

Using Human Ratings for Feedback Control: A Supervised Learning Approach with Application to Rehabilitation Robotics

Journal Article**Author(s):**

Menner, Marcel; Neuner, Lukas; Lünenburger, Lars; Zeilinger, Melanie N.

Publication date:

2020-06

Permanent link:

<https://doi.org/10.3929/ethz-b-000424388>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

IEEE Transactions on Robotics 36(3), <https://doi.org/10.1109/TRO.2020.2964147>

Funding acknowledgement:

157601 - Safety and Performance for Human in the Loop Control (SNF)

Using Human Ratings for Feedback Control: A Supervised Learning Approach with Application to Rehabilitation Robotics

Marcel Menner, Lukas Neuner, Lars Lünenburger, and Melanie N. Zeilinger

This work has been accepted for publication in the IEEE Transactions on Robotics.
Digital Object Identifier 10.1109/TRO.2020.2964147.

Please cite this work as

M. Menner, L. Neuner, L. Lünenburger, and M. N. Zeilinger, "Using Human Ratings for Feedback Control: A Supervised Learning Approach with Application to Rehabilitation Robotics," *IEEE Transactions on Robotics*, 2020, doi:10.1109/TRO.2020.2964147.

Using Human Ratings for Feedback Control: A Supervised Learning Approach with Application to Rehabilitation Robotics

Marcel Menner, Lukas Neuner, Lars Lünenburger, and Melanie N. Zeilinger

Abstract—This article presents a method for tailoring a parametric controller based on human ratings. The method leverages supervised learning concepts in order to train a reward model from data. It is applied to a gait rehabilitation robot with the goal of teaching the robot how to walk patients physiologically. In this context, the reward model judges the physiology of the gait cycle (instead of therapists) using sensor measurements provided by the robot and the automatic feedback controller chooses the input settings of the robot to maximize the reward. The key advantage of the proposed method is that only a few input adaptations are necessary to achieve a physiological gait cycle. Experiments with nondisabled subjects show that the proposed method permits the incorporation of human expertise into a control law and to automatically walk patients physiologically.

Index Terms—Human feedback-based control, human-centered robotics, learning and adaptive systems, rehabilitation robotics.

I. INTRODUCTION

HUMANS can perform very complex tasks that are difficult to achieve with autonomous systems. The dependency on human supervision or expertise still restricts efficient operation of many complex systems. An important domain where human expertise is usually needed is rehabilitation robotics, where we consider the robot-assisted gait trainer Lokomat[®] [1] in this paper, see Fig. 1. Robotic systems like the Lokomat have recently been introduced in gait rehabilitation following neurological injuries with the goal of mitigating the limitations of conventional therapy [2]–[6]. However, training with such robots still requires the supervision and interaction of experienced therapists [1].

Gait rehabilitation with the Lokomat currently requires physiotherapists to manually adjust the mechanical setup and input settings, e.g., the speed of the treadmill or the range of motion, in order to bring patients into a physiological and safe gait cycle. Therapists have to be trained specifically for the device and acquire substantial experience in order to achieve good input settings. Although there are guidelines for their adjustment [7], it remains a heuristic process, which strongly depends on the knowledge and experience of the therapist. Automatic adaptation of input settings can reduce the duration



Fig. 1. Lokomat[®] gait rehabilitation robot (Hocoma AG, Volketswil, CH).

of therapists' schooling, improve patient training, make the technology more broadly applicable, and can be more cost effective. In this work, we propose a method to automatically adapt the input settings. Although the motivation behind this work is in the domain of rehabilitation robotics, the proposed method addresses general human-in-the-loop scenarios, where expert knowledge can improve system operation.

In this paper, we propose a two-step approach to achieve automatic input adaptations: First, we define a feature vector to characterize the gait cycle and postulate a reward model to judge the physiology of the gait cycle using the feature vector. The reward model is trained with therapists' ratings using a supervised learning technique, where the feature vector is obtained from sensor measurements provided by the robot. The sensor measurements are the angle, torque, and power of both the hip and knee joints of the robot. Second, we use the gradient of the reward model to determine input adaptations that achieve the desired gait cycle. This involves a steady-state model to relate the gradient of the reward model with respect to the feature vector (high dimensional) to input settings (low dimensional) that adjust the gait cycle. A key component in the proposed formulation is that the reward model and its gradient are formulated as functions of the feature vector rather than the input settings. The high dimensionality of the feature vector allows us to use one model for all human subjects with very different body types, which enables very efficient online application of the proposed method. In order to train both the reward model and the steady-state model, we collected

This work was supported by the Swiss National Science Foundation under Grant PP00P2_157601/1.

M. Menner, L. Neuner, and M.N. Zeilinger are with the Institute for Dynamic Systems and Control, ETH Zurich, 8092 Zurich, Switzerland (e-mail: mmenner@ethz.ch; lneuner@ethz.ch; mzeilinger@ethz.ch).

L. Lünenburger is with Hocoma AG, 8604 Volketswil, Switzerland (e-mail: lars.luenenburger@hocoma.com).

Digital Object Identifier 10.1109/TRO.2020.2964147

data with various physiological and non-physiological input settings from 16 nondisabled subjects. The subjects were instructed to be passive while being walked by the robot in order to imitate patients with limited or no ability to walk in the early stages of recovery. Experiments with ten nondisabled subjects highlighted the ability of the proposed method to improve the walking pattern within few adaptations starting from multiple initially non-physiological gait cycles.

Related Work

Adaptive control strategies have been the subject of a body of research in robotic gait trainers with the goal of improving the therapeutic outcome of treadmill training [8]–[13]. The work in [8] presents multiple strategies for automatic gait cycle adaptation in robot-aided gait rehabilitation based on minimizing the interaction torques between device and patient. Biomechanical recordings providing feedback about a patient’s activity level are introduced in [9], [10]. Automated synchronization between treadmill and orthosis based on iterative learning is introduced in [11]. In [12], a path control method is proposed to allow voluntary movements along a physiological path defined by a virtual tunnel. An algorithm to adjust the mechanical impedance of an orthosis joint based on the level of support required by a patient is proposed in [13]. Further research in the domain of rehabilitation robotics is presented, e.g., in [14], [15]. In [14], the human motor system is modeled and analyzed as approximating an optimization problem trading off effort and kinematic error. In [15], a patient’s psychological state is estimated to judge their mental engagement. Different from the work in rehabilitation robotics presented in [8]–[15], we present a method for input setting adaptation of a rehabilitation robot based on a feedback controller, which is derived from human ratings.

In the following, we discuss research directions related to the techniques employed in the proposed approach.

Gait cycle classification is often used to distinguish human subjects according to two classes [16]–[19], e.g., young/elderly or healthy/impaired. In [16], a supervised learning method for automatic recognition of movement patterns is presented to discriminate gait patterns of young and elderly people. In order to improve classification performance, [17] employs a kernel-based principle component analysis for the extraction of features. Gait patterns are also used to diagnose diseases that symptomatically cause gait abnormalities, e.g., [18], [19]. Different from [16]–[19], this paper does not aim to identify or classify human individuals but to generalize from data of multiple individuals by classifying gait patterns according to their physiology. Further, the obtained classifier is not used to predict discrete/binary classes but as a continuous reward, which is maximized using feedback control.

Reinforcement learning uses a trial and error search to find a control policy [20]–[28]. The framework proposed in [21] allows human trainers to shape a policy using approval or disapproval. In [22], human-generated rewards in a reinforcement learning framework are employed for a 2-joint velocity control task. In [23], a policy shaping method is presented where human feedback is not used as a reward signal but

directly as a label for the policy. In [24], human preferences are learned through ratings based on a pairwise comparison of trajectories with the goal of reducing human feedback. In [25], a robot motion planning problem is considered, where users provide a ranking of paths that enable the evaluation of the importance of different constraints. In [26], a method is presented that actively synthesizes queries to a user to update a distribution over reward parameters. In [27], user preferences in a traffic scenario are learned based on human guidance by means of feature queries. In [28], human ratings are used to learn a probability distribution of individual preferences modeled as a Markov decision process. While a reinforcement learning approach could in principle be applied to the considered problem, the online application of these methods typically requires a few hundred human ratings to learn a policy. This is infeasible when working with a patient, where a comparatively small number of feedback rounds has to be sufficient. The main difference between our method and reinforcement learning is that we do not use trial and error search but we build a reward model that is maximized online.

Inverse optimal control and *inverse reinforcement learning* aim at learning a reward model or cost model from demonstrations of human behavior [29]–[46]. Inverse optimal control methods model demonstrations to be the result of an optimal control problem [33]–[40] and often aim at transferring human expertise to an autonomous system, e.g., for humanoid locomotion [33], [34], identifying human movements [35]–[37], robot manipulation tasks [38], or autonomous driving [39]. In inverse reinforcement learning [41]–[46], demonstrations are typically modeled to be the result of probabilistic decision-making in a Markov decision process. The fundamental difference between the proposed method and inverse optimal control/inverse reinforcement learning methods is the utilization of ratings instead of demonstrations to learn a reward model.

II. HARDWARE DESCRIPTION & PROBLEM DEFINITION

The Lokomat[®] gait rehabilitation robot (Hocoma AG, Volketswil, CH) is a bilaterally driven gait orthosis that is attached to the patient’s legs by Velcro straps. In conjunction with a bodyweight support system, it provides controlled flexion and extension movements of the hip and knee joints in the sagittal plane. Leg motions are repeated based on predefined but adjustable reference trajectories. Additional passive foot lifters ensure ankle dorsiflexion during swing. The bodyweight support system partially relieves patients from their bodyweight via an attached harness. A user interface enables gait cycle adjustments by therapists via a number of input settings [1], [10].

