2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

2021 IEEE International Conference on Robotics and Automation (ICRA) | 978-1-7281-9077-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICRA48506.2021.9560840

ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes

Kejun Li[1], Maegan Tucker[1], Erdem Bıyık[2], Ellen Novoseller[1],
Joel W. Burdick[1], Yanan Sui[3], Dorsa Sadigh[2], Yisong Yue[1], and Aaron D. Ames[1]

***Abstract*— Characterizing what types of exoskeleton gaits are comfortable for users, and understanding the science of walking more generally, require recovering a user's utility landscape. Learning these landscapes is challenging, as walking trajectories are defined by numerous gait parameters, data collection from human trials is expensive, and user safety and comfort must be ensured. This work proposes the Region of Interest Active Learning (ROIAL) framework, which actively learns each user's underlying utility function over a region of interest that ensures safety and comfort. ROIAL learns from ordinal and preference feedback, which are more reliable feedback mechanisms than absolute numerical scores. The algorithm's performance is evaluated both in simulation and experimentally for three non-disabled subjects walking inside of a lower-body exoskeleton. ROIAL learns Bayesian posteriors that predict each exoskeleton user's utility landscape across four exoskeleton gait parameters. The algorithm discovers both commonalities and discrepancies across users' gait preferences and identifies the gait parameters that most influenced user feedback. These results demonstrate the feasibility of recovering gait utility landscapes from limited human trials.**

I. Introduction

Lower-body exoskeleton research aims to restore mobility to people with paralysis, a group with nearly 5.4 million people in the US alone [1]. Currently, the relationship between exoskeleton users' preferences and the exoskeleton's walking parameters is poorly understood. On the scientific front, such an understanding could yield insight into the science of walking, for instance, why exoskeleton users prefer certain gaits to others. On the direct clinical side, identifying the gaits that users prefer is critical for rehabilitation and assistive device design. Existing approaches for customizing exoskeleton walking include optimizing factors such as body parameters and targeted walking speeds [2], [3], minimizing metabolic cost [4], [5], and optimizing user comfort [6]–[8]. More specifically, the work in [6]–[8] demonstrated the notion of optimizing exoskeleton gaits based on user preferences to find the optimal gait for each exoskeleton user. Learning from preferences is beneficial because it has been shown that pairwise preferences (e.g. "Does the user prefer A or B?") are often more reliable than numerical scores [9].

Major challenges of learning exoskeleton users' preferences include: working with limited data from time-intensive human subject experiments, ensuring user comfort and safety, accounting for user feedback reliability, and exploring the vast action space of all possible walking trajectories. Broadly speaking, methods for preference-based learning can be designed for two distinct goals. The first is **optimization**: finding optimal gaits for specific users. The second is **understanding**: reliably learning entire preference landscapes. Given the need to maximize sample efficiency from limited trials, each choice of goal implies a different sampling strategy for data collection. Previous work [6], [7] focused on the first goal of direct function optimization, and so their approach did not reliably learn the *entire* utility landscape governing user preferences across gaits. Thus, we propose an alternative approach aiming for the second goal of characterizing the entire landscape, albeit with less fine-grained data in the region close to the optimal gait.

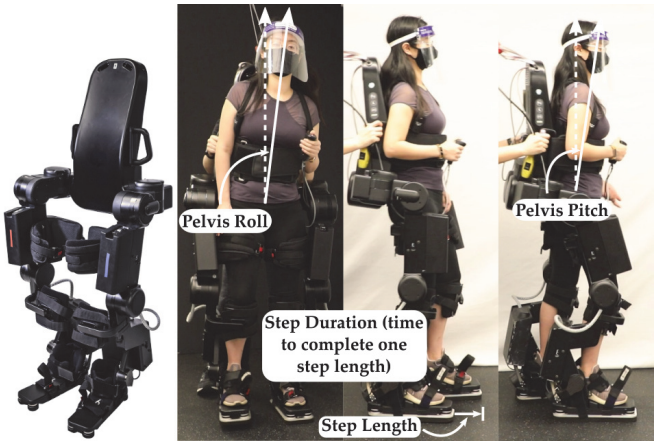


Fig. 1: The Atalante exoskeleton, designed by Wandercraft, has 12 actuated joints, 6 on each leg. The experiments explore four gait parameters: step length, step duration, pelvis roll, and pelvis pitch.

A consequence of exploring the entire gait parameter space is that users may be repeatedly exposed to gaits that make them feel unsafe or uncomfortable. In this work, we denote this region of undesirable gaits the "Region of Avoidance" (ROA) and the region of remaining gaits the "Region of Interest" (ROI). In prior work on the highly-related area of safe exploration [10]–[13], unsafe actions are considered to be catastrophically bad and therefore must be avoided completely. However, the resulting algorithms can be overly conservative in settings such as ours, where occasionally sampling from bad regions is tolerable.

This work proposes the Region of Interest Active Learning (ROIAL) algorithm, a novel active learning framework which queries the user for qualitative or preference feedback to: 1) locate the ROI, and 2) estimate the utility function

This research was supported by NIH grant EB007615, NSF NRI award 1924526 and CMMI award 1923239, NSF Graduate Research Fellowship No. DGE-1745301, and the Caltech Big Ideas and ZEITLIN Funds.
This work was conducted under IRB No. 16-0693.
[1]California Inst. of Technology, [2]Stanford University, [3]Tsinghua University

as accurately and quickly as possible over the ROI. The algorithm selects samples by modeling a Bayesian posterior over the utility function using Gaussian processes and maximizing the information gain (over the ROI) with respect to this posterior. Information gain maximization for preference elicitation is a sample-efficient, state-of-the-art approach that generates preference queries that are easy for users to answer accurately [14]–[16]. To our knowledge, our approach is the first to tackle such a region of interest active learning task.

