

Distributionally Robust Control via Stein Variational Inference for Contact-Rich Manipulation

Hrshikesh Sathyanarayan*, Victor Vantilborgh†, Harish Ravichandar‡, Tom Lefebvre†, and Ian Abraham*§
*Yale University, †Ghent University, ‡Georgia Tech, §University of Sydney

Abstract—Reliable robotic manipulation requires control policies that can accurately represent and adapt to uncertainty arising from contact-rich interactions. Modern data-driven methods mitigate uncertainty through large-scale training and computation, and degrade significantly in performance with limited number of training samples. By contrast, classical model-based controllers are computationally efficient and reliable, but their limited ability to represent task-relevant uncertainty can hinder performance in contact-rich interactions.

In this work, we propose to expand the capabilities of model-based manipulation control through more flexible uncertainty modeling that retains performance while exactly adapting to uncertainty. Our approach casts the manipulation problem as a distributionally robust control optimization and proposes a novel deterministic formulation based on Stein variational inference that preserves performance while explicitly modeling task-sensitive parameter uncertainty. As a result, the derived controllers are more aware of task sensitivities to uncertainty, yielding high reliability without compromising performance. Experimental results demonstrate up to $3\times$ improved robustness across a range of contact-rich manipulation tasks under broad parametric uncertainty, outperforming existing model-based control methods.

I. INTRODUCTION

In-the-wild manipulation often requires reasoning in an environment filled with uncertainty. However, the reliability and success of robotic control for manipulation is contingent on the ability to appropriately exploit and adapt to uncertainty through interacting with their environment. Modern control approaches mitigate uncertainty through vast accumulation of data and compute [8, 37, 4, 7, 5], effectively aiming to cover all possible scenarios at the expense of precision and interpretability, and their performance rapidly degrades with a reducing number of samples [20, 6]. The unifying goal is to develop zero-shot manipulation controllers that actively adapt to physical uncertainty, without the loss of performance and robustness.

Most prior work on uncertainty-aware manipulation control under parameter uncertainty spans data-driven methods and

*H.S. is with the Department of Mechanical Engineering, Yale University, New Haven, CT, USA, email: hrishi.sathyanarayan@yale.edu

†V.V. and T.L. are with the Department of Electromechanical, Systems and Metal Engineering, Ghent University, Belgium, email: { victor.vantilborgh, tom.lefebvre }@ugent.be

‡H.R. is with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA, email: harish.ravichandar@cc.gatech.edu

§I.A. is with the Department of Electrical Engineering, University of Sydney, Australia, email: ian.abraham@sydney.edu.au

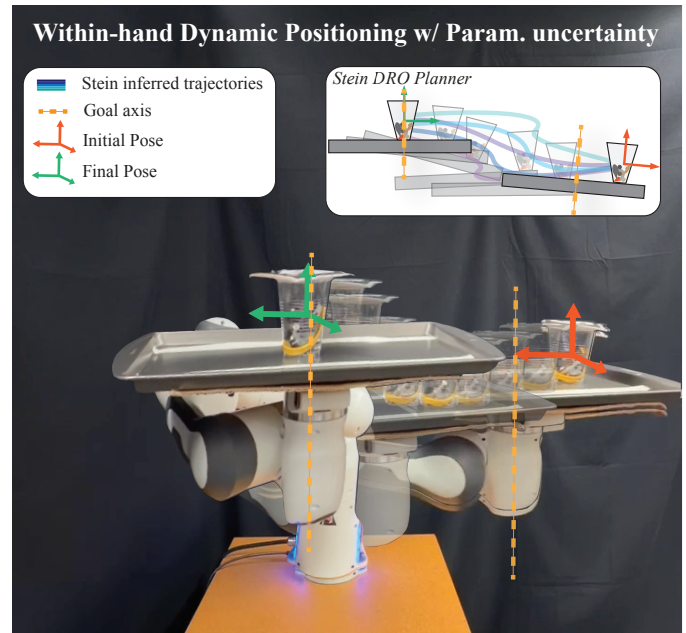


Fig. 1: Within-hand dynamic positioning of a cup with unknown mass distribution and friction coefficient. Our proposed Stein Variational Distributionally Robust Optimizer (SV-DRO) applied to a within-hand positioning task via controlled sliding. The demonstration shown requires the robot to slide an object with unknown physical parameters (inertia, mass, and friction coefficient) to a goal state located at the center of the tray. Our approach enables control policies to reason over uncertainty-sensitivity to task ensuring reliable performance.

classical adaptive control approaches, each expressing a different facet on robust optimization [44, 14, 2]. Reinforcement learning (RL) and related policy optimization methods in [10, 28] have been extensively applied to contact-rich tasks, where the ability to implicitly learn from experience offers potential adaptability in uncertainty-filled environments. These methods, however, often necessitate high compute and data quality, and sufficient coverage over all possible parameter uncertainties for high-performance guarantees [40, 30, 33]. Other works presented in [1, 25, 26] leverage sampled trajectory ensembles to mitigate the effects of parametric uncertainty through optimizing controls over average performance. These sample-based ensemble approaches have limitations, however, in that they fail to adapt robustly to parameter uncertainty under a prohibited number of samples, and compute only *average* performance that is highly dependent on the range of parameter uncertainty. Other, classical control methods

[23, 24, 36, 9] promise control policies from conservative, worst-case formulations that adapt to uncertainty via a Distributionally Robust Optimization (DRO) formulation [16]. These methods compromise task performance in response to worst-case uncertainty modeling leading to worse overall performance in contact-rich settings due to the variability of object dynamics subject to contact.

Thus, the goal of this work is to expand the capabilities of model-based control for robotic manipulation through more flexible uncertainty modeling that retains performance while adapting to uncertainty in manipulation.

To achieve this goal, we formulate robotic control under parametric uncertainty as a Distributionally Robust Optimization problem and introduce a Stein variational inference-based controller [35, 22]. Rather than optimizing against a fixed ambiguity set, we use Stein variational gradient descent (SVGD) to evolve deterministic particles drawn from a prior into a task-aware parameter posterior. This provides a non-parametric, parallelizable approximation of how parameter uncertainty affects task performance [43, 21]. Fig. 1 illustrates this idea on within-hand manipulation, where an object with uncertain mass distribution and friction is reliably moved to the tray center. Overall, SV-DRO uses exact¹ Stein inference to identify uncertainty-sensitive actions critical to contact-rich manipulation.

In practice, SV-DRO achieves strong task performance under broad parameter uncertainty and limited feedback compared with existing model-based baselines [1, 16]. The central distinction is philosophical: instead of planning for a globally worst-case model, SV-DRO reshapes uncertainty toward parameter variations that most affect the current task and optimizes against those task-sensitive particles. We demonstrate this behavior on contact-rich manipulation tasks including within-hand dynamic positioning and bimanual Push-T.

To summarize, the contributions of this paper are the following:

- Derivation of a Stein variational inference-based distributionally robust controller that accurately models uncertainty and generates control solutions that enhances task performance.
- Demonstration of $3\times$ improvement of our proposed method compared against existing DRO and ensemble worst-case manipulation controllers subject wide ranging parameter uncertainty.
- Showcasing continually reliable within-hand manipulation on a rolling experimental setting.

The remainder of this paper is structured as follows: Section II discusses prior relevant work. Section III overviews preliminary material on Distributionally Robust Control and Stein Variational Gradient Descent. Section IV outlines our proposed method that combines Stein Variational Inference with DRO, and details theoretical analyses on our method.

¹SVGD can exactly match expectations of functions within a Stein-induced function class and converges to the target distribution in the infinite-particle limit [22]; with finite particles, it gives an approximate posterior representation.

And finally, Section V discusses experimental results, Section VI covers limitations, and Section VII proposes future research directions.

II. RELATED WORK

A. Distributionally Robust Optimization

Distributionally Robust Optimization (DRO) optimizes performance against worst-case probability distributions within an ambiguity set, providing robustness to uncertainty at the cost of potential conservatism [16, 32]. Prior work has made DRO tractable for control by using duality under Kullback-Leibler ambiguity sets, reducing the original min-max problem to a single-level optimization [23, 36]. However, these approaches typically require a parametric reference distribution estimated through maximum-likelihood or related data-driven procedures, so robustness is only guaranteed near a learned nominal uncertainty model. Wasserstein DRO formulations avoid explicit distribution fitting by using empirical uncertainty samples [24], but they assume direct access to samples from the uncertain process [24, 13, Assumption 6.3], which is restrictive when uncertainty corresponds to latent physical parameters such as mass, inertia, friction, or geometry.

