

# Heuristic-free Optimization of Force-Controlled Robot Search Strategies in Stochastic Environments

Benjamin Alt<sup>1,2</sup>, Darko Katic<sup>1</sup>, Rainer Jäkel<sup>1</sup> and Michael Beetz<sup>2</sup>

**Abstract**—In both industrial and service domains, a central benefit of the use of robots is their ability to quickly and reliably execute repetitive tasks. However, even relatively simple peg-in-hole tasks are typically subject to stochastic variations, requiring search motions to find relevant features such as holes. While search improves robustness, it comes at the cost of increased runtime: More exhaustive search will maximize the probability of successfully executing a given task, but will significantly delay any downstream tasks. This trade-off is typically resolved by human experts according to simple heuristics, which are rarely optimal. This paper introduces an automatic, data-driven and heuristic-free approach to optimize robot search strategies. By training a neural model of the search strategy on a large set of simulated stochastic environments, conditioning it on few real-world examples and inverting the model, we can infer search strategies which adapt to the time-variant characteristics of the underlying probability distributions, while requiring very few real-world measurements. We evaluate our approach on two different industrial robots in the context of spiral and probe search for THT electronics assembly.\*

## I. INTRODUCTION

Despite years of research, manipulating objects whose pose can only be determined with uncertainty is still very challenging for robots. In service robotics, examples include the manipulation of occluded objects or imprecise pose estimation due to noisy perception. Industrial applications often require robots to manipulate objects whose pose in the workspace is known only within given tolerances and whose precise pose varies stochastically between cycles. In the context of electronics assembly, for example, the poor positioning accuracy of conveyor belts, manufacturing tolerances of parts from different suppliers or wear and tear of materials make it impossible to precisely know the pose of connectors or holes during offline programming [1]. Force-controlled search strategies, which probe the environment with a defined series of motions until a change in forces is detected, are a commonly used method for resolving such uncertainties [2]. While greatly increasing the robustness of robot tasks, however, force-controlled search comes at the cost of increased time required to complete the task. Determining an optimal set of search motion parameters, which maximizes robustness while keeping execution times minimal, currently requires lengthy parameter tuning by human robot programmers. We posit that a method for the

This work was supported by the German Federal Ministry of Education and Research under the grant 01DR19001B.

<sup>1</sup>ArtiMinds Robotics, Karlsruhe, Germany  
benjamin.alt@artiminds.com

<sup>2</sup>Institute for Artificial Intelligence, University of Bremen, Germany

\*See [github.com/benjaminalt/dpse](https://github.com/benjaminalt/dpse) for code and data.

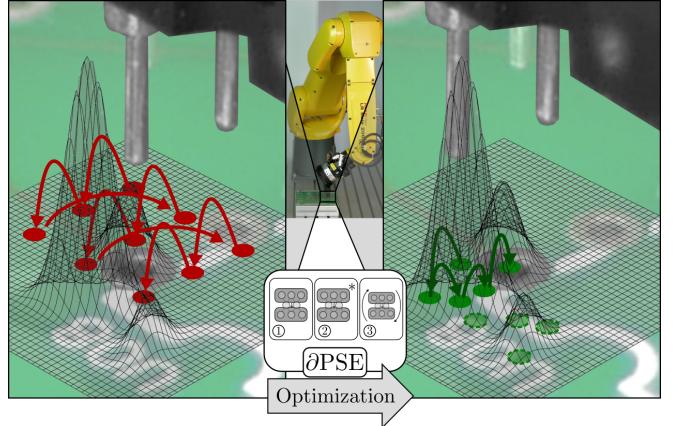


Fig. 1. Differentiable Programming in Stochastic Environments (DPSE) leverages transfer learning and differentiable robot program representations to optimize search motions in the presence of stochastic noise.

automatic parameterization of search strategies must meet four crucial requirements to be applicable for most real-world use cases:

- 1) *Heuristic-freeness*: It must make no prior assumptions about the underlying stochastic processes.
- 2) *Data efficiency*: It must require only unsupervised or self-supervised data. The amount of real-world training data must be small and data collection must not interfere with the completion of regular tasks, such as the ongoing operation of a production line.
- 3) *Handling nonstationary processes*: It must be robust to nonstationary noise processes whose characteristics change over time.
- 4) *Black-box optimization*: It must be agnostic with respect to the concrete robot program representation to ensure cross-domain applicability and compatibility with various robot frameworks.

In this paper, we introduce Differentiable Programming in Stochastic Environments (DPSE), a heuristic-free, data-efficient black-box approach to optimizing robot search motions even in the face of nonstationary stochastic processes.

Our contribution is threefold: (1) A method for *learning predictive models of robot skills* which capture the stochastic characteristics of the environments they are executed in. To that end, we present a transfer-learning based approach to efficiently condition such models on the stochastic environment at hand. By pre-training in simulated environments and fine-tuning on a small dataset of real-world action executions, the learned models predict an expected robot trajectory given

the robot skill’s parameters (e.g. the search pattern of a probe search) and the (learned) probability distributions governing the environment. We show that continuous fine-tuning can keep the learned models synchronized with time-varying, nonstationary stochastic environments.

(2) A method to *optimize motion parameters* to ensure minimal search time and maximal robustness in the face of nonstationary stochastic environments. We embed these environment-conditioned models in a state-of-the-art framework for differentiable robot programming [3], in which robot programs can be expressed as differentiable computation graphs. A gradient-based optimizer jointly optimizes the program’s input parameters with respect to relevant metrics such as cycle time or failure rate.

