IFAC International Workshop on Periodic Control Systems (PSYCO 2016)

Reinforcement Learning of Potential Fields to achieve Limit-Cycle Walking

Denise S. Feirstein * **Ivan Koryakovskiy** * **Jens Kober** **
Heike Vallery *

* *Department of BioMechanical Engineering, TU Delft, Netherlands (denise.feirstein@gmail.com, {i.koryakovskiy, h.vallery}@tudelft.nl)*
** *DCSC, TU Delft, Netherlands (j.kober@tudelft.nl)*

Abstract: Reinforcement learning is a powerful tool to derive controllers for systems where no models are available. Particularly policy search algorithms are suitable for complex systems, to keep learning time manageable and account for continuous state and action spaces. However, these algorithms demand more insight into the system to choose a suitable controller parameterization. This paper investigates a type of policy parameterization for impedance control that allows energy input to be implicitly bounded: Potential fields. In this work, a methodology for generating a potential field-constrained impedance controller via approximation of example trajectories, and subsequently improving the control policy using Reinforcement Learning, is presented. The potential field-constrained approximation is used as a policy parameterization for policy search reinforcement learning and is compared to its unconstrained counterpart. Simulations on a simple biped walking model show the learned controllers are able to surpass the potential field of gravity by generating a stable limit-cycle gait on flat ground for both parameterizations. The potential field-constrained controller provides safety with a known energy bound while performing equally well as the unconstrained policy.

Keywords: Machine learning, Energy Control, Limit cycles, Walking, Robot control

1. INTRODUCTION

The demand for robot control that is both safe and energy-efficient is greater than ever with advances in mobile robots and robots that interact in human environments. One such example is the bipedal robot which has applications ranging from home care to disaster relief. Traditional position control, common to industrial robotics, is not suitable for robots that interact in unknown environments because slight position errors can result in high contact forces that can damage the robot and its environment. In the case of humanoid robots which interact in human environments this poses a human-safety issue.

One possible solution is to employ impedance control, which attempts to enforce a dynamic relation between system variables as opposed to controlling them directly (Hogan (1984)). Impedance control based on potential fields inherently bounds the energy exchanged between the robot and the environment. Potential fields can modulate natural dynamics of a system and achieve desired behavior without requiring high-stiffness trajectory tracking. Potential fields have been developed for path planning and motion control by reformulating the objective into a potential function (Koditschek (1987)). Control torques can be represented as a vector field generated by the gradient of the potential field, such that the dimensionality of any number of actuators is essentially reduced to one, the scalar value of the potential function.

Contrasting the high energy demand of conventional, fully actuated bipedal robots, passive dynamic walkers have been developed that walk down shallow slopes using only gravity and the robot's natural dynamics (McGeer (1990)). Thus, these mechanisms exploit the natural potential field of gravity. In consequence, they possess an extremely energy-efficient gait that is remarkably similar to that of humans. The stable periodic gait of a passive dynamic walker is referred to as a Limit Cycle (LC). Rendering this gait slope-invariant and improving its disturbance rejection has been the focus of many publications including Hobbelen and Wisse (2007). For example, walking of the so-called simplest walker on flat terrain can be achieved by emulating a slanted artificial gravity field via robot actuators (Asano and Yamakita (2001)). This is a very special case of a potential field.

The design and parameterization of more generic potential fields remains challenging, particularly for systems that exhibit modeling uncertainties or are subjected to unknown disturbances. Reinforcement learning (RL) is a powerful technology to derive controllers for systems where no models are available. Policy search RL methods, also known as actor-only methods, have been found effective for robotic applications due to their ability to handle higher dimensionality and continuous state and action spaces compared to Value-based RL methods (Kober et al. (2013)). Furthermore, policy search methods have been effectively implemented on bipedal robots (Tedrake et al. (2004)).

In this work, we propose to combine RL and PF-constrained impedance control to improve robot safety for robots that operate in uncertain conditions because:

- PF-constraint provides safety with a known energy bound
- RL provides controllers for systems with modeling uncertainty.

The question arises, can policy search RL be combined with potential fields to achieve LC walking? While the theoretical advantage of a PF-constrained impedance control, specifically energy boundedness, are presented in literature, the sub-question arises, are there limitations when it comes to RL convergence?

As a first step towards answering these questions, this paper presents a methodology for defining a potential field-constrained (PF-constrained) impedance control and improving it via reinforcement learning. To achieve this, we define an impedance control as a parameterized mapping of configurations to control torques, which is analogous to a policy in Reinforcement Learning (RL) algorithms. A PF-constrained and an unconstrained parameterization of an impedance controller are compared before and after RL applied to the bipedal walking problem. These control methods are compared for three cases: the reference case of the simplest walking model (SWM), the slope-modified case of the SWM on flat ground, and the mass-modified case, of the SWM with modified foot mass on flat ground.

