Technichal Report: Optimal Control of Piecewise-Smooth Control Systems via Singular Perturbations

Tyler Westenbroek [*]   Xiaobin Xiong [†]   Aaron D. Ames [‡]   S. Shankar Sastry [§]

April 2, 2019

Abstract

This paper investigates optimal control problems formulated over a class of piecewise-smooth vector fields. Instead of optimizing over the discontinuous system directly, we instead formulate optimal control problems over a family of regularizations which are obtained by "smoothing out" the discontinuity in the original system. It is shown that the smooth problems can be used to obtain accurate derivative information about the non-smooth problem, under standard regularity conditions. We then indicate how the regularizations can be used to consistently approximate the non-smooth optimal control problem in the sense of Polak. The utility of these smoothing techniques is demonstrated in an in-depth example, where we utilize recently developed reduced-order modeling techniques from the dynamic walking community to generate motion plans across contact sequences for a 18-DOF model of a lower-body exoskeleton.

1 Introduction

Non-smooth dynamical systems naturally arise when modelling a vast array of engineering systems, and are familiar to researchers in areas ranging from the dynamic walking community [1], [2] to the domain of power systems analysis [3]. However, despite the pervasive nature of discontinuous dynamics in systems theory and control, optimizing system trajectories through unplanned sequences of discontinuities remains a distinct technical challenge.

In particular, non-smooth systems are well known to display non-differentiability with respect to initial conditions and inputs [4, Chapter 2, Section 11]. This fundamental departure from smooth control theory has meant that the powerful optimization-based techniques designed for smooth control systems [5] cannot be readily applied to the non-smooth setting. As an example, in the context of optimal control, this has traditionally led to the implementation of restrictive practices, such as forming large mixed-integer programs to reason over all possible "mode" sequences the trajectories may undergo [6], or limiting the class of system behaviors by fixed the number and sequence of discontinuities the trajectory of the system encounters *a priori* [7], [8].

1.1 Contributions

In this work, we formulate an optimal optimal control problem over a class of piecewise-smooth vector fields. Rather than reasoning through the discontinuities of the system directly, we instead study a family of regularizations which are obtained by "smoothing" the original non-smooth system along the discontinuity. We then formulate approximate optimal control problems over these smooth control systems, which can be solved using standard derivative-based techniques. Theoretical guarantees are discussed,

---

[*]Dept of EECS., University of California, Berkeley, westenbroekt@berkeley.edu. Corresponding Author.

[†]Dept. of Mechanical and Civil Engineering, California Institute of Technology, xxiong@caltech.edu

[‡]Dept. of Mechanical and Civil Engineering, California Institute of Technology, ames@caltech.edu

[§]Dept of EECS., University of California, Berkeley, sastry@berkeley.edu

demonstrating when the regularized problem can be used to obtain accurate derivative information about the non-smooth problem. We then discuss how this result can be used to derive consistent approximations in the sense of Polak [5].

Finally, we use the smoothing approach to generate trajectories for a model of a bipedal robot, specifically a powered exoskeleton, which intermittently makes contact with the ground. In particular, we build on [9], [10], which introduced a class of embedded reduced-order modeling techniques to simplify the complexity of the motion planning problem. We believe the unification of these techniques represents an important new direction for robotic trajectory optimization through contact.

1.2 Relation to the Literature

The roots of the smoothing technique considered in this paper can be traced back to the literature on *geometric singular perturbation theory* (see e.g. [11] or [12]). The precise smoothing approach considered in this paper were first considered for autonomous piecewiese-smooth systems in work pioneered by Teixeira [13], [14]. Emerging applications of these smoothing techniques in the control community can be found in [15], [16]. The authors have also recently been made aware of a contribution [17] which appears to have been made independently of the above literature and in which similar smoothing techniques were also considered in the context of optimal control. Our analysis builds on the results of [17] by requiring fewer technical assumptions, proving a concrete rate of convergence in Theorem 1, and formalizing the connection between the minimizers of the regularized and discontinuous problems in Section 4.2. Finally we mention a recent contribution [18], which builds on the geometric theory of hybrid systems [19],[20],[21], in which the smoothing techniques discussed in this paper were extended to a class of hybrid dynamical systems where the trajectories of the system undergo autonomous "jumps". We hope to extend the results presented here to this more general setting in a forthcoming article.

1.3 Notation

We now fix notation used throughout the document. Unless otherwise noted, the 2-norm is our choice of norm for finite dimensional spaces, and the L^2 norm is our norm of choice for function spaces. We say that a function $f \colon \mathbb{R}^n \to \mathbb{R}^p$ is *Lipschitz continuous* if there exists $L > 0$ such that for each $x_1, x_2 \in R^n$ we have $||f(x_1) - f(x_2)|| \leq L||x_1 - x_2||$. Given a smooth function $h \colon \mathbb{R}^n \to \mathbb{R}$ and a controlled vector field $f \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, the *Lie Derivative* of h along f is denoted $L_f h(x, u) \colon = \frac{\mathrm{d}}{\mathrm{d}x} h(x) \cdot f(x, u)$. Given a normed space V, $\mathcal{P}(V)$ denotes the set of all subsets of V. Given a subset $S \subset V$, $\overline{co}S$ denotes the convex closure of S in V.

For compactness of notation and exposition, throughout the paper all of the control systems we consider will have states belonging to $\mathbb{R}^n$ and admissible inputs taking values in $\mathbb{R}^m$, where n and m are positive integers. Moreover, we will assume that the admissible set of initial conditions for these control systems belong to a set $D \subset \mathbb{R}^n$, and that the control supplied to these systems takes on values in the set $U \subset \mathbb{R}^m$. Both D and U are assumed to be compact, connected and convex. Throughout the paper we will consider square-integrable control signals of length $T > 0$, which will be denoted $L^2([0, T], U)$. We will compactly denote the admissible data for these systems by

$$\mathcal{X} = D \times L^2([0, T], U).$$

We define the space of tangent vectors on $\mathcal{X}$ by

$$\mathcal{X}' = \mathbb{R}^n \times L^2([0, T], R^m).$$

Finally, we endow both $\mathcal{X}$ and $\mathcal{X}'$ with the norm $|| \cdot || \colon \mathbb{R}^n \times L^2([0, T], R^m) \to \mathbb{R}$ defined for each $\xi = (x_0, u) \in \mathbb{R}^n \times L^2([0, T], R^m)$ by

$$\|\xi\| = ||x_0||_2 + ||u||_2.$$

Given $\delta > 0$ and $\xi \in \mathcal{X}$, $B^\delta(\xi)$ denotes the ball of radius δ centered at ξ.

2 Piecewise Smooth Control Systems

We now introduce the class of bimodal piecewise-smooth control systems considered in this paper, briefly review Filippov's convention for defining the dynamics of this class of systems, and introduce the smooth approximations we study throughout the paper.

2.1 Piecewise-smooth Control Systems

To begin, let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a regular, smooth map with Lipschitz continuous first and second partial derivatives. Consider the two disjoint domains

$$D_1 = \{x \in \mathbb{R}^n \colon g(x) < 0\}$$

and

$$D_2 = \{x \in \mathbb{R}^n \colon g(x) > 0\},$$

which are separated by the co-dimension-1 sub-manifold

$$\Sigma = \{x \in \mathbb{R}^n \colon g(x) = 0\}.$$

Let $f_1, f_2 \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be smooth vector fields with Lipschitz continuous first and second partial derivatives, and then consider the piecewise-smooth control system

$$\dot{x} = f(x, u) = \begin{cases} f_1(x, u) & \text{if } x \in D_1 \\ f_2(x, u) & \text{if } x \in D_2, \end{cases} \tag{1}$$

where $f \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. Note that f is undefined along Σ, and in general may also be discontinuous along this surface. Due to the discontinuity in f, classical (or Caratheodory) solutions for the differential equation (1) may fail to exists along the surface of discontinuity. Thus, we turn to Filippov's convention to define the dynamics of the system.

2.2 Filippov Solutions

We now review Filippovs convention, and refer the reader to [4] for a more thorough introduction. Throughout the document we will make the following regularity assumption, which ensures the existence and uniqueness of Filippov solutions for the discontinuous system (1) [4, Chapeter 2, Section 2, Theorem 2].

Assumption 1. *For each $(x, u) \in \Sigma \times \mathbb{R}^m$ either $L_{f_1} g(x, u) > 0$ or $L_{f_2} g(x, u) < 0$.*

The assumption rules out pathological cases where $L_{f_1} g(x, u) < 0$ and $L_{f_2} g(x, u) > 0$ along Σ, in which case admissible solutions may either enter D_1 or D_2.

Filippov's convention compactly defines the dynamics of the discontinuous system by

$$\dot{x} \in F(x, u) = \begin{cases} f_1(x, u) & \text{if } x \in D_1 \\ \overline{co}\{f_1(x, u), f(x, u)\} & \text{if } x \in \Sigma \\ f_2(x, u) & \text{if } x \in D_2, \end{cases} \tag{2}$$

where $F \colon \mathbb{R}^m \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is a multivalued map [1]. For each set of data $\xi = (x_0, u) \in \mathcal{X}$, we say that the absolutely continuous function $x^\xi \colon [0, T] \to \mathbb{R}^n$ is the *Filippov solution* corresponding to this data if $x^\xi(0) = x_0$ and it satisfies the differential inclusion (2) at almost everywhere. For each $t \in [0, T]$ we then define the map $\phi_t \colon \mathcal{X} \to \mathbb{R}^n$ by

$$\phi_t(\xi) = x^\xi(t). \tag{3}$$

With Assumption 1 in effect, it is customary to distinguish the following two regions of $\Sigma \times U$:

---

[1]For a more formal derivation of this differential inclusion the reader is referred to [4, Chapter 1, Section 2]

1. The *crossing region*:

$$\{(x, u) \in \Sigma \times U : (L_{f_1}g(x,u)) \cdot (L_{f_2}g(x,u)) > 0\}$$

2. The *sliding region*:

$$\{(x, u) \in \Sigma \times U : L_{f_1}g(x,u) > 0 \text{ and } L_{f_2}g(x,u) < 0\}$$

When a Filippov solution is in the crossing region, it leaves the surface of discontinuity and enters either D_1 or D_2. However, when the solution is in the sliding region the solutions *slides* along the surface of discontinuity and obeys the differential equation

$$\dot{x} = f^s(x, u), \tag{4}$$

where Filippov's famous *sliding vector field* is defined by

$$f^s(x, u) = (1 - \alpha(x, u))f_1(x, u) + \alpha(x, u)f_2(x, u)$$

where

$$\alpha(x, u) = \frac{\nabla g(x) \cdot f_1(x, u)}{\nabla g(x) \cdot (f_1(x, u) - f_2(x, u))}$$

selects the unique convex combination of f_1 and f_2 which keeps the trajectory on Σ.

2.3   Smooth Approximations

Our family of smooth approximations are parameterized by $\varepsilon > 0$ and are obtained by smoothing (1) along the *region of regulatization*

$$\Sigma^\varepsilon = \{x \in \mathbb{R}^n : -\varepsilon < g(x) < \varepsilon\}.$$

The main idea behind the smoothing approach is to use the following class of functions to transition between f_1 and f_2 along Σ^ε.