These assumptions are especially limiting in contact-rich manipulation, where latent parameters cannot usually be sampled directly, feedback may be sparse or indirect, and contact dynamics introduce discontinuous, mode-dependent behavior outside the smooth settings considered by many DRO methods [23, 24]. Moreover, optimizing against all admissible perturbations can produce overly conservative behavior that sacrifices task performance. In this work, we extend DRO-based control by using Stein variational inference to evolve particles over latent physical parameters according to task-relevant sensitivity, rather than relying on a fixed reference distribution or a sampling oracle. This enables robust MPC for contact-dominated manipulation while focusing computation on uncertainty that directly affects task success.

B. Stein Variational Inference in Robot Control

Variational Inference (VI) approximates an intractable target distribution by optimizing a surrogate distribution to minimize the Kullback-Leibler divergence, but its effectiveness depends on choosing an expressive variational family [35], which is difficult in high-dimensional or multi-modal settings. Stein variational inference, or Stein variational gradient descent (SVGD), avoids this restriction by evolving a set of particles through deterministic functional gradient descent in a reproducing kernel Hilbert space [22]. This particle-based update combines attraction toward high-probability regions with a kernel-induced repulsion term that preserves diversity, making SVGD well suited for representing complex uncertainty distributions in control. Prior work has used SVI in MPC to maintain multi-modal control particles [17], optimize constrained trajectory distributions [31], generate informative failures [11], perform task and motion planning [18], and jointly infer controls and dynamics parameters online [3].

Our use of SVI differs from estimation-driven Stein MPC methods [3] that update parameter particles from online state-transition measurements to recover a dynamics posterior [3]. Such approaches require sufficiently informative feedback, which can be sparse, low-frequency, noisy, or only indirectly related to latent contact parameters such as mass distribution, inertia, and friction in contact-rich manipulation. In contrast, we do not use SVGD to identify the true physical parameters through a measurement-likelihood, contact-aware maximum a-posteriori, or other contact-aware active learning objective [34, 41, 48]. Instead, we construct a task-aware parameter posterior induced by the optimality gap, transport particles toward regions where the planned trajectory is most sensitive to uncertainty, and embed the resulting particle set inside a distributionally robust MPC objective. This allows the controller to optimize against uncertainty that directly affects task success, rather than uncertainty that merely best explains recent state feedback.

C. Domain Randomization

Methods in reinforcement learning (RL) [45, 15] based on domain randomization [29, 39] aim to achieve robustness under parametric uncertainty by training policies over a distribution of system parameters sampled offline. In this setting, a policy is optimized to perform well in expectation across the randomized domain, implicitly learning adapting to variations in model dynamics or environmental scenery. While effective, such approaches typically require extensive data collection and training, and their performance depends critically on the choice and coverage of the parameter distribution used during training [46, 38]. Moreover, the resulting domain randomization process is inherently *task-agnostic* with respect to parameters with larger uncertainties, as the prior distribution used for sampling is fixed a priori and not adapted based on the specific control objective at execution time.

In contrast, our approach performs online, task-aware adaptation of the parameter distribution within a model predictive control (MPC) framework, eliminating the need for offline training or repeated system interaction. By leveraging SVI, the parameter distribution is iteratively reshaped according to the optimality gap, focusing computation on uncertainty directions that directly impact task performance. This enables zero-shot control under uncertainty while maintaining computational tractability, and yields policies that are both robust and dynamically responsive to task-specific sensitivities. As such, our method can be viewed as complementary to domain randomization, offering a more data-efficient alternative that emphasizes task-relevant uncertainty at test time rather than distributional coverage during training.

III. PRELIMINARIES

A. Distributionally Robust Control

We now formalize the control problem under parametric uncertainty using a distributionally robust perspective. In this setting, the objective is to synthesize a trajectory that minimizes a task-dependent cost while remaining robust to

parameter uncertainty. To do so, we first consider the following optimization problem,

$$\begin{aligned} & \min_{\tau = \{(x_k, u_k, c_k) \forall k \in [0, t_h]\}} \mathcal{J}_\theta(\tau) \\ \text{s.t. } & \begin{cases} x_0 \in \mathcal{X} & (\text{init. state}) \\ x_{k+1} = f_\theta(x_k, u_k) & (\text{dynamical model}) \\ c_{k+1} \in \mathcal{C}(x_k) & (\text{contact model}) \end{cases} \end{aligned} \quad (1)$$

where $\mathcal{J}_\theta(\cdot) : \mathcal{X} \times \mathcal{U} \times \mathcal{C} \times \Theta \rightarrow \mathbb{R}$ is the task-relevant cost to minimize, $\tau = \{(x_k, u_k, c_k) \forall k \in [t, t + t_h]\}$ is a sequence of states $x \in \mathcal{X}$, control inputs $u \in \mathcal{U}$, and contact impulses $c_k \in \mathcal{C}$, with $t \in [0, T]$ being the current time iterate within T total time iterations and t_h being the planning horizon, and $\theta \in \Theta$ are physics model parameters. The Lagrangian $\mathcal{L} : \mathcal{X} \times \mathcal{U} \times \mathcal{C} \times \Theta \rightarrow \mathbb{R}$ of the above optimization problem can be written as,

$$\mathcal{L}(\tau, \theta) = \mathcal{J}_\theta(\tau) + \lambda^\top h_\theta(\tau) \quad (2)$$

where $\lambda \in \mathbb{R}^{n_{eq}}$ is a vector of Lagrange multipliers, and $h_\theta : \mathcal{X} \times \mathcal{U} \times \mathcal{C} \times \Theta \rightarrow \mathbb{R}^{n_{eq}}$ is the vector of equality constraints from Eq. 1. In this paper, we implement Eq. 1 via forward rollouts under the surrogate model f_θ (which we construct), which computes contact impulses using a spring-like contact model [19], and execute the first control input in a receding-horizon manner, avoiding the need to explicitly enforce equality constraints in practice. Distributionally Robust Optimization (DRO) seeks to optimize over the Lagrangian via a bi-level optimization problem,

$$\min_{\tau} \sup_{p \in \mathcal{P}} \mathbb{E}_{\theta \sim p(\theta)} [\mathcal{L}(\tau^*, \theta)] \quad (3)$$

where p is a surrogate distribution over θ , and $\mathcal{P}$ is the following ambiguity set,

$$\mathcal{P} = \{p \mid \mathbb{D}_{KL}[p \parallel q] \leq \epsilon\} \quad (4)$$

where q is true unknown target distribution. Since finding an ambiguity set $\mathcal{P}$ is challenging to find a-priori, the above DRO problem is commonly approximated as shown in [23] as,

$$\min_{\tau, \beta > 0} \beta \epsilon + \beta \log \mathbb{E}_q \left[\exp \left(\frac{1}{\beta} \mathcal{L}(\tau, \theta) \right) \right], \quad (5)$$

where $\beta \in \mathbb{R}^{n_{eq}}$ is a dual auxiliary variable. Optimizing the second term is more commonly known as the Risk-Averse Optimal Control Problem [42, 12]. Note that a similar objective was derived in [1] through the ensemble planning approach. For sufficiently large β , it holds true via second order Taylor Expansion that,

$$\beta \log \mathbb{E}_q \left[\exp \left(\frac{1}{\beta} \mathcal{L}(\tau, \theta) \right) \right] \approx \mathbb{E}_q [\mathcal{L}(\tau, \theta)] + \frac{1}{2} \frac{1}{\beta} \mathbb{V}_q [\mathcal{L}(\tau, \theta)] \quad (6)$$

Assumption 1. (KL-ball ambiguity set)[23, 24] the surrogate distribution p exists in the ϵ -neighborhood of the target distribution, where $\mathbb{D}_{KL}(p \parallel q) \leq \epsilon$ given small $\epsilon > 0$.

