A framework of human–robot coordination based on game theory and policy iteration

Article (Accepted Version)

A Framework of Human-Robot Coordination based on Game Theory and Policy Iteration

Yanan Li, *Member, IEEE*, Keng Peng Tee, *Member, IEEE*, Rui Yan, *Member, IEEE*, Wei Liang Chan and Yan Wu, *Member, IEEE*

***Abstract*—In this paper, we propose a framework to analyze the interactive behaviors of human and robot in physical interactions. Game theory is employed to describe the system under study, and policy iteration is adopted to provide a solution of Nash equilibrium. The human's control objective is estimated based on the measured interaction force, and it is used to adapt the robot's objective such that human-robot coordination can be achieved. The validity of the proposed method is verified through a rigorous proof and experimental studies.**

I. Introduction

Physical human-robot interaction is an emerging research field due to the urgent need of robotics in unstructured environments and ad-hoc human inaccessible tasks [1], [2]. 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 [3], [4]. The combination of these advantages in a common task is found to be useful and in many applications necessary, such as tele-operation [5], co-assembly [6] and co-transportation [7], etc.

To develop a natural and efficient human-robot interface is nontrivial. 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 [8]. Most of the 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 to mention the consideration of ergonomics and human factors [9]. 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 [3], [10]. 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 [11], [7], [12], beyond a simple yet robust passive leader-follower model [13]. 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, where the role is usually relevant to the balance of contributions of the human and the robot in a task [14].

Despite the aforementioned research effort, there have been few works done on rigorous analysis of interactive behaviors under an adaptive framework/model. In this paper, we aim to achieve it by integrating game theory [15] and policy iteration [16]. Game theory has been shown to be suitable for analyzing the performance of multi-agent systems [17], in which human-robot interaction is deemed as a two-agent game. In game theory, a variety of interactive behaviors can be described by different combinations of individual objective/cost functions and different optimization criteria. Given a game with known objective/cost functions for linear systems, a conventional method that solves a coupled Riccati equation can be used to obtain the optimal control [18]. In our previous work [19], we have developed an optimal control for human-robot collaboration based on this method. However, it requires solving a Riccati equation in every control loop, which is computationally intensive. Also, the method in [19] generates a fixed control police for a certain cost function, which may not achieve the equilibrium during the adaptation. Policy iteration can be employed to reduce the computational cost and continuously update control policies by evaluating the interaction performance [20]. Methods of policy iteration for games with known and unknown dynamics have been developed by several research groups [21], [22], [23]. 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 [17].

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. It will be proved that the proposed estimation method can be integrated with policy iteration, such that the robot coordinates with the human and optimal control is achieved. A rigorous solution will be provided to the problem of the system equilibrium that has not been fully addressed in our previous work [19]. Part of this work has been presented in [24] with preliminary results. The proposed formula in this work is different from that in [24], where the design parameters in the updating laws are required to satisfy some conditions to guarantee the system stability. Besides, more experimental results are presented in this work to evaluate the control performance of the proposed method.

The rest of this paper is organized as follows. In Section II,

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

All the authors are or were with the Institute for Infocomm Research (I2R), Agency for Science, Technology and Research (A*STAR), Singapore 138632. hit.li.yn@gmail.com

the human-robot interaction system under study is described, and the problem to be resolved is formulated. In Section III, policy iteration for a two-agent game is introduced, the proposed adaptive optimal control is detailed and the performance of the closed-loop system is analyzed. In Section IV, the validity of the proposed method is verified through experimental studies. The limitations of the proposed method and possible future works are discussed in Section V.

II. Problem Formulation

A. System Description

The system under study includes two parts: a human and a robot. In a typical scenario as shown in Fig. 1, 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 as the "interaction force", and it is measured by the force/torque sensor at the interaction point. The human's control objective is unknown to the robot.

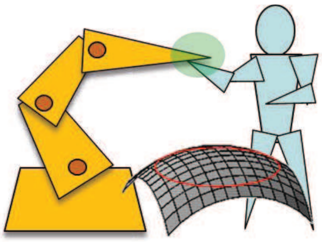


Fig. 1. Illustration of a human-robot interaction system

The robot's forward kinematics are given by

$$x(t) = \phi(q(t)) \tag{1}$$

where $x(t) \in \mathbb{R}^m$ and $q(t) \in \mathbb{R}^n$ are positions in the Cartesian space and the joint space, respectively, and m and n are degrees of freedom (DOFs). By differentiating (1) with respect to time, we have

$$\dot{x}(t) = J(q(t))\dot{q}(t) \tag{2}$$

where $J(q(t)) \in \mathbb{R}^{m\times n}$ is the Jacobian matrix. The robot's dynamics in the joint space are given by

$$\begin{aligned} & M(q(t))\ddot{q}(t) + C(q(t),\dot{q}(t))\dot{q}(t) + G(q(t)) \\ = \; & \tau(t) + J^T(q(t))f(t) \end{aligned} \tag{3}$$

where $M(q(t)) \in \mathbb{R}^{n\times n}$ is the inertia matrix, $C(q(t),\dot{q}(t))\dot{q}(t) \in \mathbb{R}^n$ the Coriolis and centrifugal forces, $G(q(t)) \in \mathbb{R}^n$ the gravitational force, $\tau(t) \in \mathbb{R}^n$ the control input and $f(t) \in \mathbb{R}^n$ the interaction force.

In the field of physical human-robot interaction, impedance control is widely used [13], [25], [26]. In this paper, we employ the following impedance model given in the Cartesian space:

$$M_d\ddot{x}(t) + C_d\dot{x}(t) = u(t) + f(t) \tag{4}$$

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, and $u(t) \in \mathbb{R}^m$ the control input in the Cartesian space.

Remark 1: To make the robot's dynamics in Eq. (3) follow the impedance model in Eq. (4) has been extensively studied in the literature [19], so it will not be detailed in this paper. The control design in the rest of the paper will focus on the impedance model in Eq. (4).

For the feasibility of control design, Eq. (4) can be rewritten in the following state-space form:

$$\dot{z}'(t) = A'z'(t) + B_1'u(t) + B_2'f(t) \tag{5}$$

$$z'(t) = \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix}, \; A' = \begin{bmatrix} \mathbf{0}_m & I_m \\ \mathbf{0}_m & -M_d^{-1}C_d \end{bmatrix}, \\ B_1' = B_2' = \begin{bmatrix} \mathbf{0}_m \\ M_d^{-1} \end{bmatrix} \tag{6}$$

with $\mathbf{0}_m$ and I_m denoting $m \times m$ zero and identity matrices, respectively. To take trajectory tracking into account, we consider that the desired trajectory and velocity are x_d and $\dot{x}_d$, respectively. Then, with the augmented state $z = [z'^T \; x_d^T]^T$, we have

$$\dot{z} = A(z) + B_1(z)u + B_2(z)f \tag{7}$$

$$A(z) = \begin{bmatrix} A'(z')z' \\ \dot{x}_d \end{bmatrix}, \; B_1(z) = B_2(z) = \begin{bmatrix} B'(z') \\ \mathbf{0}_m \end{bmatrix} \tag{8}$$

B. Problem Statement

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 the agents' interactive behaviors. While a more comprehensive list of these behaviors can be found in [15], [8], we only introduce the relevant ones.