Definition 1. *We say that $\varphi \in C^\infty(\mathbb{R}, [0, 1])$ is a transition function if 1) $\varphi(a) = 0$ if $a \leq -1$, 2) $\varphi(a) = 1$ if $a \geq 1$, 3) φ is strictly increasing on $(-1, 1)$, and 4) each of the first and second partial derivatives of φ are Lipschitz continuous.*

For the rest of the paper, we assume that as single transition function φ has been selected. For each $\varepsilon > 0$ we then define the ε-relaxation of (1) to be

$$\dot{x} = f^\varepsilon(x, u) \tag{5}$$

where $f^\varepsilon \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is given by

$$f^\varepsilon(x, u) = \left(1 - \varphi\left(\frac{g(x)}{\varepsilon}\right)\right)f_1(x, u) + \varphi\left(\frac{g(x)}{\varepsilon}\right)f_2(x, u).$$

Note that $f^\varepsilon(x, u)$ is a convex combination of $f_1(x, u)$ and $f_2(x, u)$ if $x \in \Sigma^\varepsilon$, and that $f^\varepsilon(x, u) = f_1(x, u)$ if $x \in D_1 \setminus \Sigma^\varepsilon$ and $f^\varepsilon(x, u) = f_2(x, u)$ if $x \in D_2 \setminus \Sigma^\varepsilon$.

It is straightforward to see that for each $\varepsilon > 0$ the vector field f^ε is smooth and Lipschitz continuous, and thus the trajectories corresponding to the smooth system will be unique. Thus for each $\varepsilon > 0$ and $\xi = (x_0, u) \in \mathcal{X}$ let $x^{(\varepsilon, \xi)} \colon [0, T] \to \mathbb{R}^n$ be the solution to (5) with initial condition $x^{(\xi, \varepsilon)}(0) = x_0$ and for each $t \in [0, T]$ define $\phi_t^\varepsilon \colon \mathcal{X} \to \mathbb{R}^n$ by

$$\phi_t^\varepsilon(\xi) = x^{(\varepsilon, \xi)}(t).$$

The following result from [18] establishes that the trajectories of the relaxed system converge to the Filippov solutions of the discontinuous system as $\varepsilon \to 0$ when Assumption 1 holds.

Lemma 1. *([18]) Let Assumption 1 hold for* (1). *Then there exists $C > 0$ and $\varepsilon_0 > 0$ such that for each $\varepsilon \le \varepsilon_0$, $t \in [0, T]$ and $\xi = (x_0, u) \in \mathcal{X}$*

$$\|\phi_t(\xi) - \phi_t^\varepsilon(\xi)\| \le C\varepsilon.$$

In other words, the solutions to the regularized systems converge to the solutions of the non-smooth system at a rate that is linear in ε.

3 Optimal Control Problems

In this section we formulate an optimal control problem over the piecewise-smooth system introduced in Section 2, and a family of approximate problems formulated over the smooth regularizations introduced in Section 2.3. We then discuss the challenges associated with applying standard gradient-based algorithms to arrive at minimizers for these problems.

3.1 Optimal Control Problems

To begin, let $\ell \colon \mathbb{R}^n \to \mathbb{R}$ be a cost function which is assumed to have Lipschitz continuous first and second partial derivatives, and define the cost functional $L \colon \mathcal{X} \to \mathbb{R}$ by

$$L(\xi) = \ell(\phi_T(\xi)).$$

We then define the following optimal control problem:

$$(\mathbf{P}) \quad \inf_{\xi \in \mathcal{X}} L(\xi)$$

Note that this problem formulation is quite general, since there exists well-known transformations to include additional terms, such as running costs, into the terminal cost functional we consider here [5, Chapter 5].

Next, for each $\varepsilon > 0$ we define the regularized cost functional $L^\varepsilon \colon \mathcal{X} \to \mathbb{R}$ by

$$L^\varepsilon(\xi) = \ell(\phi_T^\varepsilon(\xi)) \tag{6}$$

and subsequently define the $\varepsilon-$relaxation of P by:

$$(\mathbf{P}^\varepsilon) \quad \inf_{\xi \in \mathcal{X}} L^\varepsilon(\xi)$$

We next discuss derivative based approaches for solving these optimization problems.

3.2 Directional Derivatives

The majority of practical tools for finding local minimizers of optimal control problems rely on calculating the derivatives of the cost functional with respect to changes in the initial condition an inputs applied to the system. Once these derivatives are obtained, standard optimization procedures can be applied to iteratively improve performance [5].

3.2.1 Directional Derivatives of Regularized Problems

Since for each $\varepsilon > 0$ the vector field f^ε is smooth, we can compute directional derivatives of L^ε using standard approaches. In particular, given $\xi = (x_0, u) \in \mathcal{X}$, we define $DL^\varepsilon(\xi; \cdot) \colon \mathcal{X}' \to \mathbb{R}$ by

$$DL^\varepsilon(\xi; \delta\xi) = \lim_{\lambda \downarrow 0} \frac{L^\varepsilon(\xi + \lambda \cdot \delta\xi) - L^\varepsilon(\xi)}{\lambda} \tag{7}$$

for each direction $\delta\xi = (\delta x_0, \delta u) \in \mathcal{X}'$, which by [5, Theorem 5.6.8] is given by

$$DL^\varepsilon(\xi; \delta\xi) = p(0)^T \cdot \delta x_0 + \int_0^T p(\tau)^T B(\tau)\delta u(\tau)d\tau$$

where $p\colon [0,T] \to \mathbb{R}^n$ is the solution to the *co-state* or *adjoint* equations

$$-\dot{p}(t) = A(t)^T p(t) \text{ a.e. } t \in [0,T], \ p(T)^T = \frac{\mathrm{d}}{\mathrm{d}x}l(\phi_T(\xi))$$

and for each $t \in [0,T]$ we have $A(t) = \frac{\partial}{\partial x}f^\varepsilon(\phi_t(\xi), u(t))$ and $B(t) = \frac{\partial}{\partial u}f^\varepsilon(x, u)$. In particular $DL^\varepsilon(\xi; \delta\xi)$ provides a first-order approximation to how L^ε changes by perturbing the data supplied to the system in the direction $\delta\xi$.

However, inspecting f^ε, we see that the partial derivative $\frac{\partial}{\partial x}f^\varepsilon(x, u)$ will be of the order $\frac{1}{\varepsilon}$ when $x \in \Sigma^\varepsilon$. Thus, we should be concerned that the adjoint equations associated to the regularized system will "blow up" as we take $\varepsilon \to 0$ along trajectories which pass through the region of regularization. In other words, in the numerical setting we should be concerned that discretizations of P^ε will become ill-conditioned for small values of the regularization parameter. However, our subsequent analysis will demonstrate that this is not the case, and that the gradients of P^ε remain bounded even as we take $\varepsilon \to 0$.

3.2.2 Directional Derivatives of Non-smooth Problem

It is well-known that Filippov solutions for the discontinuous system (1) may have a non-smooth dependence on the initial conditions and inputs supplied to the system [4, Chapter 2, Section 11]. In particular, the map $\phi_t(\cdot)$ may fail to be differentiable at the point $\xi \in \mathcal{X}$ if the trajectory associated to this data arrives at the surface of discontinuity at time t. In the context of optimal control, this means that the adjoint equations associated to the discontinuous system may under go discrete "jumps" at time instances when the nominal trajectory reaches the surface of discontinuity[2]. Due to these discontinuities, the adjoint equations, and consequently the gradients of L, are difficult to approximate numerically using standard integration techniques such as Euler integration [17]. As discussed in Section 1.2, this has typically meant that derivative-based methods for solving optimal control problems over discontinuous systems have required restrictive practices such as fixing the number of times the trajectory crosses the surface of discontinuity *a priori*. This is our primary motivation for employing the smooth approximations considered in the paper.

Before proving our convergence results in Section 4, we first formalize conditions which will ensure the differentiability of L at a nominal choice of $\xi = (x_0, u) \in \mathcal{X}$. Each of these additional assumptions are standard when discussing the differentiability of Filippov solutions [4, Chapter 2, Section ll]. For compactness of notation, given $\xi = (x_0, u) \in \mathcal{X}$, we define $\mathcal{T}^\xi$ to be the set of time instances at which the Filippov solution corresponding to this data arrives at the surface of discontinuity:

$$\mathcal{T}^\xi = \{\hat{t} \in (0, T]\colon \phi_{\hat{t}}(\xi) \in \Sigma \text{ and } \phi_{\hat{t}-}(\xi) \notin \Sigma\}$$

First, we assume that the nominal trajectory does not begin on the surface of discontinuity or arrive at it at exactly time T. The first condition can lead to non-differentiability with respect to initial conditions, while the later condition can cause $\phi_T(\cdot)$ to be non-differentiable, and consequently also the terminal cost functional L [4, Chapter 2, Section ll].

Assumption 2. *The data $\xi = (x_0, u) \in \mathcal{X}$ is such that $x_0 \notin \Sigma$ and $T \notin \mathcal{T}^\xi$.*

---

[2]The "jump conditions" associated with the adjoint equations of the discontinuous system have a rich geometric interpretation which we do not have space to discuss in this article. However, the interested reader is referred to [4, Chapter 2, Section 11]) for a comprehensive discussion regarding these "jump conditions" in the context of sensitivity analysis.

Our next assumption ensures that the nominal trajectory arrives at the surface of discontinuity transversely, without "skimming" the surface, a situation which is well known to cause non-differentiability.

Assumption 3. *The data $\xi = (x_0, u) \in \mathcal{X}$ is such that if $\hat{t} \in \mathcal{T}^\xi$ then $\exists \gamma > 0$ and interval $I = (\hat{t} - \gamma, \hat{t} + \gamma) \cap [0, T]$ such that*

1. *$\phi_{\hat{t}^-}(\xi) \in D_1 \Rightarrow L_{f_1} g(\phi_t(\xi), u(t)) > 0, \forall t \in I$*

2. *$\phi_{\hat{t}^-}(\xi) \in D_2 \Rightarrow L_{f_2} g(\phi_t(\xi), u(t)) < 0, \forall t \in I$*

The final assumption we make rules our pathological cases such as "Zeno" behavior, where the trajectories of the system cross the surface of discontinuity an infinite number of times.

Assumption 4. *The data $\xi = (x_0, u) \in \mathcal{X}$ is such that $\mathcal{T}^\xi$ is a finite set.*

When the data $\xi \in \mathcal{X}$ satisfies Assumptions 2-4, the map $\phi_T(\cdot)$ will be differentiable at ξ, and thus the directional derivatives of L will be well-defined at this point. When this is the case we define $DL(\xi; \cdot): \mathcal{X}' \to \mathbb{R}$ by

$$DL(\xi; \delta\xi) = \lim_{\lambda \downarrow 0} \frac{L(\xi + \lambda \cdot \delta\xi) - L(\xi)}{\lambda}, \tag{8}$$

for each each direction $\delta\xi \in \mathcal{X}$.

4   Convergence Results

We are now ready to present the theoretical contribution of this work. In section 4.1 we demonstrate conditions under which the directional derivatives of the regularized problem converge to directional derivatives on the discontinuous problem, and in Section 4.2 we use this result to discuss how the regularized problems can be used to consistently approximate the non-smooth problem.

4.1   Convergence of Directional Derivatives

We first demonstrate that when Assumptions 2-4 are satisfied, and directional derivatives on the non-smooth cost functional are well-defined, that that the regularized problems can be used to approximate these derivatives with arbitrary precision. We first develop a few intermediate lemmas which are used in the proof of the main result. Proofs for the lemmas can be found in the appendix. The first lemma demonstrates that variations on the relaxed problems remain bounded, even as $\varepsilon \to 0$.