Input Settings: One important task of the therapist operating the Lokomat is the adjustment of the input settings to obtain a desirable gait trajectory. A total of 13 input settings can be adjusted to affect the walking behavior, which are introduced in Table I. In this work, we propose a method that can automate or assist the therapists in the adjustment of the input settings by measuring the gait cycle.

TABLE I
INPUT SETTINGS OF THE LOKOMAT

Input Setting & Description	Step-size	Range
Hip Range of Motion (Left & Right) Defines the amount of flexion and extension	3°	23°, 59°
Hip Offset (Left & Right) Shifts movements towards extension or flexion	1°	-5°, 10°
Knee Range of Motion (Left & Right) Defines amount of flexion	3°	32°, 77°
Knee Offset (Left & Right) Shifts movement into flexion for hyperextension correction	1°	0°, 8°
Speed Sets the treadmill speed	0.1km/h	0.5km/h, 3km/h
Orthosis speed Defines the orthosis and affects walking cadence	0.01	0.15, 0.8
Bodyweight Support Defines carried weight for unloading	continuous	0kg, 85kg
Guidance Force Sets amount of assistance	5%	0%, 100%
Pelvic Defines lateral movement	1cm	0cm, 4cm

A. State-of-the-Art Therapy Session

The current practice of gait rehabilitation with the Lokomat includes the preparation and setup of the patient and device, actual gait training, and finally removing the patient from the system [7]. Gait training is further divided into three phases:

1. Safe walk: The patient is gradually lowered until the dynamic range for the bodyweight support is reached. The purpose of this first phase is to ensure a safe and non-harmful gait cycle.
2. Physiological walk: After ensuring safe movements, the gait cycle is adjusted so that the patient is walked physiologically by the robot.
3. Goal-oriented walk: The gait cycle is adjusted to achieve therapy goals for individual sessions while ensuring that the patient's gait remains physiological.

In this paper, we focus on the physiological walk. In a state-of-the-art therapy session, therapists are advised to follow published heuristic guidelines on how to adjust the input settings based on observations (visual feedback) in order to reach a physiological walk. Three examples of the heuristic guidelines are as follows: If the step length does not match walking speed, then the hip range of motion or treadmill speed should be adjusted; if the heel strike is too late, then the hip offset or the hip range of motion should be decreased; if the foot is slipping, then the orthosis speed or the knee range of motion should be decreased. An extended overview of heuristics can be found in [7]. This heuristic approach requires experience and training with experts, which incurs high costs and limits the availability of the rehabilitation robot due to the small number of experienced experts. The proposed method aims to alleviate this limitation as described in the following.

B. Technological Contribution

We propose a method for automatically suggesting suitable input settings for the Lokomat based on available sensor measurements in order to walk patients physiologically. The proposed framework can be used for recommending input

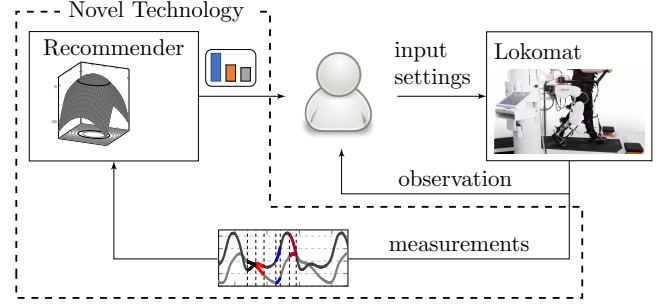


Fig. 2. Overview of the proposed method as a recommendation system. The novel technology (dashed lines) augments the state-of-the-art control loop of a therapist and the Lokomat. Sensor measurements of angle, torque, and power of both hip and knee joints provided by the Lokomat are used to compute recommendations for the input adaptations.

settings for therapists, automatic adaptation of input settings, or as an assistive method for therapists during schooling with the Lokomat. Fig. 2 illustrates the proposed method as a recommendation system. The method is derived assuming that the mechanical setup of the Lokomat is done properly, such that the purpose of adapting the input settings is the improvement of the gait cycle and not corrections due to an incorrect setup.

III. CONTROLLER DESIGN BASED ON HUMAN RATINGS

This section describes the proposed human feedback-based controller. In the setup considered, input settings $s \in \mathbb{R}^m$ of the controlled system lead to a gait cycle represented by a feature vector $x \in \mathbb{R}^n$ in steady-state:

$$x = f(s), \quad (1)$$

where f is an unknown function. For the considered application, the input settings s are given in Table I and the feature vector x is composed of statistical features of measurements, which characterize the gait cycle and are further discussed in Section IV. Here, the notion of a steady-state means that any transients due to an input adaptation have faded. The control objective is to find input settings s^* for which $x^* = f(s^*)$ represents a desired system state, i.e., a physiological gait cycle in the considered application.

Control Law and Conceptual Idea: The method is based on a reward model, reflecting the control objective, and a steady-state model, associating a feature vector with an input setting. The reward model is a function that assigns a scalar value to the feature vector estimating an expert rating of the "goodness" of the feature vector. The reward thereby provides a direction of improvement for the feature vector, which is mapped to a change in input settings via the steady-state model.

We define the control law in terms of input adaptations Δs :

$$\Delta s = f^{-1}(\alpha \Delta x + x) - s,$$

where Δx is the direction of improvement, $f^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the steady-state model (the inverse mapping of f in (1)), and $\alpha > 0$ is the gain of the control law. We compute Δx as the gradient of the reward model $r(x) \in \mathbb{R}$, i.e.,

$$\Delta x = \nabla_x r(x).$$

Fig. 3 shows an example of a reward model and indicates how its gradient is used for feedback control using the steady-state model. Both models $r(\mathbf{x})$ and $\mathbf{f}^{-1}(\cdot)$ are inferred from data. In order to train the reward model, we utilize ratings on an integer scale as samples of the reward model, i.e., $r_i = 1, \dots, S$, where $r_i = 1$ is the worst and $r_i = S$ is the best rating. Additionally, we train a steady-state model $\mathbf{f}^{-1}(\cdot)$ to relate the direction of improvement suggested by the reward model to the corresponding input adaptation (bottom part of Fig. 3). In order to build both the reward model and the steady-state model, N training samples are collected. Each training sample with index i consists of a feature vector \mathbf{x}_i , the input settings \mathbf{s}_i , and the corresponding rating $r_i \in \{1, \dots, S\}$:

$$\{\mathbf{x}_i, \mathbf{s}_i, r_i\}_{i=1}^N. \quad (2)$$

Note that throughout this paper, the feature vector \mathbf{x} is normalized using collected data \mathbf{x}_i such that the collected data are zero-mean with unit-variance in order to account for different value ranges and units, cf., [47].

Outline: The reward model is trained with the feature vector \mathbf{x}_i and its corresponding rating r_i in (2) using a supervised learning technique (Section III-A). The resulting reward model is then used to compute the gradient $\nabla_{\mathbf{x}} r(\mathbf{x})$ as direction of improvement. Finally, a steady-state model relates this direction of improvement with necessary changes in input settings \mathbf{s} . The steady-state model is computed using a regression technique (Section III-B).

A. Reward Model using Supervised Learning

The first step of the framework is the learning of a reward model reflecting the physiology of the gait based on supervised learning techniques [47]. The reward model is a continuous function, i.e., it provides a reward for all \mathbf{x} , whereas observations \mathbf{x}_i are potentially sparse.

In view of the considered application, we postulate a reward model of the form:

$$r(\mathbf{x}) = 0.5\mathbf{x}^T \mathbf{W} \mathbf{x} + \mathbf{w}^T \mathbf{x} + b, \quad (3)$$

where $\mathbf{W} = \mathbf{W}^T \prec 0$, $\mathbf{w} \in \mathbb{R}^n$, and $b \in \mathbb{R}$ are the parameters to be learned from expert ratings given in the form of integers on a scale from 1 to S . The rationale for selecting a quadratic model with negative definite \mathbf{W} is the observation that the gait degrades in all relative directions when changing input settings from a physiological gait. Important properties of this reward model are that a vanishing gradient indicates that global optimality has been reached and its computational simplicity. This motivates the gradient ascent method for optimizing performance.

In order to learn \mathbf{W} , \mathbf{w} , and b in (3), we construct $S - 1$ classification problems. These $S - 1$ classification problems share the parameters \mathbf{W} , \mathbf{w} , and b of the reward model and the corresponding classification boundaries are given by

$$r^l(\mathbf{x}) = 0.5\mathbf{x}^T \mathbf{W} \mathbf{x} + \mathbf{w}^T \mathbf{x} + b^l$$

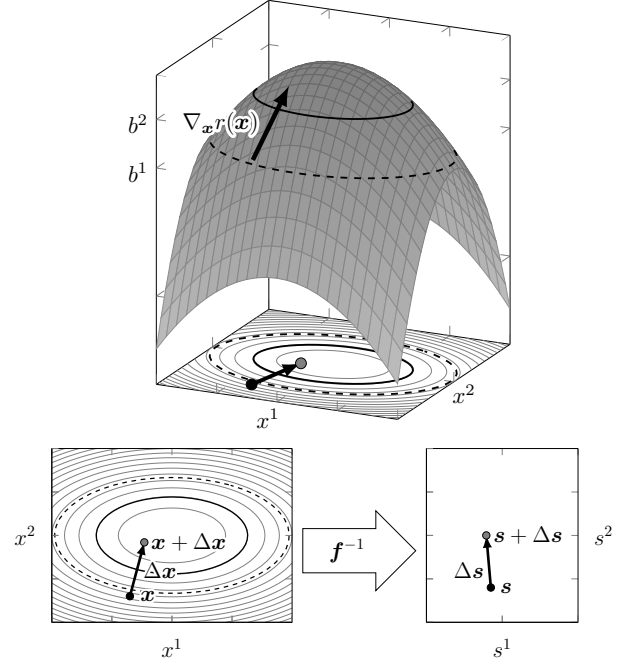


Fig. 3. Top: Example of reward model with gradient vector $\nabla_{\mathbf{x}} r(\mathbf{x})$ where $\mathbf{x} = [x^1 \ x^2]^T$ and projected level sets onto the $x^1 - x^2$ plane. The example shows a case of three ratings $r_i = 1, 2, 3$ separated by two classification boundaries indicated as solid black and dashed black ellipses. Bottom: Steady-state model to compute $\Delta \mathbf{s}$ from $\Delta \mathbf{x}$ where $\mathbf{s} = [s^1 \ s^2]^T$.

