Unconstrained Rotation for Control of Concentric Tube Robots with Deep Reinforcement Learning

Keshav Iyengar[1], Sarah Spurgeon[2], and Danail Stoyanov[1]

[1]*Welcome/EPSRC Centre for Interventional and Surgical Science, Department of Computer Science, University College London*

[2]*Department of Electronic and Electrical Engineering, University College London*

INTRODUCTION

Concentric tube robots (CTRs) are a class of continuum robot that depend on the interactions between neighbouring, concentrically aligned tubes to produce the curvilinear shapes of the robot backbone [1]. The main application of these unique robots is that of minimally invasive surgery (MIS), where most of the developments for CTRs have been focused. Due to the confined workspaces and resulting extended learning times for surgeons in MIS, dexterous, compliant continuum robots such as CTRs have been under development in preference to the mechanically rigid and limited degrees-of-freedom (DOF) robots used in interventional medicine today. The precurved tubes in CTRs, which are sometimes referred to as active cannulas or catheters, are manufactured from super-elastic materials like Nickel-Titanium alloys with each tube nested concentrically. From the base, the individual tubes can be actuated through extension and rotation, which results in the bending and twisting of the backbone as well as access to the surgical site through the channel and robot tip. Clinically, CTRs are motivated for use in brain, cardiac, gastric surgery as well other procedures [2].

Due to tube interactions, modelling and control is challenging. Position control for CTRs has relied on model development, and although a balance between computation and accuracy has been reached in the literature [1], there remain issues such as performance in the presence of tube parameter discrepancies and the impact of unmodelled physical phenomena such as friction and permanent plastic deformation. This motivates the development of an end-to-end model-free control framework for CTRs. One such model-free framework for control that is gaining popularity is reinforcement learning (RL), a paradigm of machine learning that necessitates an agent to output action that interacts with an environment. The environment then processes this action, and returns a new state and, depending on the task, a reward signal. The task we give the agent then is to control the end-effector Cartesian robot tip position by means of actions that represent changes in joint values. In Fig. 1 the components of the state are shown in relation to a illustrated CTR and further described in the next section.

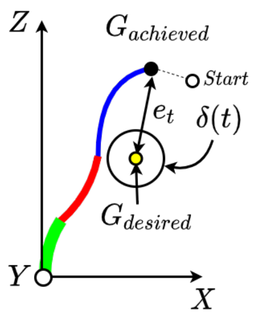


Fig. 1: State with start position and achieved goal, $G_{achieved}$ (black), desired goal, $G_{desired}$ (yellow), goal tolerance, $\delta(t)$. Outer tube (green), middle tube (red) and inner tube (blue).

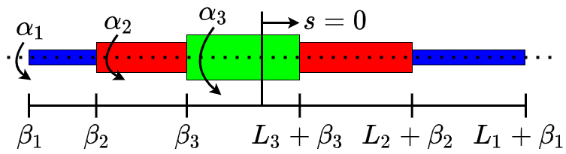


Fig. 2: Joint variables β and α of a 3 tube CTR. s is the arc-length or axis along the backbone.

In this work, we investigate how the rotational actuation affects final errors during evaluation of the learned policy. We find by avoiding constraining the rotational DOF of each tube, the agent can freely rotate to achieve goals as opposed to when constrained that result in more steps and larger error metrics.

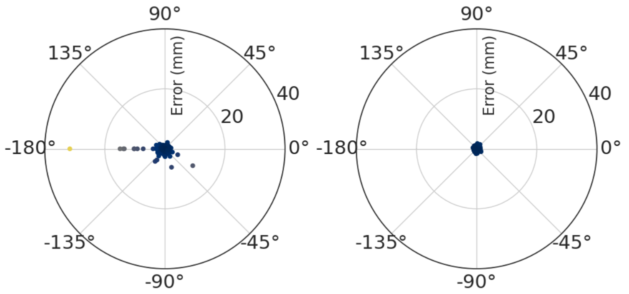
MATERIALS AND METHODS

First, the Markov Decision Process (MDP), a definition required for RL algorithms is defined as follows.

- State (s_t) : States are defined as the concatenation of the trigonometric joint representation, Cartesian goal error and current goal tolerance. As shown in Fig. 2, rotation and extension of tube i (ordered innermost to outermost) are α_i and β_i. The trigonometric representation [3] of tube i is defined as:

$$\gamma_i = \{\gamma_{1,i}, \gamma_{2,i}, \gamma_{3,i}\} = \{\cos(\alpha_i), \sin(\alpha_i), \beta_i\} \quad (1)$$

In constrained rotation, α_i for each tube is constrained to be between $-180°$ and $+180°$ during each episode



(a) (b)

Fig. 3: Achieved goal errors for constrained (a) and free rotation (b) agents. Errors are shown in a polar plot with α_1 rotation values and associated final errors.

