Active Gait Rehabilitation using Inverse Reinforcement Learning

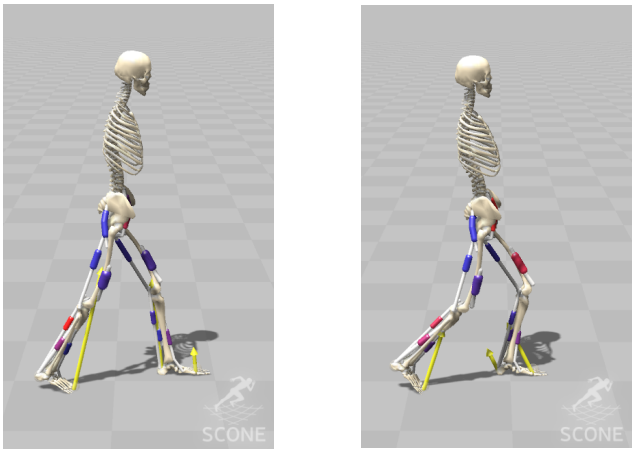
Zongwei Zhang*[1], Michael Drolet*[1], Firas Al-Hafez[1], Sebastian Hirt[2], Jan Peters[1]
[1]IAS, TU Darmstadt, Germany [2]CCPS, TU Darmstadt, Germany
zhang@ias.informatik.tu-darmstadt.de, sebastian.hirt@iat.tu-darmstadt.de
*{michael.drolet, firas.al-hafez, jan.peters}@tu-darmstadt.de *equal contribution*

1 Introduction

Exoskeletons are an attractive research direction in robotics since they involve interactions between humans and machines. Exoskeleton applications can be found in various domains, such as strength augmentation, motion assistance, and gait correction for medical use. Although many studies on exoskeleton control have been conducted, several challenges remain. For one, some methods focus solely on the exoskeleton itself and neglect the interaction/cooperation between humans and machines, resulting in the human having to adapt their motion to the exoskeleton [1]. Additionally, controls are often generated using traditional techniques that don't leverage user data. Although these methods can produce satisfactory results, one of the disadvantages is that they are often limited to providing support passively through predefined parameters without individualization, thus leading to poor adaptability across scenarios and latency in the motion [2,3]. These disadvantages can reduce the effectiveness of the wearable or even cause injury during gait correction in medical applications [4]. This paper presents a control strategy for the exoskeleton that optimizes the generation of healthy gait by correcting pathological gait based on the user's dynamic properties.

2 Methodology

The model used in this study is a bipedal musculoskeletal humanoid model, implemented in SCONE [5, 6]. The control strategy consists of two primary components: a low-level reflex controller for the muscles and a high-level joint motor controller. The low-level controller incorporates both the muscle reflex controller and the degrees of freedom (DoF) reflex controller, as proposed by Geyer and Herr [7]. The control law for the muscle reflex controller is defined as $U = C_0 + K_F[(F - F_0)]_+ + K_L[(L - L_0)]_+ + K_V[(V - V_0)]_+$, where K_F, F_0 represent the force feedback gain and offset; K_L, L_0 are the length feedback gain and offset; and K_V, V_0 correspond to the velocity feedback gain and offset. Here, F, L and V represent the real-time muscle force, muscle length, and muscle contractile velocity, respectively. The $[]_+$ operator ensures non-negative entries, and the output U represents the excitation. The control law for the DoF reflex controller is $U = C_0 + K_P(P - P_0) + K_V(V - V_0)$, where K_P, P_0 represent position feedback gain and target position, and K_V, V_0 represent velocity feedback gain and target velocity. The muscle is activated by the combined output of these reflex controllers, generating force to drive the model forward.



(a) **(b)**

Figure 1: (a) Expert gait and (b) Pathological gait with short hamstring whose MTU has 20% shorter optimal length. Both models are controlled solely by the low-level controller

Further details can be found in [8, 9].

In addition to the low-level reflex controller, four joint actuators are applied directly to the hips and knees in the sagittal plane, simulating exoskeleton functionality on the lower limbs. The high-level controller operates based on a policy trained using inverse reinforcement learning (IRL). Extending IRL to an observation-only setting is straightforward by defining the reward function solely based on observations. This approach is desirable because intrinsic variables, such as muscle activations or muscle forces, may not be directly measurable [10]. Even when these variables can be measured, variations in embodiment across cases would require extensive task-specific parameter tuning [11–13]. The process is simplified and consistent when using kinematic models to align the desired movement patterns. Therefore, IRL is a preferred method to match the feature space provided by the expert.

