

# A NUMERICAL METHOD FOR FRICTION PROBLEMS WITH MULTIPLE CONTACTS

DAVID E. STEWART<sup>1</sup>

(Received 13 August 1989; revised 31 March 1994)

## Abstract

Friction problems involving “dry” or “static” friction can be difficult to solve numerically due to the existence of discontinuities in the differential equations appearing in the right-hand side. Conventional methods only give first-order accuracy at best; some methods based on stiff solvers can obtain high order accuracy. The previous method of the author [16] is extended to deal with friction problems involving multiple contact surfaces.

## 1. Introduction

Friction problems with “dry” or “static” friction are well-known for being difficult to handle numerically. The problem is that the differential equations (ODEs) describing the motion of the masses in such a system must be discontinuous in the sense that there is a jump discontinuity in the right-hand side of the ODEs as the relative velocity of the surfaces in contact goes through zero. This, in turn, is because the *direction* of the force (being  $\mu N$ ,  $\mu$  being the coefficient of friction,  $N$  being the normal contact force) changes as the relative velocity passes through zero.

The problem of numerically solving discontinuous ODEs has only recently been investigated. Indeed, the first paper describing a systematic notion of “weak solution”, corresponding to Carathéodory’s notion for ODEs, only appeared in 1960 [8]. The first paper on the problem of numerically computing a solution is that of Taubert [18] in 1976. Since then work has been done by Taubert [19], Elliott [6], Niepage [14], Niepage and Wendt [15], Stewart [16] and Kastner-Maresch [12, 13] amongst others. For more information on the short history and development of this work, see the review article by Dontchev and Lempio [5]. Until the paper of Stewart, the best convergence results obtained were that the methods described would be at least first-order provided there were some restrictions on the problem. This was due to two factors: one is the

---

<sup>1</sup>School of Mathematical Sciences, The Australian National University, Canberra, ACT 0200, Australia.  
© Australian Mathematical Society, 1996, Serial-fee code 0334-2700/95

problem of the numerical method “chattering” about the discontinuity, and the other is that switching points need to be accurately located to obtain high accuracy. The method of Stewart assumes and uses considerable information about the structure of the discontinuities, but can give arbitrarily high-order convergence. Kastner-Maresch uses a more conventional approach based on implicit difference formulae, but is also able to obtain high-order convergence results. There is currently a gap in the theory of [12], in that there are no results on maintaining high-order accuracy when the trajectory has a “kink” at a switching point, and is not  $C^1$ .

On the other hand, the structure assumed by the method of Stewart [16] does not always exist; surprisingly it fails in what are otherwise very well-structured discontinuous ODEs arising from several masses making frictional contact. (The case of a single frictional contact is easily handled by the framework in [16].)

In this paper the representation of the discontinuous ODEs of Stewart [16] is modified to a form appropriate for friction problems with multiple bodies in contact. Convergence and related results are also proven which are analogous to those in [16].

This work extends [16] and is based on the author’s doctoral thesis [17].

The paper is structured as follows. Section 2 contains previous results and sets the framework for the remainder of the paper. Section 3 outlines how to extend Algorithm 1 of [16]; this section includes the extended results on LCPs needed for the main results. Section 4 gives a formal description of the main algorithm, and states the main results and outlines for the proofs. Section 5 describes the implementation and gives numerical results for a test friction problem.

Since the following material deals with set-valued functions, the notation  $\mathcal{P}(X)$  is used to denote the set of subsets of  $X$ , and a set valued map  $F$  from  $X$  to  $Y$  is an ordinary function  $F: X \rightarrow \mathcal{P}(Y)$ .

## 2. Previous work

Here we summarise the results of Stewart [16] and relevant information about discontinuous ODEs.

In Carathéodory’s definition of a solution of a differential equation [3]

$$x' = f(t, x), \quad (1)$$

the function  $x(\cdot)$  must be absolutely continuous (AC) and satisfy (1) at almost all values of  $t$ . Under weak continuity conditions of  $f(t, x)$  in  $x$  and integrability in  $t$ , Carathéodory showed that solutions existed, at least locally. However, this is not so for discontinuous ODEs as the following example illustrates:

$$x' = -\operatorname{sgn} x + g(t), \quad x(0) = 0$$

where  $g(\cdot)$  is a nonconstant function of time and satisfies  $|g(t)| < 1$  for all  $t$ . In this example, we cannot have  $x(t) > 0$  for any  $t > 0$  as this would imply the existence of a time  $t'$  at which  $x(t') > 0$  and  $x'(t') > 0$ , which is clearly impossible. Similarly we cannot have  $x(t) < 0$  for any  $t > 0$ . Thus we must have  $x(t) \equiv 0$ . Substituting this into the differential equation gives  $0 = -\text{sgn } 0 + g(t)$ , which cannot be true almost everywhere (even with a redefinition of “ $\text{sgn } 0$ ”) for nonconstant  $g$ .

Instead we use the regularisation of Filippov [8] which replaces the differential equation with a *differential inclusion*:

$$x' \in \mathcal{F} f(t, x) = F(t, x), \tag{2}$$

where

$$F(t, x) = \mathcal{F} f(t, x) = \bigcap_{\delta > 0} \bigcap_{\mu(N)=0} \overline{\text{co}} f(t, (x + \delta B) \setminus N)$$

and  $\mu$  is the Lebesgue measure,  $B$  is a unit ball and  $\overline{\text{co}}X$  is the closed convex hull of  $X$ .

For structure, it was assumed in [16] that there are smooth functions  $f_i$  and subsets  $R_i \subseteq \mathbb{R}^n, i = 1, 2, \dots, m$ , where

$$x' = f(x) = f_i(x) \quad \text{if } x \in R_i \tag{3}$$

and

$$\mathbb{R}^n = \overline{\bigcup_{i=1}^m R_i}.$$

These regions  $R_i$  should have nonempty interior and have piecewise smooth boundaries. At a point  $x \in \mathbb{R}^n$  the *active set* is  $I(x) = \{i \mid x \in \overline{R}_i\}$ . The regions  $R_i$  are assumed to be described by *indicator functions*  $h_i: \mathbb{R}^n \rightarrow \mathbb{R}$ , where

$$R_i = \{x \mid h_i(x) < h_j(x) \text{ for all } j \neq i\}. \tag{4}$$

This form of representation comes from optimal control theory.

For single-contact friction problems suitable indicator functions are easy to find:  $h_1(x) = -v, h_2(x) = +v$ , where  $v$  is the relative velocity of the bodies in contact. The functions  $f_1(x)$  and  $f_2(x)$  are then the right-hand sides for the differential equations with an appropriate choice of direction of the friction force. Note that the discontinuities that appear in the original ODEs for the friction do not appear in  $f_1$  or  $f_2$ .