for all $l = 1, \dots, S - 1$ with $b^l = b - l - 0.5$ separating the S different ratings such that $r^l(\mathbf{x}_i) > 0$ if $r_i > l + 0.5$. Further, for each data sample i and each l , we define

$$y_i^l = \begin{cases} 1 & \text{if } r_i > l + 0.5 \\ -1 & \text{else.} \end{cases}$$

Hence, an ideal reward model with perfect data and separation satisfies

$$y_i^l r^l(\mathbf{x}_i) \geq 0 \quad \forall i = 1, \dots, N \quad \forall l = 1, \dots, S - 1. \quad (4)$$

In order to allow for noisy data and imperfect human feedback, (4) is relaxed to find $r^l(\mathbf{x})$ that satisfies (4) "as closely as possible" by introducing a margin $\xi_i^l \geq 0$. This approach is closely related to a Support Vector Machine, cf., [47], with a polynomial kernel function of degree two. The functions $r^l(\mathbf{x})$ correspond to $S - 1$ classification boundaries in a multi-category classification framework. The parameters \mathbf{W} , \mathbf{w} , and b of the reward model (3) are computed by solving the following optimization problem using L1 regularization:

$$\underset{\mathbf{W}, \mathbf{w}, b^l, \xi_i^l}{\text{minimize}} \quad \sum_{l=1}^{S-1} \sum_{i=1}^N \xi_i^l + \lambda_1 \cdot (\|\mathbf{W}\|_1 + \|\mathbf{w}\|_1) \quad (5a)$$

$$\text{subject to} \quad y_i^l r^l(\mathbf{x}_i) \geq 1 - \xi_i^l, \quad \forall i = 1, \dots, N \quad (5b)$$

$$\xi_i^l \geq 0, \quad \forall l = 1, \dots, S - 1 \quad (5c)$$

$$r^l(\mathbf{x}_i) = 0.5\mathbf{x}_i^T \mathbf{W} \mathbf{x}_i + \mathbf{w}^T \mathbf{x}_i + b^l \quad (5d)$$

$$b^l = b - l - 0.5 \quad (5e)$$

$$\mathbf{W} = \mathbf{W}^T \prec 0 \quad (5f)$$

where $\lambda_1 > 0$ controls the trade-off between minimizing the training error and model complexity captured by the norm $\|\mathbf{W}\|_1 = \sum_{j=1}^n \sum_{k=1}^n |W_{jk}|$ (elementwise 1-norm) and $\|\mathbf{w}\|_1$, which is generally applied to avoid overfitting of a model and is sometimes also called lasso regression [47].

B. Feedback Control using Reward Model

The second step of the proposed framework is to exploit the trained reward model for feedback control. The idea is (i) to use the gradient of the reward model as the direction of improvement and (ii) to relate this gradient to a desired change in inputs with a steady-state model.

(i) *Gradient of reward model:* The gradient of the inferred reward model is the direction of best improvement. The control strategy is to follow this gradient in order to maximize reward. The gradient of the quadratic reward model in (3) is

$$\Delta \mathbf{x} = \nabla_{\mathbf{x}} r(\mathbf{x}) = \mathbf{W}\mathbf{x} + \mathbf{w}.$$

(ii) *Mapping of gradient to setting space with steady-state model:* In order to advance the system along the gradient direction, we relate the direction of improvement $\Delta \mathbf{x}$ to a change in input settings with a steady-state model \mathbf{f}^{-1} . We use a first order approximation $\mathbf{s} = \mathbf{f}^{-1}(\mathbf{x}) \approx \mathbf{M}\mathbf{x} + \mathbf{m}$ with $\mathbf{M} \in \mathbb{R}^{m \times n}$, $\mathbf{m} \in \mathbb{R}^m$ to compute the change in input settings $\Delta \mathbf{s}$ as

$$\Delta \mathbf{s} = \mathbf{M}(\alpha(\mathbf{W}\mathbf{x} + \mathbf{w}) + \mathbf{x}) + \mathbf{m} - \mathbf{s}, \quad (6)$$

where α can be interpreted as feedback gain or the learning rate in a gradient ascent method. The steady-state model is estimated as the least squares solution of the data in (2):

$$\underset{\mathbf{M}, \mathbf{m}}{\text{minimize}} \sum_{i=1}^N \|\mathbf{s}_i - \mathbf{M}\mathbf{x}_i - \mathbf{m}\|_2^2 + \lambda_2 \cdot (\|\mathbf{M}\|_1 + \|\mathbf{m}\|_1) \quad (7)$$

where, again, we use $\lambda_2 > 0$ to control the trade-off between model fit and model complexity.

Using the quadratic reward model in (3) and the linear steady-state model in (7), the application of the proposed control strategy (6) requires only matrix-vector multiplications, which is computationally inexpensive and can be performed online, cf., Algorithm 1 for an overview of the method. Additionally, as will be shown empirically, the application requires only few online input adaptations.

Algorithm 1 Training and Application of the Method

Training ▷ rating needed

- 1: Collect data set in (2).
- 2: Compute reward model $\mathbf{W}, \mathbf{w}, b$ with (5) and steady-state model \mathbf{M}, \mathbf{m} with (7).

Online Algorithm ▷ no rating needed

- 3: **do**
 - 4: Obtain feature vector \mathbf{x} from measurement.
 - 5: Apply adaptation $\Delta \mathbf{s} = \mathbf{M}(\alpha(\mathbf{W}\mathbf{x} + \mathbf{w}) + \mathbf{x}) + \mathbf{m} - \mathbf{s}$.
 - 6: Wait until steady state is reached.
 - 7: **while** stopping criterion not fulfilled ▷ cf. Section IV-D
-

Remark 1: As we will show in the analysis in Section V, the linear mapping $\mathbf{s} \approx \mathbf{M}\mathbf{x} + \mathbf{m}$ yields sufficient accuracy for the considered application. For more complex systems, one might consider a different steady-state model, e.g., higher order polynomials or a neural network to approximate \mathbf{f}^{-1} .

Remark 2: In principle, reinforcement learning could be applied to directly learn physiological settings. The proposed two-step and model-based method, in contrast, makes use of the higher dimensionality of the feature vector to characterize the gait cycle. Its key advantage is that less samples are required online and thus, less steps to find physiological settings, which is essential for the considered application.

Remark 3: The proposed method iteratively approaches the optimal settings \mathbf{s}^* with the gradient ascent method. This is important for the considered application to cautiously adapt the input settings of the robot with a human in the loop.

Remark 4: It is also possible to determine the direction of improvement using second-order derivatives of the reward model, e.g., using a Newton-Raphson method. However, as numerical second-order derivatives would be more noisy, we choose first-order derivatives, which are simple and yield a more stable estimate of the best (local) improvement.

IV. ADAPTATION OF GAIT REHABILITATION ROBOT TO WALK PATIENTS PHYSIOLOGICALLY

In this section, we show how to apply the method presented in Section III to automatically adapt, or recommend a suitable adaptation, of the Lokomat's input settings in order to walk patients physiologically. A core element is the reward model that has been built on therapists' ratings and is used to judge the physiology of the gait. For simplicity, we adjust settings for the left and right leg symmetrically. This does not pose a problem for the presented study with nondisabled subjects but might be revisited for impaired subjects in future work.

In this work, we focus on physiological walk and exclude the guidance force and the pelvic input settings as they are mainly used for goal-oriented walk [7]. This exclusion is valid for physiological walk where the guidance force and pelvic settings are kept constant at 100% and 0cm, respectively. Hence, there are seven input settings that are considered in the application of the method.

Safety: The proposed method is implemented to augment a previously developed safety controller that ensures safe operation of the Lokomat. This safety controller intervenes if the input settings exceed nominal ranges for forces and positions of the robot's joints. An additional contingency controller stops the robot when the deviation of the measured gait trajectory and the desired gait trajectory becomes too large. In this way, the overall behavior is guaranteed to have the necessary safety requirements for patient and robot, yet among the safe input settings, the ones that improve the gait cycle are chosen. The reader is referred to [48] for a more detailed description of the Lokomat's safety mechanisms.

A. Gait Cycle

The walking of a human is a repetitive sequence of lower limb motions to achieve forward progression. The gait cycle

describes such a sequence for one limb and commonly defines the interval between two consecutive events that describe the heel strike (initial ground contact) [49]. The gait cycle is commonly divided into two main phases, the stance and the swing phase. The stance phase refers to the period of ground contact, while the swing phase describes limb advancement. Fig. 4 illustrates the subdivision of these two main phases of the gait cycle into multiple sub-phases, beginning and ending with the heel strike. This results in a common description of gait using a series of discrete events and corresponding gait phases [49]. We focus on four particular phases of the gait cycle, which are emphasized in Fig. 4:

- Heel strike (HS): The moment of initial contact of the heel with the ground.
- Mid-stance (MS): The phase in which the grounded leg supports the full body weight.
- Toe off (TO): The phase in which the toe lifts off the ground.
- Mid-swing (SW): The phase in which the raised leg passes the grounded leg.

