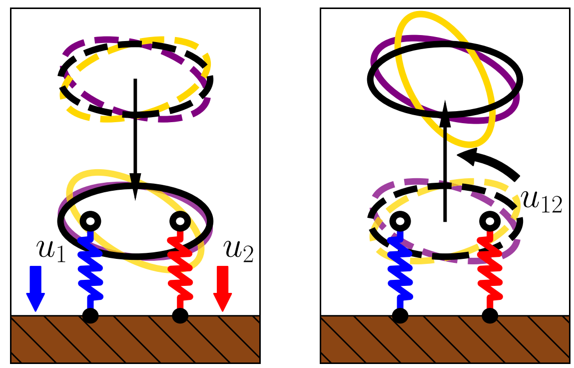
Nonsmooth Optimal Value and Policy Functions in Mechanical Systems Subject to Unilateral Constraints

Bora S. Banjanin and Samuel A. Burden

***Abstract*—State-of-the-art approaches to optimal control use smooth approximations of value and policy functions and gradient-based algorithms for improving approximator parameters. Unfortunately, we show that value and policy functions that arise in optimal control of mechanical systems subject to unilateral constraints – i.e., the contact-rich dynamics of robot locomotion and manipulation – are generally nonsmooth due to the underlying dynamics exhibiting discontinuous or piecewise-differentiable trajectory outcomes. Simple mechanical systems are used to illustrate this result and the implications for optimal control of contact-rich robot dynamics.**

***Index Terms*—Optimal control, robotics, hybrid systems.**



(a) *touchdown* maneuver (b) *liftoff* maneuver

Fig. 1. Saggital-plane biped performs two maneuvers with contact-rich dynamics – (a) *touchdown* and (b) *liftoff* – using policies that exert different forces depending on which feet are in contact with the ground. In the *touchdown* maneuver, feet are initially off the ground and trajectories terminate when the body height reaches nadir; in the *liftoff* maneuver, feet are initially on the ground and trajectories terminate when the body height reaches apex. Fig. 2 shows that the final body rotation is a nonsmooth – piecewise-differentiable or discontinuous – function of initial body rotation.

I. Introduction

THIS letter focuses on optimal control of *mechanical* systems subject to *unilateral* constraints [1], [2], which are commonly used to model contact-rich dynamics of rigid robots [3]. In an optimal control problem, a policy is sought that extremizes a given performance criterion; the performance achieved by this *optimal policy* is the *optimal value* of the problem. Two popular approaches for solving such problems are trajectory optimization [4] and reinforcement learning [5]. Although many algorithms are available in either framework, scalable algorithms in both leverage local approximations – gradients of values and/or policies – to iteratively improve toward optimality. In applications with smooth dynamics, these gradients are guaranteed to exist and can be readily computed or approximated.

Recent work has applied state-of-the-art algorithms for trajectory optimization [6]–[9] and reinforcement learning [10]–[13] to optimal control of contact-rich dynamics, producing impressive results in simulations and experiments of robot manipulation and locomotion. However, the algorithms underlying these results [4], [14] are only known to converge to stationary points in smooth systems since they rely on gradients of the functions that define costs and constraints.

We show that these gradients generally fail to exist for mechanical systems subject to unilateral constraints due to nonsmoothness in the underlying dynamics; this result is derived theoretically in Theorem 1 and demonstrated using the simple mechanical system depicted in Fig. 1. These contributions imply that additional work is required to justify applying state-of-the-art algorithms for optimal control to mechanical systems with contact-rich robot dynamics.

We begin in Section II by reviewing concepts from optimization and nonsmooth analysis that are needed to state and understand our technical results. We continue in Section III by modeling contact-rich robot dynamics using mechanical systems subject to unilateral constraints, and describe how nonsmoothness – discontintinuity or piecewise-differentiability – manifests in trajectory outcomes and (hence) costs. In Section IV we provide mathematical derivations that

Manuscript received August 27, 2019; revised November 9, 2019; accepted December 3, 2019. Date of publication December 17, 2019; date of current version January 22, 2020. This work was supported in part by the U.S. Army Research Office under Grant W911NF-16-1-0158 and in part by NSF Cyber-Physical Systems Award under Grant 1836819. Recommended by Senior Editor G. Cherubini. *(Corresponding author: Samuel A. Burden.)*

The authors are with the Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: borab@uw.edu; sburden@uw.edu).

Digital Object Identifier 10.1109/LCSYS.2019.2960442

show nonsmoothness in trajectory outcomes and costs gives rise to nonsmoothness in optimal value and (hence) policy functions, and present numerical simulations in Section V that demonstrate discontinuous or merely piecewise-differentiable optimal value and policy functions in a simple system. Finally in Section VI we discuss the prevalence of nonsmoothness in applications and how the lack of classical differentiability prevents gradient-based algorithms from converging to optimality.

II. Preliminaries

Our main result (Thm. 1) provides conditions under which the *value* and *policy* functions associated with an optimal control problem are *piecewise-differentiable*. In this section, we review concepts from optimization and nonsmooth analysis needed to state and understand this result.