Given this structure, we look for *piecewise-active* solutions of the discontinuous ODE; that is, the active set  $I(x(t))$  is a piecewise-constant function of  $t$ . In the pieces where  $I(x(t))$  is constant, an ODE can be set up which maintains the given active set and solves the discontinuous ODE. This ODE is determined by taking the intersection

of  $\mathcal{F}f(t, x)$  and the tangent plane of the surface of discontinuity. This ODE is smooth provided the intersection is nonempty and the intersection of the affine plane of  $\mathcal{F}f(t, x)$  and the tangent plane is transversal [11]. This can be tested by checking that an appropriate matrix is nonsingular. Where the active set changes we need to solve a combinatorial problem of determining what new active sets are possible. This is done in [16] by solving a linear complementarity problem (LCP). Each solution of this LCP uniquely determines the active set provided a nondegeneracy condition holds.

Unfortunately, neither LCPs nor discontinuous ODEs have unique solutions in general. (LCPs in general might not have *any* solutions; however, the LCPs that arise from discontinuous ODEs are guaranteed to have solutions.) One point that can be made is that if the LCP does not have a unique solution, then the discontinuous ODE from which it came does not have a unique solution either. This means that in the case of friction problems, where it is known *a priori* that solutions are unique from a one-sided Lipschitz condition, it follows that the LCPs that arise from these systems have unique solutions and so there is no need to somehow “guess” the correct solution. There are other situations, such as the direct solution of the Pontryagin conditions, that lead to discontinuous ODEs that do not have unique solutions to the initial value problem.

**2.1. Conditions and results** A discontinuous ODE of the form discussed by Stewart [16] is defined by  $f_i: \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $h_i: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, 2, \dots, m$ . It is assumed that both sets of functions are smooth. The following assumptions are made.

CONDITION A: For each  $x \in \mathbb{R}^n$ , the set  $\{\nabla h_i(x) \mid i \in I(x)\}$  is *geometrically independent*. That is, the affine plane generated by the set is not generated by any strict subset.

This is a generic condition on the regions  $R_i$ .

CONDITION B: For each  $x \in \mathbb{R}^n$ , the matrix  $M(x) + \alpha ee^T$ , defined by  $m_{ij}(x) = \nabla h_i(x)^T f_j(x)$ ,  $i, j \in I(x)$  and  $e = [1, 1, \dots, 1]^T$ , is nonsingular for some  $\alpha > 0$ .

The final condition is in terms of linear complementarity problems (LCPs).

DEFINITION 1. The LCP denoted  $\text{LCP}(M, q)$  with  $M \in \mathbb{R}^{m \times m}$  and  $q \in \mathbb{R}^m$  is the problem of finding  $z, w \in \mathbb{R}^m$  such that

$$Mz + q = w \geq 0, \quad z \geq 0, \quad z^T w = 0.$$

Note that the vector inequalities are understood to hold for each *component* of the vectors.

CONDITION C: For each  $x \in \mathbb{R}^n$ ,  $\text{LCP}(M(x) + \alpha ee^T, -e)$  for sufficiently large  $\alpha > 0$  only has solutions which are *strictly complementary*; that is, for each  $i \in I(x)$  either  $z_i > 0$  or  $w_i > 0$ .

This allows the unique determination of the new active set from a given solution of  $\text{LCP}(M(x) + \alpha ee^T, -e)$  given by  $\{i \mid z_i > 0\} = I_{\text{new}} = \{i \mid w_i = 0\}$ .

The two main results of [16] follow. Algorithm 1 is given in the Appendix and is “Algorithm 1” of [16].

**THEOREM 1.** *Assuming conditions A and B, if we have a sequence of approximations  $\hat{x}_h$  on  $[t_0, t_f]$  generated by Algorithm 1, then a limit point exists and furthermore, all limit points are solutions of the discontinuous ODE in the sense of Filippov.*

In the following result  $\omega(h)$  is the order of accuracy of the (smooth) ODE solver used for the regularised right-hand side. The quantity  $\epsilon(h)$  is the tolerance for approximating  $I(x)$ ;  $I_\epsilon(x) = \{i \mid h_i(x) < \min_j h_j(x) + \epsilon\}$ . The set  $I_\epsilon(x)$  is also called the  $\epsilon$ -active set. The quantity  $\eta(h)$  is the tolerance for locating zeros of switching functions along the (numerical) trajectory.

**THEOREM 2.** *Suppose that  $x(\cdot)$  is a piecewise active solution of the discontinuous ODE in the sense of Filippov. Then assuming conditions A, B and C, for  $h > 0$  sufficiently small, suitable choices of active sets in Algorithm 1 generates numerical approximations  $\hat{x}_h(\cdot)$ , where*

$$\|\hat{x}_h - x\|_\infty = O(\omega(h)),$$

provided  $\omega(h) = o(\epsilon(h))$  and  $\eta(h) = O(\omega(h))$ .

### 3. A decomposition extension

The algorithm described in Stewart [16] is inadequate to handle apparently simple discontinuous ODEs, for example

$$\begin{aligned} x_1' &= -\text{sgn } x_1, \\ x_2' &= -\text{sgn } x_2. \end{aligned} \tag{5}$$

The purpose of this section is to develop an approach for such problems that encompasses the theory developed in Stewart [16]. Such a generalisation would be particularly appropriate for problems involving friction problems with multiple friction surfaces.

Notice that in (5) there are four regions on which the right-hand side is smooth:

$$\begin{aligned} R_1 &= \{ (x_1, x_2) \mid x_1 > 0, x_2 > 0 \} & R_2 &= \{ (x_1, x_2) \mid x_1 > 0, x_2 < 0 \} \\ R_3 &= \{ (x_1, x_2) \mid x_1 < 0, x_2 > 0 \} & R_4 &= \{ (x_1, x_2) \mid x_1 < 0, x_2 < 0 \}. \end{aligned}$$

Using the formulation of the previous section to solve this problem leads to the problem of solving

$$M_\alpha(x)\hat{z} = e$$

for  $\hat{z}$ , where  $m_{ij}(x) = \nabla h_i(x) \cdot f_j(x)$ . Thus  $M(x) = \nabla h(x) \cdot F(x)$ , where  $\nabla h(x)$  is a  $4 \times 2$  matrix and  $F(x)$  is a  $2 \times 4$  matrix. The rank of  $M(x)$  is therefore not more than 2, and as  $M_\alpha(x) = M(x) + \alpha ee^T$  is a rank-1 modification, the rank of  $M_\alpha(x)$  is not more than 3. Since  $M_\alpha(x)$  is a  $4 \times 4$  matrix, it is singular and the method of Stewart [16] will fail.

The main problem in (5), strangely enough, seems to lie with the fact that the “ $-\text{sgn } x_1$ ” and “ $-\text{sgn } x_2$ ” functions are entirely independent of each other. To be more precise, one can write

$$x' = f(x) = f^1(x) + f^2(x),$$

where  $f^1(x, t) = [-\text{sgn } x_1, 0]^T$ ,  $f^2(x, t) = [0, -\text{sgn } x_2]^T$ . At  $x_1 = x_2 = 0$ , we should only need to compute 2 quantities, one for each of the two discontinuous function  $f^1$  and  $f^2$ . That is,

$$x' \in \begin{bmatrix} [-1, +1] \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ [-1, +1] \end{bmatrix}.$$

It seems then that there are more unknowns than are necessary for problems of this sort, which leads to the difficulties with singular matrices.

