Bundled Gradients through Contact via Randomized Smoothing

H.J. Terry Suh[1*], Tao Pang[1*], and Russ Tedrake[1]

***Abstract*— The empirical success of derivative-free methods in reinforcement learning for planning through contact seems at odds with the perceived fragility of classical gradient-based optimization methods in these domains. What is causing this gap, and how might we use the answer to improve gradient-based methods? We believe a stochastic formulation of dynamics is one crucial ingredient. We use tools from randomized smoothing to analyze sampling-based approximations of the gradient, and formalize such approximations through the gradient bundle. We show that using the gradient bundle in lieu of the gradient mitigates fast-changing gradients of non-smooth contact dynamics modeled by the implicit time-stepping, or the penalty method. Finally, we apply the gradient bundle to optimal control using iLQR, introducing a novel algorithm which improves convergence over using exact gradients. Combining our algorithm with a convex implicit time-stepping formulation of contact, we show that we can tractably tackle planning-through-contact problems in manipulation.**

I. Introduction

Robots interact with their environments primarily through contact, and planning of such systems (colloquially named *planning through contact*) is a long-standing challenge in manipulation and locomotion, especially in the absence of fixed mode sequences. Neither the tightly coupled problems of *modeling* the contact dynamics nor the *planning* that uses the prescribed model have straightforward answers.

A large body of previous work in planning through contact [1]–[5] has been based on gradient-based optimizers. Combining different contact models, optimization methods for planning, and a variety of relaxation schemes, these methods have shown reasonable success. However, they still tend to remain brittle: they are sensitive to hyperparameters, and the quality of initial guesses becomes a make-or-break factor. When they fail, they can be notoriously hard to debug.

In another direction, recent works in Reinforcement Learning (RL) have shown impressive results in tackling problems with contact dynamics [6]–[8]. The quality of the solutions obtained via RL is surprising, despite the fact that no underlying structure was used. This suggests that improvements can be made in time and sample complexity. More importantly; however, the fact that we do not understand the gap between the success of RL methods and the struggle of more classical gradient-based methods in planning through contact, is unsatisfying. What are the key improvements that RL made but the classical methods failed to consider? We believe that the answer lies in how RL fundamentally considers a stochastic formulation, as opposed to gradient-based methods that have largely been deterministic.

*These authors contributed equally to this work.

*This work was funded by NSF Award No.EFRI-1830901, Navy ONR Award No. N00014-18-1-2210, and Lincoln Lab PO No. 7000470769.

[1]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA {hjsuh, pangtao, russt}@mit.edu

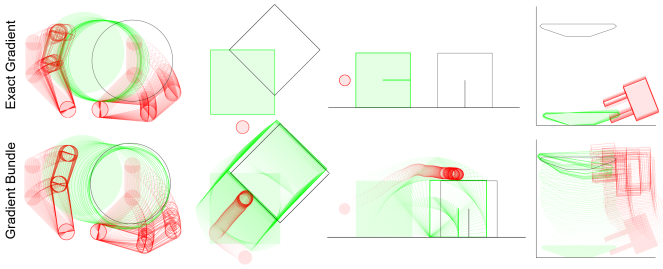

Fig. 1: Planning with exact gradients (top) vs. the gradient bundle (bottom). Actuated objects drawn in red, unactuated objects drawn in green, and the goal drawn in black. From left: planar hand, planar pushing, box pivoting, and plate pickup task ([9]).

If this is true, what would it mean to take a probabilistic formulation of contact dynamics, and how would it help the planning problem? Can we learn from the answers and combine stochasticity with gradient-based approaches to obtain better convergence behavior? We aim to answer these questions through the lens of *randomized smoothing* [10]–[12]. Results from non-smooth optimization [10] tell us that the fast-changing gradients of non-smooth problems destabilize gradient-based optimizers such as gradient descent [13]. By considering a probabilistic approach to contact, we aim to alleviate flatness of gradients, as well as stabilize fast-changing gradients of contact dynamics to produce a more globally informative direction of improvement.

We formalize gradients taken through randomized smoothing as the *gradient bundle*, and show that it has many benefits over exact gradients of contact models, such as overcoming non-smoothness, alleviating nonconvexity, and supporting a zero-order variant. The gradient bundle can readily replace the exact gradient in optimal control, policy search, state estimation, and system identification. In our work, we apply bundled gradients on optimal control using iterative LQR (iLQR) to address planning-through-contact problems.

In addition, randomized smoothing allows obtaining the gradient bundle of contact dynamics that are defined through implicit optimization problems [4], [14], [15], that are often preferred over penalty methods [16] for its ability to take longer timesteps. To demonstrate this advantage, we utilize our method on a quasi-dynamic simulator that uses an implicit time-stepping scheme [17] and show that we can tractably tackle robotic manipulation problems.

Contributions. We formalize the notion of the gradient bundle by introducing randomized smoothing, which we then apply to penalty and implicit time-stepping models of contact dynamics. We replace the exact gradients of dynamics with the gradient bundle in iLQR, showing that we can more robustly solve manipulation problems.

II. Literature Review

A. Differentiable Modeling of Contact

We mainly address two classes of approaches for modeling contact: the penalty method [16], and the implicit time-stepping method [14], [15]. Although other classes of models exist (e.g. hybrid dynamics simulation through event-detection and impacts [18]), they have yet to scale to simulating the complex geometry and dynamics of manipulation.

Works that utilize the penalty method [3], [19], [20] approximate contact behavior at each timestep with forces from stiff spring and dampers. While the method can approximate a wide range of contact behavior such as impact, the dynamics often results in a stiff ODE, which requires small timesteps to approximate well. In addition, although penalty methods often have straightforward expressions for the derivative [3], the computation of long-horizon gradients may be expensive due to large number of timesteps required.

On the other hand, implicit time-stepping methods [1], [4], [21] can take larger timesteps without sacrificing integration accuracy, but integration requires solving a Linear Complementarity Problem (LCP) [22], whose per-step computation is harder and solutions may not be unique. Nevertheless, the LCP can be converted to a convex Quadratic Program (QP) by relaxing the exact Coulomb friction constraints [17], [23], [24]. The implicit models also allow differentiation by computing the derivatives of the KKT conditions of the QP using the implicit function theorem [25].

B. Smooth Approximations of Contact for Planning

As solving the non-smooth problem directly (e.g. via mixed-integer programming [5], [26]) is difficult, many existing works consider solving a smooth approximation of the problem. For instance, penalty-based methods make explicit smooth approximations of the contact dynamics [3], [19], [20]. While the behavior of such approximations have intuitive interpretations at a dynamics-level, the choice of functions that are used is arbitrary. In addition, it is not clear how to apply such explicit smoothing to implicit dynamics whose result is defined by an optimization problem.

