



Decentralized approximated optimal control for modular robot manipulations with physical human–robot interaction: a cooperative game-based strategy

Tianjiao An · Xinye Zhu · Bing Ma ·
Mingchao Zhu · Bo Dong

Received: 23 March 2023 / Accepted: 9 February 2024 / Published online: 14 March 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract Major challenges of controlling physical human–robot interaction (pHRI)-oriented modular robot manipulator (MRM) include performance optimization and solving the coupling effect between the human and the robot as well as MRM subsystems. In this paper, a cooperative game-based decentralized optimal control approach is developed for MRMs with pHRI. The joint torque feedback (JTF) technique is utilized to form the MRM dynamic model, then, the major objective of optimal control with pHRI is transformed into approximating Pareto equilibrium by adopting cooperative game governed between the human and the MRM that are regarded as players with different tasks and optimization criteria in interaction process. On the basis of adaptive dynamic programming (ADP) algorithm,

the decentralized approximate optimal control strategy with pHRI is developed by a critic neural network (NN)-based cooperative game manner for solving the coupled Hamilton–Jacobian–Bellman (HJB) equation. The position tracking error under pHRI is verified to be ultimately uniformly bounded (UUB). Experiment results have been presented, which exhibit the superiority of the proposed method.

Keywords Adaptive dynamic programming · Modular robot manipulators · Cooperative game · Decentralized control · Physical human–robot interaction

1 Introduction

Physical human–robot interaction (pHRI) as a hotspot research area makes the design of robotic configuration and control strategy developing towards human-centered criterion [1]. As the collaborative robots (cobots) develops, the collaboration between the human and the robot has turned into a nontrivial research project, such as manufacturing [2], robotic adjuvant therapy (surgery robot [3], habilitation robot [4]) and so on. Although the traditional cobot is with superior security and accuracy, its large size and difficult assembly limit further application. Modular robot manipulator (MRM) is a kind of robot with standard modules and interfaces that can be recombined and assembled its configuration according to different task requirements. Owing to the unique modular reconfigurability,

T. An · X. Zhu · B. Ma · B. Dong (✉)
Department of Control Science and Engineering, Changchun University of Technology, Changchun 130012, China
e-mail: dongbo@ccut.edu.cn

T. An
e-mail: TianjiaoAn@hotmail.com

X. Zhu
e-mail: zhuxinye1995@outlook.com

B. Ma
e-mail: mabing@ccut.edu.cn

M. Zhu
CAS Key Laboratory of On-Orbit Manufacturing and Integration for Space Optics System, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
e-mail: zhumingchao@ciomp.ac.cn

structural flexibility and environment adaptability of MRM, it is with great significance to study the interactive control method of the human and the modular robot.

Human robot collaboration refers to the continuous contact or interaction between bilateral participants (human and robot) to achieve common goals in the same workspace [5]. Therefore, it is with great value for both participants to optimize personal goal by using available self-information [6]. However, in the process of collaboration, each participant needs to consider the decisions of other participants and make corresponding interactive actions [7]. Game theory can be traced back to 1944 [8] as a description of the rational decision-making process between participants [9], and describes the interaction behavior among participants as well as their corresponding strategies as game [10]. For the moment, game theory has been widely applied in large-scale networks [11], unmanned vehicles [12], multi-robot systems [13] and other fields. In the interaction process between a human and an MRM, bilateral participants can be regarded as players in game theory, and the interaction behavior can be described with different cost functions (tasks) and control strategies (optimization criteria) [14]. Cooperative game as a branch of game theory, multiple participants collaborate to optimize their own goals in a dynamic system, and all players need to consider the joint cost function while minimizing their own costs simultaneously, that is Pareto equilibrium, which fits well with the control objective of pHRI [15]. To obtain the optimal strategy for each participant, the Pareto equilibrium solution is attained by solving the coupled Hamilton–Jacobi–Bellman (HJB) equation of the system [16]. However, it is not always feasible to derive the optimal solution of the equation by analytical method for complex issues such as pHRI due to the occurrence of dimension disaster [17]. Therefore, the researchers utilize adaptive dynamic programming (ADP)-based learning method to approximate the ideal equilibrium results.

ADP, as a branch of reinforcement learning, is increasingly popular in the field of optimal control. It employs approximation structures such as neural networks (NNs) to approach nonlinear unknown functions through iterative learning [18]. Optimal control has already been with a wide range of applications, such as discrete-time [19–21], continuous-time [22–24], data-driven [25–27], robotic systems with input/output constraints [28–30], uncertainties distur-

bances [31–33] and multiple sensor failures [34–36]. It is with great importance to select an appropriate control mechanism to satisfy the requirements of system stability and performance for human-MRM collaboration, which involves interconnected or coupling among participants and affects subsystems as well as the whole system. Decentralized control effectively mitigates the computation burden and avoids the problem of data acquisition as well as communication blocking caused by subsystem failure, which only utilizes the local feedback information of each subsystem controller instead of the overall system information [37], and has been widely applied in ADP, such as large-scale system [38], unmanned autonomous vehicle tracking [39], robotic control [40], etc. In the application of MRM, decentralized control strategy is with the ability of decomposing the complex robot system dynamic model into several subsystems with dynamic coupling. The derivation of the local controller is capable to compensate the uncertainty and coupling of the subsystem model that can simplify the complexity of the controller as well as improve the speed of the operation simultaneously.

As noted previously, MRM is particularly appropriate for pHRI tasks. In different implementation phases among interaction tasks, the analysis of a human and a robot bilateral interaction behavior is complicated. In accordance with this issue, cooperative game approach is introduced into pHRI, as well as a human and an MRM are regarded as two participants, that is propitious to pHRI control. Simultaneously, decentralized control strategy fits well with the MRM system due to utilizing only local information of the corresponding subsystem for reducing the complexity of the controller design and ensuring the stability of the robot system. Cooperative game-based decentralized optimal control approach describes the interaction behavior via cost function and control strategy respectively, and then, the coupling effect between the human and the robot has been solved. However, the dynamic coupling effect exists not only between the human and the robot, but also in the MRM subsystem itself, that still affects the control performance and even makes the system unstable, for increases the difficulty of solving the HJB equation, when cooperative game-based decentralized optimal control approach is applied to solve Pareto equilibrium solution of human-MRM cooperative system.

Motivated by above, a cooperative game-based decentralized optimal control approach of MRMs is assessed for pHRI. The coupled MRM subsystem is

handled by using local desired control including control input matrix and local tracking error. Therefore, the assumptions on the upper bound of coupling effect and matching condition are relaxed. The main contribution of this paper is summarized as follow:

1. Unlike the distributed control mechanism in cooperative game [41], a novel decentralized cooperative game-based pHRI control approach is proposed in this paper, which not only solves the coupling effect between the human and the robot, but also deals with the intense coupling of MRM subsystem itself, and obtains the Pareto equilibrium solution of the system by solving the coupled HJB equation. Besides, the strict assumption on satisfying the norm-boundness of IDC is relaxed.
2. Unlike only considering the cooperative relation in MRM [41], a joint cooperative game concerning with human and the robot as well as MRM subsystem itself approximate optimal control method is proposed. The control torque of each module as well as the interaction torque both guarantee optimality via adaptive dynamic programming-based decentralized optimal control approach, which solves the approximate optimal interaction control problem of modular robot system under pHRI.

