Robotics: Science and Systems 2020
Corvalis, Oregon, USA, July 12-16, 2020

Learning of Sub-optimal Gait Controllers for Magnetic Walking Soft Millirobots

Utku Culha*‡, Sinan O. Demir*‡, Sebastian Trimpe†§, and Metin Sitti*¶
*Physical Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany
†Intelligent Control Systems Group, Max Planck Institute for Intelligent Systems, Stuttgart, Germany
§Institute for Data Science in Mechanical Engineering, RWTH Aachen University, Germany
‡Equally contributing authors, ¶Correspondence to sitti@is.mpg.de

***Abstract*—Untethered small-scale soft robots have promising applications in minimally invasive surgery, targeted drug delivery, and bioengineering applications as they can access confined spaces in the human body. However, due to highly nonlinear soft continuum deformation kinematics, inherent stochastic variability during fabrication at the small scale, and lack of accurate models, the conventional control methods cannot be easily applied. Adaptivity of robot control is additionally crucial for medical operations, as operation environments show large variability, and robot materials may degrade or change over time, which would have deteriorating effects on the robot motion and task performance. Therefore, we propose using a probabilistic learning approach for millimeter-scale magnetic walking soft robots using Bayesian optimization (BO) and Gaussian processes (GPs). Our approach provides a data-efficient learning scheme to find controller parameters while optimizing the stride length performance of the walking soft millirobot robot within a small number of physical experiments. We demonstrate adaptation to fabrication variabilities in three different robots and to walking surfaces with different roughness. We also show an improvement in the learning performance by transferring the learning results of one robot to the others as prior information.**

***Keywords*—Soft robotics; gait control; Bayesian optimization; transfer learning**

I. Introduction

Soft-bodied robots are composed of functional soft materials exhibiting shape-programmable properties that allow passive/active structural compliance and large degrees of freedom, which are hard to achieve using conventional rigid materials [16]. The research on soft robots is getting more attention owing to easier access to novel fabrication methods and functional materials, and potential high-impact medical and other applications [25]. Biologically inspired soft robots can be used to study their soft-bodied biological counterparts [17], and open new application areas in multi-terrain locomotion [4], adaptive manipulation [15, 28], and human-assistive wearable systems [41]. Soft robots also enable safe human-robot physical interaction due to their high compliance and limited output force, which normally require additional computational effort in conventional robotic systems [12]. Small-scale (*i.e.*, millimeter) untethered soft robots have further potential usage in medicine owing to their ability to access to enclosed small spaces non-invasively [26, 35] and the embodiment of functionalized materials enabling targeted drug delivery and bio-sensing [5].

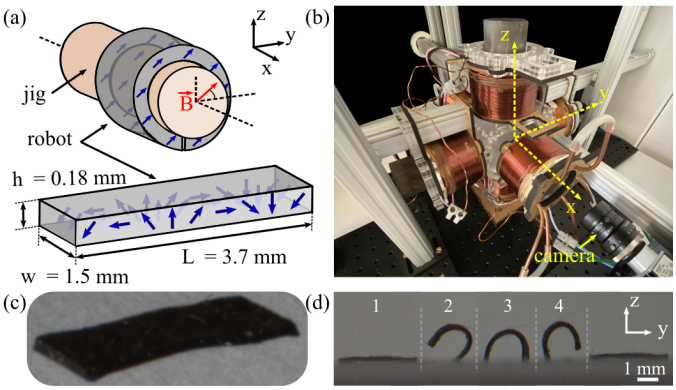

Figure 1. (a) The magneto-elastomer robot is rolled around a jig and magnetized with $|\vec{B}| = 1.8$ T field (red arrow) with a 45° angle with respect to the y-axis. The unfolded robot maintains a circular magnetization profile along its body (blue arrows). (b) Photo of the experimental setup with 6 electromagnetic coils and a high-speed camera. (c) Image of the fabricated and magnetized soft millirobot. (d) Projected planar images showing the four consecutive states of the robot walking gait: (1) relaxed, (2) front-stance, (3) double-stance, and (4) back-stance. These images are placed with a separation on the y-axis for visual clarity, *i.e.*, the robot does not jump in between states during the experiments. Numbers represent the four states.

Despite their potential, the virtual infinite degrees of freedom, the lack of accurate models, fabrication variations, and non-linear behavior (*e.g.*, hysteresis) render the application of conventional control methods challenging for soft robots [32]. So far, constant curvature (CC) models utilizing bending beam theories have been widely-used to approximately represent the deformation of continuum robots [42]. Alternatively, analytical and geometrically exact models have been suggested for continuum robots that are represented as simplified rods [29]. Finite element methods (FEM) provide numerical solutions to soft robot kinematics by utilizing a chain of rigid elements connected with tunable spring-damper mechanisms [21]. These kinematic models allow the implementation of static and dynamic controllers for continuum robots on a larger scale [9]. However, these controllers typically depend on the continuous sensing of body deformations from embedded sensors and computationally heavy model solutions, which are conditions that may not be met for untethered soft robots at the small scales [30]. The dynamic task environment, complex deformation kinematics, fabrication-dependent performance variations, and actuation/sensing limitations have further impacts on the soft mobile robots targeting medical

applications, which make adaptive and data-efficient control methods attractive for these robots [36].

In the case of uncertainty and lack of a parametric model that represents the system, data-driven control [13] and reinforcement learning [38, 19] provide promising alternatives over model-based designs in small-scale soft robotic systems. However, the need for data efficiency, *i.e.*, the ability to learn from only a few experimental trials, presents a core challenge for such methods [6]. Conversely, Bayesian optimization (BO) [10, 34] allows for the maximization of a performance function using a small number of physical experiments. BO typically employs Gaussian processes (GPs) [27] as a probabilistic model of the latent objective function. While no explicit dynamics model is needed, GPs allow for incorporating information as probabilistic priors, thus reducing data requirements. There are emerging examples that demonstrate the application of this approach to optimize the locomotion performance of robots on different length scales [3, 44, 22]. Despite its potential to address the control challenge for untethered soft robots, there are only a few examples that apply this method such as in the gait exploration of a tensegrity system [31], and the optimization of an undulating motion of a microrobot [40]. So far, a data-efficient procedure that adapts the learned controllers to different robots and environmental conditions for untethered small-scale soft robots has not been demonstrated.

