2023 20th International Conference on Ubiquitous Robots (UR)
June 25-28, 2023. Hawaii Convention Center, Honolulu, Hawaii

Human preferences' optimization in pHRI collaborative tasks

Paolo Franceschi[1,2], Marco Maccarini[3], Dario Piga[3], Manuel Beschi[1,2], Loris Roveda[3]

Abstract— **Humans and robots working together have mutual benefits. The first is great at adaptation to new situations and has very high intellectual capabilities, while robots are very effective in assisting humans with heavy/repetitive tasks. In physical Human-Robot Interaction (pHRI), one challenge is to tune the robot's controller to make the interaction with the human as comfortable and natural as possible. Indeed, robots' perception is different from human to human. Moreover, depending on the target task, different robot tuning may be preferred. In this work, an assistive Game-Theoretic based controller is presented, and its parameters are tuned according to different subjects' preferences. Preference Based Optimization (PBO) allows optimization based on human preferences and feelings. The aim of this study is twofold: present a methodology for fast tuning of a pHRI controller according to the subjective preferences in different situations, and study general human preferences according to two different tasks. The two tasks investigated are a precise path-following task and a fast-reaching task. Experimental evaluations are conducted, and subjective and objective performances are evaluated and discussed. Finally, a questionnaire is proposed to the subjects to evaluate the applicability of the proposed method.**

I. INTRODUCTION

Physical Human-Robot Interaction (pHRI) is the study of the interaction between humans and robots with explicit physical contact [1]. Applications span from teleoperation [2], co-transportation [3], rehabilitation [4], and more [5].

Impedance control [6] represents the basis for safe and smooth pHRI, with the adaptive evolution of IC in the adaptation of the impedance set-point ([7], [8]), and adaptation of the mass-spring-damper parameters ([9]–[13]), leading to hybrid controllers that combine both as in [14]–[16].

In the literature, to describe pHRI, Game Theory (GT) was proven as a valuable tool to analyze complex interactive behaviors involving multiple agents, as discussed in [17]. The use of a game-theoretical framework, mainly dealing with the non-cooperative situation, is in [18], and similarly, [19]. These works use the so-called Nash Equilibrium solution as a tool to update online the robot cost function, measuring the interaction force applied by a human. The result is a variable impedance control with damping and stiffness updated online, according to the non-cooperative game. The same differential non-cooperative game-theoretic problem is also solved in [20] and [21] by a policy iteration. A general game-theoretic framework involving an observer for the opponents' control laws identification allows addressing different game-theoretical behaviors (Cooperative and Non-Cooperative) for pHRI in [22], for the two agents game and in [23], with extension to multiple agents. Such modeling is extended in [24] for trajectory tracking in the non-cooperative scenario. Cooperative Game Theory for pHRI is studied in [25], where is proposed a control framework capable of deforming the robot's trajectory during interaction with the human. Given that GT provides suitable models to address pHRI, this work wants to investigate the use of the Cooperative formulation in order to develop an assistive controller for a human operator.

Given the control framework, different parameters have to be tuned. In particular, the control algorithm typically has parameters that depend on the user, which can only be estimated and parameters that can be set according to the required task, performances, and level of experience of the user. For example, according to previous research, [26]–[28], operators prefer low values of the impedance parameters for large movements at high velocities, while high values when performing fine movements at low velocities. In [8], these insights are used to make a variable admittance control. The damping parameter is made variable according to the velocity of the robot within a stable region.

Moreover, human subjective preferences may influence the tuning of the controllers. Different works analyze this crucial aspect of the HRI, even in domains not directly related to pHRI. The works in [29], [30] proposes the tuning of robotic prosthesis control parameters to meet the user's preference while ensuring their safety. [31] evaluates the participants' perception of the collaboration, and [32] evaluates the fluency of human-robot collaboration based on both subjective and objective evaluations. [33] defines a questionnaire to try to remove the subjective part from evaluations. Finally, [34] defines subjective metrics to assess an HRI motion planner. Therefore, it is clear that a method for control parameter tuning should be defined according to the various requirements, not least human preferences.

A novel method to optimize parameters based on human preferences is the so-called Preference Based Optimization (PBO). PBO aims at optimizing a set of parameters where a cost function cannot be directly evaluated but only recovered from preferences expressed by a user. By iteratively proposing to the user a new comparison to make, the algorithm learns a surrogate of the cost function preferences expressed by the user. Among the various, in this work, We selected the

[1]Paolo Franceschi and Manuel Beschi are with the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing of the National Research Council of Italy (CNR-STIIMA), via Alfonso Corti 12, 20133, Milano, Italy. paolo.franceschi@stiima.cnr.it

[2]Paolo Franceschi and Manuel Beschi are also with the University of Brescia, Dipartimento di Ingegneria Meccanica ed Industriale, via Branze 38, 25123, Brescia, Italy

[3] Marco Maccarini, Dario Piga and Loris Roveda are with Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Scuola Universitaria Professionale della Svizzera Italiana (SUPSI), Università della Svizzera italiana (USI), via la Santa 1, 6962, Lugano, Switzerland marco.maccarini@idsia.ch, loris.roveda@idsia.ch, dario.piga@idsia.ch

GLISp algorithm [35] to tune the robot controller's parameters. Such an algorithm has proven great results in tuning control algorithms based on human preferences as MPC controller calibration for line keeping, obstacle avoidance, continuous stirring tank reactor [36], [37], and path-based velocity planner for tuning material deposition in [38], [39].