Lemma 2. *Let Assumption 1 hold. Further, assume that the data $\xi = (x_0, u) \in \mathcal{X}$ satisfies Assumptions 2-4 hold. Then there exists $C > 0$ and $\varepsilon_0 > 0$ such that for each $\varepsilon \leq \varepsilon_0$ and $\delta\xi = (\delta x_0, \delta u) \in \mathcal{X}$*

$$\|DL^\varepsilon(\xi; \delta\xi) - DL^\varepsilon(\xi; \delta\xi)\| \leq C\|\delta\xi\|. \tag{9}$$

Next, we establish our final two preliminary results needed for the statement of our firs convergence Theorem. The next result demonstrates conditions under which directional derivatives on the regularized problems converge to a well defined limit as $\varepsilon \to 0$, and will provide the rate of convergence in Theorem 1. We remind the reader the the proof of the following two results are contained in the appendix.

Lemma 3. *Let Assumption 1 hold. Let the data $\xi \in \mathcal{X}$ satisfy assumptions 2-3. Then there exists $C > 0$ and $\varepsilon_0 > 0$ such that for each $\varepsilon_1, \varepsilon_2 \leq \varepsilon_0$ and $\delta\xi \in \mathcal{X}$ we have*

$$\|DL^{\varepsilon_1}(\xi; \delta\xi) - DL^{\varepsilon_2}(\xi; \delta\xi)\| \leq C|\varepsilon_1 - \varepsilon_2|. \tag{73}$$

The final preliminary result demonstrates that if Assumptions 2-4 hold at the point $\xi \in \mathcal{X}$, then they also hold in a neighborhood of ξ.

Lemma 4. *Let Assumption 1 hold. Let Assumptions 2-4 hold for the data $\xi \in \mathcal{X}$. Then there exists $\delta > 0$ such that assumptions 2-3 also hold for each $\xi' \in B^\delta(\xi)$.*

With these preliminary results established, we are ready to state our first convergence result.

Theorem 1. *Let Assumption 1 hold. Let $\xi = (x_0, u) \in \mathcal{X}$ satisfy Assumptions 2-4. Then there exists $C > 0$ and $\varepsilon_0 > 0$ such that for each $\delta\xi = (\delta x_0, \delta u) \in \mathcal{X}'$ and $\varepsilon \leq \varepsilon_0$*

$$\|DL(\xi; \delta\xi) - DL^\varepsilon(\xi; \delta\xi)\| \leq C\varepsilon.$$

Proof. Let $\{\varepsilon_i\}_{i \in \mathbb{N}}$ be a sequence such that $\varepsilon_i \to 0$ as $i \to \infty$. By Lemma 1 and the continuity of ℓ, we see that the sequence of cost functionals $\{L^{\varepsilon_i}\}_{i \in \mathbb{N}}$ converges point-wise to L. Using a standard uniform convergence argument (see e.g. [22, Theorem 8.6.3]), we know that if we can show that there exists $\delta > 0$ such that for each $\xi' \in B^\delta(\xi)$ the sequence of linear maps $\{DL^{\varepsilon_i}(\xi'; \cdot)\}_{i \in \mathbb{N}}$ is bounded and converges uniformly, then its limit point must be $DL(\xi'; \cdot)$. In other words, this would demonstrate that directional derivatives of $\{L^{\varepsilon_i}\}_{i \in \mathbb{N}}$ converge to the directional derivatives of L in a δ-neighborhood of ξ. Let $\delta\xi \in \mathcal{X}'$ be arbitrary. By Lemmas 2, 3 and 4, we know that there exists constants $C_1, C_2 > 0$, $\varepsilon_0 > 0$, and $\delta > 0$ such that for each $\varepsilon \leq \varepsilon_0$ and $\xi' \in B^\delta(\xi)$

$$DL^\varepsilon(\xi'; \delta\xi) \leq C_1\|\delta\xi\| \tag{10}$$

and $\bar{\varepsilon}_1, \bar{\varepsilon}_2 \leq \varepsilon_0$ we have

$$\left\| DL^{\bar{\varepsilon}_1}(\xi'; \delta\xi) - DL^{\bar{\varepsilon}_2}(\xi'; \delta\xi) \right\| \leq C_2|\bar{\varepsilon}_1 - \bar{\varepsilon}_2|, \tag{11}$$

which demonstrates that the sequence of linear maps $\{DL^\varepsilon(\xi'; \cdot)\}_{i \in N}$ converges to a bounded limit as $i \to \infty$, as desired. The linear rate of convergence follows from (11). $\qquad\square$

At this point it is prudent to compare Theorem 1 to the convergence results presented in [17]. First we note that our regularity Assumption 1 is weaker than the one-sided Lipschitz assumption made in [17, Section 3.1]. Next, we note that we provide a concrete rate of convergence in Theorem 1. Note that Theorem 1 also demonstrates that the *equations of sensitivity* [4, Chapter 2, Section ll] for the regularized system converge, at a rate that is linear in ε, to the equations of sensitivity of the non-smooth system. Many applications, such as assessing the stability of periodic orbits [2], rely on accurately assessing the equations of sensitivity with known margins of error. We anticipate that the bound in Theorem 1 will also find use in such settings.

4.2 Consistent Approximations

In practice, solutions to difficult optimization problems are often approximated by solving a *sequence* of approximate problems. The most common example are penalty methods, wherein a sequence of increasingly penalized unconstrained optimization problems are used to find minimzers for a difficult constrained optimization problem. More generally, analyzing sequential algorithms of this sort falls into the framework of *consistent approximations* as stated by Polak [5, Chapter 3.3]. In this section, we discuss how this theory can be applied to find minimizers of P using so-called "Master Algorithms" and a sequence of regularized problems $\{P^{\varepsilon_i}\}_{i \in \mathbb{N}}$.

Even though, as we saw in the proof of Theorem 1, L^ε converges to L as $\varepsilon \to 0$, this is in general not sufficient to show that a sequence of minimizers to the sequence $\{P_i^\varepsilon\}_{i \in N}$ accumulate to a minimizer of P [5, Chapter 3.3]. In particular, the consistency results we pursue will rely on comparing *optimality functions* for the non-smooth and regularized problems, which provide additional derivative information about the problems.

Definition 2. *Let G be an optimization problem on the normed space S. We say that $\theta\colon (-\infty, 0]$ is an optimality function for G iff it is continuous and $\theta(\bar{x}) = 0$ for each minimizer $\bar{x} \in S$.*

Given a difficult optimization problem and optimality function pair (G, θ), the theory of consistent approximations relies on finding a sequence of approximate optimization problems and optimality functions $\{(G_n, \theta_n)\}_{n \in \mathbb{N}}$ and being able to show that the epigraphs of the pair (G_n, θ_n) converge to the epigraphs of (G, θ) as $N \to \infty$. Then one can deploy the aforementioned Master Algorithms [5, Chapter 3.3.3] to solve of G by iteratively optimizing a subsequence of $\{G_n\}_{n \in \mathbb{N}}$. Due to space constraints, we leave a more detailed introduction to [23] or [5], and instead discuss here how to obtain such consistency results for the problems we consider.

Due to the non-differentiability of L, we modify the approach typically used to demonstrate such consistency results for approximations to smooth optimal control problems [5, Chapter 4]. In particular, we will restrict our attention to regions where Assumptions 2-4 are satisfied and the regularized problems provide accurate derivative information about the non-smooth problem in the limit. We propose the following optimality functions for the smooth and non-smooth problems, which are restricted to subsets of $\mathcal{X}$ on which Assumptions 2-4 are satisfied.

Proposition 1. *Let $S \subset \mathcal{X}$ be a connected, convex set such that Assumptions 2-4 are satisfied for each $\xi \in \mathcal{S}$. Consider the restricted optimization problems:*

$$(\mathbf{P}_S) \quad \inf_{\xi \in S} L(\xi), \tag{12}$$

$$(\mathbf{P}_S^\varepsilon) \quad \inf_{\xi \in S} L^\varepsilon(\xi) \tag{13}$$

Then the following two functions are optimality functions for P_S and P_S^ε, respectively:

$$\theta_S(\xi) = \inf\{DL(\xi; \delta\xi)\colon \delta\xi \in \mathcal{X}' \text{ and } \xi + \delta\xi \in \mathcal{X}\} \tag{14}$$

$$\theta_S^\varepsilon(\xi) = \inf\{DL^\varepsilon(\xi; \delta\xi)\colon \delta\xi \in \mathcal{X}' \text{ and } \xi + \delta\xi \in \mathcal{X}\} \tag{15}$$

Proof. We omit a detailed proof of the fact that the proposed optimality functions appropriately capture the minimizers of the respective optimization problems, as the argument closely follows the arguments made [5, Theorem 5.6.9] and [24, Proposition 4.5]. The continuity of the relaxed optimality function follows by noting that time map $\xi \to DL^\varepsilon(\xi, \delta\xi)$ is continuous for each $\varepsilon > 0$ [5, Theorem 5.6.9], and recalling Danskin's Theorem. The continuity of the proposed optimality function for the non-smooth problem follows from the uniform convergence displayed in Theorem 1. $\qquad\square$

We can now state the following consistency result:

Theorem 2. *Let $\{\varepsilon_i\}_{i \in \mathbb{N}}$ be a sequence such that $\varepsilon_i \to 0$ as $i \to \infty$, and let S be as in Proposition 1. Then the pair $(L_S^{\varepsilon_i}, \theta_S^{\varepsilon_i})$ epi-converges to (P_S, θ_S) as $i \to \infty$.*

Proof. The pointwise convergence of $\{L_S^{\varepsilon_i}\}_{i \in \mathbb{N}}$ to L_S noted in the proof of Theorem 1 also demonstrates epi-convergence of the sequence. The epi-convergence of $\{\theta_S^{\varepsilon_i}\}_{i \in \mathbb{N}}$ to θ_S follows directly from the statement of Theorem 1. $\qquad\square$

This result demonstrates that the appropriate consistency results from [5, Chapters 3-4] hold, so long as we restrict our attention to subsets where Assumptions 2-4 hold. In practice however, we do not want to restrict ourselves ahead of time to a single compact subset S on which Assumptions 2-4 are satisfied. In particular, we actually want to optimize over all of $\mathcal{X}$ and iteratively optimize over a sequence $\{L^{\varepsilon_i}\}_{i \in \mathbb{N}}$, in which case we should expect that successive iterates will sometimes produce initial conditions which cross the surface of discontinuity between iterations, or generate trajectories whose final state crosses the surface between iterations. Nonetheless, so long as an algorithm which successively optimizes over $\{L_i^\varepsilon\}_{i \in \mathbb{N}}$ accumulates to a point which satisfies Assumptions 2-4, we can invoke Theorem

2 in a neighborhood of this point to study the limiting behavior of the algorithm. Though we plan to provide a more exhaustive exploration of this topic and introduce implementable numerical algorithms for solving **P** in a forthcoming article, we believe the results we have indicated here provide a valuable basis for this important future step.

5 Trajectory Optimization via Smoothing on a Reduced-order Model

To demonstrate the utility of the smoothing technique we have investigated, we first use the approach to generate hopping trajectories for an actuated Spring-mass hopper, which was first proposed in [9] for realizing the hopping on the bipedal robot Cassie. Its nonlinear leg spring model has also shown great value for other robotic applications such as underactuated bipedal walking [1] and fully-actuated humanoid walking [10]. We note, however, that the trajectory generation procedures proposed in these prior works required fixing the sequence of contacts the robot makes with the ground *a priori*. Here, we demonstrate that the regularization techniques we consider can be used to effectively generate optimal trajectories on the Spring-mass model without pre-specifying the contact sequence ahead of time.

