Learning Human Reaching Optimality Principles from Minimal Observation Inverse Reinforcement Learning

Sarmad Mehrdad[1], Maxime Sabbah[2], Vincent Bonnet[2,3], Ludovic Righetti[1,4]

Abstract—This paper investigates the application of Minimal Observation Inverse Reinforcement Learning (MO-IRL) to model and predict human arm-reaching movements with time-varying cost weights. Using a planar two-link biomechanical model and high-resolution motion-capture data from subjects performing a pointing task, we segment each trajectory into multiple phases and learn phase-specific combinations of seven candidate cost functions. MO-IRL iteratively refines cost weights by scaling observed and generated trajectories in the maximum entropy IRL formulation, greatly reducing the number of required demonstrations and convergence time compared to classical IRL approaches. Training on ten trials per posture yields average joint-angle Root Mean Squared Errors (RMSE) of 6.4 deg and 5.6 deg for six- and eight-segment weight divisions, respectively, versus 10.4 deg using a single static weight. Cross-validation on remaining trials and, for the first time, inter-subject validation on an unseen subject's 20 trials, demonstrates comparable predictive accuracy, around 8 deg RMSE, indicating robust generalization. Learned weights emphasize joint acceleration minimization during movement onset and termination, aligning with smoothness principles observed in biological motion. These results suggest that MO-IRL can efficiently uncover dynamic, subject-independent cost structures underlying human motor control, with potential applications for humanoid robots.

I. INTRODUCTION

Understanding the optimal principles underlying simple motions like human arm reaching is crucial for progress in both neuroscience and robotics. In neuroscience, these principles shed light on how the central nervous system plans and executes goal-directed movements under constraints such as muscle redundancy, sensory noise, and biomechanical limitations. Capturing these strategies helps elucidate motor control mechanisms and supports clinical rehabilitation by identifying deviations from optimality in pathological movements. In robotics and human–robot interaction, modeling human reaching as an optimal control problem facilitates

the design of bio-inspired controllers and predictive algorithms. This is especially valuable for humanoids, assistive robotics and prosthetics, where human-likeness and intent prediction are critical. Biological motion exhibits invariant properties despite the wide range of available motor strategies. Voluntary movements tend to follow consistent patterns, suggesting that the nervous system resolves motor redundancy by adhering to specific organizational principles, the so-called optimal weights. However, the precise link between cost functions and variables encoded by the central nervous system remains unclear. Moreover, the idea of a single universal cost function may be unrealistic. The central nervous system might flexibly adjust cost weightings based on task demands [1], or even during the same task. For example, individuals may reduce velocity at the end of a reach to aim more accurately, while seeking overall speed. This suggests a balance between objective (task-related) and subjective (body-related) costs, which current models often fail to capture [2].

A widely used framework for exploring the principles underlying motor control is optimal control theory, which makes the hypothesis that biological movements arise from the minimization of specific cost or loss functions. Numerous models based on this theory have been proposed [3], [4], many of which claim to replicate experimental data with reasonable accuracy. However, these models often rely on a single cost function per task, which may not adequately capture the complexity and variability of human motion. As a result, relatively high Root Mean Square Errors (RMSEs) are frequently observed between the predicted and measured trajectories. For instance, Sylla et al. [4] reported an average RMSE of 7deg, with some angles exhibiting errors superior to 15deg, even for simple reaching movements. Moreover, many studies unfortunately do not report any quantitative comparison or use tailored metrics [3] between their model predictions and experimental data. This raises questions about the relevance and predictive power of such models, especially when considering the sensitivity of their outcomes to variations in the chosen cost function components.

Unfortunately, because of the current limitation of

[1] Machines in Motion Laboratory, New York University, USA
[2] LAAS-CNRS, Université Paul Sabatier, CNRS, Toulouse, France.
[3]Image and Pervasive Access Laboratory (IPAL), CNRS-UMI, 2955, Singapore.
[4] Artificial and Natural Intelligence Toulouse Institute (ANITI), Toulouse

Inverse Optimal Control (IOC) and Inverse Reinforcement Learning (IRL) methods used to retrieve optimal cost function weights from human optimal motion, a single set of parameters for a given task is generally used [5]. Indeed, adding time-varying weights leads to a significant increase in the number of parameters to be identified which these algorithms struggle to handle. In this paper, we leverage a new efficient IRL algorithm to instead study how time-varying weights lead to a more nuanced and accurate description of the movement.

IOC provides a model-based framework for inferring cost function weights that best explain observed human motion, assuming that the motion is optimal for some performance criterion [5]. Despite its conceptual appeal, practical applications of IOC face significant challenges. The standard bi-level formulation, in which cost weights are optimized through repeated solutions of a nested optimal control problem, is computationally expensive, often requiring several days of computation. Additionally, this approach is prone to convergence to local minima, particularly in high-dimensional problems. To address these issues, alternative formulations based on the residuals of the Karush-Kuhn-Tucker (KKT) conditions have been proposed. These methods aim to eliminate the need for repeated trajectory optimization. However, they remain highly sensitive to measurement noise and modeling errors commonly observed in human motion data [6], [7]. More recently, promising hybrid approaches that combine elements of the bi-level and residual-based formulations have been introduced [8], although these have only been validated in simulation.