2 Dynamic analysis and problem formulation

2.1 Dynamic model of MRM

By considering an MRM using the joint torque feedback (JTF) technique, the i th subsystem dynamic model is:

$$I_{im}\gamma_i\ddot{q}_i + \frac{\tau_{is}}{\gamma_i} + f_{ir}(q_i, \dot{q}_i) + I_i(q, \dot{q}, \ddot{q}) = \tau_i + \left[J_i^T f \right]_i, \tag{1}$$

where I_{im} indicates the motor’s moment of inertia, subscript i is the i th joint module subsystem, γ_i represents the gear ratio, q_i denotes the joint position, τ_{is} is the coupled joint torque, $f_{ir}(q_i, \dot{q}_i)$ represents the lumped joint friction, $I_i(q, \dot{q}, \ddot{q})$ denotes the IDC effect among subsystems, τ_i is the control torque, J_i denotes the Jacobi matrix, f represents the human force input that the interaction force exert on the end-effector, and $[\cdot]_i$ means the i th element of the vector. The analysis of the properties is summarized as follows:

(1) The lumped joint friction

The lumped friction term $f_{ir}(q_i, \dot{q}_i)$ mainly includes friction of harmonic drive speed reducer and the motor friction in each joint module, that is expressed as:

$$\begin{aligned} f_{ir}(q_i, \dot{q}_i) &= f_{ib}\dot{q}_i + \left(f_{ic} + f_{is}e^{-f_{i\tau}\dot{q}_i^2} \right) \text{sgn}(\dot{q}_i) \\ &\quad + f_{ip}(q_i, \dot{q}_i) \\ f_{ir}(q_i, \dot{q}_i) &\approx \hat{f}_{ib}\dot{q}_i + \left(\hat{f}_{is}e^{-\hat{f}_{i\tau}\dot{q}_i^2} + \hat{f}_{ic} \right) \text{sgn}(\dot{q}_i) \\ &\quad + f_{ip}(q_i, \dot{q}_i) + Y_i(\dot{q}_i)\tilde{F}_{ir}, \end{aligned} \tag{2}$$

in which

$$Y_i(\dot{q}_i) = [f_{ib} - \hat{f}_{ib}, f_{ic} - \hat{f}_{ic}, f_{is} - \hat{f}_{is}, f_{i\tau} - \hat{f}_{i\tau}]^T, \tag{3}$$

where $f_{ip}(q_i, \dot{q}_i)$ is the position dependency friction term, $f_{ib}, f_{i\tau}$ are viscous and Stribeck friction effect, f_{is}, f_{ic} are static and Coulomb friction parameters, \tilde{F}_{ir} is parameter uncertainty term. Furthermore, $\hat{f}_{ib}, \hat{f}_{ic}, \hat{f}_{is}, \hat{f}_{i\tau}$ are the estimated values of $f_{ib}, f_{ic}, f_{is}, f_{i\tau}$, respectively. \tilde{F}_{ir} is the uncertain friction. $f_{ib}, f_{ic}, f_{is}, f_{i\tau}$ and estimations are bounded because of all physical parameters of practical significance, so that \tilde{F}_{ir} is bounded as $|\tilde{F}_{ir}| \leq b_{iFrm} (m = 1, 2, 3, 4)$, and b_{iFrm} is a known positive constant. Accordingly, one can obtain $Y_i(\dot{q}_i)\tilde{F}_{ir}$ which is given as $|Y_i(\dot{q}_i)\tilde{F}_{ir}| \leq Y_i(\dot{q}_i)b_{iFrm}$. Besides, $|f_{ip}(q_i, \dot{q}_i)| \leq b_{iFp}$, in which b_{iFp} is a known positive constant bound.

(2) The interconnected dynamic coupling

The derivation of interconnected dynamic coupling mainly comes from robot reconfiguration, such as misalignment of the axes. If the IDC effect is ignored, it will reduce the tracking performance of the robot system or even lead the MRM unstable. For $i \geq 3$, IDC can be expressed as:

$$\begin{aligned} I_i(q, \dot{q}, \ddot{q}) &= I_{im} \sum_{j=1}^{i-1} v_{mi}^T v_{lj} \ddot{q}_j \\ &\quad + I_{im} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} v_{mi}^T (v_{lk} \times v_{lj}) \dot{q}_k \dot{q}_j \\ &= I_{im} \sum_{j=1}^{i-1} D_j^i \ddot{q}_j + I_{im} \sum_{j=2}^{i-1} \sum_{k=1}^{j-1} \Theta_{kj}^i \dot{q}_k \dot{q}_j \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j=1}^{i-1} \left[I_{im} \hat{D}_j^i, I_{im} \right] \left[\ddot{q}_j, \tilde{D}_j^i \ddot{q}_j \right]^T + \sum_{j=2}^{i-1} \\
 &\times \sum_{k=1}^{j-1} \left[I_{im} \hat{\Theta}_{kj}^i, I_{im} \right] \left[\ddot{q}_j, \tilde{\Theta}_{kj}^i \dot{q}_k \dot{q}_j \right]^T, \tag{4}
 \end{aligned}$$

in which v_{mi}, v_{lj}, v_{lk} represent the unit vectors along with the i th, j th and k th joint rotation axes, respectively. Accordingly, one can define $\hat{D}_j^i = v_{mi}^T v_{lj}$ and $\Theta_{kj}^i = v_{mi}^T (v_{lk} \times v_{lj})$. Moreover, we also have the relations that $\hat{D}_j^i = D_j^i - \tilde{D}_j^i$ and $\hat{\Theta}_{kj}^i = \Theta_{kj}^i - \tilde{\Theta}_{kj}^i$, in which $\hat{D}_j^i, \hat{\Theta}_{kj}^i$ denote the estimated values of D_j^i, Θ_{kj}^i as well as $\tilde{D}_j^i, \tilde{\Theta}_{kj}^i$ are alignment errors. If $i \leq 2$, the IDC term can be referenced in [42]. According to the definition of v_{mi}, v_{lk}, v_{lj} in (4), we obtain that the vector products are bounded by $|D_j^i| = |v_{mi}^T v_{lj}| < 1$ and $|\Theta_{kj}^i| = |v_{mi}^T (v_{lk} \times v_{lj})| < 1$. Moreover, we also conclude that $I_i(q, \dot{q}, \ddot{q})$ is bounded and the up-bound is given as $|I_i(q, \dot{q}, \ddot{q})| \leq b_{iI}$.

Define state vector $x_i = [x_{i1}, x_{i2}]^T = [q_i, \dot{q}_i]^T$ and the control input $u_i = \tau_i, h_i = [J_i^T f]_i$. One has the i th subsystem state space:

$$\begin{cases} \dot{x}_{i1} = x_{i2} \\ \dot{x}_{i2} = \ell_i(x) + g_i u_i + h_i \end{cases}, \tag{5}$$

where

$$\begin{aligned}
 g_i &= (I_{im} \gamma_i)^{-1} \\
 \ell_i(x) &= -g_i \begin{pmatrix} (\hat{f}_{is} e^{-\hat{f}_{i\tau} x_{i1}^2} + \hat{f}_{ic}) \operatorname{sgn}(x_{i2}) \\ + f_{ip}(x_{i1}, x_{i2}) + \hat{f}_{ib} x_{i2} \\ + Y_i(x_{i2}) \tilde{F}_{ir} + \frac{\tau_{is}}{\gamma_i} + I_i(x, \dot{x}, \ddot{x}) \end{pmatrix}. \tag{6}
 \end{aligned}$$

2.2 Human limb model and motion intention estimation

In physical human robot interaction, the human force is considered as the only external force exerting on the robot end-effector [43]. The pHRI control transfers the human force input into the motion commands of the robot manipulator:

$$-C_H \dot{z} + G_H(z_{Hd} - z) = f, \tag{7}$$

where C_H, G_H are the damper, spring unknown diagonal matrices of the human, z is the robot actual position in Cartesian space which can be calculated as $z(t) = \xi(q), q(t) = [q_1, \dots, q_i, \dots, q_n]^T$ is the position vector in the joint space, $\xi(\cdot)$ is a mapping matrix from joint space to Cartesian space, z_{Hd} denotes the trajectory planned in the human which is referred to as the motion intention of the human and robot.

Remark 1 The up-bound of interaction force f is b_f which guarantees that, once a feasible pHRI is determined, it is possible to guarantee that the MRM will converge to its goal. If the human interaction force is without a finite bound, it is not possible to guarantee the convergence that the position tracking error is UUB.

The human motion intention while interacting with a robot z_{Hd} can be expressed as [44]:

$$z_{Hd} = \Lambda(f, z, \dot{z}), \tag{8}$$

where $\Lambda(\cdot)$ is taken into account as an unknown nonlinear function.

Furthermore, it is known that z_{Hd} is difficult to obtain since the human may change its limb during the collaboration task. Considering the concept of radial basis function NN, the human motion intention while interacting with a robot and its estimation are given by as follows:

$$z_{Hd} = W_x^T S(f, z, \dot{z}) + \varepsilon, \hat{z}_{Hd} = \hat{W}_x^T S(f, z, \dot{z}), \tag{9}$$

where ε is the estimation error, \hat{W}_x denotes the estimated value of ideal weight W_x , and S represents the Gaussian function.

