Inferring Human-Robot Performance Objectives During Locomotion Using Inverse Reinforcement Learning and Inverse Optimal Control

Wentao Liu, Junmin Zhong, Ruofan Wu, Bretta L Fylstra, Jennie Si, *Fellow, IEEE*, and He (Helen) Huang, *Senior Member, IEEE*

***Abstract*—Quantitatively characterizing a locomotion performance objective for a human-robot system is an important consideration in the assistive wearable robot design towards human-robot symbiosis. This problem, however, has only been addressed sparsely in the literature. In this study, we propose a new inverse approach from observed human-robot walking behavior to infer a human-robot collective performance objective represented in a quadratic form. By an innovative design of human experiments and simulation study, respectively, we validated the effectiveness of two solution approaches to solving the inverse problem using inverse reinforcement learning (IRL) and inverse optimal control (IOC). The IRL-based experiments of human walking with robotic transfemoral prosthesis validated the realistic applicability of the proposed inverse approach, while the IOC-based analysis provided important human-robot system properties such as stability and robustness that are difficult to obtain from human experiments. This study introduces a new tool to the field of wearable lower limb robots. It is expected to be expandable to quantify joint human-robot locomotion performance objectives for personalizing wearable robot control in the future.**

***Index Terms*—Learning from demonstration, reinforcement learning, wearable robotics.**

I. Introduction

Wearable lower limb robotics, such as robotic exoskeletons and prostheses, are promising technologies for assisting locomotion in individuals with movement deficits. Researchers often focus on design and control of these robotic devices to restore normative joint kinematics and/or kinetics for improving gait performance in human users [1]. However, human gait biomechanics also play an important role in the robotic joint mechanics when human and robots are coupled in parallel (such as exoskeletons) or in series (for example robotic prosthesis) [2]. As such, we need to gain quantitative understanding of the human-robot system during walking, a problem that has been considered a major challenge in wearable robots. As the first step toward human-robot symbiosis (i.e., human and robotic limb function together for augmented locomotion), we need to consider the collaborative and interactive nature of human and lower limb robots in locomotion as well as mathematically construct a proper control objective to account for factors from both the human and the robot.

Understanding the human-robot collective goal is important to personalize wearable robots for gait assistance. This is because human motor behaviors, especially for populations with sensorimotor deficits, vary greatly between and within individuals due to differences in motor weakness and compensation strategies. Active research attempted to address this issue by personalizing the wearable robot control as an optimization problem [3]–[6], which is to fine tune wearable robot control to achieve a certain optimized performance goal in walking.

Studies have considered various performance measures as the goal to personalize the wearable robot control. Metabolic cost has been reported as a performance objective for optimization in exoskeleton control to improve walking energetic efficiency of non-disabled individuals [3], [4]. The resulting customized and optimized control allowed humans to eventually reduce overall energy expenditure in walking. However, limited success has been reported in using metabolic cost as the objective for optimizing robotic lower limb prosthesis control [7]. Some research groups have attempted to bring user preference into prosthetic device control by allowing users to self-select their preferred control parameters [8], [9]. Another research group proposed learning algorithms [10], [11] to include human feedback, such as good vs bad and preferred action 1 vs action 2, to identify optimal regions of an objective function in the exoskeleton control optimization process. One of the difficulties in incorporating user perception into the control objective is how to reliably quantify and mathematically describe this goal. Additionally, these optimization goals are based on human measurements only, which are chosen by intuition, not chosen systematically. Furthermore, it is not clear how to arbitrate between robotic prosthesis control and human control in a collective control goal

Manuscript received September 9, 2021; accepted December 30, 2021. Date of publication January 18, 2022; date of current version January 28, 2022. This letter was recommended for publication by Associate Editor Dorsa Sadigh and Editor Jens Kober upon evaluation of the reviewers' comments. This work was supported in part by the National Science Foundation under Grants 1563454, 1563921, 1808752, and 1808898. *(Wentao Liu and Junmin Zhong contributed equally to this work.) (Corresponding authors: Jennie Si; He (Helen) Huang.)*

Wentao Liu, Bretta L Fylstra, and He (Helen) Huang are with the UNC/NCSU 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: wliu29@ncsu.edu; bfylstr@ncsu.edu; hhuang11@ncsu.edu).

Junmin Zhong, Ruofan Wu, and Jennie Si are with the School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: jzhong20@asu.edu; ruofanwu@asu.edu; si@asu.edu).

Digital Object Identifier 10.1109/LRA.2022.3143579

during walking, and whether there are other confounding factors influencing the human-robot collective performance.

Given the complexity of the problem, and to make the problem meaningful and tractable, in our prosthesis control problem setting, we formulated the human-robot system objective as meeting desired prosthesis joint kinematics features. We have thus developed and demonstrated reinforcement learning (RL) based tuning of the 12 prosthesis control parameters while a human walked with the robotic prosthesis knee [5], [6]. Our results also showed that both human motor control and computerized prosthesis control have influences on the kinematics [12] or kinetics [13] of prosthesis joints. However, to our knowledge, few studies have focused on inferring a potential human-robot collective performance objective in the lower limb wearable robotics.

This study proposes a new concept and computational framework that uses observed human-robot behavior to infer a human-robot collective performance objective in locomotion. We examined two solution methods, the inverse reinforcement learning (IRL) [14], [15] and the inverse optimal control (IOC) based on which cost objective functions can be solved in closed form [16]. These methods can in principle be applied to capture general forms of collective cost functions. But as the first attempt of its kind, we demonstrate the feasibility of applying IRL and IOC in a human-prosthetic system, by considering the cost function of a quadratic form which contains two features under the influence of both human and machine. IRL and IOC were then used to specify the weighting coefficients for the performance features.