In this paper, we propose a learning procedure to find the controller parameters of magnetically actuated, untethered, soft millirobots (see Fig. 1) that generate optimum walking gaits within a small number of physical experiments. We specifically focus on these types of robots due to their biocompatible use of external magnetic actuation that supports multi-functionality in future medical tasks [14, 23] and the high-resolution magnetization methods that allow more complex deformation capabilities at the small scale [24, 7]. We produce three replicas of a previously reported soft millirobot demonstrating a hand-tuned walking gait aiming for medical applications [14] and test the repeatability of their results. We begin with finding the optimum walking gait controllers of our robots using BO with GPs; initially without any prior information about the correlation between the controller parameters and the robot performance. Later, we explore the controller parameter space of one robot and present a straightforward way in the context of BO to transfer prior information from the first robot to all three robots while finding their optimum walking gait controllers in a data-efficient fashion. We report the optimum controllers and walking gait performances in terms of achieved stride lengths for all three robots and compare the two learning approaches (*i.e.*, with and without using prior information). We also transfer this information to adapt to the changes in the task environment by finding the controller parameters for walking on rough surfaces.

The organization of this paper is as follows. We describe the robot design and its walking gaits in Section II and introduce our learning approach in Section III. Section IV presents the experiments on learning of the optimum gait controllers without using the prior information, the generation of the prior

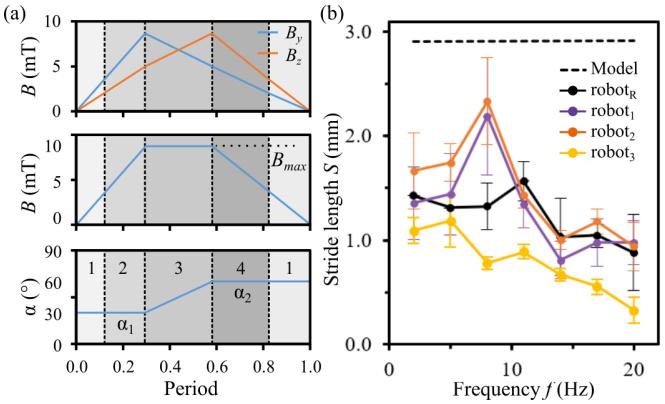

Figure 2. (a) Walking gait control parameters during a single period of a sample motion (a single period of $1/f = 90$ ms for $f = 11$ Hz is normalized to 0-1 on the abscissa). The magnetic field B is controlled on the y-z plane and shown with its y and z components (top) whose magnitude (middle) reaches B_{max} and orientation (bottom) changes from α_1 to α_2. Dashed vertical lines represent the (1) relaxed, (2) front-stance, (3) double-stance, and (4) back-stance states of the walking gait. (b) The stride length S performance of the previously reported robot ($robot_R$) vs. the performance of three replica robots using the same controller parameters ($robot_{1,2,3}$). Each data point for each robot ($robot_{1,2,3}$) represents the mean of 10 experiments and the error bars show the standard deviation. The performance of $robot_R$ and the horizontal dashed line which represents the model prediction are adapted from [14].

controller information from one robot, and the optimization of the walking gaits by transferring the learned controllers to three robots and different locomotion surfaces. We discuss the experimental results in Section V and conclude our work in Section VI.

II. Robot Design and Gait Definition

We followed the methods and materials reported in [14] and fabricated three magnetic soft millirobots with a 1:1 body mass ratio of Ecoflex 00-10 (Smooth-On Inc.) and neodymium-iron-boron (NdFeB) magnetic microparticles with around 5 μm diameter (MQP-15-7, Magnequench). We placed this pre-polymer mixture on a methyl methacrylate plate and cut the robots out of the cast using a high-resolution laser cutter (LPKF Protolaser U4) after the polymer is cured. Our robots had the final dimensions of length $L = 3.7$ mm, width $w = 1.5$ mm, and height $h = 185$ μm as shown in Fig. 1-a. We separately folded the robots around a circular jig with a circumference equal to L and magnetized them within a magnetic field with a magnitude of 1.8 T and orientation of 45° measured counterclockwise from the y-axis. Once the robots are unfolded from the jig, the magnetic particles maintained their magnetization orientation forming a circular profile along the longitudinal axis of the robot body (Fig. 1-a). We used these robots (*i.e.*, robots 1, 2, and 3) with the same nominal material properties and dimensions for our experiments (see Fig. 1-c for a sample robot image).

The walking gait of our robot is composed of four consecutive quasi-static states that are inspired by the planar quadrupedal bounding [1] and a caterpillar's inching motion [39]. These states are depicted as (1) relaxed, (2) front-stance, (3) double-stance, and (4) back-stance as shown in Fig. 1-d. We placed our magnetized robot along the y-axis of the magnetic coil setup consisting of three orthogonal pairs of

custom-made electromagnets (Fig. 1-b) that generated a 3-D uniform magnetic field within a 4x4x4 cm^3 space. We modulated the magnetic field on the y-z plane that coincided with the center of the test environment. We controlled four parameters to generate the walking gait: the maximum magnetic field magnitude (B_{max}), the frequency of the actuation cycle (f), and two magnetic field orientation angles (α_1 and α_2) measured counterclockwise from the y-axis. The plots in Fig. 2-a show the change of the control parameters during a single period of the motion for $B_{max} = 10$ mT, $f = 11$ Hz, $\alpha_1 = 30°$ and $\alpha_2 = 60°$, which are hand-tuned parameters reported in [14]. At the beginning of a single gait period, the robot started at a relaxed state for $0 \leq B \leq 4$ mT. The robot tilted forward when $\alpha = \alpha_1$ and B increased from 4 mT to $B_{max} = 10$ mT. While B remained constant at B_{max}, the orientation of the magnetic field changed from α_1 to α_2 causing the robot to initially switch to the double-stance state and then to the back-stance state when $\alpha = \alpha_2$. Then, B decreased while keeping the orientation of the magnetic field constant, and the robot gradually switched back to the relaxed state. For $B < 4$ mT, the robot assumed the relaxed state, and a single period of walking actuation ended when $B = 0$ mT. We reset B at the end of every gait cycle to avoid jerky motion when α changed from α_1 to α_2. In our experiments, the relaxed state was never skipped but its duration changed according to f. The consecutive images from a single walking gait period are shown in Fig. 1-d. We tracked the robot gait using a high-speed camera (Basler aCa2040-90uc, 60 frames per second (fps), 1 pixel $\sim$ 27 µm resolution) that is orthogonally placed to the axis of robot motion (Fig. 1-b). In every experiment, we calculated the stride length (S) of the robot by tracking the average distance covered by its center of mass in 10 consecutive steps.