5.1 Actuated Spring-mass Model of Hopping

The actuated Spring-mass hopper is depicted in Fig. 1 (a). The model has four states: the height of the mass, z, and its time derivative $\dot{z}$, as well as the natural length of the leg spring, L, and its time derivative, $\dot{L}$. The model is actuated by controlling L, i.e.,

$$\ddot{L} = u, \tag{16}$$

where u is assumed to be the actuation on the leg. The *Flight* phase is ballistic, thus the dynamics of the mass is,

$$\ddot{z} = -g \tag{17}$$

where g is the gravitational constant. When the hopper is on the ground, the dynamics of the mass is,

$$\ddot{z} = \frac{F(z, \dot{z}, L, \dot{L})}{m} - g \tag{18}$$

where m is the value of the mass and $F(z, \dot{z}, L, \dot{L}) = K(L)(L - z) + D(L)(\dot{L} - z)$ is the leg force from the spring deflection. Here, $K(L)$ and $D(L)$ are the stiffness and damping coefficient, respectively. For further details of the model can be found in [9]. The hopper lifts off the ground with $F \to 0$. In practice, the $K(L)$ is much larger than $D(L)$. Thus we simplify the assumption that the robot lifts off when $L = z$. Collecting the states of the robot as $x = (z, \dot{z}, L, \dot{L})$, we represent the dynamics using the piecewise-smooth vector field $f \colon \mathbb{R}^4 \times \mathbb{R} \to \mathbb{R}^4$ where,

$$\dot{x} = f(x, u) = \begin{cases} f_{\mathrm{F}}(x, u) & \text{if } g(z) > 0 \\ f_{\mathrm{G}}(x, u) & \text{if } g(z) < 0, \end{cases} \tag{19}$$

where,

$$f_{\mathrm{F}}(x, u) = (\dot{z}, -g, \dot{L}, u)^T \tag{20}$$

$$f_{\mathrm{G}}(x, u) = (\dot{z}, F(x, \dot{z}, L, \dot{L})/m - g, \dot{L}, u)^T \tag{21}$$
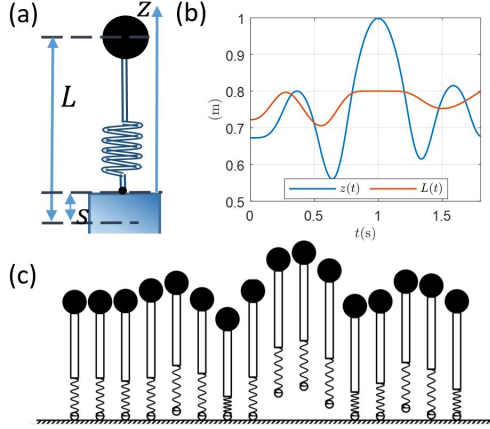
$$g(x) = z - L. \tag{22}$$

Figure 1: (a) The actuated Spring-mass model. (b) The optimized hopping trajectories. (c) Visualization of the hopping.

5.2 Trajectory Optimization Via Smoothing

Now we apply the proposed smoothing method for optimizing a hopping motion without specifying the contact modes. The hopping task we choose is to reach an apex height $z_{\text{apex}} = 1\text{m}$ at $t_{\text{apex}} = 1\text{s}$ from a static standing configuration, and then settle back to its original height at $t_{\text{f}} = 1.8\text{s}$. The initial condition is set as $x_0 = (z(0), \dot{z}(0), L(0), \dot{L}(0))^T = (.65, 0, .75, 0)^T$. The input is bounded, i.e., $u \in U = [-10, 10]$. The cost function is,

$$J(x) = (z(t_{\text{apex}}) - z_{\text{apex}})^2 + \dot{z}(t_{\text{apex}})^2 +$$
$$(z(t_{\text{f}}) - z(0))^2 + \dot{z}(t_{\text{f}})^2 + \int u^2 dt.$$

To approximate a solution to the problem numerically, we use the regularization parameter $\varepsilon = .01$, and Euler integration with 200 uniformly spaces gridpoints. The resulting finite dimensional optimization problem is solved using the Matlab fmincon function. The optimized Spring-mass trajectory is shown in Fig. 1 (b). Due to the actuation limitations of the model, the optimized hopping actually requires two hops to reach to the apex height and one additional hop to settle.

6 Hopping Embedding on the Exoskeleton

Despite the simplicity of the Spring-mass hopper, the optimized dynamics can be embedded onto complex robotic systems. This further amplifies the application of proposed smoothing technique. Here we briefly describe the process of embedding the hopping dynamics of the Spring-mass onto the exoskeleton (Fig. 2 (a)). More examples of the dynamics embedding from simple models to full robot models can be found in [9] [1] [10] and the references therein.

6.1 Robot Model and Hybrid Dynamical Model

The exoskeleton (Exo) has two legs with 12 motor joints in total. Its equation of motion can be described by the floating-base Euler-Lagrange equation,

$$M\ddot{q} + h = Bu + J_v^T F_v, \tag{23}$$
$$J_v\ddot{q} + \dot{J}_v\dot{q} = 0, \tag{24}$$

where $q \in SE(3) \times \mathbb{R}^{n=12}$. The exact meaning of each term can be found in [9]. For the hopping behavior, we assume the two feet always have same contact, i.e. either both feet are or are not in contact

with the ground. Thus the hopping is an alternation between two domains, i.e. *Ground* and *Flight* domain. We use subscript v to denote different domains. The guards are defined from the transition conditions, i.e. ground normal forces$\to 0$ and foot position$\to 0$. We model the impact as plastic impact from *Flight*$\to$*Ground*, where the joint velocities have jumps, i.e., $\dot{q}^+_{\text{Ground}} = \Delta(q)\dot{q}^-_{\text{Flight}}$.

6.2 Output Selection and CLF-QP based Force Control.

For *Ground*, the robot is fully actuated and 6 outputs are required to be defined. We first select the center of mass (COM) position and the yaw and roll angles of the pelvis/torso of the robot as relative degree 2 outputs, i.e.,

$$\mathcal{Y}_2^{\text{Ground}}(q,t) = \begin{bmatrix} p_{\text{COM}^x}(q) \\ p_{\text{COM}^y}(q) \\ p_{\text{COM}^z}(q) \\ \varphi_{\text{pelvis}}(q) \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ z^{\text{opt}}(t) \\ \mathbf{0} \end{bmatrix}, \tag{25}$$

where $z^{\text{opt}}(t)$ is the mass trajectory of the Spring-mass model for the embedding. It is also necessary to have zero centrodial angular momentum before *Flight* phase [9]. We select the last output as,

$$\mathcal{Y}_1^{\text{Ground}}(q,t) = H_{\text{pitch}}(q,\dot{q}) - 0, \tag{26}$$

where $H_{\text{pitch}}(q,\dot{q})$ is the pitch centrodial momentum.

In *Flight*, the robot is underactuated. Here, 12 outputs are required to be defined since $n_u = 12$. We selected the outputs as the positions of the feet to the COM and foot orientations. The desired vertical positions between the COM and the feet are the real leg length $L - z$ of the Spring-mass hopper.

$$\mathcal{Y}_2^{\text{Flight}}(q,t) = \begin{bmatrix} \mathbf{p}_{\text{Feet}\to\text{COM}}^x(q) \\ \mathbf{p}_{\text{Feet}\to\text{COM}}^y(q) \\ \mathbf{p}_{\text{Feet}\to\text{COM}}^z(q) \\ \varphi_{\text{Feet}}(q) \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \mathbf{c} \\ L^{\text{opt}}(t) - z^{\text{opt}}(t) \\ \mathbf{0} \end{bmatrix}, \tag{27}$$

where $\mathbf{c}$ is a constant vector and its value depends on the initial state of the flight phase.

We apply the control Lyapunov function based Quadratic programs (CLF-QP) [25] [1] for feedback-zeroing the outputs. The exponential convergence on the output dynamics is enforced by the inequality condition on the constructed Lyapunov function, i.e.,

$$\dot{V}(u,q,\dot{q}) \leq -\gamma V(q,\dot{q}), \tag{28}$$

with $\gamma > 0$. This inequality is affine with respect to u,

$$A_v^{\text{CLF}}(q,\dot{q})u \leq b_v^{\text{CLF}}(q,\dot{q}). \tag{29}$$

For external contacts as holonomic constraints, the holonomic forces F_v are affine with respect to u,

$$F_v = A_v u + b_v, \tag{30}$$

where A_v, b_v can be found in [10]. Thus contact constraints such as friction cones, non-negative normal forces can be formulated as an inequality constraints on u,

$$C_v A_v u \leq -C_v b_v, \tag{31}$$

where C_v is a constant matrix. Since it is desirable to embed the normal force of the Spring-mass hopping on the full robot, an equality constraint can be expressed for the force control,

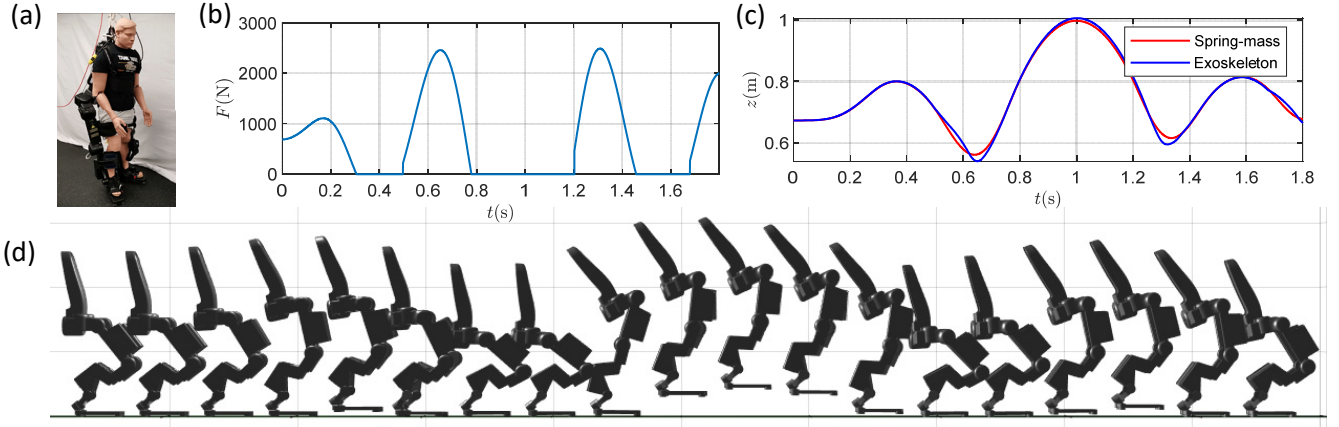$$S_v A_v u = F^{\text{opt}} - S_v b_v, \tag{32}$$

Figure 2: (a) The Exoskeleton. (b) The optimized ground force of the hopping. (c) Comparison of the COM trajectories between the Spring-mass and the Exo. (d) The snapshots of the realized hopping.

where S_v is a selection matrix to extract the vertical normal forces from the holonomic force vector and F^{opt} is the leg force from the Spring-mass optimization (Fig. 2 (b)). It is necessary to relax the force control since the vertical COM position is one of the outputs for control [10]. Thus the equality in Eq. (32) becomes,

$$\underbrace{(1-c)F^{\text{opt}} - S_v b_v}_{c_{lb}} \leq S_v A_v u \leq \underbrace{(1+c)F^{\text{opt}} - S_v b_v}_{c_{ub}}, \tag{33}$$

where $c \in (0,1)$ is a coefficient of the relaxation.