To validate the effectiveness of the proposed IRL and IOC methods, building upon the existing knowledge we learned from individuals walking with a robotic knee prosthesis, we developed two innovative procedures, involving human experiments and control theoretic analyses by simulations. For IRL approach, we conducted experiments on a human when walking with a robotic knee prosthesis. We introduced visual biofeedback as an approach to encourage human movement behavioral change while the prosthesis was controlled by RL. By observing different human-robot system behaviors under the conditions of with and without visual biofeedback, we showed how IRL was able to capture these different behavioral goals represented as different cost functions. For IOC approach, we used OpenSim [17] to simulate the human-prosthesis system in walking and conducted control theoretic analysis. Since it is difficult to simulate walking behavioral change elicited by visual feedback, instead, we introduced gait behavior change by modifying the mechanics limbs in simulation models. Then we were able to observe human-robot system behaviors which resulted in different cost functions. Both IRL and IOC have significant implication for human-robot systems in locomotion. The IRL-based human experiments validated the realistic applicability of the proposed inverse approach, while the closed form IOC-based analysis provided important human-robot system properties that were difficult to obtain from human experiments.

Our main contributions include the following. 1) We introduced a new inverse concept and approaches to identifying human-robot performance objective in locomotion. The conceptualization of the approach is general even though we demonstrated the performance cost function in a quadratic form in this study. 2) We validated the proposed concept by using IRL and IOC through an innovative design of human experiments and control theoretic analysis based on simulations. 3) While both IRL and IOC are effective realizations for quantification of a human-robot system performance in a cost function, our IRL based results demonstrated the potential of our proposed framework in realistic human application scenarios, and our IOC based results revealed new insight that explains human-robot system behavior such as stability, robustness, and control/human energy consumption.

II. Methods

This study uses two solution approaches, IRL and IOC, to infer a collective performance objective based on observed human-robot system behavior. To validate the effectiveness of both approaches, human experiments under the condition of with and without visual feedback were carried out. Due to the difficulty of simulating walking behavioral change elicited by visual feedback, instead, we introduced gait behavior change by modifying the mechanics of both intact and prosthesis limb in simulation models realized by different damping coefficients.

A. The Human-Robot System in Walking

To apply IRL and IOC, and to validate their feasibility of capturing human-robot behavioral performance and representing it in a cost function, we need to collect human-robot behavioral data under different behavioral conditions (with or without visual feedback for human experiments, and different damping coefficients for OpemSim simulations) in order to inversely compute the performance objective. Enabling normative walking requires automatically controlled robotic knee to meet a desired target profile. For this purpose, we utilized the well establish finite state machine impedance control (FSM-IC) framework. As depicted in Figs. 1(a) and (b), in FSM-IC, a single gait cycle is decomposed into four consecutive phases based on knee joint kinematics and ground reaction force: stance flexion (STF), stance extension (STE), swing flexion (SWF) and swing extension (SWE). Thus, four well-designed impedance controllers for each phase are required for continuous walking. The knee joint torque τ can then be formed to control knee joint movement based on the impedance control law,

$$\tau = K\left(\theta - \theta_e\right) - \beta\Omega \tag{1}$$

where in the above, sensors attached to the prosthesis provides real time measurements of knee joint angle θ and knee angular velocity Ω.

B. Automatic Robot Control

To enable human-robot normative walking under different experimental and simulation protocols, we used our previously developed RL-based control agent, the policy iteration with constraint embedded (PICE) [5], to automatically tune the prosthetic knee impedance parameters to generate the automatic control torque according to (1). The goal of robot control was

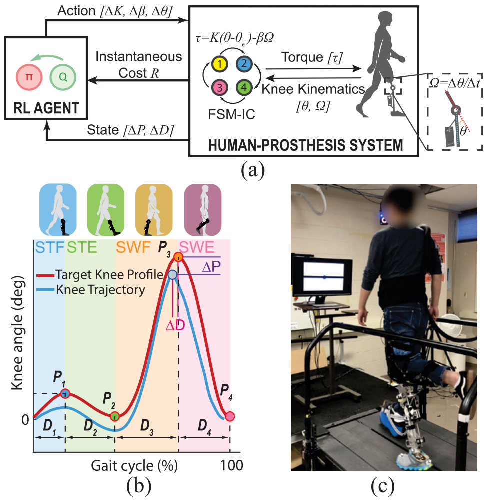

Fig. 1. Experimental procedure and experiment setup in obtaining human-robot behavioral data under two protocols of with and without visual feedback, respectively. (a) The controller was developed for the human-robot to achieve normative walking. Four reinforcement learning based controllers developed within a finite state machine (FSM) framework [5] were tuned corresponding to each of the four phases of the robotic knee. (b) Profile and features of knee kinematics were described for each of the four gait phases. An example of state definition was illustrated in the SWF phase. Two profile features, a peak error ΔP and a duration error ΔD were defined to form system state. (c) Human experiments were carried out with and without visual feedback. For experiments with visual feedback, the computer screen in front of a participant provided real-time feedback on stance duration time. The vertical displacement of the dot relative to the centrally placed horizontal bar provided the participants a measure of actual stance time on the prosthesis side relative to the desired. Human-robot behavioral data were collected during testing using the previously learned control policy. The recorded data sequence then served as inputs to an IRL agent to infer a collective performance objective of the human-robot system. Two different performance representations were expected corresponding to the two different behavioral protocols, which is a validation of IRL method.

to regulate the robotic knee joint to meet a desired knee profile [see Fig. 1(b)] characterized by four characteristic feature points. Two features are identified for each of the target point: the target angle P_d and target duration D_d. We measured the peak knee angle P and timing D in every gait cycle. The value of peak error and duration error are determined by

$$\Delta P = P - P_d$$
$$\Delta D = D - D_d \tag{2}$$

We define the state variables as

$$s = [\Delta P, \Delta D] \in \mathbb{R}^2 \tag{3}$$

Meanwhile, the control variables are the adjustments of the impedance parameters,

$$a \ = \ [\Delta K, \Delta\beta, \Delta\theta] \in \mathbb{R}^3 \tag{4}$$

corresponding to the three impedance parameters in (1).

