Appendix A Stability Analysis of the Multi-controller

In the stability analysis of the controllers, the auxiliary controller and the model-based controller together are considered as controller C. For getting an understanding of the system behavior in the presence of inaccurate dynamic models, the stability is proved on the single DOF system in this section.

The block diagram of Fig. A1 represents the dynamic behavior of the exoskeleton leg. The dynamic H represents how the kinematics of the pilot limb (e.g. velocity, position or a combination thereof) transfer to the pilot forces imposed on the exoskeleton d.

The physical properties of human dynamics determine H. G represents the transfer function from the actuator input r to the exoskeleton angular velocity v. the exoskeleton angular velocity is affected by the equivalent human torque through the sensitivity transfer function S. S maps the equivalent pilot torques d into the exoskeleton velocity v. C is the controller operating on the exoskeleton variables. The C consists of two parts, one part $\hat{C}$ is the model-based controller that based on estimated dynamic model of exoskeletons, another part M is the compensation controller that compensates the case that the dynamic model parameters of exoskeleton is greater than the estimated dynamic model.

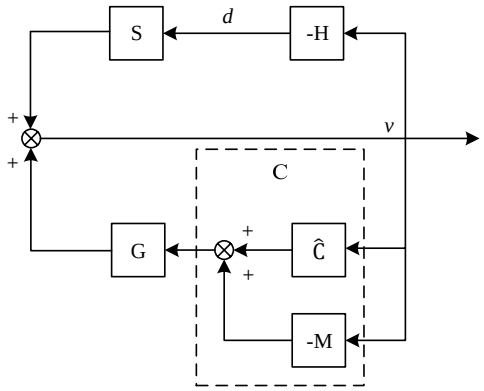


Figure A1 This block diagram represented the dynamic behavior of exoskeleton legs. The upper loop shows how the pilot moves the exoskeleton through applied forces. The lower loop shows that the exoskeleton is driven by the controller.

In Fig. A1 the feedback loop as shown, the closed-loop sensitivity transfer function is written as:

$$S_{NEW} = \frac{v}{d} = \frac{S}{1 + SH - G(\hat{G} + M)}. \tag{A1}$$

The closed-loop characteristic equation as (A1) decides the stability of the system shown in Fig. A1.

$$1 + SH - G(\hat{G} + M) = 0. \tag{A2}$$

Under the case that absence the feedback controller C, entire load included the payload and weight of the exoskeleton are carried by the pilot. The stability is decided by the characteristic equation as follow:

$$1 + SH = 0, \tag{A3}$$

characteristic equation is always stable.

When the feedback loop C is added, the stability is analyzed from two cases:

1. Accurate dynamic model: When the dynamic model of the exoskeleton is accurate, the M is regulated adaptively by RL to zero, then $\hat{C} = C$. The transfer formulation can be rewritten:

$$S_{NEW} = \frac{v}{d} = \frac{S}{1 + SH - GC}. \tag{A4}$$

   Using the Small Gain Theorem (SGT), one can be shown that as long as (A5) is satisfied, the closed-loop stability is guaranteed:

$$|GC| < |1 + SH|. \tag{A5}$$

   According to $C = (1 - \alpha^{-1})G^{-1}$ [6], the $|GC| < 1$ and without any uncertainties, (A5) is guaranteed as long as $1 < |1 + SH|$. Even though the feedback loop containing C is positive, the overall system of the pilot and the exoskeleton can be stabilized by the feedback loop containing H.

2. Inaccurate dynamic model: When the dynamic model of the exoskeleton is inaccurate, the M is regulated to make $\hat{C} - M$ approximate to C. So, in this case, the system is guaranteed as long as $1 < |1 + SH|$.

