Submitted version of the contribution presented at the IEEE ETFA 2022 conference in Stuttgart.
Published manuscript available at DOI: 10.1109/ETFA52439.2022.9921553

Inverse Optimal Control for the identification of human objective: a preparatory study for physical Human-Robot Interaction

1st Paolo Franceschi
CNR-STIIMA; UniBs
Milano, Italy; Brescia, Italy
paolo.franceschi@stiima.cnr.it

2nd Nicola Pedrocchi
CNR-STIIMA
via Alfonso Corti 12, 20133, Milano, Italy
nicola.pedrocchi@stiima.cnr.it

3rd Manuel Beschi
UniBs
via Branze 38, 25123, Brescia, Italy
manuel.beschi@unibs.it

***Abstract*—Nowadays, many applications involving humans and robots working together require physical interaction. It is known that, during an interaction, the mutual understanding and knowledge of the partner's goal improves and allows natural interaction. For this purpose, this work proposes Inverse Optimal Control (IOC) to recover the cost function of a human performing a reaching task with a robot in passive impedance control. This work presents the potentialities and limitations of the presented IOC method to describe human objectives. This work represents a preparatory study toward smooth and natural physical Human-Robot Interaction (pHRI), intending to understand the basic information on humans' behavior.**

***Index Terms*—Inverse Optimal Control, physical Human-Robot Interaction**

I. Introduction

With the large spread of collaborative robots and applications, the need to make Human-Robot Interaction as natural as possible arises, especially when they must cooperate through physical interaction, in the so-called physical Human-Robot Interaction (pHRI) [1], [2]. Indeed, in pHRI, the human operator and the robot must execute a task together, and the action of the one has immediate consequences on the other [3].

It was shown that humans understand each other's intentions while physically interacting to perform a task [4]. While this naturally comes for humans, the same does not apply to the robotic partner. Then, it is also interesting to allow the robot to understand its partner's objective.

Many techniques and models are studied and presented in the literature to describe human and human-machine behavior. Such models span from elementary models (humans as a spring) to complex neuromuscular models and Optimal Control models, depending on the need. For a complete and up-to-date review, please see [5].

Provided that Game Theory (GT) represents a robust framework to describe an individual's behavior during interaction [6], this work focuses on the identification of optimal behaviors of humans, and the recovery of the human cost function, as this modeling allows direct integration into GT frameworks. Optimal Control (OC) aims to find a control action for a dynamic system over a time window such that a cost function is minimized. Conversely, Inverse Optimal Control (IOC) techniques are adopted to recover the cost function that produced observed control actions and state histories.

The most common strategy used in IOC is to parametrize an unknown objective function as a weighted sum of relevant features (or basis functions) with unknown weights [7]. This model is also adopted to describe human behaviors in performing different tasks. In [8], human jumping is studied, highlighting the possibility that the cost function varies during the task. In [9] and [10] study human movement by detecting changes to the optimization criterion during a squat task. In [11], human arm reaching is studied, considering free-space reaching motions. The results show a trade-off between kinematics and dynamics-based controllers depending on the reaching task. Interestingly, this trade-off depends on the initial and final arm configurations. [12] also studies human arm motions, with applications to human-robot collaboration in a shared workspace. These works involve cost functions with multiple features but consider only motions in the free space, without accounting for interactions.

If the only cost function's features considered are the state and control action, the problem is reduced to a typical Linear Quadratic Regulator (LQR). In this case, more specific IOC techniques exist. Such techniques compute the weight matrices starting from the feedback matrix and not the complete control and state histories[13] [14] [15]. Interestingly, such methods do not require the entire trajectory to identify the cost function, but only the knowledge of the gain matrix is sufficient. Finally, such approaches appear suitable to study humans in interaction with autonomous agents, described by Game Theory, as in [16], [17], [18].

Motivated by the previous studies, this work aims at implementing IOC to understand the fundamental behavior of humans interacting with a variable, passive system. It is possible to obtain basic yet valuable information to develop natural pHRI controllers. The simple LQR model is studied. Indeed, despite its simplicity, as shown, LQR cost functions

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101006732 (Drapebot).

can capture essential human behaviors during the interaction. Moreover, modeling the human's cost function as the LQR cost function allows easy integration in Linear Quadratic Game-Theoretic (LQGT) framework, which provides useful tools for pHRI controllers design [19], [20], [21].

