systems I (Encyclopaedia of Mathematical Sciences, vol. 1)* (D. V. Anosov & V. I. Arnold eds). Berlin: Springer, pp. 1–148.

ARPONEN, T. & TUOMELA, J. 1996 On the numerical solution of involutive ordinary differential systems: numerical results. *Research Report A370*, Helsinki University of Technology.

BERZ, M., BISCHOF, C., CORLISS, G., & GRIEWANK, A. (eds). 1996 *Computational Differentiation*. Philadelphia: SIAM.

BOULIER, F., LAZARD, D., OLLIVIER, F., & PETITOT, M. 1995 Representation for the radical of a finitely generated differential ideal. *Proc. ISSAC 1995*, ACM.

BRENAN, K., CAMPBELL, S., & PETZOLD, L. 1989 *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. Amsterdam: North-Holland.

CARRÀ FERRO, G. 1987 Gröbner bases and differential algebra. *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (Lecture Notes in Computer Science 356)*. (L. Huguet & A. Poli eds). Berlin: Springer, pp. 129–140.

CROUCH, P. & GROSSMAN, R. 1993 Numerical integration of ordinary differential equations on manifolds. *Nonlinear Sci.* **3**, 1–33.

DARA, L. 1975 Singularités génériques des équations différentielles multiformes. *Bol. Soc. Bras. Mat.* **6**, 95–128.

DO CARMO, M. 1992 *Riemannian Geometry*. Basel: Birkhäuser.

DUDNIKOV, P. I. & SAMBORSKI, S. N. 1996 Linear overdetermined systems of partial differential equations. *Partial Differential Equations VIII (Encyclopaedia of Mathematical Sciences 65)*. (M. A. Shubin ed). Berlin: Springer, pp. 1–86.

EHRESMANN, C. 1951 Les prolongements d'une variété différentielle I: calcul des jets, prolongement principal. *C. R. Acad. Sci. Paris* **233**, 598–600.

GOLDSCHMIDT, H. 1967 Integrability criteria for systems of non-linear partial differential equations. *J. Diff. Geom.* **1**, 269–307.

HAIRER, E., LUBICH, C., & ROCHE, M. 1989 *The Numerical Solution of Differential-Algebraic Systems by Runge–Kutta Methods (Lecture Notes in Mathematics, vol. 1409)*. Berlin: Springer.

HAIRER, E. & WANNER, G. 1991 *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems (Springer Series in Computational Mathematics 14)*. Berlin: Springer.

ISERLES, A. 1996 Beyond the classical theory of computational ordinary differential equations. *State of the Art in Numerical Analysis*. (I. S. Duff & G. A. Watson eds). Oxford: Oxford University Press.

ISERLES, A. 1997 Numerical methods on (and off) manifolds. *Foundations of Computational Mathematics*. (F. Cucker ed). Berlin: Springer, pp. 180–189.

ISERLES, A. & NØRSETT, S. 1999 On the solution of linear differential equations in Lie groups. *Geometric Integration: Numerical Solution of Differential Equations on Manifolds, vol. 357*

of Philosophical Transactions of the Royal Society A. (C. J. Budd & A. Iserles eds). London Mathematical Society, pp. 983–1020.

KORVOLA, T. 1997 On the formal integrability of differential algebraic systems. *Research Report A381*, Helsinki University of Technology.

KRASILSHCHIK, I. S., LYCHAGIN, V. V., & VINOGRADOV, A. M. 1986 Geometry of jet spaces and nonlinear partial differential equations. *Advanced Studies in Contemporary Mathematics*. vol. 1, Gordon & Breach.

KUMPERA, A. & SPENCER, D. 1972 *Lie Equations General Theory (Annals of Mathematics studies 73)*. Princeton, NJ: Princeton University Press.

KURANISHI, M. 1967 *Lectures on Involutive Systems of Partial Differential Equations*. Publicações da Sociedade de Matemática de São Paulo.

LEIMKUHLER, B. & SKEEL, R. 1994 Symplectic numerical integrators in constrained Hamiltonian systems. *J. Comput. Phys.* **112**, 117–125.

LE VEY, G. 1994 Differential-algebraic equations: a new look at the index. *Research Report 808*, IRISA.

LE VEY, G. 1998 Some remarks on solvability and various indices for implicit differential equations. *Numer. Algor.* **19**, 127–145.

LYCHAGIN, V. V. (ed). 1995 *The Interplay between Differential Geometry and Differential Equations*. Am. Math. Soc. Translations, vol. 167, American Mathematical Society.

MACUTAN, Y. O. & THOMAS, G. 1998 Theory of formal integrability and DAEs: effective computations. *Numer. Algor.* **19**, 147–157.

MANSFIELD, E. 1991 Differential Gröbner bases, *PhD Thesis*, University of Sydney.

MÄRZ, R. 1992 Numerical methods for differential-algebraic equations. *Acta Numer.* **1**, 141–198.

MUNTHE-KAAS, H. 1995 Lie–Butcher theory for Runge–Kutta methods. *BIT* **35**, 572–587.

