Multiobjective Optimization for Stiffness and Position Control in a Soft Robot Arm Module

Y. Ansari, *Student Member, IEEE*, M. Manti, E. Falotico, *Member, IEEE*, M. Cianchetti, *Member, IEEE*, and C. Laschi, *Senior Member, IEEE*

***Abstract*—The central concept of this letter is to develop an assistive manipulator that can automate the bathing task for elderly citizens. We propose to exploit principles of soft robotic technologies to design and control a compliant system to ensure safe human–robot interaction, a primary requirement for the task. The overall system is intended to be modular with a proximal segment that provides structural integrity to overcome gravitational challenges and a distal segment to perform the main bathing activities. The focus of this letter is on the design and control of the latter module. The design comprises of alternating tendons and pneumatics in a radial arrangement, which enables elongation, contraction, and omnidirectional bending. Additionally, a synergetic coactivation of cables and tendons in a given configuration allows for stiffness modulation, which is necessary to facilitate washing and scrubbing. The novelty of the work is twofold: 1) Three base cases of antagonistic actuation are identified that enable stiffness variation. Each category is then experimentally characterized by the application of an external force that imposes a linear displacement at the tip in both axial and lateral directions. 2) The development of a novel algorithm based on cooperative multiagent reinforcement learning that simultaneously optimizes stiffness and position. The results highlight the effectiveness of the design and control to contribute toward the development of the assistive device.**

***Index Terms*—Assistive robotics, machine learning, robot control, soft robotics.**

I. Introduction

DESPITE increasing technological advancements and social awareness to promote/adopt a healthy lifestyle, reduced motor functionalities is common in people aged 60 and over. This restricts user autonomy in milieu of strictly intimate daily activities such as bathing, toileting, etc. [1] As a result, users become dependent upon healthcare services which entails not only financial burdens but also critical emotional challenges such as dependency upon others and loss of privacy [2].

Manuscript received February 15, 2017; accepted June 23, 2017. Date of publication July 31, 2017; date of current version August 17, 2017. This letter was recommended for publication by Associate Editor X. Liu and Editor Y. Sun upon evaluation of the reviewers' comments. This work was supported in part by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007–2013/ under REA Grant agreement Smart-e, number #608022, and in part by the European Commission through the I-SUPPORT project #643666. *(Y. Ansari and M. Manti contributed equally to this work.) (Corresponding author: Y. Ansari.)*

The authors are with Scuola Superiore Sant'Anna, BioRobotics Institute, Pontedera 56025, Italy (e-mail: y.ansari@santannapisa.it; m.manti@santannapisa.it; e.falotico@santannapisa.it; m.cianchetti@santannapisa.it; c.laschi@santannapisa.it).

Digital Object Identifier 10.1109/LRA.2017.2734247

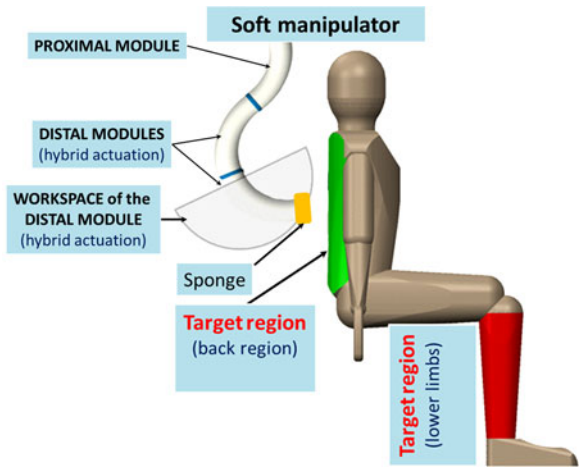

Fig. 1. General overview of the bathing scenario. The soft manipulator reaches potential target regions of human body to perform scrubbing/washing.

Advancements in robotic technologies have facilitated the development of innovative tools aimed to increase user autonomy, studied within the framework of the assistive robotics [3]. This article focuses on the design and control of a robotic manipulator that can function as an automated shower arm for the bathing task, a research topic is still in its infancy.

Existing [4], [5] systems do not automate washing/scrubbing tasks due to the lack of safe human-robot interaction. A promising solution lies in exploiting the principles of soft robotic technologies [6] to design a new generation of manipulators inspired from nature where rigid-less organisms such as the octopus tentacles are capable of advanced manipulation and variable stiffness. They comprise of deformable lightweight actuators [7] allowing for compliance which ensures safe human-robot interaction and highly dexterous motion without any kinematic singularities. Due to these desirable characteristics, these systems have already found increasing applications in industrial and medical sectors [8]–[9], however, this is the first contribution towards its application in an assistive task.

The overall vision for the soft manipulator is to concatenate three modules in a serial manner such that: (i) the proximal segment: is made up of cable-based actuation to compensate for gravitational effects and (ii) the central and distal segments: are made up of hybrid actuation to autonomously reach delicate body parts to perform the main tasks related to bathing. Fig. 1 depicts the hypothetical bathing scenario where the arm has to safely reach the target region(s).

