A SMOOTHING NEWTON METHOD FOR THE BOUNDARY-VALUED ODEs

ZHENG ZHENG

(Bsc., ECNU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2005

Acknowledgements

First of all, I would like to show great appreciation to my supervisor, Dr. Sun Defeng. His strict and patient guidance is the most impetus for me to finish this thesis. Without his critical instructions during the whole process, I could not have completed my thesis and acquired so much knowledge. Other helps from my fellows and friends are also indispensable to this work.

In addition, many thanks go to the Department of Mathematics, National University of Singapore for the Research Scholarship awarded to me, which financially supported my two years' M.Sc. candidature.

Last but not the least, the supports from my parents, my sister, and my boyfriend should not be ignored. It is their encouragement and warm care for both my study and daily life that makes me still energetic even when I was wear out in mind. In short, the thesis does not belong to my own, but to all the people who accompanied me throughout the days in Singapore.

Zheng Zheng /May 2005

Contents

Summary

The research of traditional boundary-valued ODEs has gone through a long history. With the advent of engineering systems like: multi-rigid-body dynamics with frictional contacts and constrained control systems, the smooth-coefficient differential equations are insufficient to practical utilizations. Many dynamic systems will naturally lead themselves to the ODEs with nonsmooth functions right-hand side as below

$$\begin{cases} \dot{x}(t) = f(t, x), \quad 0 \leq t \leq T \\ \Gamma(x(0), x(T)) = 0, \end{cases}$$

where f and Γ can be nonsmooth. To explore a certain method to attack this nonsmooth problem is the main goal in this thesis. In fact, the issue of solving a nonsmooth boundary-valued ODE is really a big challenge which involves interactions of different fields such as optimal control, ODE theory, nonsmooth analysis and so on. One type of the nonsmooth dynamic system: differential variational inequalities (DVIs) is worthy to mention which have been studied by Pang and Stewart for several years, as they are special case for the nonsmooth ODEs in a sense that the former can be reduced to the latter problem. Therefore, some of the

DVIs' results can be inherited and applied to the study of the nonsmooth ODEs.

One of common numerical methods for boundary value problem is the shooting method. It will provide the primary structure for the algorithm we want to develop. However, there are fundamental disadvantages mainly in that it inherits its stability properties from the stability of the initial value problems that it solves, not just the stability of the given boundary value problem. The smoothing Newton method proposed by Qi, Sun and Zhou serves as a promising modification to the shooting method because it guarantees the global convergence. More importantly, this technique is specialized for the nonsmooth equations. On the other aspect, obtained from the smoothing Newton method, the solution map $x(t)$ to the nonsmooth boundary value ODE is proved to be a semismooth (strongly semismooth) function around its nondifferentiable points, provided that f is semismooth (strongly semismooth, respectively) with respect to $x(t)$. Since the semismoothness (strongly semismoothness) is closely correlated to the superlinear (quadratic, respectively) convergence, the algorithm based on the smoothing Newton method will not lose its efficiency.

Some preliminaries are introduced in Chapter 2 as a preparation for the later discussions. In order to simplify the form of a nonsmooth ODE with parameters right-hand side as a usual ODE system and to facilitate the convergence analysis, a reformulation to the original problem is established in Chapter 3. The algorithm for the smoothing Newton method and its convergence property are given in Chapter 4, where the numerical results are also reported. Chapter 5 concerns about some final remarks and conclusions.

Chapter 1

Introduction

Ordinary Differential Equations (ODEs) with smoothing right-hand side has been quite familiar to us, since they have been studied for centuries (see [5] as a reference). Consider the standard Boundary-valued ODE form:

$$\begin{cases} \dot{x}(t) = f(t, x), \quad 0 \le t \le T \\ \Gamma(x(0), x(T)) = 0. \end{cases} \tag{1.1}$$

Here $f, \ \Gamma : \mathbb{R}^n \to \mathbb{R}^n$ are given vector functions. With the growing tendency to explore the engineering systems such as: multi-rigid-body dynamics with frictional contacts [1, 4, 6, 3] and constrained control systems [19, 12, 13, 18, 14, 8], traditional ODEs seem to be inadequate to cope with these situations, where *Nonsmooth* Boundary-value ODEs appear natural. We say an ODE is nonsmooth, when the differential and/or the boundary function (f and/or Γ) in (1.1) are/is nonsmooth.

When we cope with the nonsmooth functions, it is necessary to introduce the concept of *Generalized Jacobian*. Let $\mathbb{X}$ and $\mathbb{Y}$ be finite dimensional vector spaces, each equipped with a scalar innerproduct and an induced norm. Let $\mathcal{O}$ be an open set in $\mathbb{X}$. Suppose $H : \mathcal{O} \subseteq \mathbb{X} \to \mathbb{Y}$ is a locally Lipschitz function. According to Rademacher's Theorem , H is differentiable almost everywhere. Denote the set of points at which H is differentiable by $\mathcal{D}_H$. We write $J_x H(x)$ for the usual jacobian matrix of partial derivatives whenever x is a point at which the necessary partial derivatives exist. Let $\partial H(x)$ be the generalized Jacobian defined by Clarke in 2.6 of [11]. From the work of Warga [34, Theorem 4], the set $\partial H(x)$ is not affected if we "dig out" the sets of Lebesgue measure zero (see [11, Theorem 4] for the case $m = 1$), i.e., if S is any set of Lebesgue measure zero in $\mathbb{X}$, then

$$\partial H(x) = \text{conv}\{\lim_{k\to\infty} J_x H(x_k) : \; x_k \to x, \; x_k \in \mathcal{D}_H, \; x_k \notin S\}. \tag{1.2}$$

The nonsmooth ODE equation is definitely hard to solve and has been rarely touched until now. Nevertheless, another dynamic system *Differential Variational Inequalities* (DVIs) presented by Pang and Stewart in [23, 24, 25] can be served as a special case to the nonsmooth ODEs. The general form for the DVI is:

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)) \\ u(t) &\in \text{SOL}(K, F(t, x(t), \cdot)) \\ 0 &= \Gamma(x(0), x(T)), \end{aligned} \tag{1.3}$$

where, the second inclusion denotes the solution to the *Variational Inequalities*(VIs), for which a comprehensive reference is available [16]. According to the work from [23, 24], (1.3) can be looked upon as a special case of *Differential Algebraic Equations*(DAEs). When dealing with a DVI, one has to encounter nonsmooth functions, as the VIs always lead to nonsmooth equations. In other words, a VI can be reformulated to a nonsmooth algebraic equation. Once the solution to this algebraic equation is obtained and be substituted into the first differential equation $\dot{x} = f(t, x(t))$ we will get to a nonsmooth ODE.

Same as the motivation of studying the nonsmooth ODEs, one of the reasons to put forward the DVI as a distinctive class of dynamic system is that it also comes from those of practical engineering problems. Most applications of recent dynamic optimization take place in the context of the *Optimal Control Problem*

[11, 19, 12, 13, 18, 14, 2, 9] in standard or *Pontryagin* form. It is a formulation that has proved to be a natural one in the modeling of a variety of physical, economic, and engineering problems. In fact, the control problems act as the main source of the nonsmooth ODEs and the DVIs.

Given the dynamics, control and state constraints, and the functions $h : \mathbb{R}^n \to \mathbb{R}$ and $\varphi : [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, the optimal control problem is addressed by:

$$\begin{aligned} \min \quad & \int_0^T \varphi(t, x(t), u(t))dt + h[x(T)] \\ \text{s.t.} \quad & \dot{x}(t) = f(t, x(t), u(t)), \quad u(t) \in K \quad a.e.\ t \in [0, T] \\ & x(0) = x_0 \\ & x \in W^{1,\infty}, \quad u \in L^\infty, \end{aligned} \tag{1.4}$$

where the state $x(t) \in \mathbb{R}^n$, the control $u(t) \in \mathbb{R}^m$ and K is closed and convex. Here, L^p denotes the usual Lebesgue space of measurable functions with $p-th$ power integrable, and $W^{m,p}$ is the Sobolev space consisting of vector-valued functions whose $j-th$ derivative lies in L^p for all $0 \leq j \leq m$. Assume that (1.4) has a local minimizer (x^*, u^*) and that φ and f are twice continuously differentiable.

The Hamiltonian denoted by H is defined as:

$$H(t, x(t), u(t), \lambda(t)) = \varphi(t, x, u) + \lambda^T f(t, x, u),$$

where the variable $\lambda \in W^{2,\infty}$ is called associated *Lagrange multipliers*.

Instead of studying (1.4) directly, we examine the famous first-order necessary optimality condition (*Maximum Principle*):

Let (x^*, u^*) be a solution to the problem (1.4), then there exists a $\lambda^*(\cdot) : [0, T] \to \mathbb{R}^n$ satisfying the following at (x^*, u^*, λ^*):

$$\begin{aligned} & \dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0 \\ & \dot{\lambda} = -\nabla_x H(t, x(t), u(t), \lambda(t)), \quad \lambda(T) = h_x[x(T)] \\ & H(t, x^*(t), u^*(t), \lambda^*(t)) = \max_{u \in K} H(t, x(t), u(t), \lambda(t)), \quad a.e.\ t \in [0, T]. \end{aligned} \tag{1.5}$$

The last equation

$$H(t, x^*(t), u^*(t), \lambda^*(t)) = \max_{u \in K} H(t, x(t), u(t), \lambda(t)), \quad a.e. \ t \in [0, T]$$

can be rephrased as:

$$\langle \nabla_u (H(t, x^*(t), u^*(t), \lambda^*(t))), (u - u') \rangle \geq 0 \ \text{ for all } u' \in K.$$

Together with the definition of the VIs in [23, Section 2], this inequality can be converted into an inclusion:

$$u(t) \in \text{SOL}(K, \nabla_u H(t, x(t), u(t), \cdot, \lambda(t))).$$

By replacing the last equation in (1.5) with this inclusion, a DVI with the form of (1.3) is established. Then after substituting $u(t)$ into the two differential equations of (1.5), the DVI is reduced to a boundary-valued ODE with nonsmooth right-hand side functions.

For instance, the differential Nash game [7, 15] and multi-rigid-body dynamics with contact and friction are typical control problems that result in the nonsmooth ODEs. In [23, Section 4], Pang and Stewart provide us a careful deduction of these two systems.

