Adaptive Force and Position Control of Magnetic Endoscopes using Reinforcement Learning

Antonio Marino

Adaptive Force and Position Control of Magnetic Endoscopes using Reinforcement Learning

Antonio Marino[1], Bruno Scaglioni[1], Keith L. Obstein[2], Pietro Valdastri[1]

***Abstract*—In this work, an adaptive control framework based on model-free reinforcement learning is applied to a magnetically manipulated endoscope to simultaneously control the endoscope position and adapt to the highly variable, dynamic, unstructured environment of the human gastrointestinal tract. The ability to adapt control parameters enables the endoscope to overcome obstacles and successfully navigate the tract in the absence of a motion planning algorithm. This also facilitates maintenance of optimal contact with the tissues, which is beneficial for multiple diagnostic applications—including micro-ultrasound imaging. The approach is experimentally validated using the Magnetic Flexible Endoscope in a benchtop colon simulator through (a) execution of forward and backward motion with contact force control, (b) effectiveness at navigating a sharp turn, and (c) successfully navigating an unmodelled obstacle. Overall, this is the first example of model-free adaptive control being successfully applied to magnetic manipulation and lays the groundwork for the further development of advanced motion planning and autonomous navigation algorithms.**

***Index Terms*—magnetic control, magnetic flexible endoscopy, reinforcement learning.**

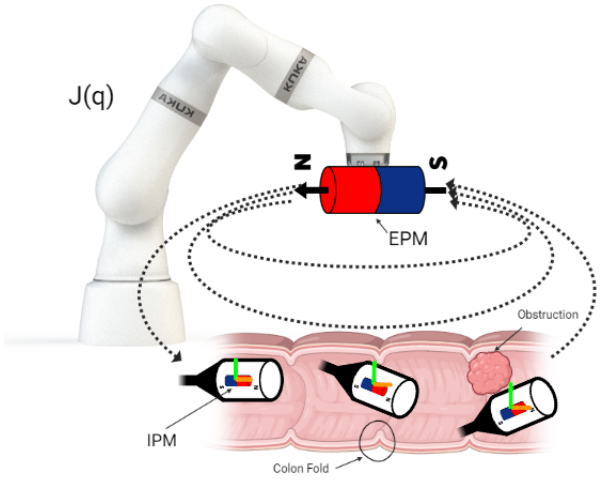


Fig. 1: Magnetic manipulation of the MFE endoscope inside the colon

I. INTRODUCTION

Colonoscopy is a common medical procedure that requires navigation of the human colon using a mechanical instrument that is actuated by pushing the endoscope from the insertion tube (the distal aspect of the endoscope) to facilitate propagation of the endoscopes proximal end. This unfortunately leads to tissue deformation due to the forces that are exerted along the colon that causes at minimum patient discomfort and at worse perforation or bleeding. Therefore, magnetic manipulation has been utilised in robotic endoscopy due to its ability to transmit forces through the skin surface for actuation of an endoscope within the gastrointestinal tract [1]–[4]. This allows for minimal interaction between the patient tissues and the endoscope thereby reducing patient discomfort and risk of perforation or bleeding. [5], [6]. The magnetic systems that are utilized are based on either permanent magnets [13] or electromagnetics [14] while some engage more complex with a combination of multi-magnets interaction [7], [8] or compliant elements [9].

While magnetic manipulation of a colonoscope has been successful, accurate control is challenging due to the nonlinear nature of magnetic coupling and the challenging, variable, dynamic environment of the human colon [10]. Nonetheless, previous control approaches [11], [26], [27] have been able to successfully navigate the colon using both open-loop and closed-loop strategies; this latter reduces navigation time and a clinicians cognitive and technical task burden. More recently, using our novel Magnetic Flexible Endoscope (MFE), our team has demonstrated how intelligent control enables endoscope teleoperation and autonomous motion [12]. While successful, the autonomous navigation mechanism is based on processing of images obtained from the endoscope camera, which can be hindered by faeces that are retained despite the use of a bowel preparation and occlusions (tight fixed angulations; stenosis). Therefore, an alternative mechanism through which the colonoscope can actively adapt to the environment of the colon is needed to facilitate autonomous obstacle navigation and endoscope propagation. Integration of micro-ultrasound has been explored as a promising avenue [16]; however, the control systems needed for coupling and navigation are unable to perform well given the highly variable nature of the colon. To overcome this challenge, our team has developed a novel adaptive algorithm that aims at controlling the endoscope position on the degrees of freedom associated with the navigation, while simultaneously controlling the

The work was supported by the Royal Society, Cancer Research UK (CRUK) Early Detection and Diagnosis Research Committee (award no. 27744), the National Institute of Biomedical Imaging, Bioengineering of the National Institutes of Health (NIH; award no. 2R01EB018992-05), by the European Research Council (ERC) (grant agreement no. 818045) and under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 952118 (AUTOCAPSULE). Any opinions, findings and conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the above funding bodies.

[1] The authors are with the STORM Lab UK, School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK. elamari, b.scaglioni, p.valdastri[at]leeds.ac.uk.

[2]STORM Lab USA, Vanderbilt University, Nashville, TN, USA. 5 Vanderbilt University Medical Center, Nashville, TN, USA.

forces resulting from the interaction with the tissue in the orthogonal direction with respect to the direction of motion. The algorithm takes advantage of the adaptability of the control system, in the absence of a trajectory planning algorithm, to provide system stability under the condition of bounded disturbances. The parallel position and force control enables use of contact-based diagnostic sensing such as micro-ultrasound during motion and enhances the system autonomy while simplifying the operator task load. Despite examples of adaptive control of magnetic endoscopes [26] in the literature, all rely on restrictive modelling assumptions about the endoscope and the environment. Given the unstructured colon environment, which can undergo significant changes with perturbation [15], these modelling assumptions can rapidly become problematic. For this reason, we explore the adaptation of the endoscope control system to the environment (Fig. 1), implemented through a model-free reinforcement learning algorithm, which does not require a model of the controlled system [19], [20]. Reinforcement learning is capable of learning complex control policies from experience [21], [22]. Although popular solutions involve actor-critic models and deep reinforcement learning [21], [23], we opt for a linear approximation approach. Although sub-optimal from a dynamic programming point of view [17], [18], this approach has similar adaptation abilities and superior execution times [24] with respect to the aforementioned techniques. Regularity condition on the exploration policy is introduced to guarantee algorithm convergence.

This paper is organized as follows: Section II presents an overview of the magnetic interaction and the system dynamics; Section III presents the control approach and defines its parameters; Section IV describes the reinforcement learning algorithm to update the control parameters; and Section V and VI present the experiments.

II. System modelling

The approach described in this paper could be applied to a wide range of magnetic systems. Without loss of generality, the discussion will be carried out on the Magnetic Flexible Endoscope (MFE) platform, developed by our group and widely validated [25]–[27]. The system consists of two components: the endoscope containing an internal permanent magnet (IPM) and a robotic arm that manipulates the External Permanent Magnet (EPM). Both are shown in Fig. 1. The MFE is equipped with a localization system capable of detecting position and orientation of the endoscope tip in a range of 50 cm from the robot end-effector in every direction, thus enabling closed-loop control of the endoscope, as shown in several papers [25]–[27]. The pose information provided by the localization system enables computation of the intermagnetic forces, which depends on the relative distance and orientation of EPM and IPM. To this end, we used the dipole model reported in (11) (12).

