How Does Noising Data Affect Learned Contact Dynamics?

H.J. Terry Suh, Max Simchowitz, Tao Pang, Russ Tedrake

***Abstract*—We investigate the effect of data noising, a regularization technique in supervised learning, in the setting of learning contact dynamics from data. We use empirical Bayes to characterize the optimal learned function under noised data in closed form, and empirically show that trained neural networks do indeed converge to this function under noised data. Using this analysis, we show that learned functions under noised data develop beneficial model biases in the setting of quasidynamic contact models that i) provide smoother gradients to facilitate better planning for gradient-based optimization, and ii) have stabilizing properties against out-of-distribution regions that allows them to stay away from the penetration region where data has not been observed before.**

I. Introduction

Noising data is a popular technique in supervised learning that regularizes a function approximator from overfitting by injecting noise in the data at training time [1], [2]. As such, practitioners in robot learning have utilized data-noising techniques for training dynamics models (i.e. system identification [3], [4]) in the context of model-based Reinforcement Learning (MBRL), which utilizes the learned model for planning and control. As MBRL is highly sensitive to model bias [5], preventing overfitting seems even more critical in this setting.

On the other hand, other works have focused on how to learn a good model *despite* being subject to noise in measuring the dynamics [6], [7]; indeed, capturing all noise in the data might lead us to overfit to noise instead of capturing the true dynamics. These two approaches share many of the same mathematical tools as they must solve how to learn dynamics under noise; despite this, the source of the noise is quite different - artificially injected in the former, and given by nature in the latter. At a cursory glance, these two stances seem mutually contradictory; yet, they tell a story of how noise can either manifest as *beneficial model bias*, or *adversarial model bias* for the learned model.

A similar contrast arises in the study of estimation of contact dynamics from data. One body of literature [8]–[11] investigates how to recover exact, non-smooth dynamics of contact from observation data by imposing inductive bias. They argue against the inherent bias of popular function approximators to discover smooth representations of contact, as it creates non-accurate bias that does not truly model contact [8], [11]. These works focus on the extent to which the smoothing bias, introduced either by noise or by the choice of popular function approximators, is *erroneous* in the setting of contact.

In contrast, other approaches have argued that smoothing bias *is beneficial* when *planning* through models of contact dynamics. Motivated by randomized smoothing approaches from non-smooth optimization [12], these works [13]–[16] purposely inject noise in known models of contact dynamics at runtime in order to smooth out the contact dynamics stochastically. Many cases have been documented [15] where smoothing out contact dynamics creates a more favorable landscape for gradient-based optimization for planning and control, as the landscapes of value functions for contact-rich problems are often filled with flatness and discontinuities.

Because the first camp considers dynamic estimation, and the second planning and control, the two camps cannot be compared directly. This paper addresses the chasm between these groups investigating how smoothing bias, specifically from intentionally augmenting data with noise, affects learning unknown contact-rich dynamics with a universal function approximator. We find that indeed, previous approaches that smooth known dynamics are not directly applicable because of the peculiarities of contact systems, and the resulting distribution of un-noised ("clean") transition data.

One key example is as follows: MBRL might reason about plans that cross over into the penetration region, regions from which we have never seen data (no "real" dynamics can penetrate). While we might expect this to degrade performance, we find that, in fact, MBRL can be quite successful for contact-rich tasks [17]. What dynamics do these models learn in the penetration region such that MBRL can still succeed? Does noising data help regularize the behavior of dynamics in this out-of-distribution (o.o.d.) regime?

To answer these questions, we utilize empirical Bayes [18], [19] to characterize the optimal learned function in the presence of data noise, and analyze the behavior of the learned dynamics under contact-rich settings. Intriguingly, our analysis finds that learned dynamics with noised data i) has beneficial model bias towards smooth gradients that allows it to improve over true known dynamics as well as learned dynamics without noise, and ii) has stabilizing properties against o.o.d. regions for quasidynamic models that allow it to recover back from penetration regions. We hope that our contribution adds a valuable perspective in the role of noise for learning models of contact dynamics, and informs practitioners in MBRL of the effect of their training decisions.