Even though it was shown that the recovered cost functions of the humans have different values if recovered from manual control or shared control [22], this work, as a preparatory study, aims at understanding the differences and things in common between various subjects cost functions. This provides useful insights into humans' behavior, which should be considered for future studies.

II. METHOD

In this section, the robot behavior is modeled as a Cartesian impedance subject to the external force applied by the human, and the system is rewritten in a state-space formulation. Optimal control to model human behavior is introduced, and the Inverse Optimal Control technique is adopted to recover the unknown human control objective.

A. System modeling

Working in the Cartesian space is more intuitive and natural for the human operator; hence the desired robot motion at the end-effector is implemented as an impedance model in the Cartesian space:

$$M_i\ddot{x}(t) + D_i\dot{x}(t) + K_i(x(t) - x_0(t)) = u_h(t) \tag{1}$$

where M_i, D_i and $K_i \in R^{6\times6}$ are the desired inertia, damping and stiffness impedance matrices, respectively, $x_0(t) \in R^6$ is the vector containing the equilibrium positions at the end-effector, $u_h(t) \in R^6$ represents the human effort applied to the system. The Cartesian coordinates in x and x_{ref} are defined according to [23], with the vector $x = [p^T\ \phi^T]^T$ where p^T are the position coordinates and ϕ^T the set of Euler angles that defines the rotation matrix describing the end-effector orientation. With $\dot{x} = [v^T\ \omega^T]^T$ is indicated the vector containing the linear and angular velocities.

After some reformulation, (1) can be rewritten in a linearized state-space formulation around the working point as

$$\dot{z} = Az + B_h u_h \tag{2}$$

where $z = [x - x_0\ \dot{x}]^T \in R^{12}$ is the state space vector, while the matrices

$$A = \begin{bmatrix} 0^{6\times6} & J_a \\ -M_i^{-1}K_i & -M_i^{-1}D_i \end{bmatrix}$$

and

$$B_h^{12\times6} = \begin{bmatrix} 0^{6\times6} \\ M_i^{-1} \end{bmatrix}$$

are the state and input matrices, respectively, with $0^{6\times6}$ denoting a 6×6 zero matrix and J_a the analytical Jacobian matrix, with the dimensions of the considered Cartesian components.

Robot controllers typically accept as control input reference positions or velocities in the joint space; hence, it is worth converting the reference velocities from the Cartesian space to the joint space.

Given the reference velocity in the Cartesian space, the following relation allows obtaining reference velocities in the joint space

$$\dot{q}_{ref}(t) = J(q)^+\dot{x}(t), \tag{3}$$

with $\dot{q}_{ref}(t) \in R^n$, where n represents the number of joints, are the reference velocities in the joint space, $J(q)^+$ is the pseudoinverse of the geometric Jacobian matrix. Simple integration allows commanding joint positions instead of velocities to the robot. Considering that today's robots have excellent tracking performance in the frequency range excitable by the operator, in this work, hypothesis $\dot{q} \simeq \dot{q}_{ref}$ is assumed to hold.

B. Optimal Control Problem

Many previous works describe human's intention as the minimization of a quadratic cost function. In particular, it is possible to assume the cost function as quadratic on the state and control input, leading to the LQR formulation of the problem, (as examples see [19], [21], [24]). Therefore, the human objective can be described as the minimization of a quadratic cost function, defined as

$$J_h = \int_0^\infty (z Q_h z + u_h R_h u_h)\, dt \tag{4}$$

where J_h is the cost that the human incurs, $Q_h \in R^{2n\times2n}$ is a matrix containing weights on the state, and $R_h \in R^{6\times6}$ is a matrix of weights on the control input.

