Robotic Knee Tracking Control to Mimic the Intact Human Knee Profile Based on Actor-Critic Reinforcement Learning

Ruofan Wu, Zhikai Yao, Jennie Si, *Fellow, IEEE*, and He (Helen) Huang, *Senior Member, IEEE*

***Abstract*—We address a state-of-the-art reinforcement learning (RL) control approach to automatically configure robotic prosthesis impedance parameters to enable end-to-end, continuous locomotion intended for transfemoral amputee subjects. Specifically, our actor-critic based RL provides tracking control of a robotic knee prosthesis to mimic the intact knee profile. This is a significant advance from our previous RL based automatic tuning of prosthesis control parameters which have centered on regulation control with a designer prescribed robotic knee profile as the target. In addition to presenting the tracking control algorithm based on direct heuristic dynamic programming (dHDP), we provide a control performance guarantee including the case of constrained inputs. We show that our proposed tracking control possesses several important properties, such as weight convergence of the learning networks, Bellman (sub) optimality of the cost-to-go value function and control input, and practical stability of the human-robot system. We further provide a systematic simulation of the proposed tracking control using a realistic human-robot system simulator, the OpenSim, to emulate how the dHDP enables level ground walking, walking on different terrains and at different paces. These results show that our proposed dHDP based tracking control is not only theoretically suitable, but also practically useful.**

***Index Terms*—Automatic tracking of intact knee, configuration of robotic knee prosthesis, direct heuristic dynamic programming (dHDP), reinforcement learning control.**

I. INTRODUCTION

POWERED lower limb prosthesis provides great promise for amputees to regain mobility in daily life. Its potential has been demonstrated for transfemoral amputees' walking ability [1], [2]. Such robotic devices rely on an impedance control framework which is designed based on human biomechanics to mimic the central nervous system controlled human joint movements to provide a natural substitute to the lost limb functions. These devices require customization of the impedance parameters for each individual user. Currently, configuration of the powered devices is performed in clinics by technicians who manually tune a subset of impedance parameters over a number of visits of the patient. This procedure is time and labor intensive for both amputees and clinicians. Therefore, an automatic approach to tuning the powered prosthesis parameters is needed.

Automatically configuring the impedance parameter settings has been attempted over the past several years. An untested idea aims at estimating the joint impedance based on biomechanical measurements and a model of the unimpaired leg [3], [4]. This idea may not be practically useful as the biomechanics and the joint activities of amputees are fundamentally different from those of the able-bodied population. Another approach is to constrain the knee kinematics via the relationship of the joint control and intrinsic measurements, which in turn requires careful modeling and thus may not be feasible [5], [6]. Such an approach relies on significant domain knowledge and is tuning time. A cyber expert system was proposed [7] to emulate the prosthetists' tuning decisions of human experts into configuring the control parameters. This approach heavily relies on the expert's experience and is not expected to scale well to more joints and different users and tasks.

As those methods all have their fundamental limitations in principled ways, new approaches to configuring the prosthesis control parameters are needed. RL based adaptive optimal control approach is a promising alternative as they have demonstrated their capability of learning from data measurements in an online or offline manner in several realistic application problems including large-scale control problems [8]–[13]. The core of the RL methods is the idea of providing approximate solutions to the Bellman equation of optimal control problems. We have successfully developed several RL algorithms to configure impedance parameter settings, including actor-critic RL [14]–[17] and policy iteration based RL approaches [18]–[21], and systematically tested them in both extensive simulations and in experiments using able-bodied and transfemoral amputee subjects.

All of our RL control approaches to date require a target knee motion profile which can only be subjectively determined. Nonetheless, those results are important as they provided the necessary understanding of the impedance

Manuscript received June 11, 2021; accepted July 14, 2021. This work was partly supported by the National Science Foundation (1563921, 1808752, 1563454, 1808898). Recommended by Associate Editor Weinan Gao. *(Ruofan Wu and Zhikai Yao contributed equally to this work. Corresponding author: Jennie Si.)*

Citation: R. F. Wu, Z. K. Yao, J. Si, and H. Huang, "Robotic knee tracking control to mimic the intact human knee profile based on actor-critic reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 1, pp. 19–30, Jan. 2022.

R. F. Wu, Z. K. Yao, and J. Si are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: ruofanwu@asu.edu; zacyao.cn@gmail.com; si@asu.edu).

H. Huang is with the Department of Biomedical Engineering, North Carolina State University, Raleigh, NC 27695 USA, and also with the University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (e-mail: hhuang11@ncsu.edu).

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/JAS.2021.1004272

control parameter configuration problem and if RL control is capable of solving this problem. Even though detailed understanding of human locomotion at neurological and biomechanical levels has long been established, the individual human subject's locomotion dynamics are still not feasible to model accurately by mathematical descriptions as individuals differ physically, biologically and neurologically. Additionally, different locomotion tasks, such as changing pace [22], sloped walking [23] and walking on uneven terrain [24] all have significant influence on human gait behavior. As such, accurately prescribing each and every locomotion behavior for control purposes is not feasible.

Tracking the intact knee joint motion by a prosthetic knee is an intuitive idea as the intact knee kinematics is the most natural and realistic target: it contains actual biological joints' inter-relational information, which makes it a good candidate to replace a subjectively defined knee profile. Studies have shown that bilateral coordination between two legs are needed in the regulation of bipedal walking to maintain stability, and that such interlimb cooperation can be accomplished at a spinal level. Since the spinal level locomotor networks are symmetrically organized [25], sensory and muscle activities of both sides are involved in rhythmic walking. Amputees usually display asymmetrical walking by relying heavily on their intact limbs because of the loss of sensory feedback.

Tracking the intact knee actually has been explored years ago. Grimes *et al.* developed a mirror control scheme for the stance phase (not a complete gait cycle). It tracked the sound limb's knee trajectory in the stance phase by multiplying a gain factor to avoid over flexion while a fixed trajectory was applied in swing phase. Bernal-Torres *et al.* copied the full gait trajectory by the Kalman filter with a biomimetic designed prosthesis but no human experiment or systematic simulations were reported [26]. Joshi *et al.* developed a control strategy by controlling the swing time to mirror the stride duration of the intact knee while the prosthesis was locked during stance phase [27]. Sahoo *et al.* aimed at mirroring the step length by controlling the push-off force [28]. The above approaches focused on either part of a gait cycle or the outcome measurement such as step length and stance time. None of them has shown feasibility of tracking a completed gait cycle.

Virtual constraints were proposed to generate coordinated joint motions as target joint motion profiles for the robotic knee to track [29]. Biomimetic virtual constraints described the joints' geometric relationships and were encoded by hybrid zero dynamics [30]. However, there are a few limitations on this approach. Virtual constraints require a simplified human model to establish the geometric relationship among joints. Such a model is difficult to establish for a human-prosthesis system. In a recent work of prosthesis control design based on the virtual constraints [29], only a proportional gain was derived and applied. The overall human-prosthesis performance during locomotion is yet to be demonstrated.

Tracking an intact knee motion poses additional challenges beyond regulation control. Individual prosthesis users demonstrate very different gait patterns, which are also very different from one another, and also from healthy subjects. This may be caused by individual's physical conditions and/or biological factors such as reduced or lost proprioception and condition of socket fitting. It has long been believed that the intact knee motion pattern changes as the user learns to walk in a prosthesis (our data below also provide corroborating evidence about this), tracking a moving target knee motion has never been demonstrated by any controller. Most existing tracking control designs based on mathematical models of the human-prosthesis dynamics and the reference trajectory, both are difficult or actually impossible to obtain. For data-driven adaptive optimal control, reports on data-driven tracking control are much less than those of regulation control. Most approaches are based on reinforcement learning to establish approximate solution to the HJB equation, yet, few of them have demonstrated feasibility in engineering problems. And even fewer results can be found to provide systematic studies on the control design to demonstrate applicability.

In this paper, we propose an RL tracking control scheme for the robotic knee to mimic the intact knee in different locomotion tasks. In a previous experiment [31], we successfully tested this pilot idea of RL tracking control to automatically configure impedance parameter settings. In this study, we formally formulate the tracking control problem, develop a complete tracking control algorithm based on dHDP, and provide an analytical framework to validate the real time control performance guarantee by using this proposed scheme. The contributions of this work include the following.

1) We provide the first systematic demonstration of an end-to-end, continuous walking enabled by automatic tracking control of a wearable lower limb robotic device beyond our previous regulation control results [14]–[21].

2) We show that our proposed tracking control possesses several important properties, such as weight convergence of the learning networks, Bellman (sub) optimality of the cost-to-go value function and control input, and practical stability of the human-robot systems.

3) We provide a systematic evaluation of the proposed tracking control using a realistic human-robot system simulator to demonstrate level ground walking, slope walking under various slope angles and walking under different paces. These results show that our proposed dHDP based tracking control is not only theoretically suitable, but also practically useful.

The remaining of this paper is organized as follows. Section II describes the human-prosthesis system and develops dHDP to solve the tracking problem. Section III gives the Lyapunov stability analysis of system. Section IV presents the implementation using Opensim simulations. Section V presents extensive simulations. Discussions and conclusion are presented in Section VI.