(3) *Experimental validation* of our method in two real-world mechanical and electronics assembly tasks for both stationary and nonstationary environments, and empirical comparison against several meta-learning approaches.

## II. RELATED WORK

*a) Robot program parameter optimization:* Several black-box optimizers for robot program parameters have been proposed, most often relying on zero-order methods such as evolutionary algorithms [4], [5], whose requirement of repeatedly executing the robot program for every optimization renders them impractical for real-world application. Bayesian optimization [6]–[8] avoids this, but requires the goal function to be known at data collection time, is difficult to extend to nonstationary processes [9], [10] and cannot exploit gradient information. First-order neural network (NN)-based optimizers have been proposed [11], [12], but remain impractical in productive environments due to their data requirements and reliance on supervised or reinforcement learning.

*b) Differentiable programming ( $\partial\mathcal{P}$ ):* In  $\partial\mathcal{P}$ , programs are composed of differentiable operations, which permits the first-order optimization of a program’s parameters with respect to its outputs [13]. PDP [14] is a  $\partial\mathcal{P}$  framework for differentiable motion control. However, PDP is not black-box, cannot represent hierarchical programs and the learned parameters become outdated in nonstationary environments. SPI [3] is a black-box  $\partial\mathcal{P}$  framework for representing robot programs, which, like Bayesian optimization, relies on surrogate models learned via unsupervised learning. It exploits gradients for efficient optimization, but assumes stationarity and requires substantial amounts of real-world data. We leverage SPI for optimization, but avoid its stationarity assumption and drastically reduce the amount of required real-world data.

*c) Transfer learning (TL):* TL is concerned with improving the performance of systems trained on some source task on different, but related target tasks [15]. In *sequential TL* schemes, a NN first learns a shared set of features on a large source dataset before fine-tuning on a much smaller target dataset [16]. In natural language processing [17]–[21] and computer vision [22]–[24], unsupervised pretraining outperforms learning from scratch [25], [26], particularly for

small target datasets [27]. In robotics, sequential TL has been applied for inter-robot learning [28], Sim2Real adaption [29] or few-shot learning [30], [31]. We show that in the context of parameter optimization, sequential TL not only greatly reduces the amount of required real-world data, but also that continuous fine-tuning as a form of “lifelong learning” enables parameter optimization in nonstationary stochastic environments.

*d) Meta-learning:* Meta-learning is concerned with efficiently learning to solve unseen tasks and has been applied with particular success in few-shot problems [32], [33]. While some approaches such as Meta Networks [34] or Prototypical Networks [35] propose specific architectures, model-agnostic methods such as MAML [36] or Reptile [37] optimize arbitrary networks by learning initial network parameters particularly conducive to test-time fine-tuning. We provide an evaluation of several model-agnostic meta-learners as alternatives to our proposed TL-based approach in the context of search strategy parameter optimization.

## III. PARAMETER OPTIMIZATION IN STOCHASTIC ENVIRONMENTS

### A. Definitions and Notation

As an example for an industrial assembly task subject to stochastic variations, we consider a peg-in-hole task, where the actual pose of the hole can be described by a random variable  $H_t$  drawn from the (possibly nonstationary) stochastic process  $\{H_t\}$ . Because the actual hole pose at time  $t$  is uncertain, an industrial robot will require the use of search strategies to perform this task reliably. We define a *search strategy* as a robot program which accepts a vector of parameters  $x \in \mathcal{X}$  and executes a sequence of end effector motions until a hole has been found or some other termination criterion was reached. Intuitively, a *robot program* can be defined as a function  $\pi$  mapping some program parameters  $x$  to the tool center point (TCP) trajectory  $y$  resulting from executing it. However, because this trajectory depends on the uncertain pose of the hole, we model it instead as a random variable  $Y_t$ , which takes values  $y \in \mathcal{Y}$  according to  $P(Y_t|\pi, x, H_t)$ , the probability distribution over trajectories resulting from the execution of program  $\pi$  at time  $t$  with inputs  $x$  and hole distribution  $H_t$ .

### B. Parameter Optimization via Differentiable Programming

In the context of robotics,  $\partial\mathcal{P}$  approaches center around learning or constructing a differentiable representation of  $\pi$ , which permits to backpropagate losses over the program’s outputs to its inputs  $x$ . SPI [3] approximates  $P(Y_t|\pi, x)$  by training an auxiliary, differentiable graph of NNs (“shadow program”,  $\hat{\pi}$ ) via unsupervised learning from past executions of  $\pi$ . It then optimizes the program’s input parameters by *inverting* the shadow program via gradient descent in the shadow program’s input space. This optimization can be performed with respect to some user-defined function  $\mathcal{L}$  of the shadow program’s output (the expected trajectory  $\hat{y}$ ), such as cycle time. SPI’s compatibility with most existing robot program or skill frameworks and exclusive reliance on