The main control law for the full-order model in each domain is thus formulated as a quadratic program as follows:

$$u^* = \underset{u \in \mathbb{R}^{12}, \delta \in \mathbb{R}}{\text{argmin}} \ u^T H u + 2F u + p\delta^2, \tag{34}$$

$$\text{s.t.} \quad A_v^{\text{CLF}}(q, \dot{q})u \leq b_v^{\text{CLF}}(q, \dot{q}) + \delta, \quad \text{(CLF)}$$
$$C_v A_v u \leq -C_v b_v, \quad \text{(Contact)}$$
$$u_{lb} \leq u \leq u_{ub}, \quad \text{(Torque Limit)}$$
$$c_{lb} \leq S_v A_v u \leq c_{ub}, \quad \text{(Force Control)}$$

where δ is a relaxation term for increasing the instantaneous feasibility of the QP, and p is a positive penalty constant.

6.3 Simulation Result

We primarily implemented the embedding in simulation to demonstrate the value of the smoothing technique. Details of the implementation can be found in [10], where same embedding was implemented for realizing walking on the Exo. Fig. 2 (c)(d) show the resulting hopping motion. The COM trajectory matches well in general. The exerted hopping corresponds to the Spring-mass hopper. A video of the simulation can also be found in [26].

7 Conclusion

This paper investigated a method for smoothing out a discontinuous differential equations. We studied optimal control problems over the discontinuous system and its regularizations, derived a number of useful properties for these problems, examined conditions under which the regularizations can be used

to obtain accurate derivative information about the non-smooth problem, and outlined how to invoked the theory of consistent approximations to study how minimizers of the regularization accumulate to minimizers of the non-smooth problem. Finally, we demonstrated the efficacy of the smoothing approach by using it in conjunction with recent reduced-order modelling techniques to generate non-trivial motion plans on a 18-DOF exoskeleton robot in simulation.

8   Acknowledgement

A   Technical Lemmas

This section contains a number of technical lemmas and proofs of lemmas that were stated earlier in the document.

Lemma 5. *Let Assumption 1 hold. Then there exists $C > 0$ and $\varepsilon_0 > 0$ such that for each $\xi \in \mathcal{X}$ and $t \in [0, T]$ we have*

$$\|\phi_t(\xi)\| \leq C \tag{35}$$

and for each $\varepsilon \leq \varepsilon_0$ we have

$$\|\phi_t^\varepsilon(\xi)\| \leq C \tag{36}$$

Proof. The proof follows closely the steps of [5, Proposition 5.6.5]. First, note that by the Lipshchitz continuity of f_1 and f_2 and the boundedness of U there exists a constant $K > 0$ such that for each $x \in \mathbb{R}^n$ and $u \in U$ we have

$$\max\{f_1(x, u), f_2(x, u)\} \leq K(1 + \|x\|). \tag{37}$$

Furthermore we also see that:

$$\|f^\varepsilon(x, u)\| = \left\|\left(1 - \varphi\left(\frac{g(x)}{\varepsilon}\right)\right)f_1(x, u) + \varphi\left(\frac{g(x)}{\varepsilon}\right)f_2(x, u)\right\| \leq \max\{\|f_1(x, u)\|, \|f_2(x, u)\|\} \leq K(1 + \|x\|) \tag{38}$$

For each $\varepsilon > 0$ and $\xi = (x_0, u) \in \mathcal{X}$ we also have

$$\|\phi_t^\varepsilon(\xi)\| \leq \|x_0\| + \int_0^t \|f^\varepsilon(\phi_t(\xi), u(t))\| \leq K(1 + \phi_t^\varepsilon(\xi)). \tag{39}$$

Using a standard Gronwall in equality [5, Lemma 5.6.4], we conclude that

$$\|\phi_t(\xi)\| \leq e^K(1 + \|x_0\|). \tag{40}$$

Recalling that D is bounded and invoking Lemma 1 demonstrates the desired results. $\square$

A.1   Proof of Lemma 2

The following proof is rather involved, so we begin by making a number of simplifying assumptions for which we incur no loss of generality. To begin, we will assume that the nominal trajectory of the discontinuous system reaches the surface of discontinuity exactly once; the generalization to the case where it reaches the surface a finite number of times is straightforward. That is, we assume there is a single time $t_1 \in \mathcal{T}^\xi$. Without loss, we assume that $x_0 \in D_1$. To ease our analysis, we will assume that the function g is such that for each $x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ we have $g(x) = x_1$. Note that by the Implicit Function Theorem there exists a set of coordinates about each point which makes this assumption true.

Furthermore, for claritrty of exposition, rather than prove the result using the adjoint equations, we will demonstrate it using the *equations of sensitivity*. In particular, by the chain rule, we may also calculate $DL^\varepsilon(\xi; \delta\xi)$ as

$$DL^\varepsilon(\xi; \delta\xi) = \frac{\mathrm{d}}{\mathrm{d}x}L(\phi_T^\varepsilon(\xi)) \cdot D\phi_T^\varepsilon(\xi; \delta\xi), \tag{41}$$

where for each $t \in [0, T]$ we define

$$D\phi_t^\varepsilon(\xi; \delta\xi) = \lim_{\lambda\downarrow 0} \frac{\phi_t^\varepsilon(\xi + \lambda\delta\xi) - \phi_t^\varepsilon(\xi)}{\lambda} \tag{42}$$

where for each $t \in [0, T]$ by [5, Theorem 5.6.8] we have

$$D\phi_t^\varepsilon(\xi; \delta\xi) = \Phi(t, 0)\delta x_0 + \int_0^t \Phi(t, \tau)\frac{\partial}{\partial u}f^\varepsilon(\phi_t(\xi), u(\tau))\delta u(\tau) \quad \delta x(0) = \delta x_0, \tag{43}$$

and, for each $\tau \in [0, T]$, $\Phi(t, \tau)$ is the solution to

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t, \tau) = \frac{\mathrm{d}}{\mathrm{d}x}f^\varepsilon(\phi_t^\varepsilon(\xi), u(t))\Phi(t, \tau), \quad \Phi(\tau, \tau) = I, \tag{44}$$

where in the above equation I denotes the identity matrix. Note that since $\frac{\mathrm{d}}{\mathrm{d}x}\phi_t(\xi)$

It is straightforward to check that, while $\frac{\partial}{\partial x}f^\varepsilon(x, u)$ is of order $\frac{1}{\varepsilon}$ in the region of regularization, $\frac{\partial}{\partial u}f^\varepsilon(x, u)$ is bounded on bounded sets. Thus, if we can show that the state transition matrix in (43) is bounded for each $t, \tau \in [0, T]$, the desired result will follow immediately.

In order to demonstrate this fact, it is sufficient to demonstrate that the solution $\delta x^\varepsilon \colon [0, T] \to \mathbb{R}^n$ to

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta x^\varepsilon(t) = \frac{\partial}{\partial x}f^\varepsilon(\phi_t^\varepsilon(\xi), u(t))\delta x^\varepsilon(t), \quad \delta x^\varepsilon(0) = \bar{x}_0 \tag{45}$$

remains bounded for each choice of initial condition $\bar{x}_0 \in \mathbb{R}^n$. Since we have assumed that the gradients of f_1 and f_2 are Lipschitz continuous, the solution to (45) will remain bounded when the nominal regularized trajectory is away from the surface of discontinutiy; thus, we need only to show that the equations of sensitivity remain bounded when the nominal trajectory passes through the region of regularization.

Throughout the proof, we will study the behavior of the sensitivity equations in three cases. We will first consider the case where he nominal trajectory of the discontinuous system crosses the surface of discontinuity transversally. In the second case, we will examine when the nominal trajectory of the discontinuous system begins to slide along the surface of discontinuity at time t_1. In the final case we consider, we will examine situations where the nominal trajectory of the discontinuous system either remains on the surface of discontinuity in an open interval containing t_1 without sliding, or instead instantaneously crosses the surface of discontinuity but does not do so transversally.

However, before addressing these various cases we first introduce notation and provide a preliminary analysis which will aid our discussion. We begin by defining for each $\varepsilon > 0$ and $t \in [0, T]$

$$\varphi^\varepsilon(t) = \varphi\left(\frac{g(\phi_t^\varepsilon(\xi))}{\varepsilon}\right) \tag{46}$$

$$\omega^\varepsilon(t) = \varphi'\left(\frac{g(\phi_t^\varepsilon(\xi))}{\varepsilon}\right) \tag{47}$$

$$A_1^\varepsilon(t) = \frac{\partial}{\partial x}f_1(\phi_t^\varepsilon(\xi), u(t)) \tag{48}$$

$$A_2^\varepsilon(t) = \frac{\partial}{\partial x}f_2(\phi_t^\varepsilon(\xi), u(t)) \tag{49}$$

$$\bar{f}^\varepsilon(t) = (f_2(\phi_t^\varepsilon(\xi), u(t)) - f_1(\phi_t^\varepsilon(\xi), u(t))) \tag{50}$$

We also define the following constant which will be used repeatedly in our analysis:

$$C_1 = \max_{a \in (-1,1)} \frac{\partial}{\partial a} \varphi(a) \tag{51}$$

where we note that C_1 exists and is finite due to our assumptions on the partial derivatives of φ. By our assumption that the nominal trajectory of the discontinuous system arrives at the surface of discontinuity transversally and Lemma 1 for each ε sufficiently small we may define

$$t^\varepsilon = \inf\{t \in [0,T] \colon \phi_t^\varepsilon(\xi) \in \Sigma^\varepsilon\} \tag{52}$$

to be the time instant at which the corresponding regularized trajectory first arrives at the region of regularization.

With the above notation we can compactly write the partial derivative

$$\frac{\partial}{\partial x} f^\varepsilon(\phi_t(\xi), u(t)) = (1 - \varphi^\varepsilon(t))A_1^\varepsilon(t) + \varphi^\varepsilon(t)A_2^\varepsilon(t) + \frac{1}{\varepsilon}\omega^\varepsilon(t)\bar{f}^\varepsilon(t)\nabla g(\phi_t^\varepsilon(\xi)) \tag{53}$$

Applying Lemma 5, and our assumption that the gradients of f_1 and f_2 are Lipzschitz, it is straightforward to bound for each $t \in [0,T]$

$$\frac{\partial}{\partial x} f^\varepsilon(\phi_t(\xi), u(t)) \leq M + M\omega^\varepsilon(t) \tag{54}$$

for some $M > 0$, by noting that $\varphi^\varepsilon(t) \in [0,1]$. Applying a standard Gronwall inequality [5, 5.6.4] we see that

$$\|\delta x(t)\| \leq \exp\left(\int_0^t M + M\frac{1}{\varepsilon}\omega^\varepsilon(\tau)\right)\|\bar{x}\| \leq C_2 \exp\left(\int_0^t \frac{1}{\varepsilon}M\omega^\varepsilon(\tau)d\tau\right)\|\bar{x}\| \tag{55}$$

for some $C_2 > 0$ sufficiently large.