Definition 1: [15] Coordination is the most cohesive form of human-robot interaction in which the human and the robot have a common cost function.

We assume that the human has the following cost function

$$\Gamma = \int_0^\infty c(t)dt \tag{9}$$

$$\begin{aligned} c(t) \; = \; & (x - x_d)^T Q_1 (x - x_d) + \dot{x}^T Q_2 \dot{x} + u^T R_1 u \\ & + f^T R_2 f \end{aligned} \tag{10}$$

In the instantaneous cost $c(t)$, $Q_1 \succeq 0$ 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. These weights can be adjusted by the human to define his/her control objective. Considering the definition of the augmented state z, Eq. (10) can be rewritten as

$$c(t) = z^T Q z + u^T R_1 u + f^T R_2 f \tag{11}$$

$$Q = \begin{bmatrix} Q_1 & \mathbf{0}_m & -Q_1 \\ \mathbf{0}_m & Q_2 & \mathbf{0}_m \\ -Q_1 & \mathbf{0}_m & Q_1 \end{bmatrix} \tag{12}$$

If the cost function Γ in (9) is known, it can be used as the cost function of the robot. According to Definition 1, coordination can be achieved with a robot controller that is designed to minimize this cost function. This two-agent game with known cost functions has been studied in the literature, e.g., [21]. Unfortunately, the cost function Γ is determined by the human so it 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, and thus human-robot coordination.

III. Adaptive Optimal Control

A. Preliminary: Nash Equilibrium

Consider a general system described by

$$\dot{\xi} = h(\xi) + g_1(\xi)u_1 + g_2(\xi)u_2 \tag{13}$$

where ξ is the system state, and u_1 and u_2 are two control policies of two agents. For $i = 1, 2$, the following cost functions are defined:

$$\Gamma_i(\xi(0), u_1, u_2) = \int_0^\infty c_i(\xi, u_1, u_2)dt \tag{14}$$

$$c_i(\xi(0), u_1, u_2) = Q(\xi) + u_1^T R_{i1} u_1 + u_2^T R_{i2} u_2 \tag{15}$$

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

$$\begin{aligned} V_i(\xi(t), u_1, u_2) &= \int_t^\infty c_i(\xi, u_1, u_2)ds \\ &= \int_t^\infty (Q(\xi) + u_1^T R_{i1} u_1 + u_2^T R_{i2} u_2)ds \end{aligned} \tag{16}$$

To minimize cost functions Γ_i, different control policies u_i can be found with different definitions of equilibrium [27]. In this paper, we focus on the Nash equilibrium defined as below.

Definition 2: [15] The Nash equilibrium policies u_1^*, u_2^* for the two-agent game satisfy the following two inequalities:

$$\begin{aligned} \Gamma_1(\xi, u_1^*, u_2^*) &\leq \Gamma_1(\xi, u_1, u_2^*) \\ \Gamma_2(\xi, u_1^*, u_2^*) &\leq \Gamma_2(\xi, u_1^*, u_2) \end{aligned} \tag{17}$$

Remark 2: From the definition of the Nash equilibrium, we understand that each agent considers its own cost function and its performance cannot be improved by changing the control policy unilaterally. Moreover, the Nash equilibrium indicates that both agents have the same hierarchical level. It is a joint strategy such that each individual strategy is a best response to the others [27]. To explain why we focus on the Nash equilibrium, we compare it with other commonly-used optimization criteria according to [15]. The Stackelberg equilibrium means that one agent always seeks to minimize its own cost function while the other also to minimize this cost function before minimizing its own. This criterion leads to a leader-follower framework (compared to the same hierarchical level) which is not desirable as already discussed in the Introduction. The Pareto equilibrium means that if either agent uses a control policy other than the equilibrium, either its own cost function or the other's will be increased. This criterion indicates that each agent tries to help the other so it requires additional agreements (cooperation) to achieve the equilibrium. In [28], it is explicitly shown that in human motor interactions, two-player interactions led predominantly to Nash solutions while individual players tended towards cooperation between the two arms. Another equilibrium can be defined such that if one agent uses a control policy other than the equilibrium, the other's cost function can be decreased. This criterion indicates that each agent tries to harm the other as much as possible, which is undesirable in the task discussed in this work. Moreover, in the literature of human motor interactions, it is shown that the performance of both individuals could be improved through interactions [29] and dyads produced much more overlapping forces than individuals so the interactions do not necessarily lead to the most efficient equilibria [30].

Assuming that the function (16) is continuously differentiable, the infinitesimal version of (16) is the so-called nonlinear Lyapunov equation

$$c_i(\xi, u_1, u_2) + \nabla V_i^T (h(\xi) + g_1(\xi)u_1 + g_2(\xi)u_2) = 0 \tag{18}$$

where $V_i(0) = 0$ and ∇V_i is the partial derivative of the value function V_i with respect to ξ. Define the Hamiltonian function

$$\begin{aligned} H_i(\xi, \nabla V_i, u_1, u_2) &= c_i(\xi, u_1, u_2) \\ &\quad + \nabla V_i^T (h(\xi) + g_1(\xi)u_1 + g_2(\xi)u_2). \end{aligned} \tag{19}$$

The optimal value function $V_i^*(\xi)$ is defined by

$$V_i^*(\xi) = \min_{u_i \in \Psi(\Omega)} \int_t^\infty (Q(\xi) + u_1^T R_{i1} u_1 + u_2^T R_{i2} u_2)ds \tag{20}$$

and satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$\min_{u_i \in \Psi(\Omega)} (H(\xi, \nabla V_i, u_1, u_2)) = 0. \tag{21}$$

If the left hand side of (21) exists and is unique, then the optimal control for the given problem is

$$u_i^*(\xi) = -\frac{1}{2} R_{ii}^{-1} g_i^T(\xi) \nabla V_i^*. \tag{22}$$

The online policy iteration algorithm developed in [21] provides a solution to achieve the Nash equilibrium, which is summarized in Algorithm 1.

B. Critic Neural Network and Adaptive Optimal Control

Comparing Eqs. (5) and (13), we see that the human-robot system (5) can be described by the general model (13). Therefore, for the known cost function Γ_i, the Nash equilibrium can be achieved if robot control u and human control f follow controls (22). 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:

$$V(z) = \int_t^\infty c(s)ds \tag{25}$$

As discussed in Section II, the problem is that Γ is unknown, as well as $c(t)$. In this section, we develop an online adaptive algorithm to solve this problem.

