Robust Learning of Sequential Motion from Human Demonstrations to Enable Robot-guided Exercise Training

PI: Momotaz Begum, Co-PI: Dain LaRoche and Sajay Arthanat
PhD Students: Paul Gesel and Mostafa Hussein

Introduction

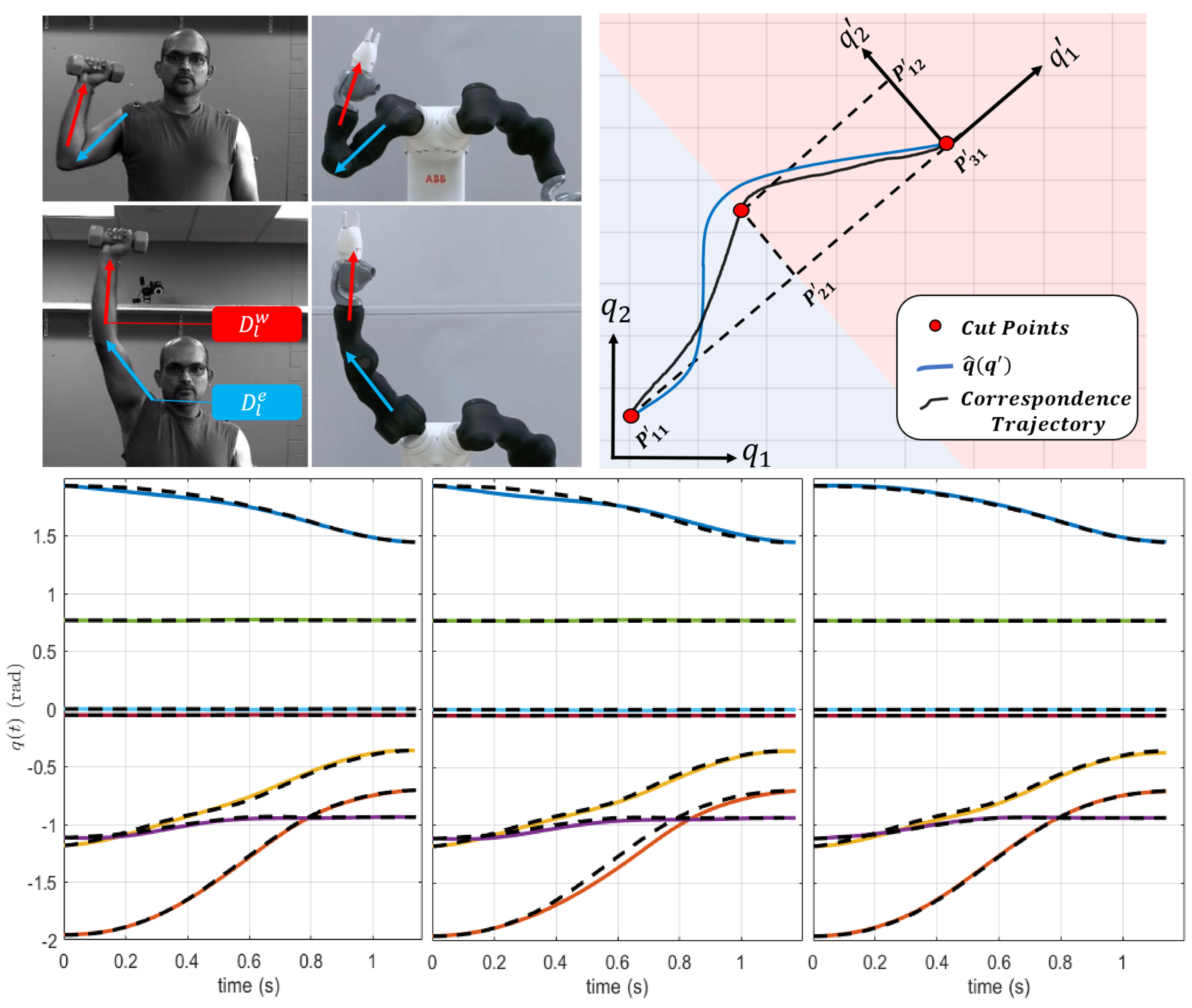
We focused on two research themes in 2020: i. Optimizing the PSM-learned trajectory and ii. Dealing with adversarial demonstrations. Our long-term goals remains the same: designing robots that can learn therapeutic exercises from demonstrations.

Learning Optimized Trajectory through PSM

PSM: A DS-based trajectory learner: We designed PSM in 2019 [1]. The PSM uses piece-wise Phase Space Transition Functions (PSTF) that drive an initial state to a desired phase space states (position, velocity). PSM-learned trajectories are not optimized.

PSM-based Trajectory Learning as an Optimal Control Problem

- Inspired by human movement literature, we hypothesize that a trajectory is optimal when the energy ($\int_0^T |\dot{\theta}\tau| dt$), torque ($\int_0^T \tau^2 dt$), and torque change ($\int_0^T \dot{\tau}^2 dt$) are minimized. We consider these quadratic functions as the basis functions of a trajectory. Three basis objective functions are then expressed in terms of PSM parameters: 1) $C_1 = \sum_{l=1}^{m} \|2E_1(Q_l^{'}) - \dot{Q}_{l1}^{'2}\|$ 2) $C_2 = \sum_{l=1}^{m} \|\tau(Q_l^{'})\|$ 3) $C_3 = \sum_{l=1}^{m-1} \|\dot{\tau}(Q_l^{'}, Q_{l+1}^{'})\|$
- The trajectory to be learned is defined as a linear combination of basis functions. We learn the trajectory through optimizing torque, jerk, and the trajectories of the robot's joints with an iterative constrained linear least squares method.
- Since PSM formulation explicitly considers robot's dynamic parameters, constraints on PSM parameters are be added to limit dynamic loads. This allows natural adaption of learned motion to varying end-effector loads.
- Since we currently rely on passive observation for demonstration data, a correspondence matrix between robot and human demonstrator is established using vector based mapping (fig. left)
- For clarity, we also express PSM parameters in rotated coordinate system (fig. right).