Appendix B Numerical Simulation on Single DOF Exoskeleton

Appendix B.1 Single DOF Exoskeleton

The single DOF exoskeleton platform is shown in Fig. B1, the single DOF exoskeleton consists of a thigh and a shank, and they are connected by a joint which is powered by a bi-directional linear hydraulic actuator. The torque τ_e gained from the model-based controller and the auxiliary controller provides the torque τ_c. A pilot leg is attached to the exoskeleton leg by compliant connections, in which compliant connections can allow the human leg to swing freely. The pilot can impose forces on the exoskeleton leg through the compliant connections. In other words, the equivalent torque τ_h resulted from the pilot's applied forces are imposed on the joint. The dynamic model of the single DOF exoskeleton is written as:

$$J\ddot{\theta_e} + B\dot{\theta_e} + mgl \cdot \sin\theta_e = \tau_e + \tau_c + \tau_h, \tag{B1}$$

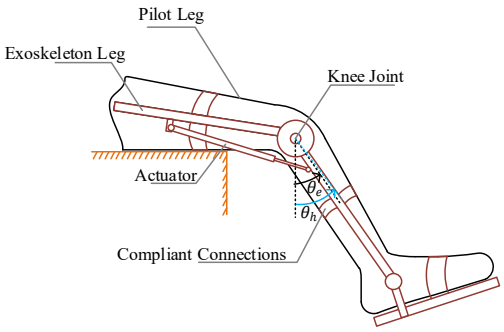


Figure B1 The single DOF exoskeleton with pilots leg.

where J, B, m and l represent the inertial moment, viscous friction coefficient, exoskeleton shank mass and length of the single DOF exoskeleton respectively. The joint states $\theta_e, \dot{\theta_e}, \ddot{\theta_e}$ represent the angle, angular velocity and angular acceleration of the knee joint, the gravitational constant is represented as g. In order to manage the case that the value of dynamic model's parameter obtained by system identification is less than the dynamic model of actual exoskeleton, we employ two controllers to control the exoskeleton system and the coefficients of two controllers are adapted by RL.

The schematic diagram is shown that the pilot's leg is attached with the exoskeleton by compliant connections in the thigh and the shank in Fig. B1. The movement of the single DOF exoskeleton is limited in swing movements with a knee joint. In the experiments, An encoder embedded in the knee joint is used to measure the angle information of exoskeleton system. In order to simplify the complexity of exoskeleton's sensor system, there are not force sensors integrated on the exoskeleton to measure the interactive forces between pilot and exoskeleton, so we model the dynamic model of pHRI as a spring-damper model which based on angle and angular velocity error between pilot and exoskeleton to label the interactive forces. The spring-damper model is designed as:

$$\tau_h = k_p(\theta_h - \theta_e) + k_d(\dot{\theta}_h - \dot{\theta}_e), \tag{B2}$$

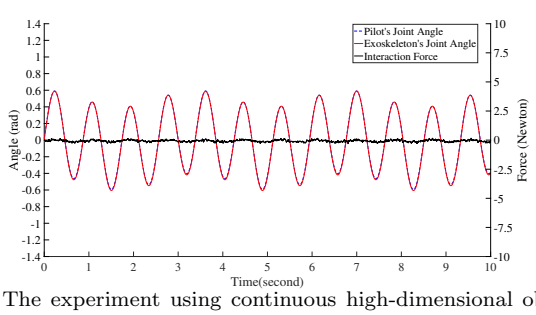
where k_p and k_d are stiffness and damper parameters, respectively which are set to a value similar to the compliant connection of the real-life exoskeleton system.