Other works [1], [2], [21], [24], [27], [28] do not directly smooth out the dynamics, but find an explicit relaxation of constraints that implicitly define the dynamics. While these works are often better connected to the underlying optimization, the interpretation of the resulting dynamics can be unintuitive which makes debugging cumbersome.

Our philosophy behind why stochasticity is necessary for planning is similar to the argument presented in [28]. However, unlike [28] which explicitly uses "rounded" complementarity constraints and interprets it as minimizing expected constraint violation, randomized smoothing is fully implicit [11], [12], in the sense that we do not require explicit smoothing formulas for dynamics, nor the constraints that define them. This allows us to make smooth approximations to a more general class of models. In addition, the bundled dynamics from randomized smoothing is more physically-interpretable than numerical smoothing strategies that "sand off" the sharp corners of non-smooth constraints.

C. Randomized Smoothing for Non-smooth Optimization

The destabilization of gradient descent for non-smooth functions was observed by early works such as Wolfe's example [13], [29]. Stabilized gradient descent through *gradient sampling* was introduced in [10], [30], [31], which proposed to approximate an instance of a subgradient in the convex hull of sampled gradients around the current iterate. Based on these works, an optimal convergence guarantee was provided in [11] where the term *randomized smoothing* was introduced, and a zero-order variant was presented in [12].

Our work strongly builds upon existing works, but we analyze the consequences of randomized smoothing in the presence of contacts and aim to use the gradient bundle for optimal control. In addition, we show potential pitfalls of gradient sampling in the presence of dynamics of discontinuities, which has not been addressed by previous works.

III. Randomized Smoothing

Motivated by [10], [11], we first introduce the idea of randomized smoothing in the context of gradient descent.

A. The Gradient Bundle and the Bundled Objective

Let $f(x)$ be a non-smooth objective function that is differentiable a.e., and denote the gradient as $\nabla f(x)$. While the gradient suggests a local direction of descent, it is myopic. For smooth non-convex functions, gradient descent could lead to bad local minima. For non-smooth functions, the gradient may be subject to large jumps, stalling the descent on the cost function [13], [29] and compromising standard guarantees of convergence. To alleviate the problems caused by the locality of the gradient, we can aggregate gradient information in some neighborhood of the current x, which we formalize with the term *gradient bundle*.

Definition 1. Gradient Bundle Let $\mu \in L^1$ be a symmetric probability density. Then, the *first-order gradient bundle* is defined as

$$\bar{\nabla} f(x) := \int \nabla f(x+w)\mu(w)dw = \mathbb{E}_{w\sim\mu}[\nabla f(x+w)]. \quad (1)$$

The gradient bundle can be thought of as a convolution of $\nabla f(x)$ and the kernel μ. Further insight of descent using the gradient bundle can be obtained by defining the *bundled objective*.

Definition 2. The Bundled Objective is defined as

$$\bar{f}(x) := \int f(x+w)\mu(w)dw = \mathbb{E}_{w\sim u}[f(x+w)]. \quad (2)$$

Lemma 1. The first-order gradient bundle $\bar{\nabla} f(x)$ is the exact gradient of the bundled objective $\bar{f}(x)$ ($\bar{\nabla} f(x) = \nabla \bar{f}(x)$)

Proof. Exchange the expectation $\mathbb{E}_w$ and the derivative ∇_x using the Dominated Convergence Theorem. □

This suggests that using the bundled gradient in place of the exact one lets us implicitly operate on a different objective function defined by the bundled objective. The bundled objective has an intuitive explanation as result of filtering via convolution, such as Gaussian smoothing. Such

smoothing schemes alleviate non-convexity by smoothing out local minima, and allows tackling non-smooth problems by stabilizing gradient descent [10]. Note that the gradient bundle is purposely biased to produce more global effects, and thus will not converge to a stationary point of the original objective, but rather that of the bundled objective. This can be mitigated by iteratively decreasing the variance $\sigma(\mu)$, as done in typical stochastic approximation schemes [32], [33].

B. Zero-Order Variation

If ∇f is not available or costly to compute, one may instead compute an expected value of finite-differences (which we denote by $\nabla_0 f(x)$) to use as a direction of descent [12]. We call this object the *zero-order gradient bundle*. Convergence of such schemes are well analyzed in Simultaneous Perturbation Stochastic Approximation (SPSA) [34].

$$\bar{\nabla}_0 f(x) := \mathbb{E}_{w\sim\mu}[\nabla_0 f(x)] = \mathbb{E}_{w\sim\mu}\left[\frac{f(x+w)-f(x)}{w}\right]. \quad (3)$$

Example 1. Function with many Local Minina: Let

$$f(x) = x^2 + 0.1\sin(20x), \quad (4)$$

which has many bad local minima. Under a suitable choice for the Gaussian kernel $\mu(w) = \mathcal{N}(w; 0, \sigma)$, this function can be smoothed out to be convex, or gradient dominant [35].

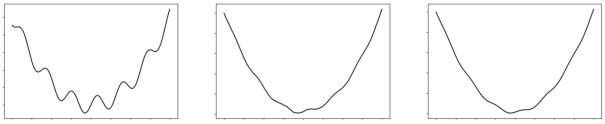

Fig. 2: Left: Objective $f(x)$. Center: Bundled objective reconstructed from first-order gradient bundle. Right: Bundled objective reconstructed from zero-order gradient bundle

C. Monte-Carlo Integration for the Gradient Bundle

Symbolic computation of the gradient bundle is prohibitive, and quadrature methods do not scale well with dimension. Thus, we propose use Monte-Carlo integration, which can scale to high dimensions under suitable conditions on the variance of the objective. This yields the gradient sampling algorithm [10] and its zero order variant [12],

$$\begin{aligned}\hat{\nabla} f(x) &\approx \frac{1}{N}\sum_i \nabla f(x+w_i) \quad w_i \sim \mu(w)\\ \hat{\nabla}_0 f(x) &\approx \frac{1}{N}\sum_i \left(\frac{f(x+w_i)-f(x)}{w_i}\right) \quad w_i \sim \mu(w).\end{aligned} \quad (5)$$

Lemma 2. Law of Large Numbers. If f has no discontinuous jumps and $\nabla f, \nabla f_0 \in L^1$, then $\hat{\nabla} f(x) \xrightarrow{a.s.} \bar{\nabla} f(x)$ and $\hat{\nabla}_0 f(x) \xrightarrow{a.s} \bar{\nabla}_0 f(x)$

However, one must be very careful with this Monte Carlo approximation when f has discontinuous jumps, due to the dirac-delta that appears in the derivative ∇f [36]. We formalize this through the following theorem:

Theorem 1. Sampling Gradients of Discontinuities. Let $f(x)$ be a function differentiable a.e. with at least one jump discontinuity; $\exists x'$ s.t. $\lim_{x\to x'^-} f(x) \neq \lim_{x\leftarrow x'^+} f(x)$. Then, the Monte-Carlo estimate of the first-order gradient bundle almost surely fails to converge to the gradient of the bundled dynamics, $\bar{\nabla} f(x) \neq^{\text{a.s.}} \nabla \bar{f}(x)$.

