Adaptive Impedance Control of Human–Robot Cooperation Using Reinforcement Learning

Zhijun Li, *Senior Member, IEEE*, Junqiang Liu, Zhicong Huang, Yan Peng, Huayan Pu, and Liang Ding, *Member, IEEE*

***Abstract*—This paper presents human–robot cooperation with adaptive behavior of the robot, which helps the human operator to perform the cooperative task and optimizes its performance. A novel adaptive impedance control is proposed for the robotic manipulator, whose end-effector's motions are constrained by human arm motion limits. In order to minimized motion tracking errors and acquire an optimal impedance mode of human arms, the linear quadratic regulation (LQR) is formulated; then, integral reinforcement learning (IRL) has been proposed to solve the given LQR with little information of the human arm model. Considering human–robot interaction force during the robot performing manipulation, a novel barrier-Lyapunov-function-based adaptive impedance control incorporating adaptive parameter learning is developed for physical limits, transient perturbations, and time-varying dynamics. Experimental results validate that the proposed controller is effective in assisting the operator to perform the human–robot cooperative task.**

***Index Terms*—Adaptive impedance control, barrier Lyapunov function (BLF), human–robot cooperation (HRC), integral reinforcement learning (IRL), linear quadratic regulation (LQR).**

Manuscript received September 9, 2016; revised January 18, 2017 and March 2, 2017; accepted March 29, 2017. Date of publication April 14, 2017; date of current version September 11, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61573147, Grant 91520201, and Grant 61625303, in part by the Guangzhou Research Collaborative Innovation Projects (2014Y2-00507), in part by the Guangdong Science and Technology Research Collaborative Innovation Projects under Grant 2013B010102010, Grant 2014B090901056, and Grant 2015B020214003, in part by the Guangdong Science and Technology Plan Project (Application Technology Research Foundation) (2015B020233006), in part by the National High-Tech Research and Development Program of China (863 Program) under Grant 2015AA042303, and in part by the State Key Laboratory of Robotics and System, Harbin Institute of Technology, under Grant SKLRS-2016-KF-04. *(Corresponding author: Zhijun Li.)*

Z. Li is with the College of Automation Science and Engineering, South China University of Technology, Guangzhou 510630, China, and with the Department of Automation, University of Science and Technology of China, Hefei 230022, China, and also with the State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China (e-mail: zjli@ieee.org).

J. Liu and Z. Huang are with the College of Automation Science and Engineering, South China University of Technology, Guangzhou 510630, China (e-mail: 979488633@qq.com; zhiconghuang1@163.com).

Y. Peng and H. Pu are with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China (e-mail: pengyan@shu.edu.cn; phygood_2001@shu.edu.cn).

L. Ding is with the Research Institute of Intelligent Control and Systems, School of Astronautics, Harbin Institute of Technology, Harbin 150001, China (e-mail: liangding@hit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2017.2694391

I. Introduction

TODAY, a vast variety of robots has been created to help humans. To realize this design goal, robots must interact with the external environments. Depending on the requirements of tasks, robots may interact with diversified objects, including humans, it should be capable of controlling not only positions but also the interaction forces. Because of the extremely complicated nature of humans, a large field of study is dedicated to understanding, designing, and evaluating robot systems for human–robot interaction. Adjustment of the impedance parameters was investigated, which considered the system stability [1], the control torques [2], minimizing a performance index [3], and the human intention estimation [8]. However, the above-mentioned approaches ignored the individual discrimination; then, the approaches called human adaptive mechatronics had been proposed in [4]–[6], which considered the individual discrimination by regulating the impedance coefficients of the robot based on estimating human arm dynamics. The works [7], [8] have been done on identifying the human arm dynamical parameters. The impedance parameters can be adjusted to assure the system stability. It should be noted that stability is a necessary condition for a controlled plant, for example, in [10] and [11], the authors adjusted the impedance based on estimating human arm impedance parameters. However, it should be preferred to tune these impedance parameters to achieve optimal performance of the system.

In this paper, we present human–robot cooperation with adaptive behavior of the robot, which helps the human operator to perform the cooperative task and optimizes its performance. Considering the human operator, we propose two control loops in the human–robot cooperation, for the inner control loop, it is a robot-oriented loop, which can handle unknown dynamics or a prescribed robotic impedance mode perceived by a human arm. For the outer loop, it is a task-oriented loop designed to optimize the parameters of the impedance model, which is formulated as a linear quadratic regulator (LQR), and reinforcement learning (RL) is employed to find the optimized parameters of the given LQR with little information of the human arm model. Therefore, the work considers that the robot's end effector remains inside a task space, which is constrained by the physical limits of a human arm. Considering barrier Lyapunov function (BLF) [18], [19], [24], [25] to maintain the constraints on the robot's end-effector within the physical limits, a novel BLF-based adaptive impedance control incorporating adaptive parameter learning is

developed for physical limits, transient perturbations, and time-varying dynamics. Experimental results validate that the proposed controller is effective in assisting the operator to perform the human robot cooperative task.

II. Overview of Human–Robot Cooperation

In the section, the overview of developing the system will first be presented; then, a structure of the proposed human–robot cooperation system control will be developed.

A. Human–Robot Cooperation

In the human–robot cooperation, robots interacting with humans are capable of assisting the humans by adjusting themselves to adapt human action and acquire the human's intention and compensating for human mistakes, because of tiredness, pressure, etc., possibly. In [12], a method based on frequency-domain stability observers was designed to detect unstable behavior, and the robot was stabilized by using online adaptation of the control gains for physical human–robot interaction. In [13], a low-impedance mechanism consisting of passive joints was specifically proposed for human–robot cooperation, and the experiments verify that effective impedance is greatly reduced. In [14], in order to enhance human–robot cooperation during training sessions, an impedance-based interactive training paradigm was proposed for a lightweight compliant ankle robot. In [15], an assistive robotic device for handling large payloads has been presented, which is based on minimized inertia, a parallel cable/belt routing system, and variable static balancing. In [16], electroencephalograph that controlled the upper limb robotic arm via a Bluetooth interface has been proposed. In [17], in order to recognize motion intention of patients, the information of the position and torque sensors in joints of an exoskeleton leg is utilized, and the dynamics modeling and identification of the leg were also investigated.

In the above works, the robot can enhance the performance of the human–robot system in terms of human force, precision, and even safety. For example, the skillful operator may be easily able to deal with high-speed movements with a small damping coefficient. Thus, it is expected that the robot can tune the damping coefficient such that the response of the system is similar to the human. Therefore, for developing such a human–robot system, one important issue is how to design the adaptive control for the assistive human–robot system so that safety and assistance can be provided for the human with better performance and few knowledge of the human and robot dynamics.

To develop such an optimized assistive human–robot cooperation system, we propose a control structure that partitioned the robot-oriented control and the task-oriented control, where the task-oriented loop includes the task information, and the optimized impedance model for the human operator, while the robot-oriented control handles unknown dynamics such that the robot behaves like an impedance model.

B. Developed System of Human–Robot Cooperation

The development of the human–robot cooperation here is inspired by the investigation of human-in-the-loop in [8] and [9]. First, a computed torque controller is used to obviate the requirement for the human and learn the robot-oriented model. Second, the assistive inputs are exerted to magnify the human's effort such that the human fulfills a prescribed task with the best performance and minimum effort.

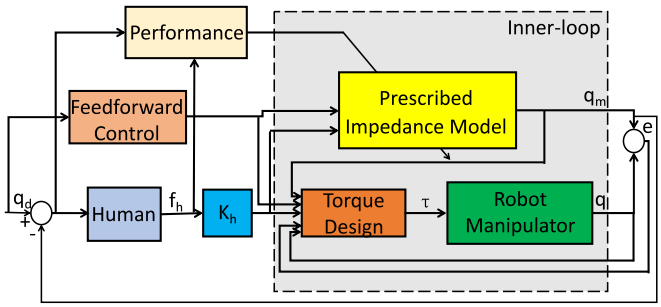

Fig. 1. Control structure with double loops for the human–robot cooperation.