Algorithm 1: Policy Iteration

```
begin
    while $\max(\|V_1^{k+1} - V_1^k\|, \|V_2^{k+1} - V_2^k\|) \geq \varepsilon$,
    where $\varepsilon > 0$ is a prescribed small scalar. do
        Start with initial controls $u_1^0$ and $u_2^0$;
        With controls $u_1^k$ and $u_2^k$, solve for costs $V_1^k(\xi)$
        and $V_2^k(\xi)$ using the following equations:
            $c_i(\xi, u_1^k, u_2^k) + (\nabla V_i^k)^T(h(\xi) + g_1(\xi)u_1^k + g_2(\xi)u_2^k) = 0$, with $V_i^k(0) = 0$   (23)
        Update the control policy using
            $u_i^{k+1}(\xi) = -\frac{1}{2}R_{ii}^{-1}g_i^T(\xi)\nabla V_i^k$   (24)
```

Assume that the value function $V(z)$ is continuously differentiable. Then, $V(z)$ is approximated on a compact set Ω by the following critic neural network (NN):

$$V(z) = W^T S(z) + \varepsilon(z) \tag{26}$$

where W is an unknown ideal weight matrix, $S(z)$ is the activation function, and $\varepsilon(z)$ is the bounded NN approximation error. The derivative of $V(z)$ with respect to z is

$$\nabla V = \frac{\partial V(z)}{\partial z} = \left(\frac{\partial S(z)}{\partial z}\right)^T W + \frac{\partial \varepsilon(z)}{\partial z} = \nabla S^T(z)W + \nabla\varepsilon$$

where ∇S and $\nabla\varepsilon$ are bounded gradients of the activation function and approximate error, respectively. $S(z)$ is selected as a complete independent basis set such that $V(z)$ and $\nabla V(z)$ are uniformly approximated, i.e., $\varepsilon \to 0$ and $\nabla\varepsilon \to 0$ for the number of neurons $N \to \infty$. Therefore, the corresponding Hamiltonian function can be written as

$$\begin{aligned} & H(z, \nabla V, u, f) \\ = \; & c(z, u, f) + W^T \nabla S(z)(A(z) + B_1(z)u + B_2(z)f) \\ & + \nabla\varepsilon^T(A(z) + B_1(z)u + B_2(z)f) \\ = \; & c(z, u, f) + W^T\sigma + \varepsilon_H = 0 \end{aligned} \tag{27}$$

where $\sigma = \nabla S(z)(A(z) + B_1(z)u + B_2(z)f)$ and $\varepsilon_H = \nabla\varepsilon^T(A(z)+B_1(z)u+B_2(z)f)$. Since the weight of the critic NN, W, is unknown, the estimated output of critic NN is

$$\hat{V}(\xi) = \hat{W}^T S(\xi), \tag{28}$$

where $\hat{W}$ is the current estimated value of the ideal critic NN weight W. In the human-robot interaction problem,

$$c(z, u, f) = z^T Q z + u^T R_1 u + f^T R_2 f$$

is also unknown due to the unknown Q, R_1 and R_2.

Since the weights in the cost function are relative, we first fix R_2 as a constant and denote the estimates of c, Q and R_1 as $\hat{c}$, $\hat{Q}$ and $\hat{R}_1$, respectively. Then, we have

$$\hat{c} = z^T \hat{Q} z + u^T \hat{R}_1 u + f^T R_2 f. \tag{29}$$

Thus, the approximate Hamiltonian function is

$$H(z, \nabla\hat{V}, u, f) = \hat{c}(z, u, f) + \hat{W}^T\sigma := e. \tag{30}$$

From Eqs. (27) and (30), we can obtain

$$e = (\hat{W} - W)^T\sigma + (\hat{c} - c) - \varepsilon_H = \tilde{W}^T\sigma + \tilde{c} - \varepsilon_H \tag{31}$$

where $\tilde{W} = \hat{W} - W$, $\tilde{c} = \hat{c} - c$ and $\tilde{c} = z^T\tilde{Q}z + u^T\tilde{R}_1 u$ with $\tilde{Q} = \hat{Q} - Q$, $\tilde{R}_1 = \hat{R}_1 - R_1$. For convenience, we denote $\tilde{\theta} = [\text{vec}^T(\tilde{Q}) \ \text{vec}^T(\tilde{R}_1)]^T$ where vec$(\cdot)$ is the column vectorization operator. Correspondingly, we denote

$$\begin{aligned} \theta &= [\text{vec}^T(Q) \ \text{vec}^T(R_1)]^T \\ \hat{\theta} &= [\text{vec}^T(\hat{Q}) \ \text{vec}^T(\hat{R}_1)]^T \end{aligned} \tag{32}$$

By denoting $Y = [\bar{z}^T \ \bar{u}^T]^T$ with

$$\begin{aligned} \bar{z} &= [z^2(1), z(1)z(2), \ldots, z(1)z(3m), z(2)z(1), z^2(2), \\ & \quad \ldots, z(2)z(3m), \ldots, z^2(3m)]^T \\ \bar{u} &= [u^2(1), u(1)u(2), \ldots, u(1)u(m), u(2)u(1), u^2(2), \\ & \quad \ldots, u(2)u(m), \ldots, u^2(m)]^T \end{aligned} \tag{33}$$

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

$$\tilde{c} = \tilde{\theta}^T Y \tag{34}$$

Thus, the estimation error e in (31) can be rewritten as follows:

$$e = \tilde{W}^T\sigma + \tilde{\theta}^T Y - \varepsilon_H. \tag{35}$$

Similarly, using the above denotations into (30), we obtain

$$e = \hat{W}^T\sigma + \hat{\theta}^T Y. \tag{36}$$

Remark 3: Since we know that there are zero sub-matrices in Q (refer to Eq. (12)), 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 and R_1 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.

Based on the policy iteration in Algorithm 1 and the critic NN, a closed-form expression for the optimal robot control is

$$\begin{aligned} u^* &= -\frac{1}{2}R_1^{-1}B_1^T\nabla V \\ &= -\frac{1}{2}R_1^{-1}B_1^T(\xi)\nabla S^T W + \nabla\varepsilon_u \end{aligned} \tag{37}$$

where $\nabla\varepsilon_u = -\frac{1}{2}R_1^{-1}B_1^T\nabla\varepsilon$. Similarly, the human control is

$$f = -\frac{1}{2}R_2^{-1}B_2^T(\xi)\nabla S^T W + \nabla\varepsilon_f \tag{38}$$

which is measured by the force/torqure sensor at the interaction point as mentioned in Section IV-A, and where $\nabla\varepsilon_f = -\frac{1}{2}R_2^{-1}B_2^T\nabla\varepsilon$. Thus, the corresponding approximate robot control u and human control $\hat{f}$ are

$$u = -\frac{1}{2}\hat{R}_1^{-1}B_1^T\nabla S^T\hat{W} \tag{39}$$

$$\hat{f} = -\frac{1}{2}R_2^{-1}B_2^T\nabla S^T\hat{W} \tag{40}$$

Given any feedback control u, our target is to design $\hat{W}$ and $\hat{\theta}$ to minimize the squared residual error

$$E = \frac{1}{2}e^T e + \frac{\beta}{2\alpha_1}e_f^T e_f. \tag{41}$$

where $e_f = \hat{f} - f$ denotes the force error, and α_1 and β are positive constants.