2. IMPEDANCE CONTROL INITIALIZATION

As opposed to conventional set-point control approaches that directly control system variables such as position and force, impedance control attempts to enforce a dynamic relation between these variables (Hogan (1984)). In this section, an impedance controller is derived for a fully actuated robot with n Degrees Of Freedom (DOF) using least squares optimization. We assume an accurate model of the robot as well as the ability to measure the position and torque at each joint as well as full collocated actuation. Each configuration of the robot can be described by a unique vector $\boldsymbol{q} = [q_1, q_2, ..., q_n]^T$ where q_n, with index $i = 1...n$, are the generalized coordinates.

If a desired trajectory, $\boldsymbol{x} = \left(\boldsymbol{q}^T, \dot{\boldsymbol{q}}^T, \ddot{\boldsymbol{q}}^T\right)^T$, is known, the idealistic control torques, $\boldsymbol{\tau}_0$, required to achieve this trajectory can be found using inverse dynamics. A function to approximate the torques applied to the system as a function of the robot's configuration, $\boldsymbol{\tau}(\boldsymbol{q}) \in \mathbb{R}^n$, can be found by solving the least squares problem

$$\min_{\boldsymbol{w}} \sum_{k=1}^{S} \|\boldsymbol{\tau}_{0,k}(\boldsymbol{x}_k) - \boldsymbol{\tau}(\boldsymbol{q}_k; \boldsymbol{w})\|^2 \qquad (1)$$

where $\boldsymbol{\tau}_{0,k}(\boldsymbol{x}_k)$ is a set of training data with S samples.

For the unconstrained case with n degrees of freedom, the vector function $\boldsymbol{\tau}(\boldsymbol{q}; \boldsymbol{w})$ is defined in terms of its components $\tau_i(\boldsymbol{q}; \boldsymbol{w}_i) = \boldsymbol{g}_i(\boldsymbol{q})^T \boldsymbol{w}_i$, each of which is approximated by a set of normalized radial basis functions (RBF) $\boldsymbol{g}_i(\boldsymbol{q})$ and corresponding weights $\boldsymbol{w}_i$, $i = 1...n$.

For the constrained case, function $\boldsymbol{\tau}(\boldsymbol{q})$ is restricted to describe a potential field by enforcing that its work is zero for any closed-path trajectory. This implies the control torques are a function of the joint variables $\boldsymbol{q}$ and are defined as the negative gradient of a potential function $\psi(\boldsymbol{q}; \boldsymbol{w}) = \boldsymbol{g}(\boldsymbol{q})^T \boldsymbol{w}$ with respect to $\boldsymbol{q}$:

$$\boldsymbol{\tau}(\boldsymbol{q}; \boldsymbol{w}) = -\nabla_q \psi(\boldsymbol{q}; \boldsymbol{w}) = -\left(\frac{\partial \boldsymbol{g}(\boldsymbol{q})}{\partial \boldsymbol{q}}\right)^T \boldsymbol{w}.$$

This is similar to the method of Generalized Elasticities presented in Vallery et al. (2009a) and Vallery et al. (2009b). For the RBF $\boldsymbol{g}(\boldsymbol{q})$ we choose to use compactly supported radial basis functions which allow for the use of a minimal number of center points in the neighborhood of the robot's position to sufficiently compute the function value. This reduces the computational resources needed during operation.

3. POLICY SEARCH REINFORCEMENT LEARNING

Reinforcement Learning (RL) is a machine learning method which attempts to find a control policy, $\pi(\boldsymbol{u}|\boldsymbol{x}, \boldsymbol{w})$, which maps states $\boldsymbol{x}$ to actions $\boldsymbol{u}$. For policy search algorithms, the policy is parameterized by a weighting vector $\boldsymbol{w}$. The policy is analogous to the impedance control laws derived in the previous section where generalized coordinates $\boldsymbol{q}$ are states and control torques $\boldsymbol{\tau}$ are actions.

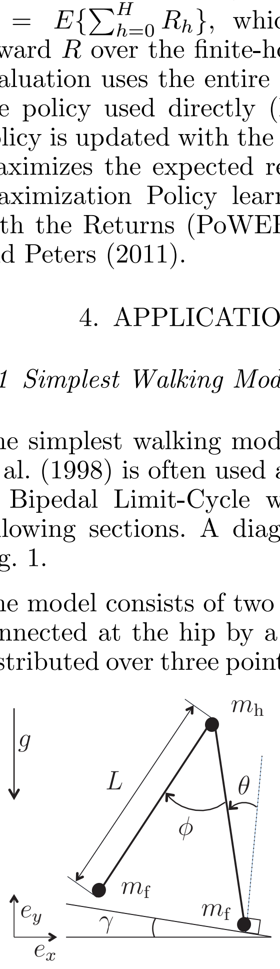
The policy space is explored by randomly perturbing the weighting vector $\boldsymbol{w}$. Batch exploration is performed where the policy is independently perturbed from the initial policy a set number of times. The perturbed policies are then evaluated by computing the expected return $J = E\{\sum_{h=0}^{H} R_h\}$, which is a sum of the expected reward R over the finite-horizon H. Episode-based policy evaluation uses the entire episode to assess the quality of the policy used directly (Deisenroth et al. (2011)). The policy is updated with the objective to find a policy which maximizes the expected return. We use the Expectation Maximization Policy learning by Weighted Exploration with the Returns (PoWER) method developed in Kober and Peters (2011).