Appendix B.2 Experimental Setup

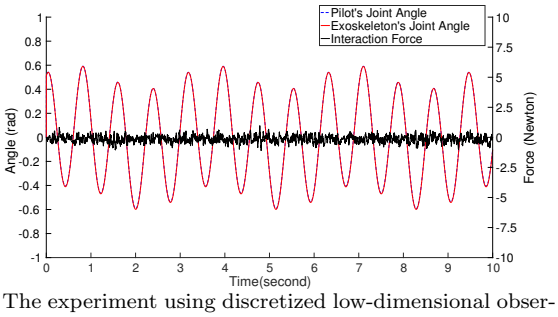
The framework of multi-controller proposed in previous works is used to adapt changing with different pilots and walking patterns in the presence of inaccurate dynamic model. So we firstly detail the advantages of ILAC through differnet comparison experiments with different walking patterns. In this section, we carry out comparison experiments using discretized low-dimensional observation spaces and using continuous high-dimensional observation spaces. This paper mainly focuses on the advantages contributed by ILAC with continuous high-dimensional observation spaces. So we only concern the learning of multi-controller, the motion trajectories of the pilots are obtained from DMPs models with parameters which have been learned. In the experiments on the single DOF exoskeleton, the pilot's motion trajectories are set as periodic sine waves but with different frequencies and amplitudes. the weight parameters k_1 and k_2 of the immediate reward r_t are chosen as 1 and 2.

In the experiments using the discretized observation spaces, the state spaces $s \in [(\theta_{e,max}, \dot{\theta}_{e,max},)$, $(\theta_{e,min}, \dot{\theta}_{e,min})]$ are identified at 100Hz (i.e. the time interval of s_t and s_{t+1} is 0.01 second) and the step size of the state space is $[(\theta_{e,max}, \dot{\theta}_{e,max},) - (\theta_{e,min}, \dot{\theta}_{e,min})]/100$. The sensitivity factors used as the action spaces are discretized with the same discretized principle of the state spaces.

In the ILAC, the continuous three-dimensional state spaces $s \in (\theta_e, \dot{\theta}_e, \ddot{\theta}_e)$ are directly used for input of the actor-critic reinforcement learning, and the three-dimensional continuous action space $a \in (\alpha, P, D)$ is the output of the actor-critic reinforcement learning in the ILAC framework. The reward discount γ is 0.9 and the step size τ is set to 0.9. The weights θ^Q and θ^μ are initialized randomly.



(a) The experiment using continuous high-dimensional observation spaces



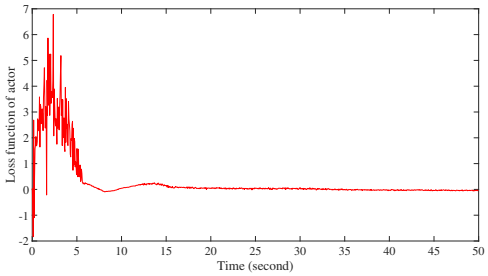
(b) The experiment using discretized low-dimensional observation spaces

Figure B2 The learning process of the proposed ILAC with a sine wave: (a) the experiment using continuous high-dimensional observation spaces; (b) the experiment using discretized low-dimensional obser-vation spaces.

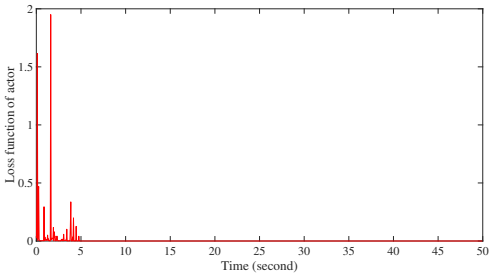
Appendix B.3 Experimental Results

In fist comparison experiments, a sine wave is defined the pilot's motion trajectory to compare the performances of the algorithms using continuous high-dimensional observation spaces and using discrete low-dimensional observation spaces. The frequency of the sine wave is approximately 2π rad/s and the amplitude is set to 0.5 rad. Fig. B2 show the performance of the two algorithms. We can find that the interaction force of ILAC is smaller and smoother than the algorithm using discrete low-dimensional observation spaces (normalized Mean Squared Errors (nMSE): 0.002 rad compare to 0.003 rad) as Tab. B1. So that, according to the comparison results of the two figures, the performance of ILAC is better than algorithm using discrete low-dimensional observation spaces.

Fig. B3 shows the learning process of ILAC, The actor's loss function took 20 seconds to converge and the critic's loss function converge to zero in 10 seconds. The conclusion from Fig. B3 is that the learning performance of ILAC can meet the requirements of our practical application.