The magnetic endoscope is modelled as a rigid body representing the endoscope tip, while the tether is considered a source of disturbances. The endoscope dynamics is described by the following equation:

$$B(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t) + C(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t))\dot{\boldsymbol{x}}(\boldsymbol{t}) + G(\boldsymbol{x}(t)) = \boldsymbol{\tau}_m(t) + \boldsymbol{v}(t) \tag{1}$$

where the matrices $B(x(t))$, $C(x(t), \dot{x}(t))$, $G(x(t))$ represent the inertia, Coriolis and gravitational terms. The endoscope dynamics is affected by exogenous disturbances, represented by the unknown input $\boldsymbol{v(t)}$, arising in the presence of contact forces, friction and interaction with the tether. The effect of the term $\boldsymbol{v(t)}$ on stability will be discussed in appendix B. The endoscope tip contains a small permanent magnet (IPM) actuated by a permanent external magnet (EPM) attached to a manipulator arm. The magnetic force-torque wrench acting on the endoscope $\boldsymbol{\tau_m}(t) \in R^6$, generated through magnetic coupling between IPM and EPM, are computed employing the widely used magnetic dipole model [39], those equations are shown in appendix A. The dipole model eqs. (11), (12) are highly non-linear; however, their local variation, approximated to the first derivative, is more suitable for the synthesis of a control algorithm because it is linear in the control variables, i.e. the joint velocities ($\dot{q}$)

$$\dot{\boldsymbol{\tau}}_{\boldsymbol{m}}(t) = J_{\boldsymbol{x}}(t)\dot{\boldsymbol{x}}(t) + J_{\boldsymbol{q}}(t)\dot{\boldsymbol{q}}(t) \tag{2}$$

where $\boldsymbol{q} \in R^m$ are the robot joint variables, $\boldsymbol{x} \in R^6$ is the endoscope pose (position and orientation), while $J_{\boldsymbol{x}}$, $J_{\boldsymbol{q}}$ are the Jacobian matrices representing the relationship between the variation of the magnetic torque and respectively the endoscope velocities ($\dot{\boldsymbol{x}}$) and the joints velocities ($\dot{\boldsymbol{q}}$). Describing the interaction between two magnets, the matrices J_x and J_q are rank deficient because the EPM magnetic field is symmetric along the magnetization axis. Hence the system controllability is reduced to five degrees of freedom. However, the MFE is designed in such a way to associate the uncontrollable DoF. with the roll of the tip. In the following, we will omit the dependency on time.

Combining (2) and (1) and defining the control input $\dot{\boldsymbol{\tau}}_{\boldsymbol{m}} = \boldsymbol{u}$, the overall system has the following expression:

$$\begin{cases} B(\boldsymbol{x})\ddot{\boldsymbol{x}} + C(\boldsymbol{x}, \dot{\boldsymbol{x}})\dot{\boldsymbol{x}} + G(\boldsymbol{x}) = \boldsymbol{\tau}_m + \boldsymbol{v} \\ \dot{\boldsymbol{\tau}}_{\boldsymbol{m}} = \boldsymbol{u} \end{cases} \tag{3}$$

the control goal is to find $\boldsymbol{u}$ such that ($\boldsymbol{x}$,$\boldsymbol{\tau_m}$) approach a desired value ($\boldsymbol{x_d}$,$\boldsymbol{\tau_d}$). To this end, the control u is converted in joint velocities by means of eq.(2) and the desired joint values are sent to the low-level robot controller. The ability to simultaneously achieve desired $\boldsymbol{x_d}$ and $\boldsymbol{\tau_d}$ is given by the actuation method. The intensity of the magnetic field generated by the EPM can be modulated by changing the intermagnetic distance while the direction and gradient of the field depend on the relative orientation, as shown in appendix A. Since we are interested only in the interaction between the tip and the colon tissues, we can assume that the attraction force between the magnet and the endoscope tip is equal to the reaction forces exerted by the colon tissues on the same tip. Therefore, we will estimate the force applied on the tissue by the endoscope tip applying the dipole model (11), (12).

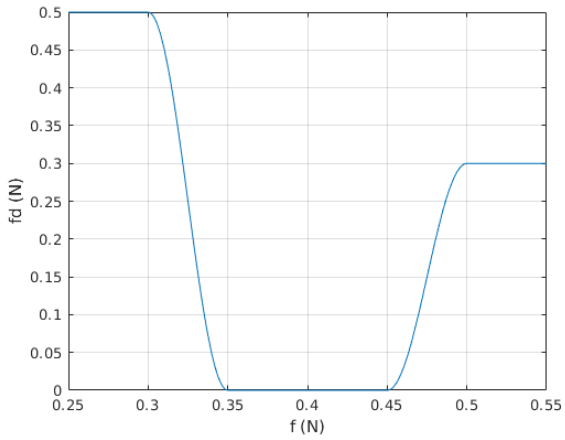


Fig. 2: desired force transition for $\gamma_1, \gamma_2 = 0.05$

III. Control Strategy

In this work, $\boldsymbol{x_d} \in R^6$ is computed as a point to point motion, therefore $\dot{\boldsymbol{x}}_{\boldsymbol{d}} = \mathbf{0}$. The acceptable range of $\boldsymbol{f_d}$ spans between $\boldsymbol{f_l} \in R^3$ and $\boldsymbol{f_u} \in R^3$, defined as lower and upper force limits. $\boldsymbol{f_l}$ is the minimum force required by the application (e.g. micro-ultrasound coupling [16]), while $\boldsymbol{f_u}$ is a limit imposed to prevent excessive tissue stretching. Any value of $\boldsymbol{f_d}$ within the range $[\boldsymbol{f_l}, \boldsymbol{f_u}]$ is acceptable. Hence, when the interaction force is inside the range, the desired force smoothly transits to $\mathbf{0}$ to avoid sharp changes. Each component i of $\boldsymbol{f_d}$ is defined as:

$$f_{di} = \begin{cases} f_{li} & |f_i| - |f_{ui}| > 0 \\ f_{sl} & -\gamma_1 \leq |f_i| - |f_{ui}| \leq 0 \\ 0 & |f_i| - |f_{ui}| < -\gamma_1 \quad and \quad |f_i| - |f_{li}| > \gamma_2 \\ f_{su} & 0 \leq |f_i| - |f_{li}| \leq \gamma_2 \\ f_{ui} & |f_i| - |f_{li}| < 0 \end{cases}$$
$$f_{su} = (\frac{1}{2} cos(\frac{(|f_i| - |f_{li}| - \gamma_2)\pi}{\gamma_2} + \pi) + \frac{1}{2}) f_{ui}$$
$$f_{sl} = (\frac{1}{2} cos(\frac{(|f_i| - |f_{ui}|)\pi}{\gamma_1}) + \frac{1}{2}) f_{li} \tag{4}$$

where f_{su}, f_{sl} are the results of a smooth transition in the range 0 and f_{li} or f_{ui}, with $|f_{ui}| > |f_{li}| > 0$. The transitions are regulated by two additional variables $\gamma_1, \gamma_2 > 0$. The higher these variables are, the more the transition is smoother and the more the range of the desired force equal to zero is reduced (Fig. 2). γ_1, γ_2 are chosen empirically. By denoting the desired wrench $\boldsymbol{\tau_d} = \begin{bmatrix} \boldsymbol{f_d} & \mathbf{0} \end{bmatrix}^T$ the torque error $\tilde{\boldsymbol{\tau}}$ can be defined as:

$$\tilde{\boldsymbol{\tau}} = \boldsymbol{\tau_d} + K_p \tilde{\boldsymbol{x}} - K_p \dot{\boldsymbol{x}} + G(\boldsymbol{x}) - \boldsymbol{\tau_m} \tag{5}$$

where $\tilde{\boldsymbol{x}}$ is the pose error computed as in [39]. The error takes into account that the magnetic manipulation of the endoscope is a five-DoF. $K_p \in R^{6\times6}$ is the matrix of control gains for position error and the velocity. Both actions are required to guarantee the system stability, as shown later. Moreover, the