II. METHOD

Our proposed RL tracking control is built upon the finite state machine (FSM) impedance controller (IC) framework. It is to mimic the torque-generating capability of biological joints to enable natural movement. Humans reportedly control muscle activity to adjust joint impedance in walking [32]–[34], and the compliant behaviors of legs are fundamen-

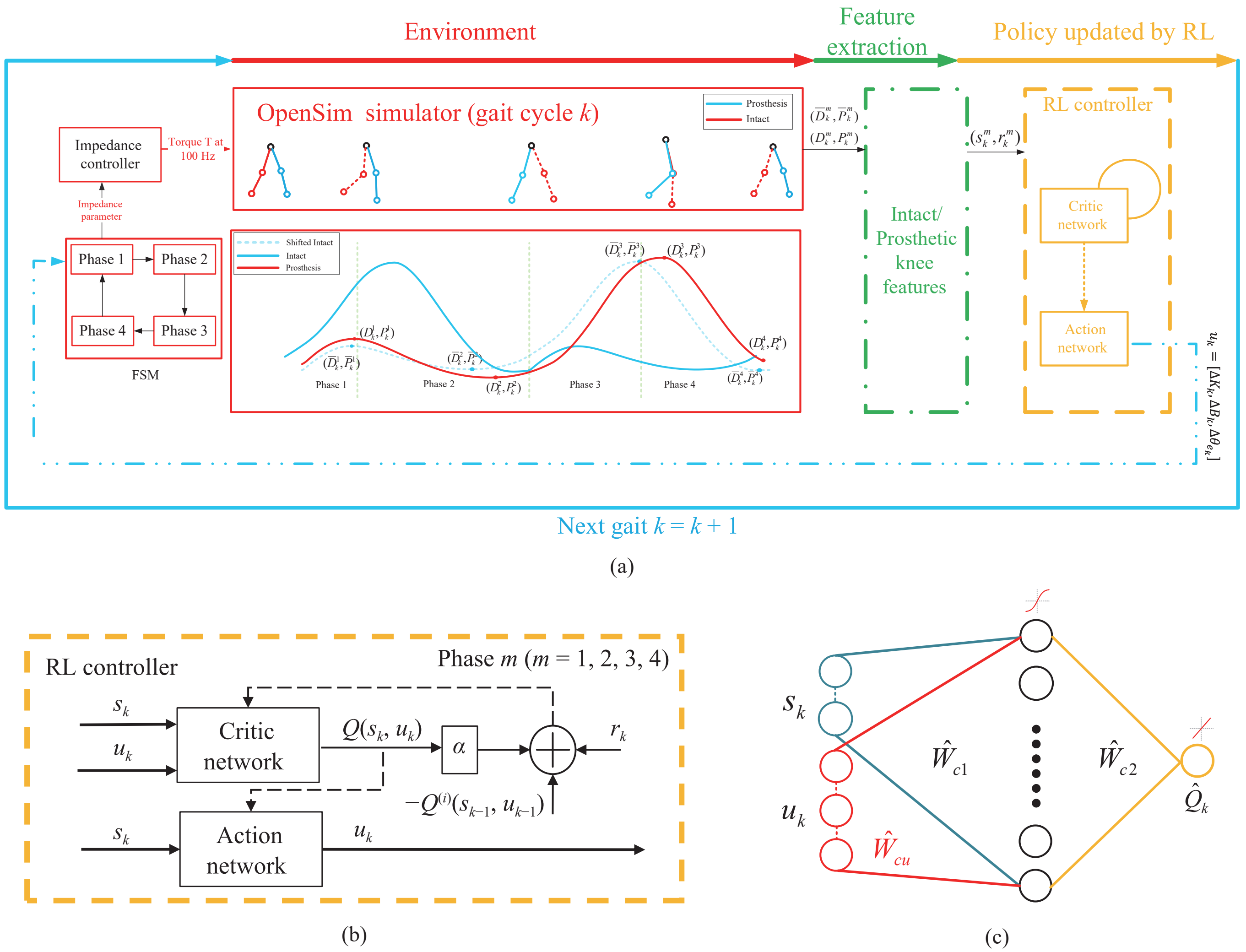

Environment
Feature extraction
Policy updated by RL
OpenSim simulator (gait cycle k)
Prosthesis
Intact
Impedance controller
Torque T at 100 Hz
Impedance parameter
Phase 1
Phase 2
Phase 3
Phase 4
FSM
Shifted Intact
Intact
Prosthesis
$(\bar{D}_k^m, \bar{P}_k^m)$
(D_k^m, P_k^m)
(s_k^m, r_k^m)
Intact/ Prosthetic knee features
RL controller
Critic network
Action network
$u_k = [\Delta K_k, \Delta B_k, \Delta\theta_{e_k}]$
Next gait $k = k + 1$
(a)
RL controller
Phase m (m = 1, 2, 3, 4)
s_k
u_k
Critic network
$Q(s_k, u_k)$
α
r_k
$-Q^{(i)}(s_{k-1}, u_{k-1})$
Action network
(b)
s_k
u_k
$\hat{W}_{c1}$
$\hat{W}_{c2}$
$\hat{W}_{cu}$
$\hat{Q}_k$
(c)

Fig. 1. Two control loops are coordinated to realize automatic control of the prosthetic knee based on an impedance control framework: i) an intrinsic impedance control (left most panel of (a)) that operates at 100 Hz that creates a control torque according to (2) based on the knee kinematics and also the impedance settings, and ii) an RL controller that is updated at each step k. The output is the adjustment of the impedance setting that is to be used in the intrinsic controller. (a) Flow chart of the automatic robotic knee control parameter tuning scheme by dHDP. (b) The dHDP controller to adjust the impedance setting. (c) The structure of the critic network in the RL controller.

tal to human locomotion [35], [36]. Within this context, impedance represents the inherent property of a mechanical joint. It describes the relationship between external force and motion produced or in other words, it is a dynamic property that governs human joint-torque relationship. In our study, the prosthesis joint is therefore characterized by the stiffness (K), damping ratio (B), and equilibrium position (θ_e). The FSM-IC provides intrinsic control in the form of adjustable control torque influenced by impedance parameters. The settings of the impedance parameters as control inputs have to be adjusted or adapted to meet individuals' needs including their different physical conditions. Our proposed RL tracking control is to automatically provide such needed impedance parameter settings. In turn, the joint impedance affects the knee kinematics such as peak value of knee angle collaboratively. External force such as ground reaction force and human reactions also affect knee angle or peak value. However, many of these factors are difficult, or nearly impossible, to be modeled. Gait duration is influenced by impedance control parameters [17], as well as human's movement control of residual hip and intact limb in gait. Meanwhile, the transition between double-stance and single-stance will influence the duration as well since they have different dynamics. Therefore, the relationship between impedance and peak knee angle/duration is not deterministic and difficult to model. These challenges motivate us to consider data-driven RL control.

A. Finite State Machine (FSM) Impedance Control (IC)

The FSM-IC is common for prosthesis intrinsic control as studies have shown that humans control the stiffness of leg muscles and therefore joint impedance while walking, and compliant behaviors of legs are instrumental for human walking. The impedance controller generates a torque input to the robotic knee based on current knee kinematics and knee joint impedance settings.

Refer to Fig. 1(b), a gait cycle is divided into four phases in the FSM-IC: stance flexion (STF, $m = 1$), stance extension (STE, $m = 2$), swing flexion (SWF, $m = 3$) and swing extension (SWE, $m = 4$). The phase transitions are determined by knee motion and gait events (heel strike and toe-off) that are obtained from vertical ground reaction forces of both legs. In each phase of the FSM, three impedance parameters (stiffness K, damping B, and equilibrium position θ_e) are provided as inputs to the FSM-IC for gait cycle k

$$I_k^m = [K_k^m, B_k^m, (\theta_e)_k^m]. \tag{1}$$

The knee joint torque T_k^m is consequently generated by the following first principle equation:

$$T_k^m = K_k^m(\theta - (\theta_e)_k^m) + B_k^m \omega. \tag{2}$$

The RL controller will adjust these impedance parameters, i.e.,

$$u_k^m = [\Delta K_k^m, \Delta B_k^m, \Delta(\theta_e)_k^m] \tag{3}$$

so that the updated impedance parameters are applied to the FSM-IC to generate knee torque according to (2)

$$I_{k+1}^m = I_k^m + u_k^m. \tag{4}$$

B. Tracking Problem Formulation

Biomechanical studies have shown that the intact knee joint movements or profiles change as amputees adapt to a prosthetic device [37]. We have observed the same in our pilot study using two human subjects [31]. Fig. 2 is an illustration of the same phenomenon using simulations where the intact knee kinematic trajectories were recorded, and changes in profile features are clearly observed. The goal of the RL controlled robotic knee is therefore to track those time varying intact knee profile features for each and every phase during each and every gait cycle. For a gait cycle k, the robotic knee motion (Fig. 1(b)) featured by the peak knee angle P_k^m (degrees) and duration D_k^m (seconds) is measured. Let

$$Z_k^m = (D_k^m, P_k^m). \tag{5}$$

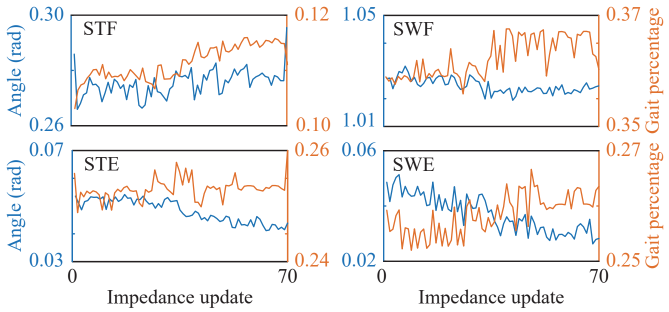


Fig. 2. An illustration of how the intact knee profile changes as the robotic knee control parameters adapt during a level ground walking simulation session. The peak angles (blue) and the phase durations (orange) are shown in all phases.

Similarly, we measure the peak knee angle and duration of the intact knee, and let

$$Y_k^m = (\bar{D}_k^m, \bar{P}_k^m). \tag{6}$$

We consider a human-robot, i.e., an amputee-prosthesis system, a discrete-time nonlinear dynamic system with unknown nonlinearities. For the ease of discussion, we let the dynamics be represented by

$$Z_{k+1}^m = F(Z_k^m, u_k^m), \ m = 1, \ldots, 4, \quad k = 0, 1, \ldots \tag{7}$$

where the nonlinear mapping F is Lipschitz continuous on the domain of $\mathcal{D} \triangleq \{(Z,u) | Z \in \mathcal{Z}, u \in \mathcal{U}\}$, where $\mathcal{Z}$ and $\mathcal{U}$ are compact sets with dimensions of N_Z and N_u, respectively. In the human-robot system under consideration, F represents the kinematics of the robotic knee, which is affected by both the human wear and the RL controller. Because of a human in-the-loop, an explicit mathematical model as (7) is intractable or impossible to obtain.