The goal of this design is to find an optimal policy which minimizes a cumulative error between target profile features and the measured profile features. In this study, PICE control [5] was used to solve this optimal control problem.

C. Inverse Reinforcement Learning Algorithm

Based on the previously defined system state and control variables, IRL problem can be formulated as follows: given a set of state s sequences, a set of actions a, a discount factor $\gamma \leqslant 1$, and a policy π; determine the cost function r that can characterize agent's control policy or behavior.

Under the guidance of a control policy π, the agent picks actions accordingly to result in a sequence of dynamic states $S = \{s_0, s_1, s_2, \cdots\}$ stemming from an initial state $s_0 \sim D$. To evaluate the policy along the trajectory, we project the state variables of each gait cycle to obtain a series of feature vectors of the states $\phi : s \rightarrow [0, 1]^k$, which represent different factors in the cost objective that we would like to trade off in the human-robot system. In this study, we consider a cost function with its features represented by quadratic errors. IRL approach can in principle be applied to other cost functions to include other performance features.

We assume that there is an unknown vector ω^* that can specify the relative weighting between different performance features, and the cost function of the system can be reconstructed by the linear combination of these features as

$$r^*(s) = \omega^* \cdot \phi(s) \tag{5}$$

The goal of IRL is to estimate this weighting vector ω^*.

For this purpose, IRL proceeds from an evaluation of a policy π. The value function for a policy π can be estimated as

$$V^\pi(s_0) = E\left[\sum_{i=0}^{+\infty} \gamma^i r(s_i)\,|\pi\right] = \omega \cdot E\left[\sum_{i=0}^{+\infty} \gamma^i \phi(s_i)\,|\pi\right] \tag{6}$$

where the state sequence starts from initial state s_0. We define the discounted accumulated feature value vector to be

$$\mu(\pi) = E\left[\sum_{i=0}^{+\infty} \gamma^i \phi(s_i)|\pi\right] \tag{7}$$

where $\mu(\pi)$ represents the feature expectation of ΔP and ΔD. Based on the notations above, the value of a policy π can be written as

$$V^\pi(s_0) = \omega \cdot \mu(\pi) \tag{8}$$

In this study, we use quadratic features and thus, the cost function in (5) can be further written as

$$r(s) = s^T H s \tag{9}$$

where $H = [\begin{smallmatrix}\omega_1 & 0\\ 0 & \omega_2\end{smallmatrix}] \in \mathbb{R}^{2\times 2}$ includes unknown weighting parameters that we would infer from human-robot behavioral data. And thus in this study, the value of an observed trajectory can be written as

$$V^\pi(s_0) = \sum_{i=0}^{\infty} \gamma^i s_i^T H s_i \tag{10}$$

The goal of IRL now becomes determining the weighting matrix H which maximizes the difference between the sampled and an optimal trajectory. This optimization problem can be written as a max-margin question

$$H^* = arg\ \max_\omega \min_\pi \ \{V^\pi(s_0) - V^{\pi_*}(s_0)\} \qquad s.t. \ \|H\|_2 \leqslant 1 \tag{11}$$

And thus, based on the feature expectations of sampled trajectory, the cost function is reversely introduced, which indicates the tradeoff among different performance factors represented by the chosen features.

D. Human Experiments for Inverse Reinforcement Learning

1) Experimental Platform Setup: Fig. 1(c) demonstrates the experiment setup. The prosthesis rotary motion is driven by a DC motor (Maxon, Switzerland), and measured by a potentiometer attached to the joint. Details of this device can be found in [6]. Three non-disabled individuals participated in the testing on the robotic knee prosthesis after providing institutionally-approved informed consent reviewed by the Institutional Review Board of UNC Chapel Hill. During each trial they wore an L-shape adapter to walk with the knee prosthesis on an instrumented split-belt treadmill. The ground reaction force were measured by loadcells (Bertec, USA) attached to the treadmill belts. All wearers received training with the powered prosthesis until they felt comfortable and confident enough to walk at a speed of 0.6 m/s without holding the handrail.

We utilized a unilateral, temporal metric (i.e. prosthesis-side stance time) [18] to display as visual feedback to assist wearer with the control of prosthesis-side stance duration. The visual feedback was created via custom code using the Vicon SDK (VICON, U.K.) and MATLAB and displayed on a computer monitor, 1 m in front of the participant on the treadmill. A dot moved up and down along the y-axis, representing the prosthesis-side stance time increasing and decreasing. The feedback displayed to the wearers was calculated from the ground reaction force in real time (with a 20 N threshold), and averaged over four strides to smooth the signal and reduce any large stride-to-stride corrections. The target stance duration time was self selected by wearers to make sure they felt comfortable with walking, and it was displayed as a black bar centered on the screen to maintain participant's perceived accuracy.