II. Preliminaries

We first introduce some relevant concepts for our analysis. In order to simplify the exposition, we deal first with the function setting with $f : \mathbb{R}^n \to \mathbb{R}^m$, and introduce the dynamics setting with states and inputs in Sec. IV.

A. Empirical Risk Minimization

Consider a typical function learning setting, where we aim to approximate an unknown function $f : \mathbb{R}^n \to \mathbb{R}^m$ with access to some dataset $\mathcal{D} = \{(x_i, f(x_i)\}$. Denoting $\hat{p}(x; \mathcal{D})$ as the empirical distribution corresponding to the dataset $\mathcal{D}$, empirical risk minimization aims to solve the following problem for approximating f,

$$\min_g \mathbb{E}_{x \sim \hat{p}(x;\mathcal{D})} \left[\|g(x) - f(x)\|^2 \right]. \tag{1}$$

Note that the optimal appoximator g^* is not unique, as long as it satisfies $g(x_i) = f(x_i)$ for all x_i.

B. Randomized Smoothing

Given some function f, randomized smoothing randomly perturbs the domain of the function according to the distribution $p_\sigma(\tilde{x}|x)$, which is often set as a Gaussian kernel $p_\sigma(\tilde{x}|x) = \mathcal{N}(\tilde{x}; x, \sigma^2\mathbf{I})$, and averages the values to define a new smoothed function.

Definition 1 (Randomized Smoothing)**.** Given some function f, its randomized smoothed function f_σ is defined as

$$\begin{aligned} f_\sigma(x) &\coloneqq \mathbb{E}_{\tilde{x}\sim p_\sigma(\tilde{x}|x)} f(\tilde{x}) \\ &= \mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2\mathbf{I})} f(x+w), \end{aligned} \tag{2}$$

where we denote w as the noise injected noise $w = \tilde{x} - x$, which is drawn from a zero-mean Gaussian with variance $\sigma^2\mathbf{I}$. Note that f_σ can be understood as a *convolution* between f and a Gaussian kernel; as a result, f_σ is smooth even if f is non-smooth or non-Lipschitz [13].

C. Data Noising

Given $x \sim \hat{p}(x;\mathcal{D})$ and its corresponding data $(x, f(x))$, data noising creates a fake data pair $(\tilde{x}; f(x))$ with $\tilde{x} = x + w$, where w is drawn from some Gaussian noise. Thus, the function approximator is forced to regress $g(\tilde{x})$ against $f(x)$, with g being the function approximator. The risk minimization under noised data is written as

$$\begin{aligned} &\min_g \mathbb{E}_{x\sim\hat{p}(x;\mathcal{D})}\mathbb{E}_{\tilde{x}\sim p_\sigma(\tilde{x}|x)}\left[\|g(\tilde{x}) - f(x)\|^2\right], \\ =&\min_g \mathbb{E}_{x\sim\hat{p}(x;\mathcal{D})}\mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2\mathbf{I})}\left[\|g(x+w) - f(x)\|^2\right]. \end{aligned} \tag{3}$$

III. How Data Augmentation Affects Learning

In this section, we review the standard empirical Bayes [18] characterizations of the optimal solution of the data-noised regression problem (Eq. (3)) in closed-form.

A. Randomized Smoothing

We first note that in the special case where we allow x to be drawn from anywhere over the reals in the absence of the empirical distribution.