B. Evaluation of Gait Cycle and Data Collection

We derive the reward model based on the four phases. For evaluating the four gait phases, we introduce a scoring criterion in consultation with experienced therapists:

Rating 1: Safe, but not physiological.

Rating 2: Safe, not entirely physiological gait cycle.

Rating 3: Safe and physiological gait cycle.

Data Collection: A total of 16 nondisabled subjects participated in the data collection. The 16 subjects were between 158cm - 193cm (5'2" - 6'4") in height, 52kg - 93kg (115lbs - 205lbs) in weight, and aged 25 - 62. Informed consent for the use of the data has been received from all human subjects. The data collection for each subject involved an evaluation of the four gait phases by therapists for several input settings in order to collect data in a wide range of gait cycles. The nondisabled subjects were instructed to be passive throughout the data collection, i.e., they were walked by the robot. This allowed us to collect data for both physiological and non-physiological gait cycles. Measurements of the Lokomat were recorded for all evaluations. For each subject, the experienced therapists first manually tuned the input settings to achieve rating 3 for all four phases (Set 0 in Table II), where the resulting input settings are referred to as initial physiological gait (IPG). Table II shows the input settings used in the data collection as deviations from the initial physiological gait. Each subject walked for approximately 60 seconds for each set of input settings, while the therapist provided evaluations of the walking pattern. The assessment started after a transient interval of approximately 15 seconds to ensure that the walking has reached a steady state. Note that the input settings resulting in a physiological gait pattern varied between subjects.

The scoring criterion and the consideration of the four phases, as well as the data collection protocol were introduced in consultation with clinical experts from Hocoma (U. Costa and P. A. Gonçalves Rodrigues, personal communication, Nov. 05, 2017). As a result, we obtained the chosen input settings,

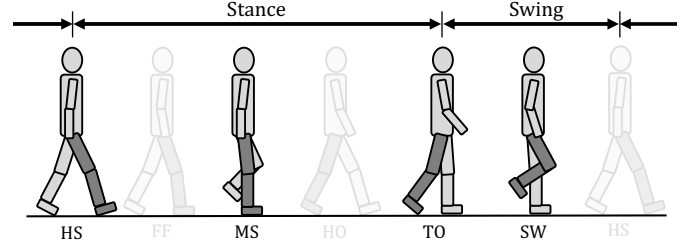


Fig. 4. Gait phases in order: Heel strike, foot flat (FF), mid-stance, heel off (HO), toe off, mid-swing. Both FF and HO phase are not rated in this work, but presented for consistency with the literature [49].

TABLE II
INPUT SETTING FOR DATA COLLECTION

Set	Input Settings	Value
0	Initial Set	IPG
1	Hip Range of Motion	IPG + 6°
2	Hip Range of Motion	IPG + 12°
3	Hip Range of Motion	IPG - 6°
4	Hip Range of Motion	IPG - 12°
5	Hip Offset	IPG + 4°
6	Hip Offset	IPG + 8°
7	Hip Offset	IPG - 5°
8	Hip Range of Motion, Hip Offset	IPG + 12° / IPG - 3°
9	Hip Range of Motion, Hip Offset	IPG + 12° / IPG + 3°
10	Hip Range of Motion, Hip Offset	IPG - 12° / IPG - 3°
11	Hip Range of Motion, Hip Offset	IPG - 12° / IPG + 3°
12	Knee Range of Motion	IPG + 6°
13	Knee Range of Motion	IPG + 12°
14	Knee Range of Motion	IPG - 9°
15	Knee Range of Motion	IPG - 15°
16	Knee Offset	IPG + 4°
17	Knee Offset	IPG + 8°
18	Knee Range of Motion / Knee Offset	IPG + 15° / IPG + 6°
19	Knee Range of Motion / Knee Offset	IPG + 21° / IPG + 6°
20	Speed	IPG + 0.5km/h
21	Speed	IPG + 1.0km/h
22	Speed	IPG - 0.5km/h
23	Speed	IPG - 1.0km/h
24	Orthosis Speed	IPG + 0.03
25	Orthosis Speed	IPG + 0.05
26	Orthosis Speed	IPG - 0.03
27	Orthosis Speed	IPG - 0.05
28	Bodyweight Support	IPG + 15%
29	Bodyweight Support	IPG + 30%
30	Bodyweight Support	IPG - 15%
31	Bodyweight Support	IPG - 30%

the corresponding ratings on an integer scale from 1 to 3, and the recording of measurements of the Lokomat. Next, we discuss the computation of the feature vector from the recorded measurements.

Feature Vector: We use the gait index signal of the Lokomat as an indicator to identify progression through the gait cycle. The gait index is a sawtooth signal and is displayed in the bottom plot in Fig. 5. It is used to determine the time-windows of the four phases, cf., the dashed lines in Fig. 5. The time-windows are used to compute the feature vector, composed of statistical features for power, angle, and torque for both hip and knee joints, cf., Table III. The result is one feature vector for each phase: $x_{HS}, x_{MS}, x_{TO}, x_{SW} \in \mathbb{R}^{12}$. The Lokomat provides measurements of all the signals listed in Table III synchronized by the gait index signal, which makes the computation of the features simple.

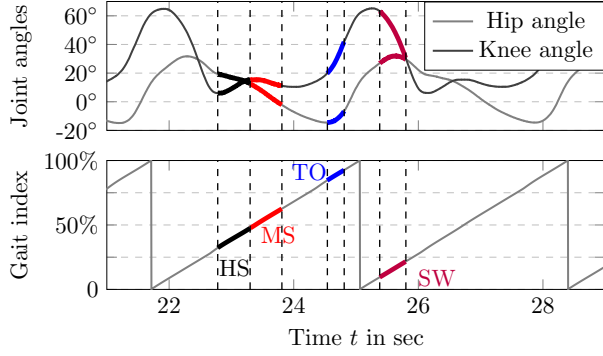


Fig. 5. Top: Joint angles. Bottom: Segmentation of time signals into four phases using the gait index with HS in 34.5%-47.5%, MS in 47.5%-65.5%, TO in 84.5%-92.5%, and SW in 9.5%-21.5% of one period of the gait index. The falling edge of the gait index does not align with the biomechanical definition of a gait cycle but enables separation of the gait cycle into phases.

TABLE III
VALUES FOR FEATURE VECTOR

#	Joint	Signal	Unit	Feature
x^1	hip	joint power	Nm/s	mean
x^2	hip	angle	rad	min
x^3	hip	angle	rad	max
x^4	hip	angle	rad	range
x^5	hip	torque	Nm	mean
x^6	hip	torque	Nm	variance
x^7	knee	joint power	Nm/s	mean
x^8	knee	angle	rad	min
x^9	knee	angle	rad	max
x^{10}	knee	angle	rad	range
x^{11}	knee	torque	Nm	mean
x^{12}	knee	torque	Nm	variance

Remark 5: For each subject, the data collection takes around one hour, including rating the gait cycle. As described in Algorithm 1, the application of the control law does not include further training and the control law is therefore not personalized to the subject.

Remark 6: Initially, we defined more features than the twelve in Table III, e.g., frequency domain features, which the supervised learning problem in (5) with L1 regularization discarded. In order to reduce the problem dimension in the online algorithm, we discarded them as well.

C. Reward Model and Steady-State Model for Lokomat

Given the data set, we apply the method in Section III to learn four reward models. We obtain a reward model for each of the four phases represented as \mathbf{W}_j , \mathbf{w}_j , and b_j from solving (5), where $j \in \{\text{HS}, \text{MS}, \text{TO}, \text{SW}\}$.

The steady-state model $\mathbf{M} \in \mathbb{R}^{7 \times 48}$, $\mathbf{m} \in \mathbb{R}^7$ in (7) is computed by stacking the features of the four phases:

$$\mathbf{x} = [\mathbf{x}_{\text{HS}}^T \quad \mathbf{x}_{\text{MS}}^T \quad \mathbf{x}_{\text{TO}}^T \quad \mathbf{x}_{\text{SW}}^T]^T.$$

D. Control Law for Gait Rehabilitation Robot

Once the four reward models and the steady-state model are trained using the data in (2), the controller automatically

chooses input setting adaptations given the current measurements, i.e., it does not require ratings from therapists. The input adaptation $\Delta \mathbf{s}$ is computed as

$$\Delta \mathbf{s} = \mathbf{M} \begin{bmatrix} \alpha(\mathbf{W}_{\text{HS}}\mathbf{x}_{\text{HS}} + \mathbf{w}_{\text{HS}}) + \mathbf{x}_{\text{HS}} \\ \alpha(\mathbf{W}_{\text{MS}}\mathbf{x}_{\text{MS}} + \mathbf{w}_{\text{MS}}) + \mathbf{x}_{\text{MS}} \\ \alpha(\mathbf{W}_{\text{TO}}\mathbf{x}_{\text{TO}} + \mathbf{w}_{\text{TO}}) + \mathbf{x}_{\text{TO}} \\ \alpha(\mathbf{W}_{\text{SW}}\mathbf{x}_{\text{SW}} + \mathbf{w}_{\text{SW}}) + \mathbf{x}_{\text{SW}} \end{bmatrix} + \mathbf{m} - \mathbf{s}.$$

While $\Delta \mathbf{s}$ yields continuous values, the input settings are adjusted in discrete steps, cf., the step-sizes in Table I. We aim to change one setting at a time, which is common practice for therapists [7] and eases the evaluation. The following suggests a method to select one single adaptation from $\Delta \mathbf{s}$.