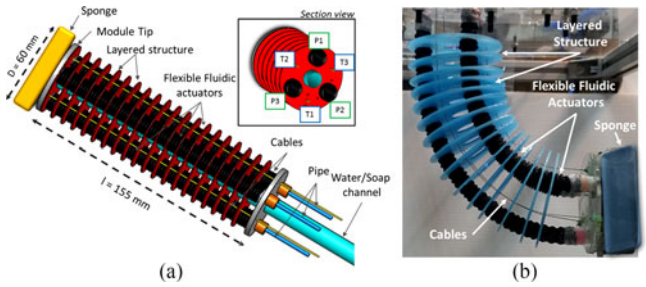

Fig. 2. (a) CAD model of the distal module based on three flexible fluidic actuators (P1, P2, P3) and three tendons (T1, T2, T3) as highlighted in the section view (b) Hardware prototype of distal module.

A robust design and control of the distal module is crucial to perform the main tasks related to bathing, and the initial focus of the research. The readers are referred to [10], where the authors presented the fabrication, kinematic characterization, and kinematic control of the distal module. This article builds on the previous work with three central contributions: (i) the design is briefly presented with a focus on aspects that have been optimized to reduce weight and improve functional capabilities; (ii) the mechanical stiffness variation due to the synergetic co-activation of the hybrid actuators is theoretically and experimentally characterized; (iii) the control algorithm is extended to incorporate multiple objectives to formulate a stiffness and position controller, a previously unaddressed challenge.

Section II briefly discusses the design of the manipulator followed by a theoretical explanation of the stiffness variation capabilities of the system in Section III. Section IV presents the control approach followed by quantitative experimentation in Section V. The problem statement, findings, and possible future outlooks are summarized in the conclusion.

II. Design

The design of the distal module comprises of three pneumatic actuators and three tendons alternately displaced at an angle of 60° along a circle with a radius of 30 mm from the center. The total length is delimited at 150 mm by an acrylic base attached to each end of the module, which has been decreased from the previous 205 mm in order to reduce the weight. A hollow chamber of 7 mm radius may be inserted through the center to facilitate the flow of water/soap. One end of the acrylic base provides a dedicated activation line for each actuator, whereas the other end functions as a base plate where a sponge is attached for washing/scrubbing. Previously, the actuators were covered with a helicoidally-shaped reinforcement structure [10] in order to constrain lateral buckling in the pneumatic chambers (henceforth referred to as chambers). However, it introduced undesirable torsional movements, limiting its bending capabilities. In order to overcome this, a layer-by-layer reinforcement structure has been introduced in this work i.e., each layer has been singularly inserted and fixed to the chambers. The distance between two consecutive layers is 10 mm (≈diameter of the actuator). Consequently, the system is capable of bending with a constant curvature in any direction as highlighted in Fig. 2(b). The total weight of a single distal module including the weight of water and soap in the central channel is approximately 120 g. The characterization of the reachable workspace is discussed and depicted in Section IV-A.

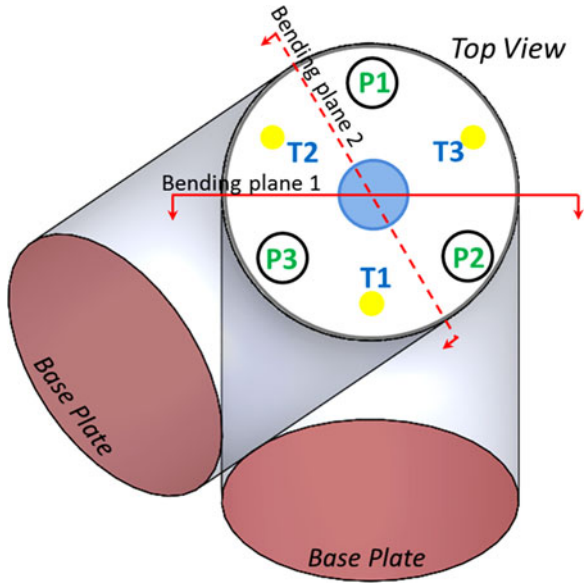

Fig. 3. (a) Bending plane of two tendons (T2-T3) and a one chamber (P1), (b) Bending plane of two chambers (P1-P2) and a one tendon (T3).

III. Stiffness Variation

Stiffness is the ability of an elastic body to resist deformation in response to an applied force. The primary motivation to employ hybrid-actuation is to enable stiffness variation to adapt to different environments. This refers to the ability of a system in a given configuration to change stiffness by modulating the pressure in the chamber(s) with respect the tension in the tendons (also known as antagonistic actuation).

The amount of modulation is dependent upon the actuation limits of the actuator. Hybrid-actuation can be found in several systems [11]–[13]. Similar to [13], there are three base configurations in the distal module in which stiffness can be varied, as explained subsequently (refer to Fig. 3 for actuator arrangement). The overall stiffness of the module in any configuration will be a combination of one or more base configuration.

A. Bending due to Two Tendons and One Chamber

Inflating a single chamber (P1) while simultaneously releasing two lateral tendons (T2, T3) results in a bending motion (refer to bending plane 1 in Fig. 3). In order to vary stiffness, either the tension or the pressure should be modulated.

B. Bending due to One Tendon and Two Chambers

Inflating two chambers (P1, P2) while simultaneously releasing tendon (T3) between the two chambers results in a bending motion (refer to bending plane 2 in Fig. 3). Similar to the previous scenario, the stiffness can be varied by modulating the tension or the pressure.

C. Elongation/Contraction

For elongation, three fluidic actuators (P1, P2, P3) should be inflated while simultaneously releasing the three tendons (T1, T2, T3). Stiffness can be varied by tightening the tendons. For contraction, the three tendons (T1, T2, T3) should be tightened. Stiffness can be varied by applying different pressure levels to the pneumatic chambers (P1, P2, P3).