Proof. Sampling from the dirac-delta is ill-defined, as it does not evaluate pointwise. In practice, however, such an attempt would always lead to 0 almost surely. Without loss of generality, let $x' = 0$. Then, $\exists f_L, f_R$ s.t.

$$\begin{aligned} f(x) &= f_R(x)\mathbb{1}_{x\geq 0} + f_L(x)\mathbb{1}_{x<0}\\ \nabla f(x) &= \nabla f_R(x)\mathbb{1}_{x>0} + \nabla f_L(x)\mathbb{1}_{x<0} + c\delta(x).\end{aligned} \quad (6)$$

where $c = f_R(0) - f_L(0) \neq 0$. Then,

$$\begin{aligned}\nabla \bar{f}(x) &= \int \nabla f(x+w)\mu(w)dw\\ &= c\mu(x) + \int \big(\nabla f(x+w) - c\delta(x+w)\big)\mu(w)dw.\end{aligned} \quad (7)$$

Given that samples drawn from $w_i \sim \mu(w)$ have zero probability of satisfying $x + w_i$,

$$\begin{aligned}\hat{\nabla} f(x) &\xrightarrow{a.s.} \int (\nabla f(x+w) - c\delta(x+w))\mu(w)dw,\\ \nabla \bar{f}(x) - \hat{\nabla} f(x) &\xrightarrow{a.s.} c\mu(x).\end{aligned} \quad (8)$$

which completes the proof. □

On the other hand, the zero-order variant does not suffer from this by using finite intervals for approximation.We illustrate an extreme case of Theorem 1 in Example 2.

Example 2. Heaviside Function: Let

$$f(x) = H(x) = \{1 \text{ if } x \geq 0, 0 \text{ if } x < 0\}. \quad (9)$$

Then, $\nabla f(x) = \delta(x)$, and the approximated gradient bundle $\bar{\nabla} f(x + w_i) = 0$ almost surely. However, the analytical computation of the bundled objective with a Gaussian μ would result in $\bar{f}(x) = \text{erf}(x)$, whose derivative is non-zero at $x = 0$. Thus $\hat{\nabla} f(x) \neq \nabla \bar{f}(x)$.

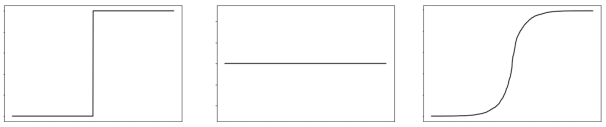

Fig. 3: Left: Objective $f(x)$. Center: Bundled objective reconstructed from first-order gradient bundle. Right: Bundled objective reconstructed from zero-order gradient bundle

Finally, first-order gradient sampling may compromise the quality of gradient bundle for non-smooth functions (even without discontinuities) in the regime of small samples.

Example 3. Quantization of Gradients: Let

$$f(x) = \{-1 + x \text{ if } x \geq 0, 1 - x \text{ if } x < 0\} \quad (10)$$

Then, $\nabla f(x) = 1$ for $x \geq 0$, and -1 for $x < 0$. With N samples, the approximated gradient bundle will be quantized and can only take a finite set of values.

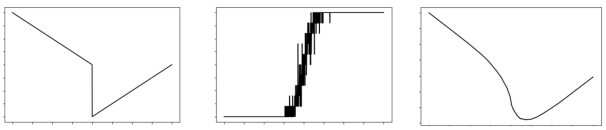

Fig. 4: Left: Original objective $f(x)$. Center: Value of the gradient bundle using gradient sampling. Right: Bundled objective reconstructed from the zero-order gradient bundle.

D. Jacobian Bundle and Bundled Dynamics

We extend the concept of randomized smoothing to dynamical systems with the *Jacobian bundle* and the *bundled dynamics*. Let f be a discrete-time dynamical system with state $x_t \in \mathbb{R}^n$, and input $u_t \in \mathbb{R}^m$,

$$x_{t+1} = f(x_t, u_t). \tag{11}$$

Recall that the linearization of the dynamical system f around the nominal trajectory $\{\bar{x}_t, \bar{u}_t\}_{t=0}^T$ can be formed by taking a first-order Taylor expansion:

$$\begin{aligned} x_{t+1} &= \underbrace{\left.\frac{\partial f}{\partial x}\right|_{x=\bar{x}_t, u=\bar{u}_t}}_{\mathbf{A}_t} x_t + \underbrace{\left.\frac{\partial f}{\partial u}\right|_{x=\bar{x}_t, u=\bar{u}_t}}_{\mathbf{B}_t} u_t + c_t, \\ c_t &= f(\bar{x}_t, \bar{u}_t) - \mathbf{A}_t \bar{x}_t - \mathbf{B}_t \bar{u}_t, \end{aligned} \tag{12}$$

While the linearization provides a local approximation, it suffers the same problem from the gradient in a sense that it is myopic. We apply randomized smoothing to the Jacobian of the linearized system, and aim to provide a more globally informative linear approximation to the system.

Definition 3. The Jacobian Bundle Let $\mu \in L^1$ be a multivariate symmetric probability density with arguments $w \in \mathbb{R}^m, v \in \mathbb{R}^n$. Then, the Jacobian bundle is defined as a convolution between $\mu(w, v)$ and the Jacobian of f:

$$\bar{\mathbf{A}}_t = \mathbb{E}_{w,v}\left[\left.\frac{\partial f}{\partial x}\right|_{(\bar{x}_t+w, \bar{u}_t+v)}\right], \quad \bar{\mathbf{B}}_t = \mathbb{E}_{w,v}\left[\left.\frac{\partial f}{\partial u}\right|_{(\bar{x}_t+w, \bar{u}_t+v)}\right]. \tag{13}$$

Definition 4. The Bundled Dynamics The Bundled dynamics are defined as a multivariate convolution between the original dynamics and the probability density μ:

$$\bar{f}(x_t, u_t) = \mathbb{E}_{w,v}\Big[f(x_t + w, u_t + v)\Big]. \tag{14}$$

Lemma 3. The linear system described by the Jacobian bundle $\bar{\mathbf{A}}_t, \bar{\mathbf{B}}_t$ is the exact linearization of the bundled dynamics $\bar{f}(x_t, u_t)$.

Proof. Identical to Lemma 1. □

The above theorem establishes that using the bundled Jacobian, as opposed to the exact linearization of the system, is equivalent to implicitly operating on a smoother version of the dynamics that has been convolved with μ.