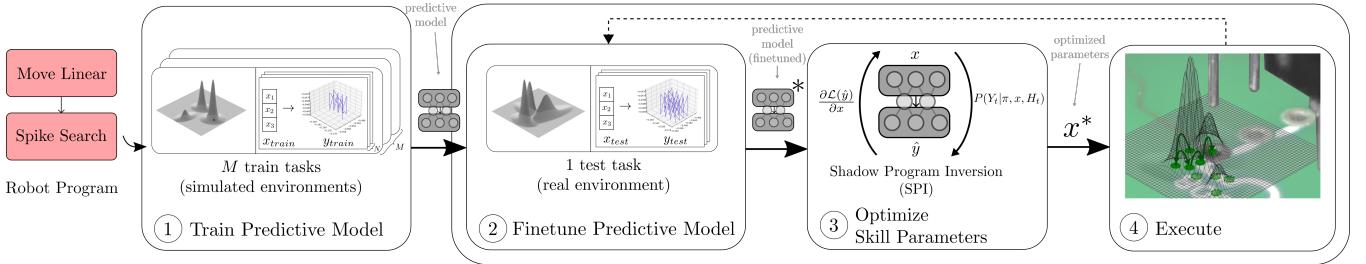


Fig. 2. Overview of our proposed approach.  $\partial\text{PSE}$  enables the data-efficient, heuristic-free optimization of arbitrary parametric search strategies by pretraining a predictive model (“shadow program”) of the search strategy on a large simulated dataset (1), finetuning it on a few samples of the concrete real-world environment (2), and inferring optimal search parameters via SPI (3). By repeatedly finetuning on data collected at test time (4), nonstationary noise processes can be addressed.

unsupervised learning are highly advantageous in industrial robot applications. SPI is a *heuristic-free* and *black-box* optimizer according to the criteria outlined in section I. In its model training phase, however, SPI requires training tuples  $(x_i, y_i)$  which cover the region of the input space  $\mathcal{X}$  relevant for optimization. This effectively requires sampling the input space during data collection, which is impractical in a running production line. Moreover, if features of the environment such as the poses of holes follow *nonstationary* stochastic processes, the learned shadow program immediately becomes outdated as the probability distributions change over time.

### C. Environment-Conditioned Skill Models

Transfer-learning methods are capable of efficiently adapting learned models to changing data distributions [38]. By treating the model learning phase of SPI as a TL problem and applying a pretraining and finetuning technique, we obtain a heuristic-free black-box differentiable program representation which supports the optimization of search motions in the face of nonstationary noise processes with high data efficiency.

We first define a *task*  $\mathcal{T} = \{\{H_t\}, P(Y_t|\pi, x, H_t)\}$ : For a stochastic process  $\{H_t\}$  generating hole distributions, SPI must learn the probability distribution over the resulting trajectories when executing program  $\pi$  with inputs  $x$  in this environment. The corresponding *task dataset*  $D_{\mathcal{T}_0} = \{(x_0, y_0), \dots, (x_N, y_N)\}$  for a task  $\mathcal{T}_0$  consists of input-trajectory pairs collected by executing  $\pi$   $N$  times, each time sampling a new hole distribution  $H_n$  from  $\{H_t\}_0$ . Following a pretraining/finetuning regime, we propose to pretrain SPI’s shadow program on a large *source dataset*  $D_S = \bigcup_{m=0}^M D_{\mathcal{T}_m}$ , the collection of  $M$  task datasets, each for a different hole-distribution-generating process  $\{H_t\}_m$  (see fig. 2 (1)). The shadow model can then be finetuned on the much smaller *target dataset*  $D_T = \{D_{\mathcal{T}_{curr}}\}$  (see fig. 2 (2)), containing only data from the current task  $\mathcal{T}_{curr}$  – in the case of a running production line, the input-trajectory pairs from the last  $N$  executions of the program. The core intuition is to learn a prior over many possible environments (hole-distribution-generating processes,  $\{H_t\}_m$ ) offline and finetune on the concrete process at hand.

**1) Data efficiency:** We found that for most peg-in-hole tasks under uncertainty, only  $N=128$  samples per task are required to effectively finetune on the concrete task. Moreover, we show that pretraining on a large ( $M=1000$ ,  $N=128$ ) source dataset collected in a simulated environment and finetuning on one single real-world target task bridges the sim-to-real-gap sufficiently for SPI to perform meaningful parameter optimization in the real-world environment. Both spiral and probe search strategies considered in experiments IV-A and IV-B involve force-sensitive interactions with the environment; in both cases, pretraining exclusively on simulated data sufficed to find near-optimal parameterizations in the real world. This efficiency can be attributed to the fact that after pretraining, the shadow program can rely on useful priors at two levels: The differentiable motion planner at the heart of the SPI shadow program architecture provides a strong prior for the expected trajectory in an ideal environment, and the new pretraining step trains a prior over a wide range of possible environments, providing a very good initialization of the shadow program for finetuning.

**2) Passive data collection:** As a corollary of these strong priors, the proposed pretraining/finetuning regime avoids the need to sample the parameter space in the real production line. Because the relationship between program inputs and outputs is learned across a wide variety of environments and for inputs  $x$  sampled from the entire parameter space, the finetuning dataset  $D_T$  does not require such diversity anymore. In fact, we find that SPI still performs meaningful parameter optimization even if all finetuning examples in  $D_T$  contain the same inputs, i.e.  $D_T = \{D_{\mathcal{T}_{curr}}\} = \{(x_0, y_0), \dots, (x_0, y_N)\}$ . Consequently, the required 128 real-world finetuning examples can be collected in a completely passive manner by simply collecting robot trajectories from the running assembly line, without needing to sample a diverse set of program parameters and potentially disrupting production.