The key point in the thesis is to apply the *smoothing Newton method* developed in [29] for the nonlinear complementarity problems and the VIs to solving the nonsmooth dynamic systems. This requires the collection of techniques from different areas. The classical *single shooting method* will be our consideration on dealing with the ODE, i.e., to “shoot” an ideal initial value $x(0; c) = c$ in order to satisfy the boundary condition

$$h(c) := \Gamma(c, x(T; c)) = 0.$$

In essence, shooting is nothing but Newton's method to find out the root of an equation $h(c) = 0$. However, the single shooting cannot have global convergence, in a sense that the terminal value $x(T;c)$ obtained from shooting would terribly deviate from the exact terminal condition, if the initial value c is not properly estimated. In order to conquer such a pitfall, other techniques should be put into use as a modification of the single shooting to insure the globalization. For this purpose, the global convergence for the smoothing Newton's method proves to be a suitable alternative, which is a big contribution even under a smoothing case.

Note that the function $f(t,x)$ in (1.1) can be nonsmooth, in order to apply the smoothing Newton method, f should be approximated by some smoothing function (the existence of such smoothing function can be obtained via convolution, see [32, 35] and the reference therein). Let us denote $f^\varepsilon(t,x)$ as the smoothing function to $f(t,x)$, in which $\varepsilon = 0$ if $f(t,x)$ is a smooth function. It follows that the solution $x(t;c)$ to (1.1) becomes $x^\varepsilon(t;c)$, which results in a new boundary equation:

$$h^\varepsilon(c) = \Gamma(c, x^\varepsilon(T;c)) = 0. \tag{1.6}$$

One significant contribution of this paper is to reformulate (1.1) along with (1.6) to a new nonsmooth dynamic system. We discuss the details in later chapter. Whatever formulation we have transferred to, finally, we have to establish an equation with respect to $h^\varepsilon(c)$ as:

$$E(\varepsilon, c) = \begin{bmatrix} \varepsilon \\ h^\varepsilon(c) \end{bmatrix} = 0,$$

which is solved by the smoothing Newton method.

To the best of our knowledge, nearly no numerical examples and results have been given for the nonsmooth boundary-value ODEs so far. Even for its special case: the DVIs, the computational work is almost blank. Therefore, all the research works on this topic are mainly at the theoretical aspect and this newly developed technique needs to be implemented. To this end, we provide the smoothing Newton algorithm and implement it with numerical examples. Results are to be reported at the end. Meanwhile, convergence analysis is also included in as a justification of this algorithm.

1.1 Overall Arrangement

In Chapter 2, firstly we introduce the classical results of the ODEs as well as the numerical methods for a boundary value problem. Then some knowledge about the nonsmoothness is presented in Section 2.2 that includes the concept of semismoothness and varies types of smoothing functions. In Section 2.3 we introduce the standard form of the a DVI and its extension problems. We construct a reformulation in Chapter 3, which is the most important part in this thesis. After all the nonsmooth functions being replaced by their smoothing approximations, we reformulated the the nonsmooth boundary ODE with parameters right-hand side to an initial value problem together with its boundary equation. In Chapter 4, an algorithm of the smoothing Newton method for solving the reformulated ODEs is established. Based on the algorithm, both the global and superlinear (quadratic) convergence are analyzed. Some numerical results are also reported at the end of this chapter. The whole thesis is ended with some conclusions and remarks given in Chapter 5.

Chapter 2

Preliminaries

In this chapter, we have two classes of preliminary discussions: ODEs and Nonsmoothness, for they are fundamental compositions in our subject. The former is mainly about the ODE sensitivity theory and numerical methods for the boundary value problems (BVP), while the latter part focuses on the *semismoothness*. Especially, the sensitivity theory and semismoothness are critical to the convergence analysis of the smoothing Newton Algorithm. In addition, knowledge about the DVIs will also be presented as a specific case.

2.1 Theories of ODEs

Consider the initial-value problems (IVP) :

$$\begin{cases} \dot{x} = f(t, x), \quad 0 \leq t \leq T \\ x(0) = c \text{ (given)}, \end{cases} \tag{2.1}$$

where $f : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is a given vector function and $c \in \mathbb{R}^n$ is an initial vector.

Lemma 2.1.1. Suppose f is Locally Lipschitz continuous in a neighborhood of the trajectory starting at c^0; i.e., there is an open neighborhood $\mathcal{N}_T$ of the set

$\Xi_T = \{x(t; c^0) : 0 \le t \le T\}$ and a scalar $L \ge 0$ such that

$$\| f(t, x(t; c)) - f(t, x(t; c')) \| \le L \| x(t; c) - x(t; c') \|, \quad \forall\ x(t; c), x(t; c') \in \mathcal{N}_T.$$

Then, there exists a neighborhood $\mathcal{N}_0$ of c^0 such that for every $c \in \mathcal{N}_0$, the ODE (2.1) has a unique solution $x(t; c)$ on $[0, T]$ which satisfies, for any c and $c' \in \mathcal{N}_0$,

$$\|x(t; c) - x(t; c')\| \le e^{Lt}\|c - c'\|, \quad \forall\ t \in [0, T]. \tag{2.2}$$

Proof. Let $x(t; c)$, $x(t; c')$ be two solutions to (2.1). From the variational rules of the derivative of the solution map, they can be written as

$$\begin{aligned} x(t; c) &= \int_0^t f(s, x(s; c))ds + c \\ x(t; c') &= \int_0^t f(s, x(s; c'))ds + c'. \end{aligned}$$

We have

$$\begin{aligned} \|x(t; c) - x(t; c')\| &\le \|c - c'\| + \int_0^t \|f(s, x(s; c)) - f(s, x(s; c'))\|ds \\ &\le \|c - c'\| + L \int_0^t \|x(s; c) - x(s; c')\|ds. \end{aligned}$$

According to the Gronwall lemma, we deduce

$$\begin{aligned} \|x(t; c) - x(t; c')\| &\le \|c - c'\| + L \int_0^t \|c - c'\|e^{L(t-s)}ds \\ &= \|c - c'\| + L\|c - c'\| \int_0^t e^{L(t-s)}ds \\ &= \|c - c'\| - \|c - c'\|(1 - e^{Lt}) \\ &= e^{Lt}\|c - c'\|, \end{aligned}$$

which gives the inequality (2.2).

The uniqueness of the solution map $x(t; c)$ for every $c \in \mathcal{N}_0$ can be directly obtained from the Lipschitz continuity with respect to initial data in (2.2). $\square$

This result is particularly true as $f(t, x)$ is globally Lipschitz continuous. For the latter case, see [5, Theorem 1.1]. More details about the locally Lipschitz characterization of ODE function and its solution map can be referred to Theorem 2.1.12 in [31].

Next, we take a further step into the boundary-value problem of ODE, which is defined as:

$$\begin{cases} \dot{x} = f(t, x), & 0 \leq t \leq T \\ \Gamma(x(0), x(T)) = 0. \end{cases} \tag{2.3}$$

We consider an initial value method: *Single Shooting* method [5, Chapter 7]. The shooting method is a straightforward extension of the initial value techniques for solving the BVPs. Essentially, one "shoots" trajectories of the same ODE with different initial values until one "hits" the correct given boundary values at the other interval end.

We denote $x(t; c) := x(t)$ as the solution of the ODE (2.3) satisfying the initial condition $x(0, c) = c$. Substituting it into the boundary equation, we have

$$h(c) := \Gamma(c, x(T; c)) = 0. \tag{2.4}$$

This gives a set of n algebraic equations for the n unknowns c. The single shooting method is that, for a given c, one solves algebraic equation (2.4) with solving the corresponding initial value ODE problem.

Consider Newton's method for finding out the root of (2.4). The iteration is:

$$c^{\nu+1} = c^{\nu} - (J_c h(c))^{-1} h(c^{\nu}),$$

where c^0 is an initial guess. In order to evaluate $J_c h(c)$ at $c = c^{\nu}$, we must differentiate the expression of h with respect to c. Denote the Jacobian matrices of $\Gamma(\mathbf{u}, \mathbf{v})$ with respect to its first and second argument vectors by

$$B_0 = J_{\mathbf{u}} \Gamma(\mathbf{u}, \mathbf{v}), \quad B_T = J_{\mathbf{v}} \Gamma(\mathbf{u}, \mathbf{v}) \tag{2.5}$$

(Often in application, Γ is linear in $\mathbf{u}, \mathbf{v}$ and the $n \times n$ matrices B_0, B_T are constant). Using the notation in (2.5), we have:

$$Q := J_c h(c) = B_0 + B_T X(T),$$

where $X(t)$ is the $n \times n$ fundamental solution matrix to the following system [5, Section 6.1]:

$$\begin{cases} \dot{X}(t) &= A(t)X(t), \qquad 0 \le t \le T \\ X(0) &= I, \end{cases} \tag{2.6}$$

with $A(t, x(t; c^\nu)) = J_x f(t, x)$. Therefore, the $n + 1$ IVPs are to be solved at each iteration by using Newton's method. Finally, once an appropriate initial value c has been found, we can use this value to obtain the solution to the original BVP from integrating the corresponding IVP.

The advantages of single shooting are conceptually simple and easy to implement. However, there are difficulties as well. Because the algorithm inherits its stability properties from that of IVPs, but not the stability of the given BVPs, the process of shooting will involve integrating a potentially unstable IVP even if the BVP is stable. Another difficulty lies in the lack of global convergence for the single shooting method, that is, there is no guarantee with the existence of solutions for an arbitrarily given initial value c. Nevertheless, the smoothing Newton method we will apply later in Chapter 4 does not have this trouble. The globalization technique can be used not only in the smoothing functions, but also in the nonsmooth problems.

Both of the disadvantages of the single shooting become worse for larger intervals of integration of IVPs. This fact leads to another type of shooting method: *Multiple* shooting, which works well in the case when single shooting is unsatisfactory. The basic idea of multiple shooting is then to restrict the size of intervals over which IVPs are integrated. After partitioning the time interval $[0, T]$ into N

subinterval: $[t_{n-1}, t_n], n = 1, \cdots, N$, we approximate the solution of the ODE by constructing an approximate solution on each $[t_{n-1}, t_n]$ and patching these approximate solutions together to form a globle one. We just give a simple introduction to this method, do not pursue this further, though.

2.2 Introduction to Nonsmoothness

2.2.1 Semismoothness

In this section, we give definition of Semismoothness, which involves the concept of generalized Jacobian $\partial H(x)$ (see (1.2) in Section 1). Semismoothness was introduced originally by Mifflin [22] for functionals. Convex functions, smooth functions, and piecewise linear functions are examples of semismooth functions. The composition of semismooth functions is still a semismooth function. Semismooth functions play an important role in the global convergence theory of nonsmooth optimization; Indeed, we need the concept to establish the superlinear convergence of smoothing Newton Methods that will be discussed in later chapter. Let us see the definition below [33, Definition 5].