Remark 1. *The human's model weights Q_h and R_h are unknown and must be identified. A reasonable estimate of the human parameters, $\hat{Q}_h$ and $\hat{R}_h$, can be obtained with the IOC method as presented in the following section.*

Given the control objective in (4), and the system dynamics in (2), the LQR Optimal Control Problem can be summarized as

$$\begin{aligned} \min_u J_h = \int_0^\infty (z Q_h z + u_h R_h u_h)\, dt \\ s.t. \dot{z} = A z + B_h u_h \\ z(t_0) = z_0 \end{aligned} \tag{5}$$

The LQR optimal control has the feedback form

$$u_h = -K z(t) \tag{6}$$

in which the human control action is described as linear feedback proportional to the system's state. In (6), the matrix K represents the feedback gain matrix, copmuted as

$$K = R_h^{-1} B^T P \tag{7}$$

in which, the matrix P is the unique solution for the feedback linear-quadratic optimal control problem, solution of the infinite horizon Continuous Algebraic Riccati Equation (CARE), given by

$$A^T P + PA - PB_h R_h^{-1} B_h^T P + Q_h = 0. \tag{8}$$

C. Human objective identification

Given the definition of the direct Optimal Control Problem as in (5), this section addresses the Inverse Optimal Control problem. That is, given observed state and control histories, denoted as $\bar{z}$ and $\bar{u}_h$, given the system dynamics in (2), recover the cost function (4) that produced such control histories.

As discussed in the Introduction, different IOC techniques exist. Among the various, this work implements the one presented in [15], considering the human as the only active player, as briefly described in this section.

1) Gain identification: The method relies on the knowledge of the feedback gain matrix. Because the gain matrix of the human is not known, it has to be recovered. If the complete trajectory (or a sufficient part of it) is known, it is possible to apply the Least Square Method (LSM), and the matrix K_h is obtained from

$$\hat{K}_h = \arg\min_{K_h} \int_{t_i}^{t_f} \|K_h \bar{z}(t) + \bar{u}_h(t)\|^2 \, dt \tag{9}$$

where t_i and t_f indicate the initial and final time of the trajectory (or a sufficient portion of it), and the symbol $\hat{(\cdot)}$ indicates an estimate of the real value.

2) Inverse Optimal Control: This section briefly presents the main steps to recover the cost function of a player. Refer to [15] for the full treatment. As preliminary, denote $\oplus$ as the Kronecker sum, defined as

$$X \oplus Y = (X \otimes I_q) + (I_r \otimes Y) \tag{10}$$

with $X \in R^{r \times r}$ and $Y \in R^{q \times q}$.

Given A, B, and $\hat{K}_h$, the Inverse Optimal Control (IOC) problem can be solved as follows. First, denote F, $F_\oplus$, $\hat{K}_\oplus$ and Z as

$$\begin{aligned} F &= A - B\hat{K}_h, \\ F_\oplus &= F^T \oplus F^T, \\ \hat{K}_\oplus &= \hat{K}_h^T \oplus \hat{K}_h^T, \\ Z &= (I_n \otimes B^T) F_\oplus^{-1} \\ M &= \begin{bmatrix} Z & Z\hat{K}_\oplus + \hat{K}^T \otimes I \end{bmatrix} \end{aligned}$$

Then, denote θ as the vector of vectorized weights as

$$\theta = [vec(\hat{Q})\ vec(\hat{R})]^T. \tag{11}$$

As demonstrated in [15], if θ satisfies

$$M\theta = 0 \tag{12}$$

the $\hat{Q}$ and $\hat{R}$ are the correct unkown parameter for (8).

Given that (12) is a reformulation of (8), the parameter set of the inverse LQ differential system is given by

$$\theta = ker(M) \tag{13}$$

with convex boundaries such that $R_h > 0$. As typical in IOC problems, a residual is defined as

$$r = M\theta \tag{14}$$

to consider non-optimal behaviors and imperfect modeling of $\hat{K}$. The following quadratic problem is formulated to minimize the residual

$$\begin{aligned} \min_{\theta} \quad & ||r||_2^2 = \frac{1}{2}\theta^T H \theta \\ \text{s.t.} \quad & I\theta \geq 0 \\ & R > 0 \end{aligned} \tag{15}$$

with $H = 2M^T M$.