Therefore, the main contribution of this work is presenting a method for tuning a pHRI controller based on the preferences of different subjects. Because different tasks may require different tuning, it is important that such a procedure is fast and easy. Moreover, interesting results can be observed as general human preferences. We tested our methodology on two different tasks, one requiring precise path following and the other requiring a fast and large motion toward a target position. The data are analyzed to see general human behavior and preferences. Finally, a questionnaire is proposed to the users to check this method's applicability in real environments, involving time required for tuning and satisfaction.

II. METHOD

This section describes the system modeled as a Cartesian Impedance Control, then formulates the problem of pHRI with an IC as a Cooperative Game. These first two sections briefly recall the modeling presented in [25], [40], [41]. Please refer to these works for a more detailed explanation. Finally, presents the PBO method for tuning the relevant parameters of the control framework.

A. System modeling

As typically happens in manual guidance (MG) applications, the robot end-effector motion is described as a Cartesian Impedance system, as Cartesian space is more natural for the human operator:

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

where M_i and $D_i \in \mathbb{R}^{6\times6}$ are the desired inertia and damping matrices (note that typically in MG there is no stiffness matrix). $u_h \in \mathbb{R}^6$ and $u_r \in \mathbb{R}^6$ represent the measured human and computed robot effort applied to the system. x is defined as in [42], 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. Considering $\Omega = [\dot{p}^T\ \omega^T]^T$ the twist with $\dot{p}$ the linear velocity, ω angular velocity, it is possible to write $\dot{x} = J_\omega \Omega$ where J_ω is the transformation matrix between $\dot{x}$ and Ω. Finally, (1) can be linearized in the state-space formulation around the working point[1] :

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

where $z = [\Delta x^T\ \dot{x}^T]^T \in \mathbb{R}^{12}$ is the state space vector, $A = \begin{bmatrix} 0^{6\times6} & J_\omega \\ 0^{6\times6} & -M_i^{-1} D_i \end{bmatrix}$, $B_h^{12\times6} = B_r^{12\times6} = \begin{bmatrix} 0^{6\times6} \\ M_i^{-1} \end{bmatrix}$, with $0^{6\times6}$ denoting a 6×6 zero matrix.

[1] the non-linearity is due to rotations, this work considers small rotations, as pHRI applications (*e.g.*, transportation of large/heavy objects) typically involve large translations and small rotations

A kinematic inversion allows converting reference velocities from Cartesian to joint space:

$$\dot{q}_{ref}(t) = J(q)^+ \dot{x}(t), \qquad \text{with} \quad \dot{q}_{ref}(t) \in \mathbb{R}^n \tag{3}$$

where n is the number of joints and $J(q)^+$ is the pseudo-inverse of the geometric Jacobian. Simple integration allows commanding joint positions instead of velocities to the robot.

B. Differential Cooperative Game Theoretic modeling

(1) can be rewritten as (2) to include it in a Differential Cooperative Game Theory (DCGT) framework, leading to

$$\dot{z} = Az + Bu \tag{4}$$

with $B \in \mathbb{R}^{12\times12} = [B_h\ B_r]$ and $u = [u_h\ u_r]^T \in \mathbb{R}^{12\times1}$.

The two agents' objectives can be modeled as the minimization of a quadratic cost function

$$J_i = \int_0^\infty \left([z - z_{ref}]^T Q_i [z - z_{ref}] + u_i^T R_i u_i \right) dt\,,\ \ i = \{h, r\} \tag{5}$$

with subscripts $i = \{h, r\}$ indicating the human and the robot respectively. z_{ref} is the reference pose, J_i is the cost that the human and the robot incur, $Q_i \in \mathbb{R}^{6\times6}$ are the matrices that weight the state and references and $R_i \in \mathbb{R}^{6\times6}$ weight the control inputs. Note that in (5) Q_i and R_i are parameters that, in the robot case, can be chosen according to desired performances, while in the human case, can only be recovered by IOC techniques [43]. In this work, performances are measured and evaluated directly by the human user. Therefore, tuning of R_r is an optimization parameter in the PBO[2].

The CDGT problem consists in the minimization of (5) subject to (4). Its solution can be found by defining a weighted cost of the different player's cost function, which turns it into a Differential Linear Quadratic problem. The cost that the two players minimize together becomes:

$$J_{gt} = \alpha J_h + (1-\alpha) J_r = \int_0^\infty \left(\tilde{z}^T Q_{gt} \tilde{z} + u^T R_{gt} u \right) dt \tag{6}$$

with $u = [u_h^T,\ u_r^T]^T$ and $\tilde{z} = z - z_{ref}$. where $Q_{gt} = \alpha \hat{Q}_h + (1-\alpha) Q_r$ and $R_{gt} = diag([\alpha \hat{R}_h, (1-\alpha) R_r])$, and $\alpha \in (0,1)$ represents the weight of each player's cost in the overall cost. The minimization of (6) has infinite solutions depending on the choice of α, all lying on the Pareto frontier. This opens a new problem called the Bargaining theory. In this work, the choice of α represents a parameter to be optimized. The LQ-CGT can be seen now as a classical LQR problem:

$$\begin{aligned} \min_u \quad & J_{gt} = \int_0^\infty \left(\tilde{z}^T Q_{gt} \tilde{z} + u^T R_{gt} u \right) dt \\ s.t. \quad & \dot{\tilde{z}} = A \tilde{z} + Bu \quad , \quad \tilde{z}(t_0) = \tilde{z}_0 \end{aligned}, \tag{7}$$

[2] Note that keeping Q_r fixed and tuning only R_r is sufficient. Indeed it is only the ratio between parameters in Q and in R that matters, as in optimization problems $min(J) = min(\lambda J)$ with λ scalar.