Remark 4: By minimizing the first component of E, i.e., $\frac{1}{2}e^T e$, the approximate Hamiltonian function $H(z, \nabla\hat{V}, u, f)$ in (30) is minimized, so the optimal control and thus the Nash equilibrium can be achieved. By minimizing the second component of E, i.e, $\frac{\beta}{2\alpha_1}e_f^T e_f$, the approximate human control $\hat{f}$ tracks the actual interaction force f, which indicates that the estimated cost $\hat{c}$ tracks the actual cost c of the human. Then, human-robot coordination is achieved.

Based on the standard gradient descent algorithm, the update law for the critic NN weights is given by

$$\begin{aligned}\dot{\hat{W}} &= -\alpha_1\frac{\partial E}{\partial \hat{W}} = -\alpha_1\sigma(\hat{W}^T\sigma + \hat{\theta}^T Y) + \beta\nabla S B_2 R_2^{-T} e_f \\ \dot{\hat{\theta}} &= -\alpha_2\frac{\partial E}{\partial \hat{\theta}} = -\alpha_2 Y(\hat{W}^T\sigma + \hat{\theta}^T Y)\end{aligned} \tag{42}$$

where α_2 is a positive constant. From (35) and (36), $W^T\sigma + \theta^T Y = -\varepsilon_H$. Thus, we obtain the error dynamics as

$$\begin{aligned}\dot{\tilde{W}} &= -\alpha_1\sigma(\tilde{W}^T\sigma + \tilde{\theta}^T Y + \varepsilon_H) + \beta\nabla S B_2 R_2^{-T} e_f \\ \dot{\tilde{\theta}} &= -\alpha_2 Y(\tilde{W}^T\sigma + \tilde{\theta}^T Y + \varepsilon_H).\end{aligned} \tag{43}$$

Denote $\Psi = [W \quad \theta]^T$, $\hat{\Psi} = [\hat{W} \quad \hat{\theta}]^T$, $\tilde{\Psi} = [\tilde{W} \quad \tilde{\theta}]^T$ and $\eta = [\sigma \quad Y]^T$. Then, Eq. (43) can be rewritten as

$$\begin{aligned}\dot{\tilde{W}} &= -\alpha_1\sigma(\tilde{\Psi}^T\eta + \varepsilon_H) + \beta\nabla S B_2 R_2^{-T} e_f \\ \dot{\tilde{\theta}} &= -\alpha_2 Y(\tilde{\Psi}^T\eta + \varepsilon_H).\end{aligned} \tag{44}$$

To guarantee the convergence of $\hat{\Psi}$ to Ψ, the following persistency of excitation (PE) assumption is necessary.

Assumption 1: Let the signal η be persistently exciting over the time interval $[t, t+T]$, i.e., there exist constants $\beta_1 > 0$ and $\beta_2 > 0$, and $T > 0$ such that for all t

$$\beta_1 I \leq \int_t^{t+T} \eta(s)\eta^T(s)ds \leq \beta_2 I. \tag{45}$$

The developed adaptive optimal control is summarized in Algorithm 2.

C. Performance Analysis

Theorem 1: Assume that $\eta(t)$ is persistently exciting and the residual error

$$\bar{\varepsilon}_H = \frac{1}{4c_1}\varepsilon_H^2 + \frac{\beta}{2c_2\alpha_1}\|\nabla\varepsilon_f\|^2$$

is bounded, i.e., $\bar{\varepsilon}_H \leq \varepsilon_m$. Considering the robot dynamics (3), the proposed robot control u in Eq. (40) and the estimated human control in (40) with the update law (42) guarantee that

$$\|\tilde{\Psi}(t)\| \leq \frac{\sqrt{\beta_2 T}}{\beta_1(1-c_1)}(1 + 2\delta\beta_2\alpha_1)\varepsilon_m, \tag{46}$$

where δ is a positive constant of the order of 1.

Algorithm 2: Adaptive Optimal Control

```
Input: Measured state $z$.
Output: Robot control $u$ in Eq. (40).
begin
    Set the weight $R_2$ and initialize estimated weights $\hat{Q}$, $\hat{R}_1$ and $\hat{W}$ in the update law (42), choose activation function $S$ in the value function (25), and set parameters $\alpha_1$, $\alpha_2$ and $\beta$ in the update law (42);
    while $t < t_f$ where $t_f$ is the terminal time do
        Design the desired trajectory $x_d$, collect the data of state $z' = [x^T \ \dot{x}^T]^T$, robot control $u$ and interaction force $f$, and form the vector $\eta$;
        Obtain robot control $u$ in Eq. (40);
        Obtain the estimated human control $\hat{f}$ in Eq. (40);
        Update $\hat{W}$ and $\hat{\theta}$ according to Eq. (42);
```

Proof 1: Consider a Lyapunov function candidate:

$$U(t) = \frac{1}{2\alpha_1}\tilde{W}^T\tilde{W} + \frac{1}{2\alpha_2}\tilde{\theta}^T\tilde{\theta}. \tag{47}$$

The time derivative of the Lyapunov function candidate is

$$\begin{aligned}\dot{U} &= \frac{1}{\alpha_1}\tilde{W}^T\dot{\tilde{W}} + \frac{1}{\alpha_2}\tilde{\theta}^T\dot{\tilde{\theta}} \\ &= -(\tilde{\Psi}^T\eta)^2 - \tilde{\Psi}^T\eta\epsilon_H + \frac{\beta}{\alpha_1}\tilde{W}^T\nabla S B_2 R_2^{-T} e_f \\ &= -(\tilde{\Psi}^T\eta)^2 - \tilde{\Psi}^T\eta\epsilon_H + \frac{\beta}{\alpha_1}\hat{W}^T\nabla S B_2 R_2^{-T} e_f \\ &\quad - \frac{\beta}{\alpha_1}W^T\nabla S B_2 R_2^{-T} e_f.\end{aligned} \tag{48}$$

Consider the representations of f and $\hat{f}$ in Eqs. (38) and (40), the last two terms of (48) can be rewritten as:

$$\begin{aligned}&\frac{\beta}{\alpha_1}\hat{W}^T\nabla S B_2 R_2^{-T} e_f - \frac{\beta}{\alpha_1}W^T\nabla S B_2 R_2^{-T} e_f \\ = & -\frac{2\beta}{\alpha_1}\|e_f\|^2 - \frac{2\beta}{\alpha_1}e_f^T\nabla\varepsilon_f.\end{aligned} \tag{49}$$

By substituting (49) into (48), we further achieve

$$\dot{U} = -(\tilde{\Psi}^T\eta)^2 - \tilde{\Psi}^T\eta\epsilon_H - \frac{2\beta}{\alpha_1}\|e_f\|^2 - \frac{2\beta}{\alpha_1}e_f^T\nabla\varepsilon_f. \tag{50}$$