In order to achieve these goals, we design a control structure with double loops, as shown in Fig. 1. The inner one is a robot-oriented inner loop that does not need any knowledge of the performed task. The outer one is a task-oriented loop that contains the dynamics of both the human and the robot and the task information. The design of the inner one is to let the unknown robot behave like a desired impedance model sensed by a human. Therefore, the human interacts with the impedance model. No trajectory needs to be tracked in the inner loop, which can give freedom to combine task information in the outer loop.

Considering the robot-oriented inner-loop design, we can obtain the impedance model learned from the task-oriented outer loop, which can make the tracking error minimized and also help the people to carry out the task with minimum effort. The unknown human dynamics must be taken into account, and the expected performance of the human–robot cooperation system is dependent on the task.

III. Robot-Oriented Control Design

Fig. 1 demonstrates the inner one of the double loop design. The purpose of the control is to make the behavior of the robot matching with a prescribed impedance model. Compared with the existing adaptive impedance control approaches, there exist two main differences in the designed inner-loop control. First, under the comparison with the traditional trajectory tracking control, the designed control does not need the desired trajectory. Namely, the model-matching error between the output of the impedance model and the motion of the robot without any task information can be minimized by the designed control. Second, in the designed control, the developed control torque requires little information of the robot impedance model, which allows us to decouple the robot-oriented control design from the task-oriented control design.

Considering the vector of joint coordinates $q \in \mathbb{R}^n$ with the joint number n, the dynamics of the human–robot cooperation could be considered as

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau + K_h f_h \tag{1}$$

where $M(q) \in \mathbb{R}^{n\times n}$ is the symmetric inertia matrix, $C(q,\dot{q})\dot{q} \in \mathbb{R}^n$ is the vector of Coriolis and centripetal forces, $G(q) \in \mathbb{R}^n$ is the vector of gravitational torques, $\tau \in \mathbb{R}^n$ is the control input vector, and f_h is the human control effort with the human input force gain K_h. In the dynamics (1), we assume that the force sensor mounted on the end effector is used to measure the human force, which is magnified by the gain K_h before exerting to the robot.

The prescribed robot impedance model can be given as

$$\overline{M}\ddot{q}_m + \overline{B}\dot{q}_m + \overline{K}q_m = K_h f_h + \bar{l}(q_d) \equiv l(f_h, q_d) \quad (2)$$

where the desired impedance model's output is denoted by q_m, and the desired inertia, damping, and stiffness parameter matrices are denoted by $\overline{M}$, $\overline{B}$, and $\overline{K}$, respectively. The auxiliary input $\bar{l}(q_d)$ is a trajectory-dependent input and will be defined later.

Let us define the sliding-mode signal as

$$e = q_m - q \quad (3)$$

$$\zeta = \int_o^t e(\iota)d\iota \quad (4)$$

$$r = \dot{e} + \Lambda_1 e + \Lambda_2 \zeta \quad (5)$$

with constant parameters Λ_1 and Λ_2. Then, the control torque can be designed as

$$\begin{aligned} \tau =&\ \tau_1 + K_v r - K_h f_h \\ \tau_1 =&\ \frac{\beta_2}{\beta_1}\sum_{j=1}^{4}\frac{|\hat{c}_j|\phi_j^2 r}{\phi_j\|r\| + \beta_1\omega_j} \end{aligned} \quad (6)$$

where K_v is the control gain, $\beta_1 = \cos^2(\frac{\pi\lambda_{\max}(M)r^T r}{2\varepsilon^2})$, and $\beta_2 = \cos^2(\frac{\pi\lambda_{\min}(M)r^T r}{2\varepsilon^2})$. Define

$$\begin{aligned} \Phi(\ddot{q}_m, \dot{e}, e, \dot{q}, q) =&\ M(q)(\ddot{q}_m + \Lambda_1\dot{e} + \Lambda_2 e) \\ &+ C(q,\dot{q})(\dot{q}_m + \Lambda_1 e + \Lambda_2\zeta) \\ &+ G(q). \end{aligned} \quad (7)$$

Then, it is easy to have the reference signal and the dynamic regressor as

$$\dot{q}_r = \dot{q}_m + \Lambda_1 e + \Lambda_2\zeta \quad (8)$$

$$\Phi(\ddot{q}_r, \dot{q}_r, \dot{q}, q) = M(q)\ddot{q}_r + C(q,\dot{q})\dot{q}_r + G(q). \quad (9)$$

According to [24], there are several unbeknown finite positive constants $c = [c_1, \ldots, c_4]^T$, $c_j \geq 0\,(j = 1, \ldots, 4)$ such that $\forall q \in R^n$, $\forall \dot{q} \in R^n$, $\|M(q)\ddot{q}_r\| \leq c_1\|\ddot{q}_r\|$, $\|C(q,\dot{q})\dot{q}_r\| \leq c_2\|\dot{q}\|\|\dot{q}_r\| + c_3\|\dot{q}_r\|$, $\|G(q)\| \leq c_4$.

Definition 1: A time-varying positive function $\delta(t)$ converges to zero as $t \to \infty$ and satisfies $\lim_{t\to\infty}\int_0^t \delta(s)ds = b < \infty$, with finite constant b. We have many choices for the function $\delta(t)$, i.e., $\delta(t) = 1/(1+t)^2$.

Choose the adaptive updating law as

$$\dot{\hat{c}}_j = -\Gamma_j\left(\sigma_j\hat{c}_j - \frac{\phi_j^2\|r\|^2}{\beta_1(\phi_j\|r\| + \beta_1\omega_j)}\right),\ \ j = 1, \ldots, 4 \quad (10)$$

where Γ_j is a positive constant to be chosen; σ_j satisfies Definition 1, and $\lim_{t\to\infty}\int_0^t \sigma_j(s)ds = b_{1j} < \infty$, with the constant $b_{1j}(j = 1, \ldots, 4)$. $\hat{c}_j$ is the estimation of c_j; ω_j satisfies Definition 1; $\phi = [\phi_1, \ldots, \phi_4]^T$, $\phi_1 = \|\ddot{q}_r\|$, $\phi_2 = \|\dot{q}\|\|\dot{q}_r\|$, $\phi_3 = \|\dot{q}_r\|$, and $\phi_4 = 1$.

Lemma 1: For any x in the interval $|x| < 1$, the following inequality holds:

$$x^2 \leq \cot\frac{\pi}{2}(1-x^2) \leq \pi x^2\csc^2\frac{\pi}{2}(1-x^2). \quad (11)$$

The proof of the lemma can be found in Appendix-A.

Integrating (4) and (7) into (1), we have

$$\begin{aligned} &M(q)(\ddot{q}_m - (\dot{r} - \Lambda_1\dot{e} - \Lambda_2 e)) \\ &\quad + C(q,\dot{q})(\dot{q}_m - (r - \Lambda_1 e - \Lambda_2\zeta)) + G(q) \\ &\quad = \tau + K_h f_h. \end{aligned} \quad (12)$$

Then, we can rewrite (12) as

$$M(q)\dot{r} = -C(q,\dot{q})r + \Phi(\ddot{q}_r, \dot{q}_r, \dot{q}, q) - \tau - K_h f_h. \quad (13)$$

Integrating (6) into (13), we have

$$M(q)\dot{r} = -C(q,\dot{q})r + \Phi(\ddot{q}_r, \dot{q}_r, \dot{q}, q) - K_v r - \tau_1. \quad (14)$$

Theorem 1: Consider the dynamics of robotic arm (1) and utilize the controller (6) and the learning updating law (10), if the initial error satisfies $\|r(0)\| < \varpi$, $\varpi = \frac{\varepsilon}{\sqrt{\lambda_{\max}(M)}}$, where ε is a small positive constant to be decided; then, the following conclusions can be made: 1) all the signals are bounded in the closed-loop system; and 2) $\forall t > 0$, the constraint $\|r(t)\| < \varpi$ holds.