Definition 1. Suppose that $H : \mathcal{O} \subseteq \mathbb{X} \to \mathbb{Y}$ is locally Lipschitz continuous function. H is said to be semismooth at $x \in \mathcal{O}$ if

(i) $H(x)$ is directionally differentiable at x; and

(ii) for any $y \to x$ and $V \in \partial H(y)$,

$$H(y) - H(x) - V(y - x) = o(\| y - x \|). \tag{2.7}$$

Part (i) and (ii) in this definition do not imply each other. H is said to be $G-$semismooth at x if condition (2.7) holds. $G-$semismooth was used in [17, 26] to obtain inverse and implicit function theorems and stability analysis for nonsmooth equations. Moreover, a stronger notion is $\gamma-$order semismoothness with $\gamma > 0$. For any $\gamma > 0$, H is said to be γ-order G-semismooth (respectively, γ-order semismooth) at x, if H is G-semismooth (respectively, semismooth) at x and for any $y \to x$ and $V \in \partial H(y)$,

$$H(y) - H(x) - V(y - x) = O(\|y - x\|^{1+\gamma}). \tag{2.8}$$

When $\gamma = 1$ (1-order G-semismooth (respectively, 1-order semismooth) at x), H is said to be strongly G-semismooth (respectively, strongly semismooth) at x.

From definition 1, one needs to consider the set of differentiable points $\mathcal{D}_H$. Sometimes this brings us much troubles in proving the semismoothness of a function. Fortunately, by the work of Warga [34, Theorem 4], the set $\partial H(x)$ remains the same if we do not consider the sets of Lebesgue measure zero. The following result cited from [33, Lemma 6] modified the original definition of semismoothness.

Theorem 2.2.1. *Let $H : \mathcal{O} \subseteq \mathbb{X} \to \mathbb{Y}$ be a locally Lipschitz near $x \in \mathcal{O}$. Let $\gamma > 0$ be a constant. If S is a set of Lebesgue measure zero in $\mathbb{X}$, then H is G-semismooth (γ-order G-semismooth) at x if and only if for any $y \to x$, $y \in \mathcal{D}_H$, and $y \notin S$,*

$$H(y) - H(x) - J_y H(y)(y - x) = o(\|y - x\|) \quad (O(\|y - x\|^{1+\gamma}). \tag{2.9}$$

Hence, those nondifferentiable points with Lebesgue measure zero can be ignored, when the semismoothness of a function is to be proved. This will save us much work in later convergence discussions.

2.2.2 Classifications to Smoothing function

We provide some computable smoothing functions [28] for variational inequality problems.

Consider the equation:

$$H(u) = 0, \tag{2.10}$$

where $H : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous but not necessarily continuous differentiable. As was mentioned in Introduction Section, H is differentiable almost everywhere by Rademacher Theorem [11]. Such nonsmooth equations arise from nonlinear complementarity problems, VIs, maximal monotone operator problems [28].

The smoothing method is to construct a smoothing function $G^\varepsilon : \mathbb{R}^n \to \mathbb{R}^n$ of H such that, for any $\varepsilon > 0$, G^ε is continuous differentiable on $\mathbb{R}^n$ and, for any $u \in \mathbb{R}^n$, it satisfies,

$$\|H(z) - G^\varepsilon(z)\| \to 0, \quad \text{as } \varepsilon \downarrow 0, z \to u. \tag{2.11}$$

To solve equation (2.10), we can approximately solve the following problems for a given positive sequence $\{\varepsilon^k\}, k = 0, 1, \ldots,$

$$G^{\varepsilon^k}(u^k) = 0. \tag{2.12}$$

In conclusion, we give a definition of smoothing function [28, Section 2.1]:

Definition 2. A function $G^\varepsilon : \mathbb{R}^n \to \mathbb{R}^m$ is called a smoothing function of a nonsmooth function $H : \mathbb{R}^n \to \mathbb{R}^m$, if any $\varepsilon > 0, G^\varepsilon(\cdot)$ is continuously differentiable and, for any $u \in \mathbb{R}^n$,

$$\|H(z) - G^\varepsilon(z)\| \to 0, \quad \text{as } \varepsilon \downarrow 0, z \to u.$$

Equation (2.11) provides a generalized definition for a smoothing function and almost all the existing smoothing functions are included in.

Usually, a convolution is involved in computing these smoothing functions. We just give a rough picture of the computation work via convolution (one can see

[28, Section 3] for further knowledge). A function $\rho : \mathbb{R} \to \mathbb{R}_+$ is called a kernel function if it is integrable (in the sense of Lebesgue) and

$$\int_{\mathbb{R}} \rho(s)ds = 1.$$

Suppose that ρ is a kernel function. Define $\Theta : \mathbb{R}^{++} \times \mathbb{R}^m \to \mathbb{R}_+$ by

$$\Theta(\varepsilon, x) := \varepsilon^{-m}\Phi(\varepsilon^{-1}x),$$

where $(\varepsilon, x) \in \mathbb{R}_{++} \times \mathbb{R}^m$ and

$$\Phi(z) := \prod_{i=1}^{m} \rho(z_i), \quad z \in \mathbb{R}^m.$$

Then, a smoothing approximation of a nonsmooth function $F : \mathbb{R}^m \to \mathbb{R}^p$ via convolution can be described by

$$\begin{aligned} F^{\varepsilon}(x) &:= \textstyle\int_{\mathbb{R}^m} F(x-y)\Theta(\varepsilon, y)dy \\ &= \textstyle\int_{\mathbb{R}^m} F(x-\varepsilon y)\Phi(y)dy \\ &= \textstyle\int_{\mathbb{R}^m} F(y)\Theta(\varepsilon, x-y)dy. \end{aligned} \tag{2.13}$$

Denote $F^0(x) = F(x)$, and $F^{|\varepsilon|}(x) = F^{-|\varepsilon|}(x)$.

Next, we introduce smoothing functions for simple nonsmooth functions, beginning with the *Plus function* first. One of the simplest but very useful nonsmooth function is the plus function $p : \mathbb{R} \to \mathbb{R}_+$, define by

$$p(t) := \max\{0, t\} \text{ for any } t \in \mathbb{R}.$$

We define $P(\varepsilon, t)$ such that

$$P(0,t) := p(t) \text{ and } P(-|\varepsilon|, t) := P(|\varepsilon|, t), (\varepsilon, t) \in \mathbb{R}^2,$$

as the smoothing function to $p(t)$.

One of the well-known smoothing functions for the plus function p is

$$P(\varepsilon, t) = \frac{1}{2}(t + \sqrt{t^2 + 4\varepsilon^2}). \tag{2.14}$$

We can derive lots of good properties from $P(\varepsilon, t)$ such as: P is globally Lipschitz continuous on $\mathbb{R}^2$ and continuously differentiable on $\mathbb{R}_{++} \times \mathbb{R}$; The directional derivative of P at $(0, t)$ exists; P is semismooth on $\mathbb{R}^2$, and so on.

Another widely used function is *Absolute Value Function*: $q : \mathbb{R} \to \mathbb{R}$, which is defined by

$$q(t) = |t| \qquad t \in \mathbb{R}.$$

Notice that $q(t)$ can be written as the linear combination of plus functions, i.e.,

$$q(t) = p(t) + p(-t).$$

Thus, based on the above discussion about P, one can easily obtain the smoothing function of q:

$$Q(\varepsilon, t) = P(\varepsilon, t) + P(\varepsilon, -t), \quad (\varepsilon, t) \in \mathbb{R}_{++} \times \mathbb{R},$$

where P is the smoothing function of the plus function$p(t)$. Analogously, we have

$$Q(0, t) := q(t) = |t| \quad \text{and} \quad Q(-|\varepsilon|, t) := Q(|\varepsilon|, t), \quad (\varepsilon, t) \in \mathbb{R}^2.$$

Finally, we study a class of computable smoothing function for the VIs, or a smoothing approximation of

$$H(u) := u - \Pi_K(u - F(u))$$

for any $u \in \mathbb{R}^n$ when K is a closed convex subset of $\mathbb{R}^n$ (we discuss more about the VIs in following Section 2.3). When K is $\mathbb{R}^n_+$, $\Pi_K(u)$ is the Euclidean projection of u onto the nonnegative orthant and satisfies

$$\begin{aligned} H(u) &= u - \Pi_K(u - F(u)) \\ &= u - \max\{0, u - F(u)\} \\ &= \min\{u, F(u)\}. \end{aligned}$$

By using (2.14) again yields the smoothing function of $H(u)$ and we have

$$\phi(\varepsilon, u) := \begin{pmatrix} \frac{1}{2}(u_1 + F(u_1) - \sqrt{(u_1 - F(u_1))^2 + 4\varepsilon^2}) \\ \vdots \\ \frac{1}{2}(u_n + F(u_n) - \sqrt{(u_n - F(u_n))^2 + 4\varepsilon^2} \end{pmatrix}.$$

and

$$\phi(0, u) := H(u); \quad \phi(-|\varepsilon|, u) := \phi(|\varepsilon|, u), \quad (\varepsilon, u) \in \mathbb{R}^{n+1}.$$

Knowledge on the smoothing functions is quite rich. It has been shown that for each semismooth function, there exists a smoothing function with semismoothness itself via convolution approach ([32], [35, Theorem 2.12]). See [30] for the smooth approximation functions for eigenvalues of a real symmetric matrix. Usually, a multivariate integral is involved in computing equation (2.13), which makes them uncomputable in practice. However, we need computable smoothing approximations for those nonsmooth functions arising from complementarity problems and variational inequality problems.

2.3 Standard formulation of DVIs

DVIs as the unusual nonsmooth dynamic systems were firstly addressed by Pang and Stewart in [23]. In their paper, it gives a formal definition. Let $f : \mathbb{R}^{1+n+m} \to \mathbb{R}^n$ and $F : \mathbb{R}^{1+n+m} \to \mathbb{R}^m$ be two continuous vector functions; Let K be a nonempty closed convex subset of $\mathbb{R}^m$. Let $\Gamma : \mathbb{R}^{2n} \to \mathbb{R}^n$ be a boundary function and $T > 0$ be a terminal time. The DVI defined by three functions: f, F and Γ, the set K, and the scalar T is to find time-dependent trajectories $x(t)$ and $u(t)$

that satisfy condition (2.15) for $t \in [0, T]$.

