Adaptive Optimal Control for Coordination in Physical Human-Robot Interaction

Yanan Li, Keng Peng Tee, Rui Yan, Wei Liang Chan, Yan Wu and Dilip Kumar Limbu

Abstract— **In this paper, we propose an adaptive optimal control for a robot to collaborate with a human. Game theory and policy iteration are employed to analyze the interactive behaviors of the human and the robot in physical interactions. The human's control objective is estimated and it is used to adapt the robot's own objective, such that human-robot coordination can be achieved. An optimal control is developed to guarantee that the robot's control objective is realized. The validity of the proposed method is verified through rigorous analysis and experiment studies.**

I. INTRODUCTION

Physical human-robot interaction is an emerging research field due to the need of robotics in unstructured environments and ad-hoc human inaccessible tasks [1]. In general, humans and robots have complementary advantages: the former excel in reasoning and problem solving, while the latter are good in execution with a guaranteed performance [2], [3]. The combination of these advantages in a common task is found to be useful and in many applications necessary, such as tele-operation, co-assembly, co-transportation, etc.

Two critical issues need to be addressed in developing a natural and efficient human-robot interface. On the one hand, analysis of interactive behaviors of two agents is difficult, which can be very complex in different tasks and in different phases within a task. Abundant research effort has been made to address this issue, in the fields of multi-agent systems and distributed intelligence [4]. Most of works in this direction focus on robots themselves, instead of considering both humans and robots. On the other hand, human-in-the-loop robotic applications introduce inevitable problems of uncertainties and unobservable states, not mentioning the consideration of ergonomics and human factors [5]. Many solutions have also been proposed to cope with these problems in the literature, including intention recognition based on different cues, e.g., haptic and visual cues [6]. While how to address these two issues individually is still an open problem, a general framework is required to take both of them into account simultaneously. Therefore, adaptive frameworks/models for human-robot interaction have been proposed in recent studies [7], [8], beyond a simple yet robust passive leader-follower model [9]. These studies point out that the robot should play an adaptive role to lead a task or to follow based on the human's intention or a specific circumstance [10].

Despite the aforementioned research effort, there is a lack of works about the rigorous analysis of the interactive behavior under an adaptive framework/model. In this paper, we aim to achieve it by integrating game theory [11] and policy iteration [12]. Game theory has been shown to be suitable for analyzing the performance of multi-agent systems [13], according to which human-robot interaction is deemed as a two-agent game. In game theory, a variety of interactive behaviors can be described by individual objective/cost functions and optimization criteria. Given a static game with fixed objective/cost functions, a conventional method which solves a coupled Riccati equation can be used to obtain the optimal control [14]. In our previous work [15], we have developed an optimal control for human-robot collaboration based on this strategy. However, it only deals with the linear systems. For general nonlinear systems, policy iteration is needed to re-compute optimal control in real time [16]. Methods of policy iteration for games with known and unknown dynamics have been developed by several research groups [17], [18]. As mentioned above, however, the human's objective is generally unknown to the robot in a typical human-robot interaction scenario. Therefore, the aforementioned methods, which presume that both agents have perfect interpretation of their partner's behaviors, are not applicable.

In this work, we consider that the collaborative task is realized through physical human-robot interaction. The human's unknown control objective is estimated by developing an adaptive estimation method. Ultimately, we will show that the proposed estimation method can be integrated with policy iteration, such that the robot coordinates with the human and optimal control is achieved.

The rest of this paper is organized as follows. In Section II, the system under study is discussed and the problem is formulated. In Section III, policy iteration for a two-agent game is introduced and the proposed adaptive optimal control is detailed with performance analysis. In Section IV, the validity of the proposed method is verified through experimental studies. Section V concludes this work.

II. PROBLEM FORMULATION

In a typical scenario, a robot arm has a predefined task to work on a workpiece, while a human arm is physically in contact with the robot arm and directly applies a force to the end-effector of the robot arm. For convenience, "human arm" and "robot arm" are denoted as "human" and "robot" from now on. The force/torque applied by the human is referred to

*This work is supported by Grant No. 1225100001 from the Science and Engineering Research Council (SERC), A*STAR, Singapore.

Y. Li, K. P. Tee, R. Yan, W. L. Chan, Y. Wu, and D. K. Limbu are with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore 138632. {liy,kptee,ryan,chanwl,wuy,dklimbu}@i2r.a-star.edu.sg

as the "interaction force", and it is available for the control design.