**3) Continuous learning and optimization for nonstationary processes:** Nonstationary processes require continuous re-optimization of program parameters, ideally at every timestep  $t$ . With the proposed pretraining/finetuning regime, this becomes straightforward: After finetuning once, running SPI’s optimization step and updating the program with the optimized parameters, the next trajectory at timestep  $t+1$

can be observed and added to the finetuning dataset. SPI's shadow program can then be finetuned again and the optimal parameters for timestep  $t + 2$  can be computed. Repeating this cycle of finetuning the shadow program and optimizing program parameters ensures that the shadow program's estimation of  $P(Y_t|\pi, x, H_t)$  does not deteriorate as  $t$  increases. We implement  $D_T = \{D_{\mathcal{T}_{curr}}\}$  as a ring buffer of fixed size  $N=128$  to ensure that outdated data is eventually removed from the finetuning dataset. To avoid catastrophic forgetting, we finetune the original pretrained model at each iteration instead of repeatedly finetuning the finetuned model.

#### IV. EXPERIMENTS

We validate our approach by comparing its performance against state-of-the-art baselines on two real-world assembly tasks involving spiral and probe search strategies, respectively. We also evaluate its performance on several simulated nonstationary processes and provide an empirical analysis of our TL scheme with respect to possible meta-learning based alternatives.

##### A. Spiral Search (stationary)

We consider the mechanical assembly task of inserting a tightly-fitting cylinder into a hydraulic valve (see fig. 3). To simulate realistic process variances, the position of the valve body in the XY-plane is sampled from a bivariate Gaussian mixture with six components, forming the stationary process  $\{H_t\} = H_t = \sum_{i=1}^6 w_i \mathcal{N}(\mu_i, \Sigma_i)$ . To successfully perform the insertion, the robot follows a *spiral search* strategy, maintaining a constant pushing force against the surface until dropping into the hole. We seek to optimize the parameters of the search strategy (spiral position, orientation, extents, number of windings, velocity and acceleration) to minimize a linear combination of cycle time  $\mathcal{L}_{cycle}(y)$  and failure probability  $\mathcal{L}_{fail}(y)$ , both of which can be expressed in terms of the expected trajectory  $y$  output by  $\partial\text{PSE}$ 's shadow program.  $\partial\text{PSE}$  is pretrained on a source dataset collected by simulating spiral searches on  $M_{train} = 1000$  different Gaussian mixtures with  $N_{train} = 128$  sampled hole poses each, and finetuned on  $N_{test} = 128$  training examples for the  $M_{test} = 1$  concrete hole distributions at hand.\* We compare against the following baselines:

- 1) A fixed parameterization by a human expert
- 2) A near-optimal Principal Component Analysis (PCA)-based heuristic fitting the spiral orientation and extents to the ground-truth valve body poses
- 3)  $\mu + \lambda$  nondominated sorting genetic algorithm (NSGA-II) [39] with  $\mu = \lambda = 25$

We repeat the experiment for 10 different distributions  $\{H_t\}$  in a purely simulated environment, and for 6 different  $\{H_t\}$  on a real UR5 robot.

The results are shown in figure 3.  $\partial\text{PSE}$  provides near-perfect success rates and significantly outperforms NSGA-II and the human expert, effectively replicating the results in [3]

\*Simulations were conducted in a lightweight scripted environment modelling Newtonian physics, damping and friction.

with significantly fewer real-world training examples. With respect to cycle time,  $\partial\text{PSE}$  outperforms both the human and NSGA-II on the real task. The poor real-world performance of NSGA-II illustrates the fundamental disadvantage of most heuristic-free, zero-order black-box optimizers, which they require a large amount of program executions to converge on a good solution. Here, was allowed 250 executions per test distribution, nearly twice  $\partial\text{PSE}$ 's real-world data requirement of 128, suggesting that  $\partial\text{PSE}$  owes its data efficiency at least partly to its exploitation of gradient information.

##### B. Probe Search (stationary)

To evaluate  $\partial\text{PSE}$  for a more complex search strategy, we consider the assembly of through-hole technology (THT) electronics components, where the pose of mounting holes on a printed circuit board (PCB) is subject to stationary noise  $\{H_t\}$  with an a priori unknown distribution, e.g. from imprecise positioning on a conveyor belt. To avoid scratching the surface of the PCB, *probe search* repeatedly touches the surface according to a predefined search pattern until the THT component's pins drop into the holes (see fig. 2 (4)). The search pattern is typically defined as a regular grid, though the search strategy can be significantly optimized by tailoring the pattern to the underlying noise distribution. Finding the optimal pattern corresponds to a high-dimensional optimization problem over the 32-dimensional input vector of the probe search strategy, describing the 16 touch points in the XY-plane. We use  $\partial\text{PSE}$  to automatically optimize the touch points without assuming knowledge about the hole distribution.  $\partial\text{PSE}$  is trained on  $M_{train} = 1000$  simulated training tasks (hole distributions) with  $N_{train} = 128$  probe searches each and finetuned on  $N_{test} = 128$  probe searches over  $M_{test} = 1$  concrete hole distribution. We compare against the following baselines:

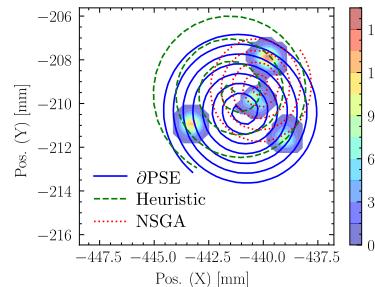
- 1) A fixed 4x4 grid covering the complete search region
- 2) A heuristic tailored specifically to probe search, which fits a 16-mode Gaussian Mixture Model (GMM) to the last points of those trajectories in the finetuning dataset where the search was successful
- 3)  $\mu + \lambda$  NSGA-II with  $\mu = \lambda = 100$  (sim) or  $\mu = \lambda = 30$  (real)