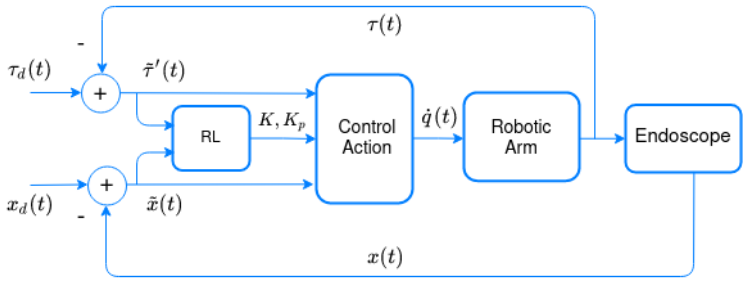


Fig. 3: Control Structure

damping velocity term is beneficial since the motion is in general discontinuous due to the colon folds. We evaluated that, although not meaningful from the physical perspective, posing the velocity gains equal to the position gains in K_p is a convenient simplification that reduces the burden on the learning algorithm. This choice allows to reduce the number of parameters while simultaneously maintaining a relation between the learning adaptation and the derivative controller. Other solutions, like fixing the velocity gain or increasing the number of parameters, would marginally improve the control performances at the expense of slower convergence of the learning procedure. By defining $\boldsymbol{\tau_d'} = \boldsymbol{\tau_d} + K_p\tilde{\boldsymbol{x}} - K_p\dot{\boldsymbol{x}} + G(\boldsymbol{x})$, and the torque error $\tilde{\boldsymbol{\tau}} = \boldsymbol{\tau_d'} - \boldsymbol{\tau_m}$, the control input is expressed as follow:

$$\boldsymbol{u} = K\tilde{\boldsymbol{\tau}} + \tilde{\boldsymbol{\tau}}^{\dagger}\boldsymbol{\tau_d}^T\dot{\boldsymbol{x}} - \dot{\boldsymbol{x}} \tag{6}$$

that ensure a torque error closed loop dynamics equal to:

$$\dot{\tilde{\boldsymbol{\tau}}} = -K\tilde{\boldsymbol{\tau}} - \tilde{\boldsymbol{\tau}}^{\dagger}\boldsymbol{\tau_d}^T\dot{\boldsymbol{x}} + \dot{\boldsymbol{x}} \tag{7}$$

given $K \in R^{6\times6}$. Eq. 7 can be developed by assuming $\ddot{x} \approx 0$:

$$\dot{\boldsymbol{\tau}}_d' - \dot{\boldsymbol{\tau}}_m = -K\tilde{\boldsymbol{\tau}} - \tilde{\boldsymbol{\tau}}^{\dagger}\boldsymbol{\tau_d}^T\dot{\boldsymbol{x}} + \dot{\boldsymbol{x}}$$
$$-K_p\dot{\boldsymbol{x}} - J_{\boldsymbol{x}}\dot{\boldsymbol{x}} - J_{\boldsymbol{q}}\dot{\boldsymbol{q}} = -K\tilde{\boldsymbol{\tau}} - \tilde{\boldsymbol{\tau}}^{\dagger}\boldsymbol{\tau_d}^T\dot{\boldsymbol{x}} + \dot{\boldsymbol{x}}$$

that leads to :

$$\dot{\boldsymbol{q}} = J_{\boldsymbol{q}}^{\dagger}(\boldsymbol{u} - (K_p + J_{\boldsymbol{x}})\dot{\boldsymbol{x}}) \tag{8}$$

where $(\cdot)^{\dagger}$ represents the Moore-Penrose pseudoinverse operator. To demonstrate the system stability subjected to the control input (6), the desired force $\boldsymbol{f_d}$ must be consistent with the desired endoscope pose $\boldsymbol{x_d}$. This is a reasonable assumption in the vast majority of navigation-related scenarios, as the tubular environment naturally leads the constrained degree of freedom to be in the orthogonal direction with respect to the direction of motion (as shown in Fig. 1). This holds true independently from the relative orientation of the endoscope and the colon. However, this condition doesn't generally hold true, for example along a slope, or if the colon presents a fold. Therefore, the desired forces $\tilde{f}_d$ are projected in the null-space (or orthogonal space [28]) of the desired position, leading to the following additional equation:

$$\tilde{\boldsymbol{f_d}} = (I - \partial x_d \partial x_d^T)\boldsymbol{f_d}$$

where ∂x_d is the direction of the desired position, for example as shown in Fig. 1, where the motion is on the horizontal plane (orange vector) and the desired force along the vertical axis (green vector).

It must be pointed out that the endoscope insertion and retraction are not robotically controlled and therefore out of the scope of this paper. The user can manually insert and retract the endoscope by exerting forces on the tether. The controller is robust to the bounded additive disturbance $\boldsymbol{v} < \infty$ when $\lambda_{min}(K_p)||\dot{x}(t)|| > ||v(t)||$ and $K_p, K > 0$. If this condition holds true, the system dynamics is uniformly ultimately bounded (UUB) [40] subjected to the control action $\boldsymbol{u}$. The details of the demonstration are provided in appendix B.

IV. Parameters update with reinforcement learning

The controller parameters $K, K_p > 0 \in R^{6\times 6}$ are continuously updated through a model-free reinforcement learning algorithm. To this end, a state–action–reward–state–action (SARSA) algorithm in LQR fashion is used [34], [35]. SARSA provides a conservative exploration policy if compared to other algorithms such as Q-learning [29] and with the linear approximation of the value function, which, although suboptimal for a nonlinear system, enables fast computation rates. Moreover, SARSA has proven a rapid adaptation to environmental changes [30], which particularly fit the colon navigation. The value function minimized by the algorithm is:

$$K, K_p \longleftarrow \underset{K,K_p}{\arg\min}\, V(\tilde{\boldsymbol{x}}_{\boldsymbol{k}}, \boldsymbol{u}_{\boldsymbol{k}}, \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k}}, \theta_k) \tag{9}$$

where $\tilde{\boldsymbol{\tau}}'_{\boldsymbol{k}} = \boldsymbol{\tau_d} - \boldsymbol{\tau_k}$, $\tilde{\boldsymbol{x}}_{\boldsymbol{k}}$ and $\boldsymbol{u}_{\boldsymbol{k}}$ are the discrete terms corresponding to the the terms introduced in the previous section, discretized at each algorithm iteration k. In the linear approximate SARSA, the optimal discrete value function V is defined as a linear regression of the quadratic combination of the terms $\tilde{\boldsymbol{\tau}}'_{\boldsymbol{k}}$, $\tilde{\boldsymbol{x}}$ and $\boldsymbol{u}_{\boldsymbol{k}}$.

$$V^*(\tilde{\boldsymbol{x}}_{\boldsymbol{k}}, \boldsymbol{u}_{\boldsymbol{k}}, \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k}}, \boldsymbol{\theta}) = \boldsymbol{\theta}^T \boldsymbol{\Phi}_{\boldsymbol{k}}$$

and the reward/cost function at each iteration k

$$r_{k+1} = \tilde{\boldsymbol{x}}_{\boldsymbol{k}}^T Q \tilde{\boldsymbol{x}}_{\boldsymbol{k}} + \boldsymbol{u}_{\boldsymbol{k}}^T R \boldsymbol{u}_{\boldsymbol{k}} + \tilde{\boldsymbol{\tau}}_{\boldsymbol{k}}'^T S \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k}}$$

where $Q, S \geq 0 \in R^{6\times 6}$ and $R > 0 \in R^{6\times 6}$ according to LQR quadratic cost. $\boldsymbol{\Phi}_{\boldsymbol{k}}$ contains quadratic terms of the pose and force error vectors and the control input at k-th instant. V is linear with respect to $\boldsymbol{\theta}$, which is unknown and can be found through temporal difference error δ, the difference between the ultimate current $\hat{V}_{k+1}$ and the previous one $\hat{V}_k$, in the linear approximate SARSA algorithm (Alg. 1).

$\boldsymbol{\theta}$ is updated at each iteration according to the value of $\hat{V}$ and r_{k+1}, applying an incremental correction. Based on the value of $\hat{V}$ computed upon the measured state $\boldsymbol{x_{k+1}}$, the exploration policy π updates the matrices K, K_p. The policy π is a Gaussian noise $\mathcal{N}$ with a mean value equal to the current gain component and standard deviation decreasing with the function $\hat{V}$

$$K_i \longleftarrow \mathcal{N}(K_i, ke^{-\frac{1}{|\hat{V}_{K_i}|}}) \quad k > 0 \tag{10}$$

where K_i are the elements of the matrices K, K_p. The semi-positive-definite condition of the matrices is ensured by only considering diagonal matrices K, K_p for which is sufficient to impose the positiveness of the diagonal elements. Two parameters can tune the algorithm: $0 << \gamma < 1$, known as a discount factor and α_k, the learning rate.

The SARSA algorithm asymptotically converges to V^* if a standard stochastic approximation condition is satisfied: $\sum_{k=0}^{\infty} \alpha_k^2$ is finite and $\sum_{k=0}^{\infty} \alpha_k$ is infinite. The learning factor α_k is chosen equal to $1/k$ [36], this choice is valid because the gains K and K_p reach the stationary point where V gets to its minimum, i.e. in the proximity of the target. The algorithm converges if the policy π is Lipschitz continuous [31], [32] for a bounded Lipschitz constant. In appendix C we demonstrate that the policy chosen is Lipschitz continuous with a Lipschitz constant $C > 6k$; therefore, k must be large enough to ensure a good grade of exploration but simultaneously small enough to guarantee the convergence of the algorithm [33]. This condition also ensures that the value function V^* is a good approximation of the optimal one. Additionally, given the slow variation of the endoscope-colon interaction dynamics, the exploration policy π computes K and K_p every n iterations. The figure 3 shows the overall control scheme.