In the field of physical human-robot interaction, impedance control is usually adopted [9] which is realized with an impedance model given in the Cartesian space:

$$M_d\ddot{x}(t) + C_d\dot{x}(t) = u(t) + f(t) \tag{1}$$

where $M_d \in \mathbb{R}^{m\times m}$ and $C_d \in \mathbb{R}^{m\times m}$ are desired inertial and damping matrices, respectively, $x(t) \in \mathbb{R}^m$ the position, $f(t) \in \mathbb{R}^m$ the interaction force and $u(t) \in \mathbb{R}^m$ the control input. For the feasibility of control design, Eq. (1) can be rewritten in the following state-space form

$$\begin{aligned}
\dot{z}'(t) &= A'z'(t) + B_1'u(t) + B_2'f(t) \qquad (2)\\
z'(t) &= \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix},\ A' = \begin{bmatrix} \mathbf{0}_m & I_m \\ \mathbf{0}_m & -M_d^{-1}C_d \end{bmatrix},\\
B_1' &= B_2' = \begin{bmatrix} \mathbf{0}_m \\ M_d^{-1} \end{bmatrix}
\end{aligned}$$

where $\mathbf{0}_m$ and I_m denote $m \times m$ zero and identity matrices, respectively. To take the trajectory tracking into account, we need to design an auxiliary system to generate the desired trajectory of the robot x_d, i.e., $\dot{x}_d = F(x_d)$ where $F(\cdot) \in \mathbb{R}^{m\times m}$ is a given matrix function. Then, with the augmented state $z = [z'^T \ x_d^T]^T$, we have

$$\begin{aligned}
\dot{z} &= A(z) + B_1(z)u + B_2(z)f \qquad (3)\\
A(z) &= \begin{bmatrix} A'(z')z' \\ F(x_d) \end{bmatrix},\ B_1(z) = B_2(z) = \begin{bmatrix} B'(z') \\ \mathbf{0}_m \end{bmatrix}
\end{aligned}$$

In game theory, the term "agent" is used to refer to the party involved in a common task, which is either the human or the robot in our context. Cost functions are usually defined to describe agents' interactive behaviors. While a more comprehensive list of these behaviors can be found in [11], we only introduce the relevant ones.

Definition 1: [11] Coordination is the most cohesive form of human-robot interaction in which that the human and the robot have a common cost function.

We assume that the human has the following cost function

$$\Gamma = \int_0^\infty c(t)dt \tag{4}$$

where the instantaneous cost is

$$\begin{aligned}
c(t) &= (x - x_d)^T Q_1 (x - x_d) + \dot{x}^T Q_2 \dot{x} + u^T R_1 u\\
&\quad + f^T R_2 f \qquad (5)
\end{aligned}$$

In the above instantaneous cost, $Q_1 \succeq 0$ and $Q_2 \succeq 0$ are the weights for trajectory tracking and velocity regulation, respectively, and $R_1 \succ 0$ and $R_2 \succ 0$ are the weights for the robot and the human controls, respectively. Considering the definition of z, Eq. (5) can be rewritten as

$$\begin{aligned}
c(t) &= z^T Q z + u^T R_1 u + f^T R_2 f \qquad (6)\\
Q &= \begin{bmatrix} Q_1 & \mathbf{0}_m & -Q_1 \\ \mathbf{0}_m & Q_2 & \mathbf{0}_m \\ -Q_1 & \mathbf{0}_m & Q_1 \end{bmatrix} \qquad (7)
\end{aligned}$$

If the cost function Γ in (4) is known, it can be used as the cost function of the robot. According to Definition 1, coordination can be achieved with a controller that is designed based on this cost function. Unfortunately, the cost function Γ may change over time and is unknown to the robot. Therefore, in the following section, we will develop a method to estimate Γ, and use the estimate as the cost function of the robot. We will show that optimal control can still be achieved with such an estimation process.

III. ADAPTIVE OPTIMAL CONTROL

A. Preliminary: Nash Equilibrium Solutions

Consider a general system described by

$$\dot{\xi} = h(\xi) + g_1(\xi)u_1 + g_2(\xi)u_2 \tag{8}$$

Denote the index of the agent as $i = 1, 2$, and define the following cost functions

$$\Gamma_i = \int_0^\infty c_i(t)dt \tag{9}$$

where the instantaneous cost functions are given by

$$c_i(t) = Q(\xi) + u_1^T R_{i1} u_1 + u_2^T R_{i2} u_2 \tag{10}$$

with $Q(\xi)$ being positive definite in ξ, $R_{i1} \succ 0$ and $R_{i2} \succ 0$. Correspondingly, the value functions in the development of the iteration policy algorithm are

$$V_i(\xi) = \int_t^\infty c_i(s)ds \tag{11}$$

To minimize cost functions Γ_i, different control policies u_i can be found under different definitions of equilibrium. In this paper, we focus on the Nash equilibrium.