(a) Evolution of the actor's loss function



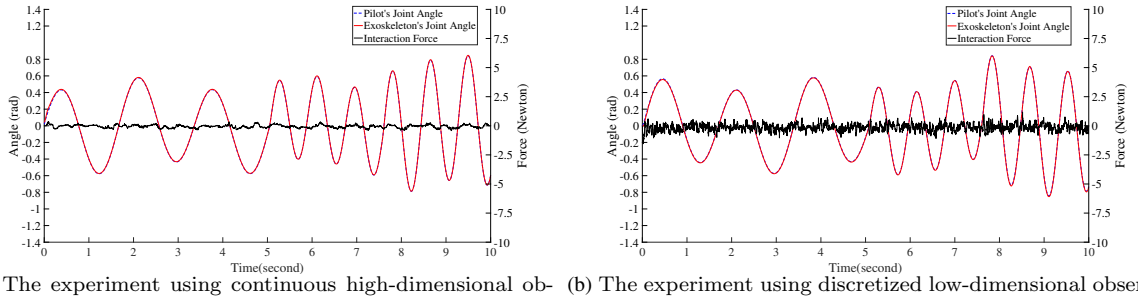
(b) Evolution of the critic's loss function

Figure B3 The learning process of the proposed ILAC with a sine wave: (a) evolution of the actor's loss function; (b) evolution of the critic's loss function.

Second experiments is that we use different frequencies to mimic the different walking speeds of pilot, and the different walking patterns are mimicked by different amplitudes. In which the frequencies are approximate 1.25π rad and 2.5π rad(3 cycles for each frequency). Fig. B4 describes the process for 8 gait cycles in the whole process of the experiments. Fig. B4(a) shows the performances of the experiment using continuous high-dimensional observation spaces, Fig. B4(b) shows the feature of the experiment using discretized low-dimensional observation spaces in a single DOF exoskeleton, where the interaction force calculated by the spring-damper model (B2) map out the advantages of the proposed ILAC. In which, the interaction force is not decreased significantly, but more smooth . The results of the experiments show that the proposed ILAC can improve the interactive performance with different pilots and different walking patterns.

During the experiments with different walking speeds and different walking patterns, the learning processes are shown as Fig. B5. In this experiment, the exoskeleton fellows the pilot's motion, after several gait cycles the loss functions of the actor and the critic convergence to zero. The actor network and critic network can converge to stable values quickly. The convergence process of the actor's loss function lasted about 24s as Fig. B5(a), the convergence process of the critic's loss function lasted about 3.5s as Fig. B5(b).

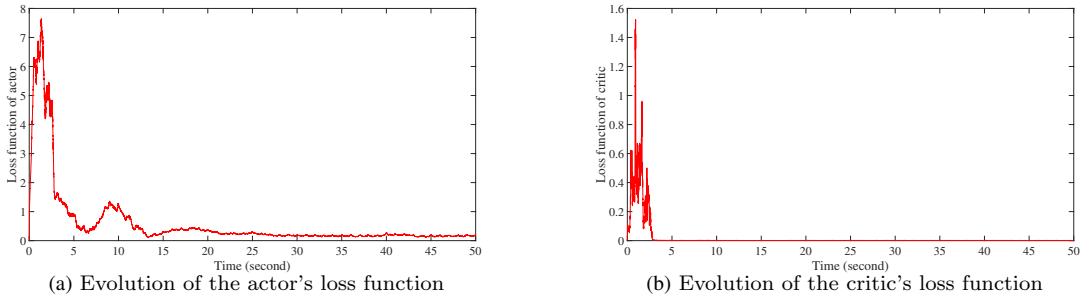
The framework of multi-controller is mainly utilized to adapted the errors between dynamic model obtained by system identification method and dyanmic model of actual exoskeleton in the presence of the changing with different pilots and different walking patterns. So we manually set dynamic model errors mimicing different pilots to verify the effectiveness of ILAC. In this experiments, the errors are set as 10% and 20%. The interaction states are visually shown in the presence of different error conditions. The quantization performance of ILAC in response to different error conditions is shown in Tab. B1. We can get the conclusion from the Fig. B6 and Tab. B1 that the performance of ILAC is better than the methods using discrete observation spaces.