Lemma 1. If $p(x)$ is a uniform distribution over all of $\mathbb{R}^n$ (an improper prior), the solution to (3) is the randomized smoothed version of f,

$$g(x) = f_\sigma(x). \tag{4}$$

Proof: Since x has a uniform improper prior, we fix x and solve pointwise for the best value of g. This allows us to use rewrite (3) as

$$\begin{aligned} &\min_g \mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2\mathbf{I})}\left[\|g(x+w) - f(x)\|^2\right] \\ =&\mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2\mathbf{I})}\left[\|g(\tilde{x}) - f(\tilde{x}-w)\|^2\right], \end{aligned} \tag{5}$$

for which the minimizer is

$$g(\tilde{x}) = \mathbb{E}_{w\sim\mathcal{N}(w;0,\sigma^2\mathbf{I})} f(\tilde{x}-w) = f_\sigma(\tilde{x}). \tag{6}$$

■

B. Reversed Smoothing

In practice, we cannot have a uniform distribution over all of $\mathbb{R}^n$ for two important reasons. First, it is impossible to decouple the learned function g from the data distribution it was trained on, $\hat{p}(x;\mathcal{D})$. In addition, even if we asymptotically collect such data, the support of the data distribution cannot be all of $\mathbb{R}^n$ as they must obey the non-penetration constraint for contact dynamics. This motivates us to write a more general solution to Eq. (3) using empirical Bayes [18].

Lemma 2 (Reversed Smoothing)**.** The solution to Eq. (3) is given by a smooth function [20] g_σ, defined as

$$g_\sigma(\tilde{x}) \coloneqq \mathbb{E}_{x\sim p_\sigma(x|\tilde{x})}\left[f(x)\right] = \int f(x) p_\sigma(x|\tilde{x}) dx \tag{7}$$

with $p_\sigma(x|\tilde{x})$ as the *reversed-smoothing distribution*, which is the distribution of data x which could have been augmented from $\mathcal{D}$ to create $\tilde{x}$,

$$p_\sigma(x|\tilde{x}) = \frac{p_\sigma(\tilde{x}|x)\hat{p}(x;\mathcal{D})}{p_\sigma(\tilde{x};\mathcal{D})}. \tag{8}$$

where $p_\sigma(\tilde{x};\mathcal{D}) \coloneqq \int p_\sigma(\tilde{x}|x)\hat{p}(x;\mathcal{D})dx$ is the *perturbed empirical distribution*, which is a randomized-smoothed version of the empirical distribution [21], [22].

Proof: We use the property of conditional expectation

$$\text{argmin}_g \mathbb{E}_{x,y}\|g(y) - f(x)\|_2^2 = \mathbb{E}_{x\sim p(x|y)}[f(x)], \tag{9}$$

which can be obtained by rewriting the objective as

$$\begin{aligned} &\iint \left[\|f(x) - g(y)\|^2\right] p(x,y) dx dy \\ =&\int \underbrace{\left[\int \left[\|f(x) - g(y)\|^2\right] p(x|y) dx\right]}_{\mathcal{L}(y)} p(y) dy \end{aligned} \tag{10}$$

and noting that $\int f(x)p(x|y)dx$ minimizes $\mathcal{L}(y)$ for every y using first-order optimality. ■

C. Closed-form solution with Empirical Bayes

When $p(x)$ describes an empirical distribution, the optimal learned function under data noising $g(\tilde{x})$ can be characterized in closed-form using standard empirical Bayes [18].

Lemma 3 (Exponential-Weighted Distance Average)**.** For an empirical data distribution $\hat{p}(x;\mathcal{D})$, the optimal function (7) can be written as a weighted average of $f(x_i)$ for $x_i \in \mathcal{D}$,

$$g_\sigma(\tilde{x}) = \left[\frac{\sum_{x_i\in\mathcal{D}} w_\sigma(\tilde{x}, x_i) f(x_i)}{\sum_{x_i\in\mathcal{D}} w_\sigma(\tilde{x}, x_i)}\right], \tag{11}$$

with $w_\sigma(\tilde{x}, x_i) \coloneqq \exp\left[-\|\tilde{x} - x_i\|^2/2\sigma^2\right]$.