The gradient descent algorithm is used to obtain \hat{W}_x in (9). In order to make the MRM actively and easily move toward its human's intended position and the interaction force f is small, \hat{W}_x is adjusted online using the following cost function $E = \frac{1}{2} \|f^2\|$.

$$\hat{W}_x = -\alpha' \frac{\partial E}{\partial \hat{W}_x} = -\alpha' f G_H S = -\alpha_A f S, \tag{10}$$

where α' is a positive scalar, $\alpha_A = \alpha' G_H$. As G_H is the parameter of human limb dynamics and unknown, it is absorbed by α_A [45].

We can get \hat{W}_x as

$$\hat{W}_x(t) = \hat{W}_x(0) - \alpha_A \int_0^t (f(v) S(v)) dv. \tag{11}$$

Therefore, one can obtain \hat{z}_{Hd} in (9).

The control object consists that optimally ensuring tracking error of MRM systems under pHRI is

UUB. Therefore, a cooperative game-based decentralized optimal control approach is presented in next section.

3 Cooperative game-based decentralized optimal control approach of MRMs

3.1 Problem transformation

In this work, a cooperative game-based decentralized optimal control approach is developed to guarantee that the end-effector of MRM can actively move toward human’s planned intention under pHRI task. In addition, in the cooperative game, the object consists in ensuring the minimization of overall performance index for all the modules as well as the interaction torque between the human and the robot. Therefore, each module in the manipulator forms cooperative game and then cooperates with the human. Hence, in order to facilitate the design of the controller, n modules can be deemed as n players.

We consider that the overall performance situation, in which all players cannot be improved simultaneously with at least one player being proved, as the Pareto equilibrium (cf. [46]).

The estimated joint space motion intention \hat{q} is defined as $\hat{q} = \xi^{-1}(z_{Hd})$. The augmented subsystem of the state space dynamic is:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = \ell(x) + \sum_{p=1}^n G_p u_p + H(x) \end{cases}, \quad (12)$$

where $x = [x_1^T, x_2^T]^T \in R^{2n}$ is the global state of the MRM system, in which the vectors x_1, x_2 are given by $x_1 = [x_{11}, \dots, x_{i1}, \dots, x_{n1}]^T \in R^n$ and $x_2 = [x_{12}, \dots, x_{i2}, \dots, x_{n2}]^T \in R^n$. Moreover, we have $\ell(x) = [\ell_1(x), \dots, \ell_i(x), \dots, \ell_n(x)]^T$, $G_p = [0, \dots, 0, g_p, 0, \dots, 0]^T$, $H(x) = [h_1, \dots, h_i, \dots, h_n]^T$, where $g_p = (I_{pm}\gamma_p)^{-1}$, $p = 1, \dots, n$. The terms $\ell(x)$, G_p and $H(x)$ are the augment of the drift dynamic, control input matrix and interaction torque, respectively.

Then, the optimal control problem of MRM system with pHRI can be transformed into a cooperative game issue, which the joint cost function is defined as follows:

$$V_p(\dot{E}, U, H) = \int_t^\infty (\dot{E}^T Q_b \dot{E} + U^T R_M U + H^T P_M H) d\tau$$

$$= \int_t^\infty \Gamma(\dot{E}, U, H) d\tau, \quad (13)$$

where the position error is $E = x_1 - x_d$ and the velocity error vector means $\dot{E} = x_2 - \dot{x}_d$, $x_d = \hat{q}(t)$ denotes the estimated human motion vectors in joint space, $Q_b, R_M = \text{diag}[R_1, R_2, \dots, R_n]$, $P_M = \text{diag}[P_1, P_2, \dots, P_n]$ are determined positive definite matrixes, and $\Gamma(\dot{E}, U, H)$ represents the utility function, $U = [U_1, U_2, \dots, U_n]$.

Remark 2 The cooperation game issue represented in (13) is considered, where each joint subsystem is responsible for the minimization of the joint cost function. By defining $G = [G_1, G_2, \dots, G_n]$, the cooperation differential game is equivalent to solving the joint cost function (13).

The mentioned MRM system is a second-order system. For obtaining the Hamiltonian function related to the control torque, we select velocity error instead of position error. Then, using the infinitesimal version of (13) with (1) and (12), the Hamiltonian function is expressed as:

$$H_{am}(\dot{E}, U, H) = (\nabla V_p)^T (\ell(x) + GU + H - \ddot{x}_d) + \Gamma(\dot{E}, U, H), \quad (14)$$

where $\nabla V_p(\dot{E}) = \frac{\partial V_p(\dot{E})}{\partial \dot{E}}$ is the partial derivative of $V_p(\dot{E})$. Moreover, the optimal cost function can be defined as:

$$V_p^*(\dot{E}, U, H) = \min_{U, H} \int_t^\infty \Gamma(\dot{E}, U, H) d\tau. \quad (15)$$

According to the stationary condition $\frac{\partial H_{am}}{\partial U} = 0$, $\frac{\partial H_{am}}{\partial H} = 0$, the local decentralized optimal control policy and interaction torque U^*, H^* are given by:

$$U^* = -\frac{1}{2} R_M^{-1} G^T \nabla V_p^*, \quad (16)$$

$$H^* = -\frac{1}{2} P_M^{-1} \nabla V_p^*. \quad (17)$$

Afterwards, by substituting (13), (16) and (17) into the Hamiltonian function (14), the coupled HJB equation is:

$$0 = (\nabla V_p^*)^T (\ell(x) - \ddot{x}_d) - \frac{1}{4} (\nabla V_p^*)^T G R_M^{-1} G^T (\nabla V_p^*) - \frac{1}{4} (\nabla V_p^*)^T P_M^{-1} (\nabla V_p^*) + \dot{E}^T Q_b \dot{E}. \quad (18)$$

As (18) can be deduced to obtain ∇V_p^* , the required Pareto optimal solution can be derived. However, in the framework of the cooperative game, the control issue is transformed to optimally ensure MRM position tracking under the pHRI task and the end-effector can actively move toward its human partner’s intended position. As the coupled HJB equation with other players is difficult to solve, we use critic NN approach to deal with it.

3.2 Approximate solution of the cooperative game-based decentralized optimal control via the implementation of critic NN

Dynamic compensation is a significant part in robotic control. In this study, a compensator controller which comprises dynamic information-based model, local desired control and cooperative game-based optimal control is designed.

$$u_i^* = u_{i1} + u_{ib} + u_{i2}^*, \tag{19}$$

where u_{i1} is dealt with the dynamic model $\ell_i(x)$, u_{ib} denotes local desired control that deals with the coupled term of MRM, and u_{i2}^* is the optimal compensation issue of the uncertainties as well as pHRI tasks.

According to (5), u_{i1} can be designed as:

$$u_{i1} = - \left(- \left(\hat{f}_{is} e^{(-\hat{f}_{i\tau} x_{i2}^2)} + \hat{f}_{ic} \right) \text{sgn}(x_{i2}) \right). \tag{20}$$

The optimal compensation control problem is then transformed into a cooperative game-based decentralized optimal control approach.

The critic NN is developed to approximate the cost function (15):

$$V_p^*(\dot{E}) = W_c^T \phi_c(\dot{E}) + \varepsilon_c, \tag{21}$$

where W_c is the critic NN vector, ε_c is the finite approximate error, and $\phi_c(\dot{E})$ represents activation function.

The gradient of the approximated cost function can be obtained as:

$$\nabla V_p^*(\dot{E}) = \nabla \phi_c(\dot{E}) W_c + \nabla \varepsilon_c, \tag{22}$$

where $\nabla \phi_c(\dot{E}) = \partial \phi_c(\dot{E}) / \partial \dot{E}$ is the gradient of the activation function, and $\nabla \varepsilon_c$ means gradient error.

By substituting (22) into (16) and (17), the optimal control policy and interaction torque are given by:

$$U^* = -\frac{1}{2} R_M^{-1} G^T \left(\nabla \phi_c^T(\dot{E}) W_c + \nabla \varepsilon_c \right), \tag{23}$$

$$H^* = -\frac{1}{2} P_M^{-1} \left(\nabla \phi_c^T(\dot{E}) W_c + \nabla \varepsilon_c \right). \tag{24}$$

According to (23) and (24), we have:

$$u_i^* = -\frac{1}{2} R_i^{-1} G_i^T \left(\nabla \phi_c^T(\dot{E}) W_c + \nabla \varepsilon_c \right), \tag{25}$$

$$h_i^* = -\frac{1}{2} P_i^{-1} \left(\nabla \phi_c^T(\dot{E}) W_c + \nabla \varepsilon_c \right). \tag{26}$$

Substituting (22), (23), (24) into (14), we have:

$$\begin{aligned} H_{am}(\dot{E}, U, H) &= \left(\dot{E}^T Q_b \dot{E} + U^T R_M U + H^T P_M H \right) \\ &+ (\nabla V_p)^T (\ell(x)) + GU + H - \ddot{x}_d \\ &- e_{cH} = 0, \end{aligned} \tag{27}$$

where e_{cH} denotes the residual error.