By considering Young's inequality, we know that there exist $c_1, c_2 \in (0, 1)$ such that

$$-\tilde{\Psi}^T\eta\varepsilon_H \leq c_1(\tilde{\Psi}^T\eta)^2 + \frac{1}{4c_1}\varepsilon_H^2,$$

$$-e_f^T\nabla\varepsilon_f \leq c_2\|e_f\|^2 + \frac{1}{4c_2}\|\nabla\varepsilon_f\|^2.$$

Thus, we obtain

$$\begin{aligned}\dot{U} &\leq -(1-c_1)(\tilde{\Psi}^T\eta)^2 - \frac{2\beta}{\alpha_1}(1-c_2)\|e_f\|^2 \\ &\quad + \left(\frac{1}{4c_1}\varepsilon_H^2 + \frac{\beta}{2c_2\alpha_1}\|\nabla\varepsilon_f\|^2\right) \\ &\leq -(1-c_1)(\tilde{\Psi}^T\eta)^2 + \bar{\varepsilon}_H.\end{aligned} \tag{51}$$

This implies that $\dot{U} \leq 0$ if $\tilde{\Psi}^T\eta > \sqrt{\frac{\varepsilon_m}{1-c_1}}$, so it gives a bound for $\tilde{\Psi}^T\eta$. Based on Technical Lemma 2 in [31], we can achieve (46).

Remark 5: In the Hamiltonian function (27), if the number of hidden layer neurons N tends to infinity, then the approximation error $\nabla\varepsilon \to 0$ uniformly. This implies that $\hat{\Psi}$ converges to the real Ψ when N tends to infinity. Thus, the proposed robot control u in Eq. (40) converges to Nash equilibrium solution u^* in (37), and the estimate of the unknown instantaneous cost of the human c, i.e., $\hat{c}$, is obtained so coordination defined in Definition 1 is realized.

Remark 6: This theorem requires an assumption that the robot control u is bounded to guarantee the boundedness of ε_H. Obviously, it is reasonable in practical applications.

D. Implementation: Variable Impedance Control

To implement the proposed control in Algorithm 2, we employ a radial basis function (RBF) NN. In (25), $S(z) = [s_1, \ldots, s_N]^T$, where N is the number of neural nodes. The activation function is chosen as the Gaussian function, i.e.,

$$s_i(z) = \exp\left[\frac{-(z-\mu_i)^T(z-\mu_i)}{\eta_i^2}\right] \tag{52}$$

where $\mu_i = [\mu_{i,1}, \mu_{i,2}, ..., \mu_{i,6}]$ is the center of the receptive field, and η_i the width of the Gaussian function, $i = 1, \ldots, N$. Then, $\nabla S(z) = [\nabla s_1, \ldots, \nabla s_N]^T$ with

$$\nabla s_i(z) = -\frac{2}{\eta_i^2}s_i(z)(z-\mu_i)^T \tag{53}$$

To interpret the proposed control, we set $\mu_i = [x_d^T, (\dot{x}+k(x_d-x))^T, x_d^T]^T$ where $k > 0$. By substituting $\nabla S(z)$ into Eq. (40), we obtain

$$\begin{aligned} u &= \hat{R}_1^{-1}B_1^T(z-\mu_i)S_\eta^T(z)\hat{W} \\ &= kS_\eta^T(z)\hat{W}\hat{R}_1^{-1}M_d^{-T}(x_d - x) \end{aligned} \tag{54}$$

where $S_\eta(z) = [\frac{s_1}{\eta_1^2}, \ldots, \frac{s_N}{\eta_N^2}]^T$. By denoting

$$K = kS_\eta^T(z)\hat{W}\hat{R}_1^{-1}M_d^{-T} \tag{55}$$

we can rewrite $u = K(x_d - x)$. By substituting it into Eq. (4), we have

$$M_d\ddot{x}(t) + C_d\dot{x}(t) + K(x - x_d) = f \tag{56}$$

which has a form of conventional impedance control [13] and where K stands for the robot stiffness. It is obvious that a smaller K will lead to a larger tracking error if $f = 0$. Conversely, a large stiffness K will be helpful in reducing the tracking error but it makes the robot more difficult to be moved by the human. From (55), we find that u is a variable impedance control where the stiffness is updated when the parameters $\hat{W}$ and $\hat{R}_1$ are updated. Under this method, we expect that the robot stiffness is high when trajectory tracking is needed while it is low when the human intervention exists. This is in line with the definition of coordination: although the robot has its own objective, it will surrender its objective and take the human's objective as its own when there is human intervention.

IV. Experiments

A. Experimental Settings

In this section, we consider an application of human-robot co-assembly as sketched in Fig. 2(a). In this application, 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., if the order of the workpiece is changed as shown in Fig. 2(a), 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 this application, we design the experiment setup as shown in Fig. 2(b). The robot is a 7-DOFs KUKA lightweight robot (LBR), which is in impedance control mode in the Cartesian space. The human may move the robot by applying a force to its end-effector, which is calculated based on the measured torque at each joint. 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. At last, the robot moves to the final position point P_1'.

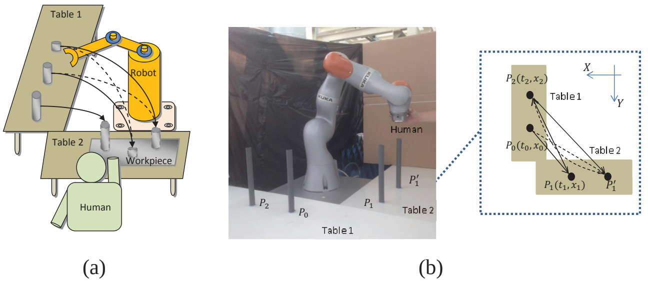


Fig. 2. a) A sketch of human-robot co-assembly. The solid arrow indicates the prescribed trajectory of the robot, while the dotted one the trajectory refined by the human to handle the unexpected situation of a changed order. For a clear illustration, only the forward trajectories are shown and the backward trajectories to the next workpieces are not shown. b) The experiment setup. The diagram on the right hand side depicts the robot setup from the top-view.