2) Experimental Protocol and Data Collection: Wearers were asked to walk at a constant speed of 0.6 m/s under two conditions: 1) without visual feedback (w/o VF), and 2) with visual feedback (w/ VF). Under the second condition, the real time averaged stance duration of the prosthetic foot was provided to the wearer. In this work we focused on the analysis of state sequences in phase 2 of a gait cycle when single support on the prosthetic leg occurs. This is because humans tend to reduce the single support duration on the prosthesis side when walking with a prosthesis [19], [20] as loading and balance in this phase both rely on prosthetic joint mechanics. When visual feedback guided the stance duration on the prosthesis side, the human-machine objective may change in order to maintain a certain level of single support duration on the prosthesis. Our previous experiments also provided evidence that visual feedback of stance duration has significant influence on knee kinematics during single support on the prosthesis side [18].

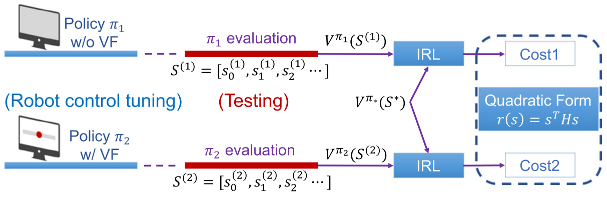

Fig. 2. Experiments to validate feasibility of IRL. Two behavioral protocols (with and without visual feedback) were used to illustrate how IRL was able to respectively capture different behavioral goals represented in a quadratic cost function form. Impedance control tuning was first performed using PICE to obtain a stationary policy when participants achieved normative walking. Testing was performed after participants took a short break. The generated state trajectories were used to obtain a quantitative representation of the cost function using IRL.

As displayed in Fig. 2, PICE control was first applied to tune impedance parameters for each participant until reaching stable normative walking. This is to determine an optimal policy based on which human-robot behavioral data could be generated during testing in order to apply IRL.

The controller and impedance parameters were randomly but identically initialized for all participants. The PICE tuning proceeded by first determining the peak error and the duration error based on knee kinematic measurements. State variable s_i were then obtained and control policy were updated using PICE. After which the impedance adjustment a_i were applied to the FSM-IC [refer to Fig. 1(a)] to enable the next gait cycle until reaching stopping criteria for tuning. We thus obtained two control policies: π_1 was generated without VF and π_2 was generated with VF. In consideration of human variability, measurement noise, and other uncertainties from the environment, we set the tolerance levels of error as ± 1.5 degree for peak errors and $\pm 3\%$ for duration errors. We considered tuning within a phase is converged if state stays within the tolerance range for 8 out of 10 consecutive impedance updates. If all four phases become converged, a trial is successful and meets the stopping criteria.

After taking a break from reaching stationary control policy for both conditions of with and without visual feedback, all three wearers were asked to complete a walking trial for at least 120 gait cycles (corresponding to around 5 minutes of walking time) and until the features of the prosthetic knee joint kinematics were within the allowed tolerance ranges. The behavioral data collected from this testing session were used as inputs to IRL to infer human-robot performance cost function.

3) Data Processing: Once human-robot behavioral data were collected using the optimal policy corresponding to normative walking, We then determined the weighting matrices H for the two policies π_1 and π_2, respectively, using the IRL algorithm in section II-A. We took the target profile as the optimal trajectory, and denote its feature value as μ^{π_*}. Then we could find the weighting matrix H^* which maximizes the difference between the sampled trajectory feature value and the optimal trajectory feature value by using the algorithm proposed in (11). The weights of feature basis vector ω_1 and ω_2 in the cost function were calculated for π_1 and π_2 separately, showing the tradeoff of these factors with its corresponding policy.

E. IOC Approach to Inferring Performance Measure

As the IRL approach, the goal of IOC algorithm is also to infer a performance objective that leads to a realized movement trajectory, which is considered optimal when the human-robot system have reached a target. IOC algorithm is thus applied to determine a cost function, which is of a quadratic form in this study. Even though both IRL and IOC are considered in the same family of inverse optimal control algorithms, the working mechanisms are slightly different, and approximations are made in both algorithms. One such distinguishing factor is that IRL algorithm we used does not require dedicated experiments to perform a system identification procedure while IOC algorithm does in order to obtain an approximate human-robot dynamic system model. As will be shown below, these two methods led to qualitatively agreeable results and IOC approach provided additional insights on the human-robot control system.

As a classical control tool, IOC has been well developed including important results such as necessary and sufficient conditions for a stable state feedback law to be optimal under some quadratic costs [16], revealing innate properties of the human-robot system for its open-loop stability by quantitatively examining pole location, closed-loop control gain, closed-loop robustness via phase margin. It is worth noting that, all these control theoretic properties are difficult to obtain from human experiments. Therefore, once we can validate agreeable results from IRL and IOC, we can expect that results from IOC analysis may benefit studies of human-robotic system properties.

IOC approach also requires the same FSM-IC and automatic robot control as in IRL (section II-A and II-B). To validate the effectiveness of IOC method, two behavioral protocols were realized by using different damping coefficients of the robotic knee as it is infeasible to simulate visual feedback in virtual environment. As in the case of IRL, different cost functions were expected for the two different behavioral protocols. While the IRL method ensured the human-robot system reach the same performance target trajectory via PICE control, the IOC method ensured a same close-loop state trajectory to assigning the same target poles. As such, we were able to infer human-robot performance measure under the same kinematic behaviors and thus, able to reflect human control and its effect on the cost function.