In order to remove the norm-boundedness assumption on interconnections for decentralized control, the desired states of coupled subsystems are employed to substitute their actual ones. Thus, the interconnection term $\ell_i(x)$ can be expressed as follows:

$$\begin{aligned} \ell_i(x) &= \ell_i(x_i, x_{bd}) + \Delta \ell_i(x, x_{bd}), \\ u_{bd} &= G_b^{-1} (\dot{x}_{b2d} - \ell_b(x_d)), b \neq i. \end{aligned} \tag{28}$$

where x_{bd} is the desired state of the coupled subsystems with $b = 1, \dots, i - 1, i + 1, \dots, n$, and $\Delta \ell_i(x, x_{bd})$ denotes the substitution error. Since the interconnection satisfies the global Lipschitz condition, it thus implies:

$$\|\Delta \ell_i(x, x_{bd})\| \leq \sum_{b=1, b \neq i}^n c_{ib} B_b, \tag{29}$$

where $B_b = \|x_b - x_{bd}\|$, and $c_{ib} \geq 0$ is an unknown global Lipschitz constant.

The estimated optimal cost function is given by:

$$\hat{V}_p^*(\dot{E}) = \hat{W}_c^T \phi_c(\dot{E}). \tag{30}$$

Based on (25), (26) and (30), the approximate optimal control is expressed as:

$$\hat{u}_{i2}^* = -\frac{1}{2} R_i^{-1} G_i^T \nabla \phi_c^T(\dot{E}) \hat{W}_c. \tag{31}$$

$$\hat{h}_i^* = -\frac{1}{2} P_i^{-1} \nabla \phi_c^T(\dot{E}) \hat{W}_c. \tag{32}$$

Remark 3 According to (28), the interconnected MRM is decoupled via the substitute technique. This facilitates the smooth implementation of the decentralized control via removing the norm-boundedness assumption

on interconnections. Therefore, the term $\phi_c(\dot{E})$ in (25) and (26) is converted to $\phi_c(\dot{e}_i)$ in (31) and (32). The similar technique can be referenced in [47].

Remark 4 [48] As the interaction torque H can be obtained by the measurable human force input via six-axis force sensor, the difference between h_i and \hat{h}_i^* in (32) is able to assess the conflict of the goal between the robot and the human, such as the clash with MRM and human in coordinate operation task.

Denote $L(x) = [\ell_1(x_1, x_{bd}), \ell_2(x_2, x_{bd}), \dots, \ell_n(x_n, x_{bd})]$. According to (27), the approximated Hamiltonian function is given by:

$$\begin{aligned} \hat{H}_{am}(\dot{E}, \hat{U}, \hat{H}) &= (\dot{E}^T Q_b \dot{E} + \hat{U}^T R_M \hat{U} + \hat{H}^T P_M \hat{H}) \\ &\quad + (\nabla V_p)^T (L(x) + G\hat{U} + \hat{H} - \ddot{x}_d) \\ &= e_c. \end{aligned} \tag{33}$$

The approximated Hamiltonian error function e_c is defined as follows:

$$e_c = \hat{H}_{am} - H_{am}, \tag{34}$$

where $e_c = \hat{H}_{am}$ is obtained using (27), (33).

When defining the vector estimate error $\tilde{W}_c = W_c - \hat{W}_c$, it can be deduced that $e_c = e_{cH} - \tilde{W}_c^T \nabla \phi_c(\dot{E}) \dot{E}$, by combining (27), (33) with (34). Base on gradient decent algorithm, define residual error function $E_c = \frac{1}{2} e_c^2$, which is minimized for adjusting the critic NN weights. The update law is designed as:

$$\begin{aligned} \dot{\tilde{W}}_c &= -\varsigma \frac{1}{(\xi^T \xi + 1)^2} \frac{\partial E_c}{\partial \tilde{W}_c} \\ &= -\varsigma \frac{\xi}{(\xi^T \xi + 1)^2} (e_{cH} - \xi^T \tilde{W}_c), \end{aligned} \tag{35}$$

where ς is the updated rate of the critic NN. ξ is denoted as $\nabla \phi_c(\dot{E}) \dot{E}$, and a positive constant ξ_L and $\|\xi\| \leq \xi_L$ is assumed.

Assumption 1 (*Persistently exciting condition* [49]) Signal $\tilde{\xi} = \frac{1}{(\xi^T \xi + 1)^2}$ is persistently excited, for $\forall t \in [0, \infty)$, then, there are with positive constants κ_1, κ_2, T , such that following equation is satisfied:

$$\kappa_1 I \leq \int_t^{t+T} \tilde{\xi} \tilde{\xi}^T dz \leq \kappa_2 I. \tag{36}$$

Theorem 1 *Considered the selected cost function which is approximated by (21), with an ideal weight W_c , and the estimated cost function given by (30) with approximated weight \hat{W}_c , if the update law of critic NN is denoted as (35), then the weight approximation error is UUB.*

Proof Choose the Lyapunov function:

$$L_a = \frac{1}{2} \tilde{W}_c^T \tilde{W}_c. \tag{37}$$

The time derivative of $L_a(t)$ is:

$$\begin{aligned} \dot{L}_a &= \tilde{W}_c^T \dot{\tilde{W}}_c = \varsigma \tilde{W}_c^T \frac{\xi}{(\xi^T \xi + 1)^2} (e_{cH} - \xi^T \tilde{W}_c) \\ &= \frac{\varsigma \tilde{W}_c^T \xi e_{cH}}{(\xi^T \xi + 1)^2} - \frac{\varsigma \tilde{W}_c^T \xi \xi^T \tilde{W}_c}{(\xi^T \xi + 1)^2} \\ &\leq -\left(\varsigma - \frac{1}{4}\right) \frac{\lambda_{\min}(\xi \xi^T) \|\tilde{W}_c\|^2}{(\xi^T \xi + 1)^2} + \frac{\varsigma^2 e_{cH}^2}{(\xi^T \xi + 1)^2}, \end{aligned} \tag{38}$$

when $\varsigma > \frac{1}{4}$ and \tilde{W}_c is outside the compact set $\Omega_a = \left\{ \tilde{W}_c : \|\tilde{W}_c\| \leq \sqrt{\frac{e_{cH}^2}{(\varsigma - \frac{1}{4}) \lambda_{\min}(\xi \xi^T)}} \right\}$, the weight approximation error of critic NN is UUB. This completes the proof. \square

Based on the compensation-based control policy (20), desired control policy (28) and the optimal control policy (31), \hat{u}_i^* is given by:

$$\begin{aligned} \hat{u}_i^* &= u_{i1} + u_{id} + \hat{u}_{i2}^* \\ &= -\left(-\left(\hat{f}_{is} e^{(-\hat{f}_{i\tau} x_{i2}^2)} + \hat{f}_{ic} \right) \text{sgn}(x_{i2}) \right) \\ &\quad - \left(-\hat{f}_{ib} x_{i2} - g_i^{-1} \ddot{x}_{id} - \frac{\tau_{is}}{\gamma_i} \right) \\ &\quad + G_i^{-1} (\dot{x}_{i2d} - \ell_i(x_d)) - \frac{1}{2} R_{ii}^{-1} G_i^T \nabla \phi_{ic}^T(\dot{e}) \hat{W}_{ic}. \end{aligned} \tag{39}$$

Remark 5 The proposed decentralized approximated optimal control fits well with the MRM due to utilize only local information of the corresponding subsystem for reducing the complexity of the controller design and ensuring the stability of the robot system. The developed scheme not only solves the coupling effect between the human and the robot, but also in the MRM itself.

The algorithm of human motion intention estimation-based decentralized optimal control is shown as below.