The orientation of the tool center point is constrained to be parallel to negative Z axis, as seen in Fig. 2(b). By setting the robot stiffness as $200I_2$N/m, damping factor as 0.7 and four position points in $X-Y$ plane: P_0 ($t_0 = 0.0$s, $x_0 = [0.47\ 0.13]^T$m), P_1 ($t_1 = 6.5$s, $x_1 = [0.18\ 0.45]^T$m), P_2 ($t_2 = 52.0$s, $x_2 = [0.47\ {-0.11}]^T$m) and P_1' ($[-0.06\ 0.45]^T$m), the robot runs in impedance control mode and the actual trajectory in this run is recorded as the desired trajectory of the robot, i.e., x_d. The parameters used in the proposed control are as follows. In Eq. (55), $\eta_i = 100$ and $k = 10$. Impedance parameters in Eq. (4) are chosen as $M_d = 3I_2$kg and $C_d = 10I_2$N/m. The initial values of the estimated weights in the estimated instantaneous cost (29) are $\hat{Q}_1 = 100I_2\text{m}^{-2}$, $\hat{Q}_2 = I_2\text{s}^2/\text{m}^2$, $\hat{R}_1 = 10^{-3}I_2\text{N}^{-2}$ and $\hat{R}_2 = 10^{-3}I_2\text{N}^{-2}$, where $\hat{Q}_1$ and $\hat{Q}_2$ are estimates of Q_1 and Q_2, respectively. In the update law (42), the initial value of the estimated weight $\hat{W}$ is calculated based on Eq. (55) with the prescribed initial value of the robot stiffness K. The parameters $\alpha_1 = 0.1$, $\alpha_2 = 10^{-8}$ and $\beta = 2 \times 10^{-6}$. Note that the parameters α_1 and α_2 can be set with different values: large values will speed up the convergence but they will also lead to fast changes of stiffness (along the lines of variable impedance control, as in Section III-D). In practice, the allowable rate of stiffness change should be moderated under considerations of human safety, as well as robot hardware and control software limitations.

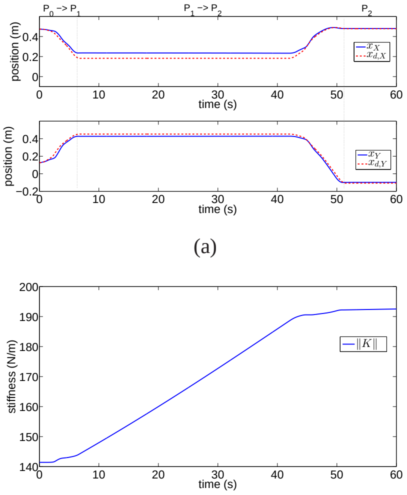
Because the force signal is noisy, the filtering of the force signal is used: the interaction force f is the average of the measured forces in the last 10 control loops. To avoid the robot stiffness becoming too large or too small, we set upper and lower bounds for the stiffness as $500I_2$N/m and $50I_2$N/m, respectively. Moreover, to satisfy the PE assumption, a sweeping frequency signal is added into u which is small enough to not cause any disturbance to the human.

B. Control Performance

In this subsection, we evaluate the control performance of the proposed method by setting different initial values of the robot stiffness and applying different forces to the robot.

In the first case, the initial value of the robot stiffness is $K = 100I_2$N/m and no force is applied to the robot. Because the robot stiffness is smaller ($100I_2$N/m) than that when recording x_d ($200I_2$N/m), there is a tracking error at the beginning, as shown in Fig. 3(a). Correspondingly, the robot stiffness increases at the beginning and keeps unchanged when there is no tracking error, as shown in Fig. 3(b). Because the control objective is trajectory tracking, the proposed method increases the robot stiffness to reduce it. When the tracking error becomes very small and gets near to zero, the increase of the robot stiffness slows down and the convergence is achieved. Note that we plot $\|K\|$ instead of two diagonal elements of K in Fig. 3(b), because the profiles of the robot stiffness in X and Y directions are similar. We also do the same in the following figures showing the robot stiffness. Also, note that although the speed of the movements is relatively slow, there is not a limit on it theoretically. The reasons that we have the speed in the experiments are: 1) the human needs to physically interact with the robot so the speed of the movements has to be limited within a certain level to guarantee the safety; and 2) the robot is moved partially by the human so the actual speed is also the speed of the human hand which is not high itself. These issues may be addressed by using a lightweight robot so that it is safe to move with a higher speed and the robot can be moved by the human more easily.

In the second case, the initial value of the robot stiffness is $K = 200I_2$N/m and there is a force applied to the robot by the human. Due to the interaction force, the robot trajectory drifts away from the desired one as shown in Fig. 4(a). Under the proposed method, the robot stiffness decreases to reduce the interaction force so that the human can move the robot with more ease, as shown in Fig. 4(b). In this case, the control objective becomes force minimization and the proposed method achieves this by decreasing the robot stiffness. This result shows the meaning of “human-robot coordination”: the robot gives up its control objective to achieve the human's.

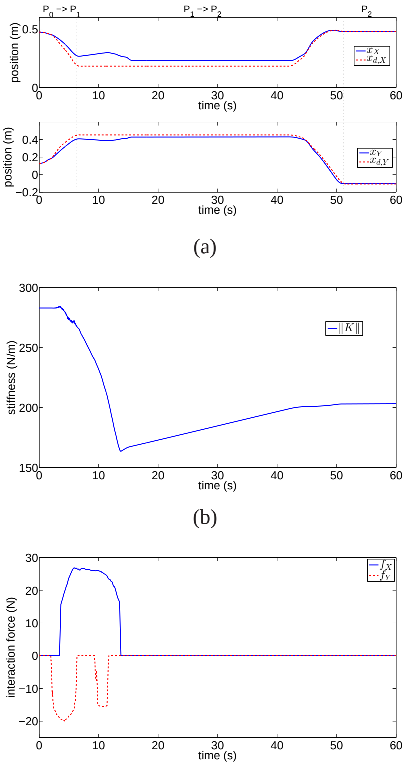


(a)

(b)

Fig. 3. The first case where the initial value of the robot stiffness is $K = 100I_2$N/m and no force is applied to the robot: (a) trajectories and (b) robot stiffness $\|K\|$.

Similarly as in the first case, when the force disappears, the control objective becomes trajectory tracking and the robot stiffness increases again to reduce the tracking error.



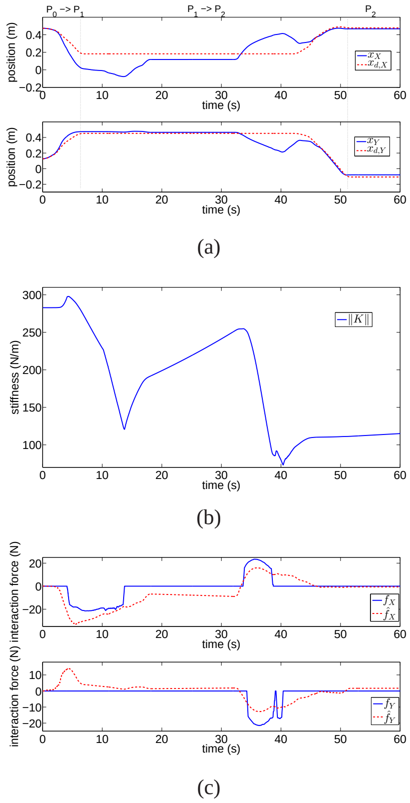
(a)

(b)

(c)

Fig. 4. The second case where the initial value of the robot stiffness is $K = 200I_2$N/m and there is a force applied to the robot by the human: (a) trajectories, (b) robot stiffness $\|K\|$ and (c) interaction forces.