To test the repeatability of the previously reported results in [14], we experimented with our fabricated robots using their suggested controller parameter sets. Fig. 2-b shows the stride length performances of our robots ($robot_{1,2,3}$) and compares them with the reported robot ($robot_R$) performance (*i.e.*, we calculated the stride length of $robot_R$ from the values reported in [14]) for $B_{max} = 10$ mT, $\alpha_1 = 30°$, $\alpha_2 = 60°$, and $2 \leq f \leq 20$ Hz. Our preliminary results revealed that:

- In this scale, the gait performance showed clear inconsistency due to the variability during fabrication and environmental factors even though the same materials, methods, controller parameters, and walking surfaces are used in the fabrication and experimentation of the millimeter-scale soft robots.
- Unlike the model prediction, the robot performance showed non-monotonic behavior along with increasing f, which rendered the design of a model-based gait controller unreliable.
- In addition to the virtual infinite degrees of freedom inherited by the soft materials, the controller parameters existed in a continuous space, making the hand-tuning of these parameters within physical experiments impractical.

These observations found the goals of our paper in which we address the necessity for a data-efficient controller learning system that is robust to the variabilities caused by the material, fabrication, and the task environment of the miniature scale, medical-oriented, untethered soft robots.

III. Learning Approach

We aim to optimize the walking gait controller parameters to maximize the stride length S of the robot. Therefore, we define the reward function as

$$S : \Theta \rightarrow \mathbb{R}, \tag{1}$$

which maps the parameter set $\theta = [B_{max}, f, \alpha_1, \alpha_2]$ to scalar reward values (*i.e.*, the experimental stride length performance of a robot). According to the definition of the reward function, we formulate the parameter learning as the (global) optimization problem

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmax}}\, S(\theta), \tag{2}$$

where Θ denotes the complete search space, θ is the parameter set, and $S(\theta)$ is the experimentally observed stride length performance of the robot for a given θ.

We define the range of the controller parameters based on the findings in [14] and the physical limitations of our magnetic actuation setup. Accordingly, B_{max} is defined between 5 mT and 12 mT, and the walking frequency, f, ranges from $f_{min} = 0.5$ Hz to $f_{max} = 20$ Hz. We limit α_1 and α_2 to $[10, 50]°$ and $[40, 80]°$ respectively and select values that satisfy $\alpha_2 > \alpha_1$ to generate the walking gait in Fig. 1-d. We use a step size of 1 mT for B, 1° for each α, and a variable step size of 0.25 Hz for $f < 2$ Hz and 2 Hz for $f \geq 2$ Hz, which yield a total number of 203520 possible parameter sets in Θ.

A. Gaussian Processes (GPs)

The magnetic soft millirobots in our paper did not have an accurate kinematic or dynamics model (see Fig. 2-b). Therefore, it is necessary to approximate the reward function based on the data collected from physical experiments rather than numerical analysis. However, the physical data has inherent uncertainty due to the noise in the measurements and the variations during the experiments. To include these uncertainties in the model, overcome the sparsity in the data, and make probabilistic predictions at unobserved locations, we model the reward function $S(\theta)$ using GPs following the previous study in [40]:

$$S(\theta) \sim GP(\mu(\theta), k(\theta, \theta')). \tag{3}$$

However, as $S(\theta)$ can only be measured with noise, we define $\tilde{S}$ as

$$\tilde{S}(\theta_i) = S(\theta_i) + n_i \tag{4}$$

where n_i is zero-mean Gaussian noise with variance σ_n^2 for each measurement i.

A GP is a non-parametric model defined by its prior mean $\mu(\theta)$ and the covariance function $\operatorname{cov}(S(\theta), S(\theta')) = k(\theta, \theta')$,

where k is the kernel. During one run of BO, the GP model is sequentially updated with $\tilde{S}(\theta)$ observed from experiments. We define one "learning run" as a run of BO until the desired stopping criterion is matched (*e.g.*, a fixed number of experiments is reached).

From the experimental data $D = \{\theta_i, \tilde{S}(\theta_i)\}_{i=1}^{N}$, the stride length of the robot for an unobserved θ can be predicted using the posterior mean and variance as follows.

$$\mu_{post}(\theta) = \mu(\theta) + k^T(\theta)K^{-1}y, \tag{5}$$

$$\sigma^2_{post}(\theta) = k(\theta,\theta) - k^T(\theta)K^{-1}k(\theta), \tag{6}$$

$$S_{post}(\theta) \mid D \sim N(\mu_{post}(\theta), \sigma^2_{post}(\theta), \tag{7}$$

where $k(\theta)$, $y \in \mathbb{R}^N$ with $k(\theta_i) = k(\theta,\theta_i)$, $y_i = \tilde{S}(\theta_i) - \mu(\theta_i)$, and $K \in \mathbb{R}^{N\times N}$ with $K_{i,j} = k(\theta_i,\theta_j) + \delta_{i,j}\sigma_n^2$, where $\delta_{i,j}$ is the Kronecker delta and σ_n^2 is the noise in the collected data set.

We select the squared exponential as the kernel function in the GPs, which is for the 1D case:

$$k_{SE}(\theta,\theta') = \sigma_f^2 exp(-(\theta-\theta')^2/2l_c^2), \tag{8}$$

where l_c is the length scale that defines the rate of variation in the modeled function for each dimension of the parameter space. Long length scales are used to model slowly-varying functions and short length scales are used to model quickly-varying functions. The signal variance σ_f^2 describes the width of distribution, *e.g.*, high σ_f^2 means higher uncertainty in the predictions of the unobserved θ.

Hyperparameters of the GPs can be listed as the noise in the collected data σ_n^2, length scale l_c for each dimension of the parameter space $\mathbb{R}^{d_c}$, and signal variance σ_f^2. To determine the value of σ_n^2, we use the maximum variance found in the experimental results shown in Fig. 2-b. We set the length scale l_c to one-fourth of the total range of each corresponding parameter. We also set the signal variance σ_f^2 to half of the body length of the robot so that the highest possible reward value (*i.e.*, L = 3.7 mm) remained inside the 95% confidence interval of the prior.