Algorithm 1 Human motion intention estimation-based decentralized optimal control

- 1: Calculate the robot actual position in Cartesian space using $z(t) = \xi(q)$.
 - 2: Obtain the human motion intention via $\hat{z}_{Hd} = \hat{W}_x^T S(f, z, \dot{z})$.
 - 3: Calculate the position tracking error by $E = x_1 - \xi^{-1}(\hat{z}_{Hd})$.
 - 4: Obtain the decentralized approximate optimal control according to the update law $\hat{W}_c = -\varsigma \frac{\xi}{(\xi^T \xi + 1)^2} (e_{cH} - \xi^T \tilde{W}_c)$.
-

Theorem 2 Given an MRM joint subsystem dynamic model developed in (1), and the state space shown in (12), the position tracking error of the closed-loop robot manipulator system is UUB under the pHRI task, by the presented cooperative game-based decentralized optimal control policy derived in (39).

Proof Choose $V_p^*(\dot{E}) = L_b(t)$ as Lyapunov function, then derivative it as follows:

$$\dot{L}_b(t) = (\nabla V_p^*)^T (L(x) + GU + H - \ddot{x}_d). \quad (40)$$

By considering the coupled HJB equation formulated in (18), it yields:

$$\begin{aligned} (\nabla V_p^*)^T (L(x) - \ddot{x}_d) &= \frac{1}{4} (\nabla V_p^*)^T G R_M^{-1} G^T (\nabla V_p^*) \\ &\quad + \frac{1}{4} (\nabla V_p^*)^T P_M^{-1} (\nabla V_p^*) \\ &\quad - \dot{E}^T Q_b \dot{E}. \end{aligned} \quad (41)$$

Combining (41) into (40), we obtain:

$$\begin{aligned} \dot{L}_b(t) &= -\dot{E}^T Q_b \dot{E} + \frac{1}{4} (\nabla V_p^*)^T G R_M^{-1} G^T (\nabla V_p^*) \\ &\quad + \frac{1}{4} (\nabla V_p^*)^T P_M^{-1} (\nabla V_p^*) + (\nabla V_p^*)^T (GU + H). \end{aligned} \quad (42)$$

Considering (42), one obtains:

$$\begin{aligned} \dot{L}_b(t) &= -\dot{E}^T Q_b \dot{E} - (\nabla V_p^*)^T \\ &\quad \times (G(U^* - U) + (H^* - H)) \end{aligned}$$

$$\begin{aligned} &+ \frac{1}{4} (\nabla V_p^*)^T G R_M^{-1} G^T (\nabla V_p^*) \\ &+ \frac{1}{4} (\nabla V_p^*)^T P_M^{-1} (\nabla V_p^*). \end{aligned} \quad (43)$$

Then, substituting (40) into (43), one yields:

$$\begin{aligned} \dot{L}_b(t) &= -\dot{E}^T Q_b \dot{E} + \frac{1}{4} (\nabla V_p^*)^T \left(G R_M^{-1} G^T (\nabla V_p^*) \right. \\ &\quad \left. + P_M^{-1} (\nabla V_p^*) \right) \\ &\quad + \frac{1}{2} \left(\nabla \phi_c^T W_c (\dot{E}) + \nabla \varepsilon_c \right)^T \\ &\quad \times \left(G R_M^{-1} \left(G^T \nabla \phi_c^T \tilde{W}_c + G^T \nabla \varepsilon_c \right) \right. \\ &\quad \left. + P_M^{-1} \left(\nabla \phi_c^T \tilde{W}_c + \nabla \varepsilon_c \right) \right) \\ &= -\dot{E}^T Q_b \dot{E} + \Pi_J, \end{aligned} \quad (44)$$

in which Π_J has up-bound:

$$\Pi_J \leq \left\| \begin{aligned} &\frac{1}{4} (\nabla V_p^*)^T \left(G R_M^{-1} G^T (\nabla V_p^*) \right. \\ &\quad \left. + P_M^{-1} (\nabla V_p^*) \right) \\ &+ \frac{1}{2} \left(\nabla \phi_c^T W_c (\dot{E}) + \nabla \varepsilon_c \right)^T \\ &\left(G R_M^{-1} \left(G^T \nabla \phi_c^T \tilde{W}_c + G^T \nabla \varepsilon_c \right) \right. \\ &\quad \left. + P_M^{-1} \left(\nabla \phi_c^T \tilde{W}_c + \nabla \varepsilon_c \right) \right) \end{aligned} \right\| \leq \pi_J, \quad (45)$$

where π_J denotes computable constant.

With (45), $\dot{L}_b(t)$ has upper bound by:

$$\dot{L}_b(t) \leq -\dot{E}^T Q_b \dot{E} + \pi_J \leq -\lambda_{\min}(Q_b) \|\dot{E}\|^2 + \pi_J. \quad (46)$$

If \dot{E} lies outside:

$$\Omega = \left\{ \dot{E} : \|\dot{E}\| \leq \sqrt{\frac{\pi_J}{\lambda_{\min}(Q_b)}} \right\}, \quad (47)$$

Equation (40) is negative. Hence, we have $\dot{L}_b(t) < 0$ for any $\dot{E} \neq 0$ when (47) is satisfied. The position tracking error under pHRI task is proved to be UUB under (39). This completes the proof. \square

4 Experiments

4.1 Experiment setup

The proposed control method is verified by two degrees of freedom MRM experimental platform (see Fig. 1).

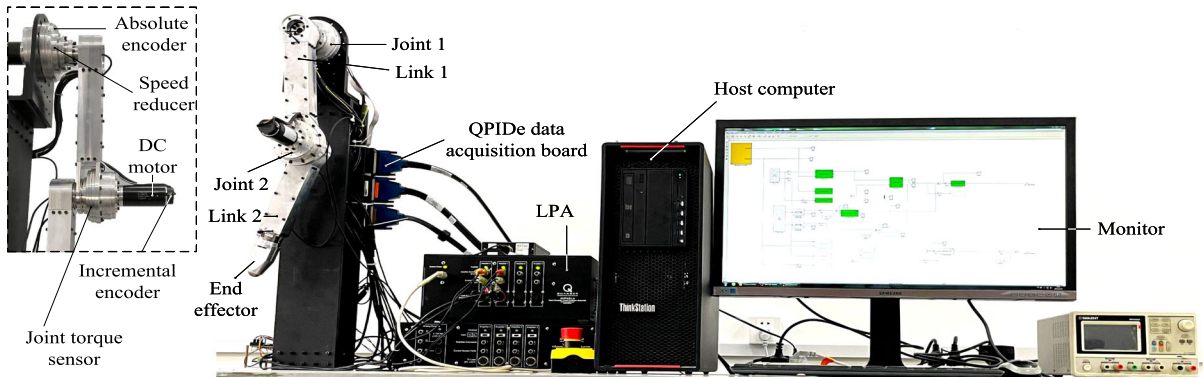


Fig. 1 Experimental platform setup

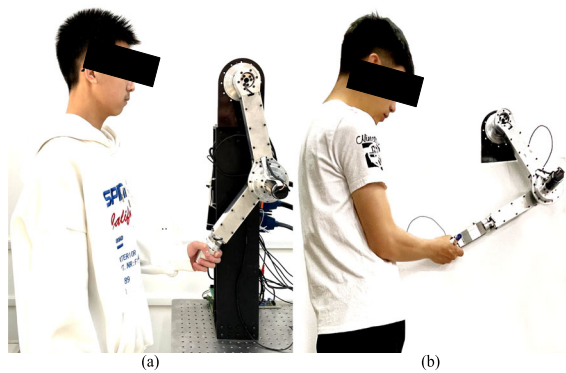


Fig. 2 Experimental with physical human robot interaction **a** handshaking tasks **b** writing-aid tasks

The situation of pHRI that is handshaking as well as writing-aid tasks with MRM is considered for simulating robotic adjuvant therapy and aided daily life (cf. Fig. 2). An experiment is performed to determine if the requirements of position tracking performance and control torque as well as interaction torque optimization are met under pHRI. The related dynamic model and control design are given by Table 1.