In the third case, the initial value of the robot stiffness is $K = 200I_2$N/m and there are forces applied to the robot by the human at two different time durations. At the time duration of 4s-14s, the force is along the X axis, so the robot trajectory along the X axis drifts away from the desired one (Fig. 5(a)). Correspondingly, Fig. 5(b) shows that the robot stiffness reduces at this time duration. In Fig. 5(c), it is shown that the estimated interaction force $\hat{f}_X$ tries to track the actual interaction force f_X. When the force disappears at 14s-33s, the robot stiffness increases due to the existence of the tracking error along the X axis. Another force is applied to the robot along both X and Y axes at 33s-40s so the stiffness reduces again, and $\hat{f}_X$ tries to track f_X as well. After 40s, the stiffness increases to reduce the tracking error till the end of this case. In this case, the estimated interaction force $\hat{f}_X$ tries to track the actual interaction force f_X, which indicates that “human-robot coordination” is achieved. Besides, the profile of the robot stiffness illustrates that the proposed method adapts to the interaction force and the tracking error continuously.



(a)

(b)

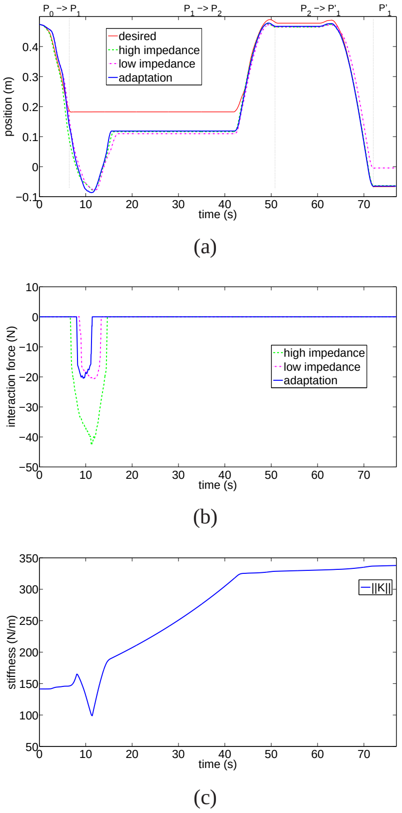
(c)

Fig. 5. The third case where the initial value of the robot stiffness is $K = 200I_2$N/m and there are forces applied to the robot by the human at two different time durations: (a) trajectories, (b) robot stiffness $\|K\|$ and (c) interaction forces.

Based on the above experimental results, we summarize that the proposed method can guarantee the following control performance: when there is an interaction force, the robot will become more compliant to make the interaction easier; when there is no force, the robot will become stiffer to achieve trajectory tracking; the robot stiffness will keep unchanged if neither the interaction force nor the tracking error exists; the change of the robot stiffness is continuous which is corresponding to the interaction force and the tracking error; and the estimated interaction force $\hat{f}_X$ tracks the actual interaction force f_X which indicates that “human-robot coordination” is achieved.

C. Comparison with Conventional Impedance Control

In this subsection, we conduct the experiment as illustrated in Fig. 2(b). As discussed in Section III-D, a fixed low (or even zero) stiffness K can be selected when the robot is expected to be compliant to the human. Conversely, a fixed high stiffness K can be selected when the robot is expected to track a desired trajectory. These two schemes are implemented for comparison with the proposed method, i.e., the following three conditions are considered: i) fixed high stiffness: $K = 200I_2$N/m; ii) fixed low stiffness: $K = 100I_2$N/m; and iii) adaptation: the proposed method with the initial value of K as $100I_2$N/m.



(a)

(b)

(c)

Fig. 6. Comparisons between conditions of “adaptation”, “fixed high stiffness” and “fixed low stiffness”: (a) trajectories, (b) interaction forces and (c) robot stiffness $\|K\|$ under the “adaptation” condition.

The experimental results are summarized in Fig. 6. Note that P_1 and P_1' are along the X axis, and an interaction force along the X axis is needed while the force along the Y axis is not. Therefore, the results of positions and forces only for the X axis are presented. From Fig. 6(a), we can see that the robot can be moved to P_1' when it tries to move to P_1 under all three conditions. However, the forces used to move the robot are different, as shown in Fig. 6(b). In particular, the interaction force under the condition of high stiffness is the largest and that under the conditions of low stiffness and adaptation are smaller. These results are in line with the expectation: a high stiffness is undesirable when there is human intervention and a low stiffness should be selected. When the forces disappear after around $t = 15$s in the three conditions, the actual trajectories are expected to track the desired one. From Fig. 6(a), we see that trajectory tracking cannot be achieved under the condition of low stiffness. In particular, there is a static state tracking error after $t = 65$s. Under the conditions of high stiffness and adaptation, trajectory tracking is finally achieved. Fig. 6(c) shows the change of the robot stiffness under the adaptation condition: when the force is applied to the robot, its stiffness reduces so the performance is similar to that under the condition of low stiffness; when the force disappears, the robot stiffness increases so the performance is similar to that under the condition of high stiffness; the increasing of the stiffness slows down when the tracking error becomes smaller; and in the whole process, the change of the robot stiffness is automatic and continuous.

Based on the above results, we conclude that conventional impedance control with a fixed stiffness can only guarantee a trade-off between human effort minimization and trajectory tracking, and the proposed method has resolved this problem by evaluating the human's force input and automatically adapting to different situations.

V. Discussion

The proposed method has an underlying assumption that human motor control is to minimize the cost function Γ in (9), and the human's control objective is estimated by minimizing the defined force error e_f. However, the interaction force f is subject to the measurement noise and the human uncertainty exists due to many factors. These issues have not been modeled in the current formula and will be investigated in our future works.

The human-robot interaction behaviors can be more complex than just coordination, which is a specific case studied in this paper. Nevertheless, it is important to note that a higher-level decision mechanism must be specified beforehand if the robot is expected to perform more complex interactive behaviors. The proposed method can only make the robot adapt to different human inputs based on a prescribed goal ("coordination" in this paper). This goal can be set by the designer based on priori knowledge of a task or through a certain learning method.

In the proposed method, the interaction force f is used as the signal to observe the human's unknown state. This is only applicable in the applications where the physical interaction exists. In other applications, e.g., tele-operation without force feedback, how to implement the proposed method is unclear at this stage. It should be expected that there is another signal such as position error to replace f which can be used to describe the human's motion intention. This is an interesting topic which will be also one of our future works.

From Eq. (56), we see that the interaction force drives a mass-damper-spring system from its current position x to the robot's reference trajectory x_d. When the robot reaches the human's desired position, there is a force trying to drag the robot to its reference trajectory x_d. There are two ways to reduce this force, namely by reducing the robot stiffness (using the proposed method in this paper) and by adapting the reference trajectory [32], [33], [34]. In particular, we may use splines to yield a smooth path that stores the deformation from the original path, similarly as in [35], [36], [37]: i) find the point in the original desired path of the robot which is nearest to the current position point x, and denote it as x_0; ii) take the following $l+1$ points as the control points of a spline: x, x_0 and another $l-1$ points that are in front of x_0 in terms of time and in the original desired path of the robot; iii) connect these control points using a spline and form a new desired path of the robot; and iv) set boundary conditions such as desired velocity and acceleration, to determine the new desired trajectory x_d. By applying reference adaptation, it is expected that the new desired trajectory x_d gets close to the robot's current position x, and the interaction force due to the difference between them is reduced. Besides this, the replanned desired trajectory can help in making the robot's motion smoother when the human releases the robot and it moves back to the original desired path. How to combine reference adaptation with the proposed method and what effects that they have on each other need to be further studied.