Without causing any confusion and for the sake of convenience, we drop the superscript m ($m = 1,2,3,4$) in the rest of the paper because all four FSM-ICs and their respective RL controllers share the same structure, although the RL controllers for each phase have different parameterizations or in other words, the control policies are different for each phase even though they have the same structure. Then, the tracking error between the intact knee and the prosthetic knee is defined as

$$e_k = Y_k - Z_k = (\Delta D_k, \Delta P_k). \tag{8}$$

C. dHDP for Tracking Control

Fig. 1(a) depicts the RL based solution approach to automatically configure the impedance parameters of the robotic knee to track the intact knee joint motion within the FSM-IC framework. Each RL control block corresponds to one of the four FSM phases. As shown in Fig. 1(a), we develop a dHDP based RL tracking control with each of the four dHDP blocks providing impedance parameter settings for each of the four gait phases. Each dHDP block has an action network and a critic network, trained for the given FSM phase only.

In the RL tracking controller, let the state be denoted by s_k, and the control input/action network output as u_k for gait cycle k, i.e.,

$$s_k = (\Delta D_k, \Delta P_k), \quad u_k = [\Delta K_k, \Delta B_k, \Delta(\theta_e)_k]. \tag{9}$$

We consider the stage cost in a quadratic form

$$U(s_k, u_k) = {s_k}^T R_s s_k + {u_k}^T R_u u_k \tag{10}$$

where $R_s \in \mathbb{R}^{2\times 2}$ and $R_u \in \mathbb{R}^{3\times 3}$ are positive definite matrices. Note that, our following results are also applicable to other stage costs such as a reinforcement signal of finite discrete levels or a general bounded sefmi-definite reinforcement signal.

We consider the tracking problem as one to devise an optimal control law via learning from observed data along the human-robot interacting system dynamics. We define the state-action Q-function or the total cost-to-go as

$$Q(s_k, u_k) = U(s_k, u_k) + \sum_{j=1}^{\infty} \gamma^j U\left(s_{k+j}, u_{k+j}\right). \tag{11}$$

Note that the $Q(s_k, u_k)$ value is a performance measure when action u_k is applied at state s_k. Such $Q(s_k, u_k)$ formulation implies that we have considered the optimal adaptive tracking control of the robotic knee as a discrete-time, infinite horizon, discounted problem without knowing an explicit mathematical description of the human-robot interacting dynamics.

For the Q-function in (11), it satisfies the Bellman equation

$$Q(s_k, u_k) = U(s_k, u_k) + \gamma Q(s_{k+1}, u_{k+1}). \tag{12}$$

Assumption 1: The state trajectory Y_k of the intact knee is bounded, and the initial robotics knee state Z_0 is bounded.

1) Critic Network: Fig. 1(c) depicts the structure of the critic network which is realized by an universal approximator with one hidden layer. Therefore, the approximated value is

$$\hat{Q}(s_k, u_k) = \hat{W}_{c2,k}\phi(\hat{W}_{c1,k}z_k) \tag{13}$$

where $\hat{W}_{c1,k}$ is the estimated weight matrix between the input layer and the hidden layer, $\hat{W}_{c2,k}$ is the estimated weight matrix between the hidden layer and the output layer, ϕ is the activation function (hyperbolic tangent) in the hidden layer, and the input $z_k = [s_k, u_k]^T$.

The approximation error of the critic network is

$$e_{c,k} = \gamma\hat{Q}(s_k, u_k) - \left[\hat{Q}(s_{k-1}, u_{k-1}) - U(s_{k-1}, u_{k-1})\right]. \tag{14}$$

In the following, we use the short-hand notation U_{k-1} for $U(s_{k-1}, u_{k-1})$ and similarly for others. The weights $\hat{W}_{c1}$ and $\hat{W}_{c2}$ are updated as

$$\hat{W}_{c,k+1} = \hat{W}_{c,k} + \Delta W_{c,k}. \tag{15}$$

According to the gradient descend rule as in [38], $\Delta W_{c1,k}$ and $\Delta W_{c2,k}$ can be written as

$$\begin{cases} \Delta\hat{W}_{c1,k} = -l_{c,k}\gamma e_{c,k}\hat{W}_{c2,k}\left[\frac{1}{2}\left(1-\phi^2_{c,k}\right)\right]z_k \\ \Delta\hat{W}_{c2,k} = -l_{c,k}\gamma e_{ck}\phi_{c,k} \end{cases} \tag{16}$$

where $\phi_{c,k}$ is the output of hidden layers in critic network, and $l_{c,k}$ is the learning rate.

2) Action Network: The output of the action network is the control input

$$u_k = \phi(\hat{W}_{a2,k}\phi(\hat{W}_{a1,k}s_k)) \tag{17}$$

where $\hat{W}_{a1,k}$ is the estimated weight matrix between the input layer and the hidden layer, $\hat{W}_{a2,k}$ is the estimated weight matrix between the hidden layer and the output layer. Note, however, that our development applies to unconstrained inputs as well.

Based on the design principle of dHDP [38], the action network is to minimize the total cost-to-go $Q(s_k, u_k)$. We defined the prediction error of the action network as

$$e_{a,k} = \hat{Q}(s_k, u_k). \tag{18}$$

Similarly to (16), the weights $\hat{W}_{a1}$ and $\hat{W}_{a2}$ are updated as

$$\hat{W}_{a,k+1} = \hat{W}_{a,k} + \Delta W_{a,k} \tag{19}$$

and $\Delta W_{a1,k}$, $\Delta W_{a2,k}$ can be written as

$$\begin{cases} \Delta\hat{W}_{a1,k} = -l_{a,k}e_{a,k}\left[\hat{W}_{c2,k}\frac{1}{2}\left(1-\phi^2_{c,k}\right)\hat{W}_{cu,k}\right]\frac{1}{2}\left(1-u^2_k\right) \\ \qquad\times\hat{W}_{a2,k}\frac{1}{2}\left(1-\phi^2_{a,k}\right)s_k \\ \Delta\hat{W}_{a2,k} = -l_{a,k}e_{a,k}\left[\hat{W}_{c2,k}\frac{1}{2}\left(1-\phi^2_{c,k}\right)\hat{W}_{cu,k}\right]\frac{1}{2}\left(1-u^2_k\right)\phi_{a,k} \end{cases} \tag{20}$$

where $\hat{W}_{cu,k}$ is the weight vector associated with the input u_k (Fig. 1(c)) from action network, i.e., the part of $\hat{W}_{c1}$ which connects with u_k, $\phi_{a,k}$ is the output of hidden layers in actor network, and $l_{a,k}$ is the learning rate.

III. LYAPUNOV STABILITY ANALYSIS

In this section, we provide a qualitative analysis for the weight convergence of the actor-critic networks, the Bellman (sub) optimality of the control policy, and practical stability of the human-prosthesis system.

A. Preliminaries

Let W^*_a, W^*_c denote the optimal weights, that is,

$$\begin{cases} W^*_a = \underset{\hat{W}_a}{\arg\min}\left\|\hat{Q}(s_k, u_k)\right\| \\ W^*_c = \underset{\hat{W}_c}{\arg\min}\left\|\gamma\hat{Q}(s_k, u_k) + U_{k-1} - \hat{Q}(s_{k-1}, u_{k-1})\right\| \end{cases} \tag{21}$$

and the optimal Q-value and optimal control policy are defined as

$$Q^* = W^*_{c2}\phi_{c,k} + \varepsilon_{c,k}, \quad u^* = \phi(W^*_{a2}\phi_{a,k}) + \varepsilon_{a,k} \tag{22}$$

where $\varepsilon_{c,k}$ and $\varepsilon_{a,k}$ are the reconstruction errors of neural networks. From (21) to (22), we implied that the persistent excitation condition is satisfied.

Assumption 2: The optimal weights for the actor-critic networks exist and they are bounded by two positive constants W_{am} and W_{cm}, respectively.

$$\left\|W^*_a\right\| \le W_{am}, \quad \left\|W^*_c\right\| \le W_{cm}. \tag{23}$$

Accordingly, the weight estimation errors of the actor-critic networks are described respectively as

$$\tilde{W}_{a,k} := \hat{W}_{a,k} - W^*_a, \quad \tilde{W}_{c,k} := \hat{W}_{c,k} - W^*_c. \tag{24}$$

Lemma 1: Consider the weight vectors of the critic network. Let

$$L_{1,k} = \frac{1}{l_c, k}tr[\tilde{W}^T_{c2,k}\tilde{W}_{c2,k}], \quad L_{2,k} = \frac{1}{l_{c,k}\alpha_1}tr\left[\tilde{W}^T_{c1,k}\tilde{W}_{c1,k}\right] \tag{25}$$

and let Assumption 2 hold. Then the first difference of $L_{1,k}$ is given by

$$\begin{aligned} \Delta L_{1,k} = &-\gamma^2\left\|\zeta_{c,k}\right\|^2 - \left(1-\gamma^2 l_{c,k}\|\phi_{c,k}\|^2\right) \\ &\times\|\gamma\hat{W}_{c2,k}\phi_{c,k} + U_{k-1} - \hat{W}_{c2,k-1}\phi_{c,k-1}\|^2 \\ &+\|\gamma W^*_{c2}\phi_{c,k} + U_{k-1} - \hat{W}_{c2,k-1}\phi_{c,k-1}\|^2 \end{aligned} \tag{26}$$

where $\zeta_{c,k} = \tilde{W}_{c2,k}\phi_{c,k}$ is an approximation error of the critic output.

And the first difference of $L_{2,k}$ is given by

$$\begin{aligned} \Delta L_{2,k} \le \frac{1}{\alpha_1}\Big(&\gamma^2 l_{c,k}\|\gamma\hat{W}_{c2,k}\phi_{c,k} + U_{k-1} \\ &-\hat{W}_{c2,k-1}\phi_{c,k-1}\|^2\|A_k\|^2\|x_{c,k}\|^2 \\ &+\gamma\|\tilde{W}_{c1,k}x_{c,k}A^T_k\|^2 + \gamma\|\gamma\hat{W}_{c2,k}\phi_{c,k} \\ &+U_{k-1} - \hat{W}_{c2,k-1}\phi_{c,k-1}\|^2\Big) \end{aligned} \tag{27}$$

where $\alpha_1 > 0$ is a weighting factor and A_k is a vector, with

$A_k = \frac{1}{2}(1-\phi_{c,k}^2)\hat{W}_{c2,k}$.