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)) \\ u(t) &\in \text{SOL}(K, F(t, x(t), \cdot)) \\ 0 &= \Gamma(x(0), x(T)). \end{aligned} \tag{2.15}$$

$\text{SOL}(K, F(t, x(t), \cdot))$ denotes the solution set to a VI $(K, \mathbf{F})$ [16]:

$$(u' - u)^T \mathbf{F}(u) \geq 0, \quad \forall\, u' \in K.$$

Moreover, $u \in \text{SOL}(K, F(t, x, \cdot))$ if and only if

$$0 = u - \Pi_K(u - F(t, x, u)),$$

where Π_K denotes the Euclidean projector onto the closed convex set K, which is the unique solution of the convex minimization problem in the variable y, where x is considered fixed:

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}(y - x)^T (y - x) \\ \text{subject to} & y \in K. \end{array}$$

For a detailed study on the differentiability properties of this operator and for references, please see ([16, Section 1.5.2]).

Assume that (see [23, Section 5.1], and [24, Section 3]):

(A) $F(t, x, u)$ is a continuous, *uniformly* P function [16] on K with modulus that is independent of (t, x); i.e., there exists a constant $\eta_F > 0$ such that

$$\max_{1 \leq \nu \leq N} (u_\nu - u'_\nu)^T (F_\nu(t, x, u) - F_\nu(t, x, u')) \geq \nu_F \|u - u'\|^2$$

for all $(t, x) \in [0, T] \times X$ and

$$u \equiv (u_\nu)_{\nu=1}^N \text{ and } u' \equiv (u'_\nu)_{\nu=1}^N \text{ in } K = \prod_{\nu=1}^N K^\nu;$$

(B) $F(\cdot, \cdot, u)$ is Lipschitz continuous with a constant that is *independent* of u;

(C) f is Lipschitz continuous and directionally differentiable on an open neighborhood $\mathcal{N}_T$ of the nominal trajectory $\Xi_T \equiv \{x(t, c^0) : \ 0 \le t \le T\}$.

Remarks: Assumption (A) is a very strong condition, however it cannot be much relaxed for the reason that "uniformly P function" is to ensure the uniqueness and Lipschitz continuity of the solution $u(t; x)$. Thus this assumption seems a reasonable one.

From [23, Theorem 1], under the assumptions (A) and (B), the solution $u(t; x)$ to the VI: $(K, F(t, x, \cdot))$ is Lipschitz continuous and unique on a close convex set K. By casting the VI as a projector Π on the close convex set K, we can expect a semismooth solution $u(t, x)$ defined with an implicit function. Since f is supposed to be a Lipschitzian, after putting $u(t, x)$ into $f(t, x(t), u(t))$, the reduced ODE function $f(t, x, u(t, x))$ is also semismooth.

When K is a cone $\mathcal{C}$. DVI (2.15) will become a differential complementarity problem (DCP):

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)) \\ \mathcal{C} \ni u(t) &\perp F(t, x(t), u(t)) \in \mathcal{C}^* \\ 0 &= \Gamma(x(0), x(T)), \end{aligned}$$

where

$$\mathcal{C}^* \equiv \{v \in \mathbb{R}^m : u^T v \ge 0 \ \ \forall u \in \mathcal{C}\}$$

is the dual cone of $\mathcal{C}$. Moreover, if the ODE function f is separable with x, u and the VI function F happens to be linear in x and u, i.e., the VI is a linear complementarity problem (LCP). The DCP yields a more specific form:

$$\begin{aligned} \dot{x}(t) &= \hat{f}(t,x) + Bu \\ 0 \le u &\perp q + Cx + Du \ge 0 \\ 0 &= \Gamma(x(0), x(T)), \end{aligned} \tag{2.16}$$

where q is a given m-vector; B, C and D are given matrices in $\mathbb{R}^{n\times m}, \mathbb{R}^{m\times n}$ and $\mathbb{R}^{m\times m}$, respectively. In this case, it is easy to see that $\mathcal{C}$ has been reduced to $\mathbb{R}^m_+$. The aim of introducing this special form is that the numerical examples in later chapter are mainly dependent on the system (2.16). For further study of the LCP problem, one can refer to [10, 20], which are Ph.D thesis by Camlibel and Heemels respectively.

Chapter 3

Reformulation of Nonsmooth ODEs

3.1 Generic Case

Let us rewrite the ODE system (1.1) presented at the very beginning

$$\begin{cases} \dot{x}(t) = f(t,x), & 0 \le t \le T \\ \Gamma(x(0), x(T)) = 0. \end{cases}$$

As is mentioned before, the main contribution to the thesis is reformulating (1.1) from nonsmooth equations to a smoothing system. There are two reasons for this transformation. One is to simplify the notation in order that the newly defined system would be more clear in variables and of uniform structure as a usual ODE. The other reason is to facilitate the proof of semismoothness for the solution set $x(t)$ so that a satisfactory convergence property could be obtained.

Given an initial value c such that $x(0; c) = c$, recall the single shooting method, whose motivation is to find out the root of equation (2.4)

$$h(c) := \Gamma(c, x(T; c)) = 0.$$

Note that $f(t, x)$ in (1.1) could be a nonsmooth function, it needs to be smoothed first. Denote $g(t, \varepsilon, x) \equiv f^{\varepsilon}(t, x)$ as the smoothing function to $f(t, x)$ ($\varepsilon = 0$ when

f is smooth). From definition 2, $g(t,\varepsilon,c)$ is continuously differentiable for any $\varepsilon > 0$. Consider the following initial ODE:

$$\begin{cases} \dot{x} = f^{\varepsilon}(t,x) \equiv g(t,\varepsilon,x), & 0 \le t \le T \\ x(0) = c. \end{cases} \tag{3.1}$$

Since an extra variable ε has been added into (3.1) due to the smoothing function $g(t,\varepsilon,x)$, the solution $x(t;c)$ is not only dependent on t and the parameter c, but also changes with ε. Hence the notation of the solution $x(t;c)$ will be altered to $x^{\varepsilon}(t;c)$. As a consequence, the boundary equation (2.4) becomes:

$$h^{\varepsilon}(c) = \Gamma(c, x^{\varepsilon}(T;c)) = 0. \tag{3.2}$$

In addition, based on the fact that $h^{\varepsilon}(c)$ might also be a nonsmooth function, again, it has to be constructed into its corresponding smoothing approximation. Let us denote this smoothing function by $\tilde{h}^{\varepsilon}(c)$. When $h^{\varepsilon}(c)$ is a smooth function, $\tilde{h}^{\varepsilon}(c)$ will be $h^{\varepsilon}(c)$ itself. Consequently, (3.2) is presented as:

$$\tilde{h}^{\varepsilon}(c) \equiv \widetilde{\Gamma}(c, x^{\varepsilon}(T;c)) = 0. \tag{3.3}$$

Apparently, it seems to be enough to get to the equation (3.3), since $\tilde{h}^{\varepsilon}(c) = 0$ is already the equation to be solved by using the smoothing Newton method. However, it is somewhat confusing in notation. To avoid such inconvenience in application, (3.1) requires further reformulation. One can easily see that ε and c play the similar roles in the process of solving equation (3.2), so it comes up quite natural that we shall take ε as another parameter in both $x^{\varepsilon}(t;c)$ and $h^{\varepsilon}(c)$, just as the same position we do on the initial value c. Actually, one advantage to do this is that all the existing results for initial-valued ODEs can be inherited to our reformulated ODE system.

The following steps are the essential part in the whole reformulation work. The thing is that ε is just a common parameter along with the smoothing function $g(t,\varepsilon,x)$. If we can make ε be the initial value for some other variable in

ODE problem (3.1), the two parameters ε and c would be taken into the identical operations during the whole solving process.

Let $\tau(t) \equiv \varepsilon$, whose initial value is always ε by its definition. Therefore, (3.1) is transformed to:

$$\begin{cases} \dot{x} = g(t, \tau, x), & 0 \le t \le T \\ \dot{\tau} \equiv 0, & 0 \le t \le T \\ x(0) = c \\ \tau(0) = \varepsilon. \end{cases}$$

This is equivalent to the initial-valued ODE below:

$$\begin{cases} \dot{y} = p(t, y), & 0 \le t \le T \\ y(0) = \begin{pmatrix} \varepsilon \\ c \end{pmatrix}, \end{cases} \tag{3.4}$$

in which the new ODE variable is defined as $y \equiv (\tau, x)^T$ (notice that $y(t; -|\varepsilon|, c) := y(t; |\varepsilon|, c)$); and $p : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ has the following form:

$$p(t, y) \equiv \begin{bmatrix} 0 \\ g(t, \tau, x) \end{bmatrix}. \tag{3.5}$$

(3.4) appears to be a standard initial value problem except for the semismoothness of the differential function $p(t, y)$.

Meanwhile, $\tilde{h}^{\varepsilon}(c)$ will also change its notation into

$$\begin{aligned} \hat{h}(\varepsilon, c) &:= \widehat{\Gamma}(y(0), y(T; \varepsilon, c)) \\ &\equiv \widehat{\Gamma}((\varepsilon, c), y(T; \varepsilon, c)). \end{aligned} \tag{3.6}$$

Finally, by constructing the function $E(\varepsilon, c)$ as

$$E(\varepsilon, c) = \begin{bmatrix} \varepsilon \\ \hat{h}(\varepsilon, c) \end{bmatrix} = 0, \tag{3.7}$$

we can utilize the smoothing Newton method to approximate the exact solution c^* to $\hat{h}(\varepsilon, c)$, or in other words, the real initial value for the original boundary-valued ODE problem (1.1). With this initial value c^*, we can proceed to work out all the numerical solutions $x(t)$ with respect to every t from 0 to T.

3.2 A Specific Case: Boundary-valued ODE with an LCP

The pure illustration based on general nonsmooth ODEs might be too abstract to fully understand. In order to give a more clear picture of the reformulation, let us consider the differential complementarity problem (DCP) (2.16):

$$\begin{aligned} \dot{x}(t) &= \hat{f}(t,x) + Bu \\ 0 \le u &\perp q + Cx + Du \ge 0 \\ 0 &= \Gamma(x(0), x(T)), \end{aligned}$$

where, the LCP matrix D is supposed to be a P_0- matrix (P_0- LCP); i.e., all of its principal minors are nonnegative. We assume $\hat{f}(t,x)$ satisfies assumption (C) in Section 2.3, hence, so does the ODE function $\hat{f}(t,x) + Bu$.