In contrast to IOC, IRL adopts a probabilistic framework to infer the underlying cost function. This approach relaxes the number of bi-level-like iterations and is especially appealing for tasks involving uncertainty and variability, such as those performed by humans. IRL defines a probability distribution over all demonstrations and seeks to identify the cost function that maximizes the likelihood of the optimal trajectories. Given this definition, IRL ideally requires all the possible trajectories for the utmost optimal cost function derivation, which is impossible. Hence, IRL's performance is heavily hinged on the trajectory space approximation accuracy obtained from a finite set of observations. There have been several efforts to circumvent this shortcoming by approximating the trajectory set through more intelligent sampling around the optimal trajectory [9], [10] and trajectory set augmentation [11].

However, the sampled trajectories often lie close to the observed ones and may not sufficiently explore the broader trajectory space. Furthermore, in order to improve cost function estimation, IRL must consider all sampled and iteratively generated trajectories in the probability maximization process. This requirement substantially increases the computational cost of IRL. To address these shortcomings, we turn to the newly proposed Minimal Observation Inverse Reinforcement Learning (MO-IRL) [12]. MO-IRL takes an iterative approach for cost function estimation, by approximating the trajectory space through scaling the effectiveness of each observed trajectory depending on its current estimate of optimality. With this added feature, even with a small observation set, MO-IRL empirically provides better iterates that lead to an improved estimation of the weights. This leads to iterative cost function learning with minimal information about the trajectory space, resulting in considerably faster convergence. To our knowledge, MO-IRL was designed and tested only with robotics tasks and fixed weights.

In this paper, we extend MO-IRL to learn tasks requiring time-varying cost weights and investigate its use in predicting accurate human joint trajectories by learning simultaneously from positions and velocities. The proposed approach is validated with a subset of reference human data from the human motor control community [3]. In particular, we show that the method can learn task weights leading to accurate movement reproduction that also generalize across movements.

II. METHODS

A. Experimental protocol and mechanical model

The human data used in this study were kindly provided by Berret al. [3]. These data were used in numerous other studies since their publication and are considered a reference. They consist of motion capture 3D marker positions, including markers located at the shoulder, elbow, and wrist level. Data from twenty right-handed naive subjects were provided. For this preliminary study, the data of only two subjects were selected. Subjects first had to sign the approved local ethical committee ASL-3 ("Azienda Sanitaria Locale", local health unit), Genoa, Italy. Then, they performed the pointing task as depicted in Fig. 1.a. While seated, participants were instructed to perform a series of pointing movements toward a vertical target bar positioned in front of the participant. Only shoulder and elbow flexion/extension were permitted during the task, as wrist movement was restricted. The shoulder-to-bar horizontal distance was set to 95% of the participant's total arm length ($L = L_1 + L_2$, with L_1 and L_2 representing upper arm and forearm lengths, respectively; Fig. 1.a). Five initial arm postures, labeled P1 through P5, were defined using reference points positioned in a vertical plane located approximately 10 cm lateral to the right shoulder. These five postures corresponded to specific predefined angular configurations of the arm and are shown in Fig. 1.b. Each individual performed 20 trials for each posture.
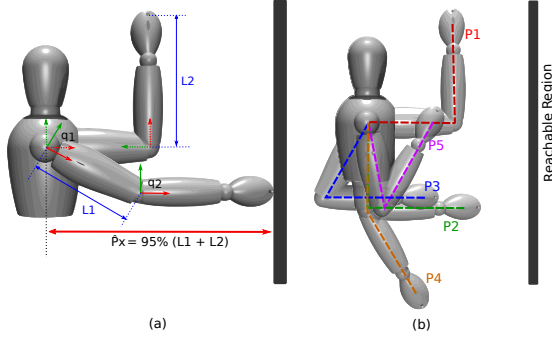
Fig. 1. (a) Biomechanical model definition, showing the beginning and the end of the pointing task. (b) Five different initial postures for the pointing task [3].

Label	Name	Equation	Reference
Φ_1	Cartesian velocity	$\sum_{t=0}^{T} \dot{P}(t)^T \dot{P}(t)\, dt$	[18]
Φ_2	Energy	$\sum_{t=0}^{T} \left| \dot{\mathbf{q}}(t)^T \mathbf{u}(t) \right|\, dt$	[19], [20]
Φ_3	Geodesic	$\sum_{t=0}^{T} \dot{\mathbf{q}}(t)^T M \dot{\mathbf{q}}(t)\, dt$	[21]
Φ_4	Joint acceleration	$\sum_{t=0}^{T} \ddot{\mathbf{q}}(t)^T \ddot{\mathbf{q}}(t)\, dt$	[22]
Φ_5	Joint torque change	$\sum_{t=0}^{T} \dot{\boldsymbol{\tau}}(t)^T \dot{\boldsymbol{\tau}}(t)\, dt$	[23], [24]
Φ_6	Joint velocity	$\sum_{t=0}^{T} \dot{\mathbf{q}}(t)^T \dot{\mathbf{q}}(t)\, dt$	[25]
Φ_7	Joint torque	$\sum_{t=0}^{T} \boldsymbol{\tau}(t)^T \boldsymbol{\tau}(t)\, dt$	[26]

A planar biomechanical model, illustrated in Fig.1.a, was developed to represent flexion/extension movements at the shoulder (q_1) and elbow (q_2) joints. The model's base frame was located at the shoulder joint, and the relative positions of successive joints in their parent frames were computed using segment lengths L_1 and L_2, estimated from marker data. Inertial parameters were calculated using anthropometric tables [13].