Proof: Consider the BLF candidate as follows:

$$V = \frac{\varepsilon^2}{\pi}\cot\left(\frac{\pi}{2} - \frac{\pi r^T M r}{2\varepsilon^2}\right) + \frac{1}{2}\mathrm{tr}\left(\tilde{c}^T\Gamma^{-1}\tilde{c}\right) \quad (15)$$

where $c = [c_1, \ldots, c_4]^T$, $\tilde{c}_j = \hat{c}_j - c_j$, $j = 1, \ldots, 4$, $\hat{c}_j$ is the estimation of the adaptive parameter c_j, and simply, we replace $M(q)$, $C(q,\dot{q})$, $G(q)$, and $\Phi(\ddot{q}_r, \dot{q}_r, \dot{q}, q)$ with M, C, G, and Φ, respectively. Let $V_1 = \frac{\varepsilon^2}{\pi}\cot(\frac{\pi}{2} - \frac{\pi r^T M r}{2\varepsilon^2})$ and $V_2 = \frac{1}{2}\mathrm{tr}(\tilde{c}^T\Gamma^{-1}\tilde{c})$.

Remark 1: It is obvious that V is positive definite and continuous in the compact set $\Omega_r = \{\|r\| < \varpi\}$, and the inequality $0 \leq \cos^2(\frac{\pi\lambda_{\max}(M)r^T r}{2\varepsilon^2}) \leq \cos^2(\frac{\pi r^T M r}{2\varepsilon^2}) \leq \cos^2(\frac{\pi\lambda_{\min}(M)r^T r}{2\varepsilon^2}) \leq 1$ is satisfied. The BLF candidate (15) is positive and obvious to approach infinity once $\|r\| \to \varpi$, with the initial tracking errors satisfying $\|r(0)\| < \varpi$.

Differentiating V_1, we have

$$\begin{aligned} \dot{V}_1 =&\ \frac{r^T(M\dot{r} + \frac{1}{2}\dot{M}r)}{\cos^2(\frac{\pi r^T M r}{2\varepsilon^2})} \\ =&\ \frac{r^T(\Phi - K_v r - \tau_1)}{\cos^2(\frac{\pi r^T M r}{2\varepsilon^2})} \\ =&\ \frac{-r^T K_v r}{\cos^2(\frac{\pi r^T M r}{2\varepsilon^2})} + \frac{r^T\left(-\frac{\beta_2}{\beta_1}\sum_{j=1}^{4}\frac{|\hat{c}_j|\phi_j^2 r}{\phi_j\|r\| + \beta_1\omega_j} + \Phi\right)}{\cos^2(\frac{\pi r^T M r}{2\varepsilon^2})}. \end{aligned} \quad (16)$$

Consider the following inequality:

$$
\begin{aligned}
&-\frac{r^T\tau_1}{\cos^2\left(\frac{\pi r^T Mr}{2\varepsilon^2}\right)}+\frac{r^T\Phi}{\cos^2\left(\frac{\pi r^T Mr}{2\varepsilon^2}\right)}+\sum_{j=1}^{4}\Gamma_j^{-1}\tilde{c}_j\dot{\tilde{c}}_j\\
&\le\sum_{j=1}^{4}\frac{1}{\beta_1}\|r\|c_j\phi_j+\sum_{j=1}^{4}\Gamma_j^{-1}\tilde{c}_j\dot{\tilde{c}}_i\\
&\quad-\sum_{j=1}^{4}\frac{\beta_2}{\beta_1\cos^2\left(\frac{\pi r^T Mr}{2\varepsilon^2}\right)}\frac{|\hat{c}_j|\phi_j^2\|r\|^2}{\phi_j\|r\|+\beta_1\omega_j}\\
&\le\sum_{j=1}^{4}\frac{1}{\beta_1}\|r\|c_j\phi_j-\sum_{j=1}^{4}\frac{1}{\beta_1}\frac{\hat{c}_j\phi_j^2\|r\|^2}{\phi_j\|r\|+\beta_1\omega_j}\\
&\quad-\sum_{j=1}^{4}\tilde{c}_j\left(\sigma_j\hat{c}_j-\frac{\phi_j^2\|r\|^2}{\beta_1(\phi_j\|r\|+\beta_1\omega_j)}\right)\\
&\le\sum_{j=1}^{4}c_j\omega_j-\sum_{j=1}^{4}\sigma_j\tilde{c}_j\hat{c}_j\\
&\le\sum_{j=1}^{4}c_j\omega_j-\sum_{j=1}^{4}\frac{1}{2}\sigma_j\tilde{c}_j^2+\sum_{j=1}^{4}\frac{1}{2}\sigma_j c_j^2. \qquad (17)
\end{aligned}
$$

Then, considering Lemma 1, we have

$$
\begin{aligned}
\dot{V}&\le\frac{-r^TK_vr}{\cos^2\left(\frac{\pi r^T Mr}{2\varepsilon^2}\right)}-\sum_{j=1}^{4}\frac{1}{2}\sigma_j\tilde{c}_j^2+\sum_{j=1}^{4}c_j\omega_j+\sum_{j=1}^{4}\frac{1}{2}\sigma_j c_j^2\\
&=-(r^TK_vr)\sec^2\left(\frac{\pi r^TMr}{2\varepsilon^2}\right)-\Upsilon\\
&=-(r^TK_vr)\csc^2\left(\frac{\pi}{2}-\frac{\pi r^TMr}{2\varepsilon^2}\right)-\Upsilon\\
&\le-\frac{\varepsilon^2}{\pi}\frac{\lambda_{\min}(K_v)}{\lambda_{\max}(M)}\pi\frac{r^TMr}{\varepsilon^2}\csc^2\left(\frac{\pi}{2}-\frac{\pi r^TMr}{2\varepsilon^2}\right)-\Upsilon\\
&\le-\frac{\varepsilon^2}{\pi}\frac{\lambda_{\min}(K_v)}{\lambda_{\max}(M)}\cot\left(\frac{\pi}{2}-\frac{\pi r^TMr}{2\varepsilon^2}\right)-\sum_{j=1}^{4}\frac{1}{2}\sigma_j\tilde{c}_j^2+\Delta\\
&\le-\rho V+\Delta \qquad (18)
\end{aligned}
$$

where

$$
\rho=\min\left(\frac{\lambda_{\min}(K_v)}{\lambda_{\max}(M)},\min_{j=1,\cdots,4}\left\{\frac{\sigma_j}{\lambda_{\max}(\Gamma_j^{-1})}\right\}\right) \qquad (19)
$$

$$
\Upsilon=\sum_{j=1}^{4}\frac{1}{2}\sigma_j\tilde{c}_j^2-\sum_{j=1}^{4}c_j\omega_j-\sum_{j=1}^{4}\frac{1}{2}\sigma_j c_j^2 \qquad (20)
$$

$$
\Delta=\sum_{j=1}^{4}c_j\omega_j+\sum_{j=1}^{4}\frac{1}{2}\sigma_j c_j^2. \qquad (21)
$$

Multiplying $e^{\rho t}$ on either side of inequality (18), and then integrating this inequality with regard to t, the following inequality can be obtained:

$$
V\le\left(V(0)-\frac{\Delta}{\rho}\right)e^{-\rho t}+\frac{\Delta}{\rho}\le V(0)+\frac{\Delta}{\rho}<\infty. \qquad (22)
$$