A. Optimization

Consider minimization of the *cost function* $c : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ with respect to an *input* $u \in \mathcal{U}$:

$$\nu(x) = \min_{u \in \mathcal{U}} c(x, u). \tag{1}$$

We restrict our attention to the case where $\mathcal{X}$ and $\mathcal{U}$ are open subsets of finite-dimensional vector spaces, and assume in what follows that the *optimal value* $\nu : \mathcal{X} \to \mathbb{R}$ in (1) is well-defined. We let $\pi : \mathcal{X} \to \mathcal{U}$ denote an *optimal policy* for (1), i.e.,

$$\forall x \in \mathcal{X} : \pi(x) \in \arg\min_{u \in \mathcal{U}} c(x, u) \tag{2}$$

or, equivalently,

$$\forall x \in \mathcal{X} : \nu(x) = c(x, \pi(x)). \tag{3}$$

B. Nonsmooth Analysis

A continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *piecewise-differentiable* on the open set $O \subset \mathbb{R}^n$ if there exists a finite collection of continuously-differentiable functions $\{f_i : O \to \mathbb{R}^m\}_{i \in I}$ such that $f(x) \in \{f_i(x)\}_{i \in I}$ for all $x \in O$ [15, Sec. 4.1]; if the functions $\{f_i\}_{i \in I}$ are r-times continuously-differentiable we write $f_i \in C^r(O, \mathbb{R}^m)$ and $f \in PC^r(O, \mathbb{R}^m)$ (or simply $f_i \in C^r$, $f \in PC^r$ when the domain and codomain are clear from context).

The preceding definition of piecewise-differentiability was originally developed by the nonsmooth analysis community to study structural stability of nonlinear programs [16]. Although the definition may appear unrelated to the intuition that a function ought to be piecewise-differentiable precisely when its "domain can be partitioned locally into a finite number of regions relative to which smoothness holds" [17, Sec. 1], the class of PC^r functions defined above are always piecewise-differentiable in this intuitive sense [17, Th. 2]. If $f \in PC^r(\mathbb{R}^n, \mathbb{R}^m)$ then f admits a non-linear first-order approximation $Df : T\mathbb{R}^n \to T\mathbb{R}^m$ termed the *Bouligand* (or *B*-) derivative [15, Secs. 3.1 and 4.1.2], and $Df(x; v)$ is the directional derivative of f in the $v \in T_x\mathbb{R}^n$ direction, but B-differentiability is a strictly stronger property that will be leveraged to obtain our main result (Thm. 1).

III. Contact-Rich Robot Dynamics

In this section, we formalize a class of models for contact-rich dynamics in robot locomotion and manipulation as *mechanical systems subject to unilateral constraints* and formulate an optimal control problem for these systems.

A. Mechanical Systems Subject to Unilateral Constraints

Consider the dynamics of a mechanical system with $d \in \mathbb{N}$ degrees-of-freedom (DOF) $q \in Q = \mathbb{R}^d$ subject to $n \in \mathbb{N}$ unilateral constraints $a(q) \geq 0$ specified by a continuously-differentiable function $a : Q \to \mathbb{R}^n$, where the inequality is enforced componentwise. Given any $J \subset \{1, \ldots, n\}$, and letting $|J|$ denote the number of elements in the set J, we let $a_J : Q \to \mathbb{R}^{|J|}$ denote the function obtained by selecting the component functions of a indexed by J. We assume in what follows that the constraints are *perfect* (*ideal* or *frictionless*), that is, they only exert force normal to the constraint surface. With $J = \{j \in \{1, \ldots, n\} : a_j(q) = 0\}$ denoting the *contact mode*, the system's dynamics take the form [3, Secs. 2.4 and 2.5]

$$M(q)\ddot{q} = f_J(q, \dot{q}, u) + Da_J(q)^\top \lambda_J(q, \dot{q}, u), \tag{4a}$$

$$\dot{q}^+ = \Delta_J(q)\dot{q}^-, \tag{4b}$$

where: $M(q) \in \mathbb{R}^{d \times d}$ is the mass matrix; $f_J(q, \dot{q}, u) \in \mathbb{R}^d$ is the vector of Coriolis, potential, and applied forces; $u \in \mathcal{U}$ is an external input; $Da_J(q) \in \mathbb{R}^{|J| \times d}$ denotes the derivative of the constraint function a_J, which we assume has full rank; $\lambda_J(q, \dot{q}, u) \in \mathbb{R}^{|J|}$ denotes the reaction forces generated in contact mode J to maintain $a_J(q) \geq 0$; $\Delta_J(q) \in \mathbb{R}^{d \times d}$ specifies the collision restitution law that instantaneously resets velocities to ensure compatibility with the constraint $a_J(q) = 0$; and $\dot{q}^+$ (resp. $\dot{q}^-$) denotes the right- (resp. left-) handed limits of the velocity with respect to time. Note that we explicitly allow the dynamics in (4) to vary with contact mode $J \subset \{1, \ldots, n\}$ to model the mode-switching optimal controllers derived below.