The proof of Lemma 1 can be found in [39].

Lemma 2: Consider the weight vector of the action network, which leads to the control input u_k of the human-prosthesis systems (7). Let

$$L_{3,k} = \frac{1}{l_{a,k}\alpha_2} tr\left[\tilde{W}_{a2,k}^T \tilde{W}_{a2,k}\right], L_{4,k} = \frac{1}{l_{a,k}\alpha_3} tr\left[\tilde{W}_{a1,k}^T \tilde{W}_{a1,k}\right]. \quad (28)$$

Then the first difference of $L_{3,k}$ is bounded by

$$\Delta L_3(k) \le \frac{1}{\alpha_2}\Big(-(1-l_{a,k}\|\phi_{a2,k}\|^2\|\hat{W}_{c2,k}C_k\|^2)\|\hat{W}_{c2,k}\phi_{c,k}\|^2 + 4\|\zeta_{c,k}\|^2 + 4\|W_{c2}^*\phi_{c,k}\|^2 + \|\hat{W}_{c2,k}C_k\zeta_{a,k}\|^2\Big) \quad (29)$$

where $\zeta_{a,k} = \tilde{W}_{a2,k}\phi_{a,k}$ is an approximation error of the actor network output, $C_k = \frac{1}{2}(1-\phi_{c,k}^2)W_{cu,k} \times \frac{1}{2}(1-u_k^2)$; $\alpha_2 > 0$ is a weighting factor.

And the first difference of $L_{4,k}$ is bounded by

$$\Delta L_4(k) \le \frac{1}{\alpha_3}\Big(l_{a,k}\|\hat{W}_{c2,k}\phi_{c,k}\|^2\|\hat{W}_{c2,k}C_kD_k^T\|^2\|s_k\|^2 + \|\hat{W}_{c2,k}\phi_{c,k}\|^2 + \|\tilde{W}_{a1,k}s_k\|^2\|\hat{W}_{c2,k}C_kD_k^T\|^2\Big) \quad (30)$$

where $D_k = \frac{1}{2}(1-\phi_{a,k}^2)\hat{W}_{a2,k}$, and $\alpha_3 > 0$ is a weighting factor. The proof of Lemma 2 can be found in Appendix.

B. Weight Convergence, (Sub) Optimality and Practical Stability

Definition 1 (Uniformly Ultimately Boundedness of a Discrete Time Dynamical System [40]): A dynamical system with states x_k is said to be uniformly ultimately bounded with ultimate bound $b > 0$, if for any $a > 0$ and $t > 0$, there exists a positive number $N = N(a,b)$ independent of t, such that $\|\tilde{x}_k\| \le b$ for all $k \ge N+t$ whenever $\|\tilde{x}_t\| \le a$.

In the following, we consider the dynamics of $\hat{W}_{a,k}$ and $\hat{W}_{c,k}$.

Theorem 1 (Weight Convergence): Let Assumptions 1 and 2 hold, and let the weights of the actor and critic neural networks be updated according to (16) and (20), respectively. Then $\tilde{W}_{c,k}$ and $\tilde{W}_{a,k}$ are uniformly ultimately bounded provided that the following conditions are met:

$$\begin{cases} l_{c,k} < \dfrac{\alpha_1-\gamma}{\gamma^2\alpha_1(\left\|\phi_{c,k}\right\|^2 + \frac{1}{\alpha_1}\left\|A_k\right\|^2\left\|z_k\right\|^2)} \\ l_{a,k} < \dfrac{\alpha_3-\alpha_2}{\alpha_3\left\|\hat{W}_{c2,k}^TC_k\right\|^2\left\|\phi_{a,k}\right\|^2 + \alpha_2\left\|\hat{W}_{c2,k}C_kD_k^T\right\|^2\left\|s_k\right\|^2}. \end{cases} \quad (31)$$

Remark 1: As γ, α_1, α_2, α_3 can be found in (11), (25), and (28), for $l_{c,k}$, $l_{a,k}$ to be positive, it is necessary that $\alpha_1 > \gamma > 0$ and $\alpha_3 > \alpha_2 > 0$. Those conditions can be satisfied in the following induction.

Proof of Theorem 1: We introduce a candidate Lyapunov function

$$L_k = L_{1,k} + L_{2,k} + L_{3,k} + L_{4,k} \quad (32)$$

where $L_{1,k}$, $L_{2,k}$, $L_{3,k}$, and $L_{4,k}$ are defined in (25) and (28). The first difference of L_k can be rewritten as

$$\begin{aligned}\Delta L_k \le & -(\gamma^2 - \frac{4}{\alpha_2})\left\|\zeta_{c,k}\right\|^2 - (1-\gamma^2 l_{c,k}\left\|\phi_{c,k}\right\|^2 \\ & - \frac{\gamma^2 l_{c,k}}{\alpha_1}\left\|A_k\right\|^2\left\|z_k\right\|^2 - \frac{\gamma}{\alpha_1})\left\|\gamma\hat{W}_{c2,k}\phi_{c,k} + U_{k-1} \right. \\ & \left. - \hat{W}_{c2,k-1}\phi_{c,k-1}\right\|^2 - \left\|\hat{W}_{c2,k}\phi_{c,k}\right\|^2(\frac{1}{\alpha_2} - \frac{l_{a,k}}{\alpha_2}\left\|\hat{W}_{c2,k}C_k\right\|^2 \\ & \times\left\|\phi_{a,k}\right\|^2 - \frac{l_{a,k}}{\alpha_3}\left\|\hat{W}_{c2,k}C_kD_k^T\right\|^2\left\|s_k\right\|^2 - \frac{1}{\alpha_3}) \\ & + \frac{4}{\alpha_2}\left\|W_{c2}^*\phi_{c,k}\right\|^2 + \frac{1}{\alpha_2}\left\|\hat{W}_{c2,k}C_k\right\|^2\left\|\zeta_{a,k}\right\|^2 \\ & + \left\|\gamma W_{c2}^*\phi_{c,k} + U_{k-1} - \hat{W}_{c2,k-1}\phi_{c,k-1}\right\|^2 \\ & + \frac{\gamma}{\alpha_1}\left\|\tilde{W}_{c1,k}z_k\right\|^2\|A_k\|^2 + \frac{1}{\alpha_3}\left\|\hat{W}_{c2,k}C_kD_k^T\right\|^2\left\|\tilde{W}_{a1,k}s_k\right\|^2.\end{aligned} \quad (33)$$

To guarantee that the second and the third terms in the last expression are negative, we need to choose learning rates in the following manner:

$$1-\gamma^2 l_{c,k}\left\|\phi_{c,k}\right\|^2 - \frac{\gamma^2 l_{c,k}}{\alpha_1}\left\|A_k\right\|^2\left\|z_k\right\|^2 - \frac{\gamma}{\alpha_1} > 0. \quad (34)$$

Therefore,

$$l_{c,k} < \min_k \frac{\alpha_1-\gamma}{\gamma^2\alpha_1(\|\phi_{c,k}\|^2 + \frac{1}{\alpha_1}\|A_k\|^2\|z_k\|^2)}. \quad (35)$$

Similarly, for the action network we obtain

$$\frac{1}{\alpha_2} - \frac{1}{\alpha_2}l_{a,k}\|(\hat{W}_{c2,k})^TC_k\|^2\|\phi_{a,k}\|^2 - \frac{l_{a,k}}{\alpha_3}\|D_kC_k^T\hat{W}_{c2,k}\|^2\|s_k\|^2 - \frac{1}{\alpha_3} > 0 \quad (36)$$

and then

$$l_{a,k} < \frac{\alpha_3-\alpha_2}{\alpha_3\|\hat{W}_{c2,k}^TC_k\|^2\|\phi_{a,k}\|^2 + \alpha_2\|\hat{W}_{c2,k}C_kD_k^T\|^2\|s_k\|^2}. \quad (37)$$

As is known, the following inequality holds:

$$\begin{aligned}&\left\|\gamma W_{c2}^*\phi_{c,k} + U_{k-1} - \hat{W}_{c2,k-1}\phi_{c,k-1}\right\|^2 \\ &\quad \le 4\gamma^2\left\|W_{c2}^*\phi_{c,k}\right\|^2 + 4U_{k-1}^2 + 2\left\|\hat{W}_{c2,k-1}\phi_{c,k-1}\right\|^2.\end{aligned} \quad (38)$$

Applying the Cauchy–Schwarz inequality, we have

$$\begin{aligned}&\frac{4}{\alpha_2}\|W_{c2}^*\phi_{c,k}\|^2 + \frac{1}{\alpha_2}\|\hat{W}_{c2,k}C_k\|^2\|\zeta_{a,k}\|^2 \\ &\quad + \|\gamma W_{c2}^*\phi_{c,k} + U_{k-1} - \hat{W}_{c2,k-1}\phi_{c,k-1}\|^2 \\ &\quad + \frac{\gamma}{\alpha_1}\|\tilde{W}_{c1,k}x_{c,k}\|^2\|A_k\|^2 + \frac{1}{\alpha_3}\|\hat{W}_{c2,k}C_kD_k^T\|^2\|\tilde{W}_{a1,k}s_k\|^2 \\ &\quad \le (\frac{4}{\alpha_2} + 4\gamma^2 + 2)(W_{2cm}\phi_{cm})^2 + 4U_m^2 + \frac{\gamma}{\alpha_1}(W_{1cm}z_mA_m)^2 \\ &\quad + \frac{1}{\alpha_2}(W_{2cm}C_mW_{2am}\phi_{am})^2 + \frac{1}{\alpha_3}(W_{2cm}C_mD_mW_{1am}s_m)^2 = M\end{aligned} \quad (39)$$

where ϕ_{am}, ϕ_{cm}, U_m, W_{1cm}, W_{2cm}, A_m, C_m, D_m, s_m, and z_m are the upper bounds of $\phi_{a,k}$, $\phi_{c,k}$, U_{k-1}, $W_{c1,k}$, $W_{c2,k}$, A_k, C_k, D_k, s_k, and z_k, respectively.