Thus, V is bounded. Due to (15) and (22), we can have

$$
\|r\|^2\le\frac{2\varepsilon^2}{\pi\lambda_{\max}(M)}\arctan\frac{\pi}{\varepsilon^2}\left(V(0)+\frac{\Delta}{\rho}\right) \qquad (23)
$$

$$
\|\tilde{c}\|^2\le\frac{2\left(V(0)+\frac{\Delta}{\rho}\right)}{\lambda_{\max}(\Gamma^{-1})}. \qquad (24)
$$

In (23), since $V(0)$ is bounded, $\arctan\frac{\pi}{\varepsilon^2}(V(0)+\frac{\Delta}{\rho})<\frac{\pi}{2}$ holds; we get $\|r\|^2<\frac{\varepsilon^2}{\lambda_{\max}(M)}$, namely $\|r(t)\|<\varpi$. Thus, r and $\tilde{c}$ are both bounded; since the initial errors satisfy $\|r(0)\|<\varpi$, $V(0)$ is bounded. Thus, V is bounded, which implies that r and $\tilde{c}$ are both bounded. First, we assume that there exists some $t=T$ such that $\|r(T)\|$ grows to their respective constraint ϖ. Then, we obtain that V rises to infinity by $r(T)$. However, we have known that (22) demonstrates that V is bounded. According to contradiction, we have that the constraint $\|r(t)\|<\varpi$ holds $\forall t>0$. Since the overall Lyapunov function V is bounded from (22), so all the signals, i.e., r and $\tilde{c}_j\,(j=1,\ldots,4)$ in the closed-loop system are bounded. ■

IV. Task-Oriented Adaptive Impedance Control

In this section, in order to minimize the effort that assists a human operator to fulfill a given task, and tracking errors, the parameters in the prescribed robot impedance model (2) need to be optimized. Without calculating the parameters of the assistive robot impedance model, we first formulate LQR using the optimizing parameters and employ RL to obtain these optimized parameters. On the other hand, in order to minimize the human control effort f_h and optimize the tracking performance depending on the task, by using the task-oriented controller, we need to obtain the optimal values of the prescribed impedance parameters $\overline{B}$, $\overline{K}$, the human operator gain K_h, (or $\overline{M}$ if $K_h=1$), and the auxiliary input $\bar{l}(q_d)$ in (2).

A. Task-Oriented Outer-Loop Control

First, we formulate an LQR problem with the optimal values of $\overline{B}$, $\overline{K}$, and K_h as, by solving an algebraic Riccati equation (ARE), it is easier to obtain these parameters.

Let us define the tracking errors as

$$
e_d=q_d-q_m\in\mathbb{R}^n \qquad (25)
$$

$$
\bar{e}_d=\bar{q}_d-\bar{q}=\left[e_d^T\ \dot{e}_d^T\right]^T\in\mathbb{R}^{2n} \qquad (26)
$$

with $\bar{q}=[q_m^T\ \dot{q}_m^T]^T\in\mathbb{R}^{2n}$, $\bar{q}_d=[q_d^T\ \dot{q}_d^T]^T\in\mathbb{R}^{2n}$.

By using (2), one can obtain

$$
\dot{\bar{q}}=\underbrace{\begin{pmatrix}0&I_{n\times n}\\0&0\end{pmatrix}}_{A_q}\bar{q}+\underbrace{\begin{pmatrix}0\\I_{n\times n}\end{pmatrix}}_{B_q}u \qquad (27)
$$

$$
u=\overline{M}^{-1}(-K_q\bar{q}+K_hf_h)+\overline{M}^{-1}\bar{l}(q_d) \qquad (28)
$$

with $K_q=[\overline{K}\ \overline{B}]\in\mathbb{R}^{n\times 2n}$, and $\overline{B},\overline{K}$, and $\overline{M}$ have been defined in (2).

From [5], it has been verified that the human dynamics are characterized by a simple linear transfer characteristics;

therefore, we choose the impedance model of human arms as

$$(K_d s + K_p) f_h = k_e e_d \tag{29}$$

with unknown gains K_d, K_p, and k_e, which vary from one individual to another and depend on the specific task. It is easy to rewrite (29) as

$$K_d \dot{f}_h + K_p f_h = k_e e_d \tag{30}$$

or equivalently

$$\dot{f}_h = -{K_d}^{-1} K_p f_h + k_e K_{d,0} \dot{e}_d \equiv A_h f_h + E_h \bar{e}_d \tag{31}$$

where $K_{d,0} = [{K_d}^{-1}\ 0] \in R^{n\times 2n}$.

By using (25) and (26), the performance index is defined as

$$J = \int_t^\infty (\bar{e}_d^T Q_d \bar{e}_d + {f_h}^T Q_h f_h + {u_e}^T R u_e) d\tau \tag{32}$$

where $Q_d = Q_d^T > 0$, $Q_h = Q_h^T > 0$, $R = R^T > 0$, and u_e is the feedback input dependent linearly on the tracking error $\bar{e}_d$ and the interaction force f_h, satisfying

$$u_e = K_1 \bar{e}_d + K_2 f_h. \tag{33}$$

The augmented state is obtained as

$$X = \begin{pmatrix} \bar{e}_d \\ f_h \end{pmatrix} \in R^{3n} \tag{34}$$

and the performance index (32) can be rewritten as

$$J = \int_t^\infty (X^T Q X + {u_e}^T R u_e) d\tau \tag{35}$$

with $Q = \mathrm{diag}(Q_d, Q_h)$, $u_e = -KX$, and $K = -[K_1\ K_2]$.

Combining (28) with (26), we have

$$u = u_e + u_d \tag{36}$$

$$u_e = \overline{M}^{-1}(K_q \bar{e}_d + K_h f_h) \tag{37}$$

$$u_d = \overline{M}^{-1}(\bar{l}(q_d) - K_q \bar{q}_d). \tag{38}$$

From (38), we can obtain

$$\bar{l}(q_d) = \overline{M} u_d + K_q \bar{q}_d = \overline{M} {B_q}^{-1}(\dot{\bar{q}}_d - A_q \bar{q}_d) + K_q \bar{q}_d. \tag{39}$$

From (27), in the steady state, we have

$$\dot{\bar{q}}_d = A_q \bar{q}_d + B_q u_d. \tag{40}$$

Differentiating e_d and using (28) and (40), we can obtain

$$\dot{\bar{e}}_d = A_q \bar{e}_d + B_q u_e. \tag{41}$$

Combining (31), (34), and (41), we have

$$\dot{X} = \underbrace{\begin{pmatrix} A_q & 0 \\ E_h & A_h \end{pmatrix}}_{A} X + \underbrace{\begin{pmatrix} B_q \\ 0 \end{pmatrix}}_{B} u_e. \tag{42}$$

Then, the control input u_e can be presented as

$$u_e = \overline{M}^{-1}(K_q \bar{e}_d + K_h f_h) = -\overline{M}^{-1} K X \tag{43}$$

$$K = \begin{pmatrix} K_q & K_h \end{pmatrix} = \overline{M} R^{-1} B^T P \tag{44}$$

where P is the solution to the ARE equation

$$0 = A^T P + PA - PBR^{-1}B^T P + Q. \tag{45}$$

Then, we can obtain the optimal feedback control as

$$u_e = \overline{M}^{-1}\overline{K} e_d + \overline{M}^{-1}\overline{B}\dot{e}_d + \overline{M}^{-1} K_h f_h. \tag{46}$$

From the above equation, we can see that the vector K consists of the impedance model parameters K_q and K_h for both robot and human arms. Therefore, the solution to the formulated LQR problem can give the optimal impedance values of both robot and human arm force.