The experiment is repeated for 10 different multimodal Gaussians  $\{H_t\}$  in a simulated environment, and for 6 additional distributions on a real Fanuc industrial robot. Both  $\partial\text{PSE}$  and NSGA-II minimize  $\mathcal{L}_{fail}(y)$ , the probability of search failure. We also assess the effects of adding additional regularization to  $\partial\text{PSE}$ , penalizing either the L1 distance from the initial parameterization ( $\mathcal{L}_{init}$ , a heuristic-free regularizer) or the smallest distance between any two optimized touch points ( $\mathcal{L}_{cdist}$ , a heuristic regularizer specific to probe search).

The results are summarized in figure 4. With both regularizers, the search patterns produced by  $\partial\text{PSE}$  reflect the underlying hole distribution without overfitting to its modes (see fig. 4b and 4c). Quantitatively, both variants of  $\partial\text{PSE}$  outperform the human and NSGA-II baselines in both simulated and real environments. While the application-specific heuristic is nearly optimal in the simulated scenario,



(a) a



(b) b

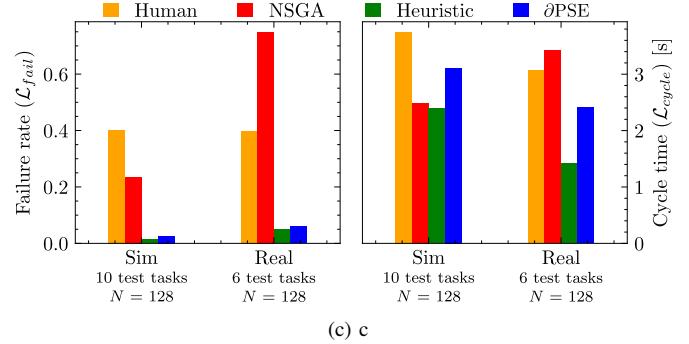
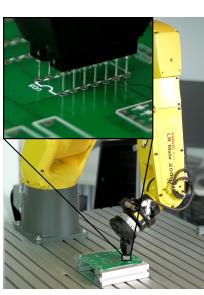
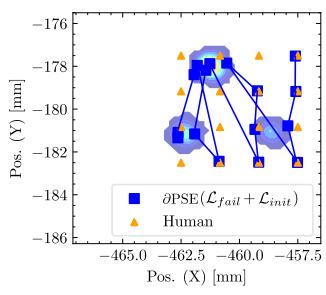


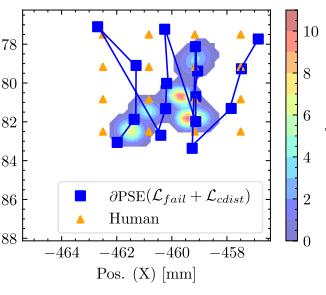
Fig. 3. Results for experiment IV-A (spiral search, stationary): Experiment setup (a) and examples for an optimized search pattern generated by ∂PSE (b). Heuristic-free ∂PSE outperforms the human and NSGA-II baselines and is competitive with an application-specific heuristic (c).



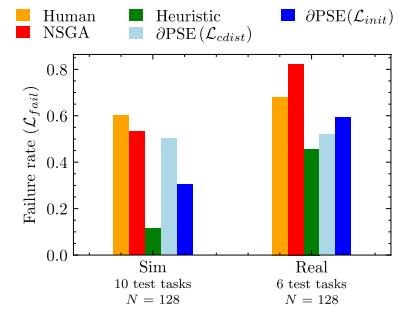
(a) a



(b) b



(c) c



(d) d

Fig. 4. Results for experiment IV-B (probe search, stationary): Experiment setup (a) and examples for optimized search patterns generated by ∂PSE with the  $\mathcal{L}_{init}$  (b) and  $\mathcal{L}_{cdistr}$  (c) regularizers. Heuristic-free ∂PSE outperforms the human and NSGA-II baselines and is competitive with an application-specific heuristic (d).

∂PSE with  $\mathcal{L}_{cdistr}$  regularization is competitive with it in the real-world environment, while avoiding any assumptions on the underlying hole distributions. The data efficiency of ∂PSE is particularly apparent when compared to NSGA-II: In the real-world scenario, NSGA-II yielded subpar results despite being allowed 300 executions, more than twice ∂PSE’s 128. ∂PSE achieves this efficiency by transferring knowledge learned from simulated data during pretraining, and by exploiting gradient information during optimization.

### C. Probe Search (nonstationary)

In a further series of experiments, we analyze the capacity of ∂PSE to optimize search strategies with respect to nonstationary processes. We consider the same task as in experiment IV-B, but omit the stationarity assumption on  $\{H_t\}$ . We evaluate ∂PSE in a simulated environment on three different nonstationary processes:

- 1) *Linear drift*, where the modes of  $\{H_t\}$  are translated by a constant offset at each timestep  $t$
- 2) *Brownian motion*, where the modes of  $\{H_t\}$  are translated by an offset sampled from a bivariate unimodal Gaussian at each timestep  $t$
- 3) *Shift*, where the modes of  $\{H_t\}$  are translated by a uniformly random offset with probability  $p_{shift} = 0.05$  at each timestep  $t$

TABLE I  
RESULTS OF EXPERIMENT IV-C: FAILURE RATES FOR DIFFERENT PARAMETER OPTIMIZERS AND STOCHASTIC PROCESSES. ∂PSE WITH THE  $\mathcal{L}_{cdistr}$  REGULARIZER REDUCES FAILURES BY UP TO 53% OVER 100 STEPS.