Substituting (38) and (39) into (33), we obtain $\Delta L_k < 0$ under the condition that $\alpha_2 > \frac{4}{\gamma^2}$, constrained $l_{a,k}$, $l_{c,k}$ in (31), and

$$\|\zeta_{c,k}\| > \sqrt{\frac{M}{\gamma^2 - \frac{4}{\alpha_2}}}. \tag{40}$$

Assumption 1 leads to that the initial tracking error s_0 is bounded, then the initial stage cost U_0 in (10) is bounded. There exist $l_{c,0}$, $l_{a,0}$ meeting condition (31) with the bounded initial weights $\hat{W}_{a,0}$, $\hat{W}_{c,0}$ and tracking error s_0. From Definition 1, $\Delta L_k < 0$ means that the estimation errors $\tilde{W}_{a,k}$ and $\tilde{W}_{c,k}$ are bounded from step k to step $k+1$, respectively. Based on Assumption 2, $\hat{W}_{a,k}$ and $\hat{W}_{c,k}$ are bounded from step k to step $k+1$. As F in (7) is Lipschitz continuous, the bounded control law u_k results in bounded Z_{k+1}, and then we obtain the bounded tracking error s_{k+1} under Assumption 1. The bounded $\hat{W}_{a,k+1}$ and $\hat{W}_{c,k+1}$ result in a bounded control law u_{k+1}, then the resulted stage cost U_{k+1} is bounded. As there exist appropriate $l_{c,k+1}$, $l_{a,k+1}$ meeting condition (31) with bounded $\hat{W}_{a,k+1}$, $\hat{W}_{c,k+1}$ and s_{k+1}, by mathematical induction, we have the estimation errors $\tilde{W}_{a,k}$ and $\tilde{W}_{c,k}$ uniformly ultimately bounded. ■

Theorem 2 ((Sub) Optimality Result): Under the conditions of Theorem 1, the Bellman optimality is achieved within finite approximation error. Meanwhile, the error between the obtained control u_k and optimal control u^* is uniformly ultimately bounded.

Proof of Theorem 2: From Assumption 2 and Theorem 1, we obtain that $\hat{Q}(s_k, u_k)$ and Q^* are bounded. Furthermore, u_k and u^* are bounded from Remark 2 and Assumption 2.

From the approximate Q-value in (13) and the optimal Q-value in (22), we have

$$\begin{aligned}\|\hat{Q}(s_k,u_k) - Q^*\| &= \|\hat{W}_{c2,k}\phi_{c,k} - W^*_{c2}\phi_{c,k} - \varepsilon_{c,k}\| \\ &\le \|\tilde{W}_{c2,k}\|\phi_{cm} + \varepsilon_{cm}.\end{aligned} \tag{41}$$

where ε_{cm} is the upper bound of $\varepsilon_{c,k}$.

For u_k and u^* from (17) and (22), we have

$$\|u_k - u^*\| = \|\phi(\hat{W}_{a2,k}\phi_{a,k}) - \phi(W^*_{a2}\phi_{a,k}) - \varepsilon_{a,k}\|. \tag{42}$$

Similarly to (50)

$$\|\hat{W}_{a1,k}\phi_{a2,k} - W^*_{a2}\phi_{a2,k}\| \le \|\tilde{W}_{a2,k}\|\phi_{a2m}. \tag{43}$$

Then, by using the Lagrange Mean Value Theorem, (51) can be rewritten as

$$\|u_k - u^*\| \le \frac{1}{2}\|\tilde{W}_{a2,k}\|\phi_{am} + \varepsilon_{am} \tag{44}$$

where ε_{am} is the upper bound of $\varepsilon_{a,k}$. This comes directly as $\|\tilde{W}_{c2}\|$ and $\|\tilde{W}_{a2}\|$ are both uniformly ultimately bounded as the time step k increases as shown in Theorem 1. It demonstrates that the Bellman optimality is achieved within finite approximation errors. ■

Remark 2 (Practical Stability): From Assumption 1, Remarks 1–3, and Theorem 1, we obtain that all the signals (such as u_k, Z_k, Y_k, and s_k) are bounded in the human-robot system (7). As shown in Theorem 2, the approximated Q-value in (13) and the resulted policy (17) achieve (sub) optimality as time step k increases. This demonstrates that the stage cost in (10) approaches zero in the sense of uniformly ultimately boundedness. As such, the tracking error (8) approaches zero in the sense of uniformly ultimately boundedness also as time step k increases. Therefore, the considered human-robot system is practically stable.

IV. OpenSim Simulation Studies

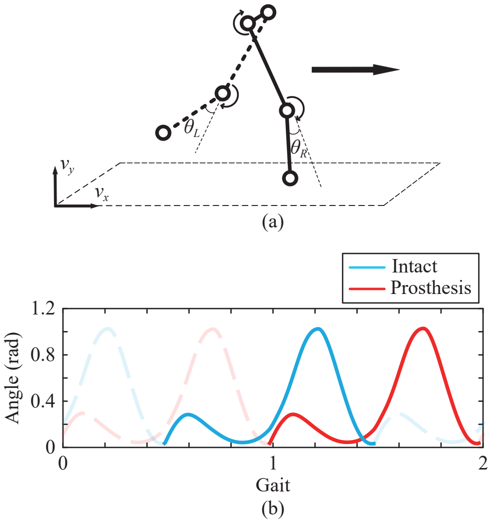
Table I is a summary of how our proposed dHDP tracking control is implemented and applied to configure the impedance parameters of the robotic knee. Pertinent information for the implementation steps is provided below.

TABLE I
IMPLEMENTATION OF SIMULATION STUDIES

```
Initialization (gait cycle $k = 0$);
Set simulation conditions such as walking speed, subject
physical conditions (body mass, height, etc.), and small
knee angles for both knees;
Random, initially feasible impedance parameters;
Random initial weights in actor-critic network;
Repeat (for a complete OpenSim simulated gait cycle)
  Obtain target feature $Y_k$ (5) and robotic feature $Z_k$ (6);
  Obtain state dynamics $s_k$ using (9);
  if $s_k >$ safety bound then
    Reset to initial impedance;
    Repeat the implementation procedure from the top
    but keep the weights of actor-critic networks;
  end if
  Obtain $u_k$ through actor network;
  Obtain stage cost $U_k$ using (12);
  Update critic network using (16);
  Update action network using (20);
  Use $u_k$ as input to the FSM-IC by updating the
  impedance parameters using (4) for the next gait cycle;
Until tracking criteria met
```

A. OpenSim Model Setup

We investigated this tracking control problem using OpenSim, a well-established simulator in the field of biomechanics [41]. A bipedal walking model, as shown in Fig. 3(a), includes a body of five rigid-segments, linked through a one degree of freedom pin joint and the pelvis was linked to the ground by a free joint to allow free movement. The model settings, such as segment length, body mass and inertial parameters, followed the lower limb OpenSim model [42]. In this study, the left knee was defined as intact while the right as prosthetic to enable locomotion in a single simulation model. The intact knee was controlled by a fixed set of impedance parameters settings while the prosthetic knee was controlled by the RL controller with its impedance parameters updated for each gait cycle.



(a)

(b)

Fig. 3. OpenSim model. (a) Five-rigid segment bipedal walking model; (b) The next gait is used to measure the tracking error as the prosthetic knee needs to copy the intact knee that is half gait ahead.

Simulating a gait cycle in OpenSim requires specifying initial model settings such as walking speed v_x and v_y, both knee angles θ_L and θ_R which are set to small numbers near stance position. We added artificial Gaussian noises to the state s and control u to simulate sensor noise and actuator noise, a consideration to make the simulations more realistic. The sensor noise magnitude was set at 20% of the performance tolerance bound and the actuator noise was 1% of the maximum action value.

B. Simulation Procedure

All simulations were carried out in trials. A trial is a continuous experiment of 500 gait cycles when RL tracking control is applied to tuning the impedance parameters of the robotic knee under different simulation scenarios. We performed two sets of simulations: training trials and testing trials. During training, we performed 30 training trials each from a randomly initialized controller, i.e., a set of randomly generated initially feasible impedance parameter settings, that allow the simulator to simulate balanced walking without falling. During testing, we randomly selected 10 successful controllers (i.e., 10 sets of impedance parameters after training) and applied those control policies (i.e., the actor network weights) as initial controller parameter settings for tracking new, untrained trajectory profiles. Then we tested each of the 10 controllers (policy network weights) to perform 30 new trials, each of which has a new set of randomly selected initial impedance parameters.

A trial is considered successful if the tracking error in (8) reached an error tolerance bound (Table II, bottom row). Specifically, for each of the 4 phases (Fig. 1(b)), if the tracking error was within the tolerance bound for 8 out of 10 consecutive gait cycles, tracking process in this phase was considered convergent. If all 4 phases had converged within 500 gait cycles, the trial was a success.

TABLE II
SAFETY BOUND AND TOLERANCE BOUND

	Phase 1	Phase 2	Phase 3	Phase 4
Safety bounds [Angle (rad), duration (%)]	[0.184, 12]	[0.131, 12]	[0.157, 12]	[0.105, 12]
Tolerance bounds [Angle (rad), duration (%)]	[0.0263, 2]	[0.0263, 2]	[0.0263, 2]	[0.0263, 2]

To ensure subjects safety (not stumble or fall), a safety bound was introduced based on the realistic conditions of balanced walking. Specifically, as shown in Table II (top row), the safety bound was set at 1.5 standard deviations above the knee kinematic peak values observed in each phase [43]. If the tracking errors exceed the safety bound, which means the prosthetic profile may place subjects in unsafe areas, the impedance parameters of the prosthetic knee will be reset to the initial impedance. Note however, the actor and critic network weights are retained for further training until meeting tracking criteria.

In obtaining all results, we set the weighting matrices in the stage cost (10) as: $R_s = \text{diag}(1,1)$ and $R_u = \text{diag}(0.1,0.1,0.1)$. For the critic network, we used 8 hidden layer neurons with hyperbolic activation function and we used a linear output layer. For the actor network, we used 6 hidden layer neurons with hyperbolic activation function and we also used a hyperbolic activation function for the output layer neuron so that the control inputs are constrained. For both networks, learning rate was 0.1. The actor-critic network was updated every gait cycle until reaching trial success.