Definition 2: [11] The Nash equilibrium policies u_1^* and u_2^* for the two-agent game under study satisfy the following two inequalities:

$$\begin{aligned}
\Gamma_1(\xi, u_1^*, u_2^*) &\le \Gamma_1(\xi, u_1, u_2^*)\\
\Gamma_2(\xi, u_1^*, u_2^*) &\le \Gamma_2(\xi, u_1^*, u_2) \qquad (12)
\end{aligned}$$

The policy iteration algorithm provides a solution to achieve the Nash equilibrium [12]. Based on it, a synchronous approximate optimal learning algorithm is developed in [18]. First, the optimal value functions are approximated as

$$V_i^*(\xi) = W_i^{*T} S_i(\xi) + \varepsilon_i(\xi) \tag{13}$$

where W_i^* are unknown ideal weight matrices, $S_i(\xi)$ activation functions and $\varepsilon_i(\xi)$ approximation errors. The optimal controls are given by

$$\begin{aligned}
u_i^*(\xi) &= -\frac{1}{2} R_{ii}^{-1} g_i^T(\xi) \nabla V_i^*\\
&= -\frac{1}{2} R_{ii}^{-1} g_i^T(\xi)(\nabla S_i^T(\xi) W_i^* + \nabla\varepsilon_i(\xi)) \ (14)
\end{aligned}$$

Denote the estimates of W_i^* as W_i, so the estimates of V_i^* are

$$V_i(\xi) = W_i^T S_i(\xi) \tag{15}$$

The update laws of W_i are given by

$$\dot{W}_i = -\alpha_i \sigma_i (W_i^T \sigma_i + c_i) \tag{16}$$

where $\alpha_i > 0$ is the learning rate, and $\sigma_i = \dot{S}_i = \nabla S_i \dot{\xi} = \nabla S_i (h(\xi) + g_1(\xi) u_1 + g_2(\xi) u_2)$. Then, the approximated controls are given by

$$u_i(\xi) = -\frac{1}{2} R_{ii}^{-1} g_i^T(\xi) \nabla S_i^T(\xi) W_i \tag{17}$$

Theorem 1: [18] Suppose that σ_i is persistently exiting (PE), then, the approximated controls $u_i(\xi)$ in Eq. (17) converge to the Nash equilibrium solutions, i.e., optimal controls u_i^*.

B. Control Design

Comparing Eqs. (3) and (8), the human-robot interaction system (3) can be described by the general model (8). Therefore, the Nash equilibrium can be achieved if robot control u and human control f follow controls (17) with update laws (16). Furthermore, coordination as defined in Definition 1 can be realized if the robot and the human have a common cost function Γ, which corresponds to the following value function:

$$V(z) = \int_t^{\infty} c(s) ds = W^T S(z) \tag{18}$$

As discussed in Section II, the problem is that Γ is unknown, and so are $c(t)$ and W. In this section, we propose a method to obtain their estimates, and then develop an optimal control to minimize them.

Denote the estimates of W and $c(t)$ as $\hat{W}$ and $\hat{c}$, respectively. In particular, we have

$$\hat{c} = z^T \hat{Q} z + u^T \hat{R}_1 u + f^T \hat{R}_2 f \tag{19}$$

where $\hat{Q}$, $\hat{R}_1$ and $\hat{R}_2$ are estimates of Q, R_1 and R_2, respectively. Therefore, according to approximated control (17) with update law (16), the following robot control is proposed:

$$u = -\frac{1}{2} \hat{R}_1^{-1} B_1^T \nabla S^T(z) \hat{W} \tag{20}$$

with the update law

$$\dot{\hat{W}} = -\alpha \sigma (\hat{W}^T \sigma + \hat{c}) \tag{21}$$

where $\alpha > 0$ is the learning rate to be designed later, and $\sigma = \nabla S(z)(A(z) + B_1 u + B_2 f)$. Note that $\hat{c}$ is unknown and will be estimated below.

Similarly, the estimated human control is

$$\hat{f} = -\frac{1}{2} \hat{R}_2^{-1} B_2^T \nabla S^T(z) \hat{W} \tag{22}$$

which is the estimate of the human control

$$f = -\frac{1}{2} R_2^{-1} B_2^T \nabla S^T(z) W \tag{23}$$

with W updated as below

$$\dot{W} = -\alpha \sigma (W^T \sigma + c) \tag{24}$$

Subtracting Eq. (21) by Eq. (24), we obtain

$$\dot{\tilde{W}} = -\alpha \sigma (\tilde{W}^T \sigma + \tilde{c}) \tag{25}$$

where $\tilde{W} = \hat{W} - W$ and $\tilde{c} = \hat{c} - c$. In particular, we have

$$\begin{aligned} \tilde{c} &= z^T \tilde{Q} z + u^T \tilde{R}_1 u + f^T \tilde{R}_2 f \qquad (26) \\ \tilde{Q} &= \hat{Q} - Q, \ \tilde{R}_1 = \hat{R}_1 - R_1, \ \tilde{R}_2 = \hat{R}_2 - R_2 \end{aligned}$$