The vast majority of prior work on preference learning obtains at most 1 bit of information per preference query [6]–[8], [14]–[22]. ROIAL additionally learns from ordinal labels [23], which assign actions to r discrete ordered categories such as "bad," "neutral," and "good." Ordinal feedback enables ROIAL to both: 1) locate the ROI by learning the boundary between the least-preferred category (ROA) and remaining actions (ROI), and 2) estimate the utility function more efficiently within the ROI. Compared to the 1 bit of information obtained per preference, each ordinal query yields up to $\log_2(r)$ bits of information. Since ordinal feedback is identical for actions within each ordinal category, preferences provide finer-grained information about the utility function's shape within the categories.

We validate ROIAL both in simulation and experimentally. We demonstrate in simulation that ROIAL estimates both the ROI and the utility function within the ROI with high accuracy. We experimentally demonstrate ROIAL on the lower-body exoskeleton Atalante (Fig. 1) to learn the utility functions of three non-disabled users over four gait parameters. The obtained landscapes highlight both agreement and disagreement in preferences among the users. Previous algorithms for exoskeleton gait optimization were incapable of drawing such conclusions; thus, this work represents progress towards establishing a better understanding of the science of walking with respect to exoskeleton gait design.

II. Problem Statement

We consider an active learning problem over a finite (but potentially-large) action space $\mathcal{A} \subset \mathbb{R}^d$ with $A = |\mathcal{A}|$. Each action $\boldsymbol{a} \in \mathcal{A}$ is assumed to have an underlying utility to the user, $f(\boldsymbol{a})$. The algorithm aims to learn the unknown utility function $f : \mathcal{A} \to \mathbb{R}$. The actions' utilities can be written in the vectorized form $\boldsymbol{f} := [f(\boldsymbol{a}^{(1)}), f(\boldsymbol{a}^{(2)}), ..., f(\boldsymbol{a}^{(A)})]^\top$, where $\{\boldsymbol{a}^{(k)} \mid k = 1, \dots, A\}$ are the actions in $\mathcal{A}$. Let $\boldsymbol{a}_i \in \mathcal{A}$ be the action selected in trial i, where $i \in \{1, \dots, N\}$. We receive qualitative information about f after each trial i, consisting of an ordinal label y_i and (possibly) a preference between $\boldsymbol{a}_i$ and $\boldsymbol{a}_{i-1}$ for $i \geq 2$. We use $\boldsymbol{a}_{k1} \succ \boldsymbol{a}_{k2}$ to denote a preference for action $\boldsymbol{a}_{k1}$ over $\boldsymbol{a}_{k2}$, and following each trial i, collect these preferences into a dataset $\mathcal{D}_p^{(i)} = \{\boldsymbol{a}_{k1} \succ \boldsymbol{a}_{k2} \mid k = 1, 2, ..., N_p^{(i)}\}$. Since preference feedback is not necessarily given for every trial, $N_p^{(i)} \leq i - 1$. The ordinal labels are similarly collected into $\mathcal{D}_o^{(i)} = \{(\boldsymbol{a}_k, y_k) \mid k = 1, 2, ..., N_o^{(i)}\}$. The full user feedback dataset after iteration i is defined as $\mathcal{D}_i := \mathcal{D}_p^{(i)} \cup \mathcal{D}_o^{(i)}$.

Ordinal feedback assigns one of r ordered labels to each sampled action. These (possibly-noisy) labels are assumed to reflect ground truth ordinal categories (e.g., "bad," "neutral," "good," etc.), which partition $\mathcal{A}$ into r sets $\mathbf{O}_j$, $j \in \{1, \dots, r\}$. We define the ROA as $\mathbf{O}_1$; for instance, in the exoskeleton setting, it consists of gaits that make the user feel unsafe or uncomfortable. Similarly, the ROA could be defined as $\bigcup_{j=1}^{n} \mathbf{O}_j$ for $n > 1$, where the choice of n is task-specific given the ordinal category definitions. We define the ROI as the complement of the ROA, $\mathcal{A} \setminus \mathbf{O}_1$.

Defining $\hat{\boldsymbol{f}}_i := [\hat{f}_i(\boldsymbol{a}^{(1)}), \dots, \hat{f}_i(\boldsymbol{a}^{(A)})]^\top$ as the maximum a posteriori (MAP) estimate of the utilities $\boldsymbol{f}$ given $\mathcal{D}_i$, we aim to adaptively select the N actions $\boldsymbol{a}_1, \dots, \boldsymbol{a}_N \in \mathcal{A}$ that minimize the error in estimating $\boldsymbol{f}$ over the ROI. Defining $\boldsymbol{u} \in \{0, 1\}^A$ as a binary vector denoting which actions are within the ROI, we model the error as $\text{Error}(N) := \boldsymbol{u}^\top |\boldsymbol{f} - \hat{\boldsymbol{f}}_N|$, where the absolute value is taken element-wise.

Algorithm 1 ROIAL Algorithm

Require: Utility prior parameters; ordinal thresholds $b_1, \dots, b_{r-1}$; subset size M; confidence parameter λ

$\mathcal{D}_0 = \emptyset$, ▷ $\mathcal{D}_i$: user feedback dataset including iteration i
Select an action $\boldsymbol{a}_1$ at random
Add ordinal feedback to data to obtain $\mathcal{D}_1$
for i = 2,..., N **do**
  Update the model posterior $P(\boldsymbol{f} \mid \mathcal{D}_{i-1})$ ▷ Eq. (1)
  Determine $\mathcal{S}^{(i)}$ by randomly selecting M actions
  Determine $\mathcal{S}_{ROI}^{(i)} \subset \mathcal{S}^{(i)}$
  $\boldsymbol{a}_i \leftarrow \arg\max_{\boldsymbol{a} \in \mathcal{S}_{ROI}^{(i)}} I(\boldsymbol{f}; s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a})$
  Add preference and ordinal feedback to data to obtain $\mathcal{D}_i$
end for

III. Active Learning Algorithm

This section describes the ROIAL algorithm (Alg. 1), which leverages qualitative human feedback to estimate the ROI and utility function (code available at [24]). We first discuss Bayesian modeling of the utility function, and then explain the procedure for rendering it tractable in high dimensions. We then detail the process for estimating the ROI and approximating the information gain to select the most informative actions.