Theorem 2: Consider the dynamics described by (42). Given a weighted matrix $Q = Q^T > 0$, $R = R^T > 0$, if there exists a symmetric positive-definite matrix $P = P^T > 0$ satisfying the ARE (45), the feedback control

$$u_e = -R^{-1} B^T P X \tag{47}$$

guarantees that all the variables of the closed-loop system are bounded and the tracking performance is achieved.

The proof of Theorem 2 can be found in Appendix-B.

B. Optimal Parameter Using Integral Reinforcement Learning

Considering the problems of optimal control of linear systems [12], [20]–[22], various model-free RL schemes have been proposed without requiring any knowledge of the system dynamics. From (45), we can see that the solution needs the matrix A and the arm model of human; however, it is impossible to acquire those information beforehand. Therefore, in the paper, the off-policy integral reinforcement learning (IRL) algorithm [12], [20], [21] has been borrowed to find the solution of the given LQR problem (45) by an iterative policy iteration algorithm, where two iteration steps are included: 1) the evaluation of the policy; and 2) the improvement of the policy. In the first step, an IRL Bellman equation is used to calculate the value function related to a fixed policy; it does not involve the dynamics of the robot. In the second step, an improved policy is employed using the value obtained in the policy evaluation step.

The IRL Bellman equation proposed in [12], [20], and [21] has been utilized in this paper; it only needs the information provided by the measurement of the system state and an integral of the utility function in finite reinforcement intervals for evaluating a control policy. Generally, the IRL Bellman equation for the given LQR problem including probing noise e_τ, i.e.,

$$\dot{X} = AX + B(u_e + e_\tau) \tag{48}$$

is given for time interval $\Delta t > 0$ as

$$X(t)^T P X(t) + \int_t^{t+\Delta t} [2X(t)^T P B e_\tau] d\tau = \int_t^{t+\Delta t} [X(t)^T Q X(t) + u_e^T R u_e] d\tau + X(t+\Delta t)^T P X(t+\Delta t). \tag{49}$$

Algorithm 1: Online IRL Algorithm.

Initialization: Begin with the control input $u^0 = K_1^0 X$

Evaluation of Policy: Given a control policy u^i, calculate P^i with the off-policy Bellman equation

$$X(t)^T P^i X(t) + \int_t^{t+\Delta t} [2X(t)^T P^i B e_\tau] d\tau = \int_t^{t+\Delta t} [X(t)^T Q X(t) + u_e^T R u_e] d\tau + X(t+\Delta t)^T P^i X(t+\Delta t). \quad (50)$$

Improvement of Policy: the control input is updated with

$$u_e^{i+1} = -R^{-1} B^T P^i X. \quad (51)$$

It should be noted that (49) explicitly contains the probing noise, which is called an off-policy Bellman equation. Therefore, we present the IRL-based algorithm as follows:

C. Online Implementation

To calculate the optimized parameters (matrix P_i) of the cost function associated with the policy K_i, we can write the term $x^T(t)P_i x(t)$ as

$$x^T(t)P_i x(t) = \overline{p}_i^T \overline{x}(t) \quad (52)$$

where $\overline{x}(t)$ denotes the Kronecker product quadratic polynomial basis vector with the elements $\{x_i(t)x_j(t)\}_{i=1,n;j=i,n}$ and $\overline{p} = v(P)$ with $v(\cdot)$. The iteration policy scheme can be performed by

$$\overline{p}_i^T (\overline{x}(t) - \overline{x}(t+T)) = \int_t^{t+T} x^T(\tau)(Q + K_i^T R K_i) x(\tau) d\tau \quad (53)$$

where $\overline{x}(t) - \overline{x}(t+T)$ can be treated as a regression vector, and $\bar{p}_i$ is an unknown parameter vector.

Define $d(\bar{x}(t), K_i)$ as

$$d(\bar{x}(t), K_i) \equiv \int_t^{t+T} x^T(\tau)(Q + K_i^T R K_i) x(\tau) d\tau \quad (54)$$

which can be obtained through comparing the system states over the time interval $[t, t+T]$.

Consider the stable function $V(t)$ and its derivative as

$$\dot{V}(t) = x^T(t) Q x(t) + u^T(t) R u(t). \quad (55)$$

Then, we can obtain $d(\overline{x}(t), K_i) = V(t+T) - V(t)$. The parameter vector $\overline{p}_i$ of the function $V_i(x_t)$ will then yield the matrix P_i, which is obtained by minimizing, through the least-squares approach, the displacement between the target function, $d(\overline{x}(t), K_i)$, and the left side of (53). During the same time interval T, the least-squares solution is obtained as

$$\bar{p}_i = (X X^T)^{-1} X Y \quad (56)$$

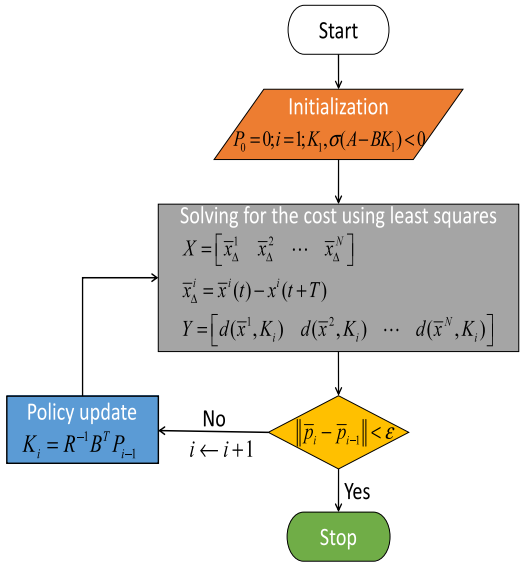

Fig. 2. Continuous-time linear policy iteration algorithm.

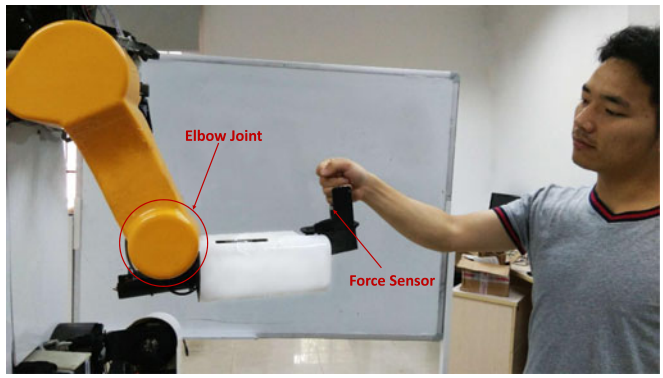

Fig. 3. Robotic exoskeleton and human.

where

$$X = \left[\bar{x}_\Delta^1 \quad \bar{x}_\Delta^2 \quad \cdots \quad \bar{x}_\Delta^N\right] \quad (57)$$

$$\bar{x}_\Delta^i = \bar{x}^i(t) - x^i(t+T) \quad (58)$$

$$Y = \left[d(\bar{x}^1, K_i) \quad d(\bar{x}^2, K_i) \quad \cdots \quad d(\bar{x}^N, K_i)\right]^T. \quad (59)$$

The least-squares problem can be solved online if sufficient data points are collected along a single-state trajectory. The algorithm is illustrated in Fig. 2.