B. Properties of Contact-Rich Robot Dynamics

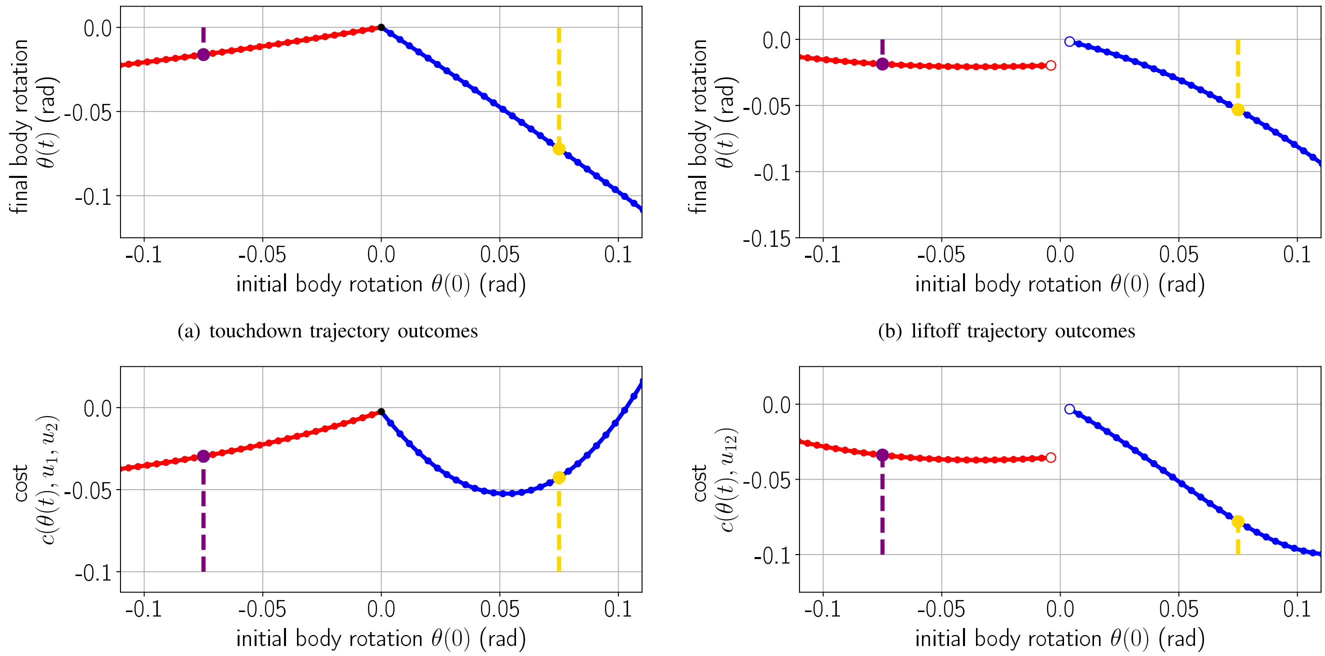
The seemingly benign equations in (4) can yield dynamics with a range of regularity properties. This issue has been investigated elsewhere [1], [18], [19]; here we focus specifically on how the design of a robot's *mechanical* and *control* systems affect (non)smoothness of trajectory outcomes.

It is common to assume that the functions in (4) are continuously-differentiable; however, as illustrated by [1, Ex. 2], this assumption alone does not ensure even existence or uniqueness of trajectories, let alone smoothness of trajectory outcomes. This case contrasts starkly with that of a smooth differential or difference equation

$$\dot{x} \text{ or } x^+ = F(x, u), \tag{5}$$

which yields unique trajectories that vary smoothly with respect to state $x \in \mathcal{X}$ and control input $u \in \mathcal{U}$ [4, Th. 5.6.8]. Since we are chiefly concerned with how properties of the dynamics in (4) affect properties of optimal value and policy functions, we will assume[1] in what follows that conditions

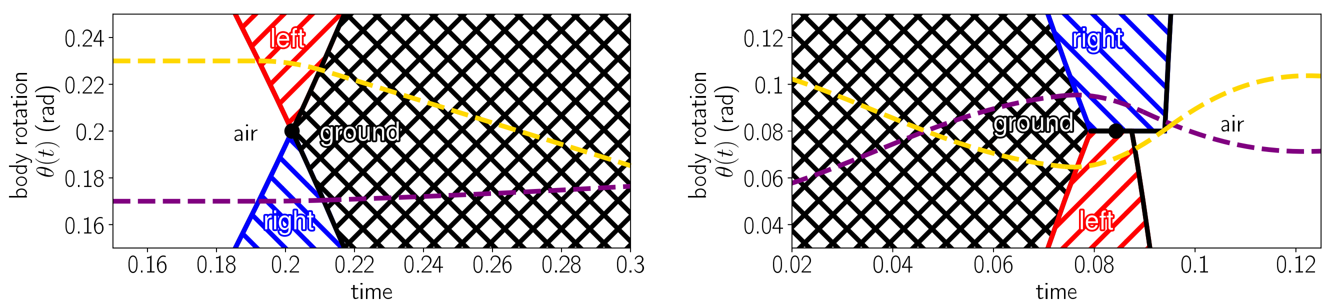
[1] We refer the interested reader to [1, Th. 10] or [3] for conditions that ensure existence and uniqueness of trajectories of (4).