Bayesian Posterior Inference. To simplify notation, this section omits the iteration i from all quantities. Given the feedback dataset $\mathcal{D} = \mathcal{D}_p \cup \mathcal{D}_o$, the utilities $\boldsymbol{f}$ have posterior:

$$P(\boldsymbol{f} \mid \mathcal{D}_p, \mathcal{D}_o) \propto P(\mathcal{D}_p \mid \boldsymbol{f}) P(\mathcal{D}_o \mid \boldsymbol{f}) P(\boldsymbol{f}), \qquad (1)$$

where $P(\boldsymbol{f})$ is a Gaussian prior over the utilities $\boldsymbol{f}$:

$$P(\boldsymbol{f}) = \frac{1}{(2\pi)^{A/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{f}^\top \Sigma^{-1} \boldsymbol{f}\right),$$

in which $\Sigma \in \mathbb{R}^{A \times A}$, $\Sigma_{ij} = \mathcal{K}(\boldsymbol{a}_i, \boldsymbol{a}_j)$, and $\mathcal{K}$ is a kernel of choice. This work uses the squared exponential kernel.

Preference feedback. We assume that the users' preferences are corrupted by noise as in [25], such that:

$$P(\boldsymbol{a}_1 \succ \boldsymbol{a}_2 \mid \boldsymbol{f}) = g_p\left(\frac{f(\boldsymbol{a}_1) - f(\boldsymbol{a}_2)}{c_p}\right),$$

where $g_p : \mathbb{R} \to (0, 1)$ is a monotonically-increasing link function, and $c_p > 0$ quantifies noisiness in the preferences.

Ordinal feedback. We define thresholds $-\infty = b_0 < b_1 < b_2 < \ldots < b_r = \infty$ to partition the action space into r ordinal categories, $\mathbf{O}_1, \ldots, \mathbf{O}_r$. For any $\boldsymbol{a} \in \mathcal{A}$, if $f(\boldsymbol{a}) < b_1$, then $\boldsymbol{a} \in \mathbf{O}_1$, and $\boldsymbol{a}$ has an ordinal label of 1. Similarly, if $b_j \leq f(\boldsymbol{a}) < b_{j+1}$, then $\boldsymbol{a} \in \mathbf{O}_{j+1}$, and $\boldsymbol{a}$ corresponds to a label of $j+1$. We assume that the users' ordinal labels are corrupted by noise as in [23], such that:

$$P(y \mid \boldsymbol{f}, \boldsymbol{a}) = g_o\left(\frac{b_y - f(\boldsymbol{a})}{c_o}\right) - g_o\left(\frac{b_{y-1} - f(\boldsymbol{a})}{c_o}\right),$$

where $g_o : \mathbb{R} \to (0,1)$ is a monotonically-increasing link function, and $c_o > 0$ quantifies the ordinal noise.

Assuming conditional independence of queries, the likelihoods $P(\mathcal{D}_p \mid \boldsymbol{f})$ and $P(\mathcal{D}_o \mid \boldsymbol{f})$ are:

$$P(\mathcal{D}_p \mid \boldsymbol{f}) = \prod_{k=1}^{N_p} g_p\left(\frac{f(\boldsymbol{a}_{k1}) - f(\boldsymbol{a}_{k2})}{c_p}\right),$$

$$P(\mathcal{D}_o \mid \boldsymbol{f}) = \prod_{k=1}^{N_o} \left[g_o\left(\frac{b_{y_k} - f(\boldsymbol{a}_k)}{c_o}\right) - g_o\left(\frac{b_{y_k-1} - f(\boldsymbol{a}_k)}{c_o}\right)\right].$$

Our simulations and experiments fix the hyperparameters c_p, c_o, and $\{b_j \mid j = 1, \ldots, r-1\}$ in advance. One could also estimate them during learning using strategies such as evidence maximization, but this can be very computationally expensive, especially in high-dimensional action spaces.

Common choices of link function (g_p and g_o) include the Gaussian cumulative distribution function [23], [25] and the sigmoid function, $g(x) = (1 + e^{-x})^{-1}$ [7]. We model feedback via the sigmoid link function because empirical results suggest that a heavier-tailed noise distribution improves performance. We use the Laplace approximation to approximate the posterior as Gaussian: $P(\boldsymbol{f} \mid \mathcal{D}_i) \approx \mathcal{N}(\hat{\boldsymbol{f}}_i, \hat{\Sigma}_i)$ [26].

High-Dimensional Tractability. Calculating the model posterior is the algorithm's most computationally-expensive step, and is intractable for large action spaces. Most existing work in high-dimensional Gaussian process learning requires quantitative feedback [27], [28]. Previous work in preference-based high-dimensional Gaussian process learning [7] restricts posterior inference to one-dimensional subspaces. However, the approach in [7] is more amenable to the regret minimization problem because each one-dimensional subspace is biased toward regions of high posterior utility. Instead, to increase ROIAL's online computing speed over high-dimensional spaces, in each iteration i we select a subset $\mathcal{S}^{(i)} \subset \mathcal{A}$ of M actions uniformly at random, and evaluate the posterior only over $\mathcal{S}^{(i)}$.

Estimating the Region of Interest. Since we lack prior knowledge about the ROI, it must be estimated during the learning process. In each iteration i, we model the ROI as the set of actions $\{\boldsymbol{a}_k\}$ that satisfy the following criterion: $\hat{f}_{i-1}(\boldsymbol{a}_k) + \lambda\hat{\sigma}_{i-1}(\boldsymbol{a}_k) > b_1$, where $\hat{\sigma}_{i-1}(\boldsymbol{a}_k)$ is the posterior standard deviation associated with $\boldsymbol{a}_k$. The variable λ is a user-defined hyperparameter that determines the algorithm's conservatism in estimating the ROI; positive λ's are optimistic, while negative λ's are more conservative in avoiding the ROA. We evaluate actions in the randomly-selected subset $\mathcal{S}^{(i)}$ and define $\mathcal{S}^{(i)}_{ROI} = \{\boldsymbol{a} \in \mathcal{S}^{(i)} \mid \hat{f}_{i-1}(\boldsymbol{a}) + \lambda\hat{\sigma}_{i-1}(\boldsymbol{a}) > b_1\}$ in each iteration i. Note that this definition is optimistic, whereas safe exploration approaches use pessimistic definitions [10]–[13].