	Drift	Brownian	Shift
None	0.679	0.679	0.600
∂PSE ( $\mathcal{L}_{fail}$ )	0.458	0.605	0.472
∂PSE ( $\mathcal{L}_{fail} + \mathcal{L}_{init}$ )	0.556	0.616	0.406
∂PSE ( $\mathcal{L}_{fail} + \mathcal{L}_{cdistr}$ )	<b>0.318</b>	<b>0.494</b>	<b>0.339</b>

The results are summarized in table I. Both unregularized and regularized ∂PSE increase the probability of success by over 20% over short time horizons, confirming the results from experiment IV-B even as the underlying distributions change over time. For long-horizon processes, the benefits of ∂PSE are more pronounced, illustrating the capacity of ∂PSE to continuously produce suitable parameterizations over longer timescales.

### D. Comparison with Meta-Learning Approaches

Meta-learning, learning to efficiently learn from few training examples, is a powerful paradigm fundamentally related to ∂PSE’s transfer-learning scheme and has become state of the art for solving few-shot learning problems. We find

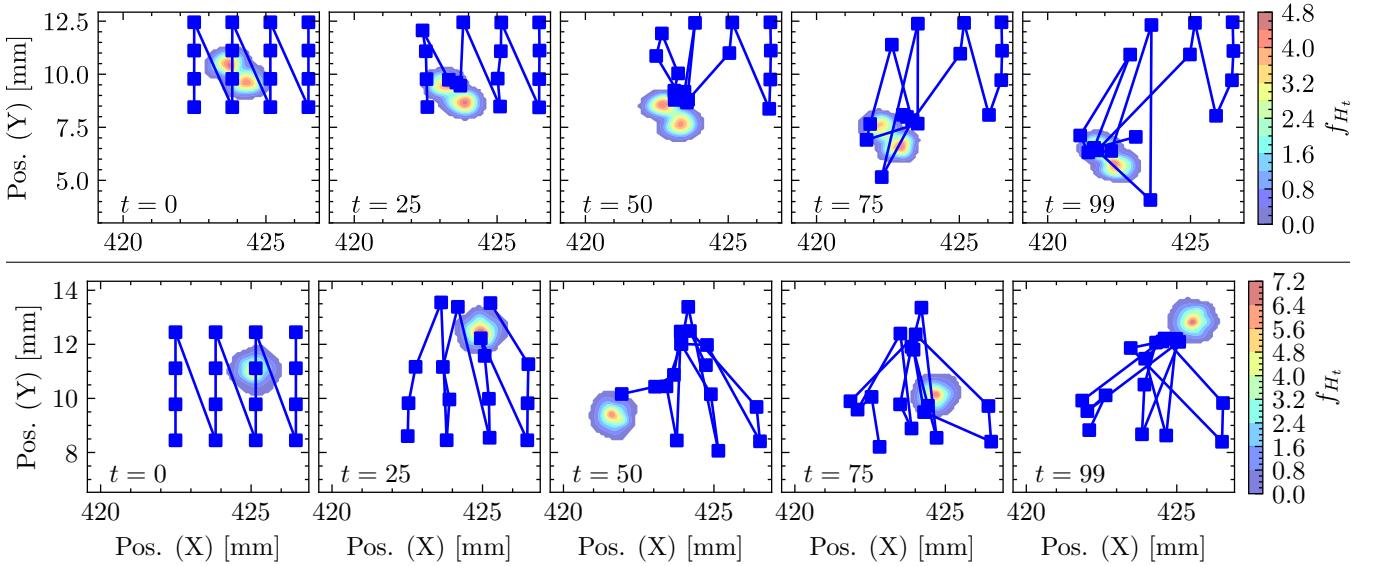


Fig. 5. Optimization of a probe search strategy ( $\partial\text{PSE}$ ,  $\mathcal{L}_{\text{init}}$  regularization) in the presence of nonstationary noise processes (top: drift, bottom: brownian motion). For both noise processes, the search pattern (blue) follows the hole distribution, while respecting the constraints imposed by the regularizer.

TABLE II  
RESULTS OF EXPERIMENT IV-D: COMPARISON OF  
PRETRAINING/FINE-TUNING AGAINST META-LEARNING ALTERNATIVES  
FOR TRAINING  $\partial\text{PSE}$ 'S SHADOW PROGRAM.