Algorithm 1 Approximate SARSA

measure initial state $\boldsymbol{x_0}$
initialize parameters θ_0
initialize $K, K_p \longrightarrow \boldsymbol{u_0}$
initialize $n = N$
for time step $k = 0, 1, 2, \ldots$ **do**
  measure next state $\tilde{\boldsymbol{x}}_{\boldsymbol{k+1}}, \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k+1}}$ and reward r_{k+1}
  if $k = n$ **then**
    update K, K_p with exploration policy (π) based on $\hat{V}\left(\tilde{\boldsymbol{x}}_{\boldsymbol{k+1}}, \boldsymbol{u_k}, , \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k+1}}, \boldsymbol{\theta_k}\right)$
    $n = n + N$
  end if
  use K, K_p to compute $\boldsymbol{u_{k+1}}$ according to eqs. 6-8
  $\hat{\delta}_k = r_{k+1} + \gamma \hat{V}\left(\tilde{\boldsymbol{x}}_{\boldsymbol{k+1}}, \boldsymbol{u_{k+1}}, \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k+1}}, \boldsymbol{\theta_k}\right) - \hat{V}\left(\tilde{\boldsymbol{x}}_{\boldsymbol{k}}, \boldsymbol{u_k}, \tilde{\boldsymbol{\tau}}'_{\boldsymbol{k}}, \boldsymbol{\theta_k}\right)$
  $\boldsymbol{\theta_{k+1}} = \boldsymbol{\theta_k} + \alpha_k \hat{\delta}_k \nabla_{\boldsymbol{\theta_k}} \hat{V}$
  apply $\boldsymbol{u_{k+1}}$
end for

V. Experimental Setup

In the MFE, the EPM is actuated by means of the KUKA LBR Med R820 robot (KUKA Roboter GMBH), the IPM is an axially magnetized N52 grade permanent magnet with 12 mm diameter, 24 mm length. The EPM is also an axially magnetized N52 grade magnet with 101.6 mm diameter and length. The experiments are carried out on a silicone colon

simulator (Kyoto Kagaku M40) widely used for benchtop experiments and clinical training. In the first experimental layout (Fig. 4), the colon phantom is arranged in an L shape (two straight lines divided by a tight angle of approximately 90°). Alternating straight tracts and tight turns is a ubiquitous scenario in clinical colonoscopies. The second layout (5) consists of a straight tract containing an obstacle, aimed at simulating a lumen obstruction. This represents a common situation in colonoscopy, as other organs or colon folds may create barriers that obstruct the endoscope motion. In both scenarios, we used lubricant to simulate the colon's mucosa low friction properties.

One of the most promising applications of force control in robotic endoscopy is micro-ultrasound scanning [16]. The force control must ensure safe interaction with the colon wall, minimizing the risk of tissues damage, simultaneously maintaining a quality contact to ensure ultrasound coupling. For these reasons, the force lower and upper limits are chosen equal to 0.3N and 0.5N. This particular range is chosen considering suggestions from the literature [16], [27]: below 0.3N the force is not sufficient to maintain contact, and 0.5N is considered a a safety limit under which the risk of tissue damage significantly increases. As mentioned in section II, we estimated the numerical value of the force by means of equations (11) (12), using the two magnet poses. The tip is the only magnetically controllable part of the device, whereas the tether is passive and highly flexible and therefore incapable of transmitting significant forces. Moreover, we assumed that the overall interaction can be sufficiently represented in a single contact point, given the tip size. In order to demonstrate the effectiveness of the force control algorithm in providing a stable and reliable contact during the experiments, a contact sensing device should be used. However, the form factor of commercially available contact sensors is not compatible with the dimensions and the presence of lubricants, which would disturb the sensors' readings. For this reason, a painted thin layer of conductive ink has been used to provide binary information (contact/no contact). Fig. 4 shows the area covered by the conductive ink. The control loop and the learning algorithm run at 100 Hz, a standard frequency for control applications of this kind.

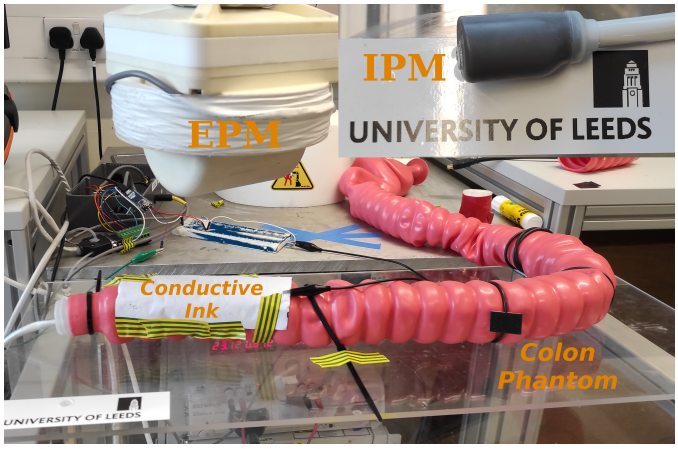


Fig. 4: Experimental setup

VI. Experimental Validation

Three experiments are performed (two with the first setup and one with the second setup), with the aim of evaluating:

- The performances of the position control.
- The ability to maintain the force in a defined range while controlling the endoscope pose.
- The ability to adapt the control and autonomously overcome obstacles and turns.

As mentioned in the previous sections, path planning is out of the scope of this work. Hence, the endoscope desired position $\boldsymbol{x_d}$ is constructed manually by combining a target location on the endoscope longitudinal direction and a desired heading. In the following, these two components of $\boldsymbol{x_d}$ will be evaluated. The first experiment is focused on controlling the endoscope forward and backward motion along a straight tract, to evaluate the accuracy of the position/force control. The experiment starts in the absence of contact between, followed by a coupling phase, before starting the reinforcement control routine. In this phase, the endoscope is required to move forward (i.e. in the direction of the camera view) of 5 cm, then return to the initial position and finally perform a rotation of 1.45 rad. The force and the pose tracking of the endoscope tip heading are shown in Figs.6a and 6b. Although the motion is not smooth due to the colon folds, the MFE follows the position and orientation setpoints.

A steady-state error of approximately 5 mm and 0.05 rads can be observed at the end of the transient phase. That is acceptable because, in the proximity of the setpoint, the adaptation given by the reinforcement learning is weak; therefore, the controller gains reach their stationary point resulting in a classic PD action that cannot erase the steady-state error. Moreover, the setpoint can coincide with a fold, so it would be unreachable. The module of magnetic force remains inside the required range throughout the motion, with an average value of 0.4 N, as shown in Fig. 6b. The aim of fig.6b is to show the ability to control the contact force in presence of significant position changes, that can be seen at t=15 and t=51. For this reason, while Fig. 6a shows the position and orientation values for t>60, 6b stops at t=65, after which the force control is not challenged by the changes of the position setpoint. Furthermore, the contact sensed by the conductive paint, shown in green, demonstrates a contact rich behaviour with a 99% of contact after the reinforcement procedure starts.