Action Selection via Information Gain Optimization. To learn the utility function in as few trials as possible, we select actions to maximize the mutual information between the utility function and the preference-based and ordinal human feedback. While optimizing the entire sequence of N actions is NP-hard [29], previous work has shown that a greedy approach which only optimizes the next immediate action achieves state-of-the-art data-efficiency [15]. Hence, we adopt the same approach to solve the following optimization in each iteration i:

$$\max_{\boldsymbol{a}_i \in \mathcal{S}^{(i)}_{ROI}} I(\boldsymbol{f}; s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i), \tag{2}$$

where s_i denotes the outcome of a pairwise preference elicitation between $\boldsymbol{a}_i$ and $\boldsymbol{a}_{i-1}$. One can rewrite (2) in terms of information entropy:

$$\max_{\boldsymbol{a}_i} H(s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i) - \mathbb{E}_{\boldsymbol{f}\mid\mathcal{D}_{i-1}}[H(s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i, \boldsymbol{f})].$$

We can interpret the first term as the uncertainty about action $\boldsymbol{a}_i$'s ordinal label and preference relative to $\boldsymbol{a}_{i-1}$. We aim to maximize this term, because queries with high model uncertainty could potentially yield significant information. The second term is conditioned on $\boldsymbol{f}$, and so represents the user's expected uncertainty. If the user is very uncertain about their feedback, then the action $\boldsymbol{a}_i$ gives only a small amount of information. Hence, we aim to minimize this second term. In this way, information gain optimization produces queries that are both informative and easy for users.

The second term is estimated via sampling from the Laplace-approximated Gaussian posterior $P(\boldsymbol{f} \mid \mathcal{D}_{i-1})$. Computing the first term requires the probability $P(s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i)$. We derive it as:

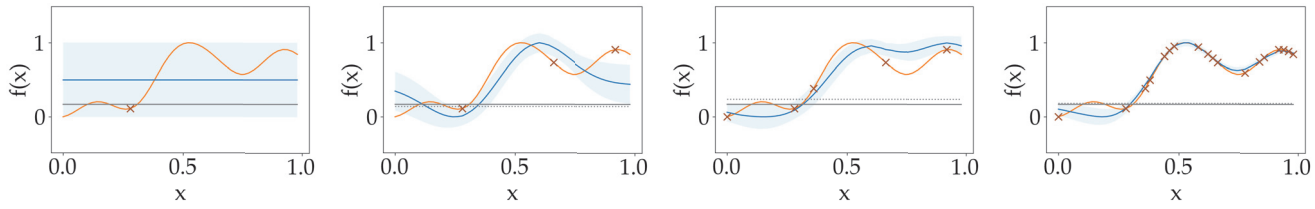
$$\begin{aligned} &P(s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i) \\ &\quad= \int_{\mathbb{R}^A} P(\boldsymbol{f} \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i) P(s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i, \boldsymbol{f}) d\boldsymbol{f} \\ &\quad= \mathbb{E}_{\boldsymbol{f}\mid\mathcal{D}_{i-1}}\left[P(s_i, y_i \mid \mathcal{D}_{i-1}, \boldsymbol{a}_i, \boldsymbol{f})\right], \end{aligned}$$

which we approximate with samples from $P(\boldsymbol{f} \mid \mathcal{D}_{i-1})$.

IV. Results

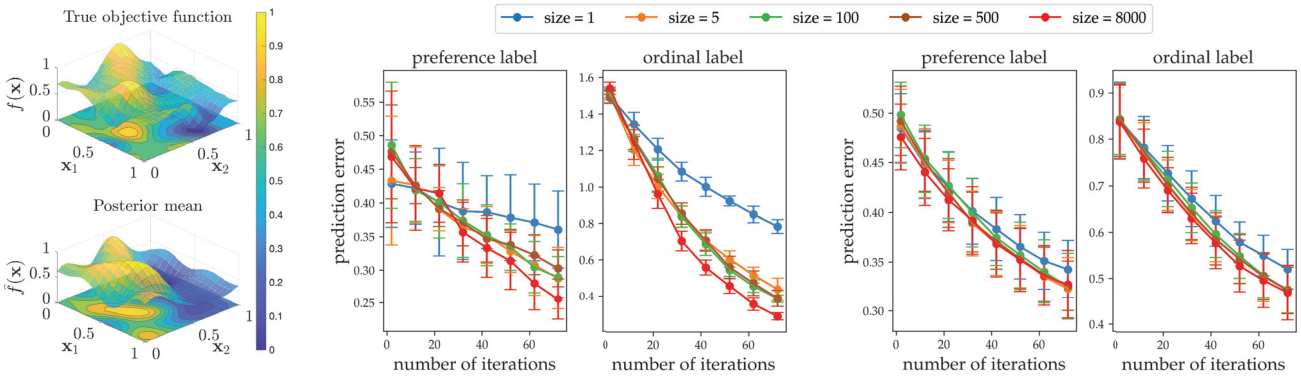
Simulation Results. We evaluate ROIAL's performance on the Hartmann3 (H3) function—which is a standard benchmark for learning non-convex, smooth functions—and on 3-dimensional synthetic functions, sampled from a Gaussian process prior over a $20 \times 20 \times 20$ grid. As evaluation metrics, we use the algorithm's errors in preference and ordinal label prediction; these allow us to quantify performance when the true utility function is unknown. The average ordinal prediction error is defined as $\overline{\text{Error}}(N) := \frac{1}{N}\sum_{k=1}^{N} |y_k^{pred} - y_k^{true}|$, and all simulations use 5 ordinal categories.[1]

[1]Unless otherwise stated, hyperparameters are held constant across simulations and experiments, and their values can be found in [24].



(a) Iteration 1 (b) Iteration 3 (c) Iteration 5 (d) Iteration 20

Fig. 2: 1D posterior illustration. The true objective function is shown in orange, and the algorithm's posterior mean is blue. Blue shading indicates the confidence region for $\lambda = 0.5$. The solid grey line indicates the true ordinal threshold b_1: the ROI is above this threshold, while the ROA is below it. The dotted grey line is the algorithm's b_1 hyperparameter. The actions queried so far are indicated with "x"s. Utilities are normalized in each plot so that the posterior mean spans the range from 0 to 1.