This work follows the approach in [13]–[15] to experimentally quantify stiffness through the application of an external force that imposes a linear displacement at the tip of the module for a given configuration, resulting in a force-displacement (F-D) graph. In order to characterize stiffness variation for each base case configuration, the tension and pressure will be modulated as follows: (i) for a given tendon actuation, the pressure will be increased gradually from minimum to maximum actuation limits (ii) the tension in the tendon will then be gradually increased and the pressure will be varied similar to (i). The F-D data for each antagonistic actuation will be collected, plotted, and mathematically analyzed revealing the stiffness variation trends in the module (implementation and results are presented in Section V-B).

IV. Control Architecture

Each actuator is considered as an agent that behaves autonomously within the environment. The manipulator is constituted of a group of agents that share this common environment forming a multiagent system (MAS) [16]. In order to reach a point, the agents need to learn to coordinate their behavior. This is facilitated through model-free RL [17] provided that each agent is an independent learner [18] synonymous to the Iterated Prisoner's Dilemma (IPD) [19]. Consequently, this formulates a cooperative multiagent reinforcement learning (MARL) [20] framework.

A. Background

In RL, a single agent interacts with an environment modelled as a Markov decision process (MDP) which comprises of the task-space *S*, the action-set *A*, system dynamics distribution $p(s_{t+1} = s' | s_t = s, a_t = a)$. The agent selects actions in a given state according to a policy $\pi(a_t|s_t)$ and receives a bounded reward $r(s_t, a_t)$. Starting from an initial state, the agent iteratively interacts with the MDP according to the policy to generate a trajectory i.e., a sequence of state-action-reward tuples $s_0 \rightarrow a_0 \rightarrow r_1 \rightarrow s_1 \rightarrow a_1 \rightarrow r_2 \rightarrow s_2 \rightarrow a_2, \ldots$. The return R_t over an infinite horizon trajectory refers to the cumulative discounted reward as $R_t = \sum_{k=0}^{\infty} \gamma^k r_{k+t+1}$ where $\gamma \in [0, 1)$. This work focuses on a finite horizon setting where, $\gamma = 1$. For control tasks, the utility of executing a state-action pair i.e., taking an action *a* in state *s* and following a policy is defined by an action-value function *Q* as follows,

$$Q\,(\mathrm{s}, \mathrm{a}) \;=\; E\{R_t \mid s_t = \; s, \;\; a_t = \; a, \; \pi\} \tag{1}$$

This is extended to multiple agents by assigning each agent a dedicated *Q-function.* The joint-policy of the MAS is $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_m]$ where *m* is the total number of agents. The goal of learning is to extract a control joint-policy from the Q-function that can reach the target from a given starting point while maximizing the accumulated reward.

B. Approximate Reinforcement Learning

Real-world scenarios constitute continuous-domain problems that cannot be represented exactly by a Q-function necessitating approximation methods. This work considers linear function approximators, given a

$$Q\,(\mathrm{s}, \mathrm{a}) \;=\; \sum_{j=1}^{k} \theta_j\,(s, a)\, w_j \tag{2}$$

where, $w \in \mathbb{R}^{\mathrm{k}}$ is the parameter vector; $\theta : S \times A \rightarrow \mathbb{R}^{\mathrm{k}}$ is a vector of binary basis functions $[\theta_1(s, a), \ldots, \theta_k(s, a)]^T$

C. Model-Free Reinforcement Learning

System dynamics of physical systems in stochastic environments is unknown. A model-free control a promising approach in such a case as demonstrated in [21], this work focuses approaches based on temporal difference (TD) based policy iteration (PI) [17] which generates a control policy from a parameterized Q-function. For linear function approximation, on-policy TD learning, widely known as the SARSA algorithm, has clear advantages over off-policy TD learning techniques [22], hence used in this work. Formally, starting from any initial position, PI is a subroutine that runs iteratively for a given duration wherein it alternates the processes of policy evaluation and policy improvement as follows: at time-step *t,* an action in the current state is chosen according to a policy $\boldsymbol{\pi}_t$ and is evaluated by an action-value function Q_t through SARSA with accumulating eligibility traces (also known as SARSA(λ) [23]) by reducing the l-$_2$ norm [refer to (3)–(5)]. In the next time-step $t + 1$, an improved policy $\boldsymbol{\pi}_{t+1}$ is generated by acting greedily with respect to the action-value function Q_t. *(Note: In PI, the Q-function is also referred to as critic and policy as actor).* This process is continued until a stationary policy is obtained i.e., $\pi_t = \pi \; \forall \; t$

$$w_{t+1} = w_t + \alpha_t^* \delta_t^* e_t \tag{3}$$

$$\delta_t = r_{t+1} + \gamma \theta(s_{t+1}, a_{t+1})^T w_t \; - \theta(s_t, a_t)^T w_t \tag{4}$$

$$e_t \;(s, a) = \begin{cases} \gamma \lambda\, e_{t-1}\,(s, a) + 1 & \text{if } \; s = s_t, a = a_t \\ \gamma \lambda\, e_{t-1}\,(s, a) & \text{otherwise} \end{cases} \tag{5}$$

where $\gamma \theta(s_{t+1}, a_{t+1})^T w_t \; - \; \theta(s_t, a_t)^T w_t$ is the temporal-difference error; α is the step-size; e_{t} is the eligibility trace from the backward view of the temporal difference learning where λ is the accumulating trace-decay error; *r* is the scalar reward indicating the value of the action taken in the given state. As mentioned in [24], this approach does not converge in the conventional sense, however, it should not be precluded from generating useful policies as demonstrated in [24]–[26].