The second experiment is aimed at demonstrating the ability to overcome a tight turn. The endoscope is teleoperated in the proximity of the turn shown in Fig. 4 and the autonomous procedure is started. If the MFE manages to navigate through the turn, the experiment is considered successful. Figs. 7a and 7b show the endoscope position, heading and interaction force during the autonomous manoeuvre as well as the desired target pose manually provided during the experiments. The MFE reaches the target poses with an acceptable error. Moreover,

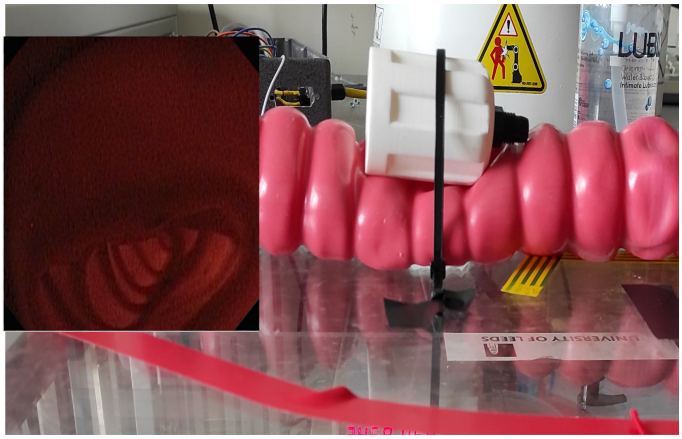
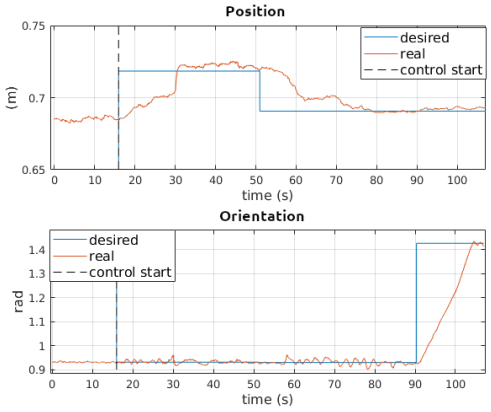
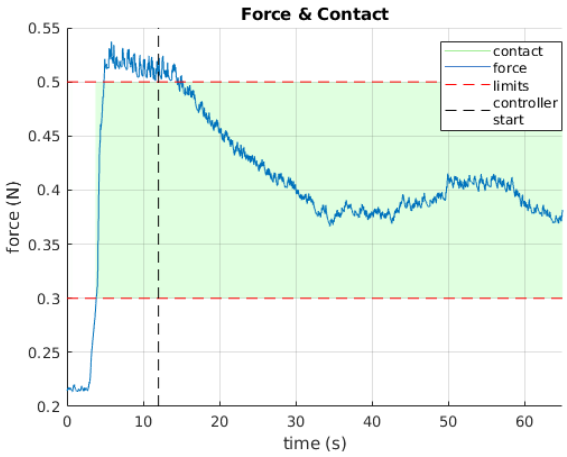


Fig. 5: Experimental setup with obstacle seen from the outside and from the endoscope camera

due to the inaccuracy of the manual target definition, the endoscope is required to reach an unreachable target between $t \approx 230s$ and $t \approx 255s$, highlighted in red in Fig. 7a. During this phase, the endoscope is stuck and does not reach the target. The manipulator, in response, moves the EPM to attract the IPM along the unfeasible direction but, since the colon hinders the endoscope motion, the EPM gets too far from the IPM, and the interaction force drops below the lower limit, as shown in Fig. 7b. As long as the desired endoscope pose becomes reachable ($t > 255$), the controller restores the position control. Although manually defining the desired endoscope pose may lead to unfeasible targets, the controller can adapt and accomplish both force and position requirements as long as the desired pose is feasible. The last experiment proposed in this work is the autonomous navigation across the obstruction created by an obstacle. This experiment aims to show the ability of the controller to overcome the obstruction without any prior knowledge. An external object has been placed on the external surface of the phantom, as shown in Fig. 5. The trajectory executed by the MFE and the contact forces are is shown in Figs. 8a and 8b. The control system must increase the distance between EPM and IPM to successfully perform the motion, thus decreasing the interaction force that falls below the lower limit. Once the endoscope passed the obstacle, the force return inside the desired range. The gains K_p and K change rapidly throughout the motion according to the value of V as shown in Fig. 9. The gains reach a stable value when the function V decreases to its lowest value in proximity of the target, i.e. when the error is low. For the sake of comparison, we repeat the same experiment with 10 static values of the control parameters ranging between 0.01 and 150. The gains have been selected to regulate the contribution of the position and the force error on the control action, balancing between these two quantities. Results are shown in Fig. 10. For every static value of the parameters, the system was not able to complete any of the trajectories. A comparison between adaptive and fixed parameters is shown in the video provided in the supplementary material.



(a)

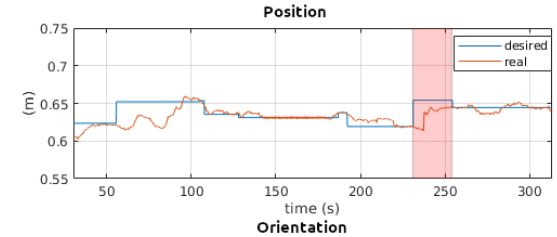


(b)

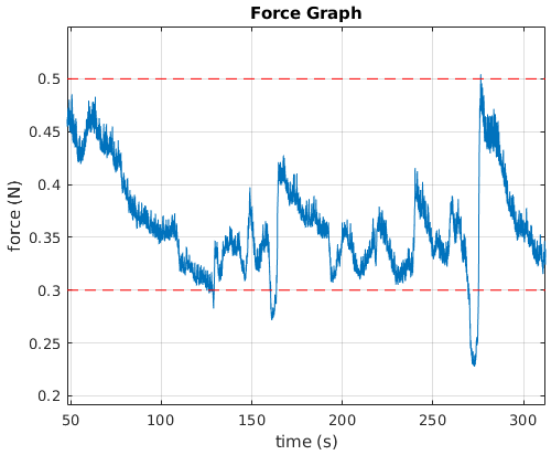
Fig. 6: Experiment in a straight colon portion. (a) Pose: desired and real position along the endoscope heading are shown in the first graph; desired and real angle of the endoscope heading are shown in the second graph. (b) force and contact graph with force limits in red : the colon-endoscope contact is highlighted in green while the black line shows when the reinforcement control starts.

VII. CONCLUSIONS

An adaptive parallel position and force control for magnetic endoscopes based on reinforcement learning is proposed in this study. This work aims to provide a dynamic controller that adapts to the unstructured environment of the colon and extends the magnetic controller capabilities, simultaneously controlling the endoscope position and the interaction forces between the endoscope and colon wall. This result will enable the adoption of contact-based diagnostic techniques such as micro-ultrasound during navigation while concurrently providing a control algorithm capable of adapting to varying conditions and overcoming unmodelled obstacles. The algorithm takes advantage of the model-free nature of reinforcement learning to adjust the control parameters. Conditions for system stability are discussed and demonstrated while the experimental results provide evidence of algorithm performance.



(a)



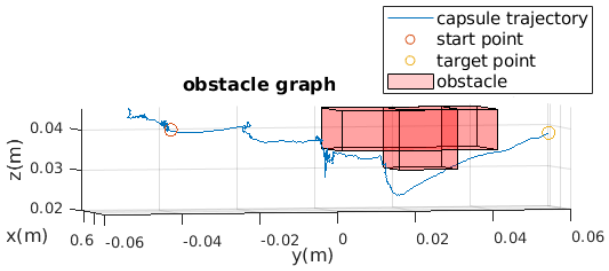
(b)

Fig. 7: Experiment in L shape colon portion. (a) Pose: the first graph shows the desired and the real position along the endoscope heading, with a red band we represented a target position impossible to reach because out of the colon; the second graph shows desired and real angles of the endoscope heading. (b) force graph with the force limits in red.

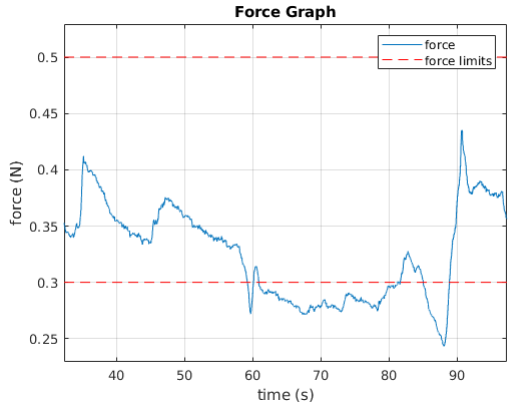
The natural evolution of the control approach presented in this work is an advanced motion planning algorithm capable of defining trajectories and boundary conditions for the endoscopes motion, considering robot joints limits, workspace bounds, camera obstruction and manipulability, with minimal user inputs. Overall, this work increases the autonomy of magnetically driven endoscopes and enables control systems to autonomously solve local navigation problems through modification of control behavior.