(a) The experiment using continuous high-dimensional observation spaces

(b) The experiment using discretized low-dimensional observation spaces

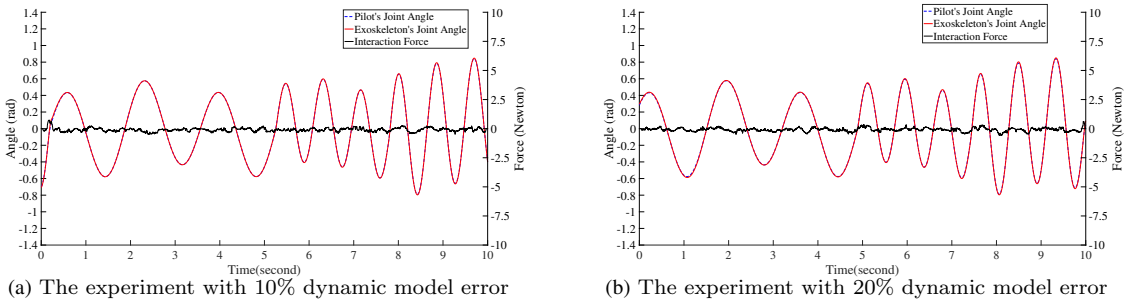
Figure B4 Comparison of performance with different pilots and different walking patterns: (a) the experiment using continuous high-dimensional observation spaces; (b) the experiment using discretized low-dimensional observation spaces.



(a) Evolution of the actor's loss function

(b) Evolution of the critic's loss function

Figure B5 The learning process of the proposed ILAC: (a) evolution of the actor's loss function; (b) evolution of the critic's loss function.



(a) The experiment with 10% dynamic model error

(b) The experiment with 20% dynamic model error

Figure B6 The experiment using continuous high-dimensional observation spaces (a) with 10% dynamic model error and (b) with 20% dynamic model error.

Table B1 Comparison Between Discrete and Continuous Domains with Sine Wave

nMSE(rad) (Domain State)	10% model error	20% model error	Fix frequency	Variable frequency
Discrete	0.009	0.010	0.003	0.004
Continuous	0.003	0.004	0.002	0.003

Appendix C Experiments on HUALEX System

Appendix C.1 HUALEX System

HUALEX system is designed as an anthropomorphic, robust and lightweight equipment for enhancing the strength and endurance of pilot with a pair of wearable robotic legs. As shown in Fig. C1, in order to enable HUALEX system to verify various algorithms, we design the hip joints and the knee joints as active joints. Each of them is activated by a hydraulic cylinder. The ankle joints are designed as an energy-storage mechanism which stores energy in stance phase and releases it in swing phase. Many compliant connections are used to connect pilot to HUALEX. In HUALEX system, all the power comes from a hydraulic system set in the backpack.

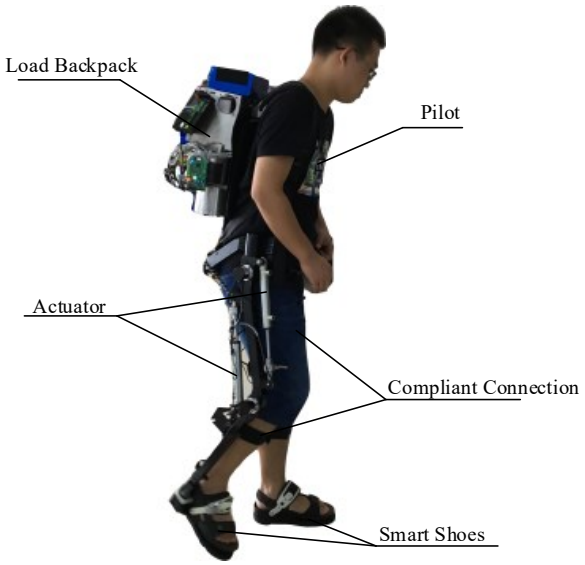


Figure C1 HUALEX with the pilot. The pilot and HUALEX are connected with compliant connections. The hip joints and knee joints are driven with bi-directional linear hydraulic cylinders. The smart shoes are used for detecting gait phases.