(a) touchdown trajectory outcomes (b) liftoff trajectory outcomes

(c) touchdown value (d) liftoff value

Fig. 2. *Piecewise-differentiable and discontinuous trajectory outcomes in the saggital-plane biped from Fig. 1.* (a,b) Trajectory outcomes (final body angle $\theta(t)$) as a function of initial body angle $\theta(0)$. (c,d) Performance of trajectories as measured by the cost functions in (14), (15). Dashed colored vertical lines indicate corresponding colored outcomes in Fig. 1 and dashed lines in Fig. 3.



(a) touchdown contact modes (b) liftoff contact modes

Fig. 3. *Contact modes for touchdown and liftoff maneuvers.* The saggital-plane biped illustrated in Fig. 1(a,b) can be in one of four *contact modes* corresponding to which subset $J \subset \{1, 2\}$ of the (two) limbs are in contact with the ground; each subset yields different dynamics in (4). (a,b) System contact mode at each time t for a given initial body rotation $\theta(0)$; the body torque input is zero ($u_{12} = 0$) and the leg forces are different ($u_1 \neq u_2$) in mode *left* ($\{1\}$) and *right* ($\{2\}$) than in *air* ($\emptyset$) or *ground* ($\{1, 2\}$). Dashed colored lines indicate corresponding colored trajectories in Fig. 1. The increase in force during the transition to modes *left* and *right* in (b) changes the ground reaction force discontinuously, delaying liftoff and causing discontinuous trajectory outcomes in Fig. 1(d).

have been imposed to ensure unique trajectories of (4) exist for time horizons, initial states, and control inputs of interest.

Assuming that unique trajectories exist for (4) does not provide any regularity properties on the trajectory outcomes; these properties are determined by the design of a robot's mechanical and control systems and their closed-loop interaction with the environment. For instance: when limbs are inertially coupled (e.g., by rigid struts and joints), so that one limb's constraint activation instantaneously changes another's velocity, trajectories can vary discontinuously near configurations where these two limbs activate constraints simultaneously [20, Table 3] [21]; when limbs are force coupled (e.g., by a mode-switching controller), so that one limb's constraint (de)activation instantaneously changes the force on another, trajectories can vary piecewise-differentiably near configurations where these two limbs (de)activate constraints simultaneously [18, Fig. 1]. It is this force coupling, explicitly permitted by the mode-dependence of the dynamics in (4), that we will leverage to obtain nonsmooth trajectory outcomes in the examples presented in Section V.

C. Properties of Optimal Value and Policy Functions

A broad class of optimal control problems for the dynamics in (4) can be formulated in terms of *final* ($\ell : \mathcal{X} \to \mathbb{R}$) and *running* ($\mathcal{L} : [0, t] \times \mathcal{X} \times \mathcal{U} \to \mathbb{R}$) costs:

$$\nu(x) = \min_{u \in \mathcal{U}} \ell(\phi^{x,u}(t)) + \int_0^t \mathcal{L}(s, \phi^{x,u}(s), u)\, ds, \tag{6}$$

where $\phi^{x,u} : [0, t] \to \mathcal{X}$ denotes the unique trajectory obtained from initial state $\phi^{x,u}(0) = x \in \mathcal{X}$ when input $u \in \mathcal{U}$ is applied. Note in (6) that we consider inputs u drawn from an open subset $\mathcal{U}$ of a finite-dimensional vector space; in the context of optimal control, a finite-dimensional input vector $u \in \mathcal{U}$ could parameterize a time- or state-dependent control law. To expose the dependence of the cost in (6) on the trajectory outcome function ϕ, we transcribe the problem in (6) to a simpler form using a standard state augmentation technique [4, Ch. 4.1.2],

$$v(x) = \min_{u\in\mathcal{U}} c(\phi(t, x, u)), \tag{7}$$

where $\phi : [0, t] \times \mathcal{X} \times \mathcal{U} \to \mathcal{X}$ is the *flow* function defined by $\phi(t, x, u) = \phi^{x,u}(t)$. As discussed in Section III-B, the properties of ϕ are determined by a robot's design: it is possible for ϕ (and hence $c \circ \phi$) to be discontinuous ($\phi \notin C^0$) or piecewise-differentiable and not continuously-differentiable ($\phi \in PC^r \backslash C^r$) depending on the properties of the robot's mechanical and control systems. In the next section, we study how continuity and differentiability properties of the cost $c \circ \phi$ affect the corresponding properties of the value v in (7).