C. Simulating Realistic Walking

To evaluate the efficacy of the proposed dHDP tracking control, we performed a systematic simulation study to evaluate human-robot walking performance under RL tracking control. We emulated three walking conditions: ground walking, walking on different terrains and at different paces.

Scenario 1 (Level Ground Tracking): The intact knee was operated by a fixed set of impedance parameters at all time while the robotic knee impedance parameters were controlled by RL controller. Note that, fixed impedance parameters still provide realistic control according to (2) as the intact knee was also controlled by FSM-IC. For both the training stage and the testing stage, the RL controller was required to successfully track the intact knee profile within 500 gait cycles, while the OpenSim setting was placed at constant walking pace under level ground condition.

Scenario 2 (Walking on Different Terrain): To simulate walking on different terrains, we provided different gait profiles for the intact knee which correspond to different impedance control parameters. The changing profiles of the intact knee were then the new moving targets for the prosthetic knee to track. Five randomly initialized impedance parameter sets that could enable balanced walking of the human-robot system were generated. The impedance parameters of the intact knee were randomly selected from this pool of 5 different gait profiles every 20 gait cycles. During training, the controller was required to successfully track the intact knee for 3 consecutive times with respective knee profiles. During testing, a new pool of 5 sets of impedance parameters was generated and used on the intact knee, the respective profiles of which were tracked by the robotic knee. Again, the controller was required to successfully track the intact knee profile for 3 consecutive times.

Scenario 3 (Changing Pace): We examined RL tracking performance when the pace changes. The simulated pace changes were implemented in a sequence of [100% → 112% → 100% → 88%] of the initial pace. Changing pace took place once the controllers successfully tracked the intact knee by meeting convergence criteria. The controller must complete the full sequence to complete the training stage. In the testing stage, the pace change was placed in a different order of [100% → 80% → 100% → 120%] which also signifies a greater variance than the respective conditions for training. Same as in the training stage, each controller must finish the full sequence to be counted as a success.

V. RESULTS

All three scenarios were simulated in both training and testing stages with the same human subject. Table III shows the tracking performance for all three scenarios.

TABLE III
SIMULATION RESULTS OF ALL THREE SCENARIOS

Scenario	Task	Success rate	Tuning steps
1	training	1	64.6 ± 84.72
1	testing	1	36.30 ± 57.22
2	training	0.97	55.76 ± 55.18
2	testing	1	19.16 ± 15.23
3	training	0.8	52.46 ± 32.75
3	testing	0.92	42.75 ± 37.87

In Scenario 1, a 100% success rate was achieved with 64.6 average steps to fully learn to track the intact knee. The RMS tracking error was reduced from 0.0588 to 0.0131 radian of peak angle and from 1.96% to 0.69% of phase duration. In the test stage, because a trained actor network was used in initialization, an improvement in tuning speed was observed from 64.6 to 36.3 average steps. The RMS tracking error has a similar performance.

In Scenario 2, a 97% success rate was achieved with 55.8 average steps for successfully tracking the intact knee each time. The RMS tracking error was reduced from 0.0581 to 0.0124 radian of peak angle and reduced from 1.82% to 0.66% of phase duration. In the test stage, the trained policy of actor network shows the ability to track the target profile rapidly. An average of 19.16 tracking steps with a 100% success rate shows the ability to track a changing target effectively. Fig. 4(a) shows an example of typical training and testing trials. The RMS tracking error has a similar outcome in both training and testing as they share the same success criteria.

Scenario 3 focuses on examining tracking performance reflected in gait duration. 80% success rate was achieved with 52.5 average steps for successfully tracking the intact knee at different paces. The RMS tracking error was reduced from 0.0573 to 0.0103 radian of peak angle and reduced from 1.90% to 0.92% of phase duration. In the test stage, the trained policy of actor network shows a slightly improvement on tracking speed. But it greatly improved the success rate from 0.8 to 0.92. Fig. 4(b) shows an example of typical training and testing trials. The RMS tracking error has a similar outcome in both training and testing as they share the same success criteria.

Fig. 5 shows RMS tracking errors for all trials in all simulation scenarios. A significant reduction in tracking error was observed under all conditions. The difference between training and testing is minor because they use the same randomization in initial impedance parameters and the same convergence criteria.

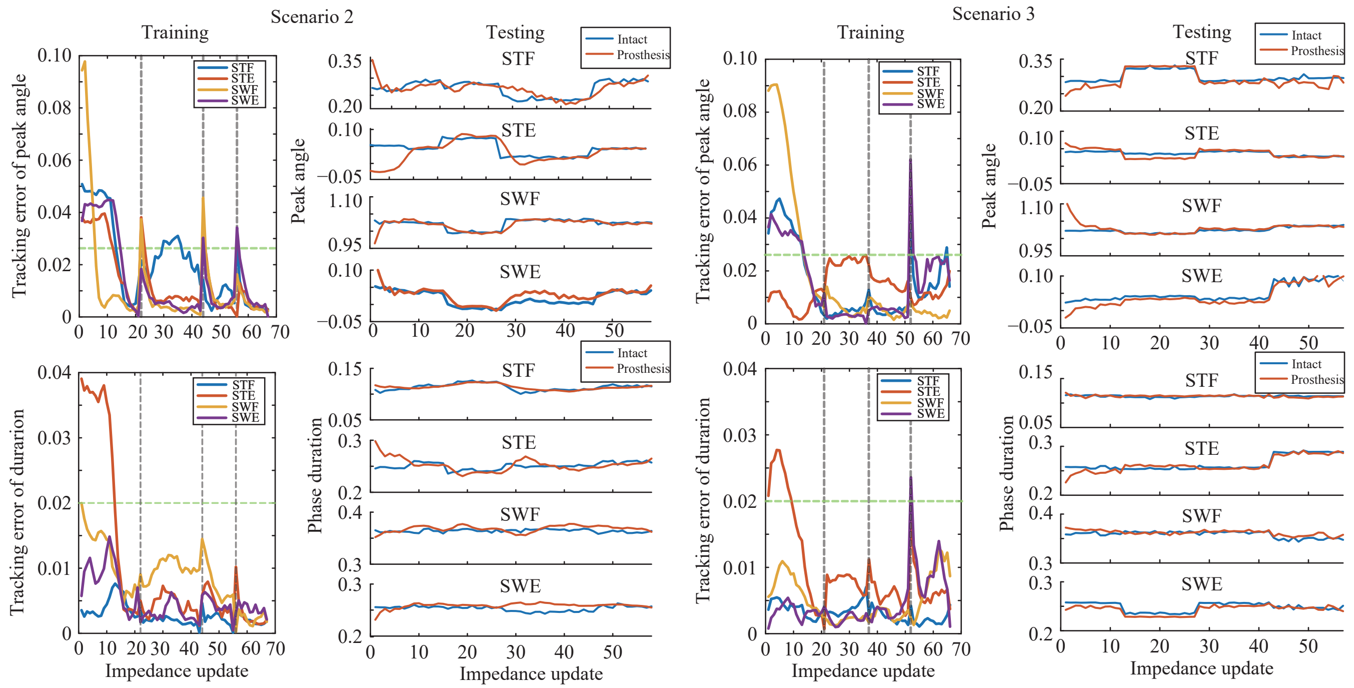
VI. CONCLUSIONS AND DISCUSSIONS

We have introduced a new RL based tracking control scheme for automatic tuning of robotic knee impedance parameter settings of a robotic knee to track the intact knee kinematics. For the first time, we successfully demonstrated stable and continuous walking in simulations of a human wearing a robotic prosthesis which was designed to track the intact knee motion.

Mirroring the intact knee motion by a prosthetic knee is an intuitive idea which has been proposed for decades, but has not been successfully demonstrated. The robotic knee control to mimic the intact knee joint is a tracking problem in classical control. Even though many control theoretic solutions exist, such as backstepping [44]–[46], observer-based control [47] and nonlinear adaptive/robust control [48], [49], they are inadequate for this problem as they require an accurate mathematical description of the system dynamics, which involve co-adapting human and robot in this case, and which are nearly impossible to obtain. Additionally, those control theoretic approaches focus on the stabilization (in the Lyapunov sense) of the nonlinear dynamic systems without addressing control performance such as the Bellman optimality.

Recently, some results emerged to tackle these issues using data-driven, learning enabled, nonlinear optimal tracking control designs [50]. Unfortunately, many of the reported results have focused on theoretical analyses, which are usually based on requiring a reference model for the desired movement trajectory and/or control trajectory. They are thus not practically useful.

In this paper, we have presented a complete tracking control algorithm based on dHDP. Additionally, we have systematically evaluated the performance of the proposed tracking controller. Our simulation results have shown effectiveness of the tracking controller for different walking tasks that emulate level ground walking, walking on different terrains and at different paces. Based on our previous work, we expect these new results to be verified in human experiments at a future time.



(a)

(b)

Fig. 4. Typical trials for Scenario 2 (a) and Scenario 3 (b). The left columns show tracking errors during training while the horizontal green dashed lines represent tolerance bounds and the grey vertical dashed lines represent task/pace transitions. The right columns show tracking trajectories during testing.

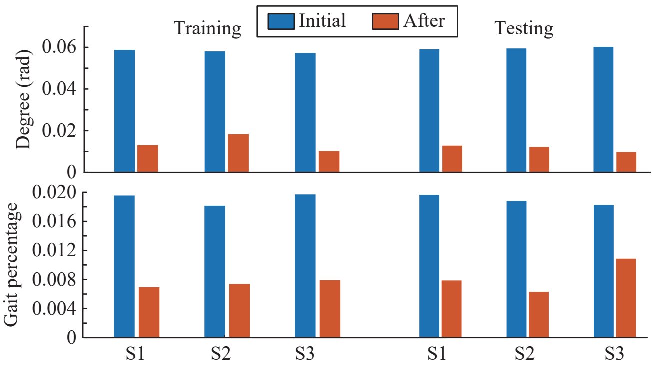


Fig. 5. Summary of RMS tracking errors of the peak angle (top) and phase duration (bottom) for all three scenarios. The blue bars represent the initial tracking errors while the red bars represent the final tracking errors.