B. Bayesian Optimization (BO)

We use BO to select the parameter set θ_{next} to be tested in the next step of the learning run using the acquisition function $\alpha_{acq}(\theta)$.

$$\theta_{next} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \alpha_{acq}(\theta) \tag{9}$$

In this work, we choose the expected improvement (EI) as the acquisition function $\alpha_{acq}(\theta)$ due to its better performance compared to its alternatives as demonstrated in [40]. EI seeks the parameter set for the next step where the expected improvement in reward function is the highest compared to the previously collected data:

$$\alpha_{acq}(\theta) = \mathbb{E}[\max(0, (S(\theta) - S(\theta^*) - \xi))], \tag{10}$$

$$\alpha_{acq}(\theta) = (\mu(\theta) - S(\theta^*) - \xi)\Phi(Z) + \sigma(\theta)\phi(Z), \tag{11}$$

where $S(\theta^*)$ is the highest reward function value collected so far, Φ and ϕ are the Gaussian cumulative density and probability density functions, respectively [2]. The term Z is described as $Z = Z(\theta) = (\mu(\theta) - S(\theta^*) - \xi)/\sigma(\theta)$, with $\mu(\theta)$ and $\sigma(\theta)$ computed from Eq. (5) and Eq. (6). In Eq. (11), the two terms define the exploitation and the exploration weights of the BO respectively. The balance between these two terms is controlled by the hyperparameter ξ . As ξ gets higher, BO focuses more on exploration and seeks the next parameter set in regions with high prediction uncertainty. Conversely, BO focuses more on exploitation and selects the next parameter set within a close range to already explored regions. We choose $\xi = 0.1$ to balance the exploration and exploitation weights.

C. Transfer of the Prior Mean

In addition to the kernel (see Section III-A), the prior mean $\mu(\theta)$ must be chosen at the beginning of a BO run. Often, $\mu = 0$ is the default choice for an uninformed prior. For the millirobot learning problem herein, we suggest and investigate the transfer of information from previous learning runs by setting the prior mean to the posterior mean of a previously trained GP model, such as from a different robot. In this way, we can approximately transfer the topology of the target function between robots, which is reasonable as long as the differences between the robots and the environment do not significantly alter the function shape. In this work, we adopt and compare both approaches of an uninformed prior ($\mu(\theta) = 0$) in Section IV-A, and the transfer of the posterior mean from robot 1's previous run to all three robots in Section IV-B.

IV. Experiments

We modulated the currents running through the electromagnetic coils and the resulting magnetic field by controlling six motor driver units (SyRen25) using an Arduino microcontroller running at 1.2 kHz. We regularly calibrated the magnetic actuation matrix inside the workspace, *i.e.*, the mapping between the applied electric current and the generated magnetic field, to maintain reliable and repeatable experiments. The learning process ran on a PC that additionally handled image processing and hardware communication tasks. One step of the learning run involved five steps:

1) BO selected a new parameter set θ that maximized the acquisition function based on the GP model,
2) The microcontroller regulated the magnetic field based on the selected θ and initiated the physical experiment,
3) The camera recorded the robot motion and measured the average stride length performance S,
4) The learning system updated the GP model using the newly collected data from the experiment,
5) The robot returned to its initial position for the next step.

A. Optimization of The Walking Gait without the Prior

To test our controller learning approach without prior information, *i.e.*, $\mu(\theta) = 0$, we experimented with all three robots in the same environmental conditions and limited the

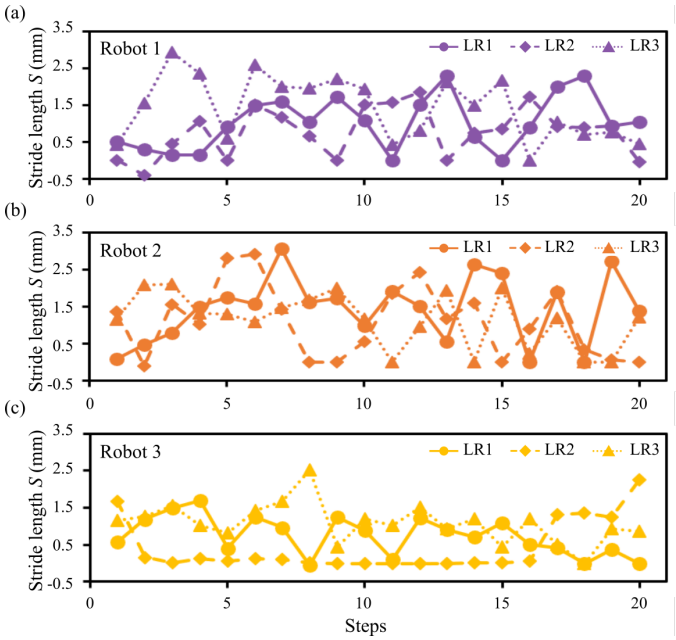

Figure 3. The learning of the controller parameters without utilizing the prior information within 20 physical experiments in 3 independent learning runs (shown as $LR_{1\text{-}3}$) for (a) robot 1, (b) robot 2, and (c) robot 3.

number of steps for each learning run to 20 experiments. We initialized the BO with the best controller parameter set reported in [14]. We performed three independent learning runs (*i.e.*, 60 experiments in total) for every robot with the same initial state, whose results are shown in Fig. 3. Each data point represents the robot's stride length performance S resulting from a different controller parameter set chosen by the BO at a given step of a learning run. As BO actively chose sample locations (*e.g.*, to explore unknown regions described in Section III), the variation in these data points was the desired behavior of the explorative learning algorithm. See Supplementary Video 1 for the gait performances of four different controller parameter sets for robot 1.

We chose the optimum controller parameter sets (θ^*) from these learning runs and repeated the walking gait five times to collect statistical information about the stride length performances. The rows designated with "no prior" in Table I shows the values for θ^* and the resulting S for each robot. It can be seen that the walking gait performances of the robots were significantly improved compared to the robot reported in [14]. Also, the standard deviation within these repeated experiments agreed with the previously reported values, highlighting the repeatability and reliability of our experimental platform. In this optimization approach, we achieved 86.6%, 94.7%, and 60.5% increase in S for robots 1 to 3 respectively (*i.e.*, compared to the S of the previous robot shown in the last row of Table I). The difference between the optimum controller parameter values in Table I demonstrates the influence of the fabrication variabilities on the robot design and performance.