B. Optimal control problem

In the context of pointing or reaching movements, Berret et al. [3] proposed a set of $N_\Phi = 7$ candidate cost functions, as detailed in Table I. Although the task may appear elementary, we posit that individuals do not adhere to a single cost function throughout the entire movement. For instance, it is intuitive to expect a deceleration near the endpoint to ensure accurate and controlled pointing. To account for such time-varying motor strategies, each recorded trajectory of duration T was segmented into N_w equal time windows, each comprising N_s samples. This segmentation was chosen based on consistent inflection points observed in the majority of trajectories. To model the temporal evolution of movement strategies, we introduced a weight matrix $\boldsymbol{\omega} \in \boldsymbol{R}^{N_\Phi \times N_w}$, which allows distinct cost function contributions across different movement phases. The full trajectory $\boldsymbol{x} \in \boldsymbol{R}^{N_s \times N_w}$ was defined as the concatenation of state vectors $\boldsymbol{x}_s = (\boldsymbol{q}_s, \dot{\boldsymbol{q}}_s)$ for each section $s \in \{1, \cdots, N_w\}$. Accordingly, $\boldsymbol{u}_s$ is defined as the control torque input to the human model joints for each section. The associated Direct Optimal Control (DOC) problem was then formulated to reconstruct the observed human motion over this multi-phase framework.

$$
\boldsymbol{x}^* = \arg\min_{\boldsymbol{u}} \ \sum_{s=1}^{N_w} \sum_{j=1}^{N_\Phi} \omega_{s,j} \Phi_j(\mathbf{x}_s, \boldsymbol{u}_s)
$$
$$
\begin{aligned}
\text{s.t.} \quad & \boldsymbol{x}(t+1) = f(\boldsymbol{x}(t), \boldsymbol{u}(t)), \\
& P_X(T) = \hat{P}_X, \\
& \boldsymbol{q}^- \leq \boldsymbol{q} \leq \boldsymbol{q}^+ \\
& \boldsymbol{q}(0) = \boldsymbol{q_0} \\
& \|\dot{\boldsymbol{q}}\| \leq \dot{\boldsymbol{q}}^+ \\
& \dot{\boldsymbol{q}}(0) = \dot{\boldsymbol{q}}(T) = 0
\end{aligned} \tag{1}
$$

where $\boldsymbol{x}(t)$ and $\boldsymbol{u}(t)$ are the state and control at discrete time t; f is the Euler time-discretized dynamics; $P(t)$ is the position of the hand obtained from forward kinematics, and $\hat{P}_X$ is the goal position on the horizontal axis the subject aims to reach as shown Fig. 1.a; $\boldsymbol{q_0}$ is the initial human joint configuration; $\boldsymbol{q}^-$, $\boldsymbol{q}^+$ are the lower and upper joint boundaries respectively; $\dot{\boldsymbol{q}}^+$ is the maximal joint velocity.

We used Pinocchio [14] for modeling the human body, together with the Croccoddyl framework [15] and the MiM_Solver nonlinear CSQP solver [16] to define and solve the constrained DOC in Eq.(1). We used MuJoCo [17] for the model simulation.

C. Minimum Observation Inverse Reinforcement Learning (MO-IRL)

As mentioned before, we aim to solve an inverse optimal control problem to derive the optimal cost function as a linear combination of explicit features. For this purpose, we use an augmented MO-IRL [12] framework to accommodate learning for several time windows with different weights. As commonly used in IRL algorithms, MO-IRL aims to maximize the probability of the optimal demonstrations (i.e. the human data)

$$
\boldsymbol{\omega}^* = \arg\max_{\boldsymbol{\omega}} P(\boldsymbol{x}^* | \boldsymbol{\omega}, \bar{\boldsymbol{x}}) \tag{2}
$$

$$
\text{where} \quad P(\boldsymbol{x}^* | \boldsymbol{\omega}, \bar{\boldsymbol{x}}) = \frac{e^{-\boldsymbol{\omega}^T \boldsymbol{\Phi}^*}}{\sum_{i=1}^{K} e^{-\boldsymbol{\omega}^T \boldsymbol{\Phi}_i}}
$$

$$
\boldsymbol{\omega} \geq 0
$$

in which $\bar{x}$ is the set of K observed trajectories. For brevity, we write $\boldsymbol{\omega}$ the concatenated weight vector for the cost features and $\boldsymbol{\Phi}$ the concatenated feature vector. In the following, $\boldsymbol{\Phi}(\boldsymbol{x}_i, \boldsymbol{u}_i)$ which is the feature costs for the ith trajectory is henceforth referred to as $\boldsymbol{\Phi}_i$. As observed in [12], all the sub-optimal trajectories are being incorporated and contribute equally in the denominator, irrespective of how close they are to optimality. This lack of distinction can create numerical issues for the optimizer. Therefore, it is important to scale them to emphasize their effectiveness so the optimizer will have better information about what the approximated trajectory set represents in terms of optimality.

MO-IRL solves Eq. (2) by iteratively improving $\boldsymbol{\omega}$ rather than optimizing it in one shot. Considering an update of cost weights at each iteration $n+1$ in the form $\boldsymbol{\omega}_{n+1} = \boldsymbol{\omega}_n + \Delta\boldsymbol{\omega}_n$, the original probability distribution can be rewritten to instead find the best $\Delta\boldsymbol{\omega}_n$:

$$
\begin{aligned}
\Delta\boldsymbol{\omega}_n^* = \underset{\Delta\boldsymbol{\omega}_n}{\arg\min} &-\log \frac{1}{1 + \sum_{\boldsymbol{x}_i \in \bar{\boldsymbol{x}}} \gamma_i e^{-\Delta\boldsymbol{w}_t^T(\boldsymbol{\Phi}_i - \boldsymbol{\Phi}^*)}} \\
\text{s.t.} \quad & \Delta\boldsymbol{\omega}_n > -\boldsymbol{\omega}_n \\
& \gamma_i = e^{-\boldsymbol{\omega}_n^T(\boldsymbol{\Phi}_i - \boldsymbol{\Phi}^*)}
\end{aligned} \tag{3}
$$