Using a decomposition type of approach, we will develop a more general form for representing discontinuous ODEs that effectively reduces the number of unknowns to be solved for. By reworking the theory in terms of this new representation, a larger class of problems becomes solvable by essentially the same methods as are developed in Stewart [16]. That is, active sets are determined through solutions of LCPs, and that the differential inclusion is replaced by a suitable ODE by selecting the element of the right-hand side to maintain the given active set. For this new representation we will, of course, have to check that the resulting systems of equations can be expected to have solutions and that the LCPs to be solved do indeed have solutions. This we will do later in this section.

**3.1. The new representation** In the above example (5), instead of dealing with the regions  $R_1, \dots, R_4$ , we can work in terms of

$$R_1^1 = \{ (x_1, x_2) \mid x_1 > 0 \}, \quad R_2^1 = \{ (x_1, x_2) \mid x_1 < 0 \},$$

$$R_1^2 = \{ (x_1, x_2) \mid x_2 > 0 \}, \quad R_2^2 = \{ (x_1, x_2) \mid x_2 < 0 \}.$$

Then

$$f^1(x) = \begin{cases} [-1, 0]^T & \text{if } x \in R_1^1 \\ [+1, 0]^T & \text{if } x \in R_2^1, \end{cases} \quad f^2(x) = \begin{cases} [0, -1]^T & \text{if } x \in R_1^2 \\ [0, +1]^T & \text{if } x \in R_2^2. \end{cases}$$

Then for  $x \in R_1^1 \cap R_1^2$  we have the smooth ODE

$$x' = \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix},$$

but for  $x \in R_1^1 \cap R_2^2$  we have

$$x' = \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ +1 \end{bmatrix}.$$

Clearly, for this example, we can write

$$x' = \sum_{j=1}^m f^j(x),$$

where  $f^j(x) = f_k^j(x) \quad \text{for } x \in R_k^j. \tag{6}$

For (5) we can put

$$f_1^1(x_1, x_2) = [-1, 0]^T, \quad f_2^1(x_1, x_2) = [+1, 0]^T,$$

$$f_1^2(x_1, x_2) = [0, -1]^T, \quad f_2^2(x_1, x_2) = [0, +1]^T.$$

More generally we consider discontinuous ODEs of the form (6), where the  $f_k^j$  are smooth functions. We further assume that there are indicator functions  $h_k^j: \mathbb{R}^n \rightarrow \mathbb{R}$  for  $k = 1, \dots, m_j$  such that

$$R_k^j = \{ x \mid h_k^l(x) < h_l^j(x) \text{ for all } l = 1, \dots, m_j, l \neq k \} \tag{7}$$

by analogy with Section 2. For (5) we can use the indicator functions

$$h_1^1(x_1, x_2) = +x_1 \quad h_2^1(x_1, x_2) = -x_1$$

$$h_1^2(x_1, x_2) = +x_2 \quad h_2^2(x_1, x_2) = -x_2.$$

As above, the discontinuous ODE (6) is understood in the Filippov sense (2). Applying the arguments of [16, Lemma 1.3] to (6) and (7) gives

$$x' \in \sum_{j=1}^m \text{co}\{f_k^j(x) \mid k \in I^j(x)\}, \quad (8)$$

where

$$I^j(x) = \{k \mid h_k^j(x) = \min_i h_i^j(x)\},$$

provided the set  $\{\nabla h_i^j(x) \mid j = 1, \dots, m; i \in I^j(x)\}$  is geometrically independent. (See Section 2.1.) In this case the “active set” is not actually a set, but rather a finite ordered sequence of sets

$$I(x) = (I^1(x), I^2(x), \dots, I^m(x))$$

with  $I^j(x) \subset \{1, \dots, m_j\}$ . In the remainder of this section we seek solutions of (8).

**3.2. Piecewise constant “active sets”** Following the same programme as laid down in [16], we consider the class of solutions of (8) with a piecewise-constant “active set”. That is, we assume that there is a finite sequence of switching points

$$t_0 < t_1 < \dots < t_N = t_f,$$

where  $I(x(t)) = I_{r-1}$  for  $t \in (t_{r-1}, t_r)$ ,  $r = 1, 2, \dots, N$ .

Let  $(t', t'')$  be an interval on which the “active set” is constant for  $x(\cdot)$ . That is, we seek solutions  $x(\cdot)$  on  $(t', t'')$ , where  $I(x(t)) = I$  for all  $t \in (t', t'')$ . That is,  $I = (I^1, I^2, \dots, I^m) = (I^1(x(t)), I^2(x(t)), \dots, I^m(x(t))) = I(x(t))$  is the *current* active set. Then by Filippov’s implicit function theorem [7], if  $x'(t)$  exists it must be equal to

$$x'(t) = \sum_{q=1}^m \sum_{p \in I^q} z_p^q(t) f_p^q(x(t)) \quad (9)$$

for some measurable  $z_p^q$  satisfying the side-conditions  $z_p^q(t) \geq 0$  and  $\sum_{p \in I^q} z_p^q(t) = 1$ . On the other hand, if  $i, k \in I^j = I^j(x(t))$  then  $h_i^j(x(t)) = h_k^j(x(t))$  for all  $t \in (t', t'')$ . Differentiating this relationship gives

$$\nabla h_i^j(x(t)) \cdot x'(t) = \nabla h_k^j(x(t)) \cdot x'(t) = \mu_j(t),$$

where the  $\mu_j(t)$  are yet to be determined. Combining this with (9) gives the system of linear equations

$$\mu_j(t) = \sum_{q=1}^m \sum_{p \in I^q} \nabla h_i^j(x(t)) \cdot f_p^q(x(t)) z_p^q(t) \quad \text{for } j = 1, \dots, m, \quad i = 1, \dots, m_j. \quad (10)$$



We now wish to fix  $t \in (t', t'')$  and  $x$  and focus on the algebraic aspects of this problem. In what follows we will show the dependence on  $x$  only when it is needed to make explicit which point on the trajectory is being considered. Put  $m_{ip}^{jq} = \nabla h_i^j \cdot f_p^q$ . Then we can form the matrices  $M^{jq} = [m_{ip}^{jq} \mid i \in I^j, p \in I^q]$  and

$$M = \begin{bmatrix} M^{11} & M^{12} & \dots & M^{1m} \\ M^{21} & M^{22} & \dots & M^{2m} \\ \vdots & \vdots & \ddots & \vdots \\ M^{m1} & M^{m2} & \dots & M^{mm} \end{bmatrix}.$$

We also form the vectors  $z^q = [z_p^q \mid p \in I^q]$  and  $z = [(z^1)^T, (z^2)^T, \dots, (z^m)^T]^T$ . Then (10) can be rewritten in more convenient form

$$Mz = \begin{bmatrix} \mu_1 e^1 \\ \mu_2 e^2 \\ \vdots \\ \mu_m e^m \end{bmatrix},$$

where  $e^j = [1, \dots, 1]^T$  with dimension card  $I^j$  and  $\mu_j \in \mathbb{R}$ . The problem remains to determine the  $\mu_j$ 's. It turns out that using the matrix  $M_\alpha = M + \alpha ee^T$  enables this to be done in a way similar to [16].