To ensure that the above assumption holds, [23] first estimates the target distribution through a data-driven maximum-likelihood estimation approach prior to k-Nearest-Neighbor sampling from the resulting inferred distribution. The above assumption already holds if terms in Eq. 6 is approximated as,

$$\begin{aligned}\mathbb{E}_q[\mathcal{L}(\tau, \theta)] &\approx \mathcal{L}(\tau, \theta) \Big|_{\theta=\bar{\theta}} \\ \mathbb{V}_q[\mathcal{L}(\tau, \theta)] &\approx \nabla_{\theta} \mathcal{L}(\tau, \theta) \Big|_{\theta=\bar{\theta}}^{\top} \Sigma_{\theta\theta} \nabla_{\theta} \mathcal{L}(\tau, \theta) \Big|_{\theta=\bar{\theta}}\end{aligned}\quad (7)$$

where $\bar{\theta}$ is the parameter mean value, and $\Sigma_{\theta\theta} \in \mathbb{R}^{\Theta \times \Theta}$ is the variance about the parameter.

Note that Assumption 1 may not always be guaranteed in the general case, and is scarce in a variety of manipulation problems. Therefore, implementing standard DRO in a manipulation setting filled with high uncertainty, our priors may not be strong enough to ensure that Assumption 1 holds. In this work, we develop a novel approach that extends DRO-based methods without the need of asserting Assumption 1 by exactly measuring uncertainty through a guided task-aware variational inference-based method.

B. Stein Variational Gradient Descent

Given a random variable $\theta \in \Theta$, the goal of variational inference is to minimize the Kullback-Leibler (KL) divergence measure over the target distribution $q(\theta)$ and surrogate $p(\theta)$,

$$\begin{aligned}p^*(\theta) &= \arg \min_{p \in \mathcal{P}} \{\mathbb{D}_{KL}(p||q)\} \\ &\equiv \mathbb{E}_{\theta \sim p(\theta)} [\log p(\theta)] - \mathbb{E}_{\theta \sim p(\theta)} [\log q(\theta)] + \log z\end{aligned}\quad (8)$$

where z is a constant and thus ignored in the optimization. Stein Variational Gradient Descent (SGVD) [22] offers a solution through kernel and gradient descent methods that avoids committing to a parametric variational family $\mathcal{P}$ entirely.

SVGD initializes a set of points $\{\theta_0^i\}_{i=1}^N$ from a prior distribution $p(\theta) = \mathcal{N}(\hat{\theta}, \Sigma_{\theta})$ in θ . The points evolve according to the one-to-one mapping,

$$\theta_{t+1}^i \leftarrow \theta_t^i + \alpha \phi_{p,q}^*(\theta_t^i) \quad (9)$$

where $\phi_{p,q}^*(\theta_t^i)$ is a smooth function that captures the steepest descent direction that minimizes the KL-divergence measure at the t^{th} step, with T being the total number of SVGD iterations, and $\alpha > 0$ is a sufficiently small perturbation magnitude. We consider $\phi_{p,q}^*$ to be the solution to the steepest descent problem,

$$\begin{aligned}\phi_{p,q}^*(\cdot) &= \arg \min_{\phi \in \mathcal{H}^d} \{-\nabla_{\theta} D_{KL}(p||q) \mid \|\phi\|_{\mathcal{H}^d} \leq 1\} \\ &= \mathbb{E}_{\theta_t \sim q} [\mathcal{A}_q k(\theta_t, \cdot)]\end{aligned}\quad (10)$$

where $\mathcal{A}_q(\cdot) : \Theta \rightarrow \mathcal{S}_{\Theta}$ is Stein's identity computed for a universal positive definite kernel function $k : \Theta \times \Theta \rightarrow \mathbb{R}$ operating in a dense $\mathcal{H}^d$ in the space of continuous functions $C(\Theta, \mathbb{R}^d)$, where $\mathcal{H}^d$ is the corresponding Reproducing Kernel Hilbert Space (RKHS). The closed form solution to Eq. 10 is computed to be,

$$\phi_{p,q}^*(\cdot) = \mathbb{E}_{\theta_t \sim p} [k(\theta_t, \cdot) \nabla_{\theta} \log q(\theta_t) + \nabla_{\theta} k(\theta_t, \cdot)] \quad (11)$$

In practice, it is common to approximate the steepest descent direction $\phi_{p,q}^*$ using Monte-Carlo sampling over random variables θ ,

$$\phi_{q,p}^*(\cdot) \approx \frac{1}{N} \sum_{t=1}^N k(\theta_t^i, \cdot) \nabla_{\theta} \log q(\theta_t^i) + \nabla_{\theta} k(\theta_t^i, \cdot). \quad (12)$$

Over sufficient number of iterations of SVGD and number of samples N , the sampled parameters become a sufficient approximation of the target $q(\theta)$. In this work, we utilize SVGD as a guiding method to evolve parameter points initially sampled from a prior based on task-aware posterior gradients in order to deterministically model the parameter distribution subject to task-sensitivities and robustly solve contact-rich manipulation tasks. Note that while contact-based control problems are non-smooth in nature, we utilize a Lagrangian function that is smooth to take gradients over via SVGD.

IV. STEIN VARIATIONAL DISTRIBUTIONALLY ROBUST CONTROL

In this section, we derive the main contribution of this work that formulates the Distributionally Robust Control problem via Stein variational inference.

A. Variational Approximation of Task-Dependent Parameter Posterior

We first formulate the parameter posterior distribution via SVI. Given a set of randomly drawn parameter points $\{\theta\}_{i=1}^N \sim p(\theta)$, where $p(\theta)$ is a prior distribution, we approximate the task-dependent parameter posterior given a likelihood output $\mathcal{O}$ as,

$$p(\mathcal{O} \mid \theta) = \exp(\delta \mathcal{L}(\tau^*, \theta)) \quad (13)$$

where τ^* is the solved optimal trajectory, and the likelihood posterior $p(\mathcal{O} \mid \theta)$ relates to the target $q^*(\theta)$ as,

$$q^*(\theta) = \arg \min_{q \in \mathcal{Q}} D_{KL}(q(\theta) \parallel p(\mathcal{O}|\theta)). \quad (14)$$

Here, $\delta \mathcal{L} = \mathcal{L}(\tau^*, \theta) - \mathcal{L}(\tau^*, \theta^*)$ denotes the optimality gap, where θ^* is the true parameter and τ^* is the solved trajectory. Since θ^* is unavailable, we use the particle mean $\theta^* \approx \frac{1}{N} \sum_{i=1}^N \theta^i$ as a stable, low-variance reference for evaluating optimality, which is a standard and consistent estimator that provides stable low-variance reference for defining the optimality gap. The key idea is not to recover the true parameter, but to infer a task-consistent posterior whose particles capture the parameter variations that matter for producing a single robust control action under uncertainty. The likelihood output $\mathcal{O}$ encodes parameter-dependent outcomes in task optimality. The posterior is then calculated as a parameter-conditioned optimality gap distribution,

$$p'(\theta) = \frac{p(\mathcal{O}|\theta)p(\theta)}{z} = \frac{\exp(\delta \mathcal{L}(\tau^*, \theta))}{z} p(\theta) \quad (15)$$

where $z = \int_{\Theta} p(\mathcal{O}|\nu)p(\nu)d\nu$ which reveals the following log posterior,

$$\log p'(\theta) = \log p(\theta) + \delta \mathcal{L}(\tau, \theta) + \log z. \quad (16)$$