For feasibility of analysis, we define

$$\tilde{\theta} = [\mathrm{vec}^T(\tilde{Q}) \ \mathrm{vec}^T(\tilde{R}_1) \ \mathrm{vec}^T(\tilde{R}_2)]^T \tag{27}$$

where $\mathrm{vec}(\cdot)$ is the column vectorization operator. Correspondingly, we denote

$$\begin{aligned} \theta &= [\mathrm{vec}^T(Q) \ \mathrm{vec}^T(R_1) \ \mathrm{vec}^T(R_2)]^T \\ \hat{\theta} &= [\mathrm{vec}^T(\hat{Q}) \ \mathrm{vec}^T(\hat{R}_1) \ \mathrm{vec}^T(\hat{R}_2)]^T \qquad (28) \end{aligned}$$

By denoting

$$\begin{aligned} &Y = [\bar{z}^T \ \bar{u}^T \ \bar{f}^T]^T \\ &\bar{z} = [z^2(1), z(1)z(2), \ldots, z(1)z(3m), z(2)z(1), z^2(2), \ldots, \\ &\quad z(2)z(3m), \ldots, z^2(3m)]^T \\ &\bar{u} = [u^2(1), u(1)u(2), \ldots, u(1)u(m), u(2)u(1), u^2(2), \ldots, \\ &\quad u(2)u(m), \ldots, u^2(m)]^T \\ &\bar{f} = [f^2(1), f(1)f(2), \ldots, f(1)f(m), f(2)f(1), f^2(2), \ldots, \\ &\quad f(2)f(m), \ldots, f^2(m)]^T \qquad (29) \end{aligned}$$

where $z(j)$, $u(j)$ and $f(j)$, $j = 1, 2, \ldots, 3m/m$, are elements of z, u and f, respectively, we obtain

$$\tilde{c} = \tilde{\theta}^T Y \tag{30}$$

Then, we can develop the following update law to obtain $\hat{c}$ in Eq. (21):

$$\begin{aligned} \dot{\hat{\theta}} &= -\beta \hat{\theta} - 2\alpha Y (\hat{R}_2 \tilde{f})^T (\nabla S(z) B_2)^{\dagger} \sigma \\ &= -\beta \hat{\theta} - 2\alpha Y (\hat{R}_2 \tilde{f})^T B_2^{\dagger} (A(z) + B_1 u + B_2 f) \quad (31) \end{aligned}$$

where $\tilde{f} = \hat{f} - f$ is the force error, β is a positive scalar, and "†" denotes the pseudo inverse of a matrix/vector.

Remark 1: Since we know that there are zero sub-matrices in Q (refer to Eq. (7)), the corresponding components in θ should also be zeros. It indicates that some coupling items, e.g., $z(1)\dot{z}(1)$, do not exist in the cost function. Thus, the corresponding components in θ and Y can be set to zeros, and only Q_1, Q_2, R_1 and R_2 need to be updated. The dimension of Y can be further reduced by ignoring the coupling effects between different directions for convenience of implementation, e.g., suppose R_1 is a diagonal matrix, then we do not take $u(i)u(j)$, $i \neq j$, into account for the computation of Y.

C. Performance Analysis

Theorem 2: Considering the robot dynamics governed by the impedance model (1), the proposed robot control u in Eq. (20) with the update law (21), and human control f in Eq. (23) with the update law (24) guarantee that $\tilde{W}$ and $\tilde{\theta}$ remain in a compact set $\Omega_1 :=$

$\left\{\tilde{W},\tilde{\theta}|\ \|[\tilde{W}^T,\tilde{\theta}^T]^T\|\leq\sqrt{\chi_1}\right\}$ and converge to another compact set $\Omega_2 := \left\{\tilde{W},\tilde{\theta}|\ \|[\tilde{W}^T,\tilde{\theta}^T]^T\|\leq\sqrt{\chi_2}\right\}$ where

$$\begin{aligned}\chi_1 &= 2(U(0)+\frac{b}{\kappa}),\ \chi_2=\frac{2b}{\kappa} \quad (32)\\ U(t) &= \frac{1}{2}\tilde{W}^T\tilde{W}+\frac{1}{2}\tilde{\theta}^T\tilde{\theta},\\ \kappa &= \min(\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y},\frac{\beta}{2}-\frac{\gamma\|W\|}{\|R_2\|}),\ b=\frac{\beta}{2}\theta^T\theta\end{aligned}$$

In above equations, $\underline{b}_\sigma$ is the lower bound of $\|\sigma\|$, $\gamma>0$, and α and β are designed as

$$\alpha=\frac{\gamma}{1+\bar{b}_\sigma\bar{b}_Y},\ \beta>\frac{2\gamma\|W\|}{\|R_2\|} \quad (33)$$

where $\bar{b}_\sigma$ and $\bar{b}_Y$ are upper bounds of $\|\sigma\|$ and $\|Y\|$, respectively.