In this case, sampled trajectories are automatically scaled depending on their cost in the previous iteration. MO-IRL solves Eq. (3) and then seeks to find an update of the form $\omega_{t+1} = \omega_t + \alpha\Delta\omega$ where α is selected using a merit function (similar to a line search procedure). Starting with $\alpha = 1$, the algorithm checks if the resulting trajectory is closer to the optimal demonstration by evaluating the merit function. If the merit function value has not been decreased with the added change to the weight, MO-IRL scales down α by factor of 0.25, and tries again for a maximum of 10 trials. If by the 10th trial there was no improvement, the algorithm stops. Otherwise, the accepted trajectory is added to the observed trajectory set $\bar{x}$, and MO-IRL moves on to the next iteration.

In the literature, algorithms to learn human motion trajectories usually minimize the gap between the estimated and the optimal trajectories in joint space without considering velocities. In this study, however, we propose to minimize the gap in both joint position ($\boldsymbol{q}$) and joint velocity ($\dot{\boldsymbol{q}}$) concurrently. Therefore, we evaluate the estimation improvement based on the full state vector $\boldsymbol{x} = [q_1, q_2, \dot{q}_1, \dot{q}_2]$. We define the merit function as $m(\boldsymbol{x}) = \frac{1}{T}||\boldsymbol{x}^* - \boldsymbol{x}||_2^2$. For this study, we discard all previously generated non-optimal trajectories from $\bar{x}$ and use only the most recent generated non-optimal trajectory as the trajectory set, i.e., $\bar{x} = \{\boldsymbol{x}_t\}$ as we empirically noticed that it leads to faster convergence. Initial weights are set to small uniform value ($\boldsymbol{\omega} = 0.05$).

We extend MO-IRL for learning multiple weight sections from multiple demonstrations:

$$
\begin{aligned}
\Delta\omega_t^* = &\underset{\Delta\omega_t}{\arg\min} \\
\sum_{d=1}^{D} &\left(-\log \frac{1}{1 + \sum_{\boldsymbol{x} \in \bar{\boldsymbol{x}}} \gamma_i e^{-C(\boldsymbol{x}_i, \Delta\boldsymbol{\omega}_t)}} \right) + \frac{\beta}{2}||\Delta\omega_t||_2^2 \\
\text{s.t.} \quad & \Delta\boldsymbol{\omega}_t > -\boldsymbol{\omega}_t \\
& \gamma_i = e^{-C(\boldsymbol{x}_i, \omega_t)} \\
& C(\boldsymbol{x}_i, \Delta\omega_t) = \sum_{s=1}^{N_\omega} \Delta\boldsymbol{\omega}_{st}^T(\Phi_{si} - \Phi_{sd}^*)
\end{aligned} \tag{4}
$$

where D is the number of provided optimal demonstrations, and N_ω is the number of weight sections as mentioned before. We also use a small $L2$ regularizer ($\beta = 10^{-10}$) for the optimization to prevent high changes in weights and overfitting. When learning from multiple demonstrations, the merit function for step acceptance is changed to $m(\boldsymbol{x}) = \frac{1}{D}\sum_{d=1}^{D}(\frac{1}{N}||\boldsymbol{x}_d^* - \boldsymbol{x}||_2^2)$.

D. Learning and Cross-Validations

This section describes the learning and cross-validation processes of the proposed method. The human trial dataset consists of 20 trials of the wall-reaching task for five initial postures for 2 human subjects. In addition to accuracy, we aim to test the generalizability of the learned weights, i.e. can one subject's trials be informative enough to predict other trials. We test our algorithm by comparing the joint value RMSE of the generated trajectory by the DOC using the learned weights against optimal trajectories executed by the human subject.

Cost weights are learned for each posture using 10 randomly selected trials from one subject and are cross-validated on the remaining 10 trials. The quality of the learned weights is evaluated by the average RMSE between the trials and the trajectories generated by the DOC using the learned weights. For further assessment of the MO-IRL's efficacy, we also perform an Inter-Subject cross-validation (ISCV) to see how the learned weights predict another subject's motion for the same posture. This helps evaluate whether the learned weights can generalize to another individual.

III. RESULTS

We evaluate the results for the two subjects, one for training and cross-validation, and another for ISCV, from the dataset provided by Berret et al. [3]. We evaluate our framework for each initial posture using 1, 6, and 8 weight sections. Note that for each posture, weights are identified separately as different strategies per posture were suggested by Berret et al. [3].