APPENDIX

Proof of Lemma 2: The first difference of $L_{2,k}$ can be written as

$$\Delta L_{3,k} = \frac{1}{l_{a,k}\alpha_2} tr\left[\tilde{W}_{a2,k+1}^T \tilde{W}_{a2,k+1} - \tilde{W}_{a2,k}^T \tilde{W}_{a2,k}\right]. \tag{45}$$

According to (19) and (20), $\tilde{W}_{a2,k+1}$ can be rewritten as

$$\begin{aligned}\tilde{W}_{a2,k+1} &= \hat{W}_{a2,k+1} - W_{a2}^*\\ &= \hat{W}_{a2,k} - l_{a,k}\phi_{a,k}\hat{W}_{c2,k}C_k[\hat{W}_{c2,k}\phi_{c,k}]^T - W_{a2}^*\\ &= \tilde{W}_{a2,k} - l_{a,k}\phi_{a,k}\hat{W}_{c2,k}C_k[\hat{W}_{c2,k}\phi_{c,k}]^T.\end{aligned} \tag{46}$$

Based on the above expression, we can obtain

$$\begin{aligned}&tr\left[\tilde{W}_{a2,k+1}^T \tilde{W}_{a2,k+1}\right]\\ &\quad= \tilde{W}_{a2,k}^T \tilde{W}_{a2,k} + l_{a,k}^2 \|\phi_{a,k}\|^2 \|\hat{W}_{c2,k}C_k\|^2 \|\hat{W}_{c2,k}\phi_{c,k}\|^2\\ &\qquad - 2l_{a,k}\hat{W}_{c2,k}C_k[\hat{W}_{c2,k}\phi_{c,k}]^T \zeta_a(k).\end{aligned} \tag{47}$$

Substituting (47) into (45), we have

$$\begin{aligned}\Delta L_{3,k} = \frac{1}{\alpha_2}&(l_{a,k}\left\|\phi_{a,k}\right\|^2 \left\|\hat{W}_{c2,k}C_k\right\|^2 \left\|\hat{W}_{c2,k}\phi_{c,k}\right\|^2\\ &+ \left\|\hat{W}_{c2,k}\phi_{c,k} - \hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2 - \left\|\hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2\\ &- \left\|\hat{W}_{c2,k}\phi_{c,k}\right\|^2).\end{aligned} \tag{48}$$

Note that

$$\begin{aligned}&\left\|\hat{W}_{c2,k}\phi_{c,k} - \hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2 - \left\|\hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2\\ &\quad\le 2\left\|\hat{W}_{c2,k}\phi_{c,k}\right\|^2 + \left\|\hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2\\ &\quad\le 2\left\|\left(\tilde{W}_{c2,k} + W_{c2}^*\right)\phi_{c,k}\right\|^2 + \left\|\hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2\\ &\quad\le 2\left(\left\|\tilde{W}_{c2,k}\phi_{c,k}\right\| + \left\|W_{c2}^*\phi_{c,k}\right\|\right)^2 + \left\|\hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2\\ &\quad\le 4\left\|\zeta_{c,k}\right\|^2 + 4\left\|W_{c2}^*\phi_{c,k}\right\|^2 + \left\|\hat{W}_{c2,k}C_k\zeta_{a,k}\right\|^2.\end{aligned} \tag{49}$$

Then we obtain (29) by substituting (49) into (48).

The first difference of $L_{4,k}$ can be written as

$$\Delta L_{4,k} = \frac{1}{l_{a,k}\alpha_3} tr\left[\tilde{W}_{a1,k+1}^T \tilde{W}_{a1,k+1} - \tilde{W}_{a1,k}^T \tilde{W}_{a1,k}\right]. \tag{50}$$

According to (19) and (20), $\tilde{W}_{a1,k+1}$ can be rewritten as

$$\begin{aligned}\tilde{W}_{a1,k+1} &= \hat{W}_{a1,k+1} - W_{a1}^*\\ &= \tilde{W}_{a1,k+1} - l_{a,k}\hat{W}_{c2,k}\phi_{c,k}D_kC_k^T\hat{W}_{c2,k}^T s_k^T.\end{aligned} \tag{51}$$

Let us consider

$$\begin{aligned}&tr\left[\tilde{W}_{a1,k}^T \tilde{W}_{a1,k}\right]\\ &\quad= \tilde{W}_{a1,k}^T \tilde{W}_{a1,k} + l_{a,k}^2 \left\|\hat{W}_{c2,k}\phi_{c,k}\right\|^2 \left\|\hat{W}_{c2,k}C_kD_k^T\right\|^2 \|s_k\|^2\\ &\qquad - 2l_{a,k}\hat{W}_{c2,k}C_kD_k^T\phi_{c,k}^T\hat{W}_{c2,k}^T\tilde{W}_{a1,k}s_k.\end{aligned} \tag{52}$$

Then, by using the cyclic property of matrix trace, the last term in (52) is bounded by

$$\begin{aligned}&-2l_{a,k}\hat{W}_{c2,k}C_kD_k^T\phi_{c,k}^T\left(\hat{W}_{c2,k}\right)^T\tilde{W}_{a1,k}s_k\\ &\quad\le l_{a,k}(\|\hat{W}_{c2,k}\phi_{c,k}\|^2+\|\hat{W}_{c2,k}C_kD_k^T\|^2\|\tilde{W}_{a1,k}s_k\|^2). \quad (53)\end{aligned}$$

We obtain (30) by substituting (52), (53) into (50). ■

REFERENCES

[1] E. C. Martinez-Villalpando and H. Herr, "Agonist-antagonist active knee prosthesis: A preliminary study in level-ground walking," *Journal of Rehabilitation Research & Development*, vol. 46, no. 3, 2009.

[2] J. L. Johansson, D. M. Sherrill, P. O. Riley, P. Bonato, and H. Herr, "A clinical comparison of variable-damping and mechanically passive prosthetic knee devices," *American Journal of Physical Medicine & Rehabilitation*, vol. 84, no. 8, pp. 563–575, 2005.

[3] E. J. Rouse, L. J. Hargrove, E. J. Perreault, and T. A. Kuiken, "Estimation of human ankle impedance during the stance phase of walking," *IEEE Trans. Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 870–878, 2014.

[4] S. Pfeifer, H. Vallery, M. Hardegger, R. Riener, and E. J. Perreault, "Model-based estimation of knee stiffness," *IEEE Trans. Biomedical Engineering*, vol. 59, no. 9, pp. 2604–2612, 2012.

[5] R. D. Gregg, T. Lenzi, N. P. Fey, L. J. Hargrove, and J. W. Sensinger, "Experimental effective shape control of a powered transfemoral prosthesis," in *Proc. IEEE 13th Int. Conf. Rehabilitation Robotics*, 2013, pp. 1–7.

[6] M. F. Eilenberg, H. Geyer, and H. Herr, "Control of a powered ankle–foot prosthesis based on a neuromuscular model," *IEEE Trans. Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, pp. 164–173, 2010.

[7] H. Huang, D. L. Crouch, M. Liu, G. S. Sawicki, and D. Wang, "A cyber expert system for auto-tuning powered prosthesis impedance control parameters," *Annals of Biomedical Engineering*, vol. 44, no. 5, pp. 1613–1624, 2016.

[8] C. Lu, J. Si, and X. Xie, "Direct heuristic dynamic programming for damping oscillations in a large power system," *IEEE Trans. Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 4, pp. 1008–1013, 2008.

[9] W. Guo, F. Liu, J. Si, D. He, R. Harley, and S. Mei, "Approximate dynamic programming based supplementary reactive power control for DFIG wind farm to enhance power system stability," *Neurocomputing*, vol. 170, pp. 417–427, 2015.

[10] W. Guo, F. Liu, J. Si, D. He, R. Harley, and S. Mei, "Online supplementary ADP learning controller design and application to power system frequency control with large-scale wind energy integration," *IEEE Trans. Neural Networks and Learning Systems,* vol. 27, no. 8, pp. 1748–1761, 2015.

[11] R. Enns and J. Si, "Helicopter flight control design using a learning control approach," in *Proc. 39th IEEE Conf. Decision and Control*, vol. 2, 2000, pp. 1754–1759.

[12] R. Enns and J. Si, "Apache helicopter stabilization using neural dynamic programming," *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 1, pp. 19–25, 2002.

[13] R. Enns and J. Si, "Helicopter trimming and tracking control using direct neural dynamic programming," *IEEE Trans. Neural Networks*, vol. 14, no. 4, pp. 929–939, 2003.

[14] Y. Wen, X. Gao, J. Si, A. Brandt, M. Li, and H. H. Huang, "Robotic knee prosthesis real-time control using reinforcement learning with human in the loop," in *Proc. Int. Conf. Cognitive Systems and Signal Processing*. Springer, 2018, pp. 463–473.

[15] Y. Wen, J. Si, X. Gao, S. Huang, and H. H. Huang, "A new powered lower limb prosthesis control framework based on adaptive dynamic programming," *IEEE Trans. Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 2215–2220, 2016.

[16] Y. Wen, J. Si, A. Brandt, X. Gao, and H. Huang, "Online reinforcement learning control for the personalization of a robotic knee prosthesis," *IEEE Trans. Cybernetics*, vol. 50, no. 6, pp. 2346–2356, 2020.

[17] Y. Wen, M. Li, J. Si, and H. H. Huang, "Wearer-prosthesis interaction for symmetrical gait: A study enabled by reinforcement learning prosthesis control," *IEEE Trans. Neural Systems and Rehabilitation*, vol. 28, no. 4, pp. 904–913, 2020.

[18] X. Gao, Y. Wen, M. Li, J. Si, and H. H. Huang, "Robotic knee parameter tuning using approximate policy iteration," in *Proc. Int. Conf. Cognitive Systems and Signal Processing*. Springer, 2018, pp. 554–563.

[19] M. Li, X. Gao, Y. Wen, J. Si, and H. H. Huang, "Offline policy iteration based reinforcement learning controller for online robotic knee prosthesis parameter tuning," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2019, pp. 2831–2837.