D. Reward Structure

The reward is implemented using a distance metric: multiple number of unequally spaced concentric hyperspheres

(with dimensions of the task-space) centered on the target are created. The hypersphere enclosing the target contains only absorbing states. The reward assigned to states between the first and second hypersphere is -1 which is exponentially reduced by a magnitude of 10. Even though this facilitates learning with rich state-space information, the high input- dimensionality combined with the discrete action-set can cause the system to become prone to jittering. In order to overcome this challenge, the policy proceeds by moving through intermediate goals in the state-space i.e., once it encounters a joint-policy that is rewarded with a scalar value from a region with a higher value than any previously encountered ones, it will be the first action taken by the system in the next episode onwards. This allows to learn temporally ordered policies.

E. Multi-Objective Reinforcement Learning

Multiple objectives can be incorporated into an RL framework, known as multi-objective reinforcement learning (MORL) [27], where instead of a scalar reward, the agent now receives a vector of rewards each representing the value for the corresponding objective which may be conflicting, complementing, or independent. The goal is to ensure that the control policy learnt by maximizing these multiple rewards is Pareto optimal. Multiple objectives can interestingly be added into our cooperative MARL framework motivated by reasoning that the joint-policy obtained when independent agents cooperate to learn from a global scalar value to achieve a target is Pareto optimal. Consequently, the underlying idea is to design a mechanism that can transform the multiple objectives into a single objective and treat the problem as a standard single-objective algorithm to generate a single-policy. Unlike the dominant approach which involves the use of a linear/non-linear scalarization function, we take advantage of the information theoretic approach of our reward structure to achieve this as follows: the distance metrics of each objectives is represented as a dimension of a multi-dimensional matrix where the ordered combination of the multiple objectives will index a single scalar value that represents the effectiveness of the joint policy for all objectives (explained in Section V-D).

F. Synthetic Roll-Outs

In order to minimize temporal differences, model-free algorithms require noisy on-policy actions. This is prohibitive for real-world applications due to large sampling complexity which must include both optimal and sub-optimal state-action pairs. One way to avoid these problems while still allowing for a large amount of on-policy exploration is to generate synthetic on-policy trajectories. In this preliminary work, we adopt a simplistic approach via hash-tables to implement this. The schematic view of the control architecture is illustrated in Fig. 4.

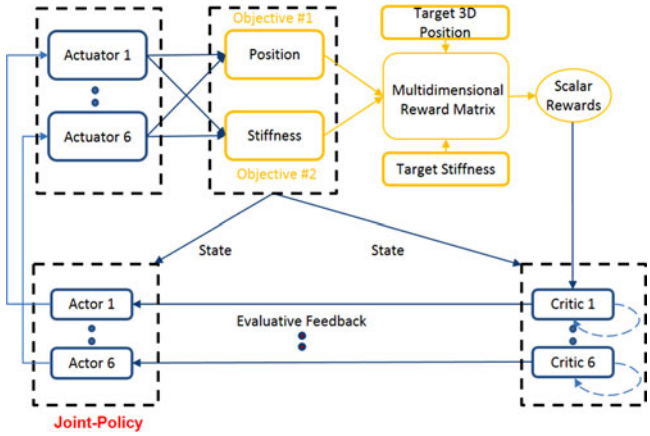

Fig. 4. Schematic view of the control architecture.

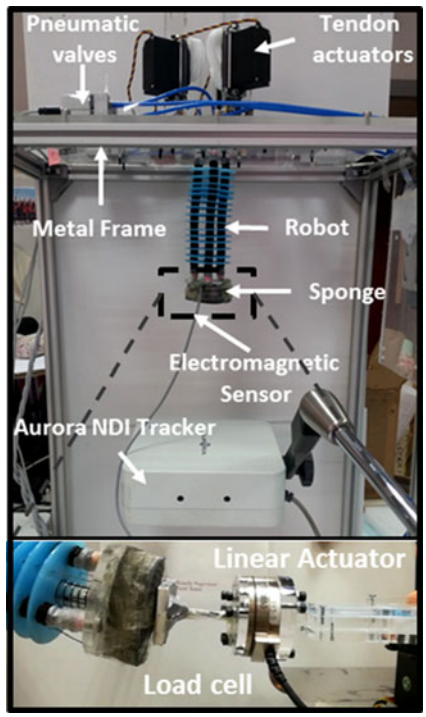

Fig. 5. The experimental set-up: (Top) Metallic frame encompassing the robot and actuation/sensing systems (Bottom) Force measurements with a custom made linear actuator that displaces the load cell 3 cm in 2.2 s.

V. Experiments and Results

The experimental set-up (see Fig. 5) consists of: (1) metal-based rectangular frame encompassing the actuation/sensing systems and the distal module; (2) pneumatic set-up: three proportional pressure-controlled electronic valves (K8P Series, EVP Systems, Output: 0 -3 bar), one filter (EVP Systems: MC-104FB0), one manometer (EVP Systems: M043-p12 0-12 bar), and one stand-alone air compressor, (3) tendon set-up: three Hs-785hb Hitech Sail Winch Motors; (4) six-DOF electromagnetic sensing probe (Aurora Northern Digital Inc.) fixed at the tip of the module base capable of sub-millimeter position tracking (5) F-D Measurements: (i) ATI Mini 45© load cell (ii) custom made linear actuator based on a slider crank mechanism using Hs 785hb and an acrylic sheet (8 mm) with a stroke of 3 cm in 2.2 s.