E. Monte-Carlo Approximation and Zero-Order Variant

We propose to approximate the Jacobian bundle using Monte-carlo integration, where $w_i, v_i \sim \mu(w, v)$:

$$\hat{\mathbf{A}}_t \approx \frac{1}{N}\sum_{i=1}^{N}\left(\left.\frac{\partial f}{\partial x}\right|_{\bar{x}_t+w_i, \bar{u}_t+v_i}\right), \quad \hat{\mathbf{B}}_t \approx \frac{1}{N}\sum_{i=1}^{N}\left(\left.\frac{\partial f}{\partial u}\right|_{\bar{x}_t+w_i, \bar{u}_t+v_i}\right). \tag{15}$$

Finally, we introduce a zero-order variant that takes a least-squares approximation to the local dynamics, which can be useful when the Jacobians are impractical to compute:

$$\hat{\mathbf{A}}_t^0, \hat{\mathbf{B}}_t^0 = \operatorname{argmin}_{\mathbf{A},\mathbf{B}} \sum_{i=1}^{N}\left\| f(\bar{x}_t + w_i, \bar{u}_t + v_i) - f(\bar{x}_t, \bar{u}_t) - \mathbf{A}w_i - \mathbf{B}v_i \right\|_2^2. \tag{16}$$

IV. Randomized Smoothing of Contact Dynamics

What does it mean to apply randomized smoothing in the presence of contacts? Intuitively, different samples will encounter different contact modes. In expectation, the effect of sampling is a smoothed behavior where normal and frictional forces are applied at a distance, and the boundaries of stick and slip blur. We analyze this intuition through two classes of contact models: implicit contact defined by complementarity constraints [15], [37], and the penalty method [16].

A. Contacts Defined by Complementarity Constraints

Implicit models of contact using complementarity constraints [14], [15] are widely used in planning through contact for their ability to stably take bigger time steps [1], [38]. However, unlike penalty methods which permit explicit smoothing of forces [3], [19], [20], smoothing results of implicit optimization problems is less straightforward; randomized smoothing is unique in that it provides the ability to do so without an explicit relaxation of constraints.

An accurate treatment of multibody dynamics with contact and friction in its full glory requires heavy notation and detracts from the intuition behind the advantage of smoothing. We instead demonstrate the effect of smoothing on contact dynamics through two simple, 2D examples, and refer the reader to [15], [17], [23], [37] for a more detailed presentation of contact dynamics.

To simplify the dynamics aspect of the modeling and focus more on contact, we keep the full second-order dynamics for un-actuated objects, but make the assumption that the actuated bodies (robots) are gravity-compensated, PD-controlled and *quasi-static* [17], [39]: the robots are always in force equilibrium, with the generalized forces exerted by contact balanced by the virtual spring of the proportional part of the PD controller. The quasi-static assumption reduces the dimension of robot states by half, and is commonly made in systems where inertial forces are negligible compared to contact forces, such as robotic manipulation [5], [40]. Despite these simplifications, our analysis on the effect of contact smoothing does transfer to the full second-order dynamics of the entire robot-object system.

Example 4. Randomized Smoothing of Normal Contact. We first consider a 1D system of two objects in Fig. 5a, where the boxes can only interact through normal contact forces. The dynamics of the robot (red box) is described by

$$-\lambda_n + hk(\tilde{x}_{t+1}^a - x_{t+1}^a) = 0, \tag{17}$$

where λ_n denotes the impulse generated by the normal force during the timestep h; $\tilde{x}_{t+1}^a$ and x_{t+1}^a are the robot's *commanded* and actual positions at the next timestep $t+1$. Constraint (17) states that the contact force $-\lambda_n/h$ is balanced by the virtual spring force $k(\tilde{x}_{t+1}^a - x_{t+1}^a)$ at $t+1$.

The dynamics of the object (green box) is given by

$$mv_{t+1}^u = \lambda_n, \tag{18a}$$

$$x_{t+1}^u = x_t^u + hv_{t+1}^u, \tag{18b}$$

where (18a) states that the gain in object momentum comes from the normal contact impulse λ_n, assuming that the robot has 0 velocity at the current timestep.

The contact impulse and the distance between the two objects satisfy the following complementarity constraint:

$$0 \leq \lambda_n \perp (x^u_{t+1} - x^a_{t+1}) \geq 0, \tag{19}$$

where $a \perp b$ for vectors a and b means that $a^\intercal b = 0$. The implication of (19) is three-fold: (**i**) the impulse can only push ($\lambda_n \geq 0$); (**ii**) the two boxes cannot penetrate at $t+1$ ($x^u_{t+1} - x^a_{t+1} \geq 0$); and (**iii**) non-zero normal impulse exists only if the two objects are in contact ($\lambda_n \perp x^u_{t+1} - x^a_{t+1}$).

The state of the 1D system at $t+1$ can be solved from the implicit contact dynamics (17) - (19). The dynamics consists of two linear pieces, depending on whether the two objects make contact at $t+1$:

$$\begin{bmatrix} x^u_{t+1} \\ x^a_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x^u_t \\ x^a_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} [\tilde{x}^a_{t+1}] \text{ (no contact)}, \tag{20a}$$

$$\begin{bmatrix} x^u_{t+1} \\ x^a_{t+1} \end{bmatrix} = \begin{bmatrix} \frac{c}{1+c} & 0 \\ \frac{c}{1+c} & 0 \end{bmatrix} \begin{bmatrix} x^u_t \\ x^a_t \end{bmatrix} + \begin{bmatrix} \frac{1}{1+c} \\ \frac{1}{1+c} \end{bmatrix} [\tilde{x}^a_{t+1}] \text{ (contact)}, \tag{20b}$$

where $c = m/(h^2 k)$. The explicit dynamics (20), which has the same form as (12), results in perfect position command being achieved for x^a_{t+1} with no movement for x^u_{t+1} if there is no contact; and $x^a_{t+1} = x^u_{t+1}$ being placed near the commanded position if there is contact.

As shown in Fig. 5b, before making contact, the control input $\tilde{x}^a$ has no effect on x^u, producing zero gradients. In contrast, when sampling around $\tilde{x}^a$, some of the samples will hit the un-actuated box, and in expectation, push them away from each other. Thus, the bundled dynamics give information that $\tilde{x}^a$ should approach x^u to push the box.