The Linear Quadratic CGT finds the input u as linear feedback on the state that minimizes (6) subject to dynamical constraints. Given (4), the control action u is finally computed as:

$$u = -K_{gt}\,\tilde{z} = -K_{gt}\,z + K_{gt}\,z_{ref} \tag{8}$$

with $K_{gt} = R_{gt}^{-1} B^T P$, and P is the solutions of the Algebraic Riccati Equation (ARE)

$$0 = A^T P + PA^T - PBR_{gt}^{-1} - B^T P + Q_{gt} \tag{9}$$

Finally, note that the control vector from (8) is composed as $u = [u_h,\ u_r]^T$. Because the human exerted force is measured, the value of u_h computed in (8) is substituted by $\bar{u_h}$, where the $(\bar{\bullet})$ defines a measured signal. Summarizing, this control scheme models the interaction as a cooperative game and the human as a rational player in order to provide the robot with the corresponding control input.

C. Human reference estimation

In hand-guiding applications, the reference trajectory is typically not known a priori, and the robot must follow as well as possible human desired reference. So the problem of estimating the human reference arises. In this work, We decided to implement an easy yet powerful method based on the interaction force. The human reference pose is updated at each cycle by the following:

$$x_{ref}^{+} = x_{ref}^{-} + K_{p,h}\,u_h \tag{10}$$

with superscripts $^+$ and $^-$ referring to the updated and previous poses, respectively, and $K_{p,h}$ defines a coefficient proportional to the human exerted force. $K_{p,h}$ is an arbitrary parameter that can be tuned based on human preference.

D. Preference-based optimization

The PBO algorithm employed in this paper is based on the methodology developed in [35] by some of the authors, where the GLISp algorithm is introduced. In the following, the algorithm is briefly recalled. Please refer to the original paper for full treatment.

E. Building a surrogate function from preferences

The objective of the GLISp algorithm is to learn and minimize a surrogate function $\hat{J} : \mathbb{R}^{n_\theta} \to \mathbb{R}$ of an (unknown) underlying performance index J based on the observed preferences. Given a set of parameters $\Theta = \{\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_N}\}$ The surrogate function $\hat{J}$ is parametrized as a linear combination of Radial Basis Functions (RBFs):

$$\hat{J}(\boldsymbol{\theta}) = \sum_{k=1}^{N} \beta_k \phi(\gamma d(\boldsymbol{\theta}, \boldsymbol{\theta}_k)), \tag{11}$$

where $d : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_\theta} \to \mathbb{R}$ is the squared Euclidean distance $d(\boldsymbol{\theta}, \boldsymbol{\theta}_i) = \|\boldsymbol{\theta} - \boldsymbol{\theta}_i\|_2^2$, $\gamma > 0$ is a scalar parameter, $\phi : \mathbb{R} \to \mathbb{R}$ is an RBF, and $\boldsymbol{\beta} = [\beta_1\ \ldots\ \beta_N]^T$ are the unknown coefficients to be computed based on the available user's preferences. In this work, the Gaussian RBF $\phi(\gamma d) = e^{-(\gamma d)^2}$ is used, for more examples and explanations of RBFs please refer to [44].

Given two sets of parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ with $i \neq j$, the *preference function* $\pi : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_\theta} \to \{-1, 0, 1\}$ is defined as:

$$\pi(\theta_i, \theta_j) = \begin{cases} -1 & \text{if } \boldsymbol{\theta}_i \text{ "better" than } \boldsymbol{\theta}_j \\ 0 & \text{if } \boldsymbol{\theta}_i \text{ "as good as" } \boldsymbol{\theta}_j \\ 1 & \text{if } \boldsymbol{\theta}_i \text{ "worst" than } \boldsymbol{\theta}_j. \end{cases} \tag{12}$$

Therefore, the surrogate function $\hat{J}$ has to satisfy the following constraints:

$$\begin{array}{ll} \hat{J}(\boldsymbol{\theta}_{i(h)}) \le \hat{J}(\boldsymbol{\theta}_{j(h)}) - \sigma + \varepsilon_h & \text{if } \pi(\boldsymbol{\theta}_{i(h)}, \boldsymbol{\theta}_{j(h)}) = -1 \\ \hat{J}(\boldsymbol{\theta}_{i(h)}) \ge \hat{J}(\boldsymbol{\theta}_{j(h)}) + \sigma - \varepsilon_h & \text{if } \pi(\boldsymbol{\theta}_{i(h)}, \boldsymbol{\theta}_{j(h)}) = 1 \\ |\hat{J}(\boldsymbol{\theta}_{i(h)}) - \hat{J}(\boldsymbol{\theta}_{j(h)})| \le \sigma + \varepsilon_h & \text{if } \pi(\boldsymbol{\theta}_{i(h)}, \boldsymbol{\theta}_{j(h)}) = 0 \end{array} \tag{13}$$

for all $h = 1, \ldots, M$ with M number of expressed preferences, where $\sigma > 0$ is a tolerance, and ε_h is a positive slack variable which is used to relax the preference constraints.