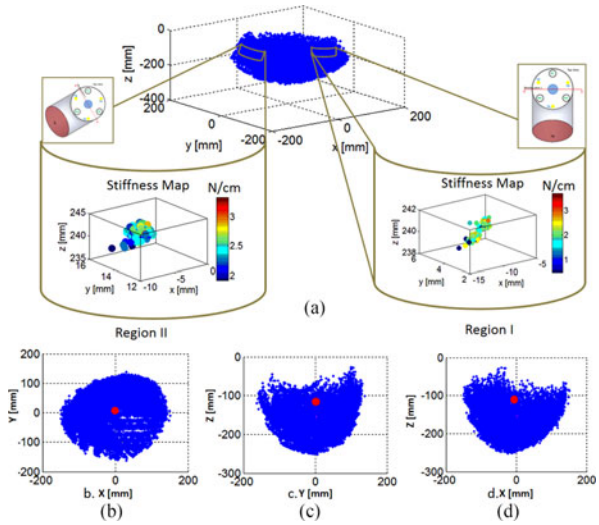

Fig. 6. (a) Isometric workspace view. The stiffness variation (referred to as the stiffness map) in two regions corresponding to the two bending base configurations is also highlighted. (b) Workspace view in the XY plane (c) Workspace view in the YZ plane (d) Workspace view in the XZ plane.

A. Workspace Identification

A set represented by a hyper-rectangle in six dimensions was used to collect a total of 46656 points at 2Hz from a combination of: (i) servomotors: 0° to 90° in steps of 15° *(Note: 10° corresponds to a tendon release of 0.3 cm)* (ii) pneumatic pressure: 0.9 bar (inferior limit of the actuator [28]) to 1.7 bar in steps of 0.2 bar. Fig. 6 depicts the resulting point cloud from an isometric [see Fig. 6(a)] and planar view [see Fig. 6(b)–(d)]. The reachable workspace size is 29 cm × 27.6 cm × 20 cm in the x-y-z axes, respectively. The red marker corresponds to the starting configuration. The overall shape is a symmetric volumetric convex with the lower limits defined by the maximum bending capabilities (approximately 180°) of the module, as expected. The minor tilt in the planar view is due to a small variation of the tension in the tendons in the initial configuration.

B. Stiffness Characterization

In order to characterize stiffness capabilities of each base configuration (Section III), axial and lateral force-displacement data have been collected for the following actuation (viewers are referred to the complementary video for better understanding):

1) *Configuration I:* Two tendons are released corresponding to 30° followed by inflating the chamber in between from 0.9 bar to 1.7 bar in increments of 0.2 bar. The other tendon and chambers remains at 0° and 0 bar, respectively
2) *Configuration II:* One tendon is released corresponding to 30° followed by inflating the adjacent lateral chambers from 0.9 bar to 1.7 bar in increments of 0.2 bar. The other two tendons and chamber remain at 0° and 0 bar, respectively
3) *Configuration III:* All tendons are released corresponding to 30° of rotation followed by inflating all chambers from 0.9 bar to 1.7 bar in increments of 0.2 bar. The results provided in this configuration corresponds to a displacement of maximum 1 cm due to buckling effects at higher displacements.
4) All experiments are then repeated with the same procedure for a tendon release of 50°. The reading for an individual experiment was repeated three times.

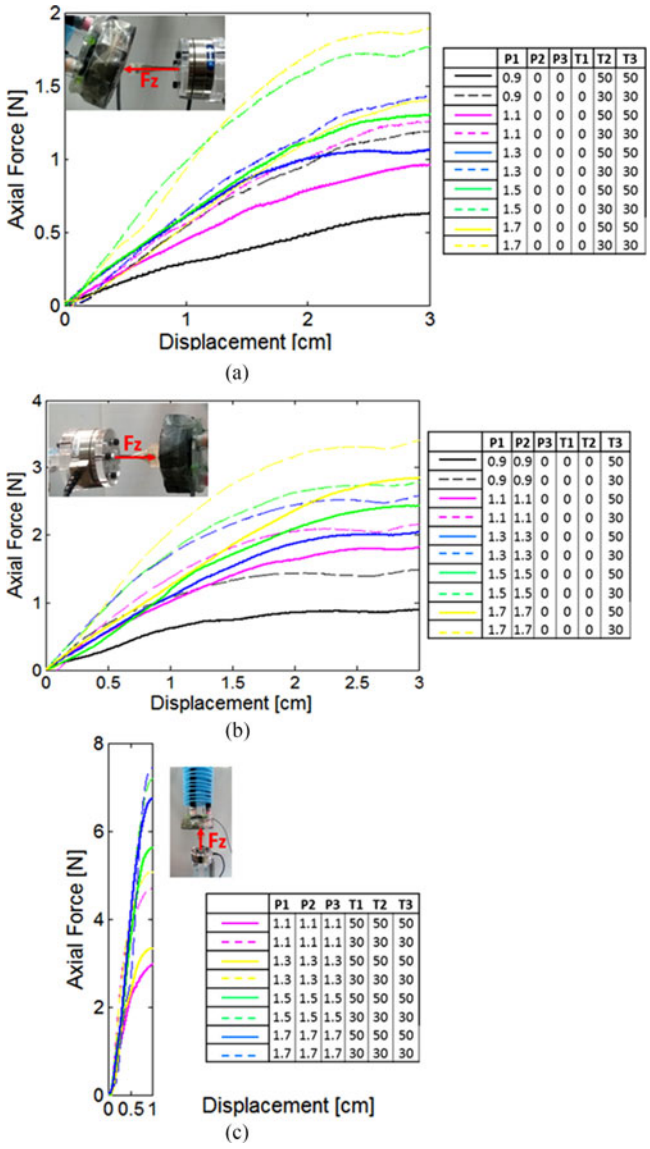

Fig. 7. Axial force-displacement measurements (a) Configuration I: bending by two tendons and one chamber (b) Configuration II: bending by one tendon and two chambers (c) Configuration III: elongation/contraction by three tendons and three chambers.