Let  $\alpha \in \mathbb{R}$  be chosen so that  $M_\alpha = M + \alpha ee^T > 0$ ; that is,  $M_\alpha$  is a matrix with strictly positive entries. Noting that  $z \geq 0$  and that  $(e^q)^T z^q = 1$  for  $q = 1, \dots, m$ , it follows that  $e^T z = \sum_{q=1}^m (e^q)^T z^q = m$ . Thus

$$M_\alpha z = \begin{bmatrix} (\mu_1 + m\alpha)e^1 \\ \vdots \\ (\mu_m + m\alpha)e^m \end{bmatrix}.$$

Put  $\tilde{\mu}_j = \mu_j + m\alpha$ . We add  $\tilde{\mu} = [\tilde{\mu}_1, \dots, \tilde{\mu}_m]^T$  to the vector of unknowns, and add the normalisation conditions  $(e^q)^T z^q = 1$  to the set of equations.

Then we obtain

$$\begin{bmatrix} M_\alpha^{11} & \dots & M_\alpha^{1m} & -e^1 & & \\ \vdots & \ddots & \vdots & & \ddots & \\ M_\alpha^{m1} & \dots & M_\alpha^{mm} & & & -e^m \\ (e^1)^T & & & & & \\ & \ddots & & & & \\ & & (e^m)^T & & & O \end{bmatrix} \begin{bmatrix} z^1 \\ \vdots \\ z^m \\ \tilde{\mu}_1 \\ \vdots \\ \tilde{\mu}_m \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

If we let  $E$  be the matrix

$$\begin{bmatrix} e^1 & & O \\ & \ddots & \\ O & & e^m \end{bmatrix},$$

then this can be written in a much more compact form

$$\begin{bmatrix} M_\alpha & -E \\ E^\top & O \end{bmatrix} \begin{bmatrix} z \\ \tilde{\mu} \end{bmatrix} = \begin{bmatrix} 0 \\ e \end{bmatrix}. \tag{11}$$

As long as the matrix  $\begin{bmatrix} M_\alpha & -E \\ E^\top & O \end{bmatrix}$  is invertible, one can compute the  $z$  vectors in a stable way and use (9) as the definition of a standard ODE, which can be solved by accurate standard methods. This leads to the following generalisations of condition B and B2.

**DEFINITION 2.** If  $I, J \in \prod_{k=1}^m \mathcal{P}(\{1, \dots, m_k\})$ , we say that  $J \subseteq I$  if  $J^k \subseteq I^k$  for  $k = 1, \dots, m$ . We also say that  $J$  is **null-free** if  $J^k \neq \emptyset$  for any  $k = 1, \dots, m$ . The **cardinality** of  $J$  is  $\text{card } J = \sum_{k=1}^m \text{card } J^k$ . The matrix  $M(J, x) = [m_{ip}^{jq}(x) \mid i \in J^j, p \in J^q]$  and  $E(J)$  is the matrix

$$\begin{bmatrix} e^1(J) & & & O \\ & e^2(J) & & \\ & & \ddots & \\ O & & & e^m(J) \end{bmatrix},$$

where  $e^k(J)$  is the column vector of 1's of size  $\text{card } J^k$ . Finally,  $M_\alpha(J, x) = M(J, x) + \alpha ee^\top$ , where  $e$  is the vector of 1's of size  $\text{card } J$ .

**CONDITION B':** Condition  $B'$  is satisfied at  $x$  if for all  $J \subseteq I(x)$ , where  $J$  is null-free, the matrix

$$N_\alpha(J, x) = \begin{bmatrix} M_\alpha(J, x) & -E(J) \\ E(J)^\top & O \end{bmatrix}$$

is nonsingular.

**LEMMA 1.** *The set of matrices  $M \in \mathbb{R}^{n \times n}$  for which  $N_\alpha = \begin{bmatrix} M_\alpha & -E \\ E^\top & O \end{bmatrix}$  is invertible is independent of  $\alpha$ .*

**PROOF.** Suppose that  $N_\alpha$  is nonsingular. We now show that  $N_{\alpha+\beta}$  must also be nonsingular by the Sherman-Morrison formula [10, page 3]. In particular, note that

$$N_{\alpha+\beta} = \begin{bmatrix} M_{\alpha+\beta} & -E \\ E^\top & O \end{bmatrix} = \begin{bmatrix} M_\alpha & -E \\ E^\top & O \end{bmatrix} + \beta \begin{bmatrix} e \\ 0 \end{bmatrix} \begin{bmatrix} e^\top \\ 0 \end{bmatrix},$$

where  $e$  is a column vector of 1s of dimension  $l$ . By the Sherman-Morrison formula,

$$N_{\alpha+\beta}^{-1} = N_{\alpha}^{-1} - \frac{\beta N_{\alpha}^{-1} \begin{bmatrix} e \\ 0 \end{bmatrix} \begin{bmatrix} e \\ 0 \end{bmatrix}^T N_{\alpha}^{-1}}{1 + \beta \begin{bmatrix} e \\ 0 \end{bmatrix}^T N_{\alpha}^{-1} \begin{bmatrix} e \\ 0 \end{bmatrix}}$$

provided the denominator is not zero. But

$$N_{\alpha} \begin{bmatrix} 0 \\ -e' \end{bmatrix} = \begin{bmatrix} M_{\alpha} & -E \\ E^T & O \end{bmatrix} \begin{bmatrix} 0 \\ -e' \end{bmatrix} = \begin{bmatrix} Ee' \\ 0 \end{bmatrix} = \begin{bmatrix} e \\ 0 \end{bmatrix},$$

where  $e'$  is a column vector of 1s of dimension  $m$ . Thus

$$\begin{bmatrix} e \\ 0 \end{bmatrix}^T N_{\alpha}^{-1} \begin{bmatrix} e \\ 0 \end{bmatrix} = \begin{bmatrix} e \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ -e' \end{bmatrix} = 0$$

and the denominator in the Sherman-Morrison formula is always 1.

Thus if  $N_{\alpha}$  is nonsingular, so is  $N_{\alpha+\beta}$  for any  $\beta \in \mathbb{R}$ . The same argument also shows the reverse implication: if  $N_{\alpha+\beta}$  is nonsingular, so is  $N_{\alpha}$ . Thus the invertibility of  $N_{\alpha}$  is independent of  $\alpha$ .

**REMARK 1.** It may be asked what the point of having an  $\alpha$  is if it has no effect on the invertibility of the matrices  $N_{\alpha}$ . There are two main reasons: it is a clearer generalisation to the results of the previous section, and more importantly, the corresponding Linear Complementarity Problem requires  $M_{\alpha} > 0$ . (See Section 3.3 below.) However, a numerical implementation would probably ignore the role of  $\alpha$  except for the LCP computations.

**3.3. At a switching point** While it is not difficult to find *conditions* that a “new active set” would have to satisfy, what is harder to see is whether or not the conditions are satisfiable. These conditions will be shown to be equivalent to finding a solution to an LCP, and solutions to such an LCP will be shown to exist, analogously to [16].