This trivial bound allows us to simply bound the sensitivity of the regularized system in the case where the nominal trajectory of the discontinuous system crosses the surface of discontinuity transversally. Indeed, consider the case where there exists $\delta > 0$ such that for almost every $t \in (t_1 - \delta, t_1 + \delta)$ we have $L_{f_1}g(\phi_t(\xi), u(t)) > 0$ and $L_{f_2}g(\phi_t(\xi), u(t)) > 0$. Note that for ε sufficiently small we will also have $L_{f_1}(\phi_t^\varepsilon(\xi), u(t)) > 0$ and $L_{f_2}(\phi_t^\varepsilon(\xi), u(t)) > 0$ for each $t \in (t_1 - \delta, t_1 + \delta)$. Since the nominal trajectory of the discontinuous system passes transversally across the surface of discontinuity, it is straightforward to see that for each ε small enough the corresponding relaxed trajectory will spend at most on the order of ε units of time in the region of regularization. In other words, $\exists K > 0$ such that $\phi_t^\varepsilon(\xi) \notin \Sigma^\varepsilon$ for each $t \notin (t_1 - K\varepsilon, t_1 + k\varepsilon)$. Combining this fact with (55), the fact that $\omega^\varepsilon(t) = 0$ if $\phi_t^\varepsilon(\xi) \notin \Sigma^\varepsilon$ and the fact that $\omega^\varepsilon(t)$ is bounded, it is straightforward to verify that in this case $\delta x^\varepsilon(t)$ remains bounded for each $t \in [0,T]$. Indeed we see that

$$\|\delta x(t)\| \leq C_2 \exp\left(\int_0^t \frac{1}{\varepsilon}M\omega^\varepsilon(\tau)d\tau\right) \leq C_2 \exp\left(\int_{t_1-K\varepsilon}^{t_1+K\varepsilon} M\frac{1}{\varepsilon}C_1 d\tau\right) \leq C \tag{56}$$

for some $C > 0$ sufficiently large, as desired.

In the second two cases that we consider, we will not be able to upperbound the time spent in the region of regularization on the order of $\frac{1}{\varepsilon}$. Before proceeding to these more delicate conditions, we introduce some additional terminology which will clarify the intuition behind our approach to obtaining a bound in this case.

First, we will further decompose

$$\delta x^\varepsilon(t) = \begin{bmatrix} \delta x_1^\varepsilon(t) \\ \delta x_2^\varepsilon(t) \end{bmatrix}, \tag{57}$$

where $\delta x_1^\varepsilon \colon [0, T] \to \mathbb{R}$ is the first entry of δx^ε, and $\delta x_2^\varepsilon \colon [0, T] \to \mathbb{R}^{n-1}$ represents the remaining n-1 entries. Given this notation, and our simplifying assumption on g which implies $\nabla g(x) = [1, 0, \ldots, 0]$, we may rewrite (45) as

$$\frac{\mathrm{d}}{\mathrm{d}t} \delta x_1^\varepsilon(t) = \left(\bar{A}_{11}^\varepsilon(t) + \frac{1}{\varepsilon} \omega^\varepsilon(t) \bar{f}_1^\varepsilon(t) \right) \delta x_1^\varepsilon(t) + \bar{A}_{12}^\varepsilon(t) \delta x_2^\varepsilon(t) \tag{58}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t} \delta x_2^\varepsilon(t) = \left(\bar{A}_{21}^\varepsilon(t) + \frac{1}{\varepsilon} \omega^\varepsilon(t) \bar{f}_2^\varepsilon(t) \right) \delta x_1^\varepsilon(t) + \bar{A}_{22}^\varepsilon(t) \delta x_2^\varepsilon(t) \tag{59}$$

where $\bar{f}_1^\varepsilon \colon [0, T] \to \mathbb{R}$ denotes the first entry of $\bar{f}^\varepsilon$ and $\bar{f}_2^\varepsilon \colon [0, T] \to \mathbb{R}^{n-1}$ denotes the $n - 1$ remaining entries, and $\bar{A}_{11}^\varepsilon \colon [0, T] \to \mathbb{R}$, $\bar{A}_{12}^\varepsilon \colon [0, T] \to \mathbb{R} \times \mathbb{R}^{(n-1)}$, $\bar{A}_{21} \colon [0, T] \to \mathbb{R}^{(n-1)} \times \mathbb{R}$ and $\bar{A}_{21}^\varepsilon \colon [0, T] \to \mathbb{R}^{(n-1)} \times \mathbb{R}^{(n-1)}$ are such that

$$(1 - \varphi^\varepsilon(t)) A_1^\varepsilon(t) + \varphi^\varepsilon(t) A_2^\varepsilon(t) = \begin{bmatrix} \bar{A}_{11}^\varepsilon(t) & \bar{A}_{12}^\varepsilon(t) \\ \bar{A}_{21}^\varepsilon(t) & \bar{A}_{22}^\varepsilon(t) \end{bmatrix}. \tag{60}$$

Solving these linear time-varying systems, we have that

$$\delta x_1^\varepsilon(t) = \exp\left(\int_0^t \bar{A}_{11}^\varepsilon(\tau) + \frac{1}{\varepsilon} \omega^\varepsilon(\tau) \bar{f}_1^\varepsilon(\tau) d\tau \right) \bar{x}_0^1$$
$$+ \int_0^t \exp\left(\int_\tau^t \bar{A}_{11}^\varepsilon(s) + \frac{1}{\varepsilon} \omega^\varepsilon(s) \bar{f}_1(s) ds \right) \bar{A}_{12}^\varepsilon(\tau) \delta x_2^\varepsilon(\tau) d\tau \tag{61}$$

and

$$\delta x_2^\varepsilon(t) = \exp\left(\int_0^t \frac{1}{\varepsilon} \bar{A}_{22}^\varepsilon(\tau) d\tau \right) \bar{x}_0^2 + \int_0^t \exp\left(\int_\tau^t \bar{A}_{22}^\varepsilon(s) ds \right) \left(\bar{A}_{12}^\varepsilon(\tau) + \frac{1}{\varepsilon} \omega^\varepsilon(\tau) \bar{f}_2^\varepsilon \right) \delta x_1^\varepsilon(\tau) d\tau \tag{62}$$

where $\bar{x}_0^1 \in \mathbb{R}$ is the first entry of $\bar{x}_0$, and $\bar{x}_0^2 \in \mathbb{R}^{(n-1)}$ represents the remaining $(n - 1)$ entries. Here we immediately see the challenge of bounding the equations of sensitivity in this case. In particular, we observe that the solutions (61) and (62) will "blow up" unless δx_1^ε goes to zero rapidly upon entering the region of regularization. More accurately, we will need to show that δx_1^ε decays exponentially to zero at a rate that is on the order of $\frac{1}{\varepsilon}$ to ensure that this does not occur.

We are now ready to consider the next case in our analysis. In this case, we assume that $\exists \delta > 0$ such that for almost every $t \in (t_1 - \delta, t_2 + \delta)$ we have $L_{f_1} g(\phi_t(\xi), u(t)) > 0$ and $L_{f_2} g(\phi_t(\xi), u(t)) < 0$. In this case the nominal trajectory of the discontinuous system begins to slide along the surface of discontinuity at time t_1. Crucially, we note that in this case for each $t \in (t_1 - \delta, t_1 + \delta)$ we will have

$$\bar{f}_1^\varepsilon(t) < -\hat{f} \tag{63}$$

for some $\hat{f} > 0$ and each ε sufficiently small, since f_1 and f_2 are both "pointing into" the surface of regularization in this case. By our assumption that $L_{f_1} g(\phi_t(\xi), u(t)) > 0$ and $L_{f_2} g(\phi_t(\xi), u(t)) < 0$, and our assumption that φ is strictly increasing on $(-1, 1)$, we see that when ε is sufficiently small that the regularized trajectory will flow to the relative interior of Σ^ε after arriving at the region of regularization at time t^ε, and remain in the relative interior of Σ^ε until at least time $t_1 + \delta$. Moreover, since $\frac{\mathrm{d}}{\mathrm{d}x} \varphi(\frac{g(x)}{\varepsilon}) > 0$ for each $x \in \mathbb{R}^n$ which lies in the relative interior of Σ^ε, we see that there exists $k_1 > 0$ such that $\omega^\varepsilon(t) > k_1$ for each $t \in (\bar{t}^\varepsilon, t_1 + \delta)$ where we define

$$\bar{t}^\varepsilon = t^\varepsilon + \varepsilon. \tag{64}$$

Thus, the differential equation (59) will be damped on the order of $\frac{1}{\varepsilon}$ when the regularized trajectory is in the region of regularization. However, the differential equation (57) will will simultaneously "blow up" exponentially at a rate that is on the order of $\frac{1}{\varepsilon}$. Due to the interconnection of these two differential

equations, it is not imediatly clear which of these effects will prevail. Nonetheless, we will still show that for ε sufficiently small δx_1^ε still decays to zero exponentially at a rate that is on the order of $\frac{1}{\varepsilon}$. Once $\delta x_2^\varepsilon(t)$ decays to the point that it is $(O(\varepsilon))$, the right hand sides of (59) and (57) will $O(1)$ for the rest of the time that the regularized trajectory remains in the region of regularization, and we will have obtained the bound that we desire.

Before proceding to our main analytic step, we first develop a number of useful intermediate bounds. By lemma 5 we see that $\bar{A}_{ij}^\varepsilon(t)$ is bounded for each $t \in [0, T]$ and $i, j \in \{1, 2\}$, and $\bar{f}_2^\varepsilon(t)$ is also bounded for each $t \in [0, T]$, and and thus using familiar properties of the exponential function we bound for each $t \in [0, T]$

$$\|\delta x_1^\varepsilon(t)\| \le \exp\left(\int_0^t \bar{A}_{11}^\varepsilon(\tau)d\tau\right)\exp\left(\int_0^t \frac{1}{\varepsilon}\omega^\varepsilon(\tau)\bar{f}_1^\varepsilon(\tau)d\tau\right)\|\bar{x}_0\|$$
$$+ \int_0^t \exp\left(\int_\tau^t \bar{A}_{11}^\varepsilon(s)ds\right)\exp\left(\int_\tau^t \frac{1}{\varepsilon}\omega^\varepsilon(s)\bar{f}_1^\varepsilon(s)ds\right)\|\delta x_2^\varepsilon(\tau)\|d\tau$$
$$\le C_3 \exp\left(\int_0^t \frac{1}{\varepsilon}\omega^\varepsilon(\tau)\bar{f}_1(\tau)d\tau\right)\|\bar{x}_0\| + C_3 \int_0^t \exp\left(\int_\tau^t \frac{1}{\varepsilon}\omega^\varepsilon(s)\bar{f}_1(s)ds\right)\|\delta x_2^\varepsilon(\tau)\|d\tau \quad (65)$$

and

$$\|\delta x_2^\varepsilon(t)\| \le \exp\left(\int_0^t \bar{A}_{22}^\varepsilon(\tau)d\tau\right)\|\bar{x}_0\| + \int_0^t \exp\left(\int_\tau^t \bar{A}_{22}^\varepsilon(s)ds\right)\frac{1}{\varepsilon}\omega^\varepsilon(\tau)\bar{f}_2^\varepsilon(\tau)$$
$$\le C_3\|\bar{x}_0\| + C_3 \int_0^t \frac{1}{\varepsilon}\omega^\varepsilon(\tau)\|\delta x_1^\varepsilon(\tau)\|d\tau \quad (66)$$

where we chose C_3 to be sufficiently large and we note that $\|\bar{x}_0^1\|, \|\bar{x}_0^2\| \le \|\bar{x}_0\|$ by the Triangle Inequality.