V. Experimental Results

The experiments are performed to verify the proposed human–robot cooperation impedance control by using a developed robotic exoskeleton. Fig. 3 demonstrates the human–robot cooperation experiment with a subject interacting with the robot. The designed exoskeleton is developed according to the kinematic structure of the human upper limb and spans the shoulder and elbow joints. The joint is equipped with Maxon DC flat brushless motor EC45 and a harmonic transmission drive for the robotic system. The motor driver is from Elmo SOLWHI5/60E01. The sampling frequency of the motion

TABLE I
INFORMATION OF THE THREE SUBJECTS

Subject	Age	Height (cm)	Weight (kg)
Subject 1	24	172	55
Subject 2	24	170	62
Subject 3	25	171	66

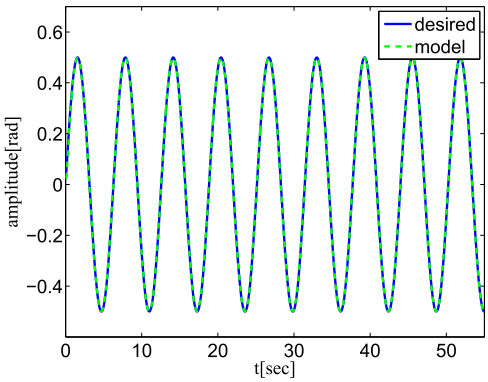

Fig. 4. Actual trajectory and desired trajectory in impedance control of Subject 1.

control system is set as 122 Hz (8.2 ms per loop). Three able-bodied subjects (three males) in the age range of 24–25 years were selected to participate in our experiments. None of them has any knowledge of the human–robot cooperation control. The subject received an explanation of the experiment in advance. Three subjects cooperated with robotic exoskeleton in this experiments. Table I shows the information of the three subjects. The desired trajectory for each joint is chosen as $q_d = 0.5\sin(t)$; the desired mass matrix for the prescribed impedance model is set to the identify of appropriated dimension. The controller parameters are chosen as $\Lambda_1 = 1.8$, $\Lambda_2 = 0.001$, $K_v = 1.2$, $\lambda_{\max} = 0.5$, $\lambda_{\min} = 0.09$, $\varepsilon = 6.0$, and $K_h = 0.03$. We choose $\hat{c}_1 = 0.08$, $\hat{c}_2 = 0.08$, $\hat{c}_3 = 0.08$, and $\hat{c}_4 = 0.08$ as the initial values of adaptive laws. Parameters in adaptive laws are set to $\Gamma_1 = 0.0001, \Gamma_2 = 0.0023, \Gamma_3 = 0.0003, \Gamma_4 = 0.0015$, and $\omega_1 = 10000/(1+0.03t)^2$, $\omega_2 = 6000/(1+0.03t)^2$, $\omega_3 = 440/(1+0.03t)^2$, $\omega_4 = 440/(1+0.03t)^2$, $\sigma_1 = 4000/(1+0.5t)^{\frac{3}{2}}$, $\sigma_2 = 100/(1+t)^{\frac{3}{2}}$, $\sigma_3 = 440/(1+t)^{\frac{3}{2}}$, and $\sigma_4 = 440/(1+t)^{\frac{3}{2}}$.

Figs. 4–14 illustrate the experiment results of three subjects. Figs. 4–6 show the trajectories of the prescribed impedance control versus the desired trajectories in the task-oriented loop. Figs. 7–9 show the trajectories of the robot versus the desired trajectories. From these figures, we can see that at the beginning of the experiments, the actual trajectory does not track the desired trajectory perfectly. However, after learning the optimal impedance model, the human–robot cooperation can be performed perfectly, and the actual trajectories track the desired trajectory with little deviation. The input control torque signal is shown in Fig. 10.

According to these figures, it can be seen that a set of nonoptimal parameters is first used to initialize the desired impedance

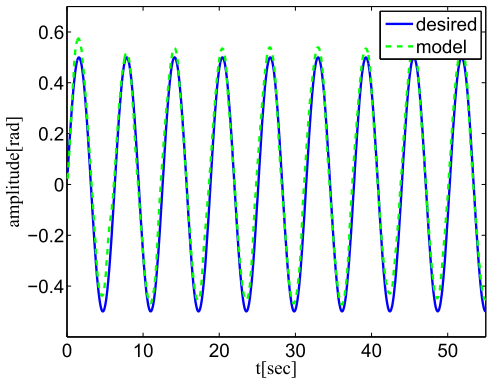

Fig. 5. Actual trajectory and the desired trajectory in impedance control of Subject 2.

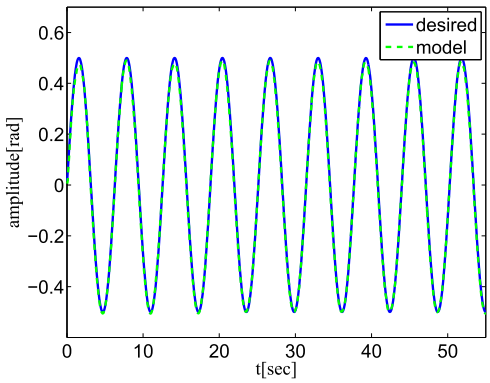

Fig. 6. The actual trajectory and the desired trajectory in impedance control of Subject 3.

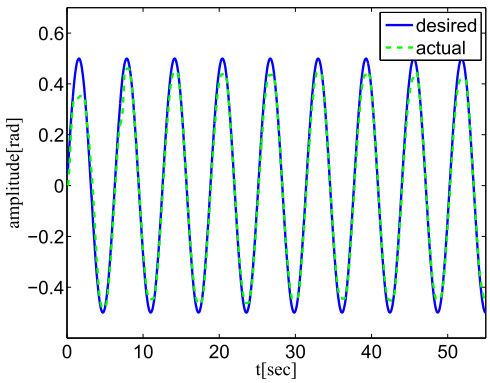

Fig. 7. The trajectory of the robot and the desired trajectory of Subject 1.

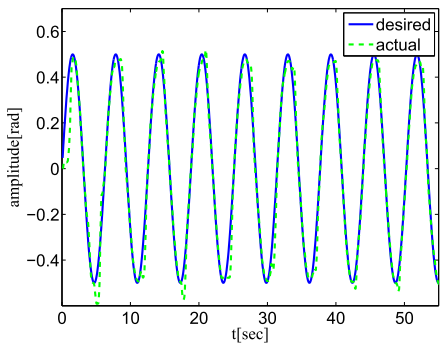

Fig. 8. Trajectory of the robot and the desired trajectory of Subject 2.

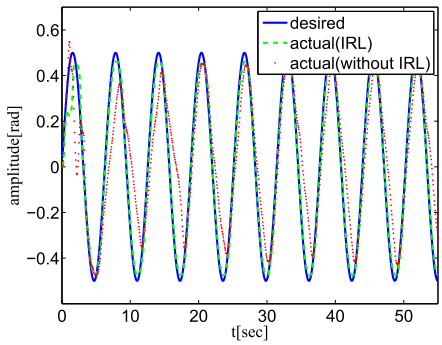

Fig. 9. Trajectory of the robot and the desired trajectory of Subject 3.

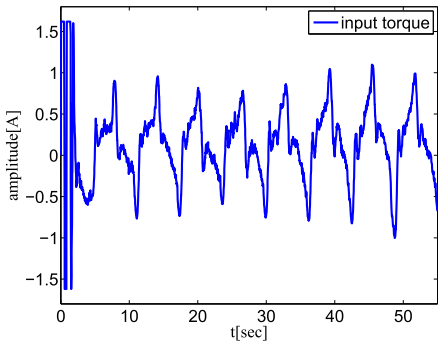

Fig. 10. Control signal.

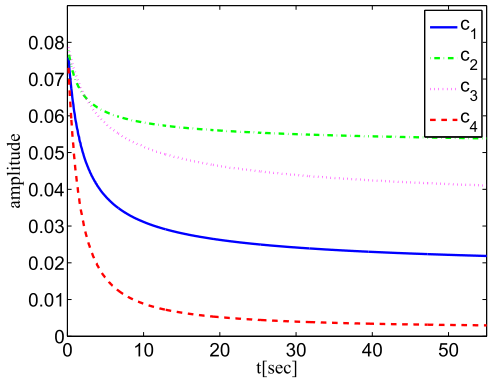

Fig. 11. Adaptive parameters.