We know that

$$0 \le u \perp q + Cx + Du \ge 0 \tag{3.8}$$

if and only if

$$0 = u - \Pi_K(u - F(t,x,u)), \tag{3.9}$$

where $F(t,x,u) \equiv q + Cx + Du$. Given an initial value c for $x(0;c)$, with the smoothing function $\phi : \mathbb{R}^3 \to \mathbb{R}$,

$$\phi(\mu, a, b) = \frac{1}{2}(a + b - \sqrt{(a-b)^2 + 4\mu^2})$$

corresponding to the nonsmooth algebraic equation (3.9), we can obtain the solution map $u(x(t;c))$ of (3.9) that is semismooth at $\mu = 0$ (see the algorithm for a P_0- LCP in [21] for the details).

However, note that once $u(x(t;c))$ is solved, the variable μ is required to approach zero. On the other hand, in Newton's method the Jacobian of the function

$$h(c) \equiv \Gamma(x(0), x(T)),$$

$J_c h(c)$ must be calculated (for simplicity of explanation, h is supposed to be smooth around any ε with $\varepsilon > 0$), which involves computing $J_x u(x(T;c))-$ the Jacobian of $u(x(T;c))$ (see (2.6)). Due to the fact that $u(x(T;c))$ is discontinuous at $\mu = 0$, if μ is such a small number as 10^{-6}, it would lead to an ill-conditioned Jacobian $J_x u(x(T;c))$ so that Newton's method could lose its efficiency. To conquer this problem, we have to set a lower bound, say ε, to μ in each inner iteration (iteration for solving LCP) in order to prevent it from being too small. By doing so, ε will act as a parameter in the ODE function $f^{\varepsilon}(t,x) \equiv \hat{f}(t,x) + Bu^{\varepsilon}(x)$. As a result, the solution to the differential equation and the boundary function will be $x^{\varepsilon}(t;c)$ and $h^{\varepsilon}(c)$, respectively.

Following the same notation that is described in Section 3.1, $f^{\varepsilon}(t,x)$ is to be replaced by $g(t,\varepsilon,x)$. Similarly, consider the initial-valued ODE (3.1), with a new variable $\tau(t) \equiv \varepsilon$ being introduced in, we get to the objective boundary equation (3.6) and approximate its solution c^* via the smoothing Newton method as the outer iteration for system (3.7):

$$E(\varepsilon, c) = \begin{bmatrix} \varepsilon \\ \hat{h}(\varepsilon, c) \end{bmatrix} = 0.$$

Within the inner steps for LCP, μ is bounded below by the parameter ε. For this purpose, throughout the entire outer iterations for solving (3.7), we can let ε be a decreasing sequence $\{\varepsilon^k\}$ in the manner that each ε^k in the kth step is imposed to be the lower bound of μ. Therefore, with the sequence $\{\varepsilon^k\}$ tending to zero, μ will become an infinitesimal at the final step. More details will be shown in the algorithm for the smoothing Newton method in chapter 4.

Chapter 4

A Smoothing Newton Method

In this chapter, we develop a smoothing Newton method for (3.7), whose fundamental is the version of the QSZ smoothing Newton algorithm in [29]. Besides, the convergence analysis and the numerical results are also reported in later sections.

4.1 Algorithm for Smoothing Newton Methods

Recall the problem we get started, which is defined by (3.7):

$$E(\varepsilon, c) = \begin{bmatrix} \varepsilon \\ \hat{h}(\varepsilon, c) \end{bmatrix} = 0,$$

where

$$\hat{h}(\varepsilon, c) \equiv \widehat{\Gamma}((\varepsilon, c), y(T; \varepsilon, c)).$$

Here $y(T; \varepsilon, c)$ is the solution to (3.4):

$$\begin{cases} \dot{y} = p(t, y), & 0 \leq t \leq T \\ y(0) = \begin{pmatrix} \varepsilon \\ c \end{pmatrix}, & \end{cases}$$

From the discussion in chapter 3, $p(t,y)$ and $\hat{h}(\varepsilon,c)$ are continuously differentiable for any $\varepsilon > 0$.

From (3.7), for any $\varepsilon > 0$ a straightforward calculation yields:

$$J_{(\varepsilon,c)}E(\varepsilon,c) = \begin{bmatrix} 1 & 0 \\ J_\varepsilon \hat{h}(\varepsilon,c) & J_c\hat{h}(\varepsilon,c) \end{bmatrix}. \tag{4.1}$$

By the definition of $\hat{h}(\varepsilon,c)$, we have:

$$J_\varepsilon \hat{h}(\varepsilon,c) = \widehat{B}_T J_\varepsilon y(T;\varepsilon,c)$$

and

$$J_c\hat{h}(\varepsilon,c) = \widehat{B}_0 + \widehat{B}_T J_c y(T;\varepsilon,c),$$

where $\widehat{B}_0 = J_{\mathbf{u}}\widehat{\Gamma}(\mathbf{u},\mathbf{v})\Big|_{\substack{\mathbf{u}=(\varepsilon,c)\\ \mathbf{v}=y(T;\varepsilon,c)}}$ and $\widehat{B}_T = J_{\mathbf{v}}\widehat{\Gamma}(\mathbf{u},\mathbf{v})\Big|_{\substack{\mathbf{u}=(\varepsilon,c)\\ \mathbf{v}=y(T;\varepsilon,c)}}$.

Next we derive the calculation of $J_\varepsilon y(T;\varepsilon,c)$ and $J_c y(T;\varepsilon,c)$. As is mentioned in Section 2.1, $J_c y(t;\varepsilon,c)$ is the $(n+1)\times n$ fundamental solution to the following system:

$$\begin{cases} \dot{Y}(t) &= \widehat{A}(t)Y(t), \qquad 0 \le t \le T \\ Y(0) &= \begin{pmatrix} \mathbf{0} \\ I \end{pmatrix} \end{cases} \tag{4.2}$$

with $\widehat{A}(t, y(t;\varepsilon,c)) = J_y p(t,y)$. Similarly, $J_\varepsilon y(t;\varepsilon,c)$ is the $(n+1)\times 1$ solution to

$$\begin{cases} \dot{Y}(t) &= \widehat{A}(t)Y(t), \qquad 0 \le t \le T \\ Y(0) &= \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \end{cases} \tag{4.3}$$

For the simplicity of derivation, we assume the differential function is separable in variable $x(t;\varepsilon,c)$ and $u(x)$. Then, from (3.5), $p(t,y)$ is defined as:

$$p(t,y) \equiv \begin{bmatrix} 0 \\ \hat{f}(t,x) + Bu(x) \end{bmatrix}, \tag{4.4}$$

where $u(x)$ is the solution to a nonsmooth problem. Hence, $J_y p(t,y)$ (the Jacobian of $p(t,y)$ with respect to $y \equiv (\tau, x)^T$) becomes:

$$J_y p(t,y) = \begin{bmatrix} 0 & 0 \\ 0 & J_x \hat{f}(t,x) + BJ_x u(x). \end{bmatrix}. \tag{4.5}$$

By observation from (4.5), one can see that $J_x u(x)$ serves as a very critical part in computing $J_y p(t,y)$ (Actually, it confirms what we have mentioned in Section 3.2 that the Jacobian of $\hat{h}(\varepsilon, c)$ will be affected by $J_x u(x)$).

Given $\bar{\varepsilon} > 0$ and $\gamma \in (0,1)$ such that $\gamma|\bar{\varepsilon}| < 1$. Let $z := (\varepsilon, c)$, $z^k := (\varepsilon_k, c^k)$ and $\bar{z} := (\bar{\varepsilon}, 0) \in \mathbb{R} \times \mathbb{R}^n$. Define $\psi : \mathbb{R}^{n+1} \to \mathbb{R}_+$ and $\beta : \mathbb{R}^{n+1} \to \mathbb{R}_+$ by

$$\psi(z) := \|E(z)\|^2 \quad \text{and} \quad \beta(z) := \gamma \min\{1, \psi(z)\}, \tag{4.6}$$

respectively.

Algorithm 4.1

Step 0: Choose $\delta \in (0,1)$ and $\sigma \in (0, 1/2)$. Let $\varepsilon^0 := \bar{\varepsilon}$, $c^0 \in \mathbb{R}^n$ be an arbitrary point and $k := 0$.

Step 1: If $\|E(z^k)\| \leq \text{tol}$, then stop. Otherwise, let $\beta_k := \beta(z^k)$.

Step 2: Compute $\triangle z^k := (\triangle\varepsilon_k, \triangle c^k) \in \mathbb{R} \times \mathbb{R}^n$ by

$$E(z^k) + J_z E(z^k) \triangle z^k \;=\; \beta_k \bar{z}. \tag{4.7}$$

Step 3: Let λ_k be the maximum of the values $1, \delta, \delta^2, \cdots$ such that

$$\psi(z^k + \lambda_k \triangle z^k) \;\leq\; [1 - 2\sigma(1 - \gamma|\bar{\varepsilon}|)\lambda_k]\psi(z^k). \tag{4.8}$$

Step 4: Set $z^{k+1} := z^k + \lambda_k \triangle z^k$ and $k := k+1$. Go to Step 1.

4.2 Convergence Analysis

4.2.1 Global Convergence

Since Algorithm 4.1 derives from QSZ method in [29], the global convergence discussed therein provides a way that we can use to show the global convergence of Algorithm 4.1.