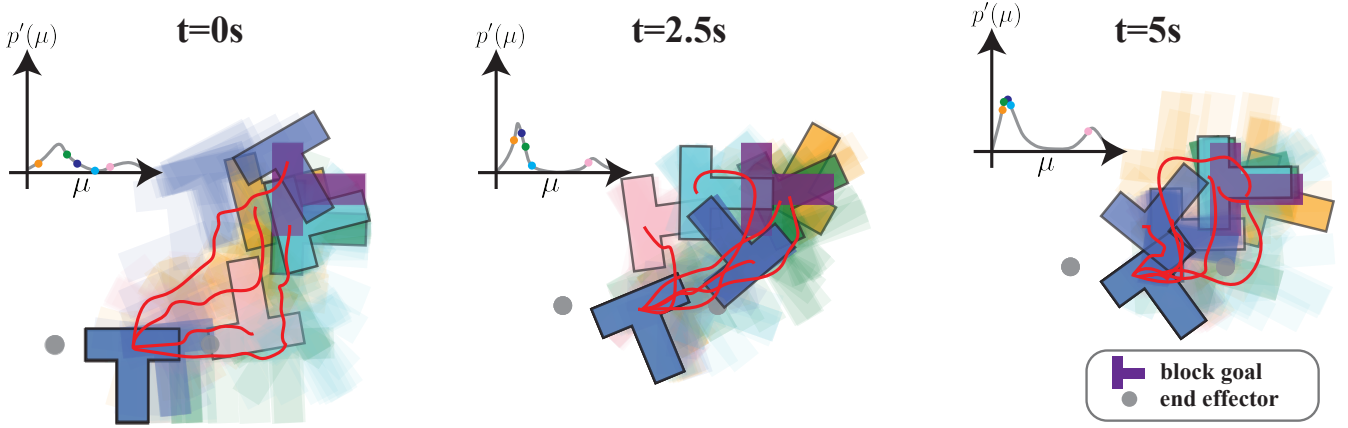


Fig. 2: Illustrative Example of SV-DRO. We visualize SV-DRO on the bimanual Push-T task, where two end effectors move a block to a goal under uncertain block-ground friction μ . Each translucent trajectory is a rollout induced by a different friction particle, initially sampled from the prior and then transported by SVGD using the task optimality gap. Rather than identifying friction before control or optimizing against a fixed worst case, SV-DRO reshapes uncertainty online toward parameter realizations most consequential for task success.

We now seek to utilize the posterior distribution to evolve the t^{th} sampled set of estimators $\{\theta_t^i\}_{i=1}^N$ drawn from a prior $p(\theta)$ by evaluating the steepest descent direction $\phi^* : \Theta \rightarrow S_{\Theta}$ using Stein’s identity. Given a prior $p(\theta)$ and a universal positive definite kernel $k : \Theta \times \Theta \rightarrow \mathbb{R}$, the resulting SGVD steepest descent solution to Eq. 10 is,

$$\begin{aligned} \phi^*(\cdot) &= \mathbb{E}_{\theta \sim p(\theta)} [k(\theta^i, \cdot) \nabla_{\theta} \log p'(\theta^i) + \nabla_{\theta} k(\theta^i, \cdot)] \\ &\approx \frac{1}{N} \sum_{i=1}^N k(\theta_t^i, \cdot) \nabla_{\theta} \log p'(\theta_t^i) + \nabla_{\theta} k(\theta_t^i, \cdot) \end{aligned} \quad (17)$$

where $k(\theta, \hat{\theta}) = \exp(-\|\theta - \hat{\theta}\|^2/h)$ is chosen as a Radius Basis Function (RBF) kernel [47].

Because the true physics model parameters θ^* are unknown and inestimable, we require a method that expands on existing DRO formulations [16, 23, 24] by *diversifying* the parametric inferences $\theta \sim p(\theta)$ through an evolving task-aware surrogate posterior that exactly adapts to the uncertainty present. The benefit of performing SVGD for parametric inference is that we can compute varied, diverse parameter-sensitive gradients to the task across a wide range θ , thus reasoning over the manipulation task over an *ensemble* of sampled parameter estimates that accurately estimates the task-aware posterior.

By utilizing SVI, we focus *limited computational resources* on task-critical uncertainties. By iteratively shaping the uncertainty distribution toward high-risk regions, we derive a robust controller that avoids the conservatism of worst-case methods while maintaining the sample efficiency of model-based approaches

B. Stein Variational DRO Formulation

Our goal in using Stein variational inference is to eliminate the restrictive Assumption 1 is not necessarily guaranteed for a weak prior $p(\theta)$. We therefore require a more robust method to optimize a control sequence u^* that reasons about *variations* of task sensitivities given that θ^* is unknown. To do so, we adopt the Stein-Variational approach [22] in the DRO formulation on the perturbed set of evolving estimators $\{\theta_t^i\}_{i=1}^N \forall t \in [0, T]$, where T is the total number of time steps.

We first define the evolving estimators $\hat{\theta}$ to follow the SVGD perturbation update,

$$\theta_{t+1}^i \leftarrow \theta_t^i + \alpha \phi^*(\theta_t^i) \quad (18)$$

where $\alpha > 0$ is the Stein evolve rate and kernel $k(\theta, \hat{\theta})$ is an RBF (commonly used in Stein Variational inference [17, 3]), and $\nabla_{\theta} \log p'(\theta)$ is the score of the log posterior in Eq. (16). The estimators are then implemented to solve the following optimization problem that extends Eq. 5,

$$\begin{aligned} \min_{\tau = \{(x_k, u_k, c_k) \mid \forall k \in [t, t+t_h]\}} \mathbb{E}_q[\mathcal{L}(\tau, \theta)] &\approx \frac{1}{N} \sum_{i=1}^N \mathcal{L}(\tau, \theta^i) \\ &= \mathcal{L}(\tau, \theta) \big|_{\theta=\bar{\theta}} + \gamma \frac{1}{N} \sum_{i=1}^N \delta \mathcal{L}(\tau, \theta_t^i) \end{aligned} \quad (19)$$

where $\bar{\theta}$ is the empirically averaged parameter value given the current parameter points, and $\mathcal{L} : \mathcal{T} \rightarrow \mathbb{R}$ where $\mathcal{T} \times \Theta \subset \mathcal{X} \times \mathcal{U} \times \mathcal{C}_{\theta}$ is the Lagrangian defined in Eq. 2. Here, tuning parameter $\gamma \in (0, 1]$ is included in order to control the trade-off between optimizing over the nominal parameters versus parameter variations. When $\gamma = 1$, the above formulation equates to the true Monte-Carlo approximation, but can be altered for numerical stability. When the auxiliary variable λ in Eq. 5 is sufficiently *large*, the variance term becomes negligible, yielding the approximation in Eq. 6 where the objective reduces to the expectation term, which recovers Eq. 19 under $\delta \mathcal{L}(\tau, \theta) \approx \mathcal{L}(\tau^*, \theta) - \mathcal{L}(\tau^*, \bar{\theta})$. We choose approximate $\mathbb{E}_{\theta \sim p(\theta)}[\mathcal{L}(\tau, \theta)]$ empirically, as this term captures the gap in task-dependent optimality with respect to parameter uncertainty, which we aim to minimize.

The core idea is to evaluate task-aware sensitivity gradients across parameter particles $\{\theta_t^i\}_{i=1}^N$, yielding a particle set that better represents variations that enlarge the optimality gap $\delta \mathcal{L}(\tau, \theta)$. Alg. 1 reflects the limited-feedback setting considered here: instead of replanning and updating particles after every control input, the controller executes the first H

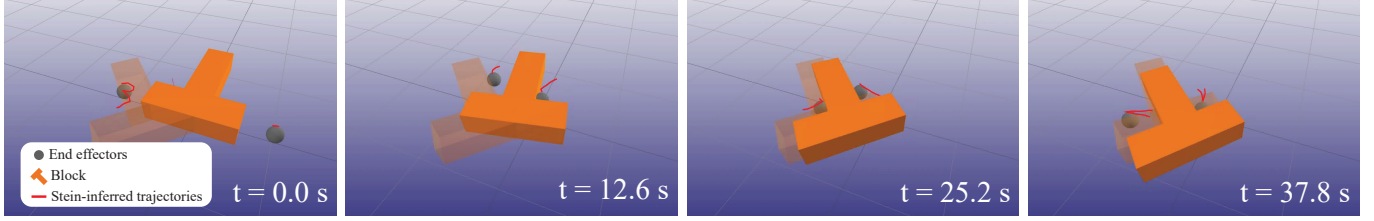


Fig. 3: Bimanual manipulation of Push-T with unknown mass distribution. Here, we demonstrate a control solution to the bimanual push-T task using our proposed SV-DGO. We observe that despite high uncertainty on mass, the bimanual end-effectors are able to successfully manipulate the T-block to the desired position while avoiding task-sensitive actions as a function of the uncertain parameters.