Separate from the optimum stride length performance of each robot, we evaluated the overall performance of the learning runs based on the achieved stride length average of all of the tested controller parameters. This performance

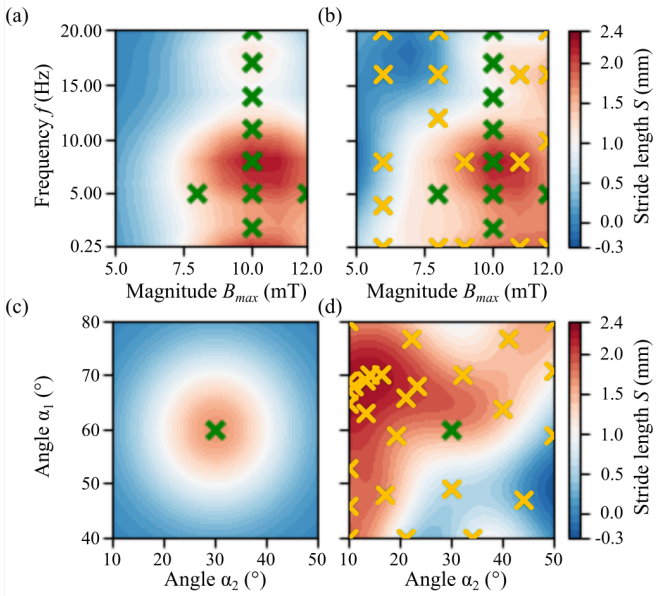

Figure 4. Approximation of the stride length performance as a function of four control parameters using GPs. The upper row shows S projected on B-f plane for $\alpha_1 = 30°$ and $\alpha_2 = 60°$, and the lower row shows S projected on the α_1-α_2 plane for $B = 10$ mT and $f = 2$ Hz. (a,c) Initial approximation of S applying the 9 hand-tuned controller parameters reported in [14] on robot 1 in our experiments. Each green cross mark represents 10 trials for the chosen parameter set. (b,d) The final probabilistic approximation of S after running the prior information generation step. Experiments with the parameters selected by our BO are represented with yellow cross marks.

metric P_{LR} shows the overall quality of the learning run's parameter selection in terms of the average of all the S and the standard deviation in 60 experiments (*i.e.*, $\text{avg}(S) \pm \text{std}$). In these experiments, the learning run for robot 1 yielded $P_{LR_1} = 1.07 \pm 0.80$ mm, $P_{LR_2} = 1.21 \pm 0.88$ mm for robot 2, and $P_{LR_3} = 0.75 \pm 0.64$ mm for robot 3. Even though there were multiple individual θs within these runs (*e.g.*, the optimums reported in Table I) that outperformed the previous study, the large standard deviation shows that the BO selected parameters that generated a wide range of performances.

B. Optimization of The Walking Gait with the Prior

1) Generation of the Prior Information: To generate useful prior information, we constructed the posterior mean $\mu_{post}(\theta)$ for robot 1 using the BO and GP described in Section III. Initially, we adopted nine different controller parameters from [14] and collected the stride length information from repeated experiments in our setup. Fig. 4-a and 4-c show the two-dimensional projection of the approximation of S function generated by the GP model (utilizing the same hyperparameters in Section III) based on these experiments. After the initial approximation, we used the BO to select new parameter sets from the unexplored parts of the 4-D search space and collected the experimental stride length performance information. We explored 123 different parameter sets in total by selectively isolating the search space dimensions. Initially, we fixed $\alpha_1 = 30°$ and $\alpha_2 = 60°$ and explored 18 different parameter values for B_{max} and f. Then, we fixed $B_{max} = 10$ mT and $f = 2$ Hz and explored 38 values for α_1 and α_2. We performed 17 additional tests for α_1 and α_2 for

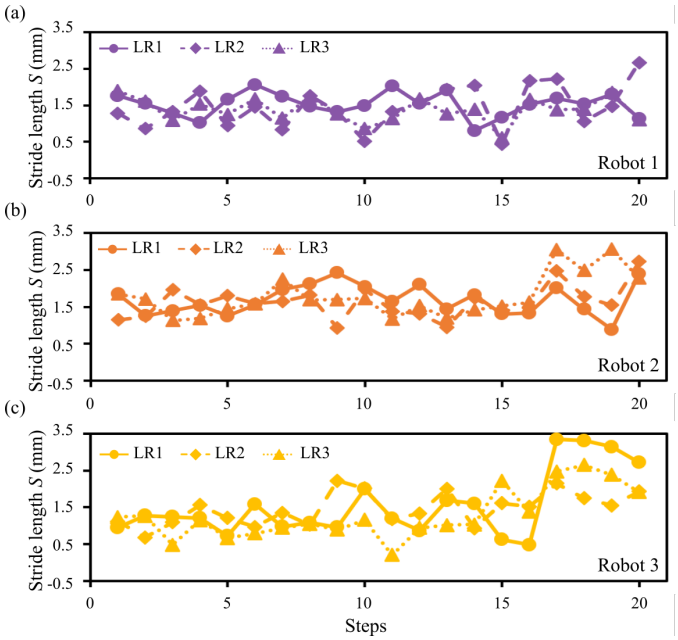

Figure 5. The learning of the controller parameters by utilizing the prior information presented in Section IV-B1 within 20 physical experiments in 3 independent learning runs for (a) robot 1, (b) robot 2, (c) robot 3.

$B_{max} = 10$ mT and $f = 8$ Hz. Finally, we explored the complete search space for four parameters with 50 more tests. For all of these tests, we stopped the exploration when the BO converged in the sense of repetitively selecting similar θs. Fig. 4 shows the two-dimensional projections of the GP-based probabilistic approximation of the performance function before (Fig. 4-a,c) and after (Fig. 4-b,d) all of the physical experiments dictated by our BO. These results show that our BO approach revealed parts of the parameter space that were not effectively explored using the hand-tuning in [14]. We used this posterior information of robot 1 and transferred it to all robots (*i.e.*, robots 1, 2, and 3) as prior information of the stride length function approximation in the remaining part of the optimization experiments.