Defining with $\Delta\hat{J}_h = \hat{J}(\boldsymbol{\theta}_{i,h}) - \hat{J}(\boldsymbol{\theta}_{j,h}) = \phi\gamma \sum_{k=1}^{N} (d(\boldsymbol{\theta}_{i(h)}, \boldsymbol{\theta}_k) - (d(\boldsymbol{\theta}_{j(h)}, \boldsymbol{\theta}_k)\beta_k$, the coefficient vector $\boldsymbol{\beta}$ describing the surrogate $\hat{J}$ is the solution of the Quadratic Programming (QP) problem, constrained by (13):

$$\begin{array}{rl} \min_{\beta,\varepsilon} & \sum_{h=1}^{M} \varepsilon_h + \frac{\lambda}{2} \sum_{k=1}^{N} \beta_k^2 \\ s.t. & \Delta\hat{J}_h \le -\sigma + \varepsilon_h, \forall h:\ b_h = -1 \\ & \Delta\hat{J}_h \ge \sigma - \varepsilon_h, \forall h:\ b_h = 1 \\ & \left|\Delta\hat{J}_h\right| \le \sigma + \varepsilon_h, \forall h:\ b_h = 0 \\ & h = 1, \ldots, M. \end{array} \tag{14}$$

with $b_h = \pi(\boldsymbol{\theta}_{i(h)}, \boldsymbol{\theta}_{j(h)})$. The scalar $\lambda > 0$ in the cost function (14) is a regularization parameter that guarantees uniqueness in the solution of the QP problem.

F. Acquisition function

To guarantee a tradeoff between *exploration* and *exploitation*, to generate a new set of parameters, an *acquisition function* $a(\boldsymbol{\theta})$ can be defined such that

$$\boldsymbol{\theta}_{N+1} = \arg\min_{\boldsymbol{\theta}\in\Theta} a(\boldsymbol{\theta}). \tag{15}$$

Let's define an *exploration function* as the Inverse Distance Weighting (IDW)

$$z(\boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \boldsymbol{\theta} \in \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\} \\ \tan^{-1}\left(\frac{1}{\sum_{i=1}^{N} w_i(\boldsymbol{\theta})}\right) & \text{otherwise,} \end{cases} \tag{16}$$

where $w_i(\boldsymbol{\theta}) = \frac{1}{d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_i)}$.

The *acquisition function* $a : \mathbb{R}^{n_\theta} \to \mathbb{R}$ is constructed as:

$$a(\boldsymbol{\theta}) = \frac{\hat{J}(\boldsymbol{\theta})}{\Delta\hat{J}} - \delta z(\boldsymbol{\theta}), \tag{17}$$

where $\delta \ge 1$ is an arbitrary exploration parameter and $\Delta\hat{J} = \max_i\{\hat{J}(\boldsymbol{\theta}_i)\} - \min_i\{\hat{J}(\boldsymbol{\theta}_i)\}$ is the range of the surrogate function on the samples in Θ and used in (17) as a normalization factor to simplify the choice of the exploration parameter δ.

After an initialization phase, the following steps are iterated for the optimization:

i) generate a new sample by (15),
ii) ask the user to express a preference $\pi(\boldsymbol{\theta}_{N+1}, \boldsymbol{\theta}_N^\star)$;
iii) update the estimate of $\hat{J}$ through (14);
iv) iterate over N.

G. Parameter for optimization

To conclude, in this work We are seeking the optimization of the robot's weight on the control action (*i.e.* R_r), the weighting factor of the cost functions (*i.e.* α), the proportional gain $K_{p,h}$ of eq. (10).

The R_r is chosen because it affects the cost the robot has on the control it can provide. As a consequence, the robot's behavior is more reactive and can put more effort into the task to track x_{ref}. This is the only parameter of the robot's cost function because it is assumed that the Q_r is always 1 on the positions and about 0 on the velocities, null all the rest. In this way, there is a strict relation between Q_r and R_r, in the sense that R_r is always a fraction of the Q_h. Note that this does not change the outcome of the optimization as $min(J) = min(\lambda J)\ \forall\ \lambda > 0$.

The parameter α represents the solution to the Bargaining problem. It directly modifies the contribution of the robot into the global cost function (6) and consequently the global optimization (7). In general, high values of α represents more robot contribution and in this work, more robot assistance.

Finally, the parameter $K_{p,h}$ modifies the set-point. At each control step, the set-point is updated according to (10). It is clear that high values of $K_{p,h}$ make the updated set-point far from the current one, while low values make it close. This means that at each control cycle, the farther the set point is, the higher the control input required to reach it is.

III. Experimental Validation

To evaluate the proposed method, We asked 5 healthy subjects, aged between 27 and 36 years old, to perform two sets of experiments. The subjects have different knowledge and experience in the pHRI field, from zero experience to some. We defined some numerical indices and a questionnaire to evaluate objective and subjective performances.

A. Design of Experiment

Two sets of experiments are proposed to evaluate the procedure in two different scenarios. The first experiment wants to evaluate the preferences of humans in performing a precise task, as a precise path following in the x–y plane. Subjects are asked to track the path visible in Fig. 1. This task is selected to mimic some typical industrial applications such as painting, cutting, material deposition, or even teaching-by-demonstration in hand-guiding applications, where the human drives the robot towards a path. In these cases, the robot can possibly hold a heavy tool relieving the operator from the weight, and the human just has to focus on the guidance asppect. The second task is a reaching task to evaluate the preferences of humans in performing relatively fast motions without the necessity to be extremely precise. Subjects are asked to reach a set point at 650 mm, in the x–y plane. This task is selected to mimic the approach phase to a target position. In general, these two tasks are selected to mimic a scenario where the human moves the robot close to a target position, then performs a precise task. It is expected that different preferences are assigned to different tasks with different goals.