We now consider a switching point  $t'$ . We suppose that  $I(x(t')) = I_0$  and that  $I(x(t)) = I$  for all  $t \in (t', t'')$ . Note that  $I_0 = (I_0^1, \dots, I_0^m)$  and  $I = (I^1, \dots, I^m)$ . In what follows we will develop a method of computing  $I$  using Linear Complementarity Problems by analogy with Stewart [16]. The existence results are constructive and based on the complementary pivoting algorithm of Cottle and Dantzig [4]. The main result of Cottle and Dantzig [4] is that if  $M$  is a *co-positive plus* matrix (that is,  $x \geq 0$  and  $x^T Mx = 0$  implies  $(M + M^T)x = 0$ ), then for any given  $q$  the complementary pivoting algorithm either terminates at a solution to  $\text{LCP}(M, q)$  or at an unbounded feasible ray.

It is easy to see that  $I \subseteq I_0$ . Now if  $i, k, l \in I_0^j$  and  $i, k \in I^j$ , but  $l \notin I^j$ , then

$$\begin{aligned} h_i^j(x(t')) &= h_k^j(x(t')) = h_l^j(x(t')) && \text{but} \\ h_i^j(x(t)) &= h_k^j(x(t)) < h_l^j(x(t)) && \text{for all } t \in (t', t''). \end{aligned}$$

As in [16] we assume that  $D^+x(t') = \lim_{h \downarrow 0} \frac{x(t'+h) - x(t')}{h}$  exists, so that

$$\mu_j = \nabla h_i^j(x(t')) \cdot D^+x(t') = \nabla h_k^j(x(t')) \cdot D^+x(t') \leq \nabla h_l^j(x(t')) \cdot D^+x(t').$$

However

$$D^+x(t') = \sum_{q=1}^m \sum_{p \in I^q} z_p^q f_p^q(x(t')) \in \sum_{q=1}^m \text{co}\{f_p^q(x(t')) \mid p \in I^q\}.$$

It should be noted that the  $z_p^q$  satisfy the usual nonnegativity conditions  $z_p^q \geq 0$ , and normalisation conditions  $\sum_{p \in I^q} z_p^q = 1$ . Combining these two facts about  $D^+x(t')$  gives

$$\begin{aligned} \mu_j &= \sum_{p \in I^q} (\nabla h_i^j(x(t')) \cdot f_p^q(x(t')) z_p^q) = \sum_{p \in I^q} (\nabla h_k^j(x(t')) \cdot f_p^q(x(t')) z_p^q) \\ &\leq \sum_{p \in I^q} (\nabla h_l^j(x(t')) \cdot f_p^q(x(t')) z_p^q), \end{aligned} \tag{12}$$

where the sum is taken over all pairs  $(p, q)$  with  $p \in I^q$ . These sums can be extended to all of  $I_0^q$  by setting  $z_p^q = 0$  whenever  $p \in I_0^q \setminus I^q$ .

To form the LCP, put  $w_i^j = \sum_{p \in I_0^q} (\nabla h_i^j(x(t')) \cdot f_p^q(x(t')) z_p^q) - \mu_j$  and the vectors  $w^j = [w_i^j \mid i \in I_0^j]$ ,  $w = [(w^1)^T, \dots, (w^m)^T]^T$ . The complementarity condition holds as if  $p \in I^q$  then  $w_p^q = 0$ , and if  $p \in I_0^q \setminus I^q$ , then  $z_p^q = 0$ . Then

$$w^T z = \sum_{j=1}^m (w^j)^T z^j = 0.$$

As above, put  $m_{ip}^{jq} = \nabla h_i^j(x(t')) \cdot f_p^q(x(t'))$  and define the matrices  $M^{jq} = [m_{ip}^{jq} \mid i \in I_0^j, p \in I_0^q]$  and  $M = [M^{jq} \mid j, q = 1, \dots, m]$ . Then (12) becomes

$$w = Mz - \begin{bmatrix} \mu_1 e^1 \\ \vdots \\ \mu_m e^m \end{bmatrix} \geq 0.$$

We would like to incorporate the  $\mu$ 's into the LCP as variables, but this is not yet possible as we cannot yet guarantee that they are nonnegative. Hence we replace  $M$  by  $M_\alpha = M + \alpha ee^T$  to obtain

$$w = M_\alpha z - \begin{bmatrix} \tilde{\mu}_1 e^1 \\ \vdots \\ \tilde{\mu}_m e^m \end{bmatrix} \geq 0,$$

where  $\tilde{\mu}_j = \mu_j + m\alpha$ , as above. Let  $E$  be the matrix as given above and introduce the column vector  $\tilde{\mu} = [\tilde{\mu}_j \mid j = 1, \dots, m]$  and a vector of complementary variables  $\beta = [\beta_j \mid j = 1, \dots, m]$  which are defined by  $\beta_j = (e^j)^T z^j - 1$ . This forms the following LCP: Find  $(z, \tilde{\mu}, w, \beta)$  such that

$$\begin{bmatrix} M_\alpha & -E \\ E^T & O \end{bmatrix} \begin{bmatrix} z \\ \tilde{\mu} \end{bmatrix} - \begin{bmatrix} 0 \\ e \end{bmatrix} = \begin{bmatrix} w \\ \beta \end{bmatrix} \geq 0, \quad \begin{bmatrix} z \\ \tilde{\mu} \end{bmatrix} \geq 0, \quad \begin{bmatrix} w \\ \beta \end{bmatrix}^T \begin{bmatrix} z \\ \tilde{\mu} \end{bmatrix} = 0. \quad (13)$$

Two questions naturally arise at this point: First, do solutions exist for this problem? Secondly, if  $(z, \tilde{\mu}, w, \beta)$  solves (13), does  $z$  satisfy all the conditions required above? Certainly the  $z_p^q \geq 0$  conditions are satisfied, but it is not immediately clear that  $\sum_{p \in I_0^q} z_p^q = 1$ . We now answer both these questions affirmatively.

LEMMA 2. *If  $M_\alpha > 0$  then solutions to (13) exist, and furthermore, every solution has  $\beta = 0$ , and therefore  $(e^j)^T z^j = 1$  for all  $j$ .*

PROOF. We first prove that (13) has solutions. Since  $N_\alpha$  is co-positive plus, either Cottle and Dantzig's complementary pivoting algorithm will find a solution, or the algorithm will terminate at a feasible complementary ray. Let this feasible ray consist of  $(z, \tilde{\mu}) = (z^0, \tilde{\mu}^0) + \theta(\tilde{z}, \tilde{\nu})$  for all  $\theta \geq 0$  with  $(\tilde{z}, \tilde{\nu}) \neq 0$ . Taking  $\theta \rightarrow +\infty$  gives

$$\begin{bmatrix} \tilde{z} \\ \tilde{\nu} \end{bmatrix}_i \left( \begin{bmatrix} M_\alpha & -E \\ E^T & O \end{bmatrix} \begin{bmatrix} \tilde{z} \\ \tilde{\nu} \end{bmatrix} \right)_i \leq 0$$

for all  $i$ . Then in particular,  $\tilde{\nu}_j((e^j)^T \tilde{z}^j) \leq 0$  for all  $j = 1, \dots, m$ . Hence, as  $\tilde{z}^j \geq 0$  and  $\tilde{\nu} \geq 0$ , either  $\tilde{\nu}_j = 0$  or  $\tilde{z}^j = 0$ . On the other hand, we also have

$$(\tilde{z}^j)^T \left( \sum_{i=1}^m M_\alpha^{ij} \tilde{z}^i - \tilde{\nu}_j e^j \right) \leq 0.$$

We now show that  $\tilde{z}^j = 0$ . This has already been shown in the case  $\tilde{\nu}_j > 0$ . If  $\tilde{\nu}_j = 0$  and  $\tilde{z}^j \neq 0$  then  $(\tilde{z}^j)^T \sum_{i=1}^m M_\alpha^{ij} \tilde{z}^i \leq 0$ , and as  $M_\alpha^{ij} > 0$  for every  $i, j$ , this implies that  $\tilde{z}^i = 0$  for every  $i$ . Thus  $\tilde{z}^j = 0$ .