Proof: According to Eqs. (25) and (30), we have

$$\begin{aligned}\dot{U}(t) &= \tilde{W}^T\dot{\tilde{W}}+\tilde{\theta}^T\dot{\tilde{\theta}}\\ &= -\alpha\sigma^T\sigma\tilde{W}^T\tilde{W}-\alpha\tilde{W}^T\sigma\tilde{\theta}^TY+\tilde{\theta}^T\dot{\hat{\theta}} \quad (34)\end{aligned}$$

where the fact $\dot{\tilde{\theta}}=\dot{\hat{\theta}}$ is used. Considering the design parameter α in (33), we obtain

$$\dot{U}(t)\leq-\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-\alpha\tilde{W}^T\sigma\tilde{\theta}^TY+\tilde{\theta}^T\dot{\hat{\theta}} \quad (35)$$

Substituting the update law (31) into Ineq. (35), we obtain

$$\begin{aligned}\dot{U}(t)\leq&-\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-\beta\tilde{\theta}^T\tilde{\theta}-2\alpha(\hat{R}_2\tilde{f})^T\\ &\times(\nabla S(z)B_2)^\dagger\sigma\tilde{\theta}^TY-\alpha\tilde{W}^T\sigma\tilde{\theta}^TY-\beta\tilde{\theta}^T\theta \quad (36)\end{aligned}$$

Considering Eqs. (22) and (23), we obtain

$$\tilde{W}^T\nabla S(z)B_2=-2(\hat{R}_2\tilde{f})^T-2(\tilde{R}_2f)^T \quad (37)$$

Substituting Eq. (37) to Ineq. (36), we have

$$\begin{aligned}\dot{U}(t) \leq& -\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-\beta\tilde{\theta}^T\tilde{\theta}-\alpha W^T(\nabla S(z)B_2)\\ &\times(\tilde{R}_2R_2^{-1})^T(\nabla S(z)B_2)^\dagger\sigma\tilde{\theta}^TY-\beta\tilde{\theta}^T\theta\\ \leq& -\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-\beta\tilde{\theta}^T\tilde{\theta}\\ &+\alpha\|\sigma\|\|Y\|\|\tilde{\theta}\|\|W\|\|\tilde{R}_2R_2^{-1}\|-\beta\tilde{\theta}^T\theta \quad (38)\end{aligned}$$

Considering that $\|\tilde{R}_2\|\leq\|\tilde{\theta}\|$ and α in (33), we obtain

$$\begin{aligned}\dot{U}(t)\leq&-\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-\beta\tilde{\theta}^T\tilde{\theta}+\frac{\gamma\|W\|}{\|R_2\|}\tilde{\theta}^T\tilde{\theta}-\beta\tilde{\theta}^T\theta\\ \leq&-\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-\beta\tilde{\theta}^T\tilde{\theta}\\ &+\frac{\gamma\|W\|}{\|R_2\|}\tilde{\theta}^T\tilde{\theta}+\frac{\beta}{2}\tilde{\theta}^T\tilde{\theta}+\frac{\beta}{2}\theta^T\theta\\ =&-\frac{\gamma\underline{b}_\sigma^2}{1+\bar{b}_\sigma\bar{b}_Y}\tilde{W}^T\tilde{W}-(\frac{\beta}{2}-\frac{\gamma\|W\|}{\|R_2\|})\tilde{\theta}^T\tilde{\theta}+\frac{\beta}{2}\theta^T\theta\end{aligned}$$

Considering definitions of κ and b in Eq. (33), we have

$$\dot{U}(t)\leq-\kappa U(t)+b \quad (39)$$

Multiplying the above inequality by $e^{\kappa t}$ leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^{\kappa t}U(t))\leq be^{\kappa t} \quad (40)$$

Integrating the above inequality, we have

$$U(t)\leq(U(0)-\frac{b}{\kappa})be^{-\kappa t}+\frac{b}{\kappa}\leq U(0)+\frac{b}{\kappa} \quad (41)$$

Therefore, $\tilde{W}$ and $\tilde{\theta}$ remain in a compact set Ω_1 and converge to another compact set Ω_2. It completes the proof.

Remark 2: In Theorem 2, sizes of compact sets Ω_1 and Ω_2 can be adjusted by choosing parameters α, β and γ. When these sizes are made small enough, we have a) $\hat{c}\approx c$ which indicates that the unknown instantaneous cost function of the human is observed and the coordination defined in Definition 1 realized, and b) $\hat{W}\approx W$ which indicates that optimal control for a 2-agent game with the Nash equilibrium is achieved, according to Theorem 1.