(a) Synthetic function posterior (b) Hartmann3 prediction error (c) Synthetic function prediction error

Fig. 3: Impact of random subset size on algorithm performance. a) Example 3D synthetic objective function and posterior learned by ROIAL with subset size = 500 after 80 iterations. Values are averaged over the 3rd dimension and normalized to range from 0 to 1. b-c) Algorithm's error in predicting preferences and ordinal labels (mean $\pm$ std). Each simulation evaluated performance at 1000 randomly-selected points; the model posterior was used to predict preferences between consecutive pairs of points and ordinal labels at each point.

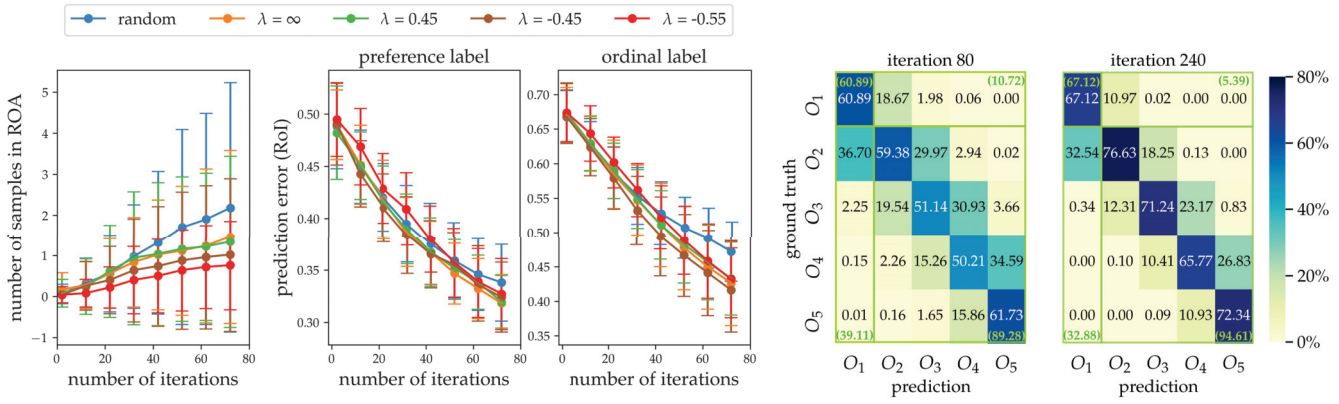
1D illustration of ROIAL. Fig. 2 illustrates the algorithm for a 1D objective function. Initially, ROIAL samples widely across the action space (Fig. 2a-2c). As seen by comparing iterations 5 and 20 (Fig. 2c-2d), the algorithm stops querying points in the ROA (actions in $\mathbf{O}_1$) because the upper confidence bound (top of the blue shaded region) there falls below the hyperparameter b_1 (dotted gray line).

Extending to higher dimensions. To characterize the impact of the random subset size on algorithmic performance, we compare performance of different sizes in simulation for both the H3 and synthetic functions. We calculate the posterior over the entire action space only every 10 steps to reduce computation time, and then use this posterior to evaluate the algorithm's error in predicting preference and ordinal labels. Fig 3a provides an example of a 3D posterior, Fig. 3b depicts the average performance for H3 over 10 simulation repetitions, and Fig. 3c shows the average performance over a set of 50 unique synthetic functions. We find that a subset size of at least 5 yields performance close to using all points.

Estimating the region of interest. We demonstrate the effect of the confidence parameter λ on the number of actions sampled from the ROA and on prediction error in the ROI. Fig. 4a demonstrates that across various values of λ, visits to the ROA decrease as λ decreases. To confirm that restricting queries to the estimated ROI does not harm performance, we also compare label prediction error in the ROI across values of λ. When $\lambda = -0.45$, ROIAL achieves similar preference prediction accuracy and slightly-improved ordinal label prediction within the ROI compared to $\lambda = \infty$, which permits sampling over the entire action space (Fig. 4a). Additionally, the confusion matrix (Fig. 4b) shows that the algorithm usually predicts either the correct ordinal label or an adjacent ordinal category. The ROI prediction accuracy (green text in Fig. 4b) indicates that ROIAL predicts whether points belong to the ROI with relatively-high accuracy.

Robustness to noisy feedback. Since user feedback is expected to be noisy, we evaluate the algorithm's robustness to noisy feedback generated from the distributions $P(y \mid \boldsymbol{f}, \boldsymbol{a}) = g_o\left(\frac{\tilde{b}_y - f(\boldsymbol{a})}{\tilde{c}_o}\right) - g_o\left(\frac{\tilde{b}_{y-1} - f(\boldsymbol{a})}{\tilde{c}_o}\right)$ and $P(\boldsymbol{a}_1 \succ \boldsymbol{a}_2 \mid \boldsymbol{f}) = g_p\left(\frac{f(\boldsymbol{a}_1) - f(\boldsymbol{a}_2)}{\tilde{c}_p}\right)$ for ordinal and preference feedback, respectively, with true ordinal thresholds $\{\tilde{b}_j | j = 1, \ldots, r-1\}$ and simulated noise parameters $\tilde{c}_p$ and $\tilde{c}_o$. We set $\tilde{c}_o > \tilde{c}_p$ because we expect ordinal labels to be noisier than preferences, as they require users to recall all past experience to give consistent feedback, whereas a preference only involves the previous and current action. The algorithm learns more slowly with noisier feedback (Fig. 5).