MUNTHE-KAAS, H. 1999 High order Runge–Kutta methods on manifolds. *Appl. Numer. Math.* **29**, 115–127.

MUNTHE-KAAS, H. & ZANNA, A. 1997 Numerical integration of differential equations on homogeneous manifolds. *Foundations of Computational Mathematics*. (F. Cucker ed). Berlin: Springer, pp. 305–315.

OLVER, P. J. 1995 *Equivalence, Invariants, and Symmetry*. Cambridge: Cambridge University Press.

PIIRILÄ, O.-P. & TUOMELA, J. 1993 Differential-algebraic systems and formal integrability. *Research Report A326*, Helsinki University of Technology, Institute of Mathematics.

POMMARET, J. F. 1978 *Systems of Partial Differential Equations and Lie Pseudogroups (Mathematics and its Applications 14)*. Gordon and Breach.

POMMARET, J. F. 1983 *Differential Galois Theory (Mathematics and its Applications 15)*. Gordon and Breach.

POMMARET, J. F. 1988 *Lie Pseudogroups and Mechanics (Mathematics and its Applications 16)*. Gordon and Breach.

POMMARET, J. F. 1994 *Partial Differential Equations and Group Theory*. Dordrecht: Kluwer.

RABIER, P. & RHEINBOLDT, W. 1991 A general existence and uniqueness theory for implicit differential-algebraic equations. *Diff. Int. Eqns.* **4**, 563–582.

RABIER, P. & RHEINBOLDT, W. 1994 A geometric treatment of implicit differential-algebraic equations. *J. Diff. Eqns.* **109**, 110–146.

RANGARAJAN, G. 1996 Symplectic completion of symplectic jets. *J. Math. Phys.* **37**, 4514–4542.

REICH, S. 1991 On an existence and uniqueness theory for nonlinear differential-algebraic equations. *Circuits, Systems Signal Proc.* **10**, 343–359.

RHEINBOLDT, W. 1984 Differential-algebraic systems as differential equations on manifolds. *Math.*

Comput. **43**, 473–482.

RIQUIER, C. 1910 *Les Systèmes d'Équations aux Dérivées Partielles*. Paris: Gauthier-Villars.

SAUNDERS, D. 1989 *The Geometry of Jet Bundles (London Math. Soc. Lecture Note Series 142)*. Cambridge: Cambridge University Press.

SANZ SERNA, J. M. & CALVO, M. P. 1994 *Numerical Hamiltonian Problems*. London: Chapman & Hall.

SEILER, W. M. 1995 Involution and constrained dynamics II: the Faddeev–Jackiw approach. *J. Phys. A Math. Gen.* **28**, 7315–7331.

SEILER, W. M. 1999 Indices and solvability for general systems of differential equations. To appear. *Computer Algebra in Scientific Computing*, CASC 99 Munich, (V. G. Ghanza, E. W. Mayr, E. V. Vorozthsov eds.) Berlin:Springer, pp. 365–385.

SEILER, W. M. & TUCKER, R. W. 1995 Involution and constrained dynamics I: the dirac approach. *J. Phys. A Math. Gen.* **28**, 4431–4451.

SPENCER, D. 1969 Overdetermined systems of linear partial differential equations. *Bull. Am. Math. Soc.* **75**, 179–239.

SPIVAK, M. 1979 *A Comprehensive Introduction to Differential Geometry vol. 1–5*. 2nd edn., Publish or Perish.

TARKHANOV, N. 1995 *Complexes of Differential Operators*. Dordrecht: Kluwer.

THOMAS, G. 1997 Contributions théoriques et algorithmiques à l'étude des équations différentielles-algébriques. *PhD Thesis*, Institut National Polytechnique de Grenoble.

TUOMELA, J. 1996 On the numerical solution of involutive ordinary differential systems. *Research Report A363*, Helsinki University of Technology.

TUOMELA, J. 1997 On singular points of quasilinear differential and differential-algebraic equations. *BIT* **37**, 966–975.

TUOMELA, J. 1998 On the resolution of singularities of ordinary differential systems. *Numer. Algor.* **19**, 247–259.

TUOMELA, J. & ARPONEN, T. On the numerical solution of involutive ordinary differential systems 2: higher order methods. submitted to BIT.

VERSHIK, A. M. & GERSHKOVICH, V. YA. 1994 Nonholonomic dynamical systems, geometry of distributions and variational problems. *Dynamical Systems VII (Encyclopaedia of Mathematical Sciences 16)*. (V. I. Arnold & S. P. Novikov eds). Berlin: Springer, pp. 1–81.

VINOGRADOV, A. M. 1981 *Math. Rev.* 81f: 58046.

WOLFRAM, S. 1991 *Mathematica: A System for Doing Mathematics by Computer*. 2nd edn., Reading, MA: Addison-Wesley.

ZANNA, A. 1998 Lie group methods for isospectral flows. *PhD Thesis*, University of Cambridge.

ZANNA, A. & MUNTHE-KAAS, H. 1997 Iterated commutators, Lie's reduction method and ordinary differential equations on matrix Lie groups. *Foundations of Computational Mathematics*. (F. Cucker ed). Berlin: Springer, pp. 434–441.