Method	Success % / Time to Completion (s)			
	Bimanual Push-T		Within-Hand Dynamic Positioning	
	≤ 10 cm	≤ 1 cm	≤ 10 cm	≤ 1 cm
SV-DRO (Ours)	93.25% / 16.45 $\pm$ 2.45	84.38% / 14.21 $\pm$ 2.22	100% / 12.00 $\pm$ 4.78	100% / 12.00 $\pm$ 4.78
DuST-MPC [3]	71.88% / 14.41 $\pm$ 3.58	59.38% / 13.32 $\pm$ 2.67	84.38% / 20.21 $\pm$ 5.62	46.88% / 20.18 $\pm$ 1.24
EMPPi [1]	43.75% / 18.88 $\pm$ 6.29	31.25% / 14.89 $\pm$ 3.12	59.38% / 22.44 $\pm$ 3.22	40.63% / 21.78 $\pm$ 3.43
MPC	28.13% / 19.56 $\pm$ 6.21	18.75% / 16.56 $\pm$ 2.21	31.25% / 18.40 $\pm$ 5.60	25% / 16.80 $\pm$ 5.14
DRO [16]	3.13% / 19.12 $\pm$ 3.13	0% / *	34.38% / 17.79 $\pm$ 5.00	28.13% / 18.55 $\pm$ 5.15

TABLE I: Simulated performance comparison over 32 trials of the bimanual Push-T and within-hand dynamic positioning experiments under unknown bounded uniform parameter distribution prior initialization. Results report success rate and completion time as mean $\pm$ standard deviation. * denotes insufficient completed trials to obtain completion-time statistics.

controls, receives a new measured state $\tilde{x}$ from the measured dynamics $\tilde{f}$, and then resolves Eq. 5. Thus, SVGD updates occur only at the lower feedback frequency, unlike estimation-driven Stein MPC methods that require continuous state-transition measurements for online system identification; here, particles evolve from task-sensitive optimality-gap gradients and must remain robust between intermittent measurements.

Fig. 2 illustrates this mechanism on bimanual Push-T. Each translucent rollout is a predicted object trajectory induced by a different friction particle, with SVGD transporting broad prior samples toward parameter regions that expose large task optimality gaps. The central principle of SV-DRO is therefore to *reshape* uncertainty online into a task-sensitive particle set, rather than fully identifying physical parameters before control.

V. RESULTS

In this section, we demonstrate and analyze the efficacy of our proposed approach on a variety of contact-based control problems. To do so, we aim to address the following research questions:

- Q1) Can our proposed approach improve task success rate for manipulation tasks under broad parameter uncertainty?
- Q2) How does Stein variational inference impact the convergence of the DRO problem?
- Q3) Does the task-aware parameter inference via Stein Variational DRO benefit over existing uncertainty mitigating approaches?
- Q4) How does the choice of kernel affect the performance of Stein Variational DRO?

Algorithm 1 Stein Variational Distributional Robust Control (SV-DRO)

Require: planning horizon t_h , parameter prior $p(\theta)$, N particles, step size α , samples $\{\theta_0^i\}_{i=1}^N \sim p(\theta)$, parameterized surrogate model $f_\theta(x_k, u_k)$, time duration T , control execution window H .

```

1:  $t = 0$ 
2: while  $t < T$  do
3:    $\tau^* \leftarrow \arg \min_{\tau} \mathcal{L}(\tau, \theta) \big|_{\theta=\bar{\theta}} + \gamma \frac{1}{N} \sum_{i=1}^N \delta \mathcal{L}(\tau, \theta_t^i)$ 
4:   for  $h = 0, \dots, H - 1$  do
5:     Apply  $u_{t+h}^*$  to the system
6:      $\tilde{x}_{t+h+1} \leftarrow \tilde{f}(\tilde{x}_{t+h}, u_{t+h}^*)$ 
7:   end for
8:    $x_{t+H} \leftarrow$  measured state after  $H$  executed controls
9:   Update SVI using the current task-aware posterior
10:   $\{\theta_{t+H}^i\}_{i=1}^N \leftarrow \text{SVGD}(\{\theta_t^i\}_{i=1}^N)$  ▷ Eq. 18
11:   $t \leftarrow t + H$ 
12: end while

```

Q5) What is the computation scaling for our Stein variational DRO approach with respect to particle parameter size and planning horizon?

We specifically answer the above questions through implementations of our approach in the following contact-based control problems:

- **Bimanual Push-T.** A bimanual manipulator is tasked with moving a T-shaped block with unknown mass and inertia to a goal pose through non-prehensile interactions.
- **Dynamic Within-Hand Positioning.** A robot manip-

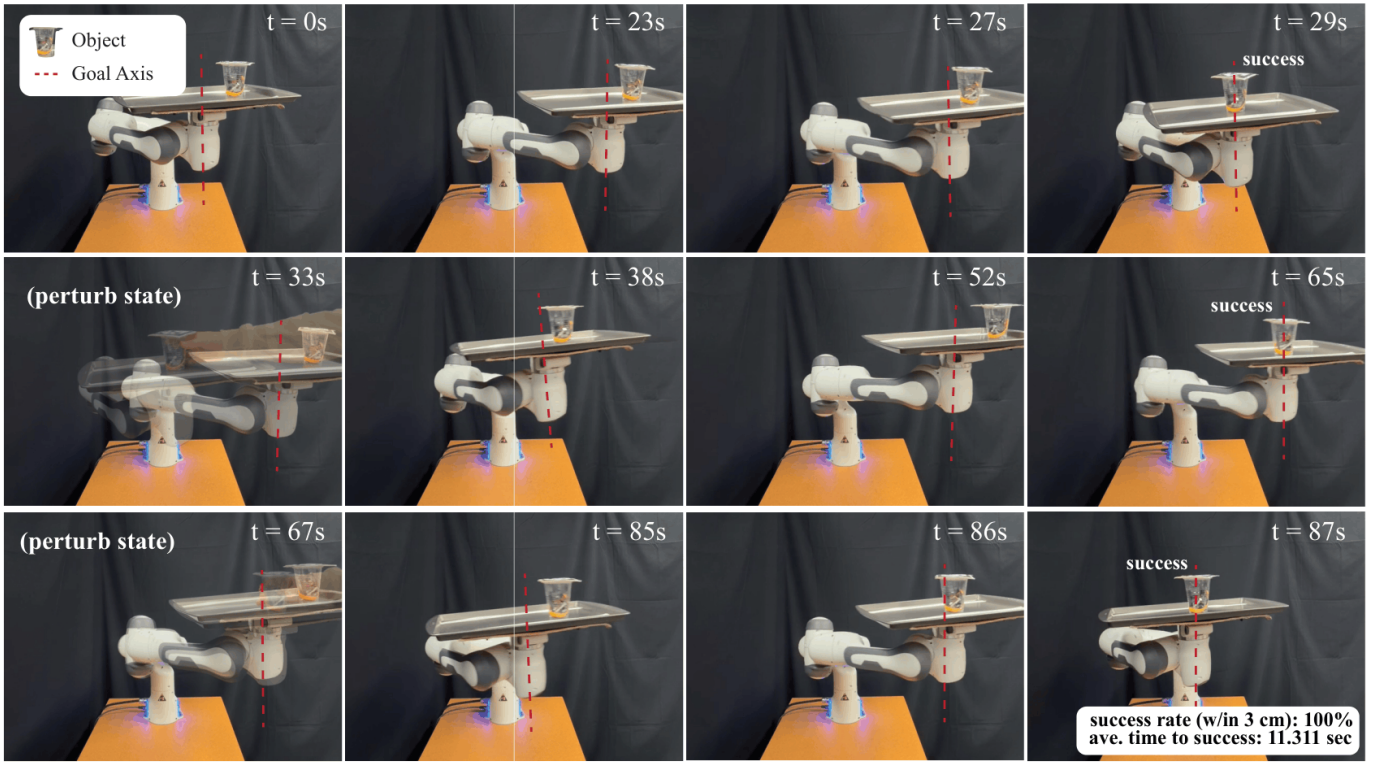


Fig. 4: Continual within-hand dynamic object positioning. SV-DRO repeatedly positions a cup with unknown mass distribution and tray friction over a continuous 3.5-minute run using AprilTag pose feedback [27]. After each successful placement within 1.5 cm of the goal axis, the cup is randomly perturbed without stopping the experiment. SV-DRO achieves the desired tolerance in approximately 100% of trials, demonstrating reliable dynamic sliding under persistent physical uncertainty.

ulates an object with unknown mass distribution (mass and inertia) and friction coefficient to another goal state within its tray hand. A reference diagram is shown in Fig. 1.