Learning algorithm	Traj. loss	Success acc.
Pretrain	<b>0.80</b>	<b>0.72</b>
Pretrain + Dropout	0.83	0.71
MAML (5) + Dropout	0.94	0.64
MAML (5)	0.95	0.64
FOMAML (5)	0.96	0.62
FOMAML (128)	0.99	0.58
Reptile (128)	1.35	0.56

that in the context of gradient-based parameter optimization, simple pretraining and fine-tuning outperforms various state-of-the-art meta-learning approaches. We substantiate our finding by comparing  $\partial\text{PSE}$  as proposed in III-C with variants of  $\partial\text{PSE}$ , where the pretraining and fine-tuning of the shadow program is replaced by the meta-learning and adaptation schemes of first- and second-order model-agnostic meta-learning (MAML) [36] and Reptile [37]. Comparing against model-agnostic approaches ensures comparability, as the network architectures as well as the training data remain unchanged. Due to the high memory requirements of second-order MAML, it could only be evaluated for meta-test sets of size  $N = 5$  (vs.  $\partial\text{PSE}$ 's  $N = 128$ ). We evaluate first-order MAML (FOMAML) for both  $N = 5$  and  $N = 128$  and Reptile for  $N = 128$ . As shown in table II, the proposed pretraining and fine-tuning scheme results in a better predictor than all tested meta-learning alternatives. The poor performance of second-order MAML is likely due to its limitation to only 5 adaptation examples, which do not suffice to learn meaningful characteristics of the underlying distributions. However, the poor performance

of FOMAML on the larger meta-test set indicates that the benefits of meta-learning diminish as the number of meta-test examples increases. This suggests that simpler, much less computationally intensive transfer-learning based approaches can compete with and even outperform meta-learning in the small-data regime (between tens and hundreds of training examples). Our findings also confirm the intuition of Sun et al. [40] that meta-learning is most effective for shallow network architectures and requires a large amount ( $M \sim 100k$ ) of meta-training tasks. The chained deep residual GRUs at the core of SPI for precise trajectory prediction and the limited amount of pretraining tasks ( $M = 1000$ ) likely limit the utility of meta-learning for gradient-based parameter optimization, at least in conjunction with  $\partial\text{PSE}$ .

## V. CONCLUSION AND OUTLOOK

In this paper, we propose Differentiable Programming in Stochastic Environments ( $\partial\text{PSE}$ ), a novel approach to optimizing robot search strategies. By conditioning learned predictive models of robot skills via transfer learning and embedding them in a differentiable program representation, we obtain a heuristic-free optimizer for arbitrary search strategies. By applying our method to two real-world assembly tasks with stationary process noise and one simulated nonstationary process, we demonstrated that  $\partial\text{PSE}$  outperforms other heuristic-free optimizers and is competitive with task-specific, hand-crafted heuristics.  $\partial\text{PSE}$  permits the effective automation of the labor-intensive optimization phase of robot workcells, while its extreme data-efficiency as well as its black-box nature renders it equally suitable for service robotics tasks. Our future work is focused on extending  $\partial\text{PSE}$  to applications beyond search strategies, such as force-controlled insertion or grasping. In addition, we are exploring the inference of task-specific goal functions from human demonstrations and symbolic knowledge.