Fig. 2 illustrates a comparison between the measured and estimated joint trajectories for posture 2, after learning the weights in 8 sections. The figure reveals that
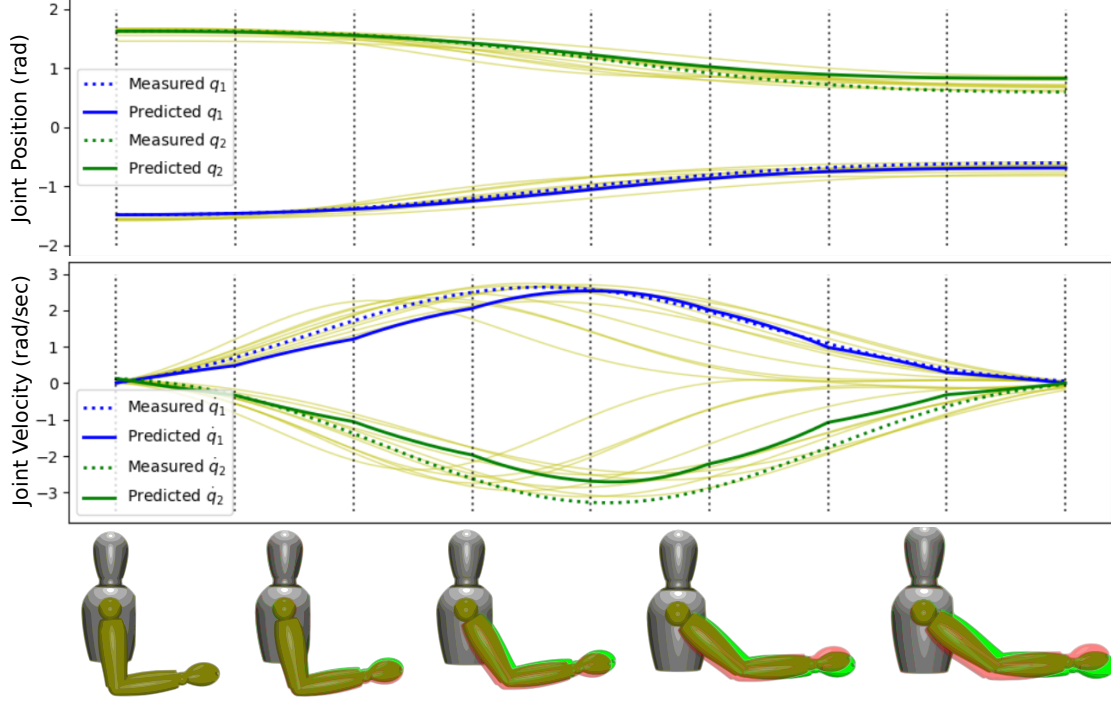
Fig. 2. Illustration of MO-IRL prediction against the actual human task execution. The cost weights are divided into 8 sections. The training data for both joint positions and velocities are shown in yellow. The dotted lines are the real trajectory performed by the human, and the solid lines are the MO-IRL predictions.

while the human demonstrations are generally similar, they exhibit noticeable variability, particularly in joint velocities, which vary more than joint positions. This observation supports our claim regarding multi-modality in IRL. Such variability is beneficial for MO-IRL, as it prevents the algorithm from fitting a cost function that reproduces a single trajectory too closely. Instead, it favors trajectories that lie within the general vicinity of the demonstrated motions. This mitigates overfitting, especially in the case of joint positions, which tend to be more consistent across trials. Furthermore, subtle changes that could be expected from segmenting the weights are visible as shown in the joint velocity profiles.

Table II shows the corresponding RMSE values for q_1, q_2, and $\boldsymbol{q} = [q_1, q_2]$, after training, cross-validation, and ISCV for 1, 6, and 8-section weights for each initial posture. When a single section is used, the average RMSE was 10.4deg, while it was 6.4deg and 5.6deg for 6- and 8-sections during training, respectively. It indicates that time-varying weights are important although adding more weight sections (from 6 to 8) does not drastically improve the overall prediction average RMSE.

One can see from these results that postures 3 and 5 are more challenging for the MO-IRL to learn, as their RMSE is nearly twice as large as the other postures. This can be attributed to the nature of these initial postures that require the human subject to have more

activity in the joint space to result in a fairly small motion in the task space. In other words, in the tasks where the initial elbow angle is more acute than others, more change in the joint space is required to result in a similar task space motion. Interestingly, both the cross-validation and ISCV trials exhibit average RMSE values that are very similar to those obtained on the training set. The achieved accuracy, about 8 degrees for cross validation, is significantly lower than results found in the literature that usually use a single set of weights [3]–[5].

Fig. 3 shows the learned weights for various postures and sections (the weights are normalized for better presentation). The varying weights in both 6- and 8-section cases, supported by the high RMSE observed with single-section weights, corroborate the fact that one uniform weight set might not be enough to correctly explain the intent of humans, even in a simple task like reaching a wall. In contrast to Berret et al. [3], the energy-related cost function was only marginally observed in postures 3, 4, and 5 of the investigated trials. This discrepancy may stem from the differing normalization techniques employed. While Berret et al. used the so-called pivot method [27], our proposed approach does not enforce normalization, thereby allowing greater flexibility in the inferred cost weights.