Proof: Since $\hat{p}(x;\mathcal{D}) = \frac{1}{N}\sum_{x_i\in\mathcal{D}}\delta(x_i)$, we can use the shifting property of dirac-deltas to write

$$\begin{aligned} \int f(x)p_\sigma(\tilde{x}|x)\hat{p}(x;\mathcal{D})dx &= \tfrac{1}{N}\textstyle\sum_{x_i\in\mathcal{D}} f(x_i)p_\sigma(\tilde{x}|x_i) \\ \int p_\sigma(\tilde{x}|x)\hat{p}(x;\mathcal{D})dx &= \tfrac{1}{N}\textstyle\sum_{x_i\in\mathcal{D}} p_\sigma(\tilde{x}|x_i). \end{aligned} \tag{12}$$

The above result is obtained by simplifying $p_\sigma(\tilde{x}|x_i)$ which are Gaussians. Note that weights act like softmin based on quadratic distance, with temperature parameter σ [22].

■

IV. Contact Dynamics under Data Augmentation

We now apply the tools from Sec. III to the setting of contact dynamics to answer the main question of the paper: how would noising data affect learned contact dynamics?

A. Quasi-dynamic Model of Contact

For our analysis, we consider the setting of [16], [23], where the dynamics of physical systems with contacts is discretized by a time-stepping scheme with timestep $h \in \mathbb{R}_{>0}$. The generalized positions of the system, $q \in \mathbb{R}^n$, is divided into actuated positions $q^{\mathsf{a}} \in \mathbb{R}^{n_{\mathsf{a}}}$ and unactuated positions $q^{\mathsf{u}} \in \mathbb{R}^{n_{\mathsf{u}}}$ such that $q = (q^{\mathsf{u}}, q^{\mathsf{a}})$. In addition, the input commands u are defined as position commands to a stiffness controller that exerts impulses $\lambda = h\mathbf{K}_{\mathsf{a}}(q^{\mathsf{a}} - u)$, with $\mathbf{K}_{\mathsf{a}} \in \mathbb{R}^{n_{\mathsf{a}} \times n_{\mathsf{a}}}$ being the stiffness values. Given the current q, the quasidynamic equations of motion in the frictionless case can then be written as finding the next relative displacements $\delta q = (\delta q^{\mathsf{u}}, \delta q^{\mathsf{a}})$, as well as contact impulses $\lambda_i \in \mathbb{R}^3$ for each possible n_c contact pairs, that satisfy the following constraints,

$$\begin{aligned} \text{find} \quad & \delta q, \lambda_i \\ \text{s.t.} \quad & h\mathbf{K}_{\mathsf{a}}(q^{\mathsf{a}} + \delta q^{\mathsf{a}} - u_t) = h\tau^{\mathsf{a}} + \textstyle\sum_{i=1}^{n_c} \mathbf{J}_{\mathsf{a}_i}(q)\lambda_i \\ & \frac{1}{h}\mathbf{M}_{\mathsf{u}}(q)\delta q^{\mathsf{u}} = h\tau^{\mathsf{u}} + \textstyle\sum_{i=1}^{n_c} \mathbf{J}_{\mathsf{u}_i}(q)\lambda_i \\ & 0 \le \phi_i(q) + \mathbf{J}_i^{\mathsf{n}}(q)\delta q \perp \lambda_i \ge 0 \quad \forall i \in \{1 \cdots n_c\}. \end{aligned} \tag{13}$$

Here, $\tau^{\mathsf{a}}, \tau^{\mathsf{u}}$ are external forces on the system such as gravity, and $\phi_i(q) \in \mathbb{R}_{>0}$ is the value of the signed-distance function. In addition, $\mathbf{J}_i = [\mathbf{J}_{\mathsf{u}_i}, \mathbf{J}_{\mathsf{a}_i}] \in \mathbb{R}^{3\times n}$ are the contact Jacobians that describe the principle directions of the contact frame, with $\mathbf{J}_i^{\mathsf{n}}(q) \in \mathbb{R}^{1\times n}$ being the normal component of the frame (i.e. linearization of $\phi_i(q)$). We refer the reader to [16] for the full case with friction. We further note that the dynamics enforce the *non-penetration constraint* by linearizing $\phi_i(q + \delta q) \ge 0$, and accurately model the fact that there the regions with $\phi_i(q) < 0$ are never seen during training.