APPENDIX A

The magnetic interaction between the EPM and the IPM is modelled through the magnetic dipole-dipole. Considering the distance between the EPM position $\boldsymbol{p_E}$ and the IPM position $\boldsymbol{p_I}$ as $\boldsymbol{p} = \boldsymbol{p_E} - \boldsymbol{p_I}$, and naming the magnetic dipole moment of the EPM $\boldsymbol{m_E} = ||\boldsymbol{m_E}||\hat{\boldsymbol{m}}_{\boldsymbol{E}}$ and of the



(a)



(b)

Fig. 8: trajectory in the presence of obstacle.(a) pose obstacle graph: the capsule trajectory starts from a point and reach the target behind the obstacle. (b) force graph in the presence of the obstacle with force limits in red.

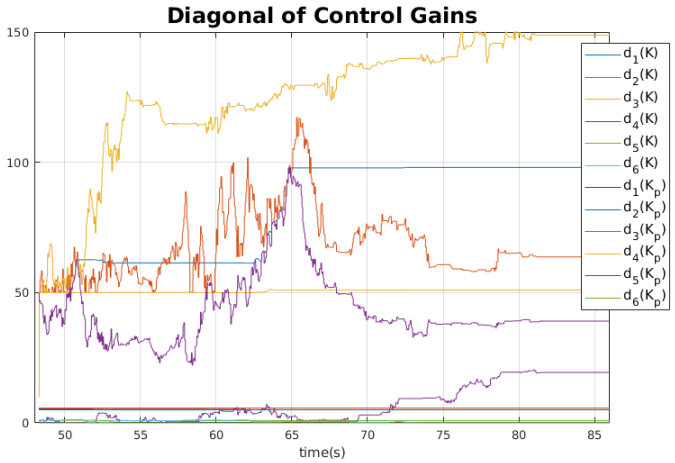


Fig. 9: diagonals of control gains for obstacle overcoming

IPM $\boldsymbol{m_I} = ||\boldsymbol{m_I}||\hat{\boldsymbol{m}}_{\boldsymbol{I}}$, the force and torques between the two magnets are described by the following equations [38]:

$$\begin{aligned}\boldsymbol{f}\left(\boldsymbol{p}, \boldsymbol{m}_E, \boldsymbol{m}_I\right) = & \frac{3\mu_0 \left\|\boldsymbol{m}_E\right\| \left\|\boldsymbol{m}_I\right\|}{4\pi\|\boldsymbol{p}\|^4} \\ & \left(\hat{\boldsymbol{m}}_E\hat{\boldsymbol{m}}_I^{\top} + \hat{\boldsymbol{m}}_I\hat{\boldsymbol{m}}_E^{\top} + \right. \\ & \left.\left(\hat{\boldsymbol{m}}_I^{\top} Z \hat{\boldsymbol{m}}_I\right) \boldsymbol{I}\right) \hat{\boldsymbol{p}}\end{aligned} \tag{11}$$

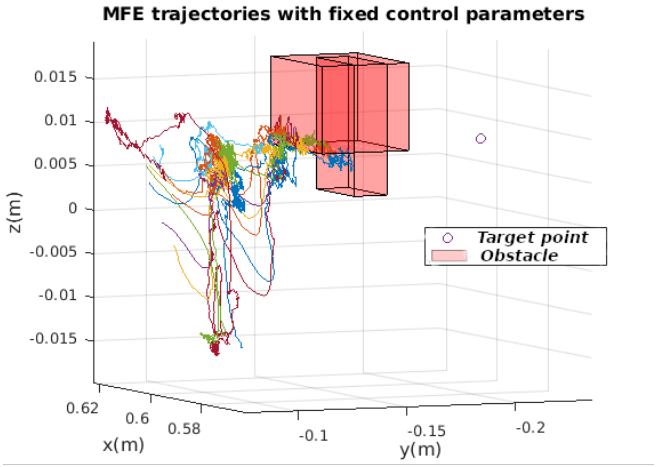


Fig. 10: MFE trajectories with different control parameters. The MFE never reach the target position.

$$\boldsymbol{\tau}(\boldsymbol{p}, \boldsymbol{m}_E, \boldsymbol{m}_I) = \frac{\mu_0 \|\boldsymbol{m}_E\| \|\boldsymbol{m}_I\|}{4\pi \|\boldsymbol{p}\|^3} \hat{\boldsymbol{m}}_I \times D(\hat{\boldsymbol{p}}) \hat{\boldsymbol{m}}_E \tag{12}$$

with $Z = I - 5\hat{\boldsymbol{p}}\hat{\boldsymbol{p}}^T$, $D = 3\hat{\boldsymbol{p}}\hat{\boldsymbol{p}}^T - I$, $\hat{\boldsymbol{p}} = \frac{\boldsymbol{p}}{\|\boldsymbol{p}\|}$ and where $I \in R^{3\times 3}$ is the identity matrix. Similarly to [39] the forces exchanged between the magnets can be described in vector form as: $\boldsymbol{\tau_m} = \begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{\tau} \end{bmatrix}$

and the time derivative of $\boldsymbol{\tau_m}$ is:

$$\begin{aligned}\dot{\boldsymbol{\tau}}_{\boldsymbol{m}} &= \begin{bmatrix} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}} & \frac{\partial \boldsymbol{f}}{\partial \hat{\boldsymbol{m}}_E} & \frac{\partial \boldsymbol{f}}{\partial \hat{\boldsymbol{m}}_I} \\ \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{p}} & \frac{\partial \boldsymbol{\tau}}{\partial \hat{\boldsymbol{m}}_E} & \frac{\partial \boldsymbol{\tau}}{\partial \hat{\boldsymbol{m}}_I} \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{p}} \\ \dot{\hat{\boldsymbol{m}}}_{\boldsymbol{E}} \\ \dot{\hat{\boldsymbol{m}}}_{\boldsymbol{I}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}} & \frac{\partial \boldsymbol{f}}{\partial \hat{\boldsymbol{m}}_E} & \frac{\partial \boldsymbol{f}}{\partial \hat{\boldsymbol{m}}_I} \\ \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{p}} & \frac{\partial \boldsymbol{\tau}}{\partial \hat{\boldsymbol{m}}_E} & \frac{\partial \boldsymbol{\tau}}{\partial \hat{\boldsymbol{m}}_I} \end{bmatrix} \left(\begin{bmatrix} \dot{\boldsymbol{p}}_{\boldsymbol{E}} \\ \dot{\hat{\boldsymbol{m}}}_{\boldsymbol{E}} \\ \boldsymbol{0} \end{bmatrix} - \begin{bmatrix} \dot{\boldsymbol{p}}_{\boldsymbol{I}} \\ \boldsymbol{0} \\ \dot{\hat{\boldsymbol{m}}}_{\boldsymbol{I}} \end{bmatrix} \right)\end{aligned}$$

the EPM and IPM are rigidly connected to the robot end-effector and to the endoscope centre of mass. Hence, considering the end-effector pose $\boldsymbol{\chi}$ and the endoscope pose $\boldsymbol{x}$, the magnets' poses are described by the following relations:

$$\begin{bmatrix} \dot{\boldsymbol{p}}_{\boldsymbol{I}} \\ \dot{\hat{\boldsymbol{m}}}_{\boldsymbol{I}} \end{bmatrix} = \begin{bmatrix} I & 0_{3,3} \\ 0_{3,3} & [\hat{\boldsymbol{m}}_{\boldsymbol{E}}\wedge]^T \end{bmatrix} \dot{\boldsymbol{x}} = M_I \dot{\boldsymbol{x}}$$
$$\begin{bmatrix} \dot{\boldsymbol{p}}_{\boldsymbol{E}} \\ \dot{\hat{\boldsymbol{m}}}_{\boldsymbol{E}} \end{bmatrix} = \begin{bmatrix} I & 0_{3,3} \\ 0_{3,3} & [\hat{\boldsymbol{m}}_{\boldsymbol{E}}\wedge]^T \end{bmatrix} \dot{\boldsymbol{\chi}} = M_E \dot{\boldsymbol{\chi}}$$

where $[\cdot\wedge] \in so(3)$ is the skew operator and $0_{i,k} \in R^{i\times k}$ is the zero matrix. The expression above is then used to compute the manipulator joint velocity ($\dot{\boldsymbol{q}}$) considering $\dot{\boldsymbol{\chi}} = J\dot{\boldsymbol{q}}$, being J the manipulator jacobian.