(left) Vector-based correspondence generation (right) Change of coordinate based on PCA (bottom) Experimental results

- We evaluated the new PSM formulation with different therapeutic exercises and an ADL task. We also compared the performance with DMP (fig. bottom)

Adversarial Demonstrations in High-Level LfD

Motivation:

- In Lfd research, a typical assumption is that the expert always provide correct demonstrations. Despite theoretical convenience, it has limited practical value.
- Only a handful of LfD frameworks consider sub-optimal demonstrations. However, all of them assume prior knowledge about which demonstrations are sub-optima.
- We are interested in directly learning a sample-efficient policy after autonomously identifying and discarding adversarial demonstrations from the training set.

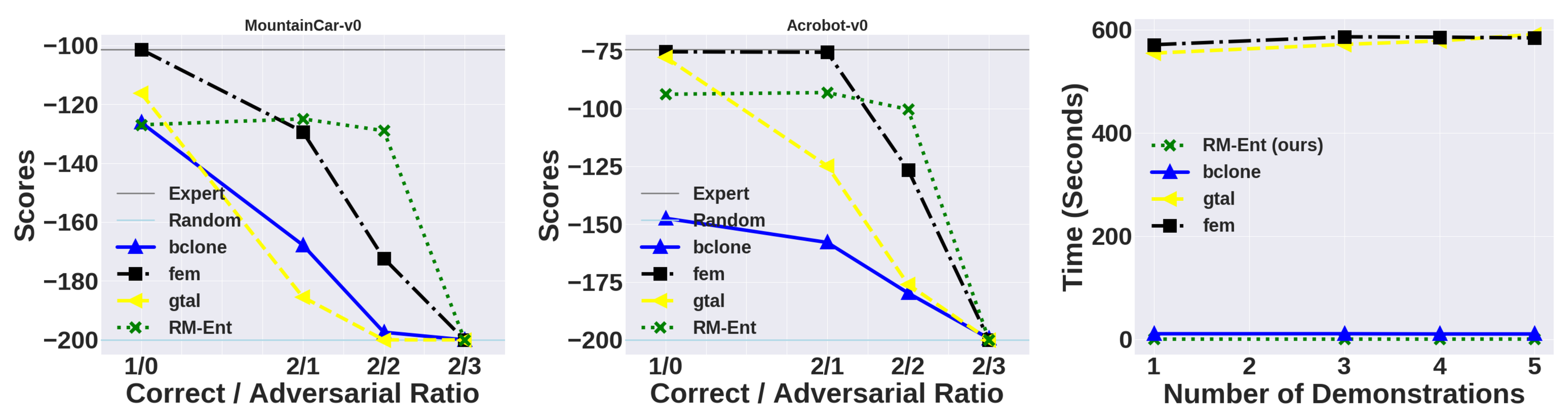
Robust Maximum Entropy Behavior Cloning (R-MaxEnt)

- From a demonstration set consisting of state-action pairs, we leverage feature expectation matching (FEM) and the maximum entropy principle to identify a base model that provides the most unbiased policy.
- We introduce a new weight variable to the base model to assign an importance weight to each demonstration.
- Optimizing for weights that minimize the entropy leads to the *min-max* problem that can be solved using Lagrange multiplier.

$$\begin{aligned} \min_{w \in^{D}} \max_{\pi \in^{\times}} \quad & -\sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} \pi(a|s)\log\pi(a|s) \sum_{d=1}^{D} w_d \cdot \tilde{p}(s,d) \\ \text{s. t.} \quad & \sum_{d=1}^{D} w_d \sum_{s\in\mathcal{S}}\sum_{a\in\mathcal{A}} f_i(s,a)\tilde{p}(s,d)\Big(\pi(a|s) - \tilde{\pi}(a|s,d)\Big) = 0, \\ & \sum_{a\in\mathcal{A}} \pi(a|s) - 1 = 0, \quad \forall s \in \mathcal{S} \qquad [\pi] \\ & \sum_{d=1}^{D} w_d = M, \quad w_d \geq 0, \quad \forall d \in \mathcal{D}, \quad w_d \leq 1 \end{aligned} \tag{1}$$

here, $\quad i = 1, \ldots, N$ and $\forall d = 1, \ldots, D$

- We evaluated R-MaxEnt on the classical control tasks *Mountain-Car* and *Acrobot* in the OpenAi-Gym simulator [3]. Results were compared against BC and a recent approach in IRL with two different objective function; (1) Linear cost function from (FEM); (2) Game-theoretic apprenticeship learning (GTAL). R-MaxEnt outperformed all other methods in the presence of adversarial demonstrations.



R-MaxEnt Evaluation with *Mountain-Car* and *Acrobot*

Reference

1) P. Gesel, M. Begum and D. LaRoche, "Learning Motion Trajectories from Phase Space Analysis of the Demonstration", IEEE ICRA 2019
2) P. Gesel et al., "Learning Optimized Human Motion via Phase Space Analysis", IEEE IROS 2020
3) M. Hussein et al., "Robust Maximum Entropy Imitation Learning", Robot Learning Workshop, NeurIPS 2020

2021 National Robotics Initiative (NRI) Principal Investigators' Meeting **Award ID: 1830597**

March 10 - 12, 2021 | Virtual Meeting