IOC approach proceeded as follows (Fig. 3). First, the same impedance control framework used in IRL was applied to tune the robotic knee using OpenSim simulations. Different stabilizing impedance control parameters were obtained previously under PICE or our other established RL controls. These control parameters were then applied to enable a system identification procedure to produce an open-loop human-robot system dynamic model. Once a system model was identified as SS_1 and SS_2, which correspond to different damping coefficients, a closed-loop pole assignment procedure was applied based on the identified model to ensure both systems had identical poles and thus resulted in the same state trajectories to the same effect of achieving target profile. With the system model and the closed-loop control in place, we were able to use IOC to infer the respective weighting coefficients in the quadratic cost measure.

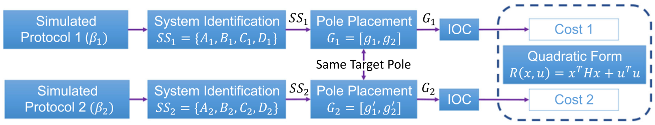
Simulated Protocol 1 (β_1)
System Identification $SS_1 = \{A_1, B_1, C_1, D_1\}$
SS_1
Pole Placement $G_1 = [g_1, g_2]$
G_1
IOC
Cost 1
Same Target Pole
Quadratic Form $R(x,u) = x^T Hx + u^T u$
Simulated Protocol 2 (β_2)
System Identification $SS_2 = \{A_2, B_2, C_2, D_2\}$
SS_2
Pole Placement $G_2 = [g_1', g_2']$
G_2
IOC
Cost 2

Fig. 3. Simulations to validate the feasibility of IOC. Two behavioral protocols (with two different damping coefficients) were used to illustrate how IOC was able to respectively capture different behavioral goals represented in a quadratic cost function form. Control excitations represented by several impedance settings were used to obtain stationary simulated human walking in OpenSim. The resulted dataset including control excitation and state trajectories were then used to identify a human-robot dynamic model. Pole placement was used to produce stationary system trajectories signifying identical system states under the influence of both human and computer controls for both behavioral protocols. This resembles reaching the same normative knee profiles with and without visual feedback in human experiment. Under such a design of simulations study, different behavioral protocols would be expected to result in different performance measures.

F. The Inverse Optimal Control Algorithm

Unlike IRL which is data driven, IOC approach requires identification of linear time invariant (LTI) model of the human-robot dynamic system. While this procedure may introduce some modeling errors, it provides additional insight about human-robot system properties such as stability, control gain, closed-loop robustness and more.

We consider a human-robot system described by the following LTI dynamics,

$$\dot{x} = Ax + Bu$$
$$s = Cx + Du \tag{12}$$

where $x \in \mathbb{R}^{2\times 1}, A \in \mathbb{R}^{2\times 2}, B \in \mathbb{R}^{2\times 1}, C \in \mathbb{R}^{2\times 2}$, and $D \in \mathbb{R}^{2\times 1}$. The system state s is as defined in (3). But unlike discussions of IRL where a stands for control action, here we denote control action by u as it was obtained from a different control mechanism.

We considered the following quadratic instantaneous cost which is common in optimal control theory,

$$R(x,u) = x^T Hx + u^T u \tag{13}$$

where $H = \mathbb{H}^T\mathbb{H}$ with $\mathbb{H} \geq 0$ and $\mathbb{H} = [h_1,\ h_2]$. We also considered an infinite horizon cumulative cost for IOC. Note that, IOC solution is based on a canonical linear quadratic regulator formalism where the energy expenditure in control u is also to be minimized. Since this term is weighed by a constant for different scenarios, it does not affect the characteristics of H. Additionally, we used $\mathbb{H} = [h_1, h_2]$ to signify the weight matrix of performance features as it is obtained using a mechanism different from IRL.

IOC problem is that, given a stable state feedback control law

$$u = -Gx \tag{14}$$

determine a condition on A, B, G such that the control law minimizes the cumulative cost based on the instantaneous cost, and determine the cost represented by $H = \mathbb{H}^T\mathbb{H} \geq 0$, which is weight factor for the peak error feature and the duration effort feature as in IRL approach.

TABLE I
OpenSim Simulation Settings

Prosthetic Stiffness K [Phase 1-4]	[2.97, 0.18, 0.082, 0.045]
Prosthetic Damping β [Phase 1-4]	[0.077, 0.085, 0.0089, 0.0059]
Prosthetic Equilibrium θ_e [Phase 1-4]	[0.39, 0.0533, 1.08, 0.27]
Intact Peak Value [Phase 1-4]	[-0.3, -0.0356, -1.017, -0.04]
Intact Duration Value [Phase 1-4]	[0.11, 0.273, 0.31, 0.2567]

The solution of IOC problem requires that (A, B) are controllable and $(A, \mathbb{H})$ are observable. The control law minimizes the cost if and only if

$$\mathcal{N}^T(-\zeta)\mathbb{H}^T\mathbb{H}\mathcal{N}(\zeta) + \mathcal{D}^T(-\zeta)\mathcal{D}(\zeta) = I \tag{15}$$

where $\mathcal{N}(\zeta)$ and $\mathcal{D}(\zeta)$ are the numerate part and the denominate part of the system transfer function expressed as

$$\begin{aligned} \mathcal{N}(\zeta) &= (\zeta I - A - BG)^{-1}B \\ D(\zeta) &= G(\zeta I - A - BG)^{-1}B + I \end{aligned} \tag{16}$$

We were thus able to solve the cost function, specifically $\mathbb{H}$ after we had identified the system dynamic model and specified the closed-loop poles which led to the same target state trajectories in our design for this simulations study.