IV. EXPERIMENT

A. Settings

In this section, we consider a scenario of human-robot co-assembly. In this scenario, the robot has a prescribed trajectory to move workpieces from a position to another in sequence. In a normal case, the robot is able to finish the task alone in such a well-defined environment. However, if there exist uncertainties, e.g., the order of the workpiece is changed, the human needs to take an online corrective action to move the robot along a new trajectory to a new destination. Once the new task is finished and the human releases the robot, the robot is able to follow its prescribed trajectory again. Motivated by the above scenario, we design the experiment setup as shown in Fig. 1. A 7-DOFs KUKA lightweight robot (LBR) is used with impedance control mode, as discussed in Section II. The interaction force f is obtained based on the measured external torque at each joint. Before the experiment starts, the desired trajectory of the robot arm, i.e., x_d, is recorded by setting three position points in a planar space, i.e., P_0, P_1 and P_2, as shown in Figs. 2 and 3. A stiffness of 300N/m and a damping factor of 0.7 are used.

We employ a radial basis function neural network to implement the proposed control. In (18), $S(z)=[s_1,\ldots,s_l]^T$, where the number of the neural nodes is $l=10$. The activation function is chosen as the Gaussian function, i.e., $s_i(z)=\exp\left[\frac{-(z-\mu_i)^T(z-\mu_i)}{\eta_i^2}\right]$ with $\mu_i=[\mu_{i,1},\mu_{i,2},...,\mu_{i,6}]$ as the center of the receptive field, and η_i the width of the Gaussian function, where $i=1,\ldots,l$. Then, $\nabla S(z)=[\nabla s_1,\ldots,\nabla s_l]^T$ with $\nabla s_i(z)=-\frac{2}{\eta_i^2}s_i(z)(z-\mu_i)^T$. Choosing $\eta_i=\eta$ and $\mu_i=[x_d^T,(\dot{x}+k(x_d-x))^T,x_d^T]^T$, and substituting $\nabla S(z)$ into Eq. (20), we obtain

$$u = \frac{k}{\eta^2}S^T(z)\hat{W}\hat{R}_1^{-1}M_d^{-T}(x_d-x) \quad (42)$$

where $\eta=100$ and $k=10$. From the above equation, we find that u is a variable impedance control with stiffness $K=\frac{k}{\eta^2}S^T(z)\hat{W}\hat{R}_1^{-1}M_d^{-T}$, which is used in the impedance

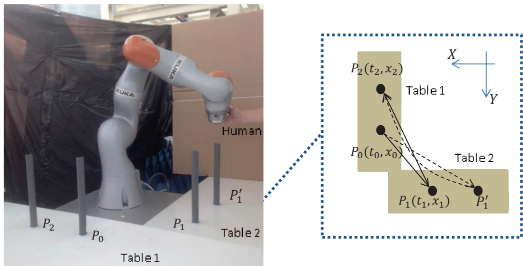

Fig. 1. The experiment setup. The diagram on the right hand side depicts the robot setup from the top-view. The solid arrow indicates the prescribed trajectory of the robot, while the dotted one indicates the trajectory refined by the human to handle the unexpected situation of a changed order. Four rods are fixed on two tables to indicate the four position points. i.e., P_0, P_1, P_1 and P_2. When the robot tries to move from P_0 to P_1, the human moves it to P_1' instead. After P_1' is reached, the human releases the robot and it moves to the next position point P_2.

control mode of LBR. Impedance parameters in Eq. (1) are $M_d = 3 \times 10^{-3} I_{2\times2}$ and $C_d = 10 I_{2\times2}$. The initial values of the estimated weights in Eq. (19) are $\hat{Q}_1 = 100 I_2$, $\hat{Q}_2 = I_2$, $\hat{R}_1 = 10^{-5} I_2$ and $\hat{R}_2 = 3 \times 10^{-4} I_2$, where $\hat{Q}_1$ and $\hat{Q}_2$ are estimates of Q_1 and Q_2, respectively. The initial value of the estimated weight in the update law (21) is $\hat{W} = 9 \times 10^{-4} \mathbf{1}_{10\times1}$, where $\mathbf{1}_{10\times1}$ denotes a 10-dimensional vector with all elements as 1. The parameters in the update law (31) are $\alpha = 10^{-3}$ and $\beta = 10^{-5}$.