$$\begin{aligned}\dot{\boldsymbol{\tau}}_{\boldsymbol{m}} &= \begin{bmatrix} \frac{\partial \boldsymbol{\tau_m}}{\partial \boldsymbol{p}} & \frac{\partial \boldsymbol{\tau_m}}{\partial \hat{\boldsymbol{m}}_E} \end{bmatrix} M_I \dot{x} - \begin{bmatrix} \frac{\partial \boldsymbol{\tau_m}}{\partial \boldsymbol{p}} & \frac{\partial \boldsymbol{\tau_m}}{\partial \hat{\boldsymbol{m}}_I} \end{bmatrix} M_E J \dot{q} \\ &= J_{\boldsymbol{x}} \dot{\boldsymbol{x}} - \hat{J}_{\boldsymbol{q}} \dot{\boldsymbol{q}}\end{aligned}$$

and naming $J_{\boldsymbol{q}} = -\hat{J}_{\boldsymbol{q}}$ we eventually get to (2):

$$\dot{\boldsymbol{\tau}}_{\boldsymbol{m}} = J_{\boldsymbol{x}} \dot{\boldsymbol{x}} + J_{\boldsymbol{q}} \dot{\boldsymbol{q}}$$

APPENDIX B

Considering the control law

$$\boldsymbol{u} = K(\boldsymbol{\tau_d} + K_p \tilde{\boldsymbol{x}} - K_p \dot{\boldsymbol{x}} + G(\boldsymbol{x}) - \boldsymbol{\tau_m}) + \tilde{\boldsymbol{\tau}}^{\dagger} \boldsymbol{\tau}_{\boldsymbol{d}}^T \dot{\boldsymbol{x}} - \dot{\boldsymbol{x}}$$

in the following, the closed loop system stability is demonstrated

Proof. Consider a positive definite Lyapunov function defined as

$$J(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) = \frac{1}{2} \dot{\boldsymbol{x}}^T B(\boldsymbol{x}) \dot{\boldsymbol{x}} + \frac{1}{2} \tilde{\boldsymbol{x}}^T K_p \tilde{\boldsymbol{x}} + \frac{1}{2} \tilde{\boldsymbol{\tau}}^T \tilde{\boldsymbol{\tau}} \tag{13}$$

Its derivative is

$$\begin{aligned}\dot{J}(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) =& \dot{\boldsymbol{x}}^T B(\boldsymbol{x}) \ddot{\boldsymbol{x}} + \frac{1}{2} \dot{\boldsymbol{x}}^T \dot{B}(\boldsymbol{x}) \dot{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^T K_p \dot{\boldsymbol{x}} + \tilde{\boldsymbol{\tau}}^T \dot{\tilde{\boldsymbol{\tau}}} \\ \dot{J}(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) =& \dot{\boldsymbol{x}}^T (\boldsymbol{\tau}_m + \boldsymbol{v} - C(\boldsymbol{x}, \dot{\boldsymbol{x}}) \dot{\boldsymbol{x}} - G(\boldsymbol{x})) \\ &+ \frac{1}{2} \dot{\boldsymbol{x}}^T \dot{B}(\boldsymbol{x}) \dot{\boldsymbol{x}} - \tilde{\boldsymbol{x}}^T K_p \dot{\boldsymbol{x}} + \tilde{\boldsymbol{\tau}}^T \dot{\tilde{\boldsymbol{\tau}}} \\ =& \dot{\boldsymbol{x}}^T (\boldsymbol{\tau}_d - \tilde{\boldsymbol{\tau}} + \boldsymbol{v}) + \tilde{\boldsymbol{\tau}}^T \dot{\tilde{\boldsymbol{\tau}}} - \dot{\boldsymbol{x}}^T K_p \dot{\boldsymbol{x}}\end{aligned}$$

where the term $\frac{1}{2} \dot{\boldsymbol{x}}^T (\dot{B}(\boldsymbol{x}) - 2C(\boldsymbol{x}, \dot{\boldsymbol{x}})) \dot{\boldsymbol{x}}$ is equal to zero according to the work-energy theorem. We know that $\dot{\tilde{\boldsymbol{\tau}}} = \boldsymbol{u}$ which erase the term $\dot{\boldsymbol{x}}^T (\boldsymbol{\tau}_d - \tilde{\boldsymbol{\tau}})$ of the previous equation and eventually we get:

$$\begin{aligned}\dot{J}(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) &= \dot{\boldsymbol{x}}^T (\boldsymbol{\tau_d} - \tilde{\boldsymbol{\tau}} + \boldsymbol{v}) + \tilde{\boldsymbol{\tau}}^T (-K\tilde{\boldsymbol{\tau}} - \tilde{\boldsymbol{\tau}}^{\dagger} \boldsymbol{\tau}_{\boldsymbol{d}}^{\boldsymbol{T}} \dot{\boldsymbol{x}} + \dot{\boldsymbol{x}}) \\ &\quad - \dot{\boldsymbol{x}}^T K_p \dot{\boldsymbol{x}} \\ &= \dot{\boldsymbol{x}}^T \boldsymbol{v} - \tilde{\boldsymbol{\tau}}^T K \tilde{\boldsymbol{\tau}} - \dot{\boldsymbol{x}}^T K_p \dot{\boldsymbol{x}}\end{aligned}$$

focusing on the most critical term with the disturbance $\boldsymbol{v}$. By assuming that the disturbance is bounded $||\boldsymbol{v}|| < \infty$, we evaluate the contribution of the disturbance as follow, given λ_{max} the highest eigenvalue of a matrix:

$$\begin{aligned}\dot{J}(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) \leq& -\lambda_{min}(K_p) ||\dot{\boldsymbol{x}}||^2 - \lambda_{min}(K) ||\tilde{\boldsymbol{\tau}}||^2 \\ &+ ||\dot{\boldsymbol{x}}|| ||\boldsymbol{v}||\end{aligned} \tag{14}$$

being $\tilde{\boldsymbol{\tau}}$ dynamics faster than the endoscope dynamics, the equilibrium point $||\tilde{\boldsymbol{\tau}}|| = 0$ is reached faster than $||\dot{\boldsymbol{x}}|| = 0$. Hence the Lyapunov function contracts under the limited condition $\lambda_{min}(K_p) ||\dot{\boldsymbol{x}}|| \geq ||\boldsymbol{v}||$. The condition guarantees uniformly unlimited bounded (UUB) [40] stability of the system. Moreover in this condition the system converges to the attractive set $B = \{(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) | \dot{J}(\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}}) = 0\}$ for $\tilde{\boldsymbol{x}}, \dot{\boldsymbol{x}}, \tilde{\boldsymbol{\tau}} = 0$

□

APPENDIX C

The definition of Lipschitz continuity function is

$$||f(x_1) - f(x_2)|| \leq C||x_1 - x_2||_2 \tag{15}$$

if there exists a Lipschitz constant $C > 0$. In the following we demonstrate the Lipschitz continuity in θ of the exploration policy π described in (10).

Proof. To demonstrate the assertion, we express the policy on the largest set of exploration, taking into account that the almost totality of gaussian noise can be described in terms of its standard deviation:

$$K = N(K, ke^{-\frac{1}{|V_{\boldsymbol{\theta}}|}}) \approx K \pm 3ke^{-\frac{1}{|V_{\boldsymbol{\theta}}|}} \tag{16}$$

applying (16) into (15) we get

$$\begin{aligned} ||K \pm 3ke^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}} - K \mp 3ke^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}|| &\leq C||\boldsymbol{\theta_1} - \boldsymbol{\theta_2}||_2 \\ || \pm 3ke^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}} \mp 3ke^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}|| &\leq \\ ||3ke^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}} + 3ke^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}|| &\leq C||\boldsymbol{\theta_1} - \boldsymbol{\theta_2}||_2 \end{aligned}$$

where we take opposite sign of the functions in the left-hand side of the last inequality in order to express the maximum policy variation.

$$\begin{aligned} ||3ke^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}} + 3ke^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}|| &\leq C||\boldsymbol{\theta_1} - \boldsymbol{\theta_2}||_2 \\ 3k||e^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}}|| + 3k||e^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}|| &\leq \\ 3k(||e^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}}|| + ||e^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}||) &\leq C||\boldsymbol{\theta_1} - \boldsymbol{\theta_2}||_2 \\ k &\leq \frac{C}{3} \frac{||\boldsymbol{\theta_1} - \boldsymbol{\theta_2}||_2}{e^{-\frac{1}{|V_{\boldsymbol{\theta_1}}|}} + e^{-\frac{1}{|V_{\boldsymbol{\theta_2}}|}}} \end{aligned}$$

we can notice that the denominator in the left-hand side of the last inequality is always less than 2 and conclude that C is finite and must be over $6k$.