Implementation details for each control problem is detailed in supplementary material. We compare our approach to three following baselines:

- 1) Ensemble Model Predictive Control (EMPPI) [1, 25],
- 2) Dual Stein Variational control (DuSt-MPC) [3]
- 3) Conventional DRO [16, 23, 24],
- 4) and a naive MPC using the prior mean.

The key distinction between Stein-DRO and EMPPI lies in sample evolution, where Stein-DRO adaptively updates parameter particles via SVI shown in Eq. 18 using task-aware sensitivities, and EMPPI relies on static sampling from the prior distribution. DuSt-MPC uses SVGD primarily to perform online Bayesian system identification over dynamics parameters and controls from repeated state-feedback measurements. We address the above questions in this section with corresponding answers ‘A(-)’. We evaluate our experiments by initializing the algorithm’s starting parameters from a bounded uniform distribution (more details in supplementary material). For EMPPI and DRO, the sampled parameters are resampled after every new plan executed by the MPC controller as defined in Eq. 5. Our method initially samples randomly from this parameter prior distribution, followed by subsequent SVGD evolution of the samples.

A1. Improved Performance Under Uncertainty

As shown in Table I, SV-DRO achieves the highest success across bimanual Push-T and within-hand positioning while improving completion times. Success is measured by final object distance to the goal, matching the task costs used by each controller.

Fig. 3 shows SV-DRO coordinating two end effectors under uncertain physical parameters, while Fig. 4 demonstrates reliable within-hand positioning under uncertain cup mass distribution and tray friction. These tasks rely on intermittent contact and sliding, making them highly sensitive to latent dynamics and difficult for methods that require persistent state feedback for system identification, such as DuSt-MPC. SV-DRO instead transports particles toward parameters that expose high optimality-gap sensitivity, allowing control to adapt to task-relevant uncertainty rather than merely fitting recent measurements. Although our theoretical convergence guarantees (see supplementary for details) assumes stationary latent parameters, we also evaluate a discrete parameter shift in Fig. 5. The robot first approaches a goal with a top-heavy wine glass of unknown mass distribution, then the glass is filled with water, abruptly changing its dynamics before the robot serves it to a second goal. SV-DRO adapts to this sudden posterior shift and successfully completes the serving task, empirically demonstrating robustness beyond fixed-parameter settings.

A2. SVI Improves Task-Aware Convergence

We analyze the convergence properties of our approach in

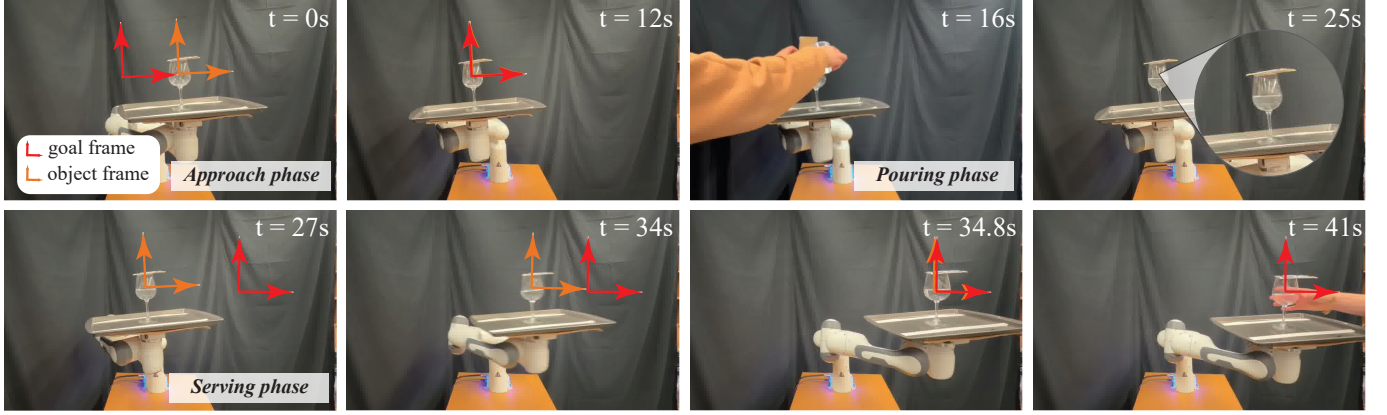


Fig. 5: Hardware demonstration of within-hand dynamic serving with a top-heavy object. The robot moves a wine glass to a goal pose and then serves it to a second goal after the glass is abruptly filled with water, inducing a sudden shift in its mass distribution. SV-DRO adapts to this posterior shift by using task-aware Stein variational gradients, enabling robust manipulation under changing physical uncertainty.

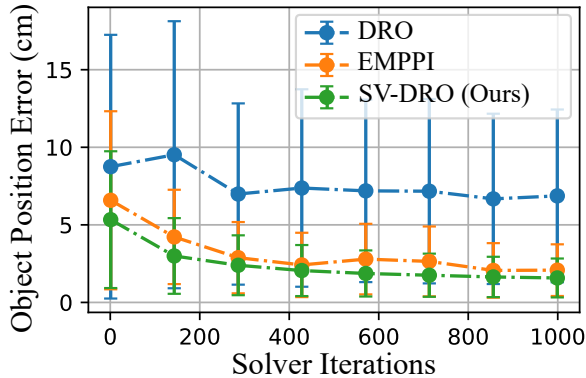


Fig. 6: Controller convergence based on prior distribution for Within-hand Positioning. The number of parameter samples, planning horizon, and prior distributions, are fixed. The results of this plot are evaluated over 5 trials per method with 5 random seeds. SV-DRO has improved reliability due to Stein inference.

solving the DRO problem to characterize the impact of SVI on the within-hand dynamic object positioning task. Our focus is only on solving the trajectory optimization component in Alg. 1 until convergence, given an initialization according to a range of parameter priors drawn from a uniform bounded distribution. We repeat this process over a set of 16 samples.

The results of the DRO objective convergence are shown in Fig. 6 in comparison to the baseline methods. We find that our approach has significantly less variance due to the deterministic evolution of the SVGD leading to consistent and reliable controllers despite the variability in parameter prior initialization. In contrast, the EMPPI approach experiences high variations of loss over solver iterations due to high stochasticity in the task-agnostic random sampling process itself, which is significantly more unreliable. The DRO approach converges to a worse-performing controller due its risk averse nature.

A3. Task-Aware Inference Improves Manipulation

We show the effect on the parameter inference technique on the overall success of solving a manipulation task. Our proposed method diversifies the initially sampled parameters

Kernel	Success Rate % (< 1 cm)	Completion Time (s)
RBF	100.0%	12.93 ± 4.78
$k = 1$	100.0%	16.54 ± 2.12
IMQ	100.0%	8.78 ± 1.34

TABLE II: Effect of kernel choice on SV-DRO performance over 32 trials for within-hand object positioning problem. We compare the standard RBF kernel, a constant kernel, and the inverse multiquadric (IMQ) kernel. The IMQ kernel achieves the fastest completion time, suggesting that its heavier-tailed distribution better preserves particle diversity over task-sensitive regions of the posterior and avoids premature mode collapse of Stein particles.

based on a task-aware posterior distribution via SVGD, which attributes to higher success rates indicated in Table I. However, what is the *exact* relationship between the discrepancy of the evolved samples to the unknown true parameters and the success rate of the manipulation task at hand?

We address this question by showing the initially sampled points from the parameter prior/posterior and their relationship to the respective performance of the within-hand dynamic object positioning experiment. Fig. 7 relates parameter discrepancy (object distance to goal) to task error for the within-hand positioning experiment. SV-DRO achieves consistently lower error across a wide range of discrepancies, showing that successful control does not require particles to converge to the true parameter θ^* . Instead, even with limited sensor feedback, SVGD transports particles toward task-relevant regions that expose high-sensitivity optimality-gap directions. In contrast, DRO and EMPPI degrade as the prior mean moves away from the true parameter, since their uncertainty representations remain conservative or task-agnostic rather than task-adaptive.