IV. Nonsmooth Optimal Value & Policy Functions

In this section, we study how continuity and differentiability properties of a cost function $c : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ relate to corresponding properties of the optimal value $v : \mathcal{X} \to \mathbb{R}$ and policy $\pi : \mathcal{X} \to \mathcal{U}$ functions as defined in Section II-A. Our main result is Thm. 1, which provides conditions that ensure the optimal value and policy functions are piecewise-differentiable (PC^1) in the sense defined in Section II-B.

A. Discontinuous Cost Functions

If the cost ($c : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$) is discontinuous with respect to its first argument, then the optimal policy ($\pi : \mathcal{X} \to \mathcal{U}$) and value ($v : \mathcal{X} \to \mathbb{R}$) are generally discontinuous as well. This observation is clear in the trivial case that the cost only depends on its first argument, but manifests more generally.

B. Piecewise-Differentiable Cost Functions

If c is piecewise-differentiable ($c \in PC^1$), then necessarily

$$\forall w \in T_u\mathcal{U} : D_2c(x, \pi(x); w) \geq 0. \tag{8}$$

Here, $D_2c(x, \pi(x)) : T_u\mathcal{U} \to \mathbb{R}$ denotes the continuous and piecewise-linear first-order approximation termed the *Bouligand* (or *B*-)derivative [15, Ch. 3.1] that exists by virtue of the cost being PC^1 [15, Lemma 4.1.3]; $D_2c(x, \pi(x); w)$ denotes the evaluation of $D_2c(x, \pi(x))$ at $w \in T_u\mathcal{U}$.

If c is two times piecewise-differentiable ($c \in PC^2$), and if a sufficient condition [22, Th. 1] for strict local optimality for (1) is satisfied at $\pi(x) \in \mathcal{U}$,

$$\begin{aligned}\forall w \in \{w \in T_u\mathcal{U} \mid w \neq 0,\ D_2c(x, \pi(x); w) = 0\}\\ : D_2^2c(x, \pi(x); w, w) > 0,\end{aligned} \tag{9}$$

and if the piecewise-linear function

$$D_2^2c(x, \pi(x)) : T_u\mathcal{U} \to T_u\mathcal{U} \text{ is invertible,} \tag{10}$$

then a PC^1 Implicit Function Theorem can be applied to choose $\pi \in PC^1$ near x [23, Corollary 3.4]. Applying the PC^1 Chain Rule [15, Th. 3.1.1] to (8) yields (see [23, Sec. 3])

$$\forall v \in T_x\mathcal{X} : D\pi(x; v) = -D_2^2c(x, \pi(x))^{-1}(D_{12}c(x, \pi(x); v)), \tag{11}$$

and applying the PC^1 Chain Rule to (3) yields

$$\begin{aligned}\forall v \in T_x\mathcal{X} : Dv(x; v) &= D_xc(x, \pi(x); v)\\ &= D_1c(x, \pi(x); v) + D_2c(x, \pi(x); D\pi(x; v)),\end{aligned} \tag{12}$$

whence we obtain B-derivatives of the optimal value and policy functions in terms of B-derivatives of the cost.

We conclude that if the cost function is two times piecewise-differentiable ($c \in PC^2$) and first-order necessary (8) and second-order sufficient (9), (10) conditions for optimality and stability of solutions to (1) are satisfied at $u = \pi(x)$, then the optimal policy and value functions are piecewise-differentiable at x ($\pi, v \in PC^1$) and their B-derivatives at x can be computed using (11), (12).

Theorem 1: If $c \in PC^2(\mathcal{X}\times\mathcal{U}, \mathbb{R})$ satisfies (8), (9), and (10) at $(\xi, \mu) \in \mathcal{X} \times \mathcal{U}$, then there exist neighborhoods $X \subset \mathcal{X}$ of ξ and $U \subset \mathcal{U}$ of μ and a function $\pi \in PC^1(X, U)$ such that $\pi(\xi) = \mu$ and, for all $x \in X$, $\pi(x)$ is the unique minimizer for

$$v(x) = \min_{u\in U} c(x, u); \tag{13}$$

the B-derivative of π is given by (11), and the B-derivative of v is given by (12).