Generative Adversarial Imitation Learning (GAIL) is a prominent method used for IRL or imitation learning [14]. As opposed to using a hand-crafted reward function, in IRL, a reward function is learned based on expert demonstrations. In GAIL, which is inspired by Generative Adversarial Networks (GAN) [15], the reward is given by the discriminator. The generator (or policy) attempts to produce actions that mimic the expert's state-action pairs, while the discrimina-

tor distinguishes between the observation through generated actions and those of the expert. A suitable policy is learned once the discriminator can no longer differentiate between the agent and the expert. The discriminator's loss is computed as $\mathbb{E}_{\tau\sim\pi}[\log(D(s,a))] + \mathbb{E}_{\tau_E}[1 - \log(D(s,a))]$ where $\tau \sim \pi$ refers to the trajectory generated by the agent's policy, τ_E represents the expert's trajectory, and $D(s,a)$ denotes the discriminator's score. In our study, we employ Proximal Policy Optimization (PPO) for the policy update step in GAIL [16].

3 Experiments

To evaluate the feasibility of the IRL control method, we simulate a scenario where an exoskeleton is deployed on a patient model exhibiting pathological gait due to shortened hamstrings. All motors are assumed to share the same rotational axis as their corresponding joints. The feasible torque of each motor ranges from -50 to 50 N•m The objective of this simulation is to correct the pathological gait and assist the patient in walking with healthy gait patterns. The model includes 9 DoF and 14 muscle tendon units (MTU) that function as muscle actuators. To simulate the pathological gait, the optimal length of the hamstring MTU in the patient model is reduced by 20%, mimicking a common medical condition. The healthy model provides a dataset of joint movements of a healthy individual walking forward at an average speed of 1.2 m/s and it is solely controlled by reflex-based controller, with no external torque applied to the joints. An example of both healthy and pathological gaits is depicted in Figure 1.

Based on the patient model which is exclusively controlled by the low-level controller, the policy of the exoskeleton is trained through IRL. The actor and critic networks are comprised of two hidden layers, each with 512 units. The actor network takes as input an 18-dimensional observation, which includes joint values and joint velocities. The output is a 4-dimensional action, representing the external torques applied to the joints. The discriminator has two hidden layers, each with 64 units. A gradient penalty is added to help regularize the discriminator [17]. Figure 2 shows the comparison of trajectories from the expert, the impaired gait, and the corrected gait. The agent is trained for 3×10^7 steps (approximately 7 hours). It can be seen in Figure 2 that the corrected gait resembles the healthy gait more closely than the pathological gait (in terms of alignment and magnitude). Table 1 compares the differences between pathological vs. healthy and corrected vs. healthy gait. The MSE Loss and MAE Loss are calculated by summing all feature terms (averaged over 10 rollouts and all timesteps). Joint trajectories consist of values of each joint in [rad]. The dissimilarity metric includes the difference in metabolic cost in [W/kg], MSE Loss of trunk velocities in [m/s], ground contact forces, and the standardized MSE loss of full joint trajectories. Here, it can be seen that the corrected gait is also quantitatively closer to the expert trajectories than the pathological gait.

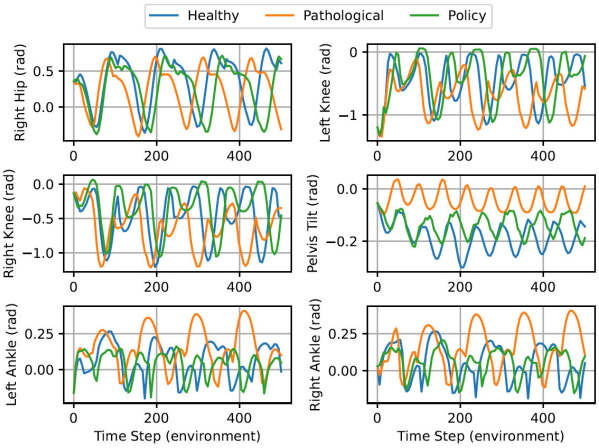


Figure 2: Trajectories of various DoF: Healthy gait (blue), Pathological gait (orange), and Corrected gait (green).

	MSE	MAE	Dissim.
Patho.	1.09 ± 0.012	2.12 ± 0.014	14.33 ± 0.002
Corrected	0.96 ± 0.001	1.82 ± 0.001	6.89 ± 0.004

Table 1: Quantification of differences between the pathological gait and corrected gait compared to the healthy gait.

We additionally study the effects of augmenting the reward function by adding task-specific rewards [18]. This combination of discriminator reward and task reward is defined as $r = w_{\text{v}} r_{\text{v}} + w_{\text{grf}} r_{\text{grf}} + w_{\text{limit}} r_{\text{limit}} + w_{\text{disc}} r_{\text{disc}}$, where r_{v}, r_{grf}, r_{limit} and r_{disc} represent the trunk velocity, ground contact force, joint limit torque, and discriminator score, respectively, and w_* terms are their corresponding weights/coefficients. We conducted a hyperparameter search over these coefficients. Notably, the discriminator score has the most significant impact on reducing dissimilarity. Among all the combinations tested, the lowest dissimilarity was achieved when w_{v} = 10.0, w_{grf} = -3.0, w_{limit} = -0.1 and w_{disc} = 10.0. Although this is perhaps unsurprising due to the definition of our dissimilarity metric, it demonstrates the feasibility of our approach compared to only using task-specific rewards.