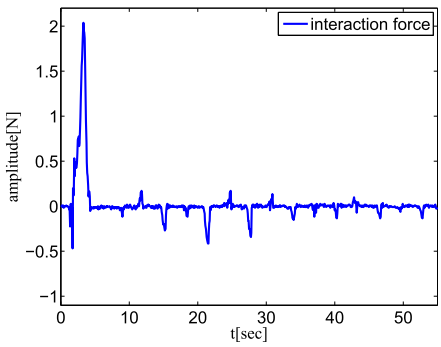

Fig. 12. Interaction force of Subject 1.

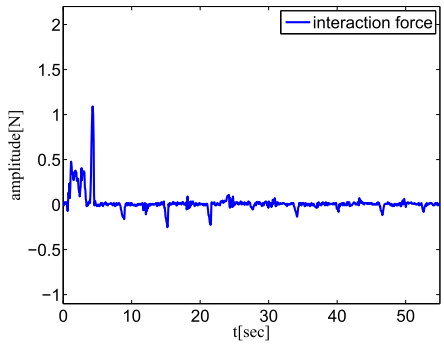

Fig. 13. Interaction force of Subject 2.

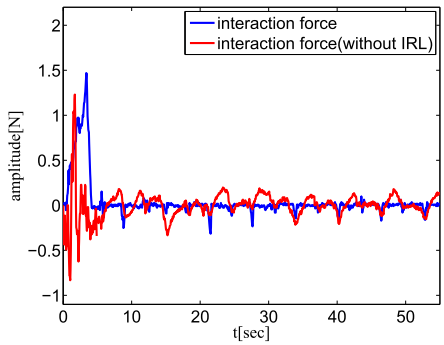

Fig. 14. Interaction force of Subject 3.

model, and the performance of the system is not acceptable. However, after a short time of cooperation between the human and the robot, the optimal parameters of the desired impedance model could be learned by the task-oriented controller; thus, the human–robot interaction system tracks the desired trajectory successfully. The evolution of the human interaction force is shown in Figs. 12–14; it can be seen that the interaction force is reduced as the learning mechanism takes effect, and the task-oriented controller could determine the optimal impedance parameters of the impedance model.

Finally, we have conducted a comparison experiment using subject 3. In the comparison, we consider the adaptive impedance control without the RL technique. The parameters $\overline{M}$, $\overline{B}$, $\overline{K}$, and K_h in the impedance model are chosen beforehand. The parameters $\overline{M} = 1.0, \overline{B} = 6.91, \overline{K} = 10.99$, and $K_h = 0.03$ in the impedance model are chosen beforehand. The comparison experiment results are shown in Figs. 9 and 14; it is obvious that the control performance is worse than the proposed adaptive control using IRL. Furthermore, we have shown that trajectory tracking is achieved without violation of the constraint, and that all closed-loop signals remain bounded, under a requirement on the initial position errors.

VI. Conclusion

In this paper, we presented human–robot cooperation with adaptive robot behavior, which fulfills the human–robot cooperation task and optimizes its performance. The novelties of the work can be summarized as follows.

1) A double-loop control framework was designed to achieve the human–robot cooperation, including robot-oriented control and task-oriented control.
2) In the task-oriented control, the LQR was formulated to acquire an optimal impedance model of an individual arm; then, IRL was proposed to solve the given LQR without information of the human arm model.
3) In the robot-oriented control, a novel BLF-based adaptive impedance control is proposed for the robotic exoskeleton under the consideration of the position constraints, such that the defined constraint region is never transgressed.

With comparison with the previous robot control methods, the proposed robot-oriented loop ignores the knowledge of the task or the prescribed impedance parameters, which separates the robot-oriented control design from the task-oriented design, while in the task-oriented loop, identified impedance parameters by RL have been developed to estimate the optimal parameters of the prescribed impedance parameters, which vary among different subjects and under different interaction forces, to assist the human to perform the task with less effort and optimal performance. Finally, extensive experiments were done to verify the effectiveness of the proposed control in the human–robot cooperative task.

APPENDIX

A. Proof of Lemma 1

Proof:

1) First, in order to prove $\cot\frac{\pi}{2}(1-x^2) \le \pi x^2 \csc^2\frac{\pi}{2}(1-x^2)$, let $g_1(x) = \pi x^2 \csc^2\frac{\pi}{2}(1-x^2) - \cot\frac{\pi}{2}(1-x^2)$ and $g_2(x) = \pi x^2 - \sin\frac{\pi}{2}(1-x^2)\cos\frac{\pi}{2}(1-x^2)$ for $x \in (-1,1)$. Then, we have $g_1(x) = g_2(x)/\sin^2\frac{\pi}{2}(1-x^2)$. The differential of $g_2(x)$ relative to x is given by
$$\frac{dg_2(x)}{dx} = \pi x[2+\cos\pi(1-x^2)]. \tag{60}$$
Since $\frac{dg_2(x)}{dx} < 0$ for $x \in (-1,0)$, $\frac{dg_2(x)}{dx} = 0$ for $x = 0$, $\frac{dg_2(x)}{dx} > 0$ for $x \in (0,1)$, and $g_2(0) = 0$, it is can be obtained that $g_2(x) \ge 0$ for any $x \in (-1,1)$. Considering $\sin^2\frac{\pi}{2}(1-x^2) > 0$ if $|x| < 1$, therefore, $g_1(x) > 0$, and furthermore, the right-hand side of inequality (11) is established when x varies in the interval $(-1,1)$.
2) Next, in order to prove $x^2 \le \cot\frac{\pi}{2}(1-x^2)$, let $g_3(x) = \cot\frac{\pi}{2}(1-x^2) - x^2$; the differential of $g_3(x)$ relative to x is given by
$$\frac{dg_3(x)}{dx} = x\left[\pi\csc^2\frac{\pi}{2}(1-x^2) - 2\right]. \tag{61}$$
Since $\pi\csc^2\frac{\pi}{2}(1-x^2) \ge \pi > 2$, it is obviously that $\pi\csc^2\frac{\pi}{2}(1-x^2) - 2 > 0$ is established under the condition of $|x| < 1$. Then, similar to 1), we have $g_3(x) \ge 0$ for any $x \in (-1,1)$, since $\frac{dg_3(x)}{dx} < 0$ for $x \in (-1,0)$, $\frac{dg_3(x)}{dx} = 0$ for $x = 0$, $\frac{dg_3(x)}{dx} > 0$ for $x \in (0,1)$, and $g_3(0) = 0$. This proves the inequality $x^2 \le \cot\frac{\pi}{2}(1-x^2)$ when $|x| < 1$. ■

B. Proof of Theorem 2

Proof: Let us choose a Lyapunov function as
$$V_2 = \frac{1}{2}X^T PX. \tag{62}$$
Taking the time derivative of V_2 along the dynamics (42), we obtain
$$\begin{aligned}\dot{V}_2 &= \frac{1}{2}(\dot{X}^T PX + X^T\dot{P}X + X^T P\dot{X}) \\ &= \frac{1}{2}\big((X^T A^T + u_e^T B^T)PX + X^T\dot{P}X \\ &\quad + X^T P(AX + Bu_e)\big) \\ &= \frac{1}{2}\big(X^T A^T PX + u_e^T B^T PX \\ &\quad + X^T\dot{P}X + X^T PAX + X^T PBu_e\big) \\ &= \frac{1}{2}X^T(A^T P + \dot{P} + PA)X + X^T PBu_e.\end{aligned} \tag{63}$$
In the steady-state case, $\dot{P} = 0$, considering (45) and (47), we have
$$\dot{V}_2 = -\frac{1}{2}X^T QX \le 0. \tag{64}$$
Therefore, by using the feedback control (47), we have
$$\dot{X} = (A - BR^{-1}B^T P)X \tag{65}$$
being globally exponentially stable about the origin, that is, X is bounded. ■