A4. Choice of Kernel for SV-DRO

We evaluate how the SVGD kernel affects SV-DRO performance. The kernel determines particle interaction during posterior transport, balancing attraction toward task-sensitive regions with repulsion that preserves diversity. As shown in Table II, all kernels achieve 100% success, indicating that final task completion is not highly kernel-sensitive in this experiment. However, completion time varies substantially: the constant kernel, which removes distance-dependent repulsion,

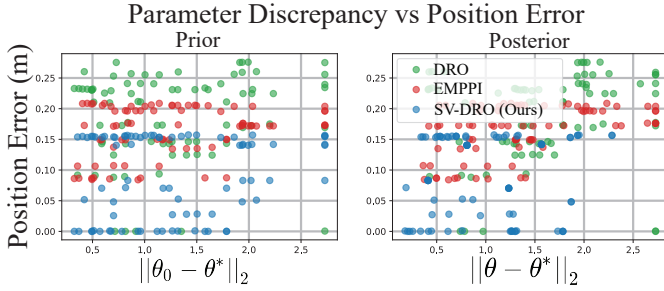


Fig. 7: Position Error over Parameter Samples. Here, we demonstrate the relationship between the performance (measured as the distance to the goal state) and the initial parameter mean discrepancy for the within-hand object positioning experiment. Our approach adapts to task-sensitive parameter uncertainty leading to lower error despite the discrepancy. We see weaker success rates for the samples from the EMPPI and DRO samples due to their task-agnostic posterior samples.

is slowest at ~ 16.54 s, suggesting that task-aware gradients alone can solve the task but less efficiently without geometry-aware particle diversity.

The IMQ kernel achieves the fastest completion time at ~ 8.78 s, outperforming the RBF kernel at ~ 12.93 s. This suggests that stronger long-range particle interactions help maintain diversity over separated task-sensitive regions and avoid premature particle collapse. Thus, kernel choice primarily affects the efficiency, rather than feasibility, of SV-DRO; heavier-tailed kernels such as IMQ may be preferable in contact-rich tasks where uncertainty-sensitive modes are distributed across parameter space.

A5. Computational Scaling of SVI

Last, we investigate the computational complexity the proposed SV-DRO approach for the within-hand dynamic positioning experiment. We evaluate the computational time scaling with respect to the parameter sample size N and the planning horizon t_h . All initial conditions, including random seeds and system state, are fixed. In evaluating over the parameter sample, we fix the planning horizon to 0.375 seconds to be constant for all computations, and vice versa. The results shown in Fig. 8 demonstrate improved computational performance with parallelization with the GPU (NVIDIA GeForce RTX 3080) compared to the CPU (AMD Ryzen Threadripper 3960X 24-Core Processor).

VI. LIMITATIONS

Model Error from Overlooked Parameters A fundamental limitation of our proposed work is its model-based nature. We require that the parameter uncertainty class be known in advance which has the risk of overlooked model parameters. In particular, inaccurate modeling, like a contact-model that is unable to capture the relevant frictional forces may have an adverse effect on the overall performance.

Model Complexity A limitation of our approach, and generally model-based approaches, is the need to construct reduced-order model to avoid mitigating computational complexity. In more general open-world settings, the need to build such models could become a bottleneck, especially when one needs

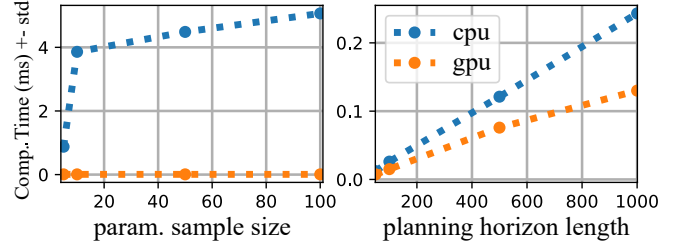


Fig. 8: Computational Analysis over Parameter Samples of SV-DRO. Here, we demonstrate the computational scaling of the proposed SV-DRO gradient step as a function of the parameter sample size N (left) and planning horizon t_h (right). With the CPU (AMD Ryzen Threadripper 3960X 24-Core Processor), we observe linear scaling with an increase of sample size, and with the GPU (NVIDIA GeForce RTX 3080), we observe constant computational scaling (about $77 \mu\text{s}$). For change in planning horizon length, we observe linear scaling with the CPU and sublinear scaling with GPU due to added parallelization.

to accurately characterize uncertainty. Model-free approaches excel in these scenarios but lack any form of interpretability. An open challenge is bridging the gap between these classes of methods, especially when reliability and robust performance is necessary. Future work will consider using the flexibility of Stein variational inference under less model-based restrictions. **Sensitivity to Informative Priors** As shown in Eq. 16, the prior score $\nabla_{\theta} \log p(\theta)$ enters every SVGD update, with the task-aware optimality-gap gradient in Alg. 1. Our bounded uniform prior has zero interior score, allowing task sensitivity to dominate particle transport. More informative priors, such as Gaussians centered at nominal estimates, can persistently pull particles toward the prior mean and suppress exploration of task-critical uncertainty directions. Future work will study structured priors to characterize this tradeoff.

VII. CONCLUSION

In summary, we present an approach that enables robots to reliably complete manipulation tasks under broad parameter uncertainty. Our approach combines Stein variational inference and Distributionally Robust Optimization to accurately model uncertainty in parameters and synthesize control solutions that enhance task performance. We demonstrate accurate modeling of task-aware posterior via Stein variational inference and show significant improvements in contact-rich manipulation problems. Last, we showcase the reliability of our approach in real-world settings via a continually running within-hand manipulation experiment.

VIII. ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation under award NSF FRR 2238066. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Ian Abraham, Ankur Handa, Nathan Ratliff, Kendall Lowrey, Todd D. Murphey, and Dieter Fox. Model-based generalization under parameter uncertainty using