G. Linear System Identification and Pole Assignment for IOC

The goal of performing a system identification is to obtain a model ((A, B, C, D) in (12)) of the human-robot system so that we could apply the inverse optimal control procedure to recover the system performance objective function given a control policy. To this end, we obtained single input multiple output (SIMO) system models from K to state ΔP as $TF_{K\rightarrow\Delta P}$, and from K to state ΔD, $TF_{K\rightarrow\Delta D}$.

The stabilizing impedance control parameters were obtained from the same automatic control algorithm used in IRL (in section II-A and II-B). Here for the purpose of system identification, a set of impedance parameters varied as follows to trigger system outputs (peak and duration errors). We kept θ_e at values summarized in Table I. The stiffness K in Table I was corrupted by white noise δK. We simulated the system dynamics with different damping ratio β to signify two different behavior protocols in OpenSim. Whereas in IRL, the two behavioral protocols were implemented by the with and without visual feedback protocols.

The pole placement technique was applied in order to reach the desired stationary closed-loop state trajectories. This step was performed after obtaining an LTI system dynamic model. Then we could obtain the corresponding control gains, which were referred to as control policies in IRL. To represent human influence on the robotic knee, we fixed the computer control gain by placing identical closed-loop poles for the human-robot system under different behavioral protocols (different damping coefficients). In IRL, this was achieved by controlled knee profile meeting the target under different behavioral protocols.

Further control theoretic analysis revealed that the optimal control gain G is related to the quadratic cost $H = \mathbb{H}^T\mathbb{H}$ such that $\mathbb{H} \approx G$.

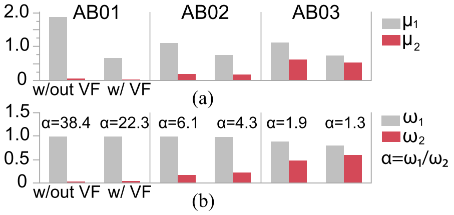

Fig. 4. Experiment results under the conditions of with and without visual feedback, respectively, for each wearer. (a) Feature values of peak error and duration error (refer to (7)). (b) Corresponding weights defined in (9).

H. OpenSim Simulations

OpenSim is a well-established platform which was used in this study to simulate level ground walking of a human-robot system. Specifically, we used it under different impedance parameter settings to generate dynamic trajectories in order to formulate the dataset for system identification introduced above (Fig. 3). The model settings for simulating lower limb to enable stable walking are summarized in Table I.

We created two simulated wearers, each had two simulated behavior protocols realized by different damping coefficients. The two simulated wearers were characteristically different as one had a fixed pelvis profile (simulated wearer #1) for the intact limb and the other had adaptive pelvis control (stimulated wearer #2) that enables walking patterns with large variance.

For simulated wearer #1, the intact knee behavior profile (characterized by peak value and duration value) was prescribed as in Table I, and we used two simulated behavioral protocols by varying β values in phase 2 ($\beta_1 = 0.08$ as behavioral protocol #1, and $\beta_2 = 0.14$ as behavioral protocol #2). For simulated wearer #2, the intact knee was enabled by impedance control specified by impedance parameters that correspond with an intact knee profile established from well-developed normative dataset [21] and the two simulated behavioral protocols are $\beta_1 = 0.11$ and $\beta_2 = 0.107$.

III. Results and Analysis

A. Experimental Results and Analysis

Fig. 4 summarizes the feature expectation μ defined by (7), weights ω_1 and ω_2, respectively, for peak error and duration error in the cost function defined in (9), as well as the ratio between the weights defined by $\alpha = \omega_1/\omega_2$ for the three wearers. We compared the values of these measurements over the two experimental conditions, with and without visual feedback.

With human intentional control under the guidance of VF, the calculated features of peak error and duration error both decreased. Peak error feature μ_1 decreased by $64.5\%, 32.7\%, 34.8\%$, and duration error feature μ_2 decreased by $38.3\%, 41.5\%, 32.8\%$, respectively for each wearer.

The weights for peak error and duration error in the cost function were also quantitatively unraveled by IRL. For all three wearers, peak error term dominated the cost function. However, when visual feedback was provided to users, duration error increased its relative prominence in the cost function. This can be seen from the relationship between the two weights

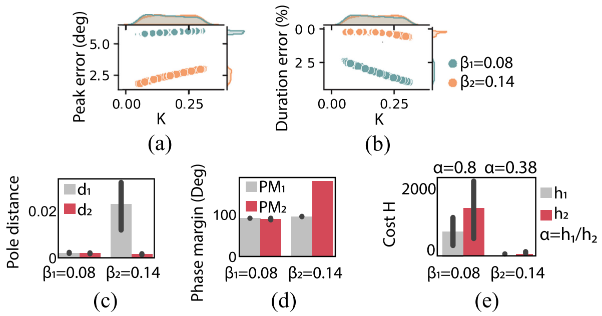

Fig. 5. Simulated walking performance results and control theoretic analyses for simulated wearer #1 under walking protocol #1 ($\beta_1 = 0.08$) and walking protocol #2 ($\beta_2 = 0.14$). All results were obtained based on 10 trials. For each trial, randomly generated input-output data pairs allowed a system identification procedure to produce a human-robot system model defined in (12). The black lines represent the 75% confidence interval. (a) and (b) Simulated walking performances represented by peak error and duration error for the two walking protocols as impedance parameter K (stiffness) varied. The shaded areas outside the top and the right edges of each panel are the histograms of the respective values. Results shown in (a) and (b) illustrate different human-robot system behaviors due to different damping coefficients. (c) Distance of the open-loop poles from the stability margin for the identified human-robot systems. (d) Phase margins which signify closed-loop system robustness or its ability to deal with uncertainties and disturbances. (e) IOC resulted weighting coefficients (h_1 and h_2 for peak error and duration error, respectively) in the quadratic cost defined in (13), and the ratio α between the two weights.

by calculating their relative ratio α with ω_1 weighs the peak error and ω_2 weighs the duration error. Overall, this relative ratio consistently decreased for all three wearers. Specifically, the relative weight ratio between peak error and duration error decreased by $41.9\%, 29.2\%, 27.5\%$, respectively for the three wearers when they were guided by visual feedback.