Remark 2. *All the $\lambda\theta$, $\lambda > 0$ solutions are acceptable, since the solution of problem* (5) *is the same for any λJ, $\lambda > 0$.*

Remark 3. *The constraints in* (15) *are imposed by Optimal Control necessary conditions, i.e. matrix Q semi-positive definite and matrix R strictly positive definite.*

III. Experiments

In this section, experimental results are presented. First, the design of experiments is presented, with comments on the evaluation procedure, then results are presented and analyzed.

A. Design of experiments

Three subjects are involved in this study. Each subject is asked to move the robot end-effector to a fixed, unknown set-point along the z-axis. The impedance parameters of the robot are set to $M = 10$, and three values of damping $D = [25, 50, 75]$ are tested to verify how different systems modify the control objective. The stiffness is set $K = 0$. Each subject is asked to perform 12 set-point reaching for each damping value for a total of 36 reaching tasks. Each reaching is along the x–axis only, with the set-point at various, random distances. Before each recording session, the subject is allowed to practice for a while to gain confidence with the system.

The measured exchanged forces, the reference position, and the actual positions compute the IOC problem. The forces are measured with a Robotiq FT300 sensor mounted on the robot tip. The robotic system gives the positions directly, and the velocities are computed by differentiating the positions.

The human cost function matrices, as typically happens [16], [18], are assumed to be nonzero only on the diagonal terms, resulting in $Q = diag([q_1, q_2])$, and $R = r$. Moreover, since the solution to the optimal control problem is the same for any λJ_h with λ scalar, as additional constraint to (15) is set $q_1 = 1$. In this way, all the recovered cost functions are comparable.

A first analysis is done on the recovered features (q_1, q_2 and r) from (15) to check their variation with respect to the variations of the system. In the ideal scenario, the weights can be different between different subjects. Despite this, each subject should have constant weight values, even when varying the system's parameters.

A measure of the control input is defined to evaluate the goodness of the recovered costs as

$$\mathscr{E}_u = \frac{\int_{t_0}^{t_f} u(t) - \hat{u}(t)\, dt}{u_{max}} \tag{16}$$

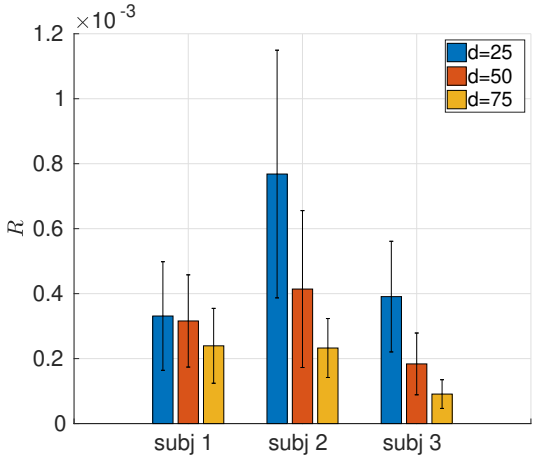


Fig. 1: Mean and standard deviations of the recovered control weights

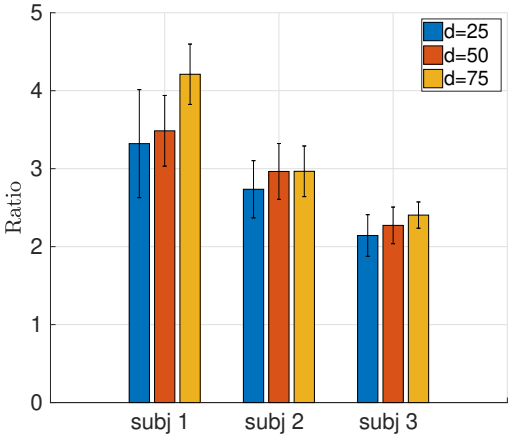


Fig. 2: risetime over setpoint computation

where u_{max} is the maximum measured control effort, and it is used to normalize the errors to make the comparison fair. The recovered control input $\hat{u}(t)$ is computed by running a system simulation. In the ideal case, this value should be zero. Defining the set-point as $x_{sp} = x_{ref} - x_0$, and the Raise Time RT as the time required to move the system from x_0 to $0.95\,x_{sp}$, it is possible to define a ratio

$$RT_{sp} = \frac{RT}{x_{sp}}, \tag{17}$$

to check how the different systems influence the raise time by removing dependency from the set-point distance. In the ideal case, the ratio should be the same for a given system with any set-point and should increase by increasing the damping.