As  $\bar{z}^j = 0$  for every  $j$ , we find  $\bar{z} = 0$ . But if we take  $\theta \rightarrow +\infty$  along this complementary ray, we also find that

$$\begin{bmatrix} \bar{w} \\ \bar{\beta} \end{bmatrix} = \begin{bmatrix} M_\alpha & -E \\ E^\top & O \end{bmatrix} \begin{bmatrix} 0 \\ \bar{v} \end{bmatrix} \geq 0.$$

Therefore  $\bar{w} = -E\bar{v} \geq 0$ . As  $E$  is a full rank matrix of zeros and ones and  $\bar{v} \geq 0$ , it follows that  $\bar{v} = 0$ . Thus the direction of the complementary ray is along  $(\bar{z}, \bar{v}, \bar{z}_0) = (0, 0, \bar{z}_0)$ . This is the complementary ray associated with the initial tableau of the complementary pivoting algorithm, which means that the basis of the terminating tableau is the same as that of the initial tableau. But this is impossible by Cottle and Dantzig [4]. Hence the complementary pivoting algorithm cannot terminate at a complementary ray and must terminate at a solution.

We now proceed to show that at a solution  $(z, \tilde{\mu}, w, \beta)$ ,  $\beta = 0$  and  $(e^j)^\top z^j = 1$ .

First, by complementarity,  $w^\top z + \beta^\top \tilde{\mu} = 0$ . In particular, either  $\beta_j = 0$  or  $\tilde{\mu}_j = 0$ . Suppose  $\beta_j \neq 0$ . Then  $\tilde{\mu}_j = 0$  and  $w^j = \sum_{i=1}^m M_\alpha^{ji} z^i$ . If  $z \neq 0$  then  $w^j > 0$  which implies that  $z^j = 0$  by complementarity, and that  $\beta_j = (e^j)^\top z^j - 1 = -1$  which is infeasible. If  $z = 0$ , then again  $\beta_j = (e^j)^\top z^j - 1 = -1$  which is infeasible. Hence  $\beta_j = 0$ . As  $\beta_j = (e^j)^\top z^j - 1$ , it follows that  $(e^j)^\top z^j = 1$  at a solution of (13).

Again assuming that the solutions to such a complementarity problem are strictly complementary, the “new active set” is completely determined by whatever solution is chosen through

$$I^j = \{i \mid z_i^j > 0\} \neq \emptyset. \tag{14}$$

Thus the method can then be restarted with this “active set”  $I = (I^1, \dots, I^m)$ . In order to guarantee strict complementarity we use the following analogue of condition C.

CONDITION C': Condition C' is satisfied at  $x$  if  $\begin{bmatrix} 0 \\ e \end{bmatrix}$  cannot be expressed as a linear combination of  $\text{card } I(x) + m - 1$  columns of  $\begin{bmatrix} M_\alpha(I(x), x) & -E(I(x)) \\ E(I(x))^\top & O \end{bmatrix}$ , where  $\alpha$  is some number in  $\mathbb{R}$  chosen so that  $M_\alpha(I(x), x)$  has only positive entries.

CONDITION C2': Condition C2' is satisfied by (6)–(8) if condition C' is satisfied at every  $x \in \mathbb{R}^n$ .

#### 4. Algorithm and convergence

We now give a modification of Algorithm 1 for solving discontinuous ODEs in this new representation. We will then give generalisations of Theorems 4.2 and 4.3 of Stewart [16].

We also need the following generalisation of  $I_\epsilon(x)$ :  $I_\epsilon(x) = (I_\epsilon^1(x), \dots, I_\epsilon^m(x))$ , where

$$I_\epsilon^k(x) = \{i \mid h_i^k(x) < \min_j h_j^k(x) + \epsilon\}.$$

Finally we need the switching function

$$\psi(I, x, t) = \min_{p=1, \dots, m} (\min_{i \in I^p} z_i^p, \min_{j \notin I^p} h_j^p(x) - \min_{i \in I^p} h_i^p(x), t^* - t),$$

where  $z_i^p$  is defined by  $(I, x)$  through (11).

Algorithm 2 for computing solutions to (6)–(8) is a modification of Algorithm 1, and is presented in the Appendix.

**THEOREM 3.** *If we have a sequence of approximations  $\hat{x}_h$  on  $[t_0, t_f]$  generated by Algorithm 2 as  $h \downarrow 0$ , then a limit point must exist, and furthermore, all limit points are solutions of (6)–(8) in the sense of Filippov.*

**PROOF.** As for [16, Theorem 4.2].

**THEOREM 4.** *Suppose  $x(\cdot)$  is a solution of the discontinuous ODE (6)–(8) and conditions  $B'$  and  $C'$  are satisfied for all  $x(t)$ . If  $x(\cdot)$  is a piecewise-constant active set solution, then Algorithm 2 can be made to give, by suitable choices of  $t^*$  and  $I$  in Step 5, numerical approximations  $\hat{x}_h(\cdot)$  such that*

$$\|\hat{x}_h - x\|_\infty = O(\omega(h))$$

on  $[t_0, t_f]$  provided  $\omega(h) = o(\epsilon(h))$  and  $\eta(h) = O(\omega(h))$ .

Note that conditions  $B'$  and  $C'$  are satisfied for all  $x(t)$  if conditions  $B2'$  and  $C2'$  hold.

**PROOF.** The proof is essentially the same as for [16, Theorem 4.3] with “ $\subseteq$ ” replaced by “ $\sqsupseteq$ ”, “ $I(x(t_k)) \setminus I_k$ ” replaced by “ $I^p(x(t_k)) \setminus I_k^p$ ”, “ $h_j(\cdot) - h_i(\cdot)$ ” replaced by “ $h_j^p(\cdot) - h_i^p(\cdot)$ ”, “ $M_\alpha(I, x)$ ” replaced by “ $N_\alpha(I, x)$ ”, and finally “LCP( $M_\alpha, -e$ )” replaced by “LCP( $N_\alpha, \begin{bmatrix} 0 \\ -e \end{bmatrix}$ )”.

### 5. Implementation and computational results

The implementation of Algorithm 2 is based on that of Algorithm 1 used in [16]; in particular, the author’s own reimplementa-tion of Gear’s DIFSUB [9] is used as the smooth ODE solver; Al-Khayyal’s “branch-and-bound” method [1] is used to find all solutions to the LCPs; and Brent’s one-dimensional root finder [2] is used to locate

zero-crossings of the switching function. For comparison purposes, Taubert’s Euler method [18], and a 4th-order Runge-Kutta scheme [19] were implemented for this decomposition approach.