C. Conclusions About Optimal Value & Policy Functions

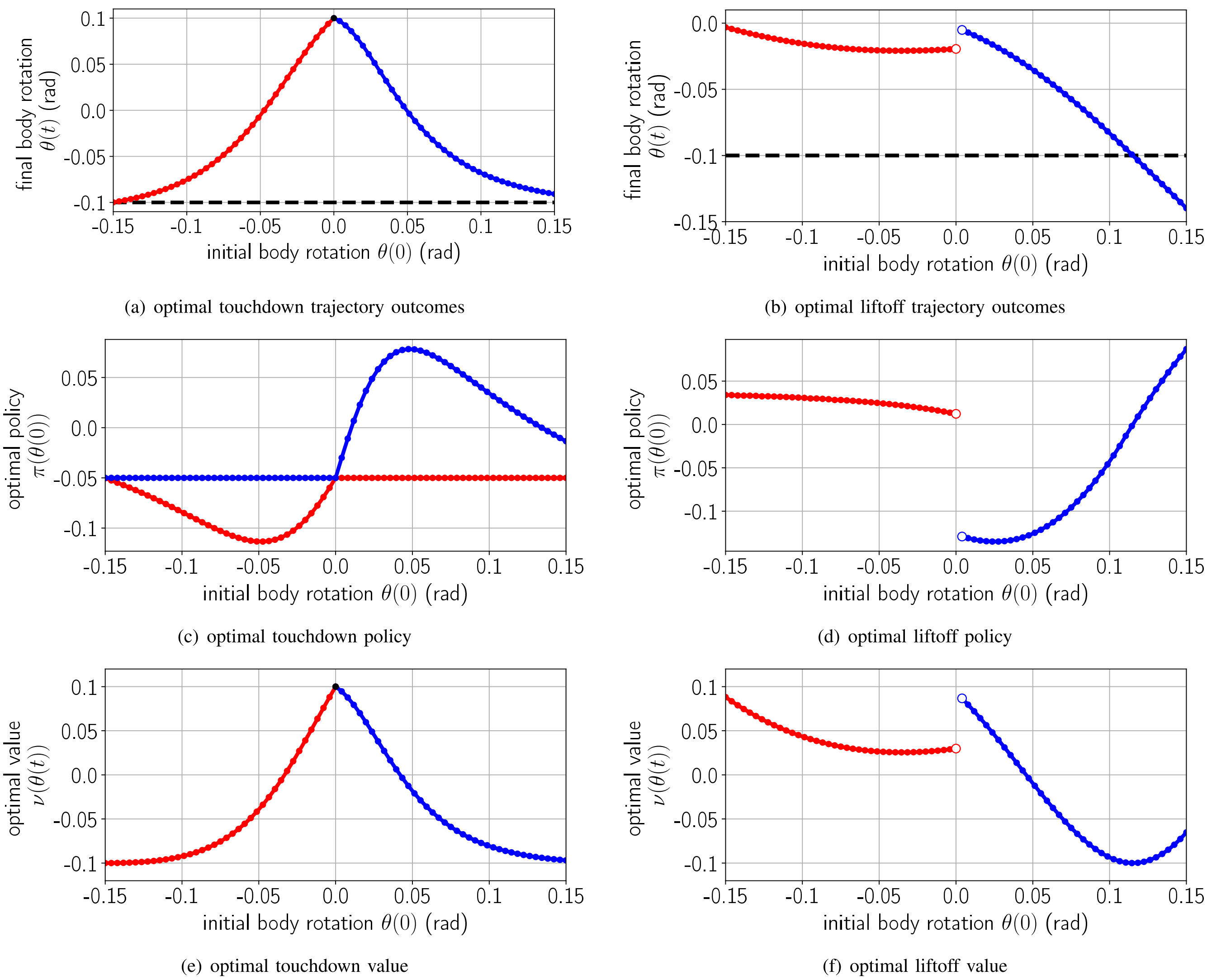
The results in Sections IV-A and IV-B suggest that we should generally expect continuity and differentiability properties of optimal value and policy functions to match that of the cost function: they should be discontinuous when the cost is discontinuous, or piecewise-differentiable when the cost is piecewise-differentiable. In Section V we demonstrate these effects in the class of models specified in Section III.

V. Nonsmooth Optimal Value & Policy Functions in Contact-Rich Robot Dynamics

We showed in the previous two sections that optimal value and policy functions inherit nonsmoothness from the underlying dynamics. To instantiate this result, we crafted a simple mechanical system subject to unilateral constraints that exhibits piecewise-differentiable and discontinuous trajectory outcomes, yielding *touchdown* and *liftoff* maneuvers described in Sections V-B and V-C and whose results are shown in Fig. 1(a,b) for the system described in Section V-A.

A. Mechanical System Subject to Unilateral Constraints

For the biped illustrated in Fig. 1: the configuration space has $d = 7$ degrees-of-freedom $q \in Q = \mathbb{R}^7$ corresponding to the horizonal and vertical positions of the body and two feet and rotation θ of the body; the constraint function $a{:}Q \to \mathbb{R}^2$ evaluates the height of the two feet above the ground; the mass and moment of inertia of the body are m, I and the mass of each foot is m'; the mass matrix $M(q)$ is diagonal and constant. Linear spring-dampers with stiffness and damping constants κ, β in parallel with prismatic actuators connect the



(a) optimal touchdown trajectory outcomes

(b) optimal liftoff trajectory outcomes

(c) optimal touchdown policy

(d) optimal liftoff policy

(e) optimal touchdown value

(f) optimal liftoff value

Fig. 4. *Optimal trajectories, values, and policies for touchdown and liftoff maneuvers.* Optimizing (14), (15) for the biped in Fig. 1 yields trajectory outcomes (a,b), policies (c,d), and values (e,f) that are nonsmooth – piecewise-differentiable (*left*) or discontinuous (*right*). Asymmetries in trajectory outcomes are due to unequal input penalty parameters ($\alpha_1 \neq \alpha_2$) in (a) and unequal leg forces ($u_1 \neq u_2$) in (b).

feet to pin joints arranged symmetrically at a distance ℓ from the body's center-of-mass. Impact of the feet with the ground is *plastic* (*purely inelastic*).