- path integral control. *IEEE Robotics and Automation Letters*, 5(2):2864–2871, April 2020. ISSN 2377-3774. doi: 10.1109/Lra.2020.2972836. URL <http://dx.doi.org/10.1109/LRA.2020.2972836>.
- [2] Bo Ai, Stephen Tian, Haochen Shi, Yixuan Wang, Tobias Pfaff, Cheston Tan, Henrik I. Christensen, Hao Su, Jiajun Wu, and Yunzhu Li. A review of learning-based dynamics models for robotic manipulation. *Science Robotics*, 10(106):eadt1497, 2025. doi: 10.1126/scirobotics.adt1497. URL <https://www.science.org/doi/abs/10.1126/scirobotics.adt1497>.
- [3] Lucas Barcelos, Alexander Lambert, Rafael Oliveira, Paulo Borges, Byron Boots, and Fabio Ramos. Dual online stein variational inference for control and dynamics, 2021. URL <https://arxiv.org/abs/2103.12890>.
- [4] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion, 2024. URL <https://arxiv.org/abs/2303.04137>.
- [5] Embodiment Collaboration and Abby O’Neill et al. Open x-embodiment: Robotic learning datasets and rt-x models, 2025. URL <https://arxiv.org/abs/2310.08864>.
- [6] Gabriel Dulac-Arnold, Nir Levine, Daniel J. Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. An empirical investigation of the challenges of real-world reinforcement learning, 2021. URL <https://arxiv.org/abs/2003.11881>.
- [7] Rishi Bommasani et. al. On the opportunities and risks of foundation models, 2022. URL <https://arxiv.org/abs/2108.07258>.
- [8] Christian Fiedler, Carsten W. Scherer, and Sebastian Trimpe. Learning-enhanced robust controller synthesis with rigorous statistical and control-theoretic guarantees, 2021. URL <https://arxiv.org/abs/2105.03397>.
- [9] Astghik Hakobyan and Insoon Yang. Distributionally robust optimization with unscented transform for learning-based motion control in dynamic environments. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3225–3232, 2023. doi: 10.1109/ICRA48891.2023.10161246.
- [10] Nicolas Heess, Greg Wayne, David Silver, Timothy Lillicrap, Yuval Tassa, and Tom Erez. Learning continuous control policies by stochastic value gradients, 2015. URL <https://arxiv.org/abs/1510.09142>.
- [11] Yixuan Huang, Novella Alvina, Mohanraj Devendran Shanathi, and Tucker Hermans. Fail2progress: Learning from real-world robot failures with stein variational inference, 2025. URL <https://arxiv.org/abs/2509.01746>.
- [12] Ian, Rhodes, and David H. Jacobson. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic Control*, 18:124–131, 1973. URL <https://api.semanticscholar.org/CorpusID:478053>.
- [13] Michael G. Kapteyn, Karen E. Willcox, and Andy Philpott. *A Distributionally Robust Approach to Black-Box Optimization*. doi: 10.2514/6.2018-0666. URL <https://arc.aiaa.org/doi/abs/10.2514/6.2018-0666>.
- [14] Jaywant P. Kolhe, Md Shaheed, T.S. Chandar, and S.E. Talole. Robust control of robot manipulators based on uncertainty and disturbance estimation. *International Journal of Robust and Nonlinear Control*, 23(1):104–122, 2013. doi: <https://doi.org/10.1002/rnc.1823>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/rnc.1823>.
- [15] Vijay Konda and John Tsitsiklis. Actor-critic algorithms. In S. Solla, T. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999. URL https://proceedings.neurips.cc/paper_files/paper/1999/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.
- [16] Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. Distributionally robust optimization, 2025. URL <https://arxiv.org/abs/2411.02549>.
- [17] Alexander Lambert, Adam Fishman, Dieter Fox, Byron Boots, and Fabio Ramos. Stein variational model predictive control, 2021. URL <https://arxiv.org/abs/2011.07641>.
- [18] Yewon Lee, Andrew Z. Li, Philip Huang, Eric Heiden, Krishna Murthy Jatavallabhula, Fabian Damken, Kevin Smith, Derek Nowrouzezahrai, Fabio Ramos, and Florian Shkurti. Stamp: Differentiable task and motion planning via stein variational gradient descent, 2024. URL <https://arxiv.org/abs/2310.01775>.
- [19] Yong-Hoon Lee, Keuntae Kim, Jaehyun Park, Chung Hyuk Park, and Hae-Won Park. Soft contact model for robust locomotion of legged robots.
- [20] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies, 2016. URL <https://arxiv.org/abs/1504.00702>.
- [21] Lei Li, Yingzhou Li, Jian-Guo Liu, Zibu Liu, and Jianfeng Lu. A stochastic version of stein variational gradient descent for efficient sampling. *Communications in Applied Mathematics and Computational Science*, 15(1):37–63, June 2020. ISSN 1559-3940. doi: 10.2140/camcos.2020.15.37. URL <http://dx.doi.org/10.2140/camcos.2020.15.37>.
- [22] Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm, 2019. URL <https://arxiv.org/abs/1608.04471>.
- [23] Rui Liu, Guangyao Shi, and Pratap Tokekar. Data-driven distributionally robust optimal control with state-dependent noise, 2025. URL <https://arxiv.org/abs/2303.02293>.
- [24] Kehan Long, Yinzhuang Yi, Zhirui Dai, Sylvia Herbert, Jorge Cortés, and Nikolay Atanasov. Sensor-based distributionally robust control for safe robot navigation in dynamic environments, 2025. URL <https://arxiv.org/abs/2405.18251>.
- [25] Kendall Lowrey, Aravind Rajeswaran, Sham Kakade, Emanuel Todorov, and Igor Mordatch. Plan online, learn offline: Efficient learning and exploration via model-based control, 2019. URL <https://arxiv.org/abs/1811.01848>.

- [26] Igor Mordatch, Kendall Lowrey, and Emanuel Todorov. Ensemble-cio: Full-body dynamic motion planning that transfers to physical humanoids. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5307–5314, 2015. doi: 10.1109/IROS.2015.7354126.
- [27] Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407, 2011. doi: 10.1109/ICRA.2011.5979561.
- [28] OpenAI, Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, Jonas Schneider, Szymon Sidor, Josh Tobin, Peter Welinder, Lilian Weng, and Wojciech Zaremba. Learning dexterous in-hand manipulation, 2019. URL <https://arxiv.org/abs/1808.00177>.
- [29] Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3803–3810, 2018. doi: 10.1109/ICRA.2018.8460528.
- [30] Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-real transfer of robotic control with dynamics randomization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, page 3803–3810. IEEE, May 2018. doi: 10.1109/icra.2018.8460528. URL <http://dx.doi.org/10.1109/ICRA.2018.8460528>.
- [31] Thomas Power and Dmitry Berenson. Constrained stein variational trajectory optimization, 2024. URL <https://arxiv.org/abs/2308.12110>.
- [32] Hamed Rahimian and Sanjay Mehrotra. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization*, 3:1–85, July 2022. ISSN 2777-5860. doi: 10.5802/ojmo.15. URL <http://dx.doi.org/10.5802/ojmo.15>.
- [33] Aravind Rajeswaran, Sarvejit Ghotra, Balaraman Ravindran, and Sergey Levine. Epopt: Learning robust neural network policies using model ensembles, 2017. URL <https://arxiv.org/abs/1610.01283>.
- [34] Hrishikesh Sathyanarayan and Ian Abraham. Behavior synthesis via contact-aware fisher information maximization, 2025. URL <https://arxiv.org/abs/2505.12214>.
- [35] Charles Stein, Persi Diaconis, Susan Holmes, and Gesine Reinert. Use of exchangeable pairs in the analysis of simulations. 46, 01 2004. doi: 10.1214/lnms/1196283797.
- [36] Tyler Summers. Distributionally robust sampling-based motion planning under uncertainty. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6518–6523, 2018. doi: 10.1109/IROS.2018.8593893.
- [37] TRI LBM Team and Jose Barreiros et. al. A careful examination of large behavior models for multitask dexterous manipulation, 2025. URL <https://arxiv.org/abs/2507.05331>.
- [38] Gabriele Tiboni, Pascal Klink, Jan Peters, Tatiana Tommasi, Carlo D’Eramo, and Georgia Chalvatzaki. Domain randomization via entropy maximization, 2024. URL <https://arxiv.org/abs/2311.01885>.
- [39] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 23–30, 2017. doi: 10.1109/IROS.2017.8202133.
- [40] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world, 2017. URL <https://arxiv.org/abs/1703.06907>.
- [41] Victor Vantilborgh, Hrishikesh Sathyanarayan, Guillaume Crevecoeur, Ian Abraham, and Tom Lefebvre. Dual control reference generation for optimal pick-and-place execution under payload uncertainty, 2025. URL <https://arxiv.org/abs/2510.20483>.
- [42] P. Whittle. Risk-sensitive linear/quadratic/gaussian control. *Advances in Applied Probability*, 13(4):764–777, 1981. doi: 10.2307/1426972.
- [43] Liang Yan and Tao Zhou. Stein variational gradient descent with local approximations. *Computer Methods in Applied Mechanics and Engineering*, 386:114087, 2021. ISSN 0045-7825. doi: <https://doi.org/10.1016/j.cma.2021.114087>. URL <https://www.sciencedirect.com/science/article/pii/S0045782521004187>.
- [44] Dan Zhang and Bin Wei. A review on model reference adaptive control of robotic manipulators. *Annual Reviews in Control*, 43:188–198, 2017. ISSN 1367-5788. doi: <https://doi.org/10.1016/j.arcontrol.2017.02.002>. URL <https://www.sciencedirect.com/science/article/pii/S1367578816301110>.
- [45] Junzi Zhang, Jongho Kim, Brendan O’Donoghue, and Stephen Boyd. Sample efficient reinforcement learning with reinforce, 2020. URL <https://arxiv.org/abs/2010.11364>.
- [46] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 737–744, 2020. doi: 10.1109/SSCI47803.2020.9308468.
- [47] Jingwei Zhuo, Chang Liu, Jiaxin Shi, Jun Zhu, Ning Chen, and Bo Zhang. Message passing stein variational gradient descent, 2018. URL <https://arxiv.org/abs/1711.04425>.
- [48] Nadina O. Zweifel, Nicholas E. Bush, Ian Abraham, Todd D. Murphey, and Mitra J. Z. Hartmann. A dynamical model for generating synthetic data to quantify active tactile sensing behavior in the rat. *Proceedings of the National Academy of Sciences*, 118(27):e2011905118, 2021. doi: 10.1073/pnas.2011905118. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2011905118>.