The test problem used here is a problem with a system of three masses connected by springs with a forcing term and friction as shown below. The actual differential

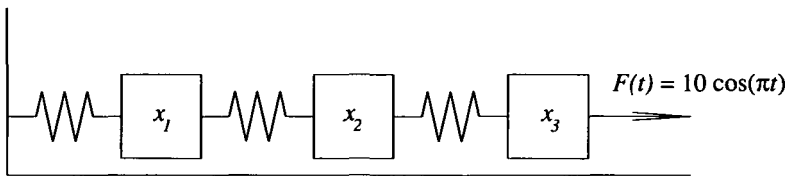


FIGURE 1. Arrangement of springs and masses.

equations to be solved are

$$\begin{aligned}
 x_1'' &= ( - x_1) + (x_2 - x_1) - x_1' - 0.3 \operatorname{sgn}(x_1'), \\
 x_2'' &= (x_1 - x_2) + (x_3 - x_2) - x_2' - 0.3 \operatorname{sgn}(x_2'), \\
 x_3'' &= (x_2 - x_3) \quad \quad \quad - x_3' - 0.3 \operatorname{sgn}(x_3') + 10 \cos(\pi t), \quad (15)
 \end{aligned}$$

with initial conditions

$$\begin{aligned}
 x_1(0) &= -1, & x_1'(0) &= -1, \\
 x_2(0) &= +1, & x_2'(0) &= +1, \\
 x_3(0) &= -1, & x_3'(0) &= +1.
 \end{aligned}$$

This system was solved over the interval  $[t_0, t_f] = [0, 10]$ . In this range there are 22 switching points. Graphs of the “exact” solution over this range are shown in Figure 2.

Table 1 has results for the case where the Euler and Runge-Kutta methods had a step size of  $h = 10^{-3}$ . Algorithm 2 had the truncation error per unit step set to  $10^{-3}$ , but with an initial step size of  $10^{-4}$ .

REMARK 2. Note that Taubert I is Taubert’s Euler method [18], Taubert II is a 4th-order Runge-Kutta method which belongs to the class of methods described in [19]. Also, to compute  $f(x)$  requires the computation of  $f_i^j(x)$  and  $h_i^j(x)$  for  $j = 1, \dots, 3$  and at least one  $i \in I^j(x)$ . Thus the number of  $f_i^j$  evaluations reported is essentially three times the number of  $f$  evaluations. Note also that the maximum error was taken over values at time intervals of 0.05 using the standard Euclidean norm. Finally, the



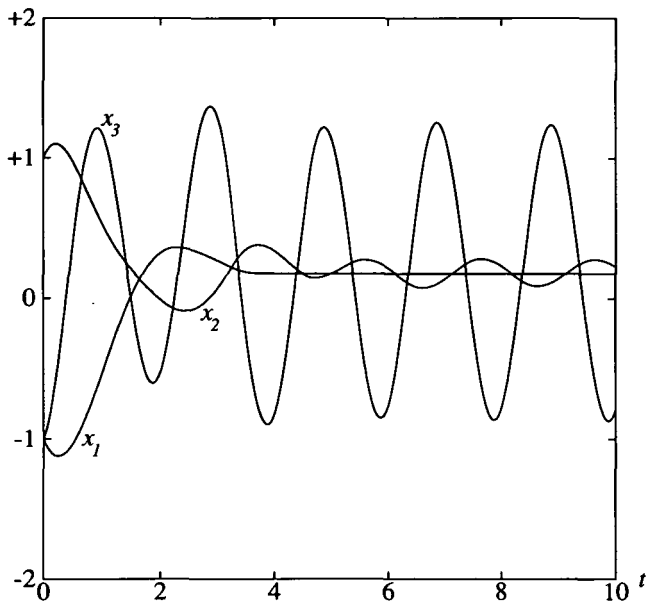


FIGURE 2. Trajectories.

“total” number of evaluations is  $\#f_i^j$  evaluations plus  $\#h_i^j$  evaluations plus  $6 \times \#\nabla h_i^j$  evaluations.

We now give a table of the number of function evaluations and errors for the case where  $h = 10^{-5}$  for Taubert’s methods, the truncation error per unit step for Algorithm 2 is set to  $10^{-5}$  and the initial step size is set to  $10^{-6}$ .

As is amply evident from these results, traditional methods, including Runge-Kutta methods, do not lead to great accuracy when applied to friction problems. Algorithm 2 on the other hand is now able to satisfactorily deal with problems with multiple friction surfaces. As noted in [16], there is considerably more overhead

Method	# Evaluations			“Total”	Max. Error
	$f_i^j$	$h_i^j$	$\nabla h_i^j$		
Taubert I	30 327	60 654	0	90 981	$9.5 \times 10^{-3}$
Taubert II	121 308	242 618	0	363 924	$6.4 \times 10^{-4}$
Algorithm 2	3 758	4 812	1 336	11 242	$5.4 \times 10^{-4}$

TABLE 1. Results for error  $\approx 10^{-3}$ .

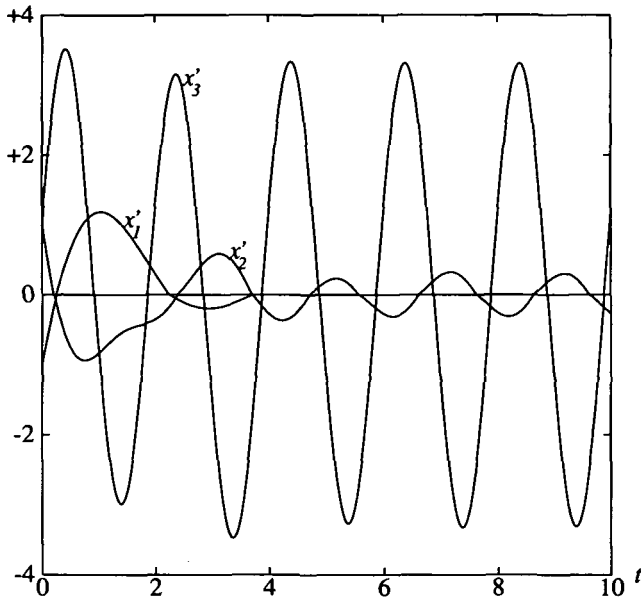


FIGURE 3. Velocity profiles.

per function evaluation in using Algorithm 2. However, this seems to be more than compensated for in situations where high accuracy is desired.

Method	# Evaluations				Max. Error
	$f'_i$	$h'_i$	$\nabla h'_i$	"Total"	
Taubert I	3 000 393	6 000 786	0	9 001 179	$9.51 \times 10^{-5}$
Taubert II	12 001 572	24 003 144	0	36 004 716	$4.98 \times 10^{-6}$
Algorithm 2	5 267	6 924	1 864	23 375	$4.68 \times 10^{-6}$

TABLE 2. Results for error  $\approx 10^{-5}$ .