Exoskeleton Experiments. After demonstrating ROIAL's performance in simulation, we experimentally deployed it on the lower-body exoskeleton Atalante, developed by Wandercraft (video: [30], ROIAL hyperparameters: [24]). Atalante, shown in Fig. 1, is an 18 degree of freedom robot designed



(a) Number of samples in the ROA and prediction error in the ROI

(b) Confusion matrices

Fig. 4: Effect of the confidence interval. All simulations are run over 50 synthetic functions with a random subset size of 500. a) Left: cumulative number of actions in the ROA ($\mathbf{O}_1$) queried at each iteration (mean $\pm$ std). Note that as λ increases, more samples are required for the confidence interval to fall below the ROA threshold, at which point ROIAL starts avoiding the ROA. Middle and right: error in predicting preference and ordinal labels for different values of λ; predictions are over 1,000 random actions (mean $\pm$ std). b) Confusion matrices (column-normalized) of ordinal label prediction over the entire action space at iterations 80 and 240 with λ = -0.45. The 2×2 confusion matrices for ROI prediction accuracy are outlined in green. Prediction accuracy increases with the number of iterations.

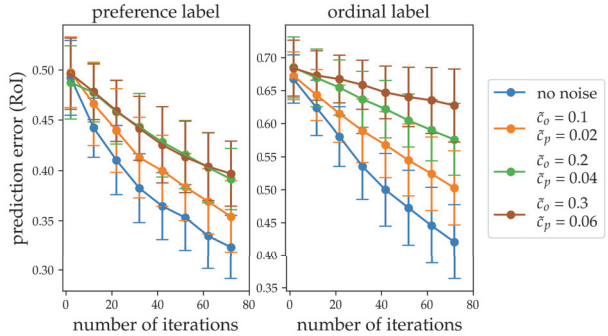


Fig. 5: Effect of noisy feedback. The ordinal and preference noise parameters, $\tilde{c}_0$ and $\tilde{c}_p$, range from 0.1 to 0.3 and 0.02 to 0.06, respectively. All cases use a random subset size of 500 and $\lambda = -0.45$, and each simulation uses 1,000 random actions to evaluate label prediction. Plots show means $\pm$ standard deviation.

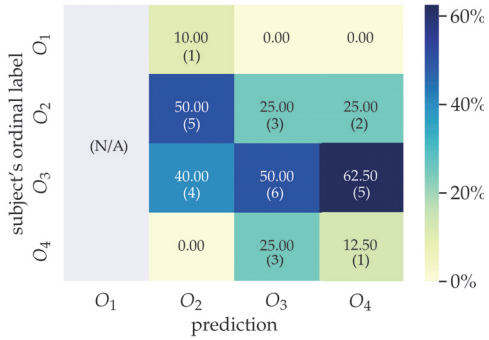


Fig. 6: Confusion matrix of the validation phase results for all three subjects. The first column is grey because actions in the ROA ($\mathbf{O}_1$) were purposefully avoided to prevent subject discomfort. Percentages are normalized across columns. Parentheses show the numbers of gait trials in each case.

to restore assisted mobility to patients with motor complete paraplegia through the control of 12 actuated joints: 3 joints at each hip, 1 joint at each knee, and 2 degrees of actuation in each ankle. For more details on Atalante, refer to [31]–[33].

Dynamically stable crutch-less exoskeleton walking gaits are generated through nonconvex optimization techniques (see Section II of [6]), based on the theory of hybrid zero dynamics (HZD) introduced by [34] and the HZD-based optimization method presented in [35]. These periodic gaits are parameterized by various features, and this studies focuses on four: step length (SL) in meters, step duration (SD) in seconds, maximum pelvis roll (PR) in degrees, and maximum pelvis pitch (PP) in degrees (Fig. 1). These parameters were selected because exoskeleton users frequently suggested modifications to SL, SD, and PR in prior work [36], and we wanted to further study the relationship between PR and PP. We discretized these parameters into bins of sizes 10, 7, 5, and 5, respectively, resulting in 1,750 actions within a 4D action space. ROIAL randomly selected 500 actions in each iteration and used $\lambda = 0.45$ to estimate the ROI.

The experimental procedure was conducted for three non-disabled subjects and consisted of 40 trials divided into a *training phase* (30 trials) and a *validation phase* (10 trials). Subjects were not informed of when the validation phase began. Subjects provided ordinal labels for all 40 gaits, and optional pairwise preferences between the current and previous gaits for all but the first trial. Four ordinal categories were considered and described to the users as:

1) **Very Bad** ($\mathbf{O}_1$): User feels unsafe or uncomfortable to the point that the user never wants to repeat the gait.
2) **Bad** ($\mathbf{O}_2$): User dislikes the gait but does not feel unsafe or uncomfortable.
3) **Neutral** ($\mathbf{O}_3$): User neither dislikes nor likes the gait and would be willing to try the gait again.
4) **Good** ($\mathbf{O}_4$): User likes the gait and would be willing to continue walking with it for a long period of time.

While including additional ordinal categories could increase the potential information gain from each query, it also increases the cognitive burden for the users and thus makes the labels less reliable. Validation actions were selected so that at least two samples were predicted to belong to $\mathbf{O}_2, \mathbf{O}_3$, and $\mathbf{O}_4$, with the remaining four validation actions sampled at random. Actions predicted to belong in $\mathbf{O}_1$ were excluded