Denote

$$\Omega := \{z = (\varepsilon, c) \in \mathbb{R} \times \mathbb{R}^n \mid \varepsilon \geq \beta(z)\bar{\varepsilon}\}.$$

Then, because for any $z \in \mathbb{R}^{n+1}$, $\beta(z) \leq \gamma < 1$, it follows that for any $x \in \mathbb{R}^n$,

$$(\bar{\varepsilon}, c) \in \Omega.$$

Under the assumption [29, page 15] that:

(D) For every $k \geq 0$, if $\varepsilon^k > 0$ and $z^k \in \Omega$, then $J_z E(z^k)$ is nonsingular; and

(E) and for any accumulation point $z^* = (\varepsilon^*, c^*) = \lim_{k\to\infty} z^k$, if $\varepsilon^* > 0$ and $z^* \in \Omega$, then $\partial E(z^*)$ is nonsingular,

the sequence $\{z^k\} := \{(\varepsilon^k, c^k)\}$ generated by Algorithm 4.1 is well-defined as a result of the proposition 7 and proposition 8 in [29, Section 4]. This means for each $k \geq 0$, if $\varepsilon^k > 0$, $z^k \in \Omega$ and $J_z E(z^k)$ is invertible, then

$$\varepsilon^{k+1} > 0 \quad \text{and} \quad z^{k+1} \in \Omega.$$

Furthermore, we have the following theorem for the global convergence according to Theorem 4 in [29]

Theorem 4.2.1. *Suppose that Assumption (D) and (E) are satisfied. Then an infinite sequence $\{z^k\}$ is generated by Algorithm 4.1 and each accumulation point $\tilde{z}$ of $\{z^k\}$ is a solution of E(z)=0.*

4.2.2 Superlinear and Quadratic Convergence

The superlinear or quadratic convergence for Algorithm 4.1 tied with the semismoothness of the function $E(\varepsilon, c)$ at its nondifferentiable point $(0, c)$, for any $c \in \mathbb{R}^n$. To show this, we need to prove the function $\hat{h}(\varepsilon, c) \equiv \widehat{\Gamma}((\varepsilon, c), y(T; \varepsilon, c))$ is semismooth at $(0, c)$. Since in Section 3.1, $\hat{h}(\varepsilon, c)$ is defined as the smoothing function of the original $h(c)$, its semismoothness, therefore only depends on $y(T; \varepsilon, c)-$ the solution map to the ODE (3.4). In this section, we prove the semismoothness of $y(T; \varepsilon, c)$ and the superlinear (quadratic) convergence.

Lemma 4.2.1. Suppose $p(t, y)$ in (3.4) is Locally Lipschitz continuous in a neighborhood of the trajectory starting at $(0, c^*)$. Then, there exists a neighborhood $\mathcal{N}_0$ of $(0, c^*)$ such that for every $(\varepsilon, c) \in \mathcal{N}_0$, the ODE (2.1) has a unique solution $y(t; \varepsilon, c)$ on $[0, T]$ which satisfies, for any (ε, c) and $(\varepsilon', c') \in \mathcal{N}_0$,

$$\|y(t; \varepsilon, c) - y(t; \varepsilon', c')\| \leq e^{Lt} \left\| \begin{pmatrix} \varepsilon - \varepsilon' \\ c - c' \end{pmatrix} \right\|, \quad \forall\, t \in [0, T].$$

Remarks: Recall the definition of $y(t; \varepsilon, c)$ in Section (3.1), we have $y(t; -|\varepsilon|, c) := y(t; |\varepsilon|, c)$. This means ε here is not restricted to be nonnegative anymore.

This Lemma (4.2.1) is an immediate consequence of Lemma (2.1.1). Because of the reformulation that transforms the parameter ε to the initial value of another variable τ, the Lipschitz continuity on initial data for traditional ODEs remains valid in the nonsmooth ODEs (2.1.1).

Theorem 4.2.2. *Suppose the functions $p(t, y)$ is continuously differentiable at $y(t; \varepsilon, c)$ for any $\varepsilon > 0$, and semismooth at $y(t; 0, c^*)$. Then the solution map $y(t; \varepsilon, c)$ to the ODE system (3.4) is semismooth at $(0, c^*)$, $c^* \in \mathbb{R}^n$ for all $t \in [0, T]$.*

(The inspiration of the proof to Theorem 4.2.4 from [24, Theorem 8])

Proof. We know that $y(t; \varepsilon, c)$ is smooth for any (ε, c) except for the set $S := \{(0, c) | c \in \mathbb{R}^n\}$, which is Lebesgue measure zero in $\mathbb{R} \times \mathbb{R}^n$. By the Theorem

2.2.1, it is sufficient to prove that for any $\varepsilon > 0$ (which indicates the existence of $J_{(\varepsilon,c)}y(t;\varepsilon,c)$),

$$y(t;\varepsilon,c) - y(t;0,c^*) - J_{(\varepsilon,c)}y(t;\varepsilon,c)\begin{pmatrix}\varepsilon\\ c-c^*\end{pmatrix} = o\left\|\begin{pmatrix}\varepsilon\\ c-c^*\end{pmatrix}\right\|. \tag{4.9}$$

For our convenience of notation, denote $z := (\varepsilon, c)$ and $z_0 := (0, c^*)$.

Then, we have

$$y(t;z) = \int_0^t p(\omega, y(\omega;z))d\omega + z, \tag{4.10}$$

$$y(t;z_0) = \int_0^t p(\omega, y(\omega;z_0))d\omega + z_0, \tag{4.11}$$

and

$$J_z y(t;z)(z-z_0) = \int_0^t [J_y p(\omega, y(\omega;z))J_z y(\omega;z)](z-z_0)d\omega + I(z-z_0). \tag{4.12}$$

Together with (4.10),(4.11) and (4.12), the left-hand side of (4.9) is:

$$\int_0^t [p(\omega, y(\omega;z)) - p(\omega, y(\omega;z_0)) - J_y p(\omega, y(\omega;z))(y(\omega;z) - y(\omega;z_0))$$
$$+ J_y p(\omega, y(\omega;z))(y(\omega;z) - y(\omega;z_0)) - J_y p(\omega, y(\omega;z))J_z y(\omega;z)(z-z_0)]\, d\omega.$$

Write

$$e(t;z,z_0) := y(t;z) - y(t;z_0) - J_z y(t;z)(z-z_0)$$
$$e_p(t;z,z_0) := p(t,y(t;z)) - p(t,y(t;z_0)) - J_y p(t,y(t;z))(y(t;z) - y(t;z_0)).$$

$p(t, y(t;z))$ is continuously differentiable hence locally Lipschitz continuous near $y(t;z) = y(t;\varepsilon,c)$ for $\varepsilon > 0$, which means $J_y p(t, y(t;z))$ is bounded with the constant L_p.

This yields

$$\|e(t;z,z_0)\| \leq \int_0^t \|e_p(\omega;z,z_0)\|d\omega \; + \; L_p \int_0^t \|e(\omega;z,z_0)\|d\omega.$$

Using the Gronwall's lemma again, we obtain

$$\|e(t;z,z_0)\| \le \int_0^t \|e_p(\omega;z,z_0)\|d\omega \;+\; L_p\int_0^t(\int_0^\omega \|e_p(s;z,z_0)\|ds)e^{L_p(t-\omega)}\,d\omega.$$

On the other side, $p(t,y)$ is semismooth at $y(t;z_0)$, and from Lemma 4.2.1 $y(t;z)$ is locally Lipschitz continuous near z_0.

We deduce

$$\lim_{z\to z_0}\frac{e_p(\omega;z,z_0)}{\|z-z_0\|} = \lim_{z\to z_0}\frac{e_p(\omega;z,z_0)}{\|y(\omega;z)-y(\omega;z_0)\|}\cdot\frac{\|y(\omega;z)-y(\omega;z_0)\|}{\|z-z_0\|} = 0.$$

An application of the Lebesgue dominated convergence theorem leads to

$$\lim_{z\to z_0}\frac{e(t;z,z_0)}{\|z-z_0\|} = 0,$$

which verify the $y(t;z) = y(t;\varepsilon,c)$ is semismooth at $z_0 = (0,c^*)$. □

Theorem 4.2.3. *Suppose all the conditions in Theorem 4.2.2 are satisfied. Moreover, the functions $p(t,y)$ is uniformly strongly semismooth at $y(t;z_0)$, i.e.,*

$$e_p(t;z,z_0) = O(\|y(t;z)-y(t;z_0)\|^2), \quad \textit{for all} \;\; t\in[0,T]. \tag{4.13}$$

Then the solution map $y(t;z)$ to the ODE system (3.4) is strongly semismooth at z_0 for all $t\in[0,T]$.

Proof. Refer to the proving process in Theorem 4.2.2, we get to

$$\|e(t;z,z_0)\| \le \int_0^t \|e_p(\omega;z,z_0)\|d\omega \;+\; L_p\int_0^t(\int_0^\omega \|e_p(s;z,z_0)\|ds)e^{L_p(t-\omega)}\,d\omega.$$

From (4.13), $e_p(s;z,z_0)$ is estimated by

$$\lim_{z\to z_0}\frac{e_p(s;z,z_0)}{\|z-z_0\|^2} = M,$$

where $M>0$ is a constant that irrelevant to the integral variable s, hence the integral can be re-evaluated and apply the Lebesgue dominated convergence theorem again, we have

$$\begin{aligned}\lim_{z\to z_0}\frac{e(t;z,z_0)}{\|z-z_0\|^2} &\le Mt + L_p\int_0^t M\omega\ e^{L_p(t-\omega)}\ d\omega \\ &= \frac{M(e^{L_p t}-1)}{L_p}.\end{aligned}$$

Since $0 \le t \le T$, we deduce

$$\lim_{z\to z_0}\frac{e(t;z,z_0)}{\|z-z_0\|^2} \le \frac{M(e^{L_p T}-1)}{L_p},$$

which establishes the strongly semismoothness of $y(t;z)$ at $y(t;z_0)$. □

Remarks: The assumption that $p(t,y)$ is uniformly strongly semismooth at $y(t;0,c^*)$ for all $t \in [0,T]$ is not unrealistic. The following simple example shows the existence of such a function p. Consider a nonsmooth initial value problem:

$$\begin{cases} \dot{p} = \max(0,p), \quad 0 \le t \le T \\ p(0) = c. \end{cases} \tag{4.14}$$

One can easily see that the solution $p(t;c)$ to the system (4.14) is

$$p(t;c) = \begin{cases} ce^t & c > 0 \\ c & c \le 0, \end{cases} \tag{4.15}$$

which is nondifferentiable at $c = 0$, so let us check the semismoothness at this point. For $c = 0$, we have

$$\begin{aligned} p(t;c') - p(t;0) - J_c p(t;c')(c'-0) = 0 \qquad c' > 0 \\ p(t;c') - p(t;0) - J_c p(t;c')(c'-0) = 0 \qquad c' < 0. \end{aligned}$$

This shows that $p(t;c)$ is strongly semismooth at $c = 0$ for all $t \in [0,T]$.