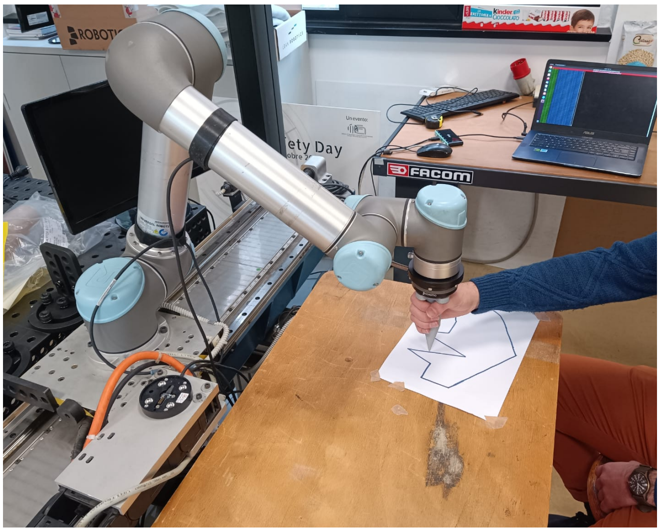

Fig. 1: A subject performing the path following task. The complete setup is visible.

For both cases, the human's cost function is composed of $Q_h = diag([1, 1, 0.0001, 0.0001])$ and $R_h = diag([0.0005, 0.0005])$, recovered via Inverse Optimal Control experimentally in a previous work [45]. The impedance parameters are $M_i = 10$ and $D_i = 50$. The only parameters related to the robot which are not optimized during the procedure are the components of the Q_r in the robot's cost function, which are set $Q_r = diag([1, 1, 0.0001, 0.0001])$. The robot used for the experiments is a Universal Robot 5 controlled in velocity at 125 Hz. The human interactive force is measured by a Robotiq FT300 force sensor mounted at the robot's end-effector.

The subjects are asked to perform the two tasks as well as possible and are asked to give a preference based on the satisfaction of performing the overall task. In the first case, the subjects are asked to evaluate their performance in the path following, along with the stress (physical and mental) required to perform it. In the second case, subjects are asked to evaluate the support that the robot is able to provide to reach the target pose quickly and with low effort.

The preferences are expressed, for both experiments, in two steps. An initial step is necessary to initialize the GLISp. Four experiments with different parameter sets are performed, and the user is asked to rank these experiments. After that, the best candidate solution found is compared with a new set, and the user is asked to express a preference. Based on the preference, the candidate solution is either updated or kept, and the process is iterated until convergence. We consider convergence when (i) three times in a row the preferences expressed are "as good as", (ii) three times in a row the new set of parameters is very similar to the ones suggested previously and discarded, (iii) a maximum number of iteration is reached.

B. Evaluation criteria

We used some performance indexes to evaluate the final results and to analyze relevant trends.

For the first case only, the Dynamic Time Warping (DTW) method [46] is used to compare the trajectories with the nominal one, to evaluate how well the preference expressed by the human is able to track it. This is not applicable to the reaching case because no predefined trajectory is used, and the subjects are asked to reach approximately a final position.

For both cases, it is possible to measure the force required to complete the task, which is computed as the Root Mean Square (RMS) of the measured force as follows:

$$f_{RMS} = \sqrt{\frac{1}{t_{end} - t_0} \int_{t_0}^{t_{end}} f(t)^2 \, dt} \tag{18}$$

We measure the f_{RMS} to evaluate how much effort humans like to put into a task. Indirectly, We want to analyze humans' preference for hard or soft haptic feedback in manual guidance applications.

Finally, a quick questionnaire is proposed to the users to evaluate their feelings on (i) how long was the tuning procedure, (ii) how tiring it was, and (iii) how much satisfied they are with the optimized parameters set. Users are asked to rate on an interval [0,4], with 0 meaning "no long at all", "no tiring at all", and "not satisfied at all", respectively. The questionnaire is asked after each optimization (path following and reaching) to check if there are differences.

IV. Results

Figure 2 shows the nominal trajectory (red-dashed), the trajectory with the best DTW (solid blue), and all the other trajectories performed by a subject with different parameters during optimization.

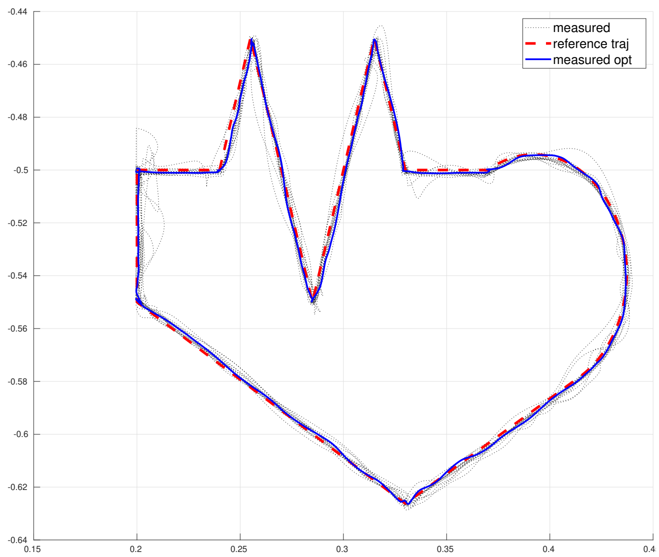

Fig. 2: Trajectories comparison relative to subject 3. Dashed red is the reference trajectory, solid blue is the measured trajectory with optimized parameters, and dotted black are all the other trajectories.