REFERENCES

[1] V. Duchaine and C. Gosselin, "Safe, stable and intuitive control for physical human-robot interaction," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, 2009, pp. 3383–3388.
[2] 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 *Proc. 11th IEEE Int. Workshop Robot Human Interact. Commun.*, 2002, pp. 500–505.
[3] S. Oh, H. Woo, and K. Kong, "Frequency-shaped impedance control for safe human-robot interaction in reference tracking application," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 6, pp. 1907–1916, Dec. 2014.
[4] S. Suzuki and K. Furuta, "Adaptive impedance control to enhance human skill on a haptic interface system," *J. Control Sci. Eng.*, vol. 2012, pp. 1–10, Jan. 2012.
[5] K. Furuta, Y. Kado, and S. Shiratori, "Assisting control in human adaptive mechatronics-single ball juggling," in *Proc. IEEE Int. Conf. Control Appl.*, Munich, Germany, 2006, pp. 545–550.
[6] S. Suzuki, K. Kurihara, K. Furuta, F. Harashima, and Y. Pan, "Variable dynamic assist control on haptic system for human adaptive mechatronics," in *Proc. 44th IEEE Conf. Decis. Control Eur. Control Conf.*, Seville, Spain, Dec. 2005, pp. 4596–4601.
[7] J. Han, Q. Ding, A. Xiong, and X. Zhao, "A state-space EMG model for the estimation of continuous joint movements," *IEEE Trans. Ind. Electron.*, vol. 62, no. 7, pp. 4267–4275, Jul. 2015.
[8] Z. Li, Z. Huang, W. He, and C.-Y. Su, "Adaptive impedance control for an upper limb robotic exoskeleton using biological signals," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1664–1674, Feb. 2017.
[9] K.-C. Chan, C.-K. Koh, and C. S. G. Lee, "An automatic design of factors in a human-pose estimation system using neural networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 875–887, Jul. 2016.
[10] Y. Kim, T. Oyabu, G. Obinata, and K. Hase, "Operability of joystick type steering device considering human arm impedance characteristics," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 2, pp. 295–306, Mar. 2012.

[11] C. Mitsantisuk, K. Ohishi, and S. Katsura, "Variable mechanical stiffness control based on human stiffness estimation," in *Proc. IEEE Int. Conf. Mechatronics*, Istanbul, Turkey, 2011, pp. 731–736.
[12] F. Dimeas and N. Aspragathos, "Online stability in human-robot cooperation with admittance control," *IEEE Trans. Haptics*, vol. 9, no. 2, pp. 267–278, Apr.–Jun. 2016.
[13] P. D. Labrecque, J.-M. Hache, M. Abdallah, and C. Gosselin, "Low-impedance physical human-robot interaction using an active-passive dynamics decoupling," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 938–945, Jul. 2016.
[14] P. K. Jamwal, S. Hussain, M. H. Ghayesh, and S. V. Rogozina, "Impedance control of an intrinsically compliant parallel ankle rehabilitation robot," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3638–3647, Jun. 2016.
[15] C. Gosselin *et al.*, "A friendly beast of burden: A human-assistive robot for handling large payloads," *IEEE Robot. Autom. Mag.*, vol. 20, no. 4, pp. 139–147, Dec. 2013.
[16] S. Qiu, Z. Li, W. He, L. Zhang, C. Yang, and C.-Y. Su, "Teleoperation control of an exoskeleton robot using brain machine interface and visual compressive sensing," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 58–69, Feb. 2017.
[17] W. Wang *et al.*, "Toward patients' motion intention recognition: dynamics modeling and identification of iLeg—An LLRR under motion constraints," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 980–992, Jul. 2016.
[18] Y.-J. Liu, S. Lu, D. Li, and S. Tong, "Adaptive controller design-based ABLF for a class of nonlinear time-varying state constraint systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2016.2633007.
[19] Y. J. Liu and S. Tong, "Barrier Lyapunov functions for Nussbaum gain adaptive control of full state constrained nonlinear systems," *Automatica*, vol. 76, pp. 143–152, Feb. 2017.
[20] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Trans. Autom. Control*, vol. 59, no. 11, pp. 3051–3056, Nov. 2014.
[21] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, Feb. 2009.
[22] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control," *IEEE Control Syst. Mag.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.
[23] J. Y. Lee, J. B. Park, and Y. H. Choi, "Integral reinforcement learning for continuous-time input-affine nonlinear systems with simultaneous invariant explorations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 916–932, May 2015.
[24] W. He, Y. Chen, and Z. Yin, "Adaptive neural network control of an uncertain robot with full-state constraints," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 620–629, Mar. 2016.
[25] W. He, Y. Dong, and C. Sun, "Adaptive neural impedance control of a robotic manipulator with input saturation,"*IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 3, pp. 334–344, Mar. 2016.

Zhijun Li (M'07–SM'09) received the Ph.D. degree in mechatronics from Shanghai Jiao Tong University, Shanghai, China, in 2002.

From 2003 to 2005, he was a Postdoctoral Fellow with the Department of Mechanical Engineering and Intelligent systems, The University of Electro-Communications, Tokyo, Japan. From 2005 to 2006, he was a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, and Nanyang Technological University, Singapore. Since 2012, he has been a Professor with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests include service robotics, teleoperation systems, nonlinear control, and neural network optimization.

Prof. Li has been the Chair of the Technical Committee on Biomechatronics and Biorobotics Systems of the IEEE Systems, Man and Cybernetics Society since 2016. He is an Editor-at-Large of the *Journal of Intelligent and Robotic Systems* and an Associate Editor of several IEEE TRANSACTIONS. He was the General Chair and Program Chair of the 2016 and 2017 IEEE Conference on Advanced Robotics and Mechatronics, respectively.

Junqiang Liu received the B.S. degree in automation from Henan University of Technology, Henan, China, in 2014. He is currently working toward the master's degree in the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China.

His research interests include adaptive control, reinforcement learning, and rehabilitation exoskeletons.

Zhicong Huang received the B.S. degree in control theory and engineering in 2014 from South China University of Technology, Guangzhou, China, where he is currently working toward the master's degree in the College of Automation Science and Engineering.

His current research interests include exoskeleton robots and EEG signal processing.

Yan Peng received the Ph.D. degree in pattern recognition and intelligent systems from Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, in 2009.

She is currently an Associate Professor with Shanghai University, Shanghai, China, where she is the Executive Dean of the Research Institute of USV Engineering. Her current research interests include modeling and control of unmanned surface vehicles, field robotics, and locomotion systems.

Huayan Pu received the M.Sc. and Ph.D. degrees in mechatronics engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and 2011, respectively.

She is currently an Associate Professor with Shanghai University, Shanghai, China. Her current research interests include modeling, control, and simulation of field robotics and locomotion systems.

Dr. Pu was awarded the Best Paper in Biomimetics at the 2013 IEEE International Conference on Robotics and Biomimetics. She was also nominated as a Best Conference Paper Finalist at the 2012 and 2014 IEEE International Conference on Robotics and Biomimetics.

Liang Ding (S'98–M'10) received the Ph.D. degree in mechanical engineering from Harbin Institute of Technology, Harbin, China, in 2010.

He is currently a Professor with the School of Mechatronics, Harbin Institute of Technology. He has authored or coauthored more than 90 papers published in journals and conference proceedings. His current research interests include planetary exploration rovers and multi-legged robots.

Prof. Ding received the 2011 National Award for Technological Invention of China and the 2009.2013, and 2015 Award for Technological Invention of Heilongjiang Province in China.