2) Transfer of Learning Between Different Robots: Similar to the experiments in Section IV-A, we performed three independent learning runs each consisting of 20 experiments for every robot. Unlike the previous learning approach, the GP model in every learning run started with the prior mean set to the posterior mean information of robot 1 that was generated in Section IV-B1. Fig. 5 shows the walking gait performance results of three robots in these learning runs. The optimum controller parameter sets (θ^*) and the resulting stride length performances S from these learning runs are reported in Table I on the rows designated with "prior". Compared to the robot in [14], we achieved optimized walking gaits with an increased performance of 70.7%, 73.9%, and 113.3% for robot 1 to 3 respectively. See Supplementary Video 2 for a comparison of the walking gaits of three robots with the optimum of the parameters found in the experiments.

The utilization of the transferred prior information can be seen as a clear improvement in the overall learning run performance P_{LR}. In these experiments, the learning runs for

Table I
SELECTED GAIT CONTROLLER PARAMETERS

Robot	Type	Controller Parameters				Stride length S
		B_{max} (mT)	f (Hz)	α_1 (°)	α_2 (°)	(mm) (avg $\pm$ std)
Robot 1	no prior	9	10	21	61	2.25 $\pm$ 0.17
	prior	12	8	20	65	2.68 $\pm$ 0.34
Robot 2	no prior	11	8	27	65	3.06 $\pm$ 0.38
	prior	9	10	32	73	2.73 $\pm$ 0.24
Robot 3	no prior	10	10	19	80	2.52 $\pm$ 0.27
	prior	12	18	10	80	3.35 $\pm$ 0.08
Robot in [14]		10	11	30	60	1.57 $\pm$ 0.38

robot 1 yielded $P_{LR_1} = 1.45 \pm 0.43$ mm, $P_{LR_2} = 1.56 \pm 0.42$ mm for robot 2, and $P_{LR_3} = 1.43 \pm 0.70$ mm for robot 3. The improved averages compared to the results in Section IV-A show that once the prior information is transferred, the BO selected parameters that yielded better performing stride lengths in the same number of limited physical experiments. Likewise, the lower deviation in the averages implies that the performance range of the selected parameters was consistent.

C. Adaptation to Different Surfaces

Similar to the variances during the fabrication, the changes in the task environment also have a significant impact on the untethered soft small-scale robots. To demonstrate the adaptation capability of our learning approach to different surfaces, we experimented with robot 1 on a surface coated with 60-grit sandpaper (Klingspor, KL385-JF), which had a higher roughness compared to the plexiglass surface used in our previous experiments (Fig. 6-a). The surface profile examination in Fig. 6-b and 6-c (Keyence VK-X260K) shows

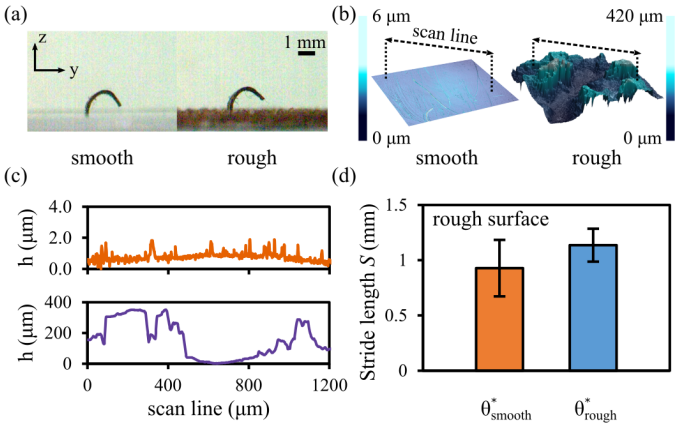

Figure 6. Adaptation to different surface roughnesses. (a) Colored images showing the robot walking on different surfaces: smooth plexiglass (left) and rough sandpaper (right). (b) Profilometer analysis showing the roughness difference between two surfaces. (c) Linear profiling along the scan line axis marked with dashed lines in (b) reveals the average height difference between surfaces (note the two orders-of-magnitude difference). (d) The average stride length performances of robot 1 on the rough surface using the parameters optimized for the smooth surface (left) vs. parameters using the prior information for the rough surface (right).

Table II
GAIT CONTROLLER PARAMETERS FOR DIFFERENT SURFACES

Surface	Controller Parameters				Stride length S
	B_{max} (mT)	f (Hz)	α_1 (°)	α_2 (°)	(mm) (avg $\pm$ std)
smooth	12	8	20	65	2.68 $\pm$ 0.34
rough	11	2	10	76	1.15 $\pm$ 0.15

that two surfaces had significant differences between their roughness Sq (root mean square height): sandpaper $S_q = 85.01$ µm compared to plexiglass surface $S_q = 0.38$ µm. Fig. 6-c shows that the terrain of the rough surface had features almost three times taller than the height of our robot.

Initially, we used the optimum control parameters of robot 1 in Section IV-B2 on the rough surface and observed that the walking gait performance dropped from $S = 2.68 \pm 0.34$ mm to $S = 0.93 \pm 0.26$ mm. Then, we optimized the robot on the rough surface using our learning approach utilizing the prior mean information generated in Section IV-B1. Within a single learning run of 20 experiments, we found a parameter set that increased the stride length performance to $S = 1.15 \pm 0.15$ mm, yielding a 24.7% optimization (Fig. 6-d). The optimum walking gait controller parameter sets for both surfaces are reported in Table II. These results show that the learning system adapted the controller parameters for the new terrain features to maintain a successful walking gait. During the optimization process, we observed that the robot typically got stuck inside the cavities that were larger than its height. To overcome this problem, BO optimized the parameters such that the robot moved slower with the lower f, and the tilted back and forward with larger α_2 and smaller α_1 to release its "legs" from the cavities. See Supplementary Video 3 for a comparison between walking gaits on two surfaces.

V. DISCUSSIONS

When the stride length performance results in Table I are compared, it can be seen that some of the controller parameters selected without the prior information outperformed the parameters selected with the prior information. Regardless of the prior information, as BO is a probabilistic optimization algorithm and promotes some exploration, these results were expected. Nonetheless, all of these optimized parameters significantly outperformed the hand-tuned values in [14], highlighting one of the major contributions of our work. As a second contribution, we showed that transferring the posterior mean of one robot as the prior mean for the learning experiments of other robots lead to benefits in terms of improved average performance of learning runs P_{LR} as shown in Fig. 7. For robot 1, the average in the P_{LR} increased by 35.5%, 29.3% for robot 2, and 91% for robot 3. We note, however, that even though our method showed positive influence for the considered robot cases, further investigation on the most appropriate means of transfer for the considered problem is interesting future work.