Theorem 4.2.4. *Suppose that Assumption (D) and (E) are satisfied and z^* is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm 4.1.*

Suppose that E is semismooth at z^* and that all $V \in \partial E(z^*)$ are nonsingular. Then the whole sequence $\{z^k\}$ converges to z^* superlinearly,

$$\|z^{k+1} - z^*\| = o(\|z^k - z^*\|) \tag{4.16}$$

Furthermore, if E is strongly semismooth at z^*, then

$$\|z^{k+1} - z^*\| = O(\|z^k - z^*\|^2) \tag{4.17}$$

Proof. From Theorem 4.2.1 that z^* is a solution of $E(z) = 0$. Then, from [27, Proposition 3.1], for all z^k sufficiently close to z^*,

$$\|(J_z E(z^k))^{-1}\| = O(1).$$

Recall the Step 2 in Algorithm 4.1, we have

$$\triangle z^k = (J_z E(z^k))^{-1}(-E(z^k) + E(z^*) + \beta_k \bar{z}),$$

and consider

$$\begin{aligned} \|z^k + \triangle z^k - z^*\| &= \| -(J_z E(z^k))^{-1}(E(z^k) - E(z^*)) + z^k - z^* + (J_z E(z^k))^{-1}\beta_k \bar{z}\| \\ &\leq \|(J_z E(z^k))^{-1}(E(z^k) - E(z^*)) - (z^k - z^*)\| + \|(J_z E(z^k))^{-1}\beta_k \bar{z})|. \end{aligned}$$

Assume E is semismooth (strongly semismooth, respectively), we have

$$(J_z E(z^k))^{-1}(E(z^k) - E(z^*)) - (z^k - z^*) = (J_z E(z^k))^{-1} o(\|z^k - z^*\|).$$

On the other side, from the definition of β_k and the fact that $z^k \to z^*$ as $k \to \infty$, for all k sufficiently large, $\beta_k = \gamma\psi(z^k)$. Hence $\|\beta_k \bar{z}\| \leq \gamma \|\bar{z}\| \|\psi(z^k)\|$. We get

$$\begin{aligned} \|z^k + \triangle z^k - z^*\| &= o(\|z^k - z^*\|) + O(\psi(z^k)) \\ (&= O(\|z^k - z^*\|) + O(\psi(z^k)). \end{aligned} \tag{4.18}$$

Because E is locally Lipschitz continuous near z^* for all z^k close to z^*,

$$\psi(z^k) = \|E(z^k)\|^2 = O(\|z^k - z^*\|^2). \tag{4.19}$$

Together with (4.18), (4.19) for all z^k sufficiently close to z^*, we deduce

$$\begin{aligned} \|z^k + \triangle z^k - z^*\| &= o(\|z^k - z^*\|) \\ (&= O(\|z^k - z^*\|^2)). \end{aligned} \tag{4.20}$$

Claim

$$z^{k+1} = z^k + \triangle z^k;$$

and after substituting z^{k+1} into (4.20), the superlinear (quadratical) convergence is verified.

Follow the proof for Theorem 8 in [29], we obtain

$$\begin{aligned} \psi(z^k + \triangle z^k) &= o(\psi(z^k)) \\ (&= O(\psi(z^k)^2)). \end{aligned} \tag{4.21}$$

Compare (4.21) with (4.8) in Step 3 of Algorithm 4.1, we conclude that the claim is true. □

4.3 Numerical Experience

In this section, we present some numerical experiments for Algorithm 4.1 implemented in Matlab to see the behavior of the smoothing Newton method. All the models we use are based on the DCP (2.16)

$$\begin{aligned} \dot{x}(t) &= \hat{f}(t,x) + Bu \\ 0 \le u &\perp q + Cx + Du \ge 0 \\ 0 &= \Gamma(x(0), x(T)), \end{aligned}$$

where the LCP is generated from a convex quadratically constrained (QP) programming problem [21, Section 6] with D being assumed to be a P_0-matrix, while the vector q is randomly given.

Recall the Jacobian $J_y p(t,y)$ (4.5) calculated in Section 4.1 which involves computing $J_x u(x)$, where $u(x)$ is the solution to the LCP for a given x. This requires us to analyze the P_0−LCP [21, Section 2].

Let $\phi : \mathbb{R}^3 \to \mathbb{R}$

$$\phi(\mu, a, b) = a + b - \sqrt{(a-b)^2 + 4\mu^2}$$

and let $\Phi : \mathbb{R}^{2n+1} \to \mathbb{R}^n$ be:

$$\Phi(\mu, u, s) := \begin{pmatrix} \phi(\mu, u_1, s_1) \\ \vdots \\ \phi(\mu, u_n, s_n). \end{pmatrix}.$$

Then, given the initial value for $x(t;c)$ solving the LCP (3.8) is equivalent to finding out the root for $H(\mu, u, s; x)$:

$$H(\mu, u, s; x) := \begin{pmatrix} \mu \\ s - Du - Cx - q \\ \Phi(\mu, u, s) + \alpha(\mu)u \end{pmatrix}, \tag{4.22}$$

in which $\alpha : \mathbb{R} \to \mathbb{R}_+$ is a twice continuously differentiable function satisfying $\alpha(\mu) > 0$ for $\mu \neq 0$, and

$$\alpha(0) = 0, \quad |\alpha(\mu)| = O(\mu^3), \quad \text{and } |\alpha'(\mu)| = O(\mu^2).$$

When $\mu \neq 0$, denote $JH(\mu, u, s; x)$ as the Jacobian with respect to the vector $(\mu, u, s)^T$.

$$JH(\mu, u, s; x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -D & I \\ J_\mu \Phi + \alpha'(\mu)u & J_u\Phi + \alpha(\mu)I & J_s\Phi \end{pmatrix}. \tag{4.23}$$

Moreover, $H(\mu, u, s; x) = 0$ infers

$$JH(\mu, u, s; x) \begin{pmatrix} J_x \mu \\ J_x u(x) \\ J_x s(x) \end{pmatrix} + J_x H(\mu, u, s; x) = 0. \tag{4.24}$$

Combing (4.23) and (4.24) yields a linear equation in $J_x u(x)$ and $J_x s(x)$, which is easy to calculate. Up to now, we have demonstrated the way to calculate $J_{(\varepsilon,c)}E(\varepsilon,c)$.

Example 4.3.1. We consider the boundary value ODE as below

$$\begin{cases} \dot{x} = Ax + q(t) + Bu(x), & 0 \le t \le 1 \\ 0 = w\max(x(0), x(1)) + (1-w)(B_0 x(0) + B_1 x(1) - b) \end{cases}$$

with $u(x)$ satisfying (3.8). The weight in the boundary condition w varies between $[0, 1]$.

The data of ODE is given by:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{pmatrix}; \qquad q(t) = \begin{pmatrix} 0 \\ 0 \\ (\pi^3 + \pi)\sin \pi t + (2 + 2\pi^2)\cos \pi t \end{pmatrix}$$

and

$$B_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad B_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \quad b = \begin{pmatrix} 1.6348 \\ 0 \\ 1.1116 \end{pmatrix}.$$

Without the LCP term $u(x)$ and the nonsmooth function $\max(x(0), x(1))$ in the boundary condition, this problem would be a very typical boundary ODE system and can be well solved by using the single shooting method. By means of adjusting the value of the coefficient matrix B, we can control the degree of how the LCP affects the pure ODE system.

Example 4.3.2. Consider the nonsmooth ODE as below

$$\begin{cases} \dot{x} = \hat{f}(t, x) + Bu(x), & 0 \le t \le 1 \\ 0 = w\max(x(0), x(1)) + (1-w)(B_0 x(0) + B_1 x(1)) \end{cases} \tag{4.25}$$

with $u(x)$ solving (3.8).

We write $x = (x_1, x_2)^T$ and define the data in the problem as

$$\hat{f}(t,y) = \begin{pmatrix} x_2(t) \\ -e^{x_1(t)+1} \end{pmatrix}; \quad B_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \quad B_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Note that both boundary functions of Example (4.3.1) and Example (4.3.2) are no longer continuously differentiable but the linear combination of a nonsmooth function and a linear operator. Such kind of problem has never been attacked before, while the numerical reports for (4.25) by means of Algorithm 4.1 is quite encouraging.

Refer to Section 3.1, in order to use the smoothing Newton method, the boundary function

$$h(c) \equiv w \max(c, x(1;c)) + (1-w)(B_0 c + B_1 x(1;,c))$$

must be smoothed first. Let

$$\begin{aligned} \tilde{h}^{\varepsilon}(c) &= \tfrac{1}{2} w (c + x^{\varepsilon}(1;c) + \sqrt{(c - x^{\varepsilon}(1;c))^2 + 4\varepsilon^2}) \\ &\quad + (1-w)(B_0 c + B_1 x^{\varepsilon}(1;c)) \end{aligned} \tag{4.26}$$

be its smoothing function, where ε arises from the procedure for solving the LCP (3.8). Denote the dimension of the variable $x(t)$ by n. Then from (4.26), for any $\varepsilon > 0$ a straightforward calculation yields

$$\begin{aligned} J_{\varepsilon}\tilde{h} &= w J_{\varepsilon} \max + (1-w) B_1 J_{\varepsilon} x^{\varepsilon}(1;c) \\ J_c \tilde{h} &= w J_c \max + (1-w)(B_0 + B_1 J_c x^{\varepsilon}(1;c)), \end{aligned}$$

where the calculations of $J_{\varepsilon} x(1)$ and $J_c x(1)$ are analogous to that of (4.2) and (4.3), respectively,

$$J_{\varepsilon} \max := \frac{1}{2} \mathrm{vec} \left\{ J_{\varepsilon} x_i^{\varepsilon}(1;c) + \frac{(c_i - x_i^{\varepsilon}(1;c_i))(-J_{\varepsilon} x_i^{\varepsilon}(1;c_i)) + 4\varepsilon}{\sqrt{(c_i - x_i^{\varepsilon}(1;c_i))^2 + 4\varepsilon^2}} : \ i = 1, 2, \cdots, n \right\},$$

and

$$J_c \max := \frac{1}{2}(\, I + J_c x^{\varepsilon}(1;c) + F - F J_c x^{\varepsilon}(1;c)),$$

where I denotes the 2×2 identity matrix,

$$F := \mathrm{diag}\left\{ \frac{c_i - x_i^{\varepsilon}(1;c_i)}{\sqrt{(c_i - x_i^{\varepsilon}(1;c_i))^2 + 4\varepsilon^2}} : \; i = 1, 2, \cdots, n \right\}.$$

For any positive integers n_1 and n_2, let rand(n_1, n_2) denote a matrix by $n_1 \times n_2$ whose each element is randomly chosen in $(0, 1)$. Throughout the computational experiments, the parameters used in Algorithm 4.1 were chosen as $\delta = 0.5$, $\sigma = 0.0001$, $\varepsilon^0 = 1$, $\gamma = 0.1 \min\{1, 1/\varepsilon^0\}$. Let the coefficient B of $u(x)$ be an identity matrix. We used $\|E(z^k)\| \leq 10^{-6}$ as the stopping rule. With varies randomly given starting points c^0, the two problems are tested ten times for different w by using algorithm 4.1. The iteration numbers are listed in Table 4.1 and Table 4.2, respectively.