The axial force measurements for each configuration is reported in Fig. 7. As a general response, pressurizing the chambers for a given tendon actuation results in an increase in stiffness. The force variability (maximum increase in axial force) in each configuration is (i) at 30°: 50%, 126%, and 60% respectively; (ii) at 50°: 100%, 222%, and 123% respectively. It is interesting to note that even though the force variability of the latter case is much higher than the former, the actual force values are overall lower. This is because a higher rotation in the servomotor corresponds to a significant release of

	CONTRACTION/ELONGATION		BENDING			
	Three Chambers, Three Tendons		Two Chambers, One Tendon		One Chamber, Two Tendons	
	30°	50°	30°	50°	30°	50°
0.9 bar	1.2 N	0.5 N	1.4 N	1.2 N	0.6 N	0.5 N
1.1 bar	1.3 N	0.5 N	1.4 N	1.2 N	0.6 N	0.5 N
1.3 bar	1.5 N	1 N	1.4 N	1.2 N	0.9 N	0.9 N
1.5 bar	1.7 N	1 N	1.4 N	1.2 N	0.9 N	0.9 N
1.7 bar	2.1 N	1 N	1.4 N	1.2 N	0.9 N	0.9 N
Total Stiffness Variability	75%	100%	0%	0%	50%	80%

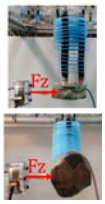

Fig. 8. Summary of lateral forces-displacement measurements in elongation/contraction and bending configurations.

tendon length which allows for more dexterity and correspondingly lower stiffness. This is useful in the showering task as the system can move towards and around the human with lower stiffness, whereas for scrubbing/washing in dedicated locations, the stiffness can be increased. Furthermore, according to the data recorded in [29] where forces from real simulated showering scenarios have been collected, the range exhibited by this system is sufficient to achieve this task. The results from all three lateral experiments are summarized in Fig. 8, where the trends are consistent with the results from the axial data.

C. Memory

For the control algorithm, firstly a motor-space and task-space mapping in the form of a hash-table was collected for each bending configuration mentioned in Section V-B as follows: (i) Motor-Space: for each bending configuration, the corresponding servomotor value was restricted to 50° after which all three chambers were activated with a pressure combination from 0.9 bar to 1.7 bar in increments of 0.2 bar constituting 125 combinations per region; (ii) Task-Space: for each actuation, the corresponding cartesian coordinates and axial stiffness at the tip was recorded. The former was collected from the Aurora NDI tracker whereas the latter was calculated by dividing the force obtained from the load cell that displaced the tip axially at 1 cm with the help of the linear actuator (considering an approximately linear response until this distance as seen in Fig. 7). The task-space values are plotted in Fig. 6(a), which we refer to as a stiffness map. For the given actuation range, it shows a localized reachability with a high stiffness variability. Thus, the challenge of learning is to achieve high positioning accuracy while simultaneously optimizing for a certain stiffness value.

D. Multidimensional Reward Matrix

This work considers two independent objectives of position and stiffness. For each objective, 5 regions are created and awarded values starting from -10e3 exponentially reduced by a power of 10 until the absorbing states. In order to implement the reward matrix, the two objectives are combined such that the absorbing state is obtained when the system is within 1.5 mm 3D space and 0.2 N/cm of the desired stiffness in that position.

E. Function Approximation

Tile coding was the opted linear function approximator, implemented through typically available software which creates infinite, axis-parallel tilings over the continuous state variable according to the following values: $\mathrm{R} = [0.36\ 0.46]$ m, $\varphi = [0^\circ\ 360^\circ]$, $\theta = [0^\circ\ 90^\circ]$, $\mathrm{S} = [0.5\ 3.7]$ N/cm. The action-set is discrete and heuristically initialized such that: (a) pneumatics: the pressure can decrease/increase in magnitude of 0.2 bar or keep unchanged within an actuation limit of [0.9 1.7] bar (b) tendons: for the two configurations under consideration, two servomotor values always remain constant while the third shifts between 0° and 50°. This implies that four critics are sufficient for this work. The parameter vector and eligibility traces are stored with the help of open-addressed hashing. Each actor follows the ubiquitous Є –greedy policy where exploration (Є) is set to 0.1 and reduced by half every time the goal is reached.

The tiling and learning parameters were selected after a systematic process that was tested on the data from memory given the following constraints [30]: (i) the starting/target position is kept constant during learning; (ii) once parameters are selected, the algorithm is repeatedly tested on 10 different starting/target positions for consistency; (iii) after heuristically selecting a large number of tiles/tilings the dimensions were gradually reduced while checking for consistent behavior in the task-space until 10, 10, 20, and 35 tiles were selected. 4 similar tilings are then offset with respect to one another such that the state space has a total of 280000 tiles; (iv) the step-size (α) was set at $\alpha = 0.16$, whereas the eligibility traces (λ) is set to 0.9 for the current investigations.