Using these new bounds, for each $t \in (\bar{t}^\varepsilon, t_1 + \delta)$ we are employ (65) and (55) and our definitions of k_1 and $\hat{f}$ we further decompose and bound

$$\|\delta x_1^\varepsilon(t)\| \le C_3 \exp\left(\int_{t^\varepsilon}^{\bar{t}^\varepsilon} \frac{1}{\varepsilon}\omega^\varepsilon(\tau)\bar{f}_1(\tau)d\tau\right)\exp\left(\int_{\bar{t}^\varepsilon}^t -\frac{1}{\varepsilon}k_1\hat{f}d\tau\right)\|\bar{x}_0\|$$
$$+ C_3 \int_{t^\varepsilon}^{\bar{t}^\varepsilon} \exp\left(\int_\tau^t \frac{1}{\varepsilon}\omega^\varepsilon(s)\bar{f}_1(s)ds\right)\|\delta x_2^\varepsilon(\tau)\|d\tau + C_3 \int_{\bar{t}^\varepsilon}^t \exp\left(\int_\tau^t -\frac{1}{\varepsilon}k_1\hat{f}ds\right)\|\delta x_2^\varepsilon(\tau)\|d\tau$$
$$\le C_3 \exp\left(\int_{t^\varepsilon}^{\bar{t}^\varepsilon} \frac{1}{\varepsilon}\omega^\varepsilon(\tau)\bar{f}_1(\tau)d\tau\right)\exp\left(\int_{\bar{t}^\varepsilon}^t -\frac{1}{\varepsilon}k_1\hat{f}d\tau\right)\|\bar{x}_0\|$$
$$+ C_3 \int_{t^\varepsilon}^{\bar{t}^\varepsilon} \exp\left(\int_\tau^t \frac{1}{\varepsilon}\omega^\varepsilon(s)\bar{f}_1(s)ds\right)C_2\exp\left(\int_0^t \frac{1}{\varepsilon}M\omega^\varepsilon(s)ds\right)\|\bar{x}_0\|d\tau + C_3 \int_{\bar{t}^\varepsilon}^t \exp\left(\int_\tau^t -\frac{1}{\varepsilon}k_1\hat{f}ds\right)\|\delta x_2^\varepsilon(\tau)\|d\tau$$
$$\le C_4 \exp\left(\int_{\bar{t}^\varepsilon}^t -\frac{1}{\varepsilon}\hat{f}k_1d\tau\right)\|\bar{x}_0\| + C_4 \int_{\bar{t}^\varepsilon}^t \exp\left(\int_\tau^t -\frac{1}{\varepsilon}\hat{f}k_1ds\right)\|\delta x_2^\varepsilon(\tau)\|d\tau \quad (67)$$

for some $C_4 > 0$ sufficiently large, where we have use (55).

We can follow a similar set of steps and employ (55) to obtain:

$$\|\delta x_2^\varepsilon(t)\| \le C_2 \exp\left(\int_0^t \frac{1}{\varepsilon}M\omega^\varepsilon(\tau)d\tau\right)\|\bar{x}_0\|$$
$$\le C_2 \exp\left(\int_{\hat{t}^\varepsilon}^{\hat{t}^\varepsilon} \frac{1}{\varepsilon}M\omega^\varepsilon(\tau)d\tau\right)\exp\left(\int_{\hat{t}^\varepsilon}^t \frac{1}{\varepsilon}M\omega^\varepsilon(\tau)d\tau\right)\|\bar{x}_0\|$$
$$\le C_5 \exp\left(\int_{\hat{t}^\varepsilon}^t \frac{1}{\varepsilon}MC_2d\tau\right)\|\bar{x}_0\| \quad (68)$$

18

for some $C_5 > 0$ sufficiently large.

Here, we pause to inspect the last inequality in (67). In order to show that δx_1^ε decays exponentially on the order of $\frac{1}{\varepsilon}$, we will need to bound $\|\delta x_1^\varepsilon\|$ on an interval of appropriate length. In particular, observe that if we can show that $\|\delta x^\varepsilon(t)\| < C$ for some $C > 0$ for each $t \in (\bar{t}^\varepsilon, \bar{t}^\varepsilon + \frac{1}{k_2 * \hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon}))$, then $\delta_1^\varepsilon(t)$ will reach a value of $O(\varepsilon)$ by then end of this interval. Once this occurs, both δx_1^ε and δx_2^ε will vary at a rate that is on the order of $O(1)$ for the rest of the time that the regularized trajectory remains in the region of regularization, and we will have obtained the bound we desire. Though we will prove this more carefully later, we provide this commentary to provide an intuition for our following steps, which my be otherwise difficult to follow.

In order to bound δx_2^ε uniformly on the desired interval, we will use the following procedure to repeatedly tighten bounds that we have already obtained for δx_1^ε and δx_2^ε in an inductive manner. In particular, we will begin by plugging in (68) into (67) and then taking several additional steps. The result will be a tighter version of (68). We will then plug this tight bound into (67) and and again repeat the intermediate steps until we have a uniform bound for $\|\delta x_1^\varepsilon(t)\|$ on our desired interval.

To begin, we plug (68) into (67), and integrate to obtain:

$$\|\delta x_1^\varepsilon(t)\| \le C_4 \exp\left(-\frac{1}{\varepsilon}\hat{f}k_1(t - \bar{t}^\varepsilon)\right)\|\bar{x}_0\| + C_4 C_5 \int_{\bar{t}^\varepsilon}^t \exp\left(-\frac{1}{\varepsilon}k_1\hat{f}(t - \tau)\right)\exp\left(\frac{1}{\varepsilon}C_2 M(\tau - \bar{t}^\varepsilon)\right)\|x_0\|d\tau$$
$$= C_4 \exp\left(-\frac{1}{\varepsilon}\hat{f}k_1(t - \bar{t}^\varepsilon)\right)\|\bar{x}_0\| + C_4 C_5 \frac{\varepsilon}{M * C_2 + k_1 * \hat{f}}\left(\exp\left(\frac{1}{\varepsilon}C_2 * M(t - \bar{t}^\varepsilon)\right) - 1\right)\|\bar{x}_0\|. \quad (69)$$

Next we plug this expression into (66) and again use $\omega^\varepsilon(t) \le C_2$ to obtain

$$\|\delta x_2^\varepsilon(t)\| \le C_3\|\bar{x}_0\| + \|\bar{x}_0\|C_3 \int_{\bar{t}^\varepsilon}^t \frac{1}{\varepsilon}C_4\left[\exp\left(-\frac{1}{\varepsilon}\hat{f}k_1(t - \bar{t}^\varepsilon)\right)\right.$$
$$\left. + C_5 \frac{\varepsilon}{M * C_2 + k_1 * \hat{f}}\left(\exp\left(\frac{1}{\varepsilon}C_2 * M(t - \bar{t}^\varepsilon)\right) - 1\right)\right]d\tau \quad (70)$$

and then pulling out constants we then obtain

$$\|\delta x_2^\varepsilon(t)\| \quad \le \quad C_6\|\bar{x}_0\|\left[1 \ + \ \int_{\bar{t}^\varepsilon}^t \frac{1}{\varepsilon}\left[\exp\left(-\frac{1}{\varepsilon}\hat{f}k_1(t - \bar{t}^\varepsilon)\right) \ + \ \varepsilon\exp\left(\frac{1}{\varepsilon}C_2 * M(t - \bar{t}^\varepsilon)\right)\right]d\tau\right] \quad (71)$$

for some $C_6 > 0$ sufficiently large, and then we integrate to obtain

$$\|\delta x_2^\varepsilon(t)\| \le C_6\|\bar{x}_0\|\left[1 + \frac{1}{\hat{f} * k_1}\left[\exp\left(-\frac{1}{\varepsilon}\hat{f}k_1(t - \bar{t}^\varepsilon)\right) - 1\right] + \frac{\varepsilon}{C_2 * M}\left[\exp\left(\frac{1}{\varepsilon}C_2 * M(t - \bar{t}^\varepsilon)\right) - 1\right]\right]. \quad (72)$$

Again pulling out constants, and noting that for each appropriate choice of t and τ we have

$$\exp\left(-\frac{1}{\varepsilon}\hat{f}k_1(t - \bar{t}^\varepsilon)\right) \le 1 \quad (73)$$

we then obtain

$$\|\delta x_2^\varepsilon(t)\| \le C_7\|\bar{x}_0\|\left(1 + \varepsilon\exp\left(\frac{1}{\varepsilon}C_2 * M(t - \bar{t}^\varepsilon)\right)\right). \quad (74)$$

for some $C_7 > 0$ sufficiently large. On the desired interval $t \in (\bar{t}^\varepsilon, \bar{t}^\varepsilon + \frac{1}{k_2 * \hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon}))$ we have that

$$\|\delta x_2^\varepsilon(t)\| \le C_7\|\bar{x}_0\|\left(1 + \varepsilon\exp\left(\frac{C_2 * M}{k_2 * \hat{f}}\ln(\frac{1}{\varepsilon})\right)\right) = C_7\|\bar{x}_0\|\left(1 + \varepsilon * \varepsilon^{-\frac{C_2 * M}{k_2 * \hat{f}}}\right) = C_7\|\bar{x}_0\|\left(1 + \varepsilon^{1 - \frac{C_2 * M}{k_2 * \hat{f}}}\right)$$
$$(75)$$

In the case that $\frac{C_2*M}{k_2*\hat{f}} < 1$ we will have that δx_2^ε is bounded on the desired interval. However, in the case that $\frac{C_2*M}{k_2*\hat{f}} > 1$ the above bound will tend to infinity as we take $\varepsilon \to 0$ and will not serve our needs. However, we can now plug (74) into our above analysis to tighten the bound. In particular, rather than plug (68) into (67) in equation (69), we can now put in the improved bound (74), and follow the steps take from (69) to (74) to improve the bound. Though we omit the details in the interest of brevity, it can be shown that this proccess yields the following bound:

$$\|\delta x_2^\varepsilon(t)\| \leq C_8 \|\bar{x}_0\| \left(1 + \varepsilon^2 \exp\left(\frac{1}{\varepsilon} C_2 * M(t - \bar{t}^\varepsilon)\right)\right), \tag{76}$$

for some $C_8 > 0$ which is chosen to be sufficiently large. If we instead repeat the above process p times, where $p > \frac{C_2*M}{k_2*\hat{f}}$, then we obtain abound of the form

$$\|\delta x_2^\varepsilon(t)\| \leq C_p \|\bar{x}_0\| \left(1 + \varepsilon^p \exp\left(\frac{1}{\varepsilon} C_2 * M(t - \bar{t}^\varepsilon)\right)\right) \tag{77}$$

for some $C_p > 0$ sufficiently large, and in this case we may bound $\delta x_2^\varepsilon(t)$ for each $t \in (\bar{t}^\varepsilon, \bar{t}^\varepsilon + \frac{1}{k_2*\hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon}))$ by

$$\|\delta x_2^\varepsilon(t)\| \leq C_p \|\bar{x}_0\| \left(1 + \varepsilon^p \exp\left(\frac{C_2 * M}{k_2 * \hat{f}} \ln(\frac{1}{\varepsilon})\right)\right) = C_p \|\bar{x}_0\| \left(1 + \varepsilon^p * \varepsilon^{-\frac{C_2*M}{k_2*\hat{f}}}\right) = C_p \|\bar{x}_0\| \left(1 + \varepsilon^{p-\frac{C_2*M}{k_2*\hat{f}}}\right), \tag{78}$$

and in this case we will have that for each $t \in (\bar{t}^\varepsilon, \bar{t}^\varepsilon + \frac{1}{k_2*\hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon}))$

$$\|\delta x_2^\varepsilon(t)\| \leq C_9 \|\bar{x}_0\|, \tag{79}$$

where $C_9 > 0$ is chosen to be sufficiently large. Combining this bound with (67), on the interval $t \in (\bar{t}^\varepsilon, \bar{t}^\varepsilon + \frac{1}{k_2*\hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon}))$ we have

$$\|\delta x_1^\varepsilon(t)\| \leq C_4 \exp\left(\int_{\bar{t}^\varepsilon}^t -\frac{1}{\varepsilon} \hat{f} k_1 d\tau\right) \|\bar{x}_0\| + C_4 \int_{\bar{t}^\varepsilon}^t \exp\left(\int_\tau^t -\frac{1}{\varepsilon} \hat{f} k_1 ds\right) C_9 \|\bar{x}_0\| d\tau$$

$$\leq C_{11} \exp\left(\int_{\bar{t}^\varepsilon}^t -\frac{1}{\varepsilon} \hat{f} k_1 d\tau\right) \|\bar{x}_0\| + \varepsilon C_{11} \|\bar{x}_0\| \tag{80}$$

for some $C_{11} > 0$ sufficiently large and setting $\hat{t}^\varepsilon = \bar{t}^\varepsilon + \frac{1}{k_2*\hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon})$ we have

$$\left\|\delta_1 x^\varepsilon(\hat{t}^\varepsilon)\right\| \leq C_4 \varepsilon \|\bar{x}_0\| + C_9 * \frac{\varepsilon}{\hat{f} k_1}. \tag{81}$$