We evaluate the DTW to check if the perception of the human on the best task is also confirmed numerically. In figure 3, it is visible for the five subjects the range of DTW measured with different parameters (for subjects 2 and 4, the range is cut because a trial was close to instability and too much high DTW is scored). In red is highlighted the execution corresponding to the preferred by the human, and in three subjects coincide with the best measured. This study is useful to check that the preferences expressed by the users can also be measured and verified so that the subjective PBO, in this case, corresponds to objective better results.

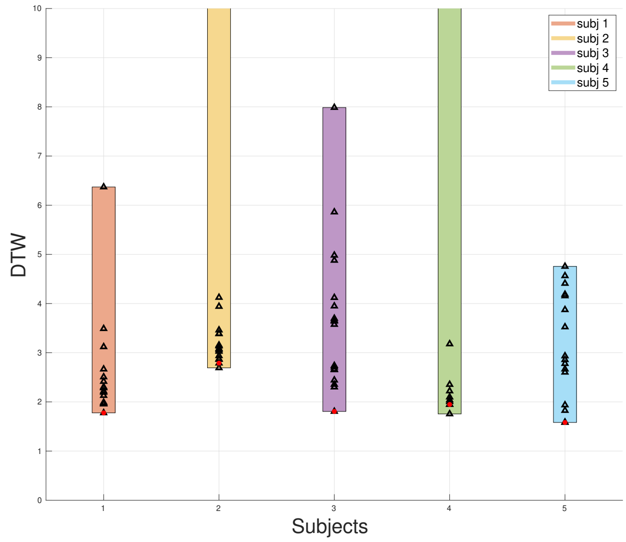

Fig. 3: The DTW relative to various repetitions of the task with different parameters. In red is highlighted the preference expressed by the user. In three cases, the preference expressed coincides with the best result, while in the other two, the preference expressed is anyway very close.

A parameter typically used to show the goodness of a proposed pHRI controller, a measurement of the force exchanged is given, assuming that the lower is the best. So, based on the preferences expressed, We want to see if this sentence is true. The f_{RMS} is visible in figure 4, for both tasks. It is, in general, true that low forces correspond to more pleasant interaction. Despite this, in a couple of cases, this is not true. This happens in particular for the reaching task. It seems that humans prefer to reach the set point quickly rather than save effort. A similar result was also recovered in [45]

Finally, the preferred parameters set are presented in figure 5 for both tasks for the three subjects. In general, high values of R are preferred for the path-following task, while low/very low values are preferred for the reaching task. Recalling that the parameter R is responsible for weighting the robot's effort in the cost function, high values of R mean low effort put into the task by the robot. So it makes sense that humans prefer high assistance for large, fast movements and very little assistance for small and precise motions. High assistance indeed means faster and easier reaching of the set point while can introduce oscillation in the case of small, precise motions.

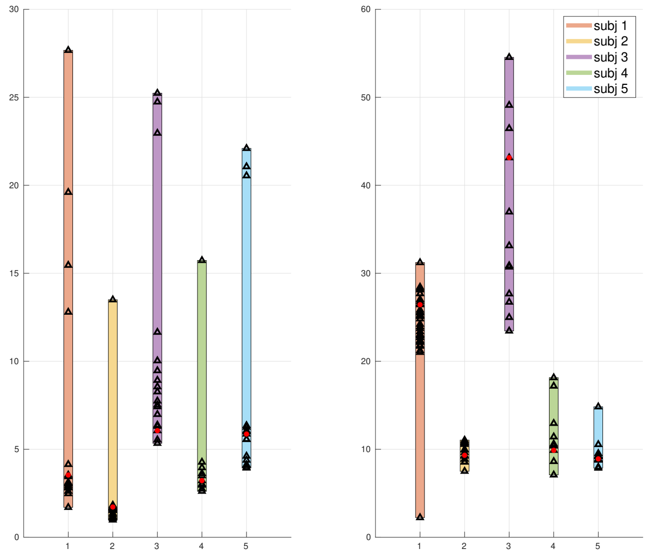

Fig. 4: Forces measured. Comparison between the range tested and the preferred one. Not in all cases, the minimum force means the human preference (indicated in red).

An opposite behavior is observed for the parameter α, which is, in general, preferred low for precise path following and high for fast motions. The reason is the same as for the R. High values of alpha mean, in the end, higher assistance.

Finally, for the $K_{h,p}$ parameter, low values are preferred for both tasks. This is due to the fact that too high values lead to jerky and unstable behavior since the set point is put too far and always overshoots the real desired target pose.