Nevertheless, as shown in Fig. 3, joint acceleration minimization appears to play a dominant role, especially
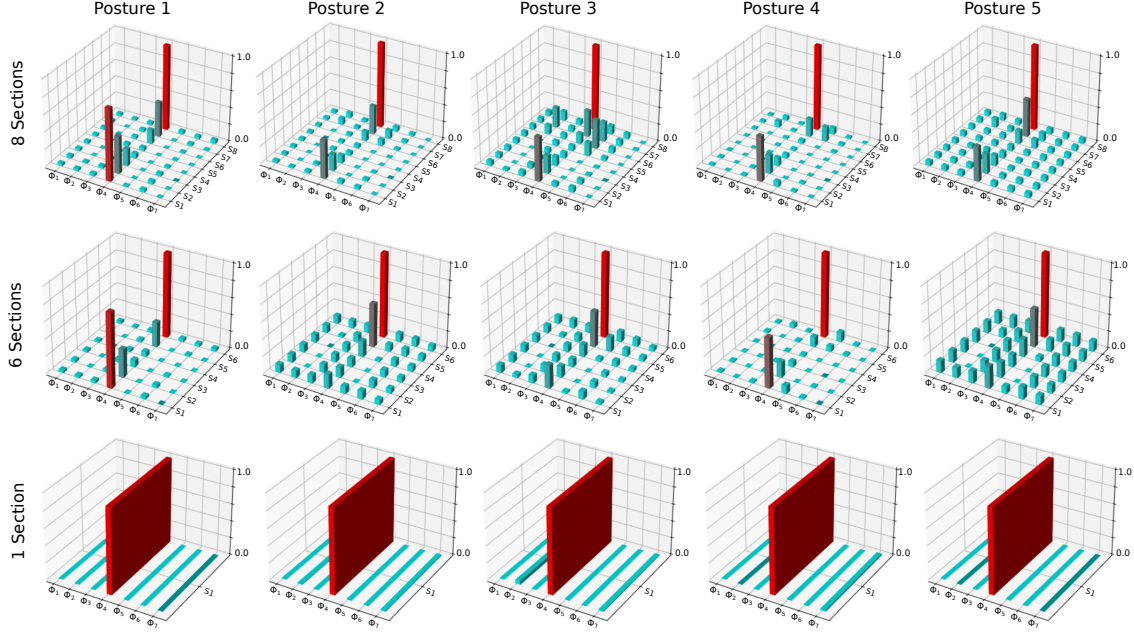
Fig. 3. Normalized weights learned by MO-IRL for each posture given 1, 6, and 8 sections.
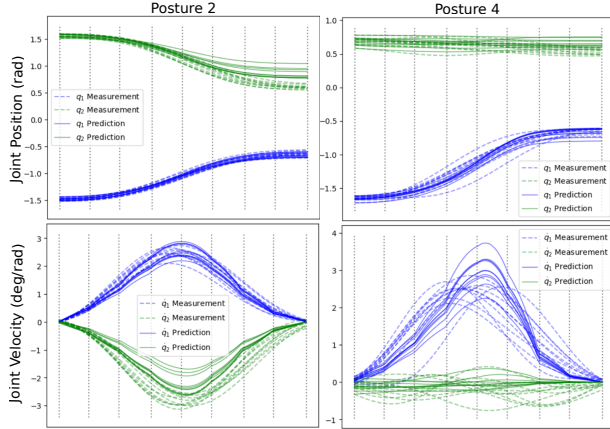

Fig. 4. Inter-Subject Cross-Validation of the learned weights (8 sections) by MO-IRL for initial postures 2 and 4. The top row shows the overlayed measured and predicted joint values from the second subject that are not used for the MO-IRL training, where q_1 and q_2 are shown by blue and green, respectively. Predictions (DOC solutions) and measured trajectories are shown with solid and dashed lines, respectively. The bottom row shows the corresponding joint velocities of the top row trajectories. The trajectories are normalized in length for clearer presentation.

during the initiation and termination phases of the movement, suggesting a strategy aimed at ensuring smooth motion onset and precise stopping. The movement onset likely reflects a strategy to ensure a smooth and stable initiation, avoiding abrupt or energetically costly changes in motor commands. This result is also consistent with prior studies showing that reaching trajectories typically exhibit bell-shaped velocity profiles and low jerk [18], suggesting implicit optimization of higher-order derivatives of position.

Reducing acceleration towards the end of the motion may serve to finely tune the final position, increasing precision, and ensuring comfortable deceleration before reaching the target. These findings support the notion that motor control is not governed by a single, static cost function across the entire trajectory, but rather by a dynamic trade-off between competing criteria such as effort minimization, accuracy, and smoothness, adapted to the temporal structure of the task.

Fig. 4 shows the ISCV results for postures 2 and 4 by comparing the actual human joint position and velocity trajectories of the second subject, with the predicted trajectories solved by DOC using the learned weights. This figure shows that not only do the predictions look very similar to the measured human motion, but also the scatteredness of the predicted trajectories resemble that of the measured trajectories. This preliminary test suggests that the learned behavior from one subject is transposable to other individuals, while retaining the variability expected from human behavior.

Another notable observation is that while human trajectories (especially in joint velocity) differ from one trial to another, even for a single subject and posture, the DOC predictions are deterministic and only depend on initial conditions $\boldsymbol{x}(0)$. The weights learned capture an average behavior based on the 10 trials it was trained on. We posit that this helps learn key common feature from the movements while discarding less relevant variations. We can see this effect in Fig. 4 where the predictions do

not necessarily match joint velocity profiles, but they all reach for the goal, same as the human intended to.

To the best of our knowledge, this study is the first to perform actual ISCV. While conclusions should be drawn with caution, as only one subject was included in the ISCV test, the results are promising: the obtained RMSE in ISCV is of the same order of magnitude as in the training and cross-validation trials. This may indicate that the identified weights generalize well and are not overfitted despite the use of time-varying weights. This is likely achieved thanks to the use of joint velocity directly in the MO-IRL's step search and also the use of weight regularization ($L2$ norm in Eq.(4)).

IV. CONCLUSION

This paper proposes a framework based on MO-IRL to predict human arm-reaching motions using time-varying weights. Our empirical results demonstrate the benefits of using time-varying weights in the cost function to learn human movements with explainable cost functions. Importantly, we achieved lower reconstruction errors than previously reported in the literature. Our results suggest a strong emphasis on joint acceleration at the beginning and end of the movement while other cost features appear less dominant. This is consistent with the

idea that humans will generally refrain from highly accelerated movements in the initial and terminal stages of the task. We also provided preliminary results for inter-subject analysis of the resulting cost function, which exhibits promising results on the ability to generalize the learned cost function from one subject to another. The reported RMSE values for the cross-validation and their close similarities to the ISCV results show promises towards a generalizable IRL framework for understanding human task intentions and also towards transferring human movements onto humanoids.