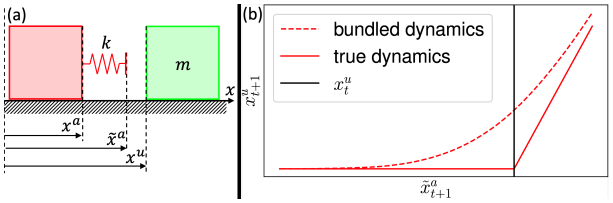

Fig. 5: (**a**) A 1D dynamical system consisting of a 1-DOF robot (red box) and an un-actuated object (green box). Both boxes are constrained to slide on a friction surface. (**b**) The piecewise-linear relationship between the free object position x^u and the commanded robot position $\tilde{x}^a$. Note that there is no contact unless $\tilde{x}^a_{t+1} \geq x^u_t$.

Example 5. Randomized Smoothing of Friction. In this example, we study the 2D system in Fig. 6 where the robot (red sphere) interacts with the green box through a frictional contact on the right of the box.

The Coulomb friction model constrains contact force to stay inside the friction cone. In addition, when the contact is sliding, the force needs to stay on the friction cone's boundary and in the opposite direction of the relative sliding velocity. Similar to the normal contact constraint (19), Coulomb friction can be modeled with additional complementarity constraints [14].

Assuming the same quasi-static dynamics for the robot, zero initial velocity for the box, and the initial condition $[x^u_t, x^a_t, y^a_t] = [0, 0, 0]$, it can be shown, by solving the dynamics subject to the Coulomb friction constraints, that the box's position at the next time step, x^u_{t+1}, is again a piece-wise linear function of the control input $[\tilde{x}^a_{t+1}, \tilde{y}^a_{t+1}]$. As shown Fig. 7a, the dynamics function consists of four

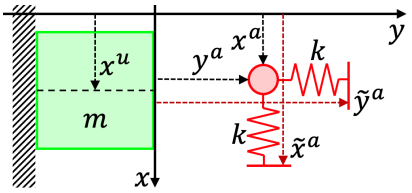

Fig. 6: A 2D dynamical system consisting of a 2-DOF robot (red sphere) and an un-actuated object (green box). The robot can move freely in the xy-plane, whereas the box is constrained to slide on the frictionless surface.

pieces, where each piece corresponds to one contact mode. Note that the exact dynamics is flat for all $\tilde{y}^a_{t+1} > 0$, which provides no information about how the control affects the state. In contrast, the bundled dynamics (Fig. 7b) has non-zero gradient even for $\tilde{y}^a_{t+1} > 0$.

For faster speed and better numerics, multi-body simulators (such as [17], [41]) frequently relax the exact complementarity-based Coulomb friction constraints in [15], [37]. One such relaxation is proposed by Anitescu [23], which allows the forward dynamics to be solved as a quadratic program (QP) instead of a linear complementarity problem (LCP). Using Anitescu's relaxed constraints, the dynamics of x^u_{t+1} with respect to $[\tilde{x}^a_{t+1}, \tilde{y}^a_{t+1}]$ (Fig. 7b) can be shown to have a very similar structure as the exact Coulomb friction contact dynamics in Fig. 7a. In particular, the relaxed dynamics is the same as the exact dynamics when the contact is sticking or separating. In sliding, a "boundary layer" is introduced, in which the ball drags the box even when they are not in contact.

Although relaxed contact models such as [23] sacrifice physical realism, the bundled dynamics of the relaxed contact models behave similar to the bundled dynamics of the exact contact models, as shown in Fig. 7c and 7d. This suggests that algorithms using bundled dynamics for planning can safely capitalize the speed and efficiency of the relaxed contact models despite their mild physical inaccuracy.

B. Penalty-based Contacts

1) Description: In the penalty method [16], the forces in the normal direction are approximated using a stiff spring. Denoting the signed distance between two objects as $\phi(q)$, we can write the normal-direction force with two modes of contact and no-contact.

$$f_n = -k \min\{\phi(q), 0\}. \tag{21}$$

In the penalty approximation of Coulomb friction, sticking is often enforced by means of very viscous damping, while slipping is approximated by a constant multiple of the normal force. Let $\psi(q, \dot{q})$ be the relative tangential velocity between two objects at the point of contact.

$$f_t = \mu f_n \qquad \mu = \begin{cases} c\psi(q,\dot{q}) & \text{if } \psi(q,\dot{q}) \leq \psi_s \\ \mu_d & \text{if } \psi(q,\dot{q}) > \psi_s \end{cases}, \tag{22}$$

where ψ_s controls the threshold of stick and slip. Simple models of Coulomb friction sets ψ_s to make the approximation continuous (i.e. $c\psi_s = \mu_d$), as done in [3]. However, the Stribeck effect [42] tells us that dynamic friction is lesser in magnitude than static friction. Continuous approximations to the discontinuities of Stribeck effect is made in Drake [43].

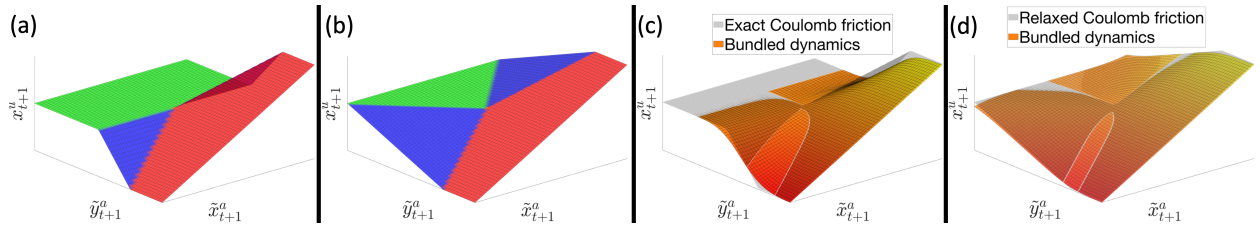

Fig. 7: Implicit contact dynamics using (**a**) the exact Coulomb friction model [37] and (**b**) the relaxed Coulomb friction model [23]. Both (**a**) and (**b**) plot the the green box position x^u_{t+1} as a function of the commanded sphere positions $[\tilde{x}^a_{t+1}, \tilde{y}^a_{t+1}]$. Both dynamics are piecewise-linear, where each piece corresponds to a contact mode: green represents separation; blue sliding and red sticking. The larger sliding region in (**b**) is a result of the "boundary layer" artifact of the relaxed Coulomb friction model. (**c**) and (**d**): bundled dynamics of (**a**) and (**b**), respectively, obtained using randomized smoothing.

2) Randomized Smoothing: We apply randomized smoothing by sampling ϕ and ψ, and plot the results on Fig.8. We note several effects of randomized smoothing: (**i**) in the normal direction, even if the objects are not in contact, repulsive forces are applied at a distance; (**ii**) tangential friction is also applied at a distance due to smoothing that occurs in the direction of ϕ in Fig.8; (**iii**) discontinuous effects from the Stribeck effect is smoothed out.