In order to promote the performance of HUALEX system, a distributed control system consisted of a main controller and four node controllers embedded in the exoskeleton system. The main controller is used for running the algorithm, and the node controllers are used for collecting the sensor data for the main controller and controlling the actuators. The main controller communicates with the node controllers via Controller Area Network (CAN) that can ensure the real-time performance of control system.

In HUALEX sensory system, a total of three kinds of sensors are embedded:

- Encoders: Encoders are integrated in active joints which can measure the current state of HUALEX system.
- IMU: The IMU sensor is installed on the backpack which aiming to measure walking velocity of the exoskeleton.
- Plantar sensors: The plantar sensors collect the foot pressure which can be utilized to judge the current phase of HUALEX system.

Appendix C.2 Dynamic models of HUALEX

The proposed ILAC is actually a model-based control strategy, the dynamic model of HUALEX system can be described with a general form, which has different parameter vectors in different walking phase:

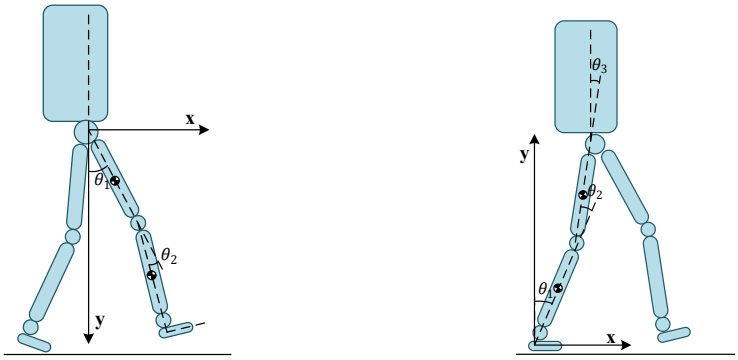
$$M(\Theta)\ddot{\Theta} + C(\Theta, \dot{\Theta})\dot{\Theta} + G(\Theta) = T_e + T_c + T_h, \tag{C1}$$

where Θ is the vector of each active joint angles, T_e, T_c and T_h represent the input torques from HUALEX and the pilot's torques respectively. $M(\Theta)$ is inertia matrix, $C(\Theta, \dot{\Theta})$ is coriolis matrix, and $G(\Theta)$ is a vector of gravitational torques. In this paper, the whole lower extremities are separated in swing leg and stance leg during locomotions as Fig. C2.

- Swing phase: The swing leg of HUALEX is modeled as a two-link mechanism with a base coordinate on the backpack, where the dimension of dynamic parameters $M(\Theta)$, $C(\Theta, \dot{\Theta})$ and $G(\Theta)$ are two.
- Stance phase: The stance leg of HUALEX is modeled as a three-link mechanism with a base coordinate on the ankle joint, where the dimension of dynamic parameters are three.

Appendix C.2.1 *Experimental Setup*

In this experiments, Three pilots (A, B, C) with different heights (168cm, 170cm, 180cm) are chosen to operate HUALEX system respectively. HUALEX system is operated in different environments (flat, slope and stair) with different walking speeds. In order to validate the effectiveness of ILAC in different environments, each pilot operates HUALEX system in



(a) Swing phase (b) Stance phase

Figure C2 Sagittal plane representation of HUALEX. The swing leg is model as a two serial link with a base coordinate fixed on its back in swing phase. The stand leg is represented as a three link with a base coordinate fixed on the ankle joint in stance phase.