Our future work includes:

- Testing the algorithm on a higher number of subjects to analyze its generalizability.
- Conducting sensitivity and occlusion tests to further understand the effect of individual cost features.
- Incorporating multimodal step-acceptance in MO-IRL (based on exerted force, joint torque, etc.) to improve convergence.
- Testing the framework on more complex tasks that include more dynamic behaviors.

V. ACKNOWLEDGMENTS

TABLE II
RMSE ON THE JOINT ANGLES (DEG) FOR 1, 6, AND 8 WEIGHT SECTIONS
FOR TRAINING, CROSS-VALIDATION, AND ISCV PROCESSES

	Training (10 trials)			Cross-Validation (10 trials)			ISCV (20 trials)		
	1 Section	6 Sections	8 Sections	1 Section	6 Sections	8 Sections	1 Section	6 Sections	8 Sections
	Posture 1								
q_1	11.32±13.9	5.87 ± 8.57	**2.54 ± 1.23**	8.71 ± 3.72	4.15 ± 3.76	**2.98 ± 3.52**	7.99 ± 5.91	**7.87 ± 5.03**	8.94 ± 5.04
q_2	14.65±12.4	9.52 ± 9.48	**5.52 ± 4.98**	5.97 ± 2.73	5.75 ± 5.19	**4.99 ± 5.29**	9.40 ± 2.29	7.94 ± 4.24	**7.55 ± 2.94**
q	13.30±13.0	8.13 ± 8.84	**4.41 ± 3.48**	7.58 ± 2.98	5.23 ± 4.28	**4.17 ± 4.44**	9.13 ± 3.60	**8.11 ± 4.28**	8.50 ± 3.63
	Posture 2								
q_1	15.10±3.61	4.91 ± 2.4	**4.46 ± 2.10**	7.32 ± 2.38	**3.54 ± 2.75**	4.19 ± 2.68	9.30 ± 4.97	3.58 ± 3.80	**3.31 ± 2.51**
q_2	13.96±2.56	7.16 ± 1.74	**6.83 ± 2.46**	13.88±6.96	9.52 ± 6.59	**8.63 ± 2.58**	15.8 ± 4.08	**8.21 ± 3.67**	8.89 ± 4.14
q	14.55±3.08	6.33 ± 1.47	**5.87 ± 2.01**	11.26±4.84	7.26 ± 4.93	**6.89 ± 2.34**	13.2 ± 3.66	**6.56 ± 3.31**	6.88 ± 3.07
	Posture 3								
q_1	11.78±3.67	**4.53 ± 1.65**	7.06 ± 3.33	14.31±2.60	**8.93 ± 2.26**	9.85 ± 2.55	13.05±19.5	**8.63 ± 4.17**	8.84 ± 5.30
q_2	12.84±3.27	8.81 ± 2.12	**7.21 ± 3.78**	17.07±3.42	18.64±6.68	**16.4 ± 7.40**	19.52±9.57	15.85±8.34	**15.8 ± 10.1**
q	12.38±3.28	**7.12 ± 1.24**	7.30 ± 3.22	15.87±2.32	14.69±4.75	**13.6 ± 5.27**	16.90±8.06	**13.0 ± 6.09**	13.03±7.76
	Posture 4								
q_1	3.97 ± 2.15	3.69 ± 1.88	**3.47 ± 1.85**	2.78 ± 1,78	2.95 ± 1.70	**2.57 ± 1.84**	4.67 ± 2.14	4.13 ± 2.03	**4.08 ± 2.04**
q_2	4.12 ± 1.94	4.15 ± 1.61	**3.88 ± 1.75**	2.76 ± 1.56	3.06 ± 1.90	**2.53 ± 1.48**	4.67 ± 2.14	4.03 ± 2.84	**4.01 ± 2.80**
q	4.31 ± 1.43	4.14 ± 1.15	**3.83 ± 1.47**	3.00 ± 1.12	3.27 ± 1.28	**2.83 ± 1.14**	4.57 ± 2.02	4.34 ± 1.97	**4.31 ± 1.96**
	Posture 5								
q_1	**7.05 ± 2.84**	8.17 ± 2.29	7.83 ± 1.62	16.86±4.73	9.83 ± 5.84	**8.26 ± 5.67**	13.80±7.80	**7.86 ± 5.33**	9.63 ± 6.43
q_2	9.34 ± 4.17	7.54 ± 3.71	**7.23 ± 3.16**	19.12±4.95	10.74±7.38	**9.81 ± 5.15**	16.27±5.91	**10.6 ± 4.66**	10.86±4.57
q	8.38 ± 3.32	8.33 ± 1.40	**7.64 ± 2.18**	18.09±4.58	10.44±6.42	**9.33 ± 4.96**	15.54±5.82	**9.80 ± 4.11**	10.73±4.62
	ALL Postures								
q_1	9.84 ± 3.90	5.43 ± 1.54	**5.07 ± 2.05**	10.00±5.03	5.88 ± 2.90	**5.57 ± 2.94**	9.76 ± 3.36	**6.41 ± 2.12**	6.96 ± 2.69
q_2	10.98±3.89	7.44 ± 1.85	**6.13 ± 1.29**	11.76±6.35	9.54 ± 5.30	**8.47 ± 4.73**	13.13±5.35	**9.33 ± 3.88**	9.42 ± 3.89
q	10.41±3.81	6.43 ± 1.42	**5.60 ± 1.57**	10.88±5.51	7.71 ± 3.86	**7.02 ± 3.77**	11.45±4.27	**7.87 ± 2.80**	8.19 ± 2.93

thorough discussions on the topic.