B. Touchdown Maneuver

For the touchdown maneuver, we seek the optimal (constant) force to exert in the left leg (u_1) when the left foot is in contact and the right foot is not; similarly, we seek the optimal choice of force in the right leg (u_2) when the right foot is in contact and the left foot is not: with θ^* denoting the desired body rotation at nadir and $\alpha_1, \alpha_2 > 0$ denoting input penalty parameters,

$$c_{\text{touchdown}}(\theta, u_1, u_2) = (\theta - \theta^*)^2 + \alpha_1 u_1^2 + \alpha_2 u_2^2. \quad (14)$$

C. Liftoff Maneuver

For the liftoff maneuver, we seek the optimal (constant) torque (u_{12}) to apply to the body while both feet are in contact: with θ^* denoting the desired body rotation at apex and $\alpha_{12} > 0$ denoting an input penalty parameter,

$$c_{\text{liftoff}}(\theta, u_{12}) = (\theta - \theta^*)^2 + \alpha_{12} u_{12}^2. \quad (15)$$

D. Numerical Results

We implemented numerical simulations[2] of the model from Section V-A and applied a scalar minimization algorithm[3] to compute optimal policies as a function of initial body rotation. As expected, the optimal value and policy functions computed for the touchdown and liftoff maneuvers are nonsmooth (Fig. 4(c,d,e,f)), owing to the nonsmoothness of the optimal trajectory outcomes (Fig. 4(a,b)). This result does not depend sensitively on the problem data; nonsmoothness is preserved after altering parameters of the model and/or cost functions. We emphasize that the nonsmoothness in Fig. 4 arises from the nonsmoothness in the underlying system dynamics (4), as the functions in (14) and (15) are smooth.

[2]using the modeling framework in [3] and simulation algorithm in [24].
[3]SciPy v0.19.0 minimize_scalar.

VI. Discussion

We conclude by discussing our focus on local rather than global sources of nonsmoothness (Section VI-A), implications of our results for trajectory optimization and reinforcement learning (Section VI-B), and generalization of our results to other behaviors and models (Section VI-C).

A. Local vs. Global Sources of Nonsmoothness

We emphasize that our results focus on *local* sources of nonsmoothness arising from zero- or first-order properties (i.e., discontinuity or piecewise-differentiability) of trajectory outcomes in contact-rich robot dynamics. Of course, such mechanical systems can also exhibit *global* sources of nonsmoothness arising from topological properties of trajectory outcomes – we refer to [25, Ch. 2] for an excellent discussion of the global perspective in feedback control.

B. Trajectory Optimization and Reinforcement Learning

Our focus on local properties of contact-rich robot dynamics is motivated by recent work that applies gradient-based algorithms in an attempt to approximate optimal value and policy functions in mechanical systems subject to unilateral constraints [6]–[13]. Such approaches presume a (possibly non-optimal) policy $\pi : \mathcal{X} \to \mathcal{U}$ has an associated value $\nu^\pi : \mathcal{X} \to \mathbb{R}$ that admits a gradient with respect to π, in which case it is natural to improve the policy by descending the cost landscape: with $\alpha > 0$ as a stepsize parameter,

$$\pi^+ = \pi + \alpha \arg \min_{\|\delta\|=1} D_\pi \nu^\pi(\delta). \tag{16}$$

The update in (16) can be interpreted as a *gradient descent* algorithm for trajectory optimization [4, Ch. 4] or as a *direct policy gradient-based* algorithm for reinforcement learning [26]. Since our results demonstrate that classical gradients fail to exist, more work is needed to justify the use of these smooth techniques in this nonsmooth setting.

C. Generalization to Other Behaviors and Models

The nonsmoothness presented in Section V manifested along trajectories that underwent simultaneous constraint (de)activations. Although such sources of nonsmoothness are confined to a subset of the state space with zero Lebesgue measure, we note that some important behaviors will reside near a large number of pieces. For instance, periodic behaviors with (near-)simultaneous (de)activation of $n \in \mathbb{N}$ constraints as in [27] could yield up to $(n!)^k$ pieces after $k \in \mathbb{N}$ periods [28, Ch. 6]. The combinatorics are similar for tasks that involve intermittently activating (a subset of) n constraints k times as in [8]. Including dry friction in the model can only increase the prevalence of nonsmoothness in trajectory outcomes [29] and, hence, optimal value and policy functions.

Acknowledgment

The authors thank S. D. Coogan and the anonymous reviewers for helpful feedback on earlier versions of this letter.

References

[1] P. Ballard, "The dynamics of discrete mechanical systems with perfect unilateral constraints," *Archive Ration. Mech. Anal.*, vol. 154, no. 3, pp. 199–274, 2000.

[2] B. Brogliato, *Nonsmooth Mechanics: Models, Dynamics, and Control*. London, U.K.: Springer-Verlag, 1999.