B. Results

Two cases are considered to evaluate the performance of the proposed method. In the first case, there is no force applied to the LBR and the experiment results are shown in Fig. 2. From Figs. 2(a) and 2(c), it is seen that the robot tracks the prescribed trajectory along both X- and Y- axes. Correspondingly, Figs. 2(b) and 2(d) illustrate that the robot stiffness obtained from the proposed method does not change much along both X- and Y- axes. In the second case, the human exerts forces to move the LBR to the position point P_1' when it moves from P_0 to P_1, as depicted in Fig. 1. Since P_1 and P_1' are along the X- axis, an interaction force along the X- axis is needed while the force along the Y- axis is not. This can be observed from Figs. 3(b) and 3(e), where, however, there is a small force noise along the Y- axis. Since the human wants to lead the task in this case, it is expected that the LBR becomes more compliant if it aims to coordinate with the human. From Figs. 3(c) and 3(f), we can see that this is achieved by the proposed method because the robot stiffness reduces from the initial value 300N/m to less than 100N/m when the interaction force is applied along the X- axis. As a result, the actual trajectory deviates from the prescribed one along the X- axis, as shown in Fig. 3(a). Interestingly, we find that the stiffness of the LBR along the X- axis does not increase even when the interaction force disappears. This is because the tracking error is acceptable (shown in Fig. 3(a)), and the LBR does not need to increase its stiffness. This feature is useful as no extra control effort is required and some energy is saved. Along the Y- axis, the results are similar to that in the first case: trajectory tracking is guaranteed and the stiffness does not change much which, however, takes a slight drop.

In conventional impedance control [9], a fixed small (or even zero) stiffness K can be selected when the robot is expected to be compliant to the human in the expense of the robot's tracking performance after the termination of human intervention. Theoretically, if $f = 0$, Eq. (1) becomes $M_d \ddot{x}(t) + C_d \dot{x}(t) + K(x - x_d) = 0$ which indicates that a smaller K will lead to a slower tracking. Conversely, a fixed large stiffness K will be helpful in reducing the tracking error but it makes the robot more difficult to be moved by the human. For comparison, we conduct the same experiment in the second case with a fixed stiffness of 300N/m, which is referred to as "impedance control" in Fig. 3. From Fig. 3, it is found that when the robot is moved to roughly the same position, the interaction force under "impedance control" is larger than that under "proposed method". Therefore, conventional impedance control with a fixed stiffness can only guarantee a trade-off between human effort minimization and trajectory tracking, as discussed in [15]. The proposed method has resolved this problem by evaluating the human's force input and automatically adapting to different situations.

The proposed method has been interpreted in the sense of variable impedance control, while reference adaptation has not been taken into consideration. When the robot's actual velocity is different from the human's desired one, additional interaction force and even wind-up will appear. In the above experiment, reference adaptation can be realized by projecting the actual position onto the reference path and using the projected position as the new reference position. In a general case with a complex reference trajectory, this issue needs to be further investigated.

V. CONCLUSIONS

Physical human-robot interaction has been considered in this paper. The human's control objective has been described by a cost function with unknown weights, which have been estimated by the proposed estimation method. The robot's cost function has been updated accordingly and an optimal control has been developed to minimize it. We have shown that the Nash equilibrium can be achieved for the human-robot interaction system by rigorous analysis. We have verified the validity of the proposed method in a human-robot co-assembly scenario. The experimental results have shown that the LBR under the proposed control can be easily moved by the human to his/her target position, while the trajectory tracking is guaranteed when the human intervention disappears.

REFERENCES

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

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

[3] Y. Li and S. S. Ge, "Force tracking control for motion synchronization in human-robot collaboration," *Robotica*, vol. FirstView, pp. 1–22, 2014.

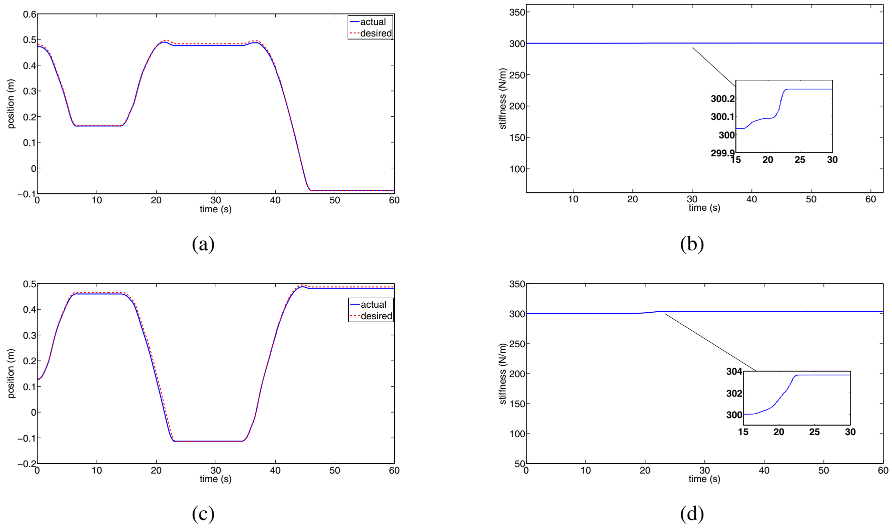

Fig. 2. The first case: trajectories and robot stiffness along X- axis (upper row) and Y- axis (lower row)