Input Setting Selection & Stopping Criterion: In order to select one single discrete change in input setting, we normalize $\Delta \mathbf{s}$ to account for different value ranges and different units per individual setting and select the input corresponding to the largest in absolute value:

$$k^* = \arg \max_{k=1, \dots, 7} \left| \frac{\Delta s^k}{\bar{s}^k - \underline{s}^k} \right|$$

with associated index k^* , where the normalization $\bar{s}^k - \underline{s}^k$ is the range of the input setting k in Table I. Hence, the algorithm chooses one adaptation with step-size in Table I. The input adaptation is stopped when the largest normalized absolute value of change is smaller than a pre-defined parameter β , i.e., $|\Delta s^{k^*} / (\bar{s}^{k^*} - \underline{s}^{k^*})| \leq \beta$. This indicates closeness to the optimum, i.e., that a physiological gait is reached.

V. MODEL EVALUATION IN SIMULATION

We first analyze the algorithm in simulation to investigate the model quality. In this simulation study, we compare two reward models: One that uses ratings on an integer scale from 1 to 3 ($S = 3$ in (5)) and one that uses only binary ratings, i.e., good and bad ($S = 2$ in (5)). For the case $S = 3$, we use the collected ratings without modification. For the case $S = 2$, we combine the data points with rating 1 and rating 2 as *samples of a bad gait* and use the data points with rating 3 as *samples of a good gait*.

A. Evaluation Metrics and Results

In order to evaluate the trained models, we split the experimentally collected data into training (80%) and validation data (20%). This split is done randomly and repeated 500 times to assess the robustness of the models. This technique is known as 5-fold cross validation [50] and ensures that the validation data are not biased by training on the same data.

1) *Evaluation of Reward Model:* We evaluate the accuracy of the reward model trained with three ratings, i.e., $S = 3$, by computing the pairwise difference in estimated rewards $r(\mathbf{x}_i) - r(\mathbf{x}_j)$ for two data samples i and j , classified with respect to their ratings r_i and r_j . The metric is motivated by the fact that two different ratings should be distinguishable. We define $\Delta \bar{r}_{nm}$ as

$$\Delta \bar{r}_{nm} = \frac{1}{|I_n||I_m|} \sum_{i \in I_n} \sum_{j \in I_m} (r(\mathbf{x}_i) - r(\mathbf{x}_j)), \quad (8)$$

where $I_n = \{i | r_i = n\}$ is an index set of data points with ratings $r_i = n$. If the trained reward model and data were perfect, $\Delta \bar{r}_{nm} = n - m$ with zero standard deviation.

We evaluate the accuracy of the reward model trained with binary ratings, i.e., $S = 2$, by computing the classification accuracy of good versus bad ratings:

$$\bar{p}_{\text{good/bad}} = \frac{1}{|I_{\text{good}}||I_{\text{bad}}|} \sum_{i \in I_{\text{good}}} \sum_{j \in I_{\text{bad}}} \mathcal{I}_{r(\mathbf{x}_i) > r(\mathbf{x}_j)} \quad (9)$$

with $\mathcal{I}_{r(\mathbf{x}_i) > r(\mathbf{x}_j)} = 1$ if $r(\mathbf{x}_i) > r(\mathbf{x}_j)$ and $\mathcal{I}_{r(\mathbf{x}_i) > r(\mathbf{x}_j)} = 0$, otherwise, and $I_{\text{good}} = I_3$ and $I_{\text{bad}} = I_1 \cup I_2 = \{i | r_i = 1 \text{ or } r_i = 2\}$.

Fig. 6 reports statistical values of both $\Delta \bar{r}_{nm}$ in (8) for evaluating the reward model with $S = 3$ and $\bar{p}_{\text{good/bad}}$ in (9) for evaluating the reward model with $S = 2$ over the 500 splits of training and validation data. For the reward model computed with $S = 3$, the overall deltas in estimated rewards match the deltas in ratings very closely with 2.00 for $\Delta \bar{r}_{31}$, 0.97 for $\Delta \bar{r}_{32}$, and 1.04 for $\Delta \bar{r}_{21}$. For the reward model computed with $S = 2$, the overall classification accuracy ($r(\mathbf{x}_i) > r(\mathbf{x}_j)$ if $r_i > r_j$) is 92.5%.

2) *Evaluation of Steady-State Model:* The steady-state model is evaluated using the prediction error \bar{e}^k defined as

$$\bar{e}^k = \frac{1}{N} \sum_{i=1}^N |s_i^k - \mathbf{M}_{k*} \mathbf{x}_i - \mathbf{m}_k|, \quad (10)$$

where k is the index of the input setting and \mathbf{M}_{k*} is the k th row of matrix \mathbf{M} and \mathbf{m}_k is the k th entry of vector \mathbf{m} . As we use normalized values for the input settings with $s_i^k \in [0, 1]$, the error \bar{e}^k can be interpreted as a percentage offset from the correct input setting.

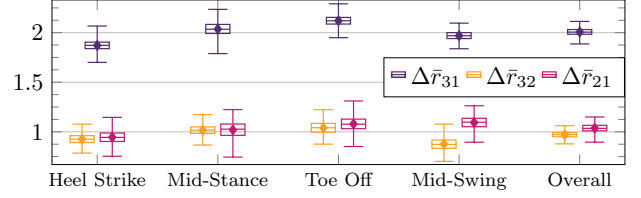
Table IV reports mean and standard deviation of the errors \bar{e}^k in (10) over the 500 random splits of training and validation data for all input settings k . It shows an overall average error of 4.17% and that the errors for all input settings are consistently lower than 6%.

3) *Evaluation of Overall Algorithm:* We evaluate the performance of the overall algorithm by comparing the collected data with the output of the algorithm. Let the changes in input settings during data collection for all data samples $i = 1, \dots, N$ be $\Delta \mathbf{s}_i^{\text{ex}} = \mathbf{s}^{\text{ex}} - \mathbf{s}_i$, where \mathbf{s}_i are the input settings of data point i and \mathbf{s}^{ex} are the physiological settings, which are set by the therapist at the beginning of the data collection. Note that \mathbf{s}^{ex} depends on the subject, however, we omit this dependency in the notation for ease of exposition. It is also important to note that \mathbf{s}^{ex} are not the only possible physiological input settings. We compare the input adaptation proposed by our algorithm $\Delta \mathbf{s}_i$ against the deviation from the physiological settings $\Delta \mathbf{s}_i^{\text{ex}}$, where we can have three different outcomes:

Case 1 (Same Setting & Same Direction): The algorithm selects the input adaptation in the same direction as during data collection, which is known to be a correct choice as it is closer to the physiological settings \mathbf{s}^{ex} .

Case 2 (Same Setting & Opposite Direction): The algorithm selects the same setting but in the opposite direction as during data collection, which is likely to be an incorrect choice.

Evaluation of reward model with three ratings



Evaluation of reward model with binary ratings

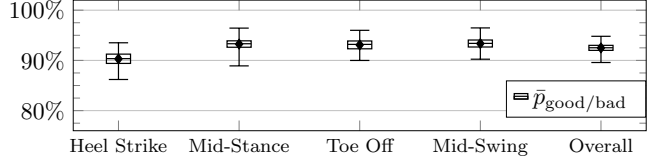


Fig. 6. Evaluation of reward models for individual phases and overall. The mean over 500 splits of training and validation data, along with the median, 25th and 75th percentiles, and maximum and minimum values are indicated by a diamond, a line, box edges, and whiskers, respectively. Top: Pairwise difference in estimated rewards $\Delta \bar{r}_{31}$, $\Delta \bar{r}_{32}$, and $\Delta \bar{r}_{21}$. Bottom: Classification accuracy $\bar{p}_{\text{good/bad}}$.

TABLE IV
MEAN AND STANDARD DEVIATION OF STEADY-STATE MODEL

s^k	Setting	Error \bar{e}^k
s^1	Hip Range of Motion	0.0578 ± 0.0019
s^2	Hip Offset	0.0370 ± 0.0008
s^3	Knee Range of Motion	0.0547 ± 0.0021
s^4	Knee Offset	0.0324 ± 0.0009
s^5	Speed	0.0307 ± 0.0009
s^6	Orthosis Speed	0.0315 ± 0.0010
s^7	Bodyweight Support	0.0542 ± 0.0017
	Overall	0.0417 ± 0.0186

Case 3 (Different Setting): The algorithm selects a different input adaptation, the implications of which are unknown and could be either correct or incorrect, which cannot be evaluated without closed-loop testing.

We compute the percentage of data points falling in each case for each setting k and for $\Delta \mathbf{s}_i^{\text{ex}} = 0$ (no adaptation), i.e., p_{C1}^k , p_{C2}^k , and p_{C3}^k for Case 1, Case 2, and Case 3, respectively, where $p_{C1}^k + p_{C2}^k + p_{C3}^k = 1$. If the algorithm replicated the data collection perfectly, then $p_{C1}^k = 1$ for all settings k . Given the discrete and unique setting selection, the overall algorithm has 15 options to choose from: An increase in one of the seven settings by one unit, a decrease in one of the seven settings by one unit, or *no adaptation*. Hence, random decision-making yields a probability of $p = 1/15 \approx 6.7\%$ for each option.