[3] A. M. Johnson, S. A. Burden, and D. E. Koditschek, "A hybrid systems model for simple manipulation and self-manipulation systems," *Int. J. Robot. Res.*, vol. 35, no. 11, pp. 1354–1392, 2016.

[4] E. Polak, *Optimization: Algorithms and Consistent Approximations*. New York, NY, USA: Springer-Verlag, 1997.

[5] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA, USA: Athena Sci., 1996.

[6] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *Int. J. Robot. Res.*, vol. 33, no. 1, pp. 69–81, 2014.

[7] A. Patel, S. L. Shield, S. Kazi, A. M. Johnson, and L. T. Biegler, "Contact-implicit trajectory optimization using orthogonal collocation," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 2242–2249, Apr. 2019.

[8] I. Mordatch, E. Todorov, and Z. Popović, "Discovery of complex behaviors through contact-invariant optimization," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, 2012.

[9] A. Hereid, C. M. Hubicki, E. A. Cousineau, and A. D. Ames, "Dynamic humanoid locomotion: A scalable formulation for HZD gait optimization," *IEEE Trans. Robot.*, vol. 34, no. 2, pp. 370–387, Apr. 2018.

[10] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, Jul. 2015, pp. 1889–1897. [Online]. Available: http://proceedings.mlr.press/v37/schulman15.html

[11] N. Heess *et al.*, "Emergence of locomotion behaviours in rich environments," *arXiv:1707.02286*. [Online]. Available: https://arxiv.org/abs/1707.02286

[12] OpenAI *et al.*, "Learning dexterous in-hand manipulation," *arXiv:1808.00177*. [Online]. Available: http://arxiv.org/abs/1808.00177

[13] J. Hwangbo *et al.*, "Learning agile and dynamic motor skills for legged robots," *Sci. Robot.*, vol. 4, no. 26, 2019, Art. no. eaau5872.

[14] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 229–256, 1992.

[15] S. Scholtes, *Introduction to Piecewise Differentiable Equations*. New York, NY, USA: Springer-Verlag, 2012.

[16] S. M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: Stability and sensitivity," *Nonlin. Anal. Optim.*, pp. 45–66, 1987.

[17] R. T. Rockafellar, "A property of piecewise smooth functions," *Comput. Optim. Appl.*, vol. 25, pp. 247–250, Apr. 2003.

[18] A. M. Pace and S. A. Burden, "Piecewise—Differentiable trajectory outcomes in mechanical systems subject to unilateral constraints," in *Proc. 20th Int. Conf. Hybrid Syst. Comput. Control (HSSC)*, Apr. 2017, pp. 243–252. [Online]. Available: http://doi.acm.org/10.1145/3049797.3049807

[19] A. M. Pace and S. A. Burden, "Decoupled limbs yield differentiable trajectory outcomes through intermittent contact in locomotion and manipulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 2261–2266.

[20] C. D. Remy, K. Buffinton, and R. Siegwart, "Stability analysis of passive dynamic walking of quadrupeds," *Int. J. Robot. Res.*, vol. 29, no. 9, pp. 1173–1185, 2010.

[21] Y. Hürmüzlü and D. B. Marghitu, "Rigid body collisions of planar kinematic chains with multiple contact points," *Int. J. Robot. Res.*, vol. 13, no. 1, pp. 82–92, 1994.

[22] R. W. Chaney, "Second-order sufficient conditions in nonsmooth optimization," *Math. Oper. Res.*, vol. 13, no. 4, pp. 660–673, 1988.

[23] S. M. Robinson, "An implicit-function theorem for a class of nonsmooth functions," *Math. Oper. Res.*, vol. 16, no. 2, pp. 292–309, 1991.

[24] S. Burden, H. Gonzalez, R. Vasudevan, R. Bajcsy, and S. Sastry, "Metrization and simulation of hybrid control systems," *IEEE Trans. Autom. Control*, vol. 60, no. 9, pp. 2307–2320, Feb. 2015.

[25] C. G. Mayhew, "Hybrid control for topologically constrained systems," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. California at Santa Barbara, Santa Barbara, CA, USA, 2010. [Online]. Available: https://search.proquest.com/docview/851557337?accountid=14784

[26] J. Baxter and P. L. Bartlett, "Infinite-horizon policy-gradient estimation," *J. Artif. Intell. Res.*, vol. 15, no. 1, pp. 319–350, 2001.

[27] R. M. Alexander, "The gaits of bipedal and quadrupedal animals," *Int. J. Robot. Res.*, vol. 3, no. 2, pp. 49–59, 1984.

[28] B. S. Banjanin, "Data-driven modeling for hybrid dynamical systems," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Washington, Seattle, WA, USA, 2019. [Online]. Available: https://digital.lib.washington.edu/researchworks/handle/1773/44689?show=full

[29] M. D. Bernardo *et al.*, "Bifurcations in nonsmooth dynamical systems," *SIAM Rev.*, vol. 50, no. 4, pp. 629–701, 2008.