F. Learning

Two different stiffness values were selected in each configuration, highlighted in magenta and cyan in Fig. 9(a) and Fig. 8(c), respectively. The goal of learning is to reach the correct region with the position and stiffness accuracy criteria mentioned in Section V-D. An experiment consists of initializing the parameter vector optimistically, and running the algorithm for a total of 300 episodes with 20 trials per episode. The initial position is fixed in the inferior most actuation values for base configuration 1. This starting position will have different distances from different targets, hence, as all the other algorithmic parameters are kept constant, reaching all the target will prove useful to qualify the effectiveness of the algorithm. Once convergence is achieved, the joint-policy is applied directly on the manipulator and the performance is then evaluated according to two criteria: (a) mean reaching error (b) time required to generate a solution. Each experimental trial is repeated five times for robustness.

The results of the experiments can be seen via the accumulated rewards and the colored markers plotted on the 2D reward matrix and in Fig. 9(a) and (c), where each dimension corresponds to an objective. The location of the marker on the reward matrix indicates that the algorithm is not only able to generate the solution while optimizing independent objectives simultaneously for the given criteria measures, but also generate a solution that behaves consistently for repeated trials. One might argue that the algorithm is only selecting the same solution repeatedly, however, the variance in the generated solution set in Fig. 9(c) for a target stiffness value of 2.8 N highlights that infact this is not the case.

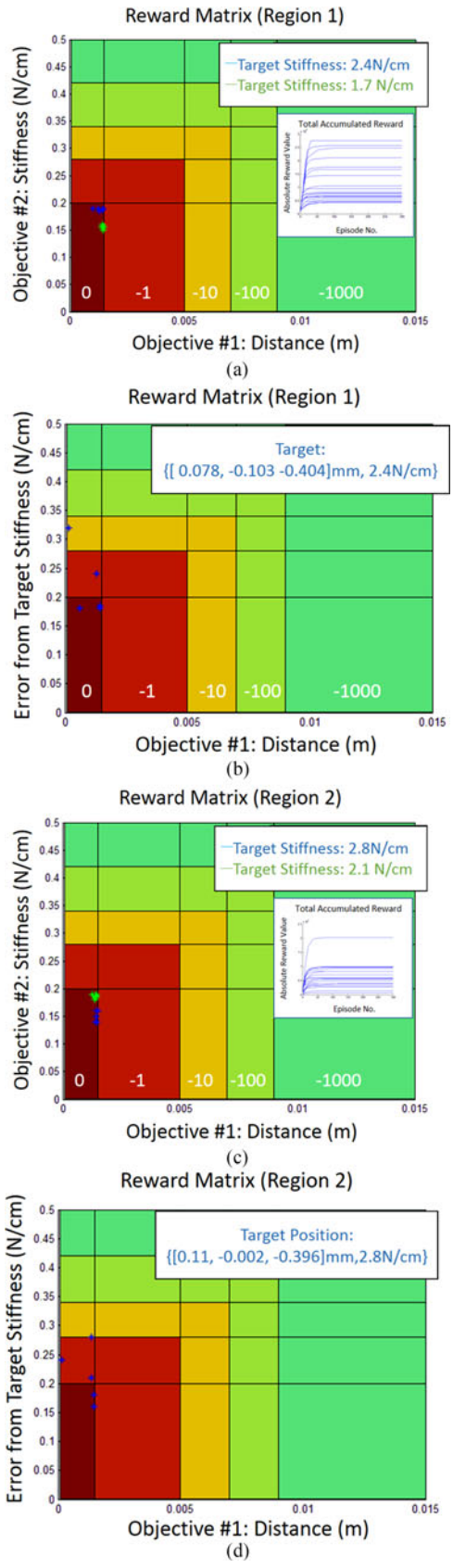

Fig. 9. (a) & (c) Configuration I and II: Optimized for position and stiffness (b) & (d) Configuration I and II: optimized for position only.

In order to validate the algorithm, a single experiment from each region was repeated again, however, only optimizing the position. The corresponding solutions are depicted in Fig. 9(c) and (d). It clearly shows that the solutions found lie within the required positional accuracy, but vary across the stiffness range. It is also worthy to note that the absorbing state can be easily changed within the matrix to actively control the priority of the obtained solutions.

VI. Conclusion

In this letter, we developed an assistive soft robotic manipulator that can automate the showering task for the elderly community. As an initial step, we investigate the design and control of a single hybrid-actuated distal module. In particular, we identify the three base cases of antagonistic actuation that enable stiffness variation, necessary to perform bathing tasks. We experimentally characterized the stiffness trends through axial and lateral force-displacement measurements. The overall stiffness range was found to lie between 0.3 N/cm and 7.2 N/cm, which, according to literature, is found to be sufficient to achieve the task at hand [29]. We also developed an approach to automate the stiffness and position capabilities simultaneously through a cooperative multiagent reinforcement learning control approach that can be leveraged to optimize multiple objectives with high accuracy.