4 Discussion

The approach presented in this paper demonstrates the potential of using GAIL to actively correct impaired gaits in exoskeleton-assisted motion, while considering the human-machine interaction. While this approach shows the ability of helping correct impaired gaits caused by hamstring shortening to match closer to the healthy gait patterns, it could be extended to address other gait impairments in the future, such as muscle weakness in the hamstring, iliopsoas and soleus. One possibility is to train separate policies for each type of impairment. Alternatively, future work would focus on developing a unified policy to adapt to various gait impairments. Future work may also include using more sample efficient IRL algorithms such as LS-IQ [19]

and more-closely accounting for the discrepancy in gait frequency using methods such as dynamic time warping. To more accurately simulate real world applications, future research could investigate scenarios where motors are kinematically misaligned. Furthermore, the gain and offset parameters of reflex controllers could be learned using other techniques, such as Reinforcement Learning (RL), to better represent real world conditions where the control methods are individually implemented.

5 Acknowledgements

This research was supported partly by the German Research Foundation (DFG) within RTG 2761 LokoAssist under grant no. 450821862.

References

[1] Shuzhen Luo, Ghaith Androwis, Sergei Adamovich, Hao Su, Erick Nunez, and Xianlian Zhou. Reinforcement learning and control of a lower extremity exoskeleton for squat assistance. *Frontiers in Robotics and AI*, 8:702845, 2021.

[2] Lingxing Chen, Chunjie Chen, Xin Ye, Zhuo Wang, Yao Liu, Wujing Cao, Shaocong Chen, and Xinyu Wu. A portable waist-loaded soft exosuit for hip flexion assistance with running. *Micromachines*, 13(2):157, 2022.

[3] Tao Xue, Ziwei Wang, Tao Zhang, and Meng Zhang. Adaptive oscillator-based robust control for flexible hip assistive exoskeleton. *IEEE Robotics and Automation Letters*, 4(4):3318–3323, 2019.

[4] Javad K Mehr, Eddie Guo, Mojtaba Akbari, Vivian K Mushahwar, and Mahdi Tavakoli. Deep reinforcement learning based personalized locomotion planning for lower-limb exoskeletons. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5127–5133. IEEE, 2023.

[5] Thomas Geijtenbeek. Scone: Open source software for predictive simulation of biological motion. *Journal of Open Source Software*, 4(38):1421, 2019.

[6] Thomas Geijtenbeek. The Hyfydy simulation software, 11 2021. `https://hyfydy.com`.

[7] Hartmut Geyer and Hugh Herr. A muscle-reflex model that encodes principles of legged mechanics produces human walking dynamics and muscle activities. *IEEE Transactions on neural systems and rehabilitation engineering*, 18(3):263–273, 2010.

[8] Hartmut Geyer, Andre Seyfarth, and Reinhard Blickhan. Positive force feedback in bouncing gaits? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1529):2173–2183, 2003.

[9] Thomas S Buchanan, David G Lloyd, Kurt Manal, and Thor F Besier. Neuromusculoskeletal modeling: estimation of muscle forces and joint moments and movements from measurements of neural command. *Journal of applied biomechanics*, 20(4):367–395, 2004.

[10] Samuel K Au, Paolo Bonato, and Hugh Herr. An emg-position controlled system for an active ankle-foot prosthesis: an initial experimental study. In *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.*, pages 375–379. IEEE, 2005.

[11] Masashi Hamaya, Takamitsu Matsubara, Tomoyuki Noda, Tatsuya Teramae, and Jun Morimoto. Learning assistive strategies for exoskeleton robots from user-robot physical interaction. *Pattern Recognition Letters*, 99:67–76, 2017.

[12] Shiyin Qiu, Wei Guo, Darwin Caldwell, and Fei Chen. Exoskeleton online learning and estimation of human walking intention based on dynamical movement primitives. *IEEE Transactions on Cognitive and Developmental Systems*, 13(1):67–79, 2020.

[13] Shuxiang Guo, Yibin Ding, and Jian Guo. Control of a lower limb exoskeleton robot by upper limb semg signal. In *2021 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1113–1118. IEEE, 2021.

[14] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. *Advances in neural information processing systems*, 29, 2016.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[16] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[17] Manu Orsini, Anton Raichuk, Léonard Hussenot, Damien Vincent, Robert Dadashi, Sertan Girgin, Matthieu Geist, Olivier Bachem, Olivier Pietquin, and Marcin Andrychowicz. What matters for adversarial imitation learning? *Advances in Neural Information Processing Systems*, 34:14656–14668, 2021.

[18] Alejandro Escontrela, Xue Bin Peng, Wenhao Yu, Tingnan Zhang, Atil Iscen, Ken Goldberg, and Pieter Abbeel. Adversarial motion priors make good substitutes for complex reward functions. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 25–32. IEEE, 2022.

[19] Firas Al-Hafez, Davide Tateo, Oleg Arenz, Guoping Zhao, and Jan Peters. Ls-iq: Implicit reward regularization for inverse reinforcement learning. In *Eleventh International Conference on Learning Representations (ICLR)*, 2023.