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] C. Passenberg, A. Peer, and M. Buss, "A survey of environment-, operator-, and task-adapted controllers for teleoperation systems," *Mechatronics*, vol. 20, no. 7, pp. 787 – 801, 2010.

[3] 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.

[4] Y. Li and S. S. Ge, "Force tracking control for motion synchronization in human-robot collaboration," *Robotica*, vol. 34, no. 6, pp. 1260–1281, 2016.

[5] S. Hirche and M. Buss, "Human-oriented control for haptic teleoperation," *Proceedings of the IEEE*, vol. 100, pp. 623–647, 2012.

[6] C. Liu and M. Tomizuka, "Modeling and controller design of cooperative robots in workspace sharing human-robot assembly teams," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1386–1391, Sept 2014.

[7] 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.

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

[9] 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.

[10] 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.

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

[12] J. R. Medina, M. Lawitzky, A. Molin, and S. Hirche, "Dynamic strategy selection for physical robotic assistance in partially known tasks," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1180–1186, 2013.

[13] 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.

[14] 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.

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

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

[17] 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.

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

[19] 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.

[20] F. L. Lewis and D. Vrabie, “Reinforcement learning and adaptive dynamic programming for feedback control,” *IEEE Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32–50, 2009.

[21] 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.

[22] H. Zhang, Q. Wei, and D. Liu, “An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum differential games,” *Automatica*, vol. 47, no. 1, pp. 207–214, 2011.

[23] 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.

[24] 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 *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (Hamburg, Germany), pp. 20–25, September 28-October 02 2015.

[25] A. A. Blank, A. M. Okamura, and L. L. Whitcomb, “Task-dependent impedance and implications for upper-limb prosthesis control,” *The International Journal of Robotics Research*, vol. 33, no. 6, pp. 827–846, 2014.

[26] J. Vogel, S. Haddadin, B. Jarosiewicz, J. D. Simeral, D. Bacher, L. R. Hochberg, J. P. Donoghue, and P. van der Smagt, “An assistive decision-and-control architecture for force-sensitive hand-arm systems driven by human-machine interfaces,” *The International Journal of Robotics Research*, vol. 34, no. 6, pp. 763–780, 2015.

[27] L. Busoniu, R. Babuska, and B. De Schutter, “A comprehensive survey of multiagent reinforcement learning,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, pp. 156–172, March 2008.

[28] D. A. Braun, P. A. Ortega, and D. M. Wolpert, “Nash equilibria in multi-agent motor interactions,” *PLoS Computational Biology*, vol. 5, p. e1000468, 8 2009.

[29] G. Ganesh, A. Takagi, R. Osu, T. Yoshioka, M. Kawato, and E. Burdet, “Two is better than one: Physical interactions improve motor performance in humans,” *Scientific Reports*, vol. 4, no. 3824, 2014.

[30] R. P. van der Wel, G. Knoblich, and N. Sebanz, “Let the force be with us: dyads exploit haptic coupling for coordination.,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 5, p. 1420, 2011.

[31] K. G. Vamvoudakis and F. L. Lewis, “Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem,” *Automatica*, vol. 46, no. 5, pp. 878 – 888, 2010.

[32] G. Ganesh, A. Albu-Schaeffer, M. Haruno, M. Kawato, and E. Burdet, “Biomimetic motor behavior for simultaneous adaptation of force, impedance and trajectory in interaction tasks,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, (Anchorage, Alaska, USA), pp. 2705–2711, May 3-8 2010.

[33] C. Yang and E. Burdet, “A model of reference trajectory adaptation for interaction with objects of arbitrary shape and impedance,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (San Francisco, CA, USA), pp. 4121–4126, September 25-30 2011.

[34] G. Ganesh, N. Jarrasse, S. Haddadin, A. Albu-Schaeffer, and E. Burdet, “A versatile biomimetic controller for contact tooling and haptic exploration,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3329–3334, May 2012.

[35] E. S. Boy, E. Burdet, C. L. Teo, and J. E. Colgate, “Investigation of motion guidance with scooter cobot and collaborative learning,” *IEEE Transactions on Robotics*, vol. 23, pp. 245–255, April 2007.

[36] Q. Zeng, C. L. Teo, B. Rebsamen, and E. Burdet, “Collaborative path planning for a robotic wheelchair,” *Disability and Rehabilitation: Assistive Technology*, vol. 3, no. 6, pp. 315–324, 2008.

[37] Q. Zeng, C. L. Teo, B. Rebsamen, and E. Burdet, “A collaborative wheelchair system,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 2, pp. 161–170, 2008.

Yanan Li (S’10-M’14) received the BEng and MEng degrees from the Harbin Institute of Technology, China, in 2006 and 2008, respectively, and the Ph.D. degree from the NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, in 2013. He has been a Research Scientist with the I2R, A*STAR, from 2013 to 2015. He is currently a Research Associate with the Department of Bioengineering, Imperial College London, UK. His research interests include physical human-robot interaction and robot control.

Keng Peng Tee (S’04-M’08) received the BEng (first class honors), MEng and PhD degrees in 2001, 2003 and 2008, respectively, all from the National University of Singapore. In 2008, he joined the I2R, A*STAR, where he is now a group leader. He has been an Associate Editor for the Conference Editorial Board of the IEEE Control Systems Society since 2012. His current research interests include adaptive control, robot manipulation, human-robot interfaces and human-robot collaboration.

Rui Yan (M’11) received the BS and MS degrees from the Department of Mathematics, Sichuan University, Chengdu, China, in 1998 and 2001, respectively, and the PhD degree from the Department of Electrical and Computer Engineering, National University of Singapore, in 2006. She was a postdoctoral research fellow in the University of Queensland, Australia, from 2006 to 2008, and a Research Scientist with the I2R, A*STAR from 2009 to 2014. Now she is a professor in the College of Computer Science, Sichuan University, China. Her research interests include intelligent robots, nonlinear control, neural computation, and complex systems analysis.

Wei Liang Chan received the BS degree in electrical engineering from the National University of Singapore in 2011. He is currently a Research Engineer with the I2R, A*STAR. His research interests include human-robot interaction, telerobotic systems, and user interface design for robotics.

Yan Wu received the BA degree in Engineering from the University of Cambridge, UK, in 2007, and the PhD degree in Electrical Engineering from Imperial College London, UK, in 2013. In 2012, he joined both the Institute of Child Health, University College London as an Honorary Research Associate and NHS Great Ormond Street Hospital for Children as a Research Fellow. Since 2013, he has been with the I2R, A*STAR, where he serves as a Research Scientist. Yan’s research interests lie in imitation learning and human-robot interaction.