References

[1] T. Bock, C. Georgoulas, and T. Linner, "Towards robotic assisted hygienic services: Concept for assisting and automating daily activities in the bathroom," *Gerontechnology*, vol. 11, no. 2, p. 362, Jun. 2012.
[2] S. Bedaf, H. Draper, G. J. Gelderblom, T. Sorell, and L. de Witte, "Can a service robot which supports independent living of older people disobey a command? The views of older people, informal carers and professional caregivers on the acceptability of robots," *Int. J. Social Robot.*, vol. 8, no. 3, pp. 409–420, Jun. 2016.
[3] E. de Sousa Leite *et al.*, "Influence of assistive technology for the maintenance of the functionality of elderly people: An integrative review," *Int. Arch. Med.*, vol. 9, Mar. 2016.
[4] 2016. [Online]. Available: http://www.seatedshower.com/
[5] 2016. [Online]. Available: http://www.drivemedical.co.uk/sections/bathroom-toilet-aids
[6] C. Laschi, B. Mazzolai, and M. Cianchetti, "Soft robotics: Technologies and systems pushing the boundaries of robot abilities," *Sci. Robot.*, vol. 1, no. 1, 2016, p. eaah3690.
[7] C. Laschi and M. Cianchetti,"Soft robotics: New perspectives for robot bodyware and control," *Frontiers Bioeng. Biotechnol.*, vol. 2, no. 3, 2014.
[8] A. Grzesiak, R. Becker, and A. Verl, "The bionic handling assistant—A success story of additive manufacturing," *Assem. Autom.*, vol. 31, no. 4, 2011.
[9] T. Ranzani, M. Cianchetti, G. Gerboni, I. De Falco, and A. Menciassi, "A soft modular manipulator for minimally invasive surgery: Design and characterization of a single module," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 187–200, Feb. 2016.
[10] Y. Ansari, M. Manti, E. Falotico, Y. Mollard, M. Cianchetti, and C. Laschi, "Towards the development of a soft manipulator as an assistive robot for personal care of elderly people," *Int. J. Adv. Robot. Syst.*, vol. 14, no. 2, 2017.
[11] G. Immega and K. Antonelli, "The KSI tentacle manipulator,". in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1995, vol. 3, pp. 3149–3154.
[12] W. McMahan, B. A. Jones, and I. D. Walker, "Design and implementation of a multi-section continuum robot: Air-Octor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Aug. 2005, pp. 2578–2585.

[13] A. Shiva *et al.*, "Tendon-based stiffening for a pneumatically actuated soft manipulator," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 632–637, Jul. 2016.
[14] M. Cianchetti, T. Ranzani, G. Gerboni, I. De Falco, C. Laschi, and A. Menciassi, "STIFF-FLOP surgical manipulator: Mechanical design and experimental characterization of the single module," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Nov. 2013, pp. 3576–3581.
[15] J. Li *et al.*, "Stiffness characteristics of soft finger with embedded SMA fibers," *Composite Struct.*, vol. 160, pp. 758–764, 2017.
[16] M. Wooldridge, *An Introduction to Multiagent Systems*. Hoboken, NJ, USA: Wiley, 2009.
[17] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.
[18] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," in *Proc. 15th Proc. Nat./10th Conf. Artif. Intell./Innov. Appl. Artif. Intell.*, Jul. 1998, pp. 746–752.
[19] T. W. Sandholm and R. H. Crites, "Multiagent reinforcement learning in the iterated prisoner's dilemma," *Biosystems*, vol. 37, no. 1, pp. 147–166, 1996.
[20] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 156–172, Mar. 2008.
[21] V. Vishesh, P. Grover, and B. Trimmer "Model-free control framework for multi-limb soft robots," in *Proc. IEEE/RSJ Int. Conf. IEEE Intell. Robot. Syst.*,, 2015, pp. 1111–1116.
[22] L. Baird, "Residual algorithms: Reinforcement learning with function approximation," in *Proc. 12th Int. Conf. Mach. Learn.*, Tahoe City, CA, USA, Jul. 1995, pp. 30–37.
[23] G. A. Rummery and M. Niranjan, "On-line Q-learning using connectionist systems," Dept. Eng., Univ. Cambridge, Cambridge, U.K., Rep. TR-166, 1994.
[24] S. Peter and R. S. Sutton, "Scaling reinforcement learning toward RoboCup soccer," in *Proc. Int. Conf. Mach. Learn.*, 2001, vol. 1, pp. 537–544.
[25] Y. Ansari, E. Falotico, Y. Mollard, B. Busche, M. Cianchetti, and C. Laschi, "A multi-agent reinforcement learning approach for inverse kinematics of high dimensional manipulators with precision positioning," in *Proc. 6th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics*, Jun. 2016, pp. 457–463.
[26] S. Adam, L. Busoniu, and R. Babuska, "Experience replay for real-time reinforcement learning control," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 2, pp. 201–212, Mar. 2012.
[27] D. M. Roijers *et al.*, "A survey of multi-objective sequential decision-making," *J. Artif. Intell. Res.*, vol. 48, pp. 67–113, 2013.
[28] M. Manti, A. Pratesi, E. Falotico, M. Cianchetti, and C. Laschi, "Soft assistive robot for personal care of elderly people," in *Proc. 6th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics*, Jun. 2016, pp. 833–838.
[29] C. Mandery, Ö. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "The KIT whole-body human motion database," in *Proc. Int. Conf. Adv. Robot.*, 2015, pp. 329–336.
[30] Y. Ansari, E. Falotico, M. Cianchetti, and C. Laschi, "Point-to-point motion controller for soft robotic manipulators," in *Proc. IEEE Int. Conf. Simul., Model. Program. Auton. Robots*, Dec. 2016, pp. 49–54.