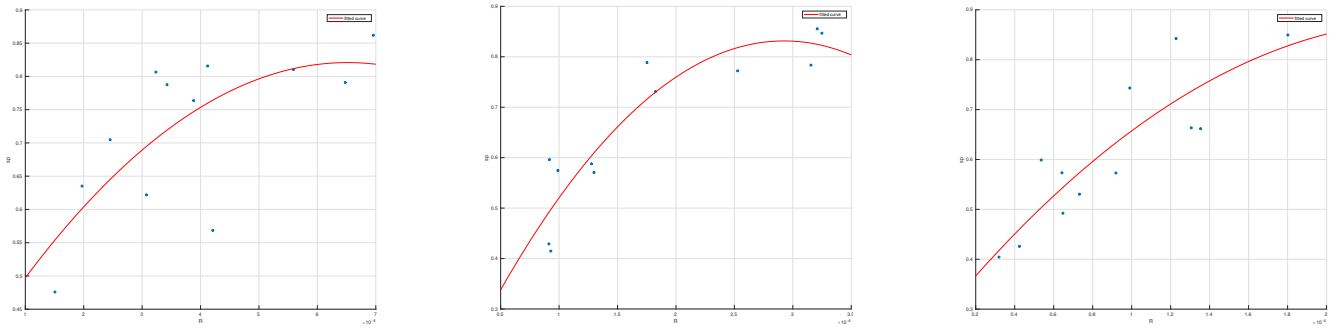
Finally, a possible relation between the distance from set-point and weight on control is investigated. In the ideal case, the values of R should always be the same, without regard for the set-point.

B. Results

As previously discussed, the weight q_1 is constrained to be $q_1 = 1$, and the other weights q_2 and r are computed consequently. From the experiments, all the subjects have a negligible q_2, with values in the order of $1e-4$. This means that the human feedback is based mainly on position, as also happens in [22]. Since the values of q_2 are very close to 0, the rest of this paper will approximate them with 0 without further analysis.

Given $q_2 = 0$, the ratio between q_1 and r drives the system's response. The analysis of the values of R, visible in figure 1, shows that humans have a low weight also in how much effort is put into the task. The low value of r, compared with q_1, means that a fast set-point reaching is more favorable than saving effort.

Moreover, three subjects tend to decrease the value of r as the damping increase, which means that reaching the desired position at a specific time should remain almost constant against changes in the system.



(a) control weight identified with D=25

(b) control weight identified with D=50

(c) control weight identified with D=75

Fig. 3: Correlation between control weight identified and different setpoints, for subject 3

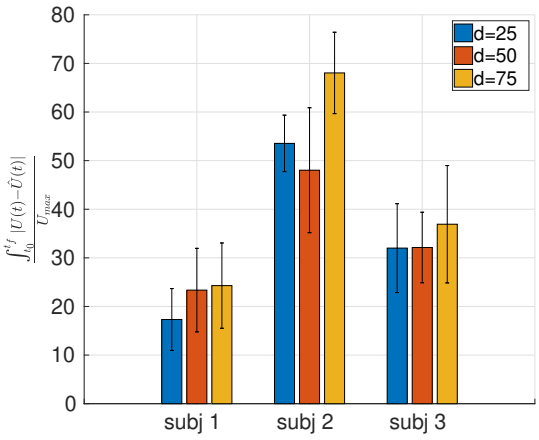


Fig. 4: Mean and standard deviations of the control inputs

Figure 2 also displays such a behavior. shows the ratio RT_{sp}.

The Figure shows a minimal increment in the ratio compared with the increment of damping, particularly in subjects 2 and 3, confirming that humans appear more interested in time than in the effort required.

This study also shows a correlation between the set-point distance and the weight a human gives to the control effort. The farther the set-point is, the higher the control weight, as shown in figure 3.

The error between the measured and reconstructed control input is analyzed. Figure 4 shows the normalized error between the measured and the simulated control effort of the three subjects in the three scenarios.