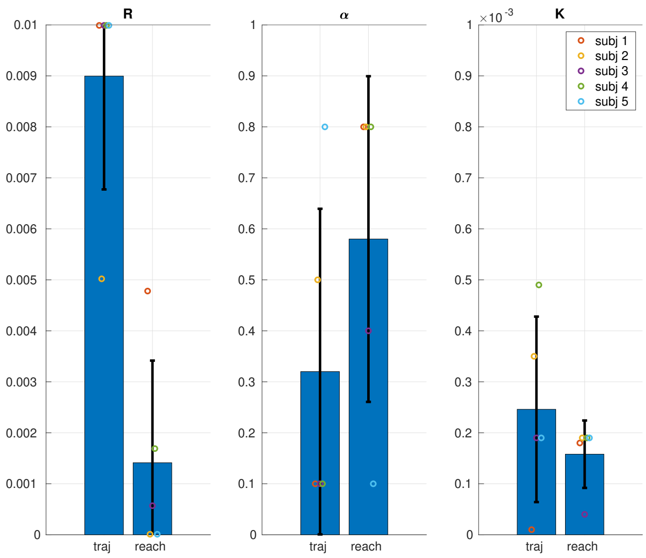

Fig. 5: Mean optimal values found for the three parameters. On the left are the parameters related to the path-following task, on the right to the reaching task

In conclusion, the questionnaire results are presented in table I for the two tasks. On average, users are pretty satisfied with the results obtained with the PBO procedure for both cases. Users also found the procedure not so long and not so tiring for both tasks. The only slight variation observed is in the length of the procedure, which was found to be a little shorter for the reaching case. This can be explained easily because the time required to perform a single reaching task was way shorter than a complete path-following task.

	Long	Tiring	Satisfied
Path following	1.25	1.00	2.75
Reaching	0.75	1.00	2.75

TABLE I: Questionnaire evaluation results.

V. Conclusion

This work presents a method for tuning an assistive game-theoretical cooperative controller for pHRI. The GLISp algorithm is used to perform PBO for two different sets of tasks. Results show that the procedure is appreciated by the users and is not tiring. Moreover, it is shown that subjective human preferences are, in general, coherent with numerical performance indexes. It was also shown that for each subject, different sets of parameters are preferred for the different tasks, while similar results are shared between the same task. This suggests that it is possible to extract general preferences for tuning such controllers based on the task. To make this approach more general, online variations of such parameters will be considered in future works for different tasks. For instance, a complete task may require fast motion to a target pose and then precise path following, so parameters can be changed smoothly on-the-fly. In that case, optimization of the transition function will be investigated.

References

[1] 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.

[2] V. Girbés-Juan, V. Schettino, Y. Demiris, and J. Tornero, "Haptic and visual feedback assistance for dual-arm robot teleoperation in surface conditioning tasks," *IEEE Transactions on Haptics*, vol. 14, no. 1, pp. 44–56, 2021.

[3] L. Roveda, N. Castaman, S. Ghidoni, P. Franceschi, N. Boscolo, E. Pagello, and N. Pedrocchi, "Human-robot cooperative interaction control for the installation of heavy and bulky components," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 339–344.

[4] A. Mohebbi, "Human-robot interaction in rehabilitation and assistance: a review," *Current Robotics Reports*, vol. 1, no. 3, pp. 131–144, 2020.

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

[6] N. Hogan, "Impedance Control: An Approach to Manipulation: Part I—Theory," *J of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 1–7, 03 1985.

[7] J. Dong, J. Xu, Q. Zhou, and S. Hu, "Physical human–robot interaction force control method based on adaptive variable impedance," *J of the Franklin Institute*, vol. 357, no. 12, pp. 7864–7878, 2020.

[8] F. Ficuciello, L. Villani, and B. Siciliano, "Variable impedance control of redundant manipulators for intuitive human–robot physical interaction," *IEEE Tran on Robotics*, vol. 31, no. 4, pp. 850–863, 2015.

[9] D. P. Losey and M. K. O'Malley, "Trajectory deformations from physical human–robot interaction," *IEEE Tran on Robotics*, vol. 34, no. 1, pp. 126–138, 2018.

[10] S. S. Ge, Y. Li, and H. He, "Neural-network-based human intention estimation for physical human-robot interaction," in *2011 8th Int Conf on Ubiquitous Robots and Ambient Intelligence*, 2011, pp. 390–395.

[11] L. Roveda, S. Haghshenas, M. Caimmi, N. Pedrocchi, and L. Molinari Tosatti, "Assisting operators in heavy industrial tasks: On the design of an optimized cooperative impedance fuzzy-controller with embedded safety rules," *Frontiers in Robotics and AI*, vol. 6, p. 75, 2019.

[12] L. Roveda, "A user-intention based adaptive manual guidance with force-tracking capabilities applied to walk-through programming for industrial robots," in *2018 15th International Conference on Ubiquitous Robots (UR)*. IEEE, 2018, pp. 369–376.

[13] L. Roveda, J. Maskani, P. Franceschi, A. Abdi, F. Braghin, L. Molinari Tosatti, and N. Pedrocchi, "Model-based reinforcement learning variable impedance control for human-robot collaboration," *J of Intelligent & Robotic Systems*, vol. 100, no. 2, pp. 417–433, Nov 2020.

[14] L. Roveda, S. Haghshenas, A. Prini, T. Dinon, N. Pedrocchi, F. Braghin, and L. M. Tosatti, "Fuzzy impedance control for enhancing capabilities of humans in onerous tasks execution," in *2018 15th Int Conf on Ubiquitous Robots (UR)*, 2018, pp. 406–411.