4.2 Experimental results

The experimental results demonstrate the performance of position tracking, tracking error, control torque, interaction torque and critic NN weight. Two types of control methods are used to verify the validity of the proposed method: the existed learning-based optimal control without cooperative game method (e.g. critic only policy iteration-based zero-sum neuro-optimal

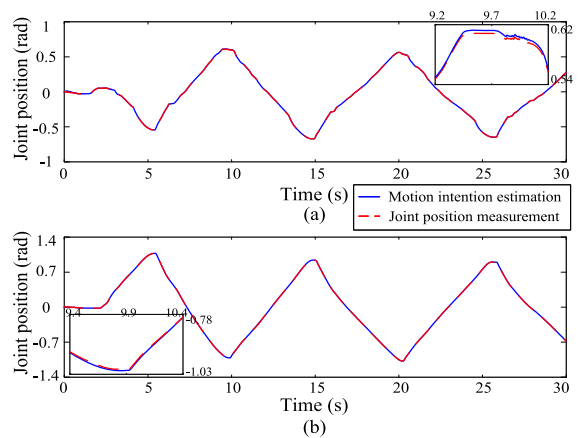


Fig. 3 Motion intention estimation and position tracking curves under handshaking tasks via the proposed control approach **a** joint one **b** joint two

control [50], compensator-critic-based non-zero-sum optimal control [51], discounted guaranteed cost control [52]), and the proposed cooperative game-based decentralized optimal control approach. The two control methods are both used in handshaking as well as writing-aid pHRI tasks.

(1) Position tracking performance

Figures 3 and 5 present the position tracking and tracking error curves in joint space under handshaking tasks, obtained by the existed learning-based optimal control and proposed cooperative game-based decentralized optimal control approach, respectively. It can be seen that the amplitude of position tracking error under the proposed cooperative game-based decentralized optimal control approach is fairly smaller than the one under the existed control method, as the proposed

Table 1 Parameter definition

Parameter type	Name	Value	Name	Value	Name	Value
Model parameters	\hat{f}_{ib}	0.012 Nms/rad	\hat{f}_{ic}	0.03 Nm	\hat{f}_{is}	0.04 Nm
	$\hat{f}_{i\tau}$	$20 \text{ s}^2/\text{rad}^2$	γ_i	100	I_{im}	0.12 kgcm^2
Control parameters	b_{iFr1}	0.03 Nms/rad	b_{iFr2}	0.06 mNm	b_{iFr3}	0.08 mNm
	b_{iFr4}	$50 \text{ s}^2/\text{rad}^2$	b_{iFp}	0.5	b_{iI}	4.9
	b_f	1 N	n	2	ζ	0.82
	α'	0.92	π_J	4.25	ξ_L	3.5
	P_M	$0.65I$	Q_b	$0.75I$	R_M	$0.98I$

Table 2 Performance comparisons

Control method	Mean absolute value of position error	Mean absolute value of control torque	Mean absolute value of interaction torque
The existed method with handshaking (Joint 1)	$4.26\text{e}-3$ rad	2.09 Nm	0.10 Nm
The proposed method with handshaking (Joint 1)	$2.67\text{e}-3$ rad	1.72 Nm	0.07 Nm
The existed method with handshaking (Joint 2)	$6.02\text{e}-3$ rad	3.06 Nm	0.09 Nm
The proposed method with handshaking (Joint 2)	$3.28\text{e}-3$ rad	2.53 Nm	0.07 Nm
The existed method with writing-aid (Joint 1)	$3.85\text{e}-3$ rad	2.45 Nm	0.09 Nm
The proposed method with writing-aid (Joint 1)	$2.01\text{e}-3$ rad	1.86 Nm	0.06 Nm
The existed method with writing-aid (Joint 2)	$6.35\text{e}-3$ rad	0.46 Nm	0.12 Nm
The proposed method with writing-aid (Joint 2)	$3.77\text{e}-3$ rad	0.39 Nm	0.09 Nm

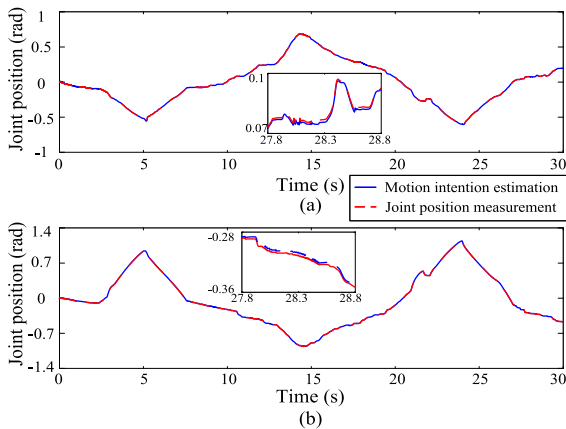


Fig. 4 Motion intention estimation and position tracking curves under writing-aid tasks via the proposed control approach **a** joint one **b** joint two

cooperative game-based decentralized optimal control approach accurately solves the human motion intention estimation problem. Figures 4 and 6 are the position tracking and error curves in joint space under the writing-aid tasks. The similar results can be obtained.

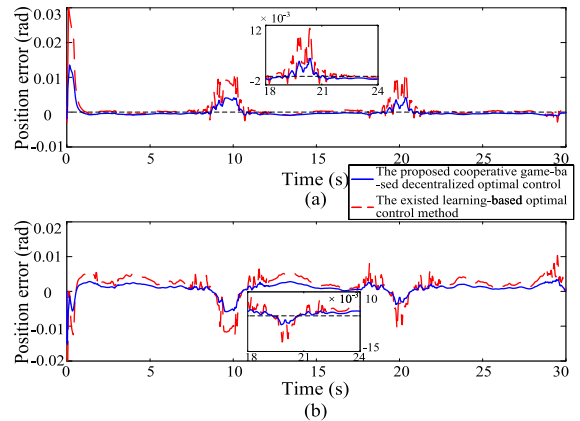


Fig. 5 Position tracking error curves in joint space under handshaking tasks via the existed and the proposed control method **a** joint one **b** joint two

(2) Control torque

Figure 7 shows the control torque curves under handshaking tasks using the existed learning-based optimal control method and the proposed cooperative game-based decentralized optimal control approach. Further-

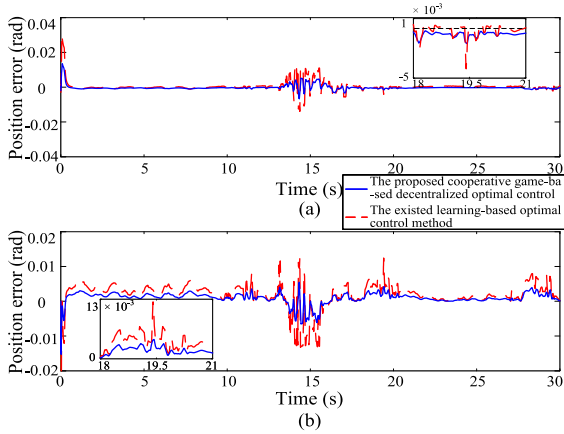


Fig. 6 Position tracking error curves in joint space under writing-aid tasks via the existed and the proposed control method **a** joint one **b** joint two



Fig. 7 Control torque curves under handshaking tasks via the existed and the proposed control method **a** joint one **b** joint two

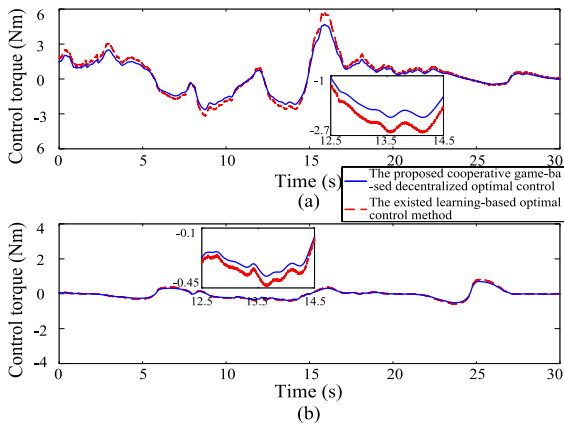


Fig. 8 Control torque curves under writing-aid tasks via the existed and the proposed control method **a** joint one **b** joint two

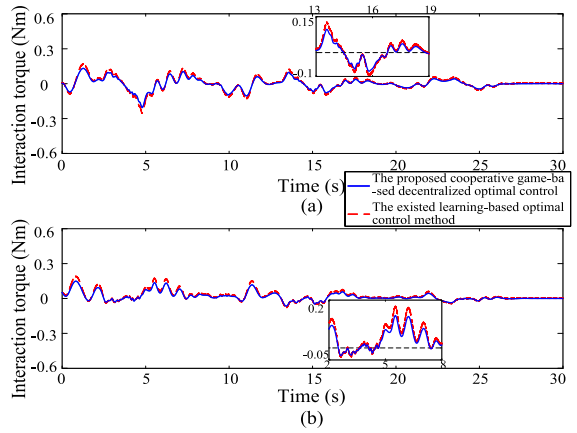


Fig. 9 Interaction torque curves under handshaking tasks via the existed and the proposed control method **a** joint one **b** joint two