Finally, figure 5 presents one control history measured and recovered for each subject in the study, with the corresponding normalized error.

C. Discussion

This work's analysis suggests human modeling as an optimal controller and hints about real human intention. The most relevant result is that humans tend to keep the required time as constant as possible when performing a task. This is probably a consequence of the fact that a human control works as feedback on a visual stimulus and keeps this control -which is not considered in the presented study- constant.

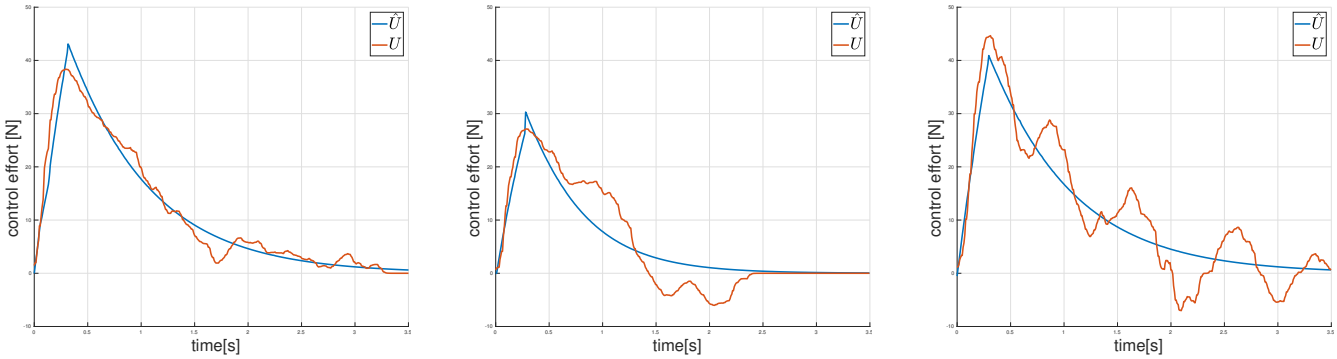
Despite this, optimal control modeling of a human interacting with a robot can still be adopted, keeping in mind some baseline rules shown in the results section. The human's weight on control input decreases if the system becomes more rigid against motions and decreases as the desired position becomes closer to the current one.

IV. Conclusions

This work presents a study of the human behavior modeled as optimal control during physical Human-Robot Interaction. Theoretical modeling of the system and control model is presented, and the formulation of the Inverse Optimal Control is presented. IOC is used to recover the cost functions of three subjects performing a reaching task interacting with different robot behavior. The analysis shows that OC modeling of a human may be used and gives good results with some caution.

Such a model presents applications in designing game-theoretical-based controllers for pHRI, in which the knowledge of the other's cost function is necessary. Moreover, it can also be used as a reference in modeling human behavior for digital twin simulations.

Future works will study more complex models with different cost functions that consider other features, such as acceleration, jerk, time that results relevant, and energy. Moreover, different interacting strategies, such as cooperative and non-cooperative game theory, will be analyzed to understand how two humans behave when performing interacting tasks and which one model interaction better.



(a) Computed and measured effort of subject 1. $\mathscr{E}_u = 17.81$

(b) Computed and measured effort of subject 2. $\mathscr{E}_u = 46.32$

(c) Computed and measured effort of subject 3. $\mathscr{E}_u = 40.02$

Fig. 5: Example of one trial of each subject.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101006732 (Drapebot).

REFERENCES

[1] S. Haddadin and E. Croft, *Physical Human–Robot Interaction*. Cham: Springer International Publishing, 2016, pp. 1835–1874.

[2] A. De Santis, B. Siciliano, A. De Luca, and A. Bicchi, "An atlas of physical human–robot interaction," *Mechanism and Machine Theory*, vol. 43, no. 3, pp. 253–270, 2008.