[15] L. Roveda, A. Testa, A. A. Shahid, F. Braghin, and D. Piga, "Q-learning-based model predictive variable impedance control for physical human-robot collaboration," *Art Intell*, vol. 312, p. 32, 2022.

[16] Y. Li and S. S. Ge, "Human–robot collaboration based on motion intention estimation," *IEEE-ASME Trans. Mechatron.*, vol. 19, no. 3, pp. 1007–1014, 2014.

[17] 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.

[18] 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 Tran on Robotics*, vol. 31, no. 3, pp. 672–681, 2015.

[19] W. Bi, X. Wu, Y. Liu, and Z. Li, "Role adaptation and force, impedance learning for physical human-robot interaction," in *2019 IEEE 4th Int Conf on Advanced Robotics and Mechatronics*, 2019, pp. 111–117.

[20] Y. Li, K. P. Tee, R. Yan, W. L. Chan, Y. Wu, and D. K. Limbu, "Adaptive optimal control for coordination in physical human-robot interaction," in *2015 IEEE/RSJ Int Conf on Intelligent Robots and Systems (IROS)*, 2015, pp. 20–25.

[21] 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 Tran on Robotics*, vol. 32, no. 6, pp. 1408–1418, 2016.

[22] 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.

[23] R. Zou, Y. Liu, J. Zhao, and H. Cai, "A framework for human-robot-human physical interaction based on n-player game theory," *Sensors*, vol. 20, no. 17, 2020.

[24] S. Musić and S. Hirche, "Haptic shared control for human-robot collaboration: A game-theoretical approach," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 10 216–10 222, 2020, 21st IFAC World Congress.

[25] P. Franceschi, N. Pedrocchi, and M. Beschi, "Adaptive impedance controller for human-robot arbitration based on cooperative differential game theory," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 7881–7887.

[26] R. Ikeura, T. Moriguchi, and K. Mizutani, "Optimal variable impedance control for a robot and its application to lifting an object with a human," in *Proceedings. 11th IEEE International Workshop on Robot and Human Interactive Communication*, 2002, pp. 500–505.

[27] A. Lecours, B. Mayer-St-Onge, and C. Gosselin, "Variable admittance control of a four-degree-of-freedom intelligent assist device," in *2012 IEEE International Conference on Robotics and Automation*, 2012, pp. 3903–3908.

[28] V. Duchaine, B. Mayer St-Onge, D. Gao, and C. Gosselin, "Stable and intuitive control of an intelligent assist device," *IEEE Transactions on Haptics*, vol. 5, no. 2, pp. 148–159, 2012.

[29] A. Alili, V. Nalam, M. Li, M. Liu, J. Si, and H. H. Huang, "User controlled interface for tuning robotic knee prosthesis," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 6190–6195.

[30] A. Alili, V. Nalam, M. Li, M. Liu, J. Feng, J. Si, and H. Huang, "A novel framework to facilitate user preferred tuning for a robotic knee prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 895–903, 2023.

[31] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015, pp. 51–58.

[32] G. Hoffman, "Evaluating fluency in human–robot collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, 2019.

[33] J. Schmidtler, K. Bengler, F. Dimeas, and A. Campeau-Lecours, "A questionnaire for the evaluation of physical assistive devices (quead): Testing usability and acceptance in physical human-robot interaction," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2017, pp. 876–881.

[34] F. Dehais, E. A. Sisbot, R. Alami, and M. Causse, "Physiological and subjective evaluation of a human–robot object hand-over task," *Applied Ergonomics*, vol. 42, no. 6, pp. 785–791, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687011000044

[35] A. Bemporad and D. Piga, "Global optimization based on active preference learning with radial basis functions," *Machine Learning*, vol. 110, no. 2, pp. 417–448, 2021.

[36] M. Zhu, D. Piga, and A. Bemporad, "C-glisp: Preference-based global optimization under unknown constraints with applications to controller calibration," *IEEE Transactions on Control Systems Technology*, vol. 30, no. 5, pp. 2176–2187, 2022.

[37] M. Zhu, A. Bemporad, and D. Piga, "Preference-based mpc calibration," in *2021 European Control Conference (ECC)*, 2021, pp. 638–645.

[38] L. Roveda, B. Maggioni, E. Marescotti, A. A. Shahid, A. Maria Zanchettin, A. Bemporad, and D. Piga, "Pairwise preferences-based optimization of a path-based velocity planner in robotic sealing tasks," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6632–6639, 2021.

[39] L. Roveda, P. Veerappan, M. Maccarini, G. Bucca, A. Ajoudani, and D. Piga, "A human-centric framework for robotic task learning and optimization," *Journal of Manufacturing Systems*, vol. 67, pp. 68–79, 2023.

[40] J. Engwerda, *LQ dynamic optimization and differential games*. John Wiley & Sons, 2005.

[41] ——, "The regular convex cooperative linear quadratic control problem," *Automatica*, vol. 44, no. 9, pp. 2453–2457, 2008.

[42] B. Siciliano and L. Villani, *Robot Force Control*. Springer US, 1999.

[43] 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.

[44] A. Bemporad, "Global optimization via inverse distance weighting and radial basis functions," *arxiv*, vol. 8, no. 1, pp. 1–18, 2020.

[45] P. Franceschi, N. Pedrocchi, and M. Beschi, "Inverse optimal control for the identification of human objective: a preparatory study for physical human-robot interaction," in *2022 IEEE ETFA*, 2022, pp. 1–6.

[46] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.