Table V reports mean and standard deviation of the percentage values of the three cases. The algorithm chooses the input adaptations for hip range of motion, hip offset, knee range of motion, and knee offset very often when their adaptation leads to \mathbf{s}^{ex} (86.7%–100.0%). Also, it often chooses *no adaptation* when the gait is physiological, with input settings \mathbf{s}^{ex} . Table V also shows that decision-making with the proposed algorithm is more ambiguous for the input adaptations of speed, orthosis speed, and bodyweight support. Overall, the algorithm proposes a setting that is closer to \mathbf{s}^{ex} (Case 1) in 80.7% and 80.6% for the reward models trained

TABLE V
EVALUATION OF OVERALL ALGORITHM IN SIMULATION

Three ratings (1, 2, or 3) $S = 3$				
s^k	Setting	p_{C1}^k in %	p_{C2}^k in %	p_{C3}^k in %
	No Adaptation	77.6 ± 3.6	-	22.4 ± 3.6
s^1	Hip Range of Motion	86.7 ± 1.8	0	13.3 ± 1.8
s^2	Hip Offset	96.4 ± 1.0	0	3.6 ± 1.0
s^3	Knee Range of Motion	91.0 ± 1.7	0	9.1 ± 1.7
s^4	Knee Offset	100.0 ± 0.0	0	0
s^5	Speed	71.1 ± 2.8	0	29.0 ± 2.8
s^6	Orthosis Speed	33.3 ± 3.7	3.5 ± 1.6	63.2 ± 3.8
s^7	Bodyweight Support	55.6 ± 3.8	0	44.5 ± 3.8
	Overall accuracy	80.7 ± 1.0	0.3 ± 0.1	19.0 ± 1.0

Binary ratings (good or bad) $S = 2$				
s^k	Setting	p_{C1}^k in %	p_{C2}^k in %	p_{C3}^k in %
	No Adaptation	76.8 ± 3.7	-	23.2 ± 3.7
s^1	Hip Range of Motion	88.6 ± 1.7	0	11.4 ± 1.7
s^2	Hip Offset	95.6 ± 1.1	0	4.4 ± 1.1
s^3	Knee Range of Motion	90.0 ± 2.0	0	10.0 ± 2.0
s^4	Knee Offset	100.0 ± 0.0	0	0
s^5	Speed	71.3 ± 2.9	0	28.7 ± 2.9
s^6	Orthosis Speed	32.1 ± 3.8	2.9 ± 1.4	65.0 ± 3.9
s^7	Bodyweight Support	53.9 ± 3.9	0	46.2 ± 3.9
	Overall accuracy	80.6 ± 1.1	0.2 ± 0.1	19.2 ± 1.0

with $S = 3$ and $S = 2$, respectively. The algorithm suggests a probably incorrect input adaptation in less than 1% (Case 2). In around 19%, the algorithm suggests a different input adaptation (Case 3).

Remark 7: The same evaluation using exclusively kinematic features ($x^2, x^3, x^4, x^8, x^9, x^{10}$ in Table III) yields slightly different results with overall $p_{C1} = 79.1\% \pm 1.0\%$, $p_{C2} = 0.2\% \pm 0.1\%$, and $p_{C3} = 20.6\% \pm 1.0\%$ for binary ratings. A purely kinematic feature vector might be important when working with impaired patients, where power/torque features might be an indication of individual impairments rather than a characterization of a physiological gait.

B. Discussion

The rewards predicted with the reward model trained with three ratings (two classification boundaries at 1.5 and 2.5), match the true ratings very closely. The reward model trained with binary ratings (one classification boundary at 2.5) is able to distinguish *good* from *bad* gait patterns confidently with an overall classification accuracy of more than 90%. The steady-state model shows an average error of 5%. As we will show in Section VI, this accuracy suffices for the considered application. For example, the expected error of 3.07% of s^5 translates into an error in treadmill speed of 0.075m/s and the expected error of 5.78% of s^1 translates into an error in hip range of motion of 2.08° , which is less than one input setting step-size, cf., Table I. Even though another model may increase accuracy, it may come at the expense of increased complexity in the computation. Our linear model only requires matrix-vector multiplication, which can easily be implemented on the controller of the Lokomat and is chosen as a suitable compromise of simplicity and accuracy.

The evaluation of both components, the reward model and the steady-state model, in simulation allow us to conclude that

TABLE VI
TEST SCENARIOS OF EXPERIMENT

Scenario: Observations	Therapists' heuristic rules (expectation)
1: Limited foot clearance, foot dropping	Increase knee range of motion ($s^3 \uparrow$)
2: Short steps	Increase hip range of motion, speed ($s^1 \uparrow, s^5 \uparrow$)
3: Foot dragging	Decrease speed, increase orthosis speed ($s^5 \downarrow, s^6 \uparrow$)
4: Large steps, late heel strike	Decrease hip range of motion, hip offset ($s^1 \downarrow, s^2 \downarrow$)
5: Short steps, hip extension	Increase hip range of motion, hip offset ($s^1 \uparrow, s^2 \uparrow$)
6: Bouncing	Decrease speed, bodyweight support ($s^5 \downarrow, s^7 \downarrow$)
7: Foot slipping	Decrease knee range of motion, orthosis speed ($s^3 \downarrow, s^6 \downarrow$)
8: Knee buckling	Increase knee range of motion, bodyweight support ($s^3 \uparrow, s^7 \uparrow$)
9: Large steps, early heel strike	Decrease hip range of motion, increase hip offset, increase speed ($s^1 \downarrow, s^2 \uparrow, s^5 \uparrow$)
10: Large steps, late heel strike, foot slipping	Decrease hip range of motion, hip offset, knee range of motion, orthosis speed ($s^1 \downarrow, s^2 \downarrow, s^3 \downarrow, s^6 \downarrow$)

they provide suitable models for the considered application. For the overall algorithm, Case 1 is known to result in an improved physiology of the gait cycle. Case 3, however, does not imply that the suggested adaptation will not lead to an improved gait cycle as there may be multiple different input adaptations that lead to a physiological gait (not only s^{ex}). In these cases, we do not know if the suggested adaptation would have led to an improvement in gait without closed-loop testing. Hence, the probabilities 80.7% and 80.6% of Case 1 for the two reward models can be interpreted as a lower bound for the overall improvement. The relatively low standard deviation for all settings indicates that the learning is robust against variation in the training data. The use of binary ratings eases the data collection and has been shown to perform similarly well. Therefore, we proceed with closed-loop testing of the algorithm using a reward model trained with binary ratings.

VI. EXPERIMENTAL RESULTS - CLOSED-LOOP TESTING

The proposed algorithm was implemented as a recommendation system on the Lokomat for closed-loop evaluation. We implemented the algorithm using the reward model trained with binary ratings (good and bad) of the gait cycle. It is important to note, that no data from the respective test person was used for training of the reward or the steady-state model.

A. Experiment Setup

We conducted experiments with ten nondisabled subjects and ten different sets of initial non-physiological gait cycles (test scenarios). Table VI describes the ten test scenarios and outlines input adaptations that therapists are expected to make (according to the heuristic guidelines). Scenario 1 through 8 are very common observations of a patient's gait cycle on the Lokomat. Scenario 9 and 10 are combinations of the more common flaws and are included to challenge the algorithm with more complex scenarios.

The selected scenarios cover the most common observations of the gait cycle of a passive subject (without muscle activity) on the Lokomat and, therefore, are expected to adequately evaluate the proposed algorithm experimentally (with nondisabled subjects). The initial input settings to achieve non-physiological gait patterns (scenarios and observations in Table VI) were chosen manually and purposefully by experienced therapists individually for each subject until the respective observation was present. The guidance force was set to 100% for all trials, i.e., the subjects were walked by the Lokomat. The treadmill speed was varied between 1.4km/h and 2.3km/h.

We conducted 63 experimental trials with the proposed algorithm in closed-loop with the ten nondisabled subjects, where each subject underwent at least five trials. The difference in the number of experimental trials is due to each subject's availability. However, the test scenarios were chosen so that each scenario was tested comparably often. Similarly to the data collection, the subjects were instructed to be as passive as possible. Two therapists assessed the input adaptations suggested by the algorithm and rated whether the gait was physiological. The therapists implemented the input adaptations until the algorithm indicated that a physiological gait cycle had been reached. Additionally, the therapists indicated when they thought that a physiological gait had been reached and the algorithm should be stopped.

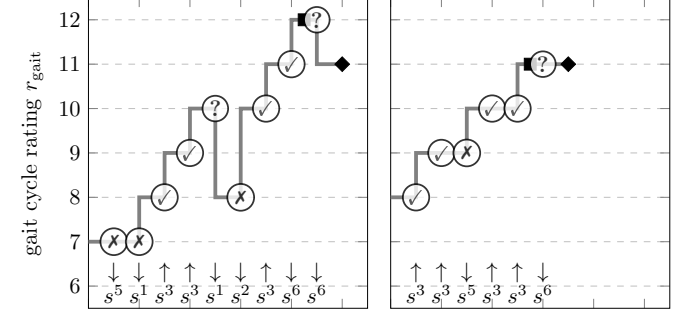
B. Results

Fig. 7 illustrates eight representative trials with the first subject. It contains four types of information and is separated by therapist in columns and by test scenario in rows:

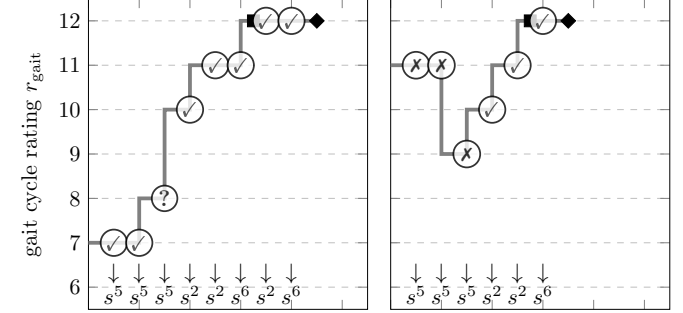
- i) The gait cycle rating r_{gait} (y-axis), calculated as the sum of the individual phase ratings $r_{\text{gait}} = r_{\text{HS}} + r_{\text{MS}} + r_{\text{TO}} + r_{\text{SW}}$, over the number of input adaptations (x-axis);
- ii) the applied input adaptations and their direction, e.g., $s_1 \uparrow$ represents an increase of Setting 1 by one unit;
- iii) a statement from the therapists about the algorithm's suggested input adaptation, i.e., agreement with the suggestion as check mark \checkmark , disagreement as cross \times , and uncertainty about the suggestion as question mark $?$; and
- iv) the reaching of a physiological gait judged by the therapist with square markers \blacksquare (for the usage as recommendation system) and by the algorithm with diamond markers \blacklozenge (for the usage as automatic adaptation system).