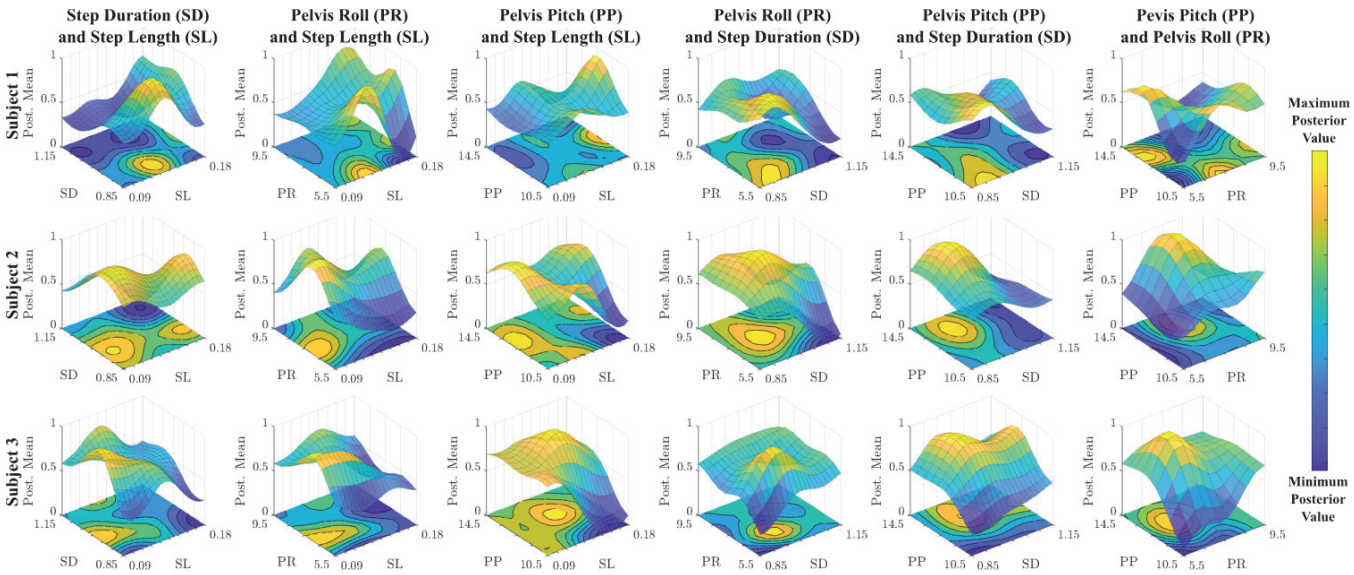


Fig. 7: 4D posterior mean utility across exoskeleton gaits. Utilities are plotted over each pair of gait space parameters, with the values averaged over the remaining 2 parameters in each plot. Each row corresponds to a subject: Subject 1 is the most experienced exoskeleton user, Subject 2 is the second-most experienced user, and Subject 3 never used the exoskeleton prior to the experiment.

because they are likely to make the user feel uncomfortable or unsafe, and actions sampled during the training phase were explicitly excluded from the validation trials.

Experimental results. Figure 6 depicts the results of the validation phase for all three subjects. These results show a reliable correlation between the predicted categories and the users' reported ordinal labels, in which the majority of the predicted ordinal labels are within one category of the true ordinal labels. Since less than 2% of the action space was explored during the experiment, we expect that the prediction accuracy would increase with additional exoskeleton trials as observed in simulation (Fig 4b). Overall, these results suggest that ROIAL can yield reliable preference landscapes within a moderate number of samples.

Figure 7 depicts the final posterior mean for each of the subjects. These utility functions highlight both regions of agreement and disagreement among the subjects. For example, all subjects strongly dislike gaits at the lower bound of PP and lower bound of PR. However, all subjects disagree in their utility landscapes across SL and SD. This type of insight could not be derived from direct gait optimization, which mostly obtains information near the optimum.

We also evaluated the effect of each gait parameter on the posterior utility using the permutation feature importance metric. The results of this test for each respective subject across the four gait parameters (SL, SD, PR, PP) are: (0.20, 0.30, 0.33, 0.27), (0.26, 0.36, 0.38, 0.29), and (0.23, 0.16, 0.21, 0.45). These values suggest that the preferences of more experienced users (Subjects 1 and 2) may be most influenced by SD and PR, while the least-experienced user's feedback may be most weighted by PP (Subject 3). The code for this test is available on GitHub [24]. These results demonstrate that ROIAL is capable of obtaining preference landscapes within relatively-few exoskeleton trials while avoiding gaits that make users feel unsafe or uncomfortable.

V. CONCLUSIONS

This work presents the ROIAL framework for actively learning utility functions within a region of interest from pairwise preferences and ordinal feedback. The ROIAL algorithm is experimentally demonstrated on the lower-body exoskeleton Atalante for three non-disabled subjects (video: [30]). In simulation, ROIAL predicts utilities in the ROI while learning to stay away from the ROA. In experiments, ROIAL typically predicts subjects' ordinal labels correctly to within one ordinal category. Furthermore, the results illustrate that gait preference landscapes vary across subjects. In particular, a feature importance test suggests that the two more-experienced users prioritized step duration and pelvis roll, while a new user prioritized pelvis pitch.

Making conclusive claims about gait preference landscapes requires conducting these experiments on patients with motor complete paraplegia, as well as scaling up the experiments. Another limitation of this work is the high noise in users' ordinal labels, which may depend on factors such as prior experience and bias due to the gait execution order. Thus, future work includes designing a study to directly quantify the noise in exoskeleton users' ordinal labels. Future work also includes continuing the experiments over more trials, as prediction accuracy is expected to improve with additional data. To conclude, the ROIAL algorithm provides a principled methodology for characterizing exoskeleton users' preference landscapes in high-dimensional action spaces. This work contributes to better understanding the mechanisms behind user-preferred walking and optimizing future gait generation for user comfort.

ACKNOWLEDGMENTS

The authors would like to thank the experiment volunteers and the entire Wandercraft team that designed Atalante and continues to provide technical support for this project.

REFERENCES

[1] B. S. Armour, E. A. Courtney-Long, M. H. Fox, H. Fredine, and A. Cahill, "Prevalence and causes of paralysis—United States, 2013," *American Journal of Public Health*, vol. 106, no. 10, pp. 1855–1857, 2016.

[2] X. Wu, D.-X. Liu, M. Liu, C. Chen, and H. Guo, "Individualized gait pattern generation for sharing lower limb exoskeleton robot," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 4, pp. 1459–1470, 2018.

[3] S. Ren, W. Wang, Z.-G. Hou, B. Chen, X. Liang, J. Wang, and L. Peng, "Personalized gait trajectory generation based on anthropometric features using random forest," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2019.

[4] M. Kim, Y. Ding, P. Malcolm, J. Speeckaert, C. J. Siviy, C. J. Walsh, and S. Kuindersma, "Human-in-the-loop Bayesian optimization of wearable device parameters," *PloS one*, vol. 12, no. 9, e0184054, 2017.

[5] J. Zhang, P. Fiers, K. A. Witte, R. W. Jackson, K. L. Poggensee, C. G. Atkeson, and S. H. Collins, "Human-in-the-loop optimization of exoskeleton assistance during walking," *Science*, vol. 356, no. 6344, pp. 1280–1284, 2017.