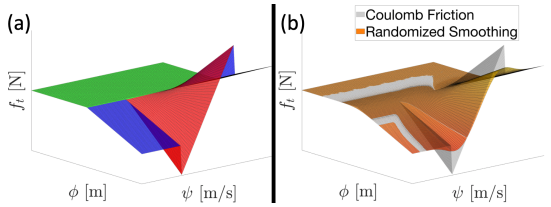

Fig. 8: (**a**) Penalty-based Coulomb friction model with Stribeck effect. Note the discontinuity between the blue (sliding) and red (sticking) pieces. (**b**) Discontinuity in friction is removed by randomized smoothing.

C. Key Insights

1) Problems with Exact Gradients: The non-smoothness of both the implicit time-stepping scheme and the penalty method suggests that the algorithms that rely on the local validity of linearization will suffer. The flatness of the gradients in non-penetrating configurations also suggests no direction of improvement.

2) Relaxing Coulomb friction leads to better planning: Relaxations of the LCP version of Coulomb friction (e.g. Anitescu) allow the application of friction at a distance and coincidentally provide meaningful gradients useful for planning. For instance, the slope of sliding in Fig.7.b suggests that the ball should be closer to the box in the direction of y order to drag the box. This suggests that regularization used in popular simulators such as Mujoco [41] may fundamentally help planning algorithms achieve better performance.

3) Discontinuities of Contact Dynamics: Implicit time-stepping and the penalty method both result in non-smooth, yet *continuous* contact behavior with the exception of the Stribeck effect. The intuition behind the continuity of implicit time-stepping methods lies in the *inelastic* approximation of contact behavior, while the continuity of penalty methods come from considering an infinitesimal section of time.

However, we have yet to conclude the consequences of Theorem 1 to be a minor nuisance. In more complex examples with non-smooth geometries, the normal of the object surface can also change in a discontinuous manner [44], [45]. In contrast, the zero-order gradient bundle can be free from the consequences of Theorem 1, which might suggest that the zero-order nature of popular RL algorithms such as the policy gradient [46] might coincidentally help in overcoming pitfalls of sampling gradients directly.

V. Iterative Randomized Smoothing-LQR

The gradient bundle obtained via randomized smoothing can be used to replace the exact gradient in a variety of optimization problems, suggesting a new family of algorithms. In this work, we choose to replace the exact linearization with the Jacobian bundle in the trajectory optimization algorithm of iterative LQR (iLQR). [47]–[49]. We name our variant iterative Randomized Smoothing-LQR (iRS-LQR). Our resulting variant is similar to SaDDP [50] or Unscented DDP [51], but we uniquely apply it in the presence of contact dynamics with the understanding of how sampling helps.

We first modify iLQR to handle more general constraints by solving a Model Predictive Control (MPC) problem on the locally linear dynamics at every time step, which results in solving a QP of decreasing length at every time step. Such approaches fit well within the computational budget of modern QP solvers [52].

$$\begin{aligned}
&\text{MPC} \quad (\bar{x}_j) = u^*_j, \text{ where} \\
&\min_{x_t, u_t} \quad \left[\|x_T - x^d_T\|^2_{\mathbf{Q}_d} + \sum_{t=j}^{T-1} \left(\|x_t - x^d_t\|^2_{\mathbf{Q}_t} + \|u_t\|^2_{\mathbf{R}_t} \right) \right] \\
&\text{s.t.} \quad x_{t+1} = \mathbf{A}_t x_t + \mathbf{B}_t u_t + c_t \quad \forall t \in \{j, \cdots, T-1\}, \\
&\qquad \mathbf{C}^u_t u_t \le d^u_t \qquad \mathbf{C}^x_t x_t \le d^x_t \quad \forall t \in \{j, \cdots, T\}, \\
&\qquad x_j = \bar{x}_j.
\end{aligned} \tag{23}$$

Here, $\{\mathbf{C}^u_t, d^u_t\}$ are parameters of linear inequalities on state, and $\{\mathbf{C}^x_t, d^x_t\}$ are parameters of linear inequalities on inputs, and $\{x^d_t\}$ represents the desired trajectory. Note that the time indices depend on j; the horizon of the MPC decreases with our progress along the trajectory. Using this subroutine, we roll out the system using MPC at every timestep (whose horizons get shorter at every timestep) and iteratively repeat until a satisfactory solution has been reached (Alg.1).

We introduce a variance scheduling function $\eta(\Sigma_0, k)$ which takes the initial variance Σ_0 and the current iteration number k to produce a reduced variance Σ_k [32], [34]. Convergence to a local minima of the original problem is guaranteed as long as $\sum_k \eta(\Sigma_0, k)^2 < \infty$.

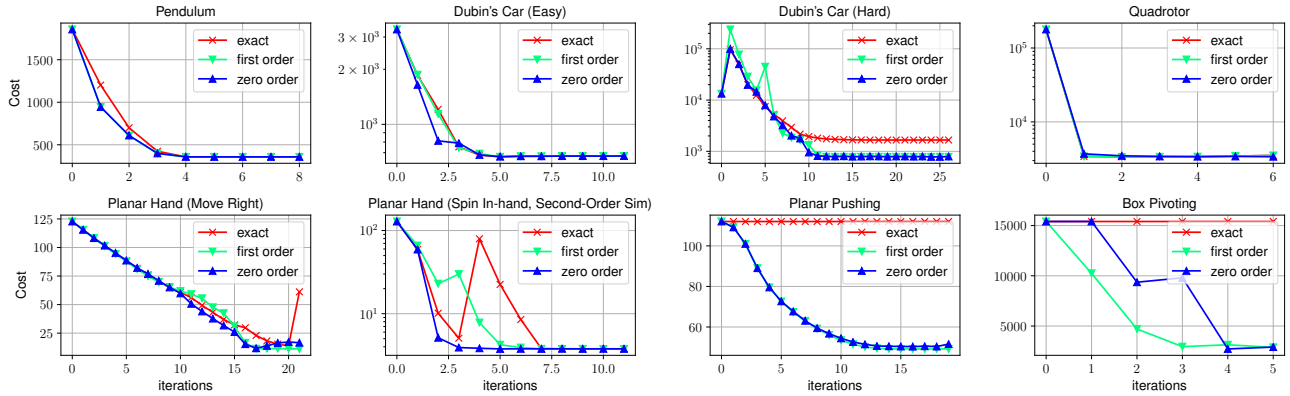

Fig. 9: Results for various systems on which iRS-LQR has been tested. Top row: Smooth Systems. Bottom row: Systems with Contact. Results of the final iterations are illustrated in Fig.1.