In all eight illustrated experiments, the algorithm provides a reliable, although not monotonic, improvement in the physiology of the gait. The input adaptations suggested by the algorithm led to a physiological gait for both the usage as recommendation system (square marker) and automatic adaptation system (diamond marker) in less than 10 adaptations with an overall rating of greater than or equal to 11, where 12 is the maximum possible rating. The input adaptations during the test of Scenario 1 with both therapists (first row) are similar to the heuristic guidelines in Table VI, i.e., an increase in the knee range of motion ($s^3 \uparrow$). The input adaptations for Scenario 2 (second row) are different from the heuristic guidelines. Here, the algorithm converges to a kinematically

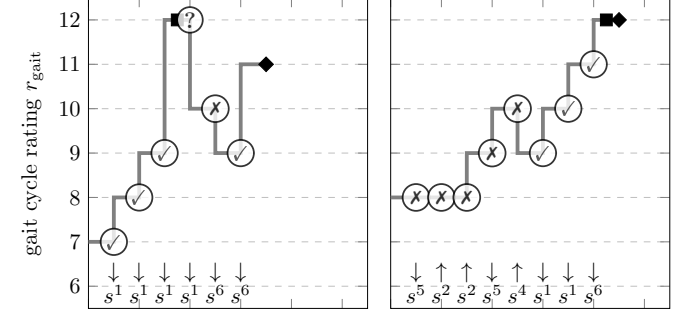
Scenario 1: Limited foot clearance, foot dropping



Scenario 2: Short steps



Scenario 9: Large steps, early heel strike



Scenario 10: Large steps, late heel strike, foot slipping

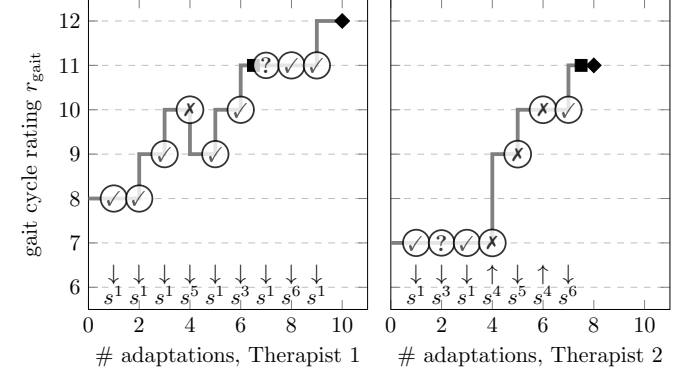


Fig. 7. Experimental evaluation of the closed-loop recommendation system. Averaged for the eight experiments, a physiological gait was reached after 6.0 input adaptations (until square marker). The diamond marker indicates that the algorithm assessed the gait as physiological (stopping criterion).

different but physiological gait that is achieved through a slower treadmill speed and input settings that are adjusted accordingly. For Scenario 9 and 10 (third and fourth row), the algorithm achieved a physiological gait through adaptations that are similar to the heuristic guidelines. In all illustrated cases, the algorithm converges to a physiological gait.

TABLE VII
SUMMARY OF ALL EXPERIMENTAL TRIALS

Subject (body type)	Trials	APT	Physiology of gait	
			improved	not degraded
1 (193cm, 93kg, male)	8	6.0	65%	92%
2 (195cm, 100kg, male)	5	3.8	89%	100%
3 (163cm, 53kg, female)	6	6.8	68%	98%
4 (175cm, 85kg, female)	5	4.8	58%	92%
5 (172cm, 68kg, female)	9	4.9	57%	89%
6 (190cm, 85kg, male)	5	7.4	54%	97%
7 (167cm, 85kg, male)	6	7.2	60%	93%
8 (180cm, 75kg, male)	5	7.0	60%	83%
9 (167cm, 64kg, female)	7	5.7	65%	97%
10 (161cm, 48kg, female)	7	6.9	71%	92%
Overall	63	6.0	63%	93%

Table VII summarizes all 63 experimental trials with ten subjects. On average, after a proposed input adaptation, the gait cycle improved in 63% and did not degrade in 93% of adaptations. The latter percentage is important as sometimes, changing an input setting by only one unit is too small to make a noticeable change in the gait cycle and a couple of consecutive adaptations are necessary, e.g., for the orthosis speed (s^6). Most importantly, a physiological gait cycle was reached for all trials within 6.0 adaptations per trial (APT) on average for all subjects combined, and between 3.8 and 7.4 for each individual subject.

C. Discussion

In general, the algorithm reached a physiological gait cycle within very few adaptations. This is achieved as the proposed algorithm reasons about the gait cycle using the reward model in a higher dimensional feature space rather than the space of input settings. As a result, the controller does not rely on trial and error search and, therefore, does not require to be tuned individually for each patient, which makes the approach more efficient, e.g., compared to classical reinforcement learning methods. The majority of times, the therapists agreed with the suggestions from the algorithm, i.e., the suggested adaptations were conform with the heuristic tuning guidelines and their experience. Consequently, the resulting gait cycle was mostly kinematically similar to the one that the therapists would have chosen. In some notable instances, the therapists disagreed or were uncertain about the proposition and were surprised by the improvement in the gait cycle, e.g., Row 1, Therapist 1, Adaptation 6; Row 2, Therapist 1, Adaptation 3; or Row 4, Therapist 2, Adaptation 4 in Fig. 7. These instances are examples of situations where the algorithm chooses input adaptations, which were unknown to the therapists. In these cases, the resulting gait cycle was sometimes kinematically different to the heuristic guidelines, e.g., a gait with slower treadmill speed. Table VII shows that the algorithm is able to cope with various body types with similar results for all individuals.

It is worth noting that the differences between similar scenarios with two different therapists in Fig. 7 and the same initial input settings do not necessarily lead to the same adjustment of input settings. This observation can be explained

as the physiology of the gait does not only depend on the chosen input settings but also on the hardware setup, e.g., the tightness of the straps, which differs slightly between therapists. However, even though the hardware was setup slightly differently by the two therapists, the algorithm managed to find input settings that walk the subject physiologically, indicating a certain robustness to slight variations in the hardware.

VII. CONCLUSION AND FUTURE WORK

This paper has derived a supervised learning-based method utilizing human ratings for learning parameters of a feedback controller. The approach was applied to the Lokomat robotic gait trainer with the goal of automatically adjusting the input settings to reach a physiological gait cycle by encapsulating the therapists' expertise in a reward model. Feedback control was enabled by this reward model and a steady-state model, which allows for converting desired changes in gait into input adaptations. Experiments with human subjects showed that the therapists' expertise in the form of ratings of four gait phases provides sufficient information to discriminate between physiological and non-physiological gait cycles. Furthermore, the provided adaptations led to an improvement of the gait cycle towards a physiological one within fewer than ten adaptations. The physiological gait cycle was partly reached by changes in input settings that domain experts would not have chosen themselves, suggesting that the proposed method might also be capable of generalizing from ratings and proposing improved settings for unseen scenarios. This observation remains to be confirmed with more data in future work.

Future work involves the data collection, evaluation, and validation of the proposed method with impaired patients. This will include the assessment of asymmetric gait adaptations for the right and left legs, which can readily be achieved by considering one feature vector for each leg. Further, physical limitations and/or constraints in the patients' movements could be assessed online using sensor measurements of the Lokomat and considered for the selection of input settings.

ACKNOWLEDGMENT

We gratefully acknowledge Patricia Andreia Gonçalves Rodrigues, Ursula Costa, and Serena Maggioni for their clinical support and invaluable discussions; Luca Somaini and Nils Reinert for their technical support; and Liliana Pavel for implementing the user interface.

REFERENCES

- [1] G. Colombo, M. Joerg, R. Schreier, and V. Dietz, "Treadmill training of paraplegic patients using a robotic orthosis," *J. Rehab. Res. and Develop.*, vol. 37, no. 6, pp. 693–700, 2000.
- [2] D. J. Reinkensmeyer, J. L. Emken, and S. C. Cramer, "Robotics, motor learning, and neurologic recovery," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 497–525, 2004.
- [3] J. L. Emken and D. J. Reinkensmeyer, "Robot-enhanced motor learning: accelerating internal model formation during locomotion by transient dynamic amplification," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 13, no. 1, pp. 33–39, 2005.
- [4] L. Marchal-Crespo and D. J. Reinkensmeyer, "Review of control strategies for robotic movement training after neurologic injury," *J. Neuroeng. and Rehabil.*, vol. 6, no. 1, p. 20, 2009.