[6] M. Tucker, E. Novoseller, C. Kann, Y. Sui, Y. Yue, J. W. Burdick, and A. D. Ames, "Preference-based learning for exoskeleton gait optimization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[7] M. Tucker, M. Cheng, E. Novoseller, Y. Yue, J. Burdick, and A. D. Ames, "Human preference-based learning for high-dimensional optimization of exoskeleton walking gaits," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[8] N. Thatte, H. Duan, and H. Geyer, "A method for online optimization of lower limb assistive devices with high dimensional parameter spaces," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[9] T. Joachims, L. A. Granka, B. Pan, H. Hembrooke, and G. Gay, "Accurately interpreting clickthrough data as implicit feedback," in *SIGIR*, vol. 5, 2005, pp. 154–161.

[10] Y. Sui, A. Gotovos, J. Burdick, and A. Krause, "Safe exploration for optimization with Gaussian processes," in *International Conference on Machine Learning*, 2015.

[11] J. Schreiter, D. Nguyen-Tuong, M. Eberts, B. Bischoff, H. Markert, and M. Toussaint, "Safe exploration for active learning with Gaussian processes," in *Joint European conference on machine learning and knowledge discovery in databases*, Springer, 2015, pp. 133–149.

[12] F. Berkenkamp, A. P. Schoellig, and A. Krause, "Safe controller optimization for quadrotors with Gaussian processes," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016.

[13] Y. Sui, J. Burdick, Y. Yue, *et al.*, "Stagewise safe Bayesian optimization with Gaussian processes," in *International Conference on Machine Learning*, PMLR, 2018, pp. 4781–4789.

[14] N. Houlsby, F. Huszár, Z. Ghahramani, and M. Lengyel, "Bayesian active learning for classification and preference learning," *arXiv preprint arXiv:1112.5745*, 2011.

[15] E. Bıyık, N. Huynh, M. J. Kochenderfer, and D. Sadigh, "Active preference-based Gaussian process regression for reward learning," in *Proceedings of Robotics: Science and Systems (RSS)*, 2020.

[16] E. Bıyık, M. Palan, N. C. Landolfi, D. P. Losey, and D. Sadigh, "Asking easy questions: A user-friendly approach to active reward learning," in *Conference on Robot Learning (CoRL)*, 2019.

[17] Z. Xu, K. Kersting, and T. Joachims, "Fast active exploration for link-based preference learning using Gaussian processes," in *European Conference on Machine Learning*, 2010, pp. 499–514.

[18] J. Fürnkranz and E. Hüllermeier, "Preference learning and ranking by pairwise comparison," in *Preference learning*, Springer, 2010, pp. 65–82.

[19] N. Wilde, D. Kulic, and S. L. Smith, "Active preference learning using maximum regret," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[20] L. Qian, J. Gao, and H. Jagadish, "Learning user preferences by adaptive pairwise comparison," *Proceedings of the VLDB Endowment*, vol. 8, no. 11, pp. 1322–1333, 2015.

[21] Y. Sui, V. Zhuang, J. W. Burdick, and Y. Yue, "Multi-dueling bandits with dependent arms," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2017.

[22] V. Bengs, R. Busa-Fekete, A. El Mesaoudi-Paul, and E. Hüllermeier, "Preference-based online learning with dueling bandits: A survey," *Journal of Machine Learning Research*, vol. 22, no. 7, pp. 1–108, 2021.

[23] W. Chu and Z. Ghahramani, "Gaussian processes for ordinal regression," *Journal of machine learning research*, vol. 6, no. Jul, pp. 1019–1041, 2005.

[24] K. Li, *Repository for ROIAL*. https://github.com/kli58/ROIAL.

[25] W. Chu and Z. Ghahramani, "Preference learning with Gaussian processes," in *Proceedings of the 22nd international conference on machine learning*, 2005, pp. 137–144.

[26] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, 3. MIT press, 2006, vol. 2.

[27] K. Kandasamy, J. Schneider, and B. Póczos, "High dimensional Bayesian optimisation and bandits via additive models," in *International conference on machine learning*, 2015, pp. 295–304.

[28] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, N. De Freitas, *et al.*, "Bayesian optimization in high dimensions via random embeddings.," in *IJCAI*, 2013, pp. 1778–1784.

[29] N. Ailon, "An active learning algorithm for ranking from pairwise preferences with an almost optimal query complexity," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 137–164, 2012.

[30] *Supplementary video*, https://www.youtube.com/watch?v=04lMJmKmZrQ.

[31] A. Agrawal, O. Harib, A. Hereid, S. Finet, M. Masselin, L. Praly, A. D. Ames, K. Sreenath, and J. W. Grizzle, "First steps towards translating HZD control of bipedal robots to decentralized control of exoskeletons," *IEEE Access*, vol. 5, pp. 9919–9934, 2017.

[32] O. Harib, A. Hereid, A. Agrawal, T. Gurriet, S. Finet, G. Boeris, A. Duburcq, M. E. Mungai, M. Masselin, A. D. Ames, *et al.*, "Feedback control of an exoskeleton for paraplegics: Toward robustly stable, hands-free dynamic walking," *IEEE Control Systems Magazine*, vol. 38, no. 6, pp. 61–87, 2018.

[33] T. Gurriet, S. Finet, G. Boeris, A. Duburcq, A. Hereid, O. Harib, M. Masselin, J. Grizzle, and A. D. Ames, "Towards restoring locomotion for paraplegics: Realizing dynamically stable walking on exoskeletons," in *International Conference on Robotics and Automation*, IEEE, 2018, pp. 2804–2811.

[34] A. D. Ames, "Human-inspired control of bipedal walking robots," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1115–1130, 2014.

[35] A. Hereid, C. M. Hubicki, E. A. Cousineau, and A. D. Ames, "Dynamic humanoid locomotion: A scalable formulation for HZD gait optimization," *IEEE Transactions on Robotics*, vol. 34, no. 2, pp. 370–387, 2018.

[36] *Supplementary website*, https://sites.google.com/view/roial-icra2021.