[20] M. Li, Y. Wen, X. Gao, J. Si, and H. Huang, "Toward expedited impedance tuning of a robotic prosthesis for personalized gait assistance by reinforcement learning control," *IEEE Trans. Robotics*, 2021. DOI: 10.1109/TRO.2021.3078317

[21] X. Gao, J. Si, Y. Wen, M. Li, and H. Huang, "Reinforcement learning control of robotic knee with human-in-the-loop by flexible policy iteration," *IEEE Trans. Neural Networks and Learning Systems*, 2021. DOI: 10.1109/TNNLS.2021.3071727

[22] B. Pietraszewski, S. Winiarski, and S. Jaroszczuk, "Three-dimensional human gait pattern–reference data for normal men," *Acta of Bioengineering and Biomechanics*, vol. 14, no. 3, pp. 9–16, 2012.

[23] A. N. Lay, C. J. Hass, and R. J. Gregor, "The effects of sloped surfaces on locomotion: A kinematic and kinetic analysis," *Journal of Biomechanics*, vol. 39, no. 9, pp. 1621–1628, 2006.

[24] A. S. Voloshina, A. D. Kuo, M. A. Daley, and D. P. Ferris, "Biomechanics and energetics of walking on uneven terrain," *Journal of Experimental Biology*, vol. 216, no. 21, pp. 3963–3970, 2013.

[25] G. D'Angelo, Y. Thibaudier, A. Telonio, M.-F. Hurteau, V. Kuczynski, C. Dambreville, and A. Frigon, "Modulation of phase durations, phase variations, and temporal coordination of the four limbs during quadrupedal split-belt locomotion in intact adult cats," *Journal of Neurophysiology*, vol. 112, no. 8, pp. 1825–1837, 2014.

[26] M. G. Bernal-Torres, H. I. Medellín-Castillo, and J. C. Arellano-González, "Design and control of a new biomimetic transfemoral knee prosthesis using an echo-control scheme," *Journal of Healthcare Engineering*, vol. 2018, 2018.

[27] D. Joshi, R. Singh, R. Ribeiro, S. Srivastava, U. Singh, and S. Anand, "Development of echo control strategy for AK prosthesis: An embedded system approach," in *Proc. IEEE Int. Conf. Systems in Medicine and Biology*, 2010, pp. 143–147.

[28] S. Sahoo, D. K. Pratihar, and S. Mukhopadhyay, "A novel supervisory control scheme to tackle variations in step length for walking with powered ankle prosthesis," *Biomedical Signal Processing and Control*, vol. 46, pp. 212–220, 2018.

[29] S. Kumar, A. Mohammadi, D. Quintero, S. Rezazadeh, N. Gans, and R. D. Gregg, "Extremum seeking control for model-free auto-tuning of powered prosthetic legs," *IEEE Trans. Control Systems Technology*, vol. 28, no. 6, pp. 2120–2135, 2020.

[30] E. R. Westervelt, J. W. Grizzle, and D. E. Koditschek, "Hybrid zero dynamics of planar biped walkers," *IEEE Trans. Automatic Control*, vol. 48, no. 1, pp. 42–56, 2003.

[31] R. Wu, M. Li, Z. Yao, J. Si, *et al.*, "Reinforcement learning enabled automatic impedance control of a robotic knee prosthesis to mimic the intact knee motion in a co-adapting environment," arXiv preprint arXiv: 2101.03487, 2021.

[32] C. T. Leonard, R. Craik, and C. Oatis, "The neurophysiology of human locomotion," in *Gait Analysis: Theory and Application*. Mosby-Year, 1995.

[33] N. Hogan, "Impedance control: An approach to manipulation: Part II–Implementation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 8–16, 1985.

[34] G. Taga, Y. Yamaguchi, and H. Shimizu, "Self-organized control of bipedal locomotion by neural oscillators in unpredictable environment," *Biological Cybernetics*, vol. 65, no. 3, pp. 147–159, 1991.

[35] H. Geyer, A. Seyfarth, and R. Blickhan, "Positive force feedback in bouncing gaits?" *Proc. the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. 1529, pp. 2173–2183, 2003.

[36] K. Shamaei, G. S. Sawicki, and A. M. Dollar, "Estimation of quasistiffness of the human knee in the stance phase of walking," *PloS One*, vol. 8, no. 3, p. e59993, 2013.

[37] S. Huang, J. P. Wensman, and D. P. Ferris, "Locomotor adaptation by transtibial amputees walking with an experimental powered prosthesis under continuous myoelectric control," *IEEE Trans. Neural Systems and Rehabilitation Engineering*, vol. 24, no. 5, pp. 573–581, 2015.

[38] J. Si and Y.-T. Wang, "Online learning control by association and reinforcement," *IEEE Trans. Neural Networks*, vol. 12, no. 2, pp. 264–276, 2001.

[39] Z. Yao, J. Si, R. Wu, and J. Yao, "Toward reliable designs of data-driven reinforcement learning tracking control for euler-lagrange systems," arXiv preprint arXiv: 2101.00068, 2020.

[40] A. N. Michel and R. K. Miller, *Qualitative Analysis of Large Scale Dynamical Systems*. Academic Press, 1977.

[41] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "OpenSim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Trans. Biomedical Engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.

[42] D. Jacobs, "From the ground up: Building a passive dynamic walker model," 2014. [Online]. Available: https://simtkconfluence.stanford.edu:8443/display/OpenSim33

[43] M. P. Kadaba, H. Ramakrishnan, and M. Wootten, "Measurement of lower extremity kinematics during level walking," *Journal of Orthopaedic Research*, vol. 8, no. 3, pp. 383–392, 1990.

[44] M. Krstic, P. V. Kokotovic, and I. Kanellakopoulos, *Nonlinear and Adaptive Control Design*. John Wiley & Sons, Inc., 1995.

[45] Z.-P. Jiang and H. Nijmeijer, "Tracking control of mobile robots: A case study in backstepping," *Automatica*, vol. 33, no. 7, pp. 1393–1399, 1997.

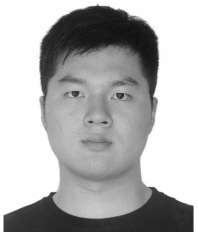
[46] K. D. Do, Z.-P. Jiang, and J. Pan, "Simultaneous tracking and stabilization of mobile robots: An adaptive approach," *IEEE Trans. Automatic Control*, vol. 49, no. 7, pp. 1147–1151, 2004.

[47] H. K. Khalil and J. W. Grizzle, *Nonlinear Systems*. Prentice Hall Upper Saddle River, NJ, 2002, vol. 3.

[48] H. Nijmeijer and A. Van der Schaft, *Nonlinear Dynamical Control Systems*. Springer, 1990, vol. 175.

[49] A. Isidori, *Nonlinear Control Systems*. Springer Science & Business Media, 2013.

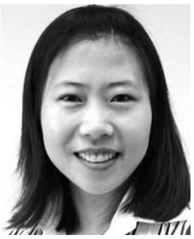
[50] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3–35, 2013.



Ruofan Wu received the B.S. degree in telecommunications engineering from Beijing University of Posts and Telecommunications, China, in 2014, and the M.S. degree in Electrical Engineering from Arizona State University, USA, in 2017. He is currently a Ph.D. candidate at the School of Electrical, Computer and Energy Engineering, Arizona State University, USA. His research interests include adaptive control, reinforcement learning and robotics control.

Zhikai Yao received the B.Tech. degree in 2014 from the Qingdao University of Technology, Qingdao, China, and the Ph.D. degree in 2021 from Nanjing University of Science and Technology, Nanjing, China. He has been an exchange student with the School of Electrical, Computer and Energy Engineering, Arizona State University, USA, since 2019. His research interests include intelligent control of mechanical systems, reinforcement learning and human-robot interaction.

Jennie Si (Fellow, IEEE) received the B.S. degree in 1985, the M.S. degree in 1988 from Tsinghua University, China, and the Ph.D. degree in 1992 from the University of Notre Dame, USA. She has been a Faculty Member with the School of Electrical, Computer and Energy Engineering, Arizona State University, USA, since 1991. Her research focuses on reinforcement learning control which utilizes machine learning and neural network approximations. She is also interested in fundamental neuroscience of the frontal cortex and its role in decision and control processes. Dr. Si was a Recipient of the NSF/White House Presidential Faculty Fellow Award in 1995 and the Motorola Engineering Excellence Award in 1995. She is a Distinguished Lecturer of the IEEE Computational Intelligence Society. She consulted for Intel, Arizona Public Service, and Medtronic. She has served on several professional organizations' executive boards and international conference committees. She was an Advisor to the NSF Social Behavioral and Economical Directory. She served on several proposal review panels. She was an Associate Editor for the *IEEE Transactions on Semiconductor Manufacturing*, the *IEEE Transactions on Automatic Control*, and an Action Editor for *Neural Networks*. She is currently an Associate Editor for the *IEEE Transactions on Neural Networks and Learning Systems*.



He (Helen) Huang (Senior Member, IEEE) received the B.S. degree in electronic and information engineering from Xi'an Jiaotong University in 2000, and the M.S. and Ph.D. degrees in biomedical engineering from Arizona State University, USA, in 2002 and 2006, respectively. She is currently the Jackson Family Distinguished Professor with the joint Department of Biomedical Engineering, North Carolina State University, USA, and the University of North Carolina at Chapel Hill, USA, and the Director for the Closed-Loop Engineering for Advanced Rehabilitation (CLEAR) core. Her research interest lies in neural-machine interfaces for prostheses and exoskeletons, human-robot interaction, adaptive and optimal control of wearable robots, and human movement control. Dr. Huang was the Recipient of the Delsys Prize for Innovation in Electromyography, the Mary E. Switzer Fellowship with NIDRR, and a NSF CAREER Award. She is a Fellow of AIMBE and Member of the Society for Neuroscience, BMES, and ASB. She is currently an Associate Editor for the *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, the *Journal of Neuroengineering and Rehabilitation*, and the *Wearable Techologies*, and is a Guest Associate Editor for the *IEEE Transactions on Robotics*.