□

REFERENCES

[1] Toggweiler, Stefan, et al. "Management of vascular access in transcatheter aortic valve replacement: part 2: vascular complications." JACC: Cardiovascular Interventions 6.8 (2013): 767-776.

[2] Liu, Xiaolong, Gregory J. Mancini, and Jindong Tan. "Design and analysis of a magnetic actuated capsule camera robot for single incision laparoscopic surgery." 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015.

[3] Jeon, Sungwoong, et al. "A magnetically controlled soft microrobot steering a guidewire in a three-dimensional phantom vascular network." Soft robotics 6.1 (2019): 54-68.

[4] Abbott, Jake J., Eric Diller, and Andrew J. Petruska. "Magnetic methods in robotics." Annual Review of Control, Robotics, and Autonomous Systems 3 (2020): 57-90.

[5] Mahoney, Arthur W., and Jake J. Abbott. "Generating rotating magnetic fields with a single permanent magnet for propulsion of untethered magnetic devices in a lumen." IEEE Transactions on Robotics 30.2 (2013): 411-420.

[6] Son, Donghoon, Hunter Gilbert, and Metin Sitti. "Magnetically actuated soft capsule endoscope for fine-needle biopsy." Soft robotics 7.1 (2020): 10-21.

[7] Yim, Sehyuk, Kartik Goyal, and Metin Sitti. "Magnetically actuated soft capsule with the multimodal drug release function." IEEE/ASME Transactions on Mechatronics 18.4 (2013): 1413-1418.

[8] Popek, Katie M., Tucker Hermans, and Jake J. Abbott. "First demonstration of simultaneous localization and propulsion of a magnetic capsule in a lumen using a single rotating magnet." 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017.

[9] Yim, Sehyuk, and Metin Sitti. "Design and rolling locomotion of a magnetically actuated soft capsule endoscope." IEEE Transactions on Robotics 28.1 (2011): 183-194.

[10] Ciuti, G., et al. "Robotic versus manual control in magnetic steering of an endoscopic capsule." Endoscopy 42.2 (2010): 148.

[11] Edelmann, Janis, Andrew J. Petruska, and Bradley J. Nelson. "Estimation-based control of a magnetic endoscope without device localization." Journal of Medical Robotics Research 3.01 (2018): 1850002.

[12] Martin, James W., et al. "Enabling the future of colonoscopy with intelligent and autonomous magnetic manipulation." Nature machine intelligence 2.10 (2020): 595-606.

[13] P. Slawinski, A. Taddese, K. B. Musto, S. Sarker, P. Valdastri, K. L. Obstein, "Autonomously controlled magnetic flexible endoscope for colon exploration", Gastroenterology, 2018, Vol. 154, No. 6, pp. 1577-1579.

[14] Lucarini, Gioia, et al. "A new concept for magnetic capsule colonoscopy based on an electromagnetic system." International Journal of Advanced Robotic Systems 12.3 (2015): 25.

[15] Barducci, Lavinia, et al. "Fundamentals of the gut for capsule engineers." Progress in Biomedical Engineering 2.4 (2020): 042002.

[16] Norton, Joseph C., et al. "Intelligent magnetic manipulation for gastrointestinal ultrasound." Science robotics 4.31 (2019).

[17] Buşoniu, Lucian, Bart De Schutter, and Robert Babuška. "Approximate dynamic programming and reinforcement learning." Interactive collaborative information systems. Springer, Berlin, Heidelberg, 2010. 3-44.

[18] Bertsekas, Dimitri P. "Dynamic programming and suboptimal control: A survey from ADP to MPC." European Journal of Control 11.4-5 (2005): 310-334.

[19] Degris, Thomas, Patrick M. Pilarski, and Richard S. Sutton. "Model-free reinforcement learning with continuous action in practice." 2012 American Control Conference (ACC). IEEE, 2012.

[20] Yarats, Denis, et al. "Improving sample efficiency in model-free reinforcement learning from images." arXiv preprint arXiv:1910.01741 (2019).

[21] Turan, Mehmet, et al. "Learning to navigate endoscopic capsule robots." IEEE Robotics and Automation Letters 4.3 (2019): 3075-3082.

[22] Huang, Tao, et al. "The Control of Magnetic Levitation System Based on Improved Q-network." 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2019.

[23] Pane, Yudha P., et al. "Reinforcement learning based compensation methods for robot manipulators." Engineering Applications of Artificial Intelligence 78 (2019): 236-247.

[24] Lewis, Frank L., and Kyriakos G. Vamvoudakis. "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 41.1 (2010): 14-25.

[25] Pittiglio, Giovanni, et al. "Magnetic levitation for soft-tethered capsule colonoscopy actuated with a single permanent magnet: a dynamic control approach." IEEE robotics and automation letters 4.2 (2019): 1224-1231.

[26] Barducci, Lavinia, et al. "Adaptive dynamic control for magnetically actuated medical robots." IEEE robotics and automation letters 4.4 (2019): 3633-3640.

[27] Scaglioni, Bruno, et al. "Explicit model predictive control of a magnetic flexible endoscope." IEEE robotics and automation letters 4.2 (2019): 716-723

[28] Siciliano B. (1996) Parallel Force/Position Control of Robot Manipulators. In: Giralt G., Hirzinger G. (eds) Robotics Research. Springer, London.

[29] Corazza, Marco, and Andrea Sangalli. "Q-Learning and SARSA: a comparison between two intelligent stochastic control approaches for financial trading." University Ca'Foscari of Venice, Dept. of Economics Research Paper Series No 15 (2015).

[30] Perrusquía, Adolfo, Wen Yu, and Alberto Soria. "Position/force control of robot manipulators using reinforcement learning." Industrial Robot: the international journal of robotics research and application (2019).

[31] Theodore J. Perkins and Doina Precup. 2002. A convergent form of approximate policy iteration. In Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS'02). MIT Press, Cambridge, MA, USA, 1627–1634.

[32] Melo, Francisco S., Sean P. Meyn, and M. Isabel Ribeiro. "An analysis of reinforcement learning with function approximation." Proceedings of the 25th international conference on Machine learning. 2008.
[33] Zou, Shaofeng, Tengyu Xu, and Yingbin Liang. "Finite-sample analysis for sarsa with linear function approximation." arXiv preprint arXiv:1902.02234 (2019).
[34] Bradtke, Steven J. "Reinforcement learning applied to linear quadratic regulation." Advances in neural information processing systems. 1993.
[35] Lewis, Frank L., and Draguna Vrabie. "Reinforcement learning and adaptive dynamic programming for feedback control." IEEE circuits and systems magazine 9.3 (2009): 32-50.
[36] Jaakkola, Tommi, Michael I. Jordan, and Satinder P. Singh. "On the convergence of stochastic iterative dynamic programming algorithms." Neural computation 6.6 (1994): 1185-1201.
[37] A. Mamunes, F. Campisano, J. Martin, B. Scaglioni, E. Mazomenos, P. Valdastri and K. Obstein, "Magnetic flexible endoscope for colonoscopy: an initial learning curve analysis" Endoscopy International Open 09, 2021.
[38] Mahoney, Arthur W., and Jake J. Abbott. "Five-degree-of-freedom manipulation of an untethered magnetic device in fluid using a single permanent magnet with application in stomach capsule endoscopy." The International Journal of Robotics Research 35.1-3 (2016): 129-147.
[39] Taddese, Addisu Z., et al. "Nonholonomic closed-loop velocity control of a soft-tethered magnetic capsule endoscope." 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016.
[40] Krstic, Miroslav, Petar V. Kokotovic, and Ioannis Kanellakopoulos. Nonlinear and adaptive control design. John Wiley & Sons, Inc., 1995.