## REFERENCES

- [1] M. Metzner, S. Leurer, A. Handwerker, E. Karlidag, A. Blank, F. Hefner, and J. Franke, “High-precision assembly of electronic devices with lightweight robots through sensor-guided insertion,” *Procedia CIRP*, vol. 97, pp. 337–341, Jan. 2021.
- [2] I. F. Jasim, P. W. Plapper, and H. Voos, “Position Identification in Force-Guided Robotic Peg-in-Hole Assembly Tasks,” *Procedia CIRP*, vol. 23, pp. 217–222, Jan. 2014.
- [3] B. Alt, D. Katic, A. K. Bozcuoglu, R. Jäkel, and M. Beetz, “Robot Program Parameter Inference via Differentiable Shadow Program Inversion,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 2021.
- [4] J. A. Marvel, W. S. Newman, D. P. Gravel, G. Zhang, Jianjun Wang, and T. Fuhlbrigge, “Automated learning for parameter optimization of robotic assembly tasks utilizing genetic algorithms,” in *2008 IEEE International Conference on Robotics and Biomimetics*, Feb. 2009, pp. 179–184.
- [5] S. Chernova and M. Veloso, “An evolutionary approach to gait learning for four-legged robots,” in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, vol. 3, Sep. 2004, pp. 2562–2567 vol.3.
- [6] R. Calandra, A. Seyfarth, J. Peters, and M. P. Deisenroth, “Bayesian optimization for learning gaits under uncertainty,” *Ann Math Artif Intell*, vol. 76, no. 1, pp. 5–23, Feb. 2016.
- [7] R. Akrour, D. Sorokin, J. Peters, and G. Neumann, “Local Bayesian Optimization of Motor Skills,” in *International Conference on Machine Learning*. PMLR, Jul. 2017, pp. 41–50.
- [8] F. Berkenkamp, A. Krause, and A. P. Schoellig, “Bayesian Optimization with Safety Constraints: Safe and Automatic Parameter Tuning in Robotics,” *ArXiv160204450 Cs*, Apr. 2020.
- [9] J. Snoek, K. Swersky, R. Zemel, and R. Adams, “Input Warping for Bayesian Optimization of Non-Stationary Functions,” in *International Conference on Machine Learning*. PMLR, Jun. 2014, pp. 1674–1682.
- [10] A. Hebbal, L. Brevault, M. Balesdent, E.-G. Talbi, and N. Melab, “Bayesian Optimization using Deep Gaussian Processes,” *ArXiv190503350 Cs Stat*, May 2019.
- [11] Y. Zhou, J. Gao, and T. Asfour, “Movement Primitive Learning and Generalization: Using Mixture Density Networks,” *IEEE Robot. Automat. Mag.*, vol. 27, no. 2, pp. 22–32, Jun. 2020.
- [12] J. Kulk and J. S. Welsh, “Evaluation of walk optimisation techniques for the NAO robot,” in *2011 11th IEEE-RAS International Conference on Humanoid Robots*, Oct. 2011, pp. 306–311.
- [13] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, “Automatic Differentiation in Machine Learning: A Survey,” *J. Mach. Learn. Res.*, vol. 18, no. 153, pp. 1–43, 2018.
- [14] W. Jin, Z. Wang, Z. Yang, and S. Mou, “Pontryagin Differentiable Programming: An End-to-End Learning and Control Framework,” *ArXiv191212970 Cs Eess Math*, Jun. 2020.
- [15] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, “A Comprehensive Survey on Transfer Learning,” *Proceedings of the IEEE*, vol. PP, pp. 1–34, Jul. 2020.
- [16] S. Ruder, “Neural Transfer Learning for Natural Language Processing,” Ph.D. dissertation, National University of Ireland, Feb. 2019.
- [17] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving, “Fine-Tuning Language Models from Human Preferences,” *ArXiv190908593 Cs Stat*, Jan. 2020.
- [18] C. Sun, X. Qiu, Y. Xu, and X. Huang, “How to Fine-Tune BERT for Text Classification?” in *Chinese Computational Linguistics*, ser. Lecture Notes in Computer Science, M. Sun, X. Huang, H. Ji, Z. Liu, and Y. Liu, Eds. Cham: Springer International Publishing, 2019, pp. 194–206.
- [19] M. Mosbach, M. Andriushchenko, and D. Klakow, “On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines,” in *International Conference on Learning Representations*, Sep. 2020.
- [20] Y. Liu, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” *CoRR*, vol. abs/1907.11692, 2019.
- [21] Y. Peng, S. Yan, and Z. Lu, “Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets,” in *Proceedings of the 18th BioNLP Workshop and Shared Task*. Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 58–65.
- [22] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, “Pre-Trained Image Processing Transformer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 299–12 310.
- [23] X. Li, Y. Grandvalet, F. Davoine, J. Cheng, Y. Cui, H. Zhang, S. Belongie, Y.-H. Tsai, and M.-H. Yang, “Transfer learning in computer vision tasks: Remember where you come from,” *Image and Vision Computing*, vol. 93, p. 103853, Jan. 2020.
- [24] V. Miglani and M. Bhatia, “Skin Lesion Classification: A Transfer Learning Approach Using EfficientNets,” in *Advanced Machine Learning Technologies and Applications*, ser. Advances in Intelligent Systems and Computing, A. E. Hassanien, R. Bhatnagar, and A. Darwish, Eds. Singapore: Springer, 2021, pp. 315–324.
- [25] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, “Why Does Unsupervised Pre-training Help Deep Learning?” *J. Mach. Learn. Res.*, vol. 11, no. 19, pp. 625–660, 2010.
- [26] K. Zhang and S. Bowman, “Language Modeling Teaches You More than Translation Does: Lessons Learned Through Auxiliary Syntactic Task Analysis,” in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018, pp. 359–361.
- [27] S. Wang, M. Khabsa, and H. Ma, “To Pretrain or Not to Pretrain: Examining the Benefits of Pretraining on Resource Rich Tasks,” *ArXiv200608671 Cs Stat*, Jun. 2020.
- [28] S. Dasari, F. Ebert, S. Tian, S. Nair, B. Bucher, K. Schmeckpeper, S. Singh, S. Levine, and C. Finn, “RoboNet: Large-Scale Multi-Robot Learning,” *ArXiv191011215 Cs*, Oct. 2019.
- [29] R. Julian, B. Swanson, G. S. Sukhatme, S. Levine, C. Finn, and K. Hausman, “Never Stop Learning: The Effectiveness of Fine-Tuning in Robotic Reinforcement Learning,” *ArXiv200410190 Cs Stat*, Jul. 2020.
- [30] E. D. Coninck, T. Verbelen, P. V. Molle, P. Simoens, and B. D. IDLab, “Learning to Grasp Arbitrary Household Objects from a Single Demonstration,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019, pp. 2372–2377.
- [31] L. Yen-Chen, A. Zeng, S. Song, P. Isola, and T.-Y. Lin, “Learning to See before Learning to Act: Visual Pre-training for Manipulation,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 7286–7293.
- [32] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey, “Meta-Learning in Neural Networks: A Survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2021.
- [33] A. Goyal and Y. Bengio, “Inductive Biases for Deep Learning of Higher-Level Cognition,” *ArXiv201115091 Cs Stat*, Feb. 2021.
- [34] T. Munkhdalai and H. Yu, “Meta Networks,” *ArXiv170300837 Cs Stat*, Jun. 2017.
- [35] J. Snell, K. Swersky, and R. S. Zemel, “Prototypical Networks for Few-shot Learning,” *ArXiv170305175 Cs Stat*, Jun. 2017.
- [36] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in *ICML*, Aug. 2017, pp. 1126–1135.
- [37] A. Nichol, J. Achiam, and J. Schulman, “On First-Order Meta-Learning Algorithms,” *ArXiv180302999 Cs*, Oct. 2018.
- [38] I. Prapas, B. Derakhshan, A. R. Mahdiraji, and V. Markl, “Continuous Training and Deployment of Deep Learning Models,” *Datenbank Spektrum*, vol. 21, no. 3, pp. 203–212, Nov. 2021.
- [39] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
- [40] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele, “Meta-Transfer Learning for Few-Shot Learning,” *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. CVPR*, 2019.