Our results can reveal design guidelines to improve the kinematic models of the small-scale robots while utilizing the

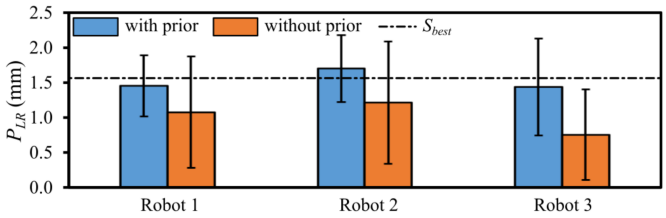

Figure 7. Improvement in the controller learning procedure represented with the overall stride length average P_{LR} and its standard deviation from the learning runs with and without using prior information. The dashed line (S_{best}) refers to the best stride length performance of the robot in [14].

constant curvature (CC) approximations [42], analytical models [29], and FEM methods [21]. Additionally, recent studies suggesting fabrication methods with higher magnetization resolution on a smaller scale [43, 18] may address the fabrication variability problem owing to their automated procedures. However, especially for robots designed for non-invasive medical operations, the interaction with the dynamic task environment may still have degrading effects on the robot's soft body and change its performance unpredictably. In the absence of an adaptive online controller with a high-bandwidth feedback system, a data-efficient controller learning system may adapt the previously optimum controller parameters to the changes in the robot. For example, such an adaptive learning system may be applied for endoscopic soft robots within or outside the gastrointestinal (GI) tract [37, 11] using a small number of trials. Contributing to this idea, our paper demonstrated the data-efficient learning of controller parameters and adaptation to different task environments without depending on the robot models whose results are shown in Table II.

In our experiments, we noticed that the duty factor of the double-stance state reduced with the increased actuation frequencies, which is commonly observed in legged locomotion in nature [1]. The highest stride length performances for all three robots were lower than the body-length (*i.e.*, $L = 3.7$ mm) of the robot, which also suggests that robots were following the walking gait state sequence by avoiding ballistic flight as in running. However, our approach can be extended to investigate the switch between dynamic gaits and the change of controller parameters accordingly.

In this paper, we focused on finding the optimal walking gait parameters inside this Θ using only physical experiments with BO and GPs. The systematic comparison of our experimental approach to alternative methods supported with simulations such as intelligent trial and error [8], evolution algorithms [20], or policy gradients [33] is also an interesting future work.

VI. CONCLUSIONS

The results in this paper show the potential of a control learning system that can learn the new robot parameters quickly, and adapt to variabilities in the absence of a model-based control for soft robots. Our experimental results suggest that the boundaries for the parameter search space may be widened further to explore richer behaviors in future studies. This study can be further extended to involve the design parameters, such as the magnetic particle density in our robots,

and guide the task-oriented design strategies for medical-oriented robots. Our long term vision is to build a completely autonomous system that can actuate, track, evaluate, and optimize a complex soft robot with minimum human involvement.

ACKNOWLEDGMENTS

U.C. thanks the Alexander von Humboldt Foundation for the Humboldt Postdoctoral Research Fellowship and the Federal Ministry for Education and Research. S.O.D. thanks the Ministry of National Education of the Republic of Turkey for the Doctoral Scholarship. We thank Wenqi Hu for his contributions in developing the magnetic coil setup. This work was funded in part by the Cyber Valley Initiative, Grassroots Initiative, the Max Planck Society, and the European Research Council (ERC) Advanced Grant "SoMMoR" Project with Grant No: 834531.

REFERENCES

[1] R. McN. Alexander. The Gaits of bipedal and quadrupedal animals. *The International Journal of Robotics Research*, 3(2):49–59, 1984.

[2] Eric Brochu, Vlad M. Cora, and Nando de Freitas. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv:1612.06830 [cs/LG]*, 2010.

[3] Roberto Calandra, André Seyfarth, Jan Peters, and Marc Peter Deisenroth. Bayesian optimization for learning gaits under uncertainty. *Annals of Mathematics and Artificial Intelligence*, 76(1-2):5–23, 2016.

[4] Marcello Calisti, Giacomo Picardi, and Cecilia Laschi. Fundamentals of soft robot locomotion. *Journal of The Royal Society Interface*, 14(130):20170101, 2017.

[5] Hakan Ceylan, Immihan C Yasa, Ugur Kilic, Wenqi Hu, and Metin Sitti. Translational prospects of untethered medical microrobots. *Progress in Biomedical Engineering*, 1(1):012002, 2019.

[6] Konstantinos Chatzilygeroudis, Vassilis Vassiliades, Freek Stulp, Sylvain Calinon, and Jean-Baptiste Mouret. A survey on policy search algorithms for learning robot controllers in a handful of trials. *IEEE Transactions on Robotics*, 36(2):328 – 347, 2019.

[7] Jizhai Cui, Tian-Yun Huang, Zhaochu Luo, Paolo Testa, Hongri Gu, Xiang-Zhong Chen, Bradley J Nelson, and Laura J Heyderman. Nanomagnetic encoding of shape-morphing micromachines. *Nature*, 575(7781):164–168, 2019.

[8] Antoine Cully, Jeff Clune, Danesh Tarapore, and Jean-Baptiste Mouret. Robots that can adapt like animals. *Nature*, 521(7553):503–507, 2015.

[9] Thomas George Thuruthel, Yasmin Ansari, Egidio Falotico, and Cecilia Laschi. Control strategies for soft robotic manipulators: A survey. *Soft Robotics*, 5(2):149–163, 2018.

[10] Zoubin Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521(7553):452–459, 2015.

[11] Evin Gultepe, Jatinder S Randhawa, Sachin Kadam, Sumitaka Yamanaka, Florin M Selaru, Eun J Shin, Anthony N Kalloo, and David H Gracias. Biopsy with thermally-responsive untethered microtools. *Advanced materials*, 25(4):514–519, 2013.