[3] E. Matheson, R. Minto, E. G. G. Zampieri, M. Faccio, and G. Rosati, "Human–robot collaboration in manufacturing applications: A review," *Robotics*, vol. 8, no. 4, 2019. [Online]. Available: https://www.mdpi.com/2218-6581/8/4/100

[4] A. Takagi, G. Ganesh, T. Yoshioka, M. Kawato, and E. Burdet, "Physically interacting individuals estimate the partner's goal to enhance their movements," *Nature Human Behaviour*, vol. 1, no. 3, pp. 1–6, 2017.

[5] A. Scibilia, N. Pedrocchi, and L. Fortuna, "Human control model estimation in physical human;machine interaction: A survey," *Sensors*, vol. 22, no. 5, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/5/1732

[6] N. Jarrassé, T. Charalambous, and E. Burdet, "A framework to describe, analyze and generate interactive motor behaviors," *PLOS ONE*, vol. 7, no. 11, pp. 1–13, 11 2012.

[7] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 1.

[8] K. Westermann, J. F.-S. Lin, and D. Kulić, "Inverse optimal control with time-varying objectives: application to human jumping movement analysis," *Scientific reports*, vol. 10, no. 1, pp. 1–15, 2020.

[9] W. Jin, D. Kulić, J. F.-S. Lin, S. Mou, and S. Hirche, "Inverse optimal control for multiphase cost functions," *IEEE Transactions on Robotics*, vol. 35, no. 6, pp. 1387–1398, 2019.

[10] J. F.-S. Lin, V. Bonnet, A. M. Panchea, N. Ramdani, G. Venture, and D. Kulić, "Human motion segmentation using cost weights recovered from inverse optimal control," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 1107–1113.

[11] O. S. Oguz, Z. Zhou, S. Glasauer, and D. Wollherr, "An inverse optimal control approach to explain human arm reaching control based on multiple internal models," *Scientific reports*, vol. 8, no. 1, pp. 1–17, 2018.

[12] J. Mainprice, R. Hayne, and D. Berenson, "Goal set inverse optimal control and iterative replanning for predicting human reaching motions in shared workspaces," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 897–908, 2016.

[13] M. Menner and M. N. Zeilinger, "Convex formulations and algebraic solutions for linear quadratic inverse optimal control problems," in *2018 European Control Conference (ECC)*, 2018, pp. 2107–2112.

[14] F. Jean and S. Maslovskaya, "Inverse optimal control problem: the linear-quadratic case," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 888–893.

[15] J. Inga, E. Bischoff, T. L. Molloy, M. Flad, and S. Hohmann, "Solution sets for inverse non-cooperative linear-quadratic differential games," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 871–876, 2019.

[16] F. Köpf, J. Inga, S. Rothfuß, M. Flad, and S. Hohmann, "Inverse reinforcement learning for identification in linear-quadratic dynamic games," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14 902–14 908, 2017.

[17] S. Rothfuß, J. Inga, F. Köpf, M. Flad, and S. Hohmann, "Inverse optimal control for identification in non-cooperative differential games," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14 909–14 915, 2017, 20th IFAC World Congress. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896317334602

[18] J. Inga, M. Flad, and S. Hohmann, "Validation of a human cooperative steering behavior model based on differential games," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 3124–3129.

[19] Y. Li, G. Carboni, F. Gonzalez, D. Campolo, and E. Burdet, "Differential game theory for versatile physical human–robot interaction," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 36–43, Jan 2019.

[20] Y. Li, K. P. Tee, R. Yan, W. L. Chan, and Y. Wu, "A framework of human–robot coordination based on game theory and policy iteration," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1408–1418, 2016.

[21] Y. Li, K. P. Tee, W. L. Chan, R. Yan, Y. Chua, and D. K. Limbu, "Continuous role adaptation for human–robot shared control," *IEEE Transactions on Robotics*, vol. 31, no. 3, pp. 672–681, 2015.

[22] J. Inga, M. Eitel, M. Flad, and S. Hohmann, "Evaluating human behavior in manual and shared control via inverse optimization," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 2699–2704.

[23] B. Siciliano and L. Villani, "Robot force control," 2000.

[24] X. Na and D. J. Cole, "Linear quadratic game and non-cooperative predictive methods for potential application to modelling driver–afs interactive steering control," *Vehicle system dynamics*, vol. 51, no. 2, pp. 165–198, 2013.