- [5] O. Lamberg, L. Lünenburger, R. Gassert, and M. Bolliger, “Robots for measurement/clinical assessment,” in *Neurorehabil. Technol.* London: Springer, 2012, pp. 443–456.
- [6] H. M. Van der Loos, D. J. Reinkensmeyer, and E. Guglielmelli, “Rehabilitation and health care robotics,” in *Springer handbook of robotics*. Springer, 2016, pp. 1685–1728.
- [7] Hocoma, “Instructor script Lokomat training,” Hocoma AG, Tech. Rep., 2015.
- [8] S. Jezernik, G. Colombo, and M. Morari, “Automatic gait-pattern adaptation algorithms for rehabilitation with a 4-dof robotic orthosis,” *IEEE J. Robot. Automat.*, vol. 20, no. 3, pp. 574–582, 2004.
- [9] R. Riener, L. Lünenburger, S. Jezernik, M. Anderschitz, G. Colombo, and V. Dietz, “Patient-cooperative strategies for robot-aided treadmill training: First experimental results,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 13, no. 3, pp. 380–394, 2005.
- [10] R. Riener, L. Lünenburger, and G. Colombo, “Human-centered robotics applied to gait training and assessment,” *J. Rehab. Res. and Develop.*, vol. 43, no. 5, pp. 679–694, 2006.
- [11] A. Duschau-Wicke, J. Von Zitzewitz, R. Banz, and R. Riener, “Iterative learning synchronization of robotic rehabilitation tasks,” *Proc. IEEE Int. Conf. Rehabil. Robot.*, pp. 335–340, 2007.
- [12] A. Duschau-Wicke, J. von Zitzewitz, A. Caprez, L. Lünenburger, and R. Riener, “Path control: a method for patient-cooperative robot-aided gait rehabilitation,” *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 18, no. 1, pp. 38–48, 2010.
- [13] S. Maggioni, L. Lünenburger, R. Riener, and A. Melendez-Calderon, “Robot-aided assessment of walking function based on an adaptive algorithm,” *Proc. IEEE Int. Conf. Rehabil. Robot.*, pp. 804–809, 2015.
- [14] J. L. Emken, R. Benitez, A. Sideris, J. E. Bobrow, and D. J. Reinkensmeyer, “Motor adaptation as a greedy optimization of error and effort,” *J. Neurophysiol.*, vol. 97, no. 6, pp. 3997–4006, 2007.
- [15] A. Koenig, X. Omlin, L. Zimmerli, M. Sapa, C. Krewer, M. Bolliger, F. Müller, and R. Riener, “Psychological state estimation from physiological recordings during robot-assisted gait rehabilitation,” *J. Rehabil. Res. & Develop.*, vol. 48, no. 4, pp. 367–386, 2011.
- [16] R. Begg and J. Kamruzzaman, “A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data,” *J. Biomechanics*, vol. 38, no. 3, pp. 401–408, 2005.
- [17] J. Wu, J. Wang, and L. Liu, “Feature extraction via KPCA for classification of gait patterns,” *Human Movement Sci.*, vol. 26, no. 3, pp. 393–411, 2007.
- [18] M. Yang, H. Zheng, H. Wang, S. McClean, J. Hall, and N. Harris, “A machine learning approach to assessing gait patterns for complex regional pain syndrome,” *Med. Eng. & Physics*, vol. 34, no. 6, pp. 740–746, 2012.
- [19] P. Tahafchi, R. Molina, J. A. Roper, K. Sowalsky, C. J. Hass, A. Gunduz, M. S. Okun, and J. W. Judy, “Freezing-of-gait detection using temporal, spatial, and physiological features with a support-vector-machine classifier,” *IEEE Int. Conf. Eng. in Medicine and Biol. Soc.*, pp. 2867–2870, 2017.
- [20] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 1998.
- [21] W. B. Knox and P. Stone, “Interactively shaping agents via human reinforcement: The TAMER framework,” *Proc. 5th Int. Conf. Knowledge Capture*, pp. 9–16, 2009.
- [22] P. M. Pilarski, M. R. Dawson, T. Degris, F. Fahimi, J. P. Carey, and R. S. Sutton, “Online human training of a myoelectric prosthesis controller via actor-critic reinforcement learning,” *Proc. IEEE Int. Conf. Rehabil. Robot.*, pp. 134–140, 2011.
- [23] S. Griffith, K. Subramanian, and J. Scholz, “Policy shaping: Integrating human feedback with reinforcement learning,” *Adv. Neural Inform. Process. Syst.*, pp. 2625–2633, 2013.
- [24] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, “Deep reinforcement learning from human preferences,” in *Adv. Neural Inform. Process. Syst.*, 2017, pp. 4302–4310.
- [25] N. Wilde, D. Kulić, and S. L. Smith, “Learning user preferences in robot motion planning through interaction,” in *Proc. IEEE Int. Conf. Robot. and Automat.*, 2018, pp. 619–626.
- [26] A. D. D. Dorsa Sadigh, S. Sastry, and S. A. Seshia, “Active preference-based learning of reward functions,” in *Robot.: Sci. and Syst.*, 2017.
- [27] C. Basu, M. Singhal, and A. D. Dragan, “Learning from richer human guidance: Augmenting comparison-based learning with feature queries,” in *Proc. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2018, pp. 132–140.
- [28] M. Menner and M. N. Zeilinger, “A user comfort model and index policy for personalizing discrete controller decisions,” in *Proc. Eur. Control Conf.*, 2018, pp. 1759–1765.
- [29] R. E. Kalman, “When is a linear control system optimal?” *J. Basic Eng.*, vol. 86, no. 1, pp. 51–60, 1964.
- [30] A. Y. Ng and S. J. Russell, “Algorithms for inverse reinforcement learning,” in *Proc. 17th Int. Conf. Mach. Learning*, 2000, pp. 663–670.
- [31] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, “Learning-based model predictive control: Toward safe learning in control,” *Annu. Rev. Control, Robot., and Auton. Syst.*, vol. 3, 2020.
- [32] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, “Recent advances in robot learning from demonstration,” *Annu. Rev. Control, Robot., and Auton. Syst.*, vol. 3, 2020.
- [33] K. Mombaur, A. Truong, and J.-P. Laumond, “From human to humanoid locomotion: An inverse optimal control approach,” *Auton. Robots*, vol. 28, no. 3, pp. 369–383, 2010.
- [34] D. Clever, R. M. Schemschat, M. L. Felis, and K. Mombaur, “Inverse optimal control based identification of optimality criteria in whole-body human walking on level ground,” in *6th IEEE Int. Conf. Biomed. Robot. and Biomechanics*, 2016, pp. 1192–1199.
- [35] A. L. E. N. Kleesattel and K. Mombaur, “Inverse optimal control based enhancement of sprinting motion analysis with and without running-specific prostheses,” in *7th IEEE Int. Conf. Biomed. Robot. and Biomechanics*, 2018, pp. 556–562.
- [36] M. Menner, P. Worsnop, and M. N. Zeilinger, “Constrained inverse optimal control with application to a human manipulation task,” *IEEE Trans. Control Syst. Technol.*, 2019, doi:10.1109/TCST.2019.2955663.
- [37] J. F.-S. Lin, V. Bonnet, A. M. Panchea, N. Ramdani, G. Venture, and D. Kulić, “Human motion segmentation using cost weights recovered from inverse optimal control,” in *IEEE-RAS 16th Int. Conf. Humanoid Robots*, 2016, pp. 1107–1113.
- [38] P. Englert, N. A. Vien, and M. Toussaint, “Inverse KKT: Learning cost functions of manipulation tasks from demonstrations,” *Int. J. Robot. Res.*, vol. 36, no. 13–14, pp. 1474–1488, 2017.
- [39] M. Menner, K. Berntorp, M. N. Zeilinger, and S. Di Cairano, “Inverse learning for human-adaptive motion planning,” in *Proc. 58th IEEE Conf. Decision and Control*, 2019, pp. 809–815.
- [40] M. Menner and M. N. Zeilinger, “Convex formulations and algebraic solutions for linear quadratic inverse optimal control problems,” in *Proc. Eur. Control Conf.*, 2018, pp. 2107–2112.
- [41] P. Abbeel and A. Y. Ng, “Apprenticeship learning via inverse reinforcement learning,” in *Proc. 21st Int. Conf. Mach. Learning*, 2004, p. 1.
- [42] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, “Maximum entropy inverse reinforcement learning,” in *Proc. 23rd AAAI Conf. Artif. Intell.*, vol. 8, 2008, pp. 1433–1438.
- [43] S. Levine, Z. Popovic, and V. Koltun, “Nonlinear inverse reinforcement learning with Gaussian processes,” in *Adv. Neural Inform. Process. Syst.*, 2011, pp. 19–27.
- [44] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, “Cooperative inverse reinforcement learning,” in *Adv. Neural Inform. Process. Syst.*, 2016, pp. 3909–3917.
- [45] K. Bogert, J. F.-S. Lin, P. Doshi, and D. Kulić, “Expectation-maximization for inverse reinforcement learning with hidden data,” in *Proc. Int. Conf. Auton. Agents & Multiagent Syst.*, 2016, pp. 1034–1042.
- [46] V. Joukov and D. Kulić, “Gaussian process based model predictive controller for imitation learning,” in *IEEE-RAS 17th Int. Conf. Humanoid Robot.*, 2017, pp. 850–855.
- [47] C. M. Bishop, *Pattern Recognition and Machine Learning*. Secaucus, NJ: Springer, 2006.
- [48] R. Riener, L. Lünenburger, I. C. Maier, G. Colombo, and V. Dietz, “Locomotor training in subjects with sensori-motor deficits: an overview of the robotic gait orthosis lokomat,” *J. Healthcare Eng.*, vol. 1, no. 2, pp. 197–216, 2010.
- [49] J. Perry, *Gait Analysis: Normal and Pathological Function*. Thorofare, NJ: SLACK Incorporated Inc., 1992.
- [50] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. New York: Springer, 2013, vol. 112.