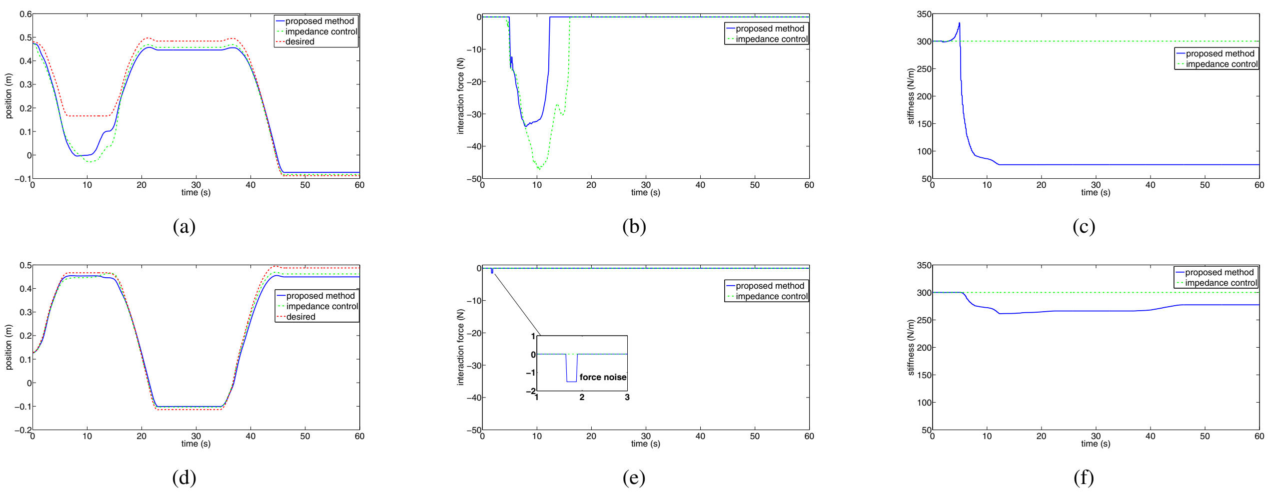

Fig. 3. The second case: trajectories, interaction force and robot stiffness along X- axis (upper row) and Y- axis (lower row)

[4] L. E. Parker, "Distributed intelligence: Overview of the field and its application in multi-robot systems," *Journal of Physical Agents*, pp. 5–14, 2008.

[5] B. D. Argall and A. G. Billard, "A survey of tactile human-robot interactions," *Robotics and Autonomous Systems*, vol. 58, no. 10, pp. 1159–1176, 2010.

[6] T. Wojtara, M. Uchihara, H. Murayama, S. Shimoda, S. Sakai, H. Fujimoto, and H. Kimura, "Human-robot collaboration in precise positioning of a three-dimensional object," *Automatica*, vol. 45, no. 2, pp. 333–342, 2009.

[7] P. Evrard and A. Kheddar, "Homotopy switching model for dyad haptic interaction in physical collaborative tasks," in *3rd Joint Euro-Haptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, (Salt Lake City, Utah, USA), pp. 45–50, 2009.

[8] A. Mortl, M. Lawitzky, A. Kucukyilmaz, M. Sezgin, C. Basdogan, and S. Hirche, "The role of roles: Physical cooperation between humans and robots," *The International Journal of Robotics Research*, vol. 31, no. 13, pp. 1656–1674, 2012.

[9] N. Hogan, "Impedance control: an approach to manipulation-Part I: Theory; Part II: Implementation; Part III: Applications," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 1–24, 1985.

[10] N. Jarrasse, V. Sanguineti, and E. Burdet, "Slaves no longer: review on role assignment for human-robot joint motor action," *Adaptive Behavior*, vol. 22, no. 1, pp. 70–82, 2014.

[11] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory, 2nd Edition*. Society for Industrial and Applied Mathematics, 1998.

[12] R. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton University Press, 1 ed., 1957.

[13] N. Jarrasse, T. Charalambous, and E. Burdet, "A framework to describe, analyze and generate interactive motor behaviors," *PLoS ONE*, vol. 7, no. 11, p. e49945, 2013.

[14] D. E. Kirk, *Optimal Control Theory: An Introduction*. Dover Books on Electrical Engineering, Dover Publications, 2012.

[15] 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.

[16] G. Hudas, K. G. Vamvoudakis, D. Mikulski, and F. L. Lewis, "Online adaptive learning for team strategies in multi-agent systems," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, vol. 9, no. 1, pp. 59–69, 2012.

[17] K. G. Vamvoudakis and F. L. Lewis, "Multi-player non-zero-sum games: Online adaptive learning solution of coupled hamilton-jacobi equations," *Automatica*, vol. 47, no. 8, pp. 1556–1569, 2011.

[18] D. Liu, H. Li, and D. Wang, "Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, pp. 1015–1027, Aug 2014.