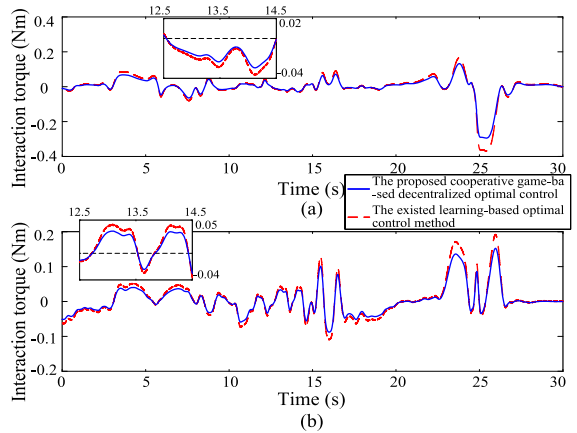


Fig. 10 Interaction torque curves under writing-aid tasks via the existed and the proposed control method **a** joint one **b** joint two

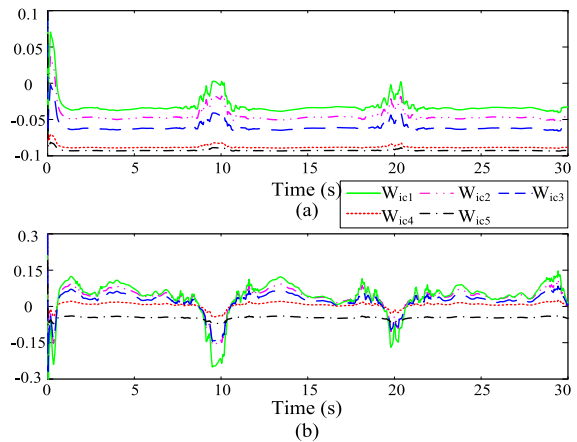


Fig. 11 NN curves under handshaking tasks via the proposed cooperative game-based decentralized optimal control approach **a** joint one **b** joint two

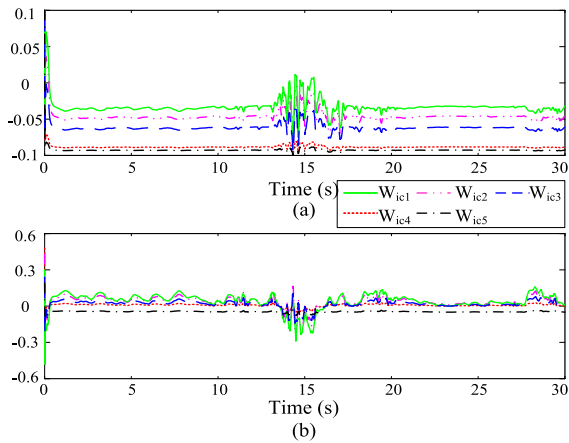


Fig. 12 NN curves under writing-aid tasks via the proposed cooperative game-based decentralized optimal control approach **a** joint one **b** joint two

more, the control torque curves show serious chattering effects under the existed control method, which may reduce the precision of the position tracking. The control torque curves of the writing-aid tasks are illustrated in Fig. 8. The effectiveness of the proposed cooperative game-based decentralized optimal control approach is demonstrated, which is attributed to the approximate optimal control that realizes output torque optimization.

(3) Interaction torque

Figure 9 shows the interaction torque curves under handshaking tasks by the existed learning-based optimal control method and the proposed cooperative game-based decentralized optimal control approach. The existed learning-based optimal control method is without coupling compensation, therefore the interaction torque is a touch of quavering and massive than the proposed method. It can be seen that the interaction torque is less than or equal to 0.25 Nm for both joints, under the cooperative game-based decentralized optimal control approach, and without strong chattering phenomenon. This facilitates the accurate implementation of the cooperative game technique, which guarantees the comfort and security of the human demonstrated from the interaction torque. Figure 10 provides the interaction torque curves under writing-aid tasks via the proposed cooperative game-based decentralized optimal control approach. Making the full use of cooperative game and decentralized control mechanism, the coupling effect between the MRM and the human

has been worked out faultlessly. The writing process for human perceives gentle and effortless. Besides, on account of the approximate optimal control approach, the interaction torque is minor than the existed learning-based optimal control method.

(4) Critic NN weight

Figures 11 and 12 show the critic NN weight curves under handshaking and writing-aid tasks obtained by the proposed cooperative game-based decentralized optimal control approach, respectively. The cooperative game-based decentralized optimal control approach can be obtained using the converged weights (cf. Figs. 7 and 8). Thus, the weight can reflect the human intention in real time.

Based on the experimental results, the closed-loop MRM systems under the pHRI task have better performance than the existed methods in terms of position tracking, control torque, and interaction torque under the proposed cooperative game-based decentralized optimal control approach (cf. Table 2).

5 Conclusions

This paper proposed a cooperative game-based decentralized optimal control approach for MRMs under pHRI tasks. Based on the differential game strategy, the optimal control problem of closed-loop robotic systems is transformed into a cooperative game issue between the human and the MRM. Using the ADP algorithm, the cost function is developed by critic NN and implemented to solve the coupled HJB equation, which facilitates the acquisition of Pareto equilibrium solutions. The position tracking error is UUB under the pHRI task. Finally, the performed experiments demonstrate the efficiency of the proposed method. In pHRI, human's comfort via subjective scales is with great importance. Inspired by [53, 54], the quantification of human's comfort will be our future work.

Acknowledgements This work is supported by the National Natural Science Foundation of China (62173047), the Scientific Technological Development Plan Project in Jilin Province of China (YDZJ202201ZYTS508).

Funding The authors have not disclosed any funding.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship and publication of this article.