Thus, we see that $\|\delta x^\varepsilon\|$ decays exponentially to zero at a rate of decay that is on the order of $\frac{1}{\varepsilon}$, as desired. Moreover, combing (80) with (80) and integrating through we obtain for each $t \in (\bar{t}^\varepsilon, \bar{t}^\varepsilon + \frac{1}{k_2*\hat{f}} * \varepsilon \ln(\frac{1}{\varepsilon}))$

$$\|\delta x_2^\varepsilon(t)\| \leq C_3 \|\bar{x}_0\| + C_3 \int_0^t \frac{1}{\varepsilon} C_2 \left(C_{11} \exp\left(\int_{\bar{t}^\varepsilon}^t -\frac{1}{\varepsilon} \hat{f} k_1 ds\right) \|\bar{x}_0\| + \varepsilon C_{11}\right) d\tau \leq C_{12} \|\bar{x}_0\| \tag{82}$$

for some $C_{12} > 0$ sufficiently large. From here it is straightforward to verify that the equations of sensitivity remain bounded for the rest of the time that the nominal trajectory is in the region of regularization. We omit a detailed argument in the interest of brevity, since it closely follows the argument made in [17].

Finally, we discuss the case where the nominal trajectory of the discontinuous system either stays on the surface of discontinuity with out sliding, and the case where the nominal trajectory of the discontinuous system instantaneously crosses the surface of discontinuity, but does not do so transversally. This

covers the remaining situations we must consider under Assumptions 2-4. We provide only a sketch of the proofs in these case, since the arguments follow largely from our previous analysis.

First, suppose that the nominal trajectory of the discontinuous system stays on the surface of discontinuity on the interval (t_1, t_2) where $t_2 > 0$ but does not slide. For almost every time $t \in (t_1, t_2)$, we will have either $L_{f_1} g(\phi_t(\xi), u(t)) > 0$ and $L_{f_1} g(\phi_t(\xi), u(t)) \leq 0$ or $L_{f_1} g(\phi_t(\xi), u(t)) \geq 0$ and $L_{f_1} g(\phi_t(\xi), u(t)) < 0$, otherwise the trajectory would leave the surface of discontinuity before time t_2. Since either f_1 or f_2 is always "pointing into" the surface of discontinuity, for each ε small enough, we will again have $\bar{f}_1^\varepsilon(t) < 0$ for almost every time $t \in (t_1, t_2)$, and we will again have exponential dampening of variations that are normal to the surface of discontinuity. Thus, the analysis to show that the equations of sensitivity are bounded in this case follows closely the previous case with a few minor modifications.

In the case where the nominal trajectory crosses the surface of discontinuity but does not do so transversally, it will be the case that there exists δ such that for almost every $t \in (t_1 - \delta, t_1 + \delta)$ we have $L_{f_1} g(\phi_t(\xi), u(t)) > 0$ and $L_{f_2} g(\phi_t(\xi), u(t)) \geq 0$. Moreover, given any $\gamma > 0$ we may choose δ to be to be small enough so that $\gamma \leq L_{f_2} g(\phi_t(\xi), u(t))$ for almost every $t \in (t_1 - \delta, t_1 + \delta)$. This is a consequence of the trajectory not leaving the surface of discontinuity transversally. Thus, for ε small enough, will again have that $\bar{f}_1^\varepsilon(t) < 0$ for almost every t such that $\sigma_t^\varepsilon(\xi) \in \Sigma^\varepsilon$. Thus, the previous analysis can again can be applied to this situation with minor modifications to show that the variations on the regularized trajectory remain bounded as the trajectory flows through the interior of the region of regularization. We omit the details of the proof for these final two cases, because it would involve reintroducing the large number of bounds and notations introduced in the case of sliding without providing much new insight.

A.2   Proof of Lemma 4

By the proof of Lemma 2 we see that that there exists a neighborhood of ξ on which the map ϕ_t is continuous for each $t \in [0, T]$ (this is a well known property of Filippov solutions [4, Chapter 2, Section 7]). This demonstrates assumptions 2 and 4 hold in a neighborhood of ξ. We then see that 3 holds in a neighborhood of ξ by noting that f_1 and f_2 are assumed to be continuous.

A.3   Proof of Lemma 3

In order to again simplify our analysis, we again make the simplifying assumptions made in the proof of Lemma 2. That is, we again assume that the nominal trajectory of the discontinuous system reaches the surface of discontinuity exactly once at time $t_1 \in \mathcal{T}^\xi$ and that $x_0 \in D_1$. We also reuse all of the terminology and notation from the proof of Lemma 2. We again prove the desired result by studying the equations of sensitivity for the relaxed system, and demonstrating that the map $\varepsilon \to D\phi_t(\xi; \delta\xi)$ is Lipschitz continuous when we restrict our attention to values of ε that are small enough.

For each $\varepsilon > 0$, we define the *fast time scale*

$$\tau = \frac{t - t^\varepsilon}{\varepsilon} \tag{83}$$

on this timescale, the the dynamics of the sensitivity equations are given by

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \delta x^\varepsilon(\tau) = \varepsilon[(1 - \varphi^\varepsilon(\tau))A_{11}^\varepsilon(\tau) + \varphi^\varepsilon(\tau)]\delta x^\varepsilon(\tau) + \omega^\varepsilon(\tau)f_1(\tau)\delta x_1^\varepsilon(\tau). \tag{84}$$

We again begin by studying the case where the nominal trajectory of the discontinuous systems crosses the surface of discontinuity transversally. That is, we assume that there exists $\delta > 0$ such that for each $t \in (t_1 - \delta, t_1 + \delta)$ we have $L_{f_1} g(\phi_t(\xi), u(t)) > 0$ and $L_{f_2} g(\phi_t(\xi), u(t)) > 0$. Similarly, to the proof of Lemma 2, our analysis in this case largely relies on the fact that for each ε small enough the corresponding regularized trajectory will spend on the order of ε units of time in the region of regularization. In particular, we begin by bounding:

a

References

[1] X. Xiong and A. D. Ames, "Coupling reduced order models via feedback control for 3d underactuated bipedal robotic walking," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2018, pp. 1–9.

[2] E. R. Westervelt, J. W. Grizzle, and D. E. Koditschek, "Hybrid zero dynamics of planar biped walkers," *IEEE transactions on automatic control*, vol. 48, no. 1, pp. 42–56, 2003.

[3] I. A. Hiskens and M. Pai, "Trajectory sensitivity analysis of hybrid systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 2, pp. 204–220, 2000.

[4] A. F. Filippov, *Differential equations with discontinuous righthand sides: control systems*. Springer Science & Business Media, 2013, vol. 18.

[5] E. Polak, *Optimization: algorithms and consistent approximations*. Springer Science & Business Media, 2012, vol. 124.

[6] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[7] M. Kamgarpour and C. Tomlin, "On optimal control of non-autonomous switched systems with a fixed mode sequence," *Automatica*, vol. 48, no. 6, pp. 1177–1181, 2012.

[8] M. S. Shaikh and P. E. Caines, "On the hybrid optimal control problem: theory and algorithms," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1587–1603, 2007.

[9] X. Xiong and A. D. Ames, "Bipedal hopping: Reduced-order model embedding via optimization-based control," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3821–3828.

[10] ——, "Motion decoupling and composition via reduced order model optimization for dynamic humanoid walking with clf-qp based active force control," in *Submitted to 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

[11] S. Sastry, *Nonlinear systems: analysis, stability, and control*. Springer Science & Business Media, 2013, vol. 10.

[12] H. K. Khalil, "Noninear systems."

[13] J. Llibre, P. R. da Silva, M. A. Teixeira *et al.*, "Sliding vector fields via slow–fast systems," *Bulletin of the Belgian Mathematical Society-Simon Stevin*, vol. 15, no. 5, pp. 851–869, 2008.

[14] J. Llibre and M. A. Teixeira, "Regularization of discontinuous vector fields in dimension three," *Discrete & Continuous Dynamical Systems-A*, vol. 3, no. 2, pp. 235–241, 1997.

[15] D. Fiore, M. Coraggio, and M. di Bernardo, "Observer design for piecewise smooth and switched systems via contraction theory," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 2959–2964, 2017.

[16] D. Fiore, S. J. Hogan, and M. di Bernardo, "Contraction analysis of switched systems via regularization," *Automatica*, vol. 73, pp. 279–288, 2016.

[17] D. E. Stewart and M. Anitescu, "Optimal control of systems with discontinuous differential equations," *Numerische Mathematik*, vol. 114, no. 4, pp. 653–695, 2010.

[18] T. Westenbroek, H. Gonzalez, and S. S. Sastry, "A new solution concept and family of relaxations for hybrid dynamical systems," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 743–750.

[19] S. N. Simic, K. H. Johansson, J. Lygeros, and S. Sastry, "Towards a geometric theory of hybrid systems," *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms*, vol. 12, no. 5-6, pp. 649–687, 2005.

[20] S. A. Burden, H. Gonzalez, R. Vasudevan, R. Bajcsy, and S. S. Sastry, "Metrization and simulation of controlled hybrid systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 9, pp. 2307–2320, 2015.

[21] A. D. Ames and S. Sastry, "A homology theory for hybrid systems: Hybrid homology," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2005, pp. 86–102.

[22] J. Dieudonné, *Foundations of modern analysis*. Read Books Ltd, 2013.

[23] T. Westenbroek, X. Xiong, A. D. Ames, and S. S. Sastry, "Technical report: Optimal control of piece-wise smooth control systems via singular perturbations," *Arxiv Preprint*.

[24] E. Polak and Y. Y. Wardi, "A study of minimizing sequences," *SIAM Journal on Control and Optimization*, vol. 22, no. 4, pp. 599–609, 1984.

[25] A. D. Ames and M. Powell, "Towards the unification of locomotion and manipulation through control lyapunov functions and quadratic programs," in *Control of Cyber-Physical Systems*. Springer, 2013, pp. 219–240.

[26] https://youtu.be/FeIjtTDrbGs.