Table C1 Compensation of loss fuction's convergence time

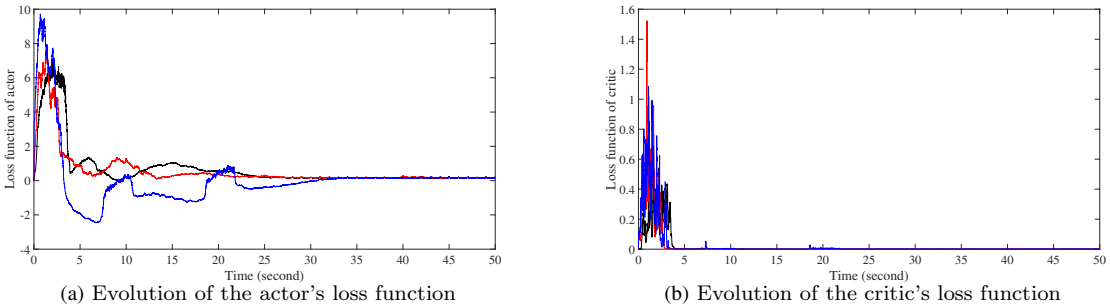
Time(Second) (Actor|critic)	Pilot A	Pilot B	Pilot C
Actor	23	24	32
Critic	4	3.5	3

three environments in different orders. Before each pilot operates HUALEX system, the weights of the actor network and critic network are initialized randomly.

Before starting the experiments, we have employed the DMPs to model the motion trajectories of pilots and the DMPs have been updated incrementally by LWR from clinical gait analysis dataset. Then we start the experiments which learn the controllers with different pilots in different environments.

Appendix C.2.2 *Experimental Results*

As Fig. C3 shows the learning process of the proposed ILAC on HUALEX system with different pilots (the hip joint and knee joint of right leg) in different environments. Since the weights of the networks are initialized randomly before the learning begins, the convergence time of the learning process with different pilots is slightly different. The specific convergence time with different pilots is shown in Tab. C1.

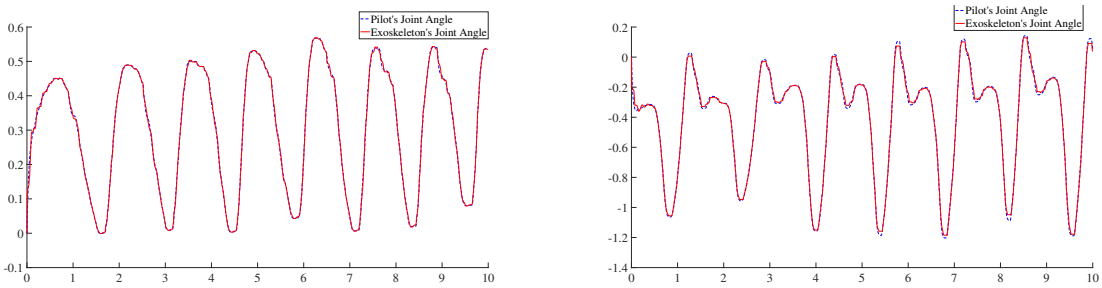


(a) Evolution of the actor's loss function (b) Evolution of the critic's loss function

Figure C3 The learning process of the proposed ILAC with three pilots: (a) evolution of the actor's loss function; (b) evolution of the critic's loss function.

After the online learning of the controller parameters, HUALEX system has a good performance in dealing with different pilot and different walking patterns in different environments. Fig. C4 shows the control performance of ILAC in hip right joint and knee right joint of pilot B. As shown in Fig. C4, The proposed algorithm can make the exoskeleton system follow the pilot's motion with little tracking error, in which the continuous optimal policies are found in continuous high-dimensional domains.

The experiments results on both single DOF exoskeleton and HUALEX system show that the proposed ILAC can make exoskeleton system track the pilot'motion with different pilots in different environments through the online learning process in continuous high-dimensional domains.



(a) Hip joint (b) Knee joint

Figure C4 Control performance of ILAC on HUALEX system with pilot B for 10 seconds walking of the whole experiment: (a) right hip joint; (b) right knee joint.