c^0	rand(0, 1)			10*rand(0, 1)			100*rand(0, 1)		
w	0	0.5	1	0	0.5	1	0	0.5	1
N	13	15	14	16	14	13	16	18	25
	13	20	26	13	12	18	19	15	19
	14	17	17	22	19	26	19	16	20
	13	19	16	12	21	20	15	17	20
	10	18	19	14	16	18	18	19	21
	15	12	15	12	14	13	18	19	20
	12	13	19	15	18	15	20	22	24
	17	13	16	11	13	13	22	21	21
	14	15	12	17	17	18	15	15	15
	14	16	21	16	14	16	20	16	14

Table 4.1: The numerical results for Example 4.3.1, where N is iteration number

c^0	rand(0, 1)			$-$rand(0, 1)			$5 *$ rand(0, 1)		
w	0	0.5	1	0	0.5	1	0	0.5	1
N	6	5	4	8	6	5	7	5	4
	6	5	4	6	5	5	5	6	4
	10	6	4	4	7	6	6	8	4
	6	5	4	9	6	5	4	5	4
	5	6	4	8	6	5	11	7	4
	4	6	4	7	6	6	6	6	4
	5	5	4	6	6	6	5	7	4
	5	5	4	11	5	5	6	6	4
	6	5	5	5	6	4	7	5	4
	5	4	4	11	5	5	6	8	4

Table 4.2: The numerical results for Example 4.3.2, where N is iteration number

Chapter 5

Conclusions

Example 4.3.1 and Example 4.3.2 are only of three or two dimensions, which cannot fully show the advantages of Algorithm 4.1 for the nonsmooth ODEs, and one can test some large scaled boundary valued problems arising from the constrained control systems. Furthermore, During the procedure of the computation, not all the problems can be observed the quadratic convergence but only the superlinear convergence, whereas the former is provided theoretically. One of the reasons comes from the inexact finite difference method for the initial value ODE, which means $E(\varepsilon^k, c^k)$ is just an approximation to its real value in each step k. Due to this fact, an inexact smoothing Newton method might be developed to make up for the computational errors. In addition, the uniformly strong semismoothness assumed in Theorem 4.2.3 still needs a deep discussion. At least, we have shown the reasonability of this assumption. One may derive certain conditions under which the assumption can be satisfied.

Another point worthy of discusstion is that the whole structure of solving the nonsmooth ODE is based on the single shooting method. Meanwhile, there are many other classical ways for the boundary value ODEs such as: implicit Runge-Kutta methods, finite element methods, etc. that could be combined with the smoothing Newton method. We do not choose these methods mainly because of the potential large scale which might be involved in the numerical practice. Nevertheless, there is still feasibility of this new idea, which remains to be a further study.

Finally, even if the boundary condition $h(c) \equiv \Gamma(c, x(T; c)) = 0$ is already smooth in $x(t; c)$, we can still add a regularization term $\alpha(\varepsilon)c$ in the smoothing function $E(\varepsilon, c)$ as we do in the third term of (4.22) to improve the Jacobian condition $J_{(\varepsilon,c)}E$ and accelerate the convergence rate of Algorithm 4.1. As an extension, chances are that the nonsmooth boundary-valued ODE system could be expected to a more general form: a differential algebraic equation (DAE) problem [5]

$$\begin{cases} \dot{x} = f(t, x, u) \\ 0 = g(t, x, u), \end{cases}$$

so that the robustness of the smoothing Newton method for the differential systems can be fully tested. All that have been mentioned above leave us significant research topics in the future.

Bibliography

[1] M. Anitescu and G.D. Hart. A constraint-stabilized time-stepping for multibody dynamics with contact and friction. Preprint ANL/MCS-P, Dvivision of Mathematics and Computer Science, Argonne National Laborary, (2002) 1002-1002.

[2] U. Ascher and P. Lin. Sequential regularization methods for simulating mechanical systems with many closed loops. *SIAM J. Sci. Comput.*, 21 (1999) 1244-1262.

[3] M. Anitescu and F.A. Potra. Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems. *ASME Nonlinear Dynamics*, 4 (1997) 231-247.

[4] M. Anitescu, F.A. Potra and D.Stewart. Time-stepping for three-dimensional rigid-body dynamics. *Computer Methods in Applied Mechanics and Engineering*, 177 (1999) 183-197.

[5] U.M. Ascher and L.R. Petzold. Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. SIAM (1988).

[6] B. Brogliato. Nonsmooth Impact Mechanics. *Models, Dynamics and Control*, Springer-Verlag (London 1999).

[7] T. Basar and G.J. Olsder. Dynamic Noncooperative Game Theory. SIAM Series in Classics in Applied Mathematics (Philadelphia 1999) [Revised, updated version of the 1995 Academic Press book with the same title].

[8] B. Brogliato, A.A. ten Dam, L. Paoli, and M. Abadie. Numerical simulation of finite dimensional multibody nonsmooth mechanical system. Manuscript, Laboratoire d'Automatique de Grenoble (May 2000).

[9] A.E. Bryson and Y. Ho. Applied Optimal Control: Optimization, Estimation, and Control. John Wiley and Sons (New York 1975).

[10] M.K. Camlibel. Complementarity Methods in the Analysis of Piecewise Linear Dynamical System. Ph.D. thesis, Center for Economic Research, Tilburg University, The Netherlands (May 2001).

[11] F.H. Clarke. Optimization and Nonsmooth Analysis. Wiley (New York 1983).

[12] A.L. Dontchev and W.W. Hager. The Euler Approximation in State Constrained Optimal Control. *Mathematics of Computation*, 70 (2000) 173-203.

[13] A.L. Dontchev, W.W. Hager and K. Malanowski. Error bounds for the Euler approximation of a state and control constrained optimal control problem. *Numerical Functional Analysis and Optimization*, 21 (2000) 653-682.

[14] A.L. Dontchev, W.W. Hager and V.M. Veliov. Second-order Runge-Kutta approximations in constrained optimal control. *SIAM Journal on Numerical Analysis*, 38 (2000) 202-226.

[15] E. Dockner, S. Jergensen, N.V. Long and G. Sorger. Differential Games in Economics and Management Science. Cambridge University Press (Cambridge 2000).

[16] F. Facchinei and J.S. Pang. Finite-dimensional Inequalies and Complementarity Problems. Springer-Verlag (New York 2003).

[17] M.S. Gowda. Inverse and implicit function theorems for H-differentiable and semismooth functions. *Optimization Methods and Software*, 19 (2004) 443-460.

[18] W.W. Hager. Lipschitz continuity for constrained processes. *SIAM J. Contr. Optim.*, 17 (1979) 321-338.

[19] W.W. Hager. Numerical Analysis on Optimal Control. *International Series of Numerical Mathematics*, Vol. 139 Birkhauser Verlag, Basel/Switzerland (2001) 83-93.

[20] W.P.H. Heemels. Linear Complementarity Systems: A Study in Hybrid dynamics. Ph.D. thesis, Department of Electrical Engineering, Eindhoven University of Technology (November 1999).

[21] Z. Huang, L. Qi and D. Sun. Sub-Quadratic Convergence of a Smoothing Newton Algorithm for the P_0- and Monotone LCP. *Mathematical Programming*, 99 (2004) 423–441.

[22] R. Mifflin. Semismooth and Semiconvex functions in constrained optimization. *SIAM J. Control and Optim.*, Vol. 15, No.6 (November 1997).

[23] J.S. Pang and D.E. Stewart. Differential Variational Inequalities. Technical Reports, (2003).

[24] J.S. Pang and D.E. Stewart. Solution Dependence on Initial Conditions in Differential Variational Inequalities. Technical Reports, (November 2004).

[25] J.S. Pang and D.E. Stewart. Strongly regular variational systems. Technical Reports, (January 2005).

[26] J.S. Pang, D. Sun and J. Sun. Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Mathematics of Operations Research*, 28 (2003) 39-63.

[27] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Mathematical Programming*, 58 (1993) 353-367.

[28] L. Qi and D. Sun. Smoothing Functions and a Smoothing Newton Method for Complementarity and Variational Inequality Problems. *Journal of Optimization Theory and Applications*, 113 (2002) 121-147.

[29] L. Qi, D. Sun and G. Zhou. A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities. *Mathematical Programming*, 87 (2000) 1-35.

[30] S. Shi. Smooth Convex Approximation and Its Applications. Master thesis, Department of Mathematics, National University of Singapore, (July 2004).

[31] A.M. Stuart and A.R. Humphries. Dynamical Systems and Numerical Analysis. Cambridge University Press, (Cambridge; New York 1998).

[32] D. Sun and L. Qi. Solving variational inequality problems via smoothing-nonsmooth reformulations. *Journal of Computational and Applied Mathematics*, 129 (2001) 37-62.

[33] D. Sun and J. Sun. Löewner's operator and spectral functions in Euclidean Jordan algebras. Technical Reports, (December 2004).

[34] J. Warga. Fat homeomorphisms and unbounded derivative containers. *Journal of Mathematical Analysis and Applications*, 81 (1981) 545-560.

[35] J. Zhao. The Smoothing Function of the Nonsmooth Matrix Valued Function. Master thesis, Department of Mathematics, National University of Singapore, (July 2004).

Name: Zheng Zheng

Degree: Master of Science

Department: Mathematics

Thesis Title: A Smoothing Newton Method for the Nonsmooth Boundary-valued ODEs

Abstract

In this thesis, we focus on the Nonsmooth Boundary-valued ODEs, whose right-hand functions are parameterized by algebraic variables that solve the initial-valued problems and the nonsmooth equations. Based on the idea of the initial value techniques used in traditional ODEs, a smoothing Newton method originated from the QSZ method is applied to the nonmsmooth dynamic systems. Most significantly, we address a reformulation of the nonsmooth ODEs to make them applicable to the smoothing algorithm and facilitate the convergence analysis. Especially, the differential variational inequalities are served as the special case for discussion.

Keywords:

Nonsmooth ODEs, optimal control, semismoothness, differential variational inequalities, shooting method, superlinear convergence.

A SMOOTHING NEWTON METHOD FOR THE BOUNDARY-VALUED ODEs

ZHENG ZHENG

NATIONAL UNIVERSITY OF SINGAPORE

2005

A Smoothing Newton Method for the Boundary-valued ODEs **Zheng Zheng** **2005**