### References

- [1] F. Al-Khayyal, "An implicit enumeration procedure for the general linear complementarity problem", *Math. Prog. Study* 31 (1987) 1–20.
- [2] R. P. Brent, "Algorithms for minimization without derivatives", in *Automatic Computation*, (Prentice-Hall, Englewood Cliff, NJ, 1973).
- [3] E. A. Coddington and N. Levinson, *Theory of ordinary differential equations* (Tata McGraw-Hill Publishing, New Delhi, 1972).

- [4] R. W. Cottle and G. B. Dantzig, "Complementary pivot theory of mathematical programming", *Lin. Alg. and its Appl.* **1** (1969) 103–125.
- [5] A. Dontchev and F. Lempio, "Difference methods for differential inclusions: A survey", *SIAM Rev.* **34** (1992) 263–294.
- [6] C. M. Elliott, "On the convergence of a one-step method for the numerical solution of ordinary differential inclusions", *IMA J. Numer. Anal.* **5** (1985) 3–21.
- [7] A. F. Filippov, "On certain questions in the theory of optimal control", *SIAM J. on Control and Optim.* **1** (1962) 76–84.
- [8] A. F. Filippov, "Differential equations with discontinuous right-hand side", *Amer. Math. Soc. Transl.* **42** (1964) 199–231.
- [9] C. W. Gear, "Numerical initial value problems in ordinary differential equations", in *Automatic Computation*, (Prentice-Hall, Englewood Cliff, NJ, 1973).
- [10] G. Golub and C. Van Loan, *Matrix computations*, first ed. (North Oxford Academic, 1983).
- [11] V. Guilleman and A. Pollack, *Differential topology* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
- [12] A. Kastner-Maresch, "Diskretisierungsverfahren zur lösung von differentialinklusionen", Ph. D. Thesis, Universität Bayreuth, 1990.
- [13] A. Kastner-Maresch, "Implicit Runge-Kutta methods for differential inclusions", *Numer. Funct. Anal. and Optim.* **11** (1991) 937–958.
- [14] H.-D. Niepage, "Inverse stability and convergence of difference approximations for boundary value problems for differential inclusions", *Numer. Funct. Anal. and Optim.* **9** (1987) 761–778.
- [15] H.-D. Niepage and W. Wendt, "On the discrete convergence of multistep methods for differential inclusions", *Numer. Funct. Anal. and Optim.* **9** (1987) 591–617.
- [16] D. E. Stewart, "A high accuracy method for solving ODEs with discontinuous right-hand side", *Numer. Math.* **58** (1990) 299–328.
- [17] D. E. Stewart, "High accuracy methods for solving ordinary differential equations with discontinuous right-hand side", Ph. D. Thesis, University of Queensland, St. Lucia, Queensland 4072, Australia, 1990.
- [18] K. Taubert, "Differenzverfahren für schwingungen mit trocken und zäher reibung und für reglungssysteme", *Numer. Math.* **26** (1976) 379–395.
- [19] K. Taubert, "Converging multistep methods for initial value problems involving multi-valued maps", *Computing* **27** (1981) 123–136.

## Appendix: Algorithms

Algorithm 1 of [16] is

ALGORITHM 1. *Algorithm 1 to compute solutions of 2–4.*

**Given:**  $h > 0$  step size for ODE solver  
 $\epsilon > 0$  for  $I_\epsilon(x)$ ;  $\eta \geq 0$  for locating zeros  
 $x_0 \in \mathbb{R}^n$ , initial value;  $t_0 \in \mathbb{R}$ , initial time

**Variables:**  $r$ , step number;  $\tau_r$ , time at  $r$ 'th step  
 $t'$ , switching time  
 $\psi_r$  for  $\psi$  evaluated at  $\tau_r$   
 $I, I_0$ , active sets  
 $t^*$ , restart time for non-unique solutions  
 $y_r, s_r$ , temporary values of  $x$  and  $t$

**Step 1:** \*\* Initialisation \*\*  
 $r \leftarrow 0$ ;  $\tau_0 \leftarrow t_0$ ;  $t^* \leftarrow +\infty$   
 $I_0 \leftarrow I_\epsilon(x_0)$   
 Perform step 5 to determine  $I$   
 Initialise ODE solver with  $x_0, t_0$  and step size  $h$

**Step 2:** \*\* Tentative step \*\*  
 $y_{r+1} \leftarrow$  result of single step of ODE solver with  
 right-hand side given by 9, 10  
 $s_{r+1} \leftarrow \tau_r + h$ ;  $\psi_{r+1} \leftarrow \psi(I, y_{r+1}, s_{r+1})$   
 If  $\psi_{r+1} < 0$  and  $\psi_{r+1} < \psi_r$  then go to step 4.

**Step 3:** \*\* Accept step \*\*  
 $x_{r+1} \leftarrow y_{r+1}$ ;  $\tau_{r+1} \leftarrow s_{r+1}$ ;  $r \leftarrow r + 1$   
 Go to step 2

**Step 4:** \*\* Switching point step \*\*  
 Locate an interval  $[a, b] \subset (\tau_r, s_{r+1})$  with  $|b - a| \leq \eta$   
 such that  $\psi(I, x(\tau), \tau) = 0$  for some  $\tau \in [a, b]$   
 $t' \leftarrow \tau_{r+1} \leftarrow b$ ; This is the switching point  
 $x_{r+1} \leftarrow x(\tau_{r+1}) = x(t')$   
 Perform step 5 to determine the new  $I$   
 Initialise ODE solver with  $x_{r+1}, t'$  and step size  $h$   
 Go to step 2

**Step 5:** \*\* Determine active set \*\*  
 Find all solutions of the LCP 13 for given  $I_0$ .  
 For each solution  $z^{(p)}, w^{(p)}, p = 1, 2, \dots$  construct  
 $(I^{(p)}) = \{i \mid \hat{z}^{(p)}_i > 0\}$ .

If the solution of 13 is unique then  $I \leftarrow I^{(1)}$

else choose  $p$  and  $I \leftarrow I^{(p)}$ .

If  $\text{LCP}(M_\alpha(I), -e)$  does not have a unique solution, then

choose  $t^* > \tau_{r+1}$ .

Return.

This is modified to form Algorithm 2:

**ALGORITHM 2.** *Modification of Algorithm 1 to compute solutions to 6–8.*

**Step 5:** **\*\* Determine active set \*\*** is modified to be

Find all solutions of the LCP 13 for given  $I_0$ .

For each solution  $\begin{bmatrix} z^{(p)} \\ \tilde{\mu}^{(p)} \end{bmatrix}, \begin{bmatrix} w^{(p)} \\ \beta^{(p)} \end{bmatrix}, p = 1, 2, \dots$  construct

$$(I^{(p)})^k = \{ i \mid \hat{z}^{(p)}_i > 0 \},$$

$$I^{(p)} = ((I^{(p)})^1, \dots, (I^{(p)})^m).$$

If the solution of 13 is unique then  $I \leftarrow I^{(1)}$

else choose  $p$  and  $I \leftarrow I^{(p)}$ .

If  $\text{LCP}(N_\alpha(I), \begin{bmatrix} 0 \\ -e \end{bmatrix})$  does not have a unique solution, then

choose  $t^* > \tau_{r+1}$ .

Return.