[12] Sami Haddadin, Alessandro De Luca, and Alin Albu-Schäffer. Robot collisions: A survey on detection, isolation, and identification. *IEEE Transactions on Robotics*, 33(6):1292–1312, 2017.

[13] Zhong-Sheng Hou and Zhuo Wang. From model-based control to data-driven control: Survey, classification and perspective. *Information Sciences*, 235:3–35, 2013.

[14] Wenqi Hu, Guo Zhan Lum, Massimo Mastrangeli, and Metin Sitti. Small-scale soft-bodied robot with multimodal locomotion. *Nature*, 554(7690):81–85, 2018.

[15] Josie Hughes, Utku Culha, Fabio Giardina, Fabian Guenther, Andre Rosendo, and Fumiya Iida. Soft manipulators and grippers: a review. *Frontiers in Robotics and AI*, 3: 69, 2016.

[16] Filip Ilievski, Aaron D Mazzeo, Robert F Shepherd, Xin Chen, and George M Whitesides. Soft robotics for chemists. *Angewandte Chemie International Edition*, 50 (8):1890–1895, 2011.

[17] Sangbae Kim, Cecilia Laschi, and Barry Trimmer. Soft robotics: a bioinspired evolution in robotics. *Trends in Biotechnology*, 31(5):287–294, 2013.

[18] Yoonho Kim, Hyunwoo Yuk, Ruike Zhao, Shawn A Chester, and Xuanhe Zhao. Printing ferromagnetic domains for untethered fast-transforming soft materials. *Nature*, 558(7709):274–279, 2018.

[19] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[20] Sam Kriegman, Douglas Blackiston, Michael Levin, and Josh Bongard. A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences*, 117(4):1853–1859, 2020.

[21] Frederick Largilliere, Valerian Verona, Eulalie Coevoet, Mario Sanz-Lopez, Jeremie Dequidt, and Christian Duriez. Real-time control of soft-robots using asynchronous finite element modeling. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2550–2555, 2015.

[22] Thomas Liao, Grant Wang, Brian Yang, Rene Lee, Kristofer Pister, Sergey Levine, and Roberto Calandra. Data-efficient learning of morphology and controller for a microrobot. In *2019 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2488–2494, 2019.

[23] Haojian Lu, Mei Zhang, Yuanyuan Yang, Qiang Huang, Toshio Fukuda, Zuankai Wang, and Yajing Shen. A bioinspired multilegged soft millirobot that functions in both dry and wet conditions. *Nature Communications*, 9 (1):1–7, 2018.

[24] Guo Zhan Lum, Zhou Ye, Xiaoguang Dong, Hamid Marvi, Onder Erin, Wenqi Hu, and Metin Sitti. Shape-programmable magnetic soft matter. *Proceedings of the National Academy of Sciences*, 113(41):E6007–E6015, 2016.

[25] Carmel Majidi. Soft robotics: a perspective—current trends and prospects for the future. *Soft Robotics*, 1(1): 5–11, 2014.

[26] Bradley J Nelson, Ioannis K Kaliakatsos, and Jake J Abbott. Microrobots for minimally invasive medicine. *Annual Review of Biomedical Engineering*, 12:55–85, 2010.

[27] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer, Berlin, Heidelberg, 2003.

[28] Ziyu Ren, Wenqi Hu, Xiaoguang Dong, and Metin Sitti. Multi-functional soft-bodied jellyfish-like swimming. *Nature Communications*, 10(1):1–12, 2019.

[29] Federico Renda, Michele Giorelli, Marcello Calisti, Matteo Cianchetti, and Cecilia Laschi. Dynamic model of a multibending soft robot arm driven by cables. *IEEE Transactions on Robotics*, 30(5):1109–1122, 2014.

[30] Steven I Rich, Robert J Wood, and Carmel Majidi. Untethered soft robotics. *Nature Electronics*, 1(2):102, 2018.

[31] John Rieffel and Jean-Baptiste Mouret. Adaptive and resilient soft tensegrity robots. *Soft Robotics*, 5(3):318–329, 2018.

[32] Daniela Rus and Michael T Tolley. Design, fabrication and control of soft robots. *Nature*, 521(7553):467–475, 2015.

[33] Frank Sehnke, Christian Osendorfer, Thomas Rückstieß, Alex Graves, Jan Peters, and Jürgen Schmidhuber. Parameter-exploring policy gradients. *Neural Networks*, 23(4):551–559, 2010.

[34] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

[35] Metin Sitti. Miniature soft robots—road to the clinic. *Nature Reviews Materials*, 3(6):74–75, 2018.

[36] Metin Sitti, Hakan Ceylan, Wenqi Hu, Joshua Giltinan, Mehmet Turan, Sehyuk Yim, and Eric D Diller. Biomedical applications of untethered mobile milli/microrobots. *Proceedings of the IEEE*, 103(2):205–224, 2015.

[37] Donghoon Son, Hunter Gilbert, and Metin Sitti. Magnetically actuated soft capsule endoscope for fine-needle biopsy. *Soft Robotics*, 7(1):10–21, 2019.

[38] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[39] Barry A. Trimmer and Huai Ti Lin. Bone-free: Soft mechanics for adaptive locomotion. *Integrative and Comparative Biology*, 54(6):1122–1135, 2014.

[40] Alexander von Rohr, Sebastian Trimpe, Alonso Marco, Peer Fischer, and Stefano Palagi. Gait learning for soft microrobots controlled by light fields. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6199–6206, 2018.

[41] Conor Walsh. Human-in-the-loop development of soft wearable robots. *Nature Reviews Materials*, 3(6):78–80, 2018.

[42] Robert J Webster III and Bryan A Jones. Design and kinematic modeling of constant curvature continuum robots: A review. *The International Journal of Robotics Research*, 29(13):1661–1683, 2010.

[43] Tianqi Xu, Jiachen Zhang, Mohammad Salehizadeh, Onaizah Onaizah, and Eric Diller. Millimeter-scale flexible robots with programmable three-dimensional magnetization and motions. *Science Robotics*, 4(29):eaav4494, 2019.

[44] Brian Yang, Grant Wang, Roberto Calandra, Daniel Contreras, Sergey Levine, and Kristofer Pister. Learning flexible and reusable locomotion primitives for a microrobot. *IEEE Robotics and Automation Letters*, 3 (3):1904–1911, 2018.