B. Analyzing within-distribution behavior

To verify the results of Lem. 1, we first set up a 1D system with two bodies, one actuated and one without. Given a fixed initial condition, the next configuration of the unactuated body q^{u}_{t+1} as a function of the position command on the actuated body u describes a ReLU-like function, as the unactuated body will not move if not in contact, and move close to the position command of the actuated body if it makes contact [13].

We see in Fig.1 that when we have uniformly dense data coverage over all of u_t, the Exponential Weighted Distance Average (EWDA) $g_\sigma(x)$ does create a smoothing effect equivalent to randomized smoothing f_σ [13]. In addition, in order to empirically verify the results of Lem. 3, we also train a neural network with noised data, which we call h_σ. In both the sparse data regime and the dense data regime, we observe that the trained network and the EWDA solution shows a good match. Combining the two insights, we believe that in the asymptotic regime, the learned dynamics will have similar beneficial model biases as smoothed contact dynamics in the context of gradient-based optimization [13], [16].

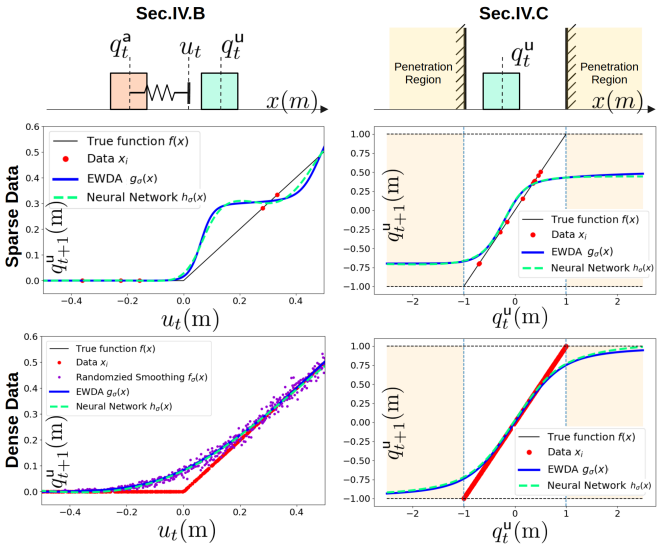

Fig. 1: Left: Figure for Sec. IV-B. Right: Figure for Sec. IV-C. The first row visualizes the examples, the second row shows results in the sparse data regime, and the third row displays results for the dense data regime. Note that the purple dots on the left correspond to Monte-Carlo estimation of f_σ. The right does not have these as f is not properly defined in the penetration region.

C. Analyzing out-of-distribution behavior

To study out-of-distribution behavior, we set up a simple 1D example in the absence of any actuation. In this case the dynamics is simply the identity function $q^{\mathsf{u}} = f(q^{\mathsf{u}})$, as without any actuation or external forces, the block will simply remain still. However, we set up walls on both side of the block that acts as a penetration region. While no data is collected in this region, the EWDA can still be evaluated.

In the dense data regime, we find that the EWDA takes states that are in the penetration region and moves them to the non-penetration region at the next step. This stabilizing behavior is consistent with our intuition, since we regress the perturbed data back to data within the distribution. We believe that this effect provides noised dynamics with additional stability against going into penetrating regions during rollouts.

Finally, we study how the learned function changes in the regime of sparse data. We find that the EWDA simply takes the support of the data and behaves as if the walls were placed on each side of the support. Indeed, from the given data, there is no way to tell if the region belongs to a penetrating region, or if it simply outside the support of data. Intriguingly, this duality implies that the learned dynamics will try to push away from o.o.d. regions, which might provide some benefits in combating distribution shift in MBRL [22].

V. Gradient-Based Optimal Control

Using insights from Sec. III, we finally ask if noising data offers benefits in gradient-based planning through learned contact dynamics, as injecting noise at runtime did in the setting of planning with known physics models [13].