Algorithm 1: iRS-LQR

Given: Initial state x_0, initial input trajectory $\bar{u}_t$;
Given: Variance scheduler η, initialize $k = 0$.;
Given: Sampling distribution μ with variance Σ_0;
$\{\bar{x}_t, \bar{u}_t\}_{t=0}^{T} \leftarrow$ Rollout f from x_0 with $\bar{u}_t$;
while *Convergence* **do**
  for $0 \leq t < T$ **do**
    $\{\bar{\mathbf{A}}_t, \bar{\mathbf{B}}_t\}_{t=0}^{T} \leftarrow$ Use (15) or (16) around $(\bar{x}_t, \bar{u}_t)$ with variance $\Sigma_k = \eta(\Sigma_0, k)$;
    $\bar{c}_t \leftarrow f(\bar{x}_t, \bar{u}_t) - \bar{\mathbf{A}}_t \bar{x}_t - \bar{\mathbf{B}}_t \bar{u}_t$;
  for $0 \leq t < T$ **do**
    Observe new nominal state $\bar{x}_t$;
    Compute new nominal input $\bar{u}_t \leftarrow \text{MPC}(\bar{x}_t)$;
    Apply $\bar{u}_t$ on the system. $\bar{x}_{t+1} = f(\bar{x}_t, \bar{u}_t)$
  $k \leftarrow k + 1$ ▷ Decrease variance

VI. Experiments and Results

Here, we show experimental comparisons of iRS-LQR (Alg. 1) against the baseline of using the same algorithm with exact gradients, which we refer to as iLQR.

A. Simulation and Parallelization

We test iRS-LQR on a quasi-dynamic simulation that relies on convex implicit time-stepping models of contact [17], as well as the simulation of Drake [43]. We also parallelize iRS-LQR on the CPU by having different threads per knot point in $\{\bar{x}_t, \bar{u}_t\}_{t=0}^{T}$, which are responsible for sampling around the designated points. We use 100 samples throughout all the experiments to demonstrate the effectiveness of Monte-Carlo integration in higher dimensions.

B. Smooth Systems

Will the gradient bundle be more effective than the gradient even for smooth systems? We investigate this question for three smooth systems: the pendulum, Dubin's car [53], and the quadrotor. In many cases, the performance of iRS-LQR is comparable to or better than that of exact linearization. However, when the initial guess supplied is not informative, the gradient bundle can power through local minima and still arrive at much better minima compared to exact gradients, as demonstrated by the example of Dubin's car.

C. Non-Smooth Systems with Contact

To test our hypothesis that the gradient bundle results in more stable behavior compared to using the exact gradient in the presence of contacts, we test iRS-LQR on three systems in robotic manipulation: a planar two-finger manipulation example with gravity, a planar pushing example [5], and a box flipping (pivoting) example with gravity (Fig.1).

In the planar hand example, we start with an initial guess that makes contact. Although the exact gradient is able to drive the cost down, the fast changing gradients due to the switching of contact modes between iterations noticeably destabilize the algorithm. On the other hand, the first-order and zero-order gradient bundle are much more stable.

For the planar pushing and box pivoting example, we set the initial guess such that the robot is not in contact with the object. Due to flat gradients, iLQR is no longer able to make progress from this bad initialization. In contrast, iRS-LQR still succeeds in finding a valid descent direction and stably drives down the cost. These results demonstrate our insights in the problems of using exact gradients in Sec.IV-C.

D. First-order vs. Zero-Order Gradient Bundle

Throughout our work, we were constantly surprised to see that the zero-order gradient bundle was on par with, sometimes outperformed, the first-order one. Despite only having 100 samples, the zero-order variant also produced high-quality solutions in high-dimensional systems (e.g. quadrotor with 12 states). which is consistent with the findings of [50].

Combined with the robustness of zero-order optimization for sampling non-smooth functions (Example 2), this may suggest that zero-order methods may be a promising alternative to first-order methods in tackling non-smooth contact problems. As a consequence, simulators, given enough computation speed and parallelizability, may not necessarily need to focus on end-to-end differentiability.

E. Contact-dynamics formulations

We show that randomized smoothing is also able to stabilize second-order (fully-dynamic) systems with penalty-based contact models, whose gradients are obtained through Drake's autodiff feature. However, using the implicit quasi-dynamic formulation for planning resulted in much faster (x36) computation compared to the second-order penalty method due to less number of timesteps.

VII. Conclusion

In this work, we have presented a method to obtain the gradient bundle through contact dynamics by using randomized smoothing. By applying randomized smoothing on two different contact models of implicit time-stepping and penalty, we answered what it means to take a stochastic formulation of contact dynamics. We then applied the gradient bundle to planning-through-contact problems, and showed that the gradient bundle enables much more stable convergence behavior compared to exact gradients. Through this, we answered how a stochastic formulation of dynamics might help the planning process.

In addition to the qualitative analysis, our contribution includes taking the gradient bundle of implicit quasi-dynamic dynamics, whose explicit smooth approximations are difficult to make. Combining the stability of our gradients and the computational benefits of the quasi-dynamic simulation, we took a step towards stable and real-time solving of contact-rich manipulation problems.

References

[1] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.

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

[3] M. Geilinger, D. Hahn, J. Zehnder, M. Bächer, B. Thomaszewski, and S. Coros, "Add: Analytically differentiable dynamics for multi-body systems with frictional contact," *ACM Trans. Graph.*, vol. 39, no. 6, Nov. 2020.

[4] R. Budhiraja, J. Carpentier, C. Mastalli, and N. Mansard, "Differential dynamic programming for multi-phase rigid contact dynamics," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 1–9.

[5] F. R. Hogan and A. Rodriguez, *Feedback Control of the Pusher-Slider System: A Story of Hybrid and Underactuated Contact Dynamics*. Cham: Springer International Publishing, 2020, pp. 800–815.

[6] J. Fu, S. Levine, and P. Abbeel, "One-shot learning of manipulation skills with online dynamics adaptation and neural network priors," 2016.

[7] A. Nagabandi, K. Konoglie, S. Levine, and V. Kumar, "Deep dynamics models for learning dexterous manipulation," 2019.

[8] N. Heess, D. TB, S. Sriram, J. Lemmon, J. Merel, G. Wayne, Y. Tassa, T. Erez, Z. Wang, S. M. A. Eslami, M. Riedmiller, and D. Silver, "Emergence of locomotion behaviours in rich environments," 2017.

[9] P. Florence, L. Manuelli, and R. Tedrake, "Self-supervised correspondence in visuomotor policy learning," *Robotics and Automation Letters*, 2020.

[10] J. Burke, F. E. Curtis, A. Lewis, M. Overton, and L. Simões, *Gradient Sampling Methods for Nonsmooth Optimization*, 02 2020, pp. 201–225.

[11] J. Duchi, M. Jordan, M. Wainwright, and A. Wibisono, "Optimal rates for zero-order convex optimization: The power of two function evaluations," *IEEE Transactions on Information Theory*, vol. 61, 12 2013.