REFERENCES

[1] S. Cao, Z. Luo, and C. Quan, "Online inverse optimal control for time-varying cost weights," *Biomimetics*, vol. 9, no. 2, p. 84, 2024.

[2] K. E. Zelik and A. D. Kuo, "Mechanical work as an indirect measure of subjective costs influencing human movement," *PLoS one*, vol. 7, no. 2, e31143, 2012.

[3] B. Berret, E. Chiovetto, F. Nori, and T. Pozzo, "Evidence for composite cost functions in arm movement planning: an inverse optimal control approach," *PLoS Comput. Biol.*, vol. 7, no. 10, p. e1002183, 2011.

[4] N. Sylla, V. Bonnet, G. Venture, N. Armande, and P. Fraisse, "Human arm optimal motion analysis in industrial screwing task," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2014, pp. 964–969.

[5] J. F.-S. Lin, P. Carreno-Medrano, M. Parsapour, M. Sakr, and D. Kulić, "Objective learning from human demonstrations," *Annu. Rev. Control.*, vol. 51, pp. 111–129, 2021.

[6] J. Colombel, D. Daney, and F. Charpillet, "On the reliability of inverse optimal control," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8504–8510.

[7] F. Bečanović, J. Miller, V. Bonnet, K. Jovanović, and S. Mohammed, "Assessing the quality of a set of basis functions for inverse optimal control via projection onto global minimizers," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 7598–7605.

[8] F. Becanovic, K. Jovanović, and V. Bonnet, "Reliability of single-level equality-constrained inverse optimal control," in *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*, Nancy, France, Nov. 2024, pp. 623–630.

[9] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, "Stomp: Stochastic trajectory optimization for motion planning," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 4569–4574.

[10] M. Kalakrishnan, P. Pastor, L. Righetti, and S. Schaal, "Learning objective functions for manipulation," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 1331–1336.

[11] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization," in *International conference on machine learning*. PMLR, 2016, pp. 49–58.

[12] S. Mehrdad, A. Meduri, and L. Righetti, "Cost function estimation using inverse reinforcement learning with minimal observations," 2025, submitted to IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2025. [Online]. Available: https://arxiv.org/abs/2505.08619

[13] R. Dumas, L. Cheze, and J.-P. Verriest, "Adjustments to mcconville et al. and young et al. body segment inertial parameters," *J. Biomech.*, vol. 40, pp. 543–553, 2007.

[14] J. Carpentier, G. Saurel, G. Buondonno, J. Mirabel, F. Lamiraux, O. Stasse, and N. Mansard, "The pinocchio c++ library: A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives," in *2019 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2019, pp. 614–619.

[15] C. Mastalli, R. Budhiraja, W. Merkt, G. Saurel, B. Hammoud, M. Naveau, J. Carpentier, L. Righetti, S. Vijayakumar, and N. Mansard, "Crocoddyl: An efficient and versatile framework for multi-contact optimal control," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2536–2542.

[16] A. Jordana, S. Kleff, A. Meduri, J. Carpentier, N. Mansard, and L. Righetti, "Stagewise implementations of sequential quadratic programming for model-predictive control," *Subm. IEEE TRO*, 2023.

[17] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.

[18] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *J. Neurosci.*, vol. 5, pp. 1688–1703, 1985.

[19] J. Nishii, "Energetic optimicality of arm trajectory," in *Proc. Int. Conf. on Biomechanics of Man, 2002*, 2002.

[20] B. Berret, C. Darlot, F. Jean, T. Pozzo, C. Papaxanthis, and J. P. Gauthier, "The inactivation principle: mathematical solutions minimizing the absolute work and biological implications for the planning of arm movements," *PLoS Comput. Biol.*, vol. 4, p. e1000194, 2008.

[21] A. Biess, D. G. Liebermann, and T. Flash, "A computational model for redundant human three-dimensional pointing movements: integration of independent spatial and temporal motor plans simplifies movement dynamics," *J. Neurosci.*, vol. 27, pp. 13 045–13 064, 2007.

[22] S. Ben-Itzhak and A. Karniel, "Minimum acceleration criterion with constraints implies bang-bang control as an underlying principle for optimal trajectories of arm reaching movements," *Neural Comput.*, vol. 20, pp. 779–812, 2008.

[23] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement," *Biol. Cybern.*, vol. 61, pp. 89–101, 1989.

[24] E. Nakano, H. Imamizu, R. Osu, Y. Uno, H. Gomi, T. Yoshioka, and M. Kawato, "Quantitative examinations of internal representations for arm trajectory planning: minimum commanded torque change model," *J. Neurophysiol.*, vol. 81, pp. 2140–2155, 1999.

[25] C. G. Atkeson and J. M. Hollerbach, "Kinematic features of unrestrained vertical arm movements," *J. Neurosci.*, vol. 5, pp. 2318–2330, 1985.

[26] W. L. Nelson, "Physical principles for economies of skilled movements," *Biol. Cybern.*, vol. 46, pp. 135–147, 1983.

[27] A. M. Panchea, N. Ramdani, V. Bonnet, and P. Fraisse, "Human arm motion analysis based on the inverse optimization approach," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. IEEE, 2018, pp. 1005–1010.