A. Gradient-based Single-Shooting

Given some running cost matrices $\mathbf{Q} \in \mathbb{S}_+^{n\times n}, \mathbf{R} \in \mathbb{S}_+^{n_{\mathsf{a}}\times n_{\mathsf{a}}}$ and a terminal cost matrix $\mathbf{Q}_T \in \mathbb{S}_+^{n\times n}$, we simultaneously solve for a nominal trajectory $\bar{x}_t$ and time-varying gains $\mathbf{K}_t \in$

$\mathbb{R}^{n_a \times n}$, such that the feedback law becomes $u_t = \mathbf{K}_t(x_t - \bar{x}_t)$. Reparametrizing $-\mathbf{K}_t\bar{x}_t = k_t$ allows us to rewrite $u_t = \mathbf{K}_t x_t + k_t$, resulting in the following problem:

$$\begin{aligned} \min_{\mathbf{K}_t, k_t, x_t, u_t, \delta u_t} \quad & x_T^\top \mathbf{Q}_T x_T + \textstyle\sum_{t=0}^{T-1} \left(x_t^\top \mathbf{Q} x_t + \delta u_t^\top \mathbf{R} \delta u_t \right) \\ \text{s.t.} \quad & x_{t+1} = f(x_t, u_t) \quad \forall t \in \{0 \cdots T-1\}, \\ & \delta u_t = u_t - u_{t-1} \\ & u_t = \mathbf{K}_t x_t + k_t \quad \forall t \in \{0 \cdots T-1\}, \end{aligned} \tag{14}$$

where δu_t comes u_t being position commands in [16].

Given a differentiable f, we can obtain the gradients of the value with respect to decision variables $\mathbf{K}_t, k_t$ using automatic differentiation and pass it to a first-order optimizer such as Adam [24]. We solve the optimal control problem under 5 different representations of dynamics: (i) true dynamics f, (ii) true dynamics with randomized smoothing f_σ, (iii) EWDA under noise g_σ, (iv) learned neural network (NN) without data noising h, and (v) learned NN with data noising h_σ. We perform this comparison on two systems of `PlanarPushing`, and `PlanarHand` from [16], where we collect a dataset of size $100,000$ in the non-penetrating region. We initialize from a no-contact configuration to highlight pathologies of using exact dynamics for gradient-based planning.

B. Results & Discussion

Our results are plotted in Fig. 2. In both cases, using gradients from exact dynamics f does not make any progress as the objects are not initially in contact. On the other hand, in both cases, the randomized smoothed version f_σ does the best, which is consistent with the findings of [13].

Unlike the true dynamics f and the smoothed dynamics f_σ, the EWDA and the NNs both need to take the empirical data distribution into account. We find that in this setting, the NN that was trained without noised data can be prone to the same pathologies of exact dynamics f in the `PlanarPushing` case and not able to make any progress. However, it makes improvements in the `PlanarHand` example, which may indicate that the bias of NNs to learn smooth functions might allow it to improve over using gradients of exact dynamics f.

Our experiments nevertheless indicate that EWDA and the NN trained with noised data, which are models that noise the training data, show superior performance compared to the NN trained without noised data, as well as exact dynamics f, in both examples. We note that the EWDA performs better than the NN trained with noise data, as it is not prone to any approximation error. These preliminary experiment results suggest that noising data can indeed provide *beneficial model bias* in the context of gradient-based optimization for planning.

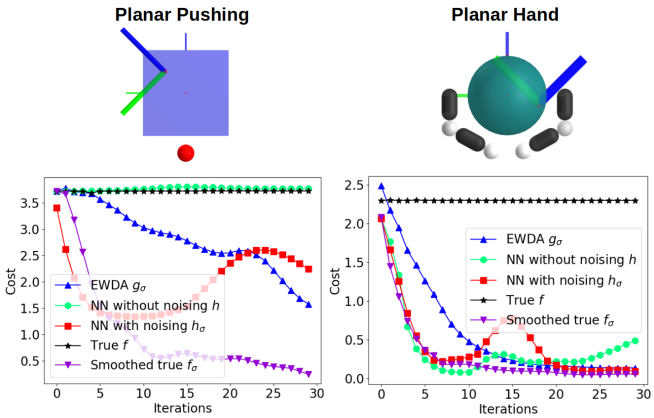

Fig. 2: Left: Cost vs. iteration plot for planar pushing. Right: Same plot for planar hand. The top figure illustrates initial and goal poses, where goal is denoted with a thicker frame.