References

1. Bednarczyk, M., Omran, H., Bayle, B.: EMG-based variable impedance control with passivity guarantees for collaborative robotics. *IEEE Robotics Autom. Lett.* **7**(2), 4307–4312 (2022)
2. Villani, V., Pini, F., Leali, F., et al.: Survey on human-robot collaboration in industrial settings: safety, intuitive interfaces and applications. *Mechatronics* **55**, 248–266 (2018)
3. Bogue, R.: Rehabilitation robots. *Ind. Robot An Int. J.* **45**(3), 301–306 (2018)
4. Weber, L., Stein, J.: The use of robots in stroke rehabilitation: A narrative review. *NeuroRehabilitation* **43**(1), 99–110 (2018)
5. Wang, Q., Liu, D., Carmichael, M., et al.: Computational model of robot trust in human co-worker for physical human–robot collaboration. *IEEE Robotics Autom. Lett.* **7**(2), 3146–3153 (2022)
6. Cappello, D., Mylvaganam, T.: Distributed differential games for control of multi-agent systems. *IEEE Trans. Control Netw*
7. Jin, Z., Liu, A., Zhang, W., et al.: A learning based hierarchical control framework for human–robot collaboration. *IEEE Trans. Autom. Sci. Eng*
8. Von Neumann, J., Morgenstern, O.: *Theory of Games and Economic Behavior*, 2nd edn. Princeton Univ, Princeton, NJ, USA (1947)
9. Nekouei, E., Nair, G.N., Alpcan, T., et al.: Sample complexity of solving non-cooperative games. *IEEE T. Inform. Theory.* **66**(2), 1261–1280 (2020)
10. Fudenberg, D., Tirole, J.: *Game Theory*. MIT Press, Cambridge (1991)
11. Kearns, M., Littman, M., Singh, S.: Graphical models for game theory. *Proc. UA I*, 253–260 (2001)
12. Qi, N., Huang, Z., Zhou, F., et al.: A task-driven sequential overlapping coalition formation game for resource allocation in heterogeneous UAV networks. *IEEE Trans. Mob. Comput*
13. Xiao, W., Zhou, Q., Liu, Y., et al.: Distributed reinforcement learning containment control for multiple nonholonomic mobile robots. *IEEE Trans. Circuits Syst. I Regul. Pap.* **69**(2), 896–907 (2022)
14. Li, Y., Tee, K., Yan, R., et al.: A framework of human–robot coordination based on game theory and policy iteration. *IEEE Trans. Robotics* **32**(6), 1408–1418 (2016)
15. Stalford, H.: Criteria for Pareto-optimality in cooperative differential games. *J. Optim. Theory Appl.* **9**(6), 391–398 (1972)
16. Zhang, H., Jiang, H., Luo, Y., et al.: Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method. *IEEE Trans. Ind. Electron.* **64**(5), 4091–4100 (2017)
17. Song, R., Li, J., Lewis, F.: Robust optimal control for disturbed nonlinear zero-sum differential games based on single NN and least squares. *IEEE Trans. Syst. Man. Cybern.* **50**(11), 4009–4019 (2020)
18. Mu, C., Wang, K., Ni, Z.: Adaptive learning and sampled-control for nonlinear game systems using dynamic event-triggering strategy. *IEEE Trans. Neural Net. Learn. Syst*
19. Wang, D., Ha, M., Zhao, M.: The intelligent critic framework for advanced optimal control. *Artif. Intell. Rev.* **55**, 1–22 (2022)
20. Ha, M., Wang, D., Liu, D.: A novel value iteration scheme with adjustable convergence rate. *IEEE Trans. Neural Netw. Learn*
21. Wei, Q., Lu, J., Zhou, T., et al.: Event-triggered near-optimal control of discrete-time constrained nonlinear systems with application to a boiler-turbine system. *IEEE Trans. Ind. Inform.* **18**(6), 3926–3935 (2022)
22. Gao, X., Bai, W., Li, T., et al.: Broad learning system-based adaptive optimal control design for dynamic positioning of marine vessels. *Nonlinear Dyn.* **105**, 1593–1609 (2021)
23. Beuchat, P., Warrington, J., Lygeros, J.: Accelerated pointwise maximum approach to approximate dynamic programming. *IEEE Trans. Autom. Control* **67**(1), 251–266 (2022)
24. Adams, S., Cody, T., Beling, P.: A survey of inverse reinforcement learning. *Artif. Intell. Rev.* **55**, 4307–4346 (2022)
25. Li, Z., Wu, L., Xu, Y., et al.: Multi-stage real-time operation of a multi-energy microgrid with electrical and thermal energy storage assets: a data-driven MPC-ADP approach. *IEEE Trans. Smart Grid* **13**(1), 213–226 (2022)
26. Yang, R., Wang, D., Qiao, J.: Policy gradient adaptive critic design with dynamic prioritized experience replay for wastewater treatment process control. *IEEE T. Ind. Inform.* **18**(5), 3150–3158 (2022)
27. Liu, Y., Zhang, H., Yu, R., et al.: Data-driven optimal tracking control for discrete-time systems with delays using adaptive dynamic programming. *J. Frankl. Inst.* **355**(13), 5649–5666 (2018)
28. Li, Y., Wei, C., An, T., et al.: Event-triggered-based cooperative game optimal tracking control for modular robot manipulator with constrained input. *Nonlinear Dyn.* **109**, 2759–2779 (2022)
29. Yang, H., Hu, Q., Dong, H., et al.: ADP-based spacecraft attitude control under actuator misalignment and pointing constraints. *IEEE Trans. Ind. Electron.* **69**(9), 9342–9352 (2022)
30. Huang, J., Zhang, Z., Cai, F., et al.: Optimized formation control for multi-agent systems based on adaptive dynamic programming without persistence of excitation. *IEEE Control Syst. Lett.* **6**, 1412–1417 (2022)
31. Dong, B., An, T., Zhou, F., et al.: Decentralized robust zero-sum neuro-optimal control for modular robot manipulators in contact with uncertain environments: theory and experimental verification. *Nonlinear Dyn.* **97**, 503–524 (2019)
32. Tazi, K., Abbou, F., Abdi, F.: Multi-agent system for microgrids: design, optimization and performance. *Artif. Intell. Rev.* **53**, 1233–1292 (2020)

33. Li, K., Li, Y.: Adaptive NN optimal consensus fault-tolerant control for stochastic nonlinear multiagent systems. *IEEE Trans. Neural Netw. Learn.*
34. Ma, B., Dong, B., Zhou, F., et al.: Adaptive dynamic programming-based fault-tolerant position-force control of constrained reconfigurable manipulators. *IEEE Access* **8**, 183286–183299 (2020)
35. Han, K., Feng, J., Yao, Y.: An integrated data-driven Markov parameters sequence identification and adaptive dynamic programming method to design fault-tolerant optimal tracking control for completely unknown model systems. *J. Frankl. Inst.* **354**(13), 5280–5301 (2017)
36. Xue, S., Luo, B., Liu, D., et al.: Constrained event-triggered H_∞ control based on adaptive dynamic programming with concurrent learning. *IEEE Trans. Syst. Man. Cybern.* **52**(1), 357–369 (2022)
37. Liu, Y., Li, X.: Decentralized robust adaptive control of nonlinear systems with unmodeled dynamics. *IEEE Trans. Autom. Control* **47**(5), 848–856 (2002)
38. Yang, X., He, H.: Adaptive dynamic programming for decentralized stabilization of uncertain nonlinear large-scale systems with mismatched interconnections. *IEEE Trans. Syst. Man. Cybern.* **50**(8), 2870–2882 (2020)
39. Zhou, Z., Xu, H.: Decentralized adaptive optimal tracking control for massive autonomous vehicle systems with heterogeneous dynamics: a stackelberg game. *IEEE Trans. Neural Netw. Learn.* **32**(12), 5654–5663 (2021)
40. Dong, B., Zhou, F., Liu, K., et al.: Decentralized robust optimal control for modular robot manipulators via critic-identifier structure-based adaptive dynamic programming. *Neural Comput. Appl.* **32**, 3441–3458 (2020)
41. An, T., Wang, Y., Liu, G., et al.: Cooperative game-based approximate optimal control of modular robot manipulators for human–robot collaboration. *IEEE Trans. Cybern.* **53**(7), 4691–4703 (2023)
42. Liu, G., Abdul, S., Goldenberg, A.A.: Distributed control of modular and reconfigurable robot with torque sensing. *Robotica* **26**(1), 75–84 (2008)
43. Rahman, M., Ikeura, R., Mizutani, K.: Investigation of the impedance characteristic of human arm for development of robots to cooperate with humans. *JSME Int. J. Ser. C* **45**(2), 510–518 (2002)
44. Yu, X., Li, Y., Zhang, S., et al.: Estimation of human impedance and motion intention for constrained human-robot interaction. *Neurocomputing* **390**, 268–279 (2020)
45. Li, Y., Ge, S.: Human-robot collaboration based on motion intention estimation. *IEEE-ASME Trans. Mech.* **19**(3), 1007–1014 (2014)
46. Mu, C., Wang, K., Ni, Z., et al.: Cooperative differential game-based optimal control and its application to power systems. *IEEE Trans. Ind. Inform.* **16**(8), 5169–5179 (2020)
47. Zhao, B., Wang, D., Shi, G., Liu, D., Li, Y.: Decentralized control for large-scale nonlinear systems with unknown mismatched interconnections via policy iteration. *IEEE Trans. Syst Man Cybern.* **48**(10), 1725–1735 (2018)
48. Li, Y., Tee, K., Chan, W., et al.: Continuous role adaptation for human–robot shared control. *IEEE Trans. Robotics* **31**(3), 672–681 (2017)
49. Vamvoudakis, K., Lewis, F.: Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica* **46**, 878–888 (2010)
50. Dong, B., An, T., Zhu, X., et al.: Zero-sum game-based neuro-optimal control of modular robot manipulators with uncertain disturbance using critic only policy iteration. *Neurocomputing* **450**(2), 183–196 (2021)
51. Ma, B., Li, Y., An, T., et al.: Compensator-critic structure-based neuro-optimal control of modular robot manipulators with uncertain environmental contacts using non-zero-sum games. *Knowl. Based Syst.* **224**(13), 107100 (2021)
52. Wang, D., Qiao, J., Cheng, L.: An approximate neuro-optimal solution of discounted guaranteed cost control design. *IEEE Trans. Cybern.* **52**(1), 77–86 (2022)
53. Li, Q., Wang, Z., Wang, W., et al.: A human-centered comprehensive measure of take-over performance based on multiple objective metrics. *IEEE Trans. Intell. Transp. Syst.* **24**(4), 4235–4250 (2023)
54. Li, Q., Su, Y., Wang, W., et al.: Latent hazard notification for highly automated driving: Expected safety benefits and driver behavioral adaptation. *IEEE Trans. Intell. Transp. Syst.* <https://doi.org/10.1109/TITS.2023.3280955>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.