[12] J. Duchi, P. Bartlett, and M. Wainwright, "Randomized smoothing for stochastic optimization," *SIAM Journal on Optimization*, vol. 22, 03 2011.

[13] Y. Chen, "Lecture notes for ele522: Large-scale optimization for data science," 2019. [Online]. Available: https://www.princeton.edu/∼yc5/ele522_optimization/

[14] D. Stewart and J. J. Trinkle, "An implicit time-stepping scheme for rigid body dynamics with coulomb friction." vol. 1, 01 2000, pp. 162–169.

[15] M. Anitescu and F. Potra, "Formulating dynamic multi-rigid-body contact problems with friction as solvable linear complementarity problems," *Nonlinear Dynamics*, vol. 14, 03 1997.

[16] K. H. Hunt and F. R. E. Crossley, "Coefficient of Restitution Interpreted as Damping in Vibroimpact," *Journal of Applied Mechanics*, vol. 42, no. 2, pp. 440–445, 06 1975. [Online]. Available: https://doi.org/10.1115/1.3423596

[17] T. Pang and R. Tedrake, "A convex quasistatic time-stepping scheme for rigid multi-body systems with contact and friction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021.

[18] B. V. Mirtich, "Impulse-based dynamic simulation of rigid body systems," Ph.D. dissertation, 1996, aAI9723116.

[19] T. Marcucci, M. Gabiccini, and A. Artoni, "A two-stage trajectory optimization strategy for articulated bodies with unscheduled contact sequences," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 104–111, 2017.

[20] H. T. Suh and Y. Wang, "Comparing effectiveness of relaxation methods for warm starting trajectory optimization through soft contact," 2019. [Online]. Available: http://people.csail.mit.edu/felixw/soft_contact/soft_contact.pdf

[21] A. O. Onol, P. Long, and T. Padır, "Contact-implicit trajectory optimization based on a variable smooth contact model and successive convexification," in *ICRA*, 2019, pp. 2447–2453.

[22] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The linear complementarity problem*. SIAM, 2009.

[23] M. Anitescu, "Optimization-based simulation of nonsmooth rigid multibody dynamics," *Mathematical Programming*, vol. 105, no. 1, pp. 113–143, 2006.

[24] E. Todorov, "A convex, smooth and invertible contact model for trajectory optimization," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1071–1076.

[25] J. C. Boot, "On sensitivity analysis in convex quadratic programming problems," *Operations Research*, vol. 11, no. 5, pp. 771–786, 1963.

[26] T. Marcucci and R. Tedrake, "Mixed-integer formulations for optimal control of piecewise-affine systems," in *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 230–239.

[27] Z. Manchester, N. Doshi, R. J. Wood, and S. Kuindersma, "Contact-implicit trajectory optimization using variational integrators," *The International Journal of Robotics Research*, vol. 38, no. 12-13, pp. 1463–1476, 2019.

[28] Y. Tassa and E. Todorov, "Stochastic complementarity for local control of discontinuous dynamics," *Robotics: Science and Systems*, 06 2010.

[29] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Society for Industrial and Applied Mathematics, 1990.

[30] J. V. Burke, A. S. Lewis, and M. L. Overton, "Approximating subdifferentials by random sampling of gradients," *Mathematics of Operations Research*, vol. 27, no. 3, pp. 567–584, 2002.

[31] F. E. Curtis and M. L. Overton, "A sequential quadratic programming algorithm for nonconvex, nonsmooth constrained optimization," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 474–500, 2012. [Online]. Available: https://doi.org/10.1137/090780201

[32] H. Robbins and S. Monro, "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400 – 407, 1951.

[33] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462 – 466, 1952.

[34] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332–341, 1992.

[35] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition," in *Machine Learning and Knowledge Discovery in Databases*. Cham: Springer International Publishing, 2016, pp. 795–811.

[36] S. P. Bangaru, J. Michel, K. Mu, G. Bernstein, T.-M. Li, and J. Ragan-Kelley, "Systematically differentiating parametric discontinuities," *ACM Trans. Graph.*, vol. 40, no. 4, Jul. 2021.

[37] D. E. Stewart, "Rigid-body dynamics with friction and impact," *SIAM review*, vol. 42, no. 1, pp. 3–39, 2000.

[38] B. Landry, J. Lorenzetti, Z. Manchester, and M. Pavone, "Bilevel optimization for planning through contact: A semidirect method," *International Symposium on Robotics Research (ISRR)*, vol. abs/1906.04292, 2019.

[39] J.-S. Pang, J. C. Trinkle, and G. Lo, "A complementarity approach to a quasistatic multi-rigid-body contact problem," *Computational Optimization and Applications*, vol. 5, no. 2, pp. 139–154, 1996.

[40] N. Chavan-Dafle and A. Rodriguez, "Stable prehensile pushing: In-hand manipulation with alternating sticking contacts," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 254–261.

[41] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 5026–5033.

[42] R. Stribeck, *Die wesentlichen Eigenschaften der Gleit- und Rollenlager*, ser. Mitteilungen über Forschungsarbeiten auf dem Gebiete des Ingenieurwesens, insbesondere aus den Laboratorien der technischen Hochschulen. Julius Springer, 1903.

[43] R. Tedrake and the Drake Development Team, "Drake: Model-based design and verification for robotics," 2019. [Online]. Available: https://drake.mit.edu

[44] R. Tedrake, "Underactuated robotics: Algorithms for walking, running, swimming, flying, and manipulation (course notes for 6.832)," 2021.

[45] R. Elandt, E. Drumwright, M. Sherman, and A. Ruina, "A pressure field model for fast, robust approximation of net contact force and moment between nominally rigid objects," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8238–8245, 2019.

[46] R. Sutton, D. Mcallester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Adv. Neural Inf. Process. Syst*, vol. 12, 02 2000.

[47] D. H. JACOBSON and D. Q. MAYNE, "Differential dynamic programming. modern analytic and computational methods in science and mathematics, no. 24. american elsevier publ. co., inc., new york 1970. xvi, 208 s., 17 abb., dfl. 51.50." *Biometrische Zeitschrift*, vol. 15, no. 5, pp. 363–364, 1973.

[48] W. Li and E. Todorov, "Iterative linear quadratic regulator design for nonlinear biological movement systems." vol. 1, 01 2004, pp. 222–229.

[49] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4906–4913.

[50] J. Rajamäki, K. Naderi, V. Kyrki, and P. Hämäläinen, "Sampled differential dynamic programming," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 1402–1409.

[51] B. Plancher, Z. Manchester, and S. Kuindersma, "Constrained unscented dynamic programming," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 5674–5680.

[52] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2021.

[53] L. Dubins, "On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents," *American Journal of Mathematics*, vol. 79, p. 497, 1957.