References

[1] C. M. Bishop, "Training with noise is equivalent to tikhonov regularization," *Neural Computation*, vol. 7, no. 1, pp. 108–116, 1995.

[2] A. Orvieto, A. Raj, H. Kersting, and F. Bach, "Explicit regularization in overparametrized models via noise injection," in *26th International Conference on Artificial Intelligence and Statistics*, vol. 206. PMLR, 25–27 Apr 2023, pp. 7265–7287.

[3] S. Pitis, E. Creager, A. Mandlekar, and A. Garg, "Mocoda: Model-based counterfactual data augmentation," 2022.

[4] R. Khraishi and R. Okhrati, "Simple noisy environment augmentation for reinforcement learning," 2023.

[5] J. Schneider, "Exploiting model uncertainty estimates for safe dynamic control learning," in *Advances in Neural Information Processing Systems*, vol. 9. MIT Press, 1996.

[6] P. Goyal and P. Benner, "Learning dynamics from noisy measurements using deep learning with a runge-kutta constraint," 2021.

[7] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse identification of nonlinear dynamics with control (sindyc)," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016, 10th IFAC Symposium on Nonlinear Control Systems NOLCOS 2016.

[8] S. Pfrommer, M. Halm, and M. Posa, "Contactnets: Learning discontinuous contact dynamics with smooth, implicit representations," 2020.

[9] M. Parmar, M. Halm, and M. Posa, "Fundamental challenges in deep learning for stiff contact dynamics," 2021.

[10] B. Bianchini, M. Halm, N. Matni, and M. Posa, "Generalization bounded implicit learning of nearly discontinuous functions," in *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, vol. 168. PMLR, 23–24 Jun 2022, pp. 1112–1124.

[11] W. Jin, A. Aydinoglu, M. Halm, and M. Posa, "Learning linear complementarity systems," in *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, vol. 168. PMLR, 23–24 Jun 2022.

[12] J. C. Duchi, P. L. Bartlett, and M. J. Wainwright, "Randomized smoothing for stochastic optimization," 2012.

[13] H. J. T. Suh, T. Pang, and R. Tedrake, "Bundled gradients through contact via randomized smoothing," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4000–4007, 2022.

[14] Q. L. Lidec, L. Montaut, C. Schmid, I. Laptev, and J. Carpentier, "Leveraging randomized smoothing for optimal control of nonsmooth dynamical systems," 2022.

[15] H. J. T. Suh, M. Simchowitz, K. Zhang, and R. Tedrake, "Do differentiable simulators give better policy gradients?" in *Proceedings of the 39th International Conference on Machine Learning*, vol. 162. PMLR, 17–23 Jul 2022, pp. 20 668–20 696.

[16] T. Pang, H. J. T. Suh, L. Yang, and R. Tedrake, "Global planning for contact-rich manipulation via local smoothing of quasi-dynamic contact models," 2023.

[17] A. Nagabandi, K. Konolige, S. Levine, and V. Kumar, "Deep dynamics models for learning dexterous manipulation," in *Proceedings of the Conference on Robot Learning*, vol. 100. PMLR, 30 Oct–01 Nov 2020, pp. 1101–1112.

[18] H. E. Robbins, *An Empirical Bayes Approach to Statistics*. New York, NY: Springer New York, 1992, pp. 388–394.

[19] S. Saremi, R. K. Srivastava, and F. Bach, "Universal smoothed score functions for generative modeling," 2023.

[20] A. Block, D. Pfrommer, and M. Simchowitz, "Imitating arbitrary demonstrators: Bridging low-level stability and high-level planning," 2023.

[21] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[22] H. J. T. Suh, G. Chou, H. Dai, L. Yang, A. Gupta, and R. Tedrake, "Fighting uncertainty with gradients: Offline reinforcement learning via diffusion score matching," 2023.

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

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.