Individuals walking with and without visual feedback are two characteristically different behaviors. With visual feedback, prosthesis wearers were intentionally involved in the control process. Even though it is unclear how would wearers internalize the visual feedback in order to generate improved gait timing, what we have demonstrated suggests that IRL was capable of capturing different behavioral performance goals and quantitatively represent the goals in a quadratic cost function with appropriate weighting on performance features that we could measure.

B. Simulation Results and Control Theoretic Analysis

In addition to providing validations for IOC as another feasible approach to capturing behavioral performance goals and representing them in a quadratic cost function with appropriate weightings, the results obtained in this section via control theoretical analysis allowed us to gain additional insight that could not be devised from human experiments.

We analyzed performance of two simulated subjects. In each trial, a system model was identified based on 800 measured input-output data pairs as Fig. 3 shows. The respective results were obtained from 10 random trials. The average fit rates of the identified system models for the two simulated wearers are 71% and 73.8% respectively. Fig. 5 summarizes results of simulated wearer #1, the one with fixed pelvis profile.

Figs. 5(a) and (b) are walking performances obtained after system identification procedures where input-output data pairs were generated via OpenSim simulations. The respective averaged peak error and duration error are 5.92° and 3.2% for protocol #1, 2.5° and 0.32% for protocol #2. This illustrates behavioral differences between the two protocols. In addition, based on target profile, the peak error change rate from protocol #1 to protocol #2 is much greater than the duration error which indicates that the peak errors are more sensitive to damping impedance change.

Fig. 5(c) examines open-loop poles of the identified human-robot system models for stability. Greater pole distance from the origin means a more stable system. As shown, the pole distance associated with peak error d_1 is more sensitive to the protocol change from #1 to #2 than that associated with duration error d_2. This indicates that a more stable human-robot system (signified by farther left open-loop poles) is more responsive to impedance control over the peak error than over the duration error, and that walking protocol #2 possesses better stability property than walking protocol #1.

Fig. 5(d) shows phase margins (PM) of closed-loop systems associated with peak error PM_1 and duration error PM_2. Large phase margin is associated with good robustness with respect to phase perturbation and allows the closed-loop system to have good low frequency command following and nice low frequency disturbance attenuation [22]. As shown, both PM_1 and PM_2 have at least $\pm 90^\circ$ phase margin. Additionally, the PM for protocol #2 is greater than protocol #1 which again indicates that system behavior associated with protocol #2 is more stable and more robust than protocol #1.

Fig. 5(e) shows the result of the inferred weighting coefficients (h_1 and h_2), as well as the ratio $\alpha = h_1/h_2$ for the quadratic cost function represented. Protocol #2 is associated with average h_1 and h_2 values that are significantly reduced from those associated with protocol #1 by 97.9% and 94.85%, respectively. Consequently, relative ratio α decreased by 45.7%.

Similar to the analysis of simulated wearer #1, we obtained qualitatively agreeable results for simulated wearer #2, the one with impedance controlled pelvis profile, which had large variances in movement trajectories than the wearer with fixed pelvis. The two walking protocols now are specified by protocol #1 with $\beta_1 = 0.11$ and protocol #2 with $\beta_2 = 0.107$. As expected, wearer #2 had greater variances in prosthetic kinematics than wearer #1 who had a fixed pelvis profile. Under protocol #2, the human-robot performance reflected by the peak error and duration error had absolute mean value at 0.1° and 0.68%, respectively. The human-robot system performed poorer under protocol #1 than protocol #2 where peak error and duration error were at 0.48° and 0.37%, respectively. Examining the weighting coefficients h_1 and h_2 and their ratio α in the quadratic cost, we found similar outcomes as in wearer #1. The average weighting coefficients h_1 and h_2 for protocol #2 significantly decreased from those for protocol #1 by 68.5% and 68.3%, respectively, and the ratio α decreased by 3.17%.

In summary, the control theoretic analysis allowed us to gain new insights and to provide new explanations on the human-robot system properties. Results consistently show that better kinematic performances are associated with more stable human-robot systems (open-loop pole locations), more robust closed-loop performance to deal with uncertainties and reject disturbances from the environment. Finally it is worth noting that the simulations result shows smaller control gains were also associated with better kinematic performances. This in turn implies less control energy expenditure.

IV. Conclusion and Discussion

We presented a computational approach to infer human-robot collective performance goal in a cost function of a quadratic form given human-robot behavioral measurements. We validated this approach by human experiments and by simulated human walking experiments using OpenSim. Our results showed that they are both feasible. Additionally, the human experimental results demonstrated that IRL can be used as a practical approach in realistic human walking conditions, and IOC approach on the other hand, provided additional insight on important stability and robustness properties of the human-robot system. We further hypothesize that the human-robot system can achieve better performance with realistic and accurate objective functions serving as control design goals of wearable lower limb assistive robots.