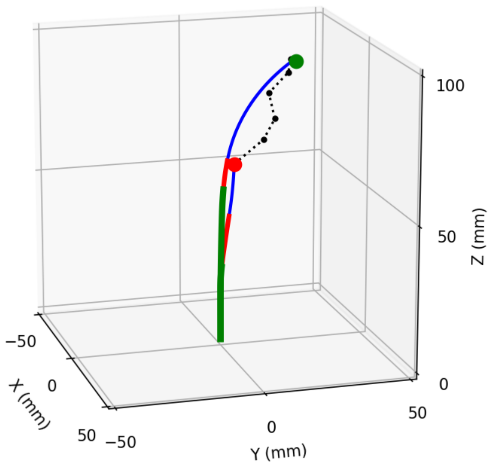


Fig. 4: Reaching a desired goal of $(20, 20, 100)$ mm with a tip error of 0.98 mm. Starting position (red), desired goal (green).

for generation of new goals. However, these constraints are non-essential in the trigonometric representation during episode steps. The extension joint β_i can be retrieved directly and has constraints

$$0 \geq \beta_3 \geq \beta_2 \geq \beta_1 \tag{2}$$

$$0 \leq L_3 + \beta_3 \leq L_2 + \beta_2 \leq L_1 + \beta_1 \tag{3}$$

from the actuation side. Lastly, the current goal tolerance, $\delta(t)$, is included in the state where t is the current timestep. A decay curriculum function was used for $250,000$ steps out of the total $500,000$ training steps. The full state, s_t, can then be defined as:

$$s_t = \{\gamma_1, \gamma_2, \gamma_3, G_{achieved} - G_{desired}, \delta(t)\} \tag{4}$$

- Action (a_t) : Actions are defined as a change in rotation and extension joint positions.

$$a_t = \{\Delta\beta_1, \Delta\beta_2, \Delta\beta_3, \Delta\alpha_1, \Delta\alpha_2, \Delta\alpha_3\} \tag{5}$$

- Goals (G) : Goals are defined as Cartesian points within the workspace of the robot. The achieved goal, $G_{achieved}$, is determined with the forward kinematics of the geometrically exact model [4] and is recomputed at each timestep as the joint configuration changes from the agents actions. The desired goal $G_{desired}$ updates at the start of every episode where a desired goal is found by sampling valid joint configurations and applying forward kinematics.
- Rewards (r_t) : The reward is a scalar value returned by the environment as feedback for the chosen action by the agent at the current timestep. The reward function used in this work is defined as:

$$r_t = \begin{cases} 0 & \text{if } e_t \leq \delta(t) \\ -1 & \text{otherwise} \end{cases} \tag{6}$$

where e_t is the Euclidean distance $||G_{achieved} - G_{desired}||$ and $\delta(t)$ is the goal-based curriculum function that determines the goal tolerance at timestep t. The workspace and various state and reward elements are illustrated in Fig. 1.

We train a policy for constraint and constraint-free rotation in simulation using the base hyperparameters and tube parameters as our previous work [5]. To evaluate each policy, 1000 evaluation episodes where the desired goal where randomized. Representing error metrics as mean error ± standard deviation, The constraint-free agent had error of 0.69 mm ± 0.24 mm with a success rate of 97.1% while constrained agent had error of 0.94 mm ± 1.44 mm and with a success rate of 87.7%. In our previous work [5], where the rotation was constrained, evaluation showed error metrics of 1.29 mm ± 0.18 mm and a success rate of 90.3%. Providing a goal of $(20, 20, 100)$ mm, the solved joint values were $[-2.36, -2.03, -0.92]$ mm for β and $[-205.6°, -108.2°, -271.4°]$ for α with a final tip error of 0.98 mm as seen in Fig. 4.

CONCLUSIONS AND DISCUSSION

Constraining the rotational DOF of the tubes results in the trained policy with worse error metrics. Moreover, with the trigonometric representation, the rotational constraints are redundant. We aim to further analysis differences in joint sampling and testing on a hardware system.

REFERENCES

[1] H. B. Gilbert, D. C. Rucker, and R. J. Webster III, "Concentric tube robots: The state of the art and future directions," *Robotics Research*, pp. 253–269, 2016.

[2] H. Alfalahi, F. Renda, and C. Stefanini, "Concentric tube robots for minimally invasive surgery: Current applications and future opportunities," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 410–424, 2020.

[3] R. Grassmann, V. Modes, and J. Burgner-Kahrs, "Learning the forward and inverse kinematics of a 6-dof concentric tube continuum robot in se(3)," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 5125–5132.

[4] D. C. Rucker, B. A. Jones, and R. J. Webster III, "A geometrically exact model for externally loaded concentric-tube continuum robots," *IEEE Transactions on Robotics*, vol. 26, no. 5, pp. 769–780, 2010.

[5] K. Iyengar and D. Stoyanov, "Deep reinforcement learning for concentric tube robot control with a goal-based curriculum," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1459–1465.