Experimental and simulation results showed that impedance control can have a greater influence on the kinematic peak error performance than on the gait duration error. As such, the peak error reduced faster than duration error if wearers were given training to behavioral protocols such as using visual feedback to guide walking. Our control theoretic analysis also revealed that a better performing behavioral protocol (such as #2) measured by kinematic errors was related to a more stable open-loop or human-robot system. This may suggest that even though different control settings can lead to same kinematic performance by meeting the trajectory target, some controllers could be more stable than others. Such analysis had not been performed previously in the context of impedance control design for robotic prosthesis. However, the linearization and the system identification procedures in IOC may introduce some error to the analyses. An integrated experimental and theoretical approach may deserve further exploration.

This work entails a quadratic cost structure involving two kinematic errors in the objective. Our experimental and simulation validations of IRL and IOC approaches have shown promise. This method can potentially be extended to more general cost structure to capture additional human-robot performance related factors beyond kinematic errors. However, we do not know all the aspects that a user cares about. And therefore, innovative designs of experiments and feature representation models of the cost function in IRL are needed. Integrated experimentation and computational validation and testing can be expected to help shed light on how a human user and a robotic limb function together for augmented locomotion.

References

[1] M. R. Tucker *et al.*, "Control strategies for active lower extremity prosthetics and orthotics: A review," *J. Neuroengineering Rehabil.*, vol. 12, no. 1, pp. 1–30, 2015.

[2] H. H. Huang, J. Si, A. Brandt, and M. Li, "Taking both sides: Seeking symbiosis between intelligent prostheses and human motor control during locomotion," *Curr. Opin. Biomed. Eng.*, vol. 20, 2021, Art. no. 100314.

[3] Y. Ding, M. Kim, S. Kuindersma, and C. J. Walsh, "Human-in-the-loop optimization of hip assistance with a soft exosuit during walking," *Sci. Robot.*, vol. 3, no. 15, 2018, Art. no. eaar5438.

[4] J. Zhang *et al.*, "Human-in-the-loop optimization of exoskeleton assistance during walking," *Science*, vol. 356, no. 6344, pp. 1280–1284, 2017.

[5] 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. Robot.*, to be published, doi: 10.1109/TRO.2021.3078317.

[6] Y. Wen, J. Si, A. Brandt, X. Gao, and H. H. Huang, "Online reinforcement learning control for the personalization of a robotic knee prosthesis," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2346–2356, Jun. 2019.

[7] C. G. Welker, A. S. Voloshina, V. L. Chiu, and S. H. Collins, "Shortcomings of human-in-the-loop optimization of an ankle-foot prosthesis emulator: A case series," *Roy. Soc. open Sci.*, vol. 8, no. 5, 2021, Art. no. 202020.

[8] T. R. Clites, M. K. Shepherd, K. A. Ingraham, L. Wontorcik, and E. J. Rouse, "Understanding patient preference in prosthetic ankle stiffness," *J. Neuroengineering Rehabil.*, vol. 18, no. 1, pp. 1–16, 2021.

[9] A. Alili, V. Nalam, M. Li, M. Liu, J. Si, and H. H. Huang, "User controlled interface for tuning robotic knee prosthesis," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 6190–6195.

[10] K. Li *et al.*, "ROIAL: Region of interest active learning for characterizing exoskeleton gait preference landscapes," in *Proc. IEEE Int. Conf. Robot. Automat.*. 2021, pp. 3212–3218.

[11] M. Tucker *et al.*, "Preference-based learning for exoskeleton gait optimization," in *Proc. IEEE Int. Conf. Robot. Automat.* 2020, pp. 2351–2357.

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

[13] B. L. Fylstra, I.-C. Lee, S. Huang, A. Brandt, M. D. Lewek, and H. H. Huang, "Human-prosthesis coordination: A preliminary study exploring coordination with a powered ankle-foot prosthesis," *Clin. Biomech.*, vol. 80, 2020, Art. no. 105171.

[14] A. Y. Ng *et al.*, "Algorithms for inverse reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2000, vol. 1, pp. 663–670.

[15] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2004, vol. 1, p. 2.

[16] H. Kong, G. Goodwin, and M. Seron, "A revisit to inverse optimality of linear systems," *Int. J. Control*, vol. 85, no. 10, pp. 1506–1514, 2012.

[17] S. L. Delp *et al.*, "OpenSim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 11, pp. 1940–1950, 2007, doi: 10.1109/TBME.2007.901024.

[18] A. Brandt, W. Riddick, J. Stallrich, M. Lewek, and H. H. Huang, "Effects of extended powered knee prosthesis stance time via visual feedback on gait symmetry of individuals with unilateral amputation: A preliminary study," *J. Neuroengineering Rehabil.*, vol. 16, no. 1, pp. 1–12, 2019.

[19] S. A. Gard, "Use of quantitative gait analysis for the evaluation of prosthetic walking performance," *J. Prosthetics Orthotics*, vol. 18, no. 6, pp. P93–P104, 2006.

[20] M.-Y. Lee, C. F. Lin, and K. S. Soon, "Balance control enhancement using sub-sensory stimulation and visual-auditory biofeedback strategies for amputee subjects," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2007, pp. 2951–2959.

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

[22] L. Yang, J. Si, K. S. Tsakalis, and A. A. Rodriguez, "Performance evaluation of direct heuristic dynamic programming using control-theoretic measures," *J. Intell. Robotic Syst.*, vol. 55, no. 2, pp. 177–201, 2009.