4. APPLICATION TO LC WALKING

4.1 Simplest Walking Model

The simplest walking model (SWM) developed in Garcia et al. (1998) is often used as a tool to study the paradigm of Bipedal Limit-Cycle walking and is detailed in the following sections. A diagram of the SWM is shown in Fig. 1.

The model consists of two massless rigid links of length L connected at the hip by a frictionless hinge. The mass is distributed over three point masses at the hip and feet such



	"R"	"S"	"M"
g ($\mathrm{m/s^2}$)	10.000	10.000	10.000
L (m)	1.000	1.000	1.000
m_h (kg)	1.000	1.000	1.000
m_f (kg)	0.001	0.001	0.010
γ (rad)	0.004	0.000	0.000

Fig. 1. Diagram of the Simplest Walking Model (SWM) and its parameters for Reference ("R"), Slope-modified ("S") and Mass-modified ("M") cases.

that the hip mass m_{h} is much larger than the foot mass m_{f}. The model is situated on a slope of angle γ and acts only under the force of gravity with acceleration constant g. The configuration of the model is given by the ankle angle θ and hip angle ϕ. The generalized coordinates are $\boldsymbol{q} = (x_{\mathrm{c}}, y_{\mathrm{c}}, \theta, \phi)^T$ where the subscripts "c" denotes the contact point of the stance foot with the ground.

The training data was found by first, scanning the initial conditions $(\boldsymbol{q}, \dot{\boldsymbol{q}})$ for cases in which the SWM converges to an LC and then the associated accelerations $\ddot{\boldsymbol{q}}$ were found using inverse dynamics. The ankle angle was varied between 0.1 and 0.2 rad with a step size of 0.005 rad, and the initial hip angle was set to twice that of the ankle so the model initializes in double support phase. The initial ankle angular velocity was varied between -0.68 and $-0.38\,\mathrm{rad\,s^{-1}}$ with a step size of $0.005\,\mathrm{rad\,s^{-1}}$, and the initial hip angular velocity was set to $0\,\mathrm{rad\,s^{-1}}$. The torques $\boldsymbol{\tau}_0$ found from training data can be used to solve the least-squares problem in (1) using the recursive least-squares method described in Sec. 2 resulting in impedance control laws of the form $\boldsymbol{\tau}(\boldsymbol{q}; \boldsymbol{w})$.

4.2 Reinforcement Learning

The resulting impedance control laws $\boldsymbol{\tau}(\boldsymbol{q}; \boldsymbol{w})$ parameterized by vector $\boldsymbol{w}$ are specific to the simplest walking model case and will likely not be effective if the model is modified or more degrees of freedom are added. If this is the case $\boldsymbol{\tau}(\boldsymbol{q}; \boldsymbol{w}_0)$ parameterized by vector $\boldsymbol{w}_0$ can be used as the initial policy for policy search RL. The policy search with episode-based evaluation strategy described in Sec. 3 can be used where one episode is H steps of the biped. For a biped robot, the state transitions from the previous state $\boldsymbol{x}$ to the next state $\boldsymbol{x}'$ caused by actions $\boldsymbol{u}$ can be modeled by solving the equations of motion using iterative methods where $\boldsymbol{x} = \left(\boldsymbol{q}^T,\ \dot{\boldsymbol{q}}^T\right)^T$ are the states and the generalized forces Q_θ, Q_ϕ are the actions $\boldsymbol{u}$.

The reward function used for each step is

$$\begin{aligned} R_h(\boldsymbol{x}, \boldsymbol{u}) = & R_{\mathrm{step}} - R_\Delta ||\Delta\theta|| - R_{\dot{\Delta}} ||\Delta\dot{\theta}|| \\ & - R_t ||t_h - t_0|| - R_{\tau,\theta} ||\boldsymbol{\tau}_\theta|| - R_{\tau,\phi} ||\boldsymbol{\tau}_\phi|| \end{aligned}$$

where $\Delta\theta = \theta_h - \theta_{h-1}$, and $\Delta\dot{\theta} = \dot{\theta}_h - \dot{\theta}_{h-1}$, and $R_{\mathrm{step}} = 1$, $R_\Delta = 10\,\mathrm{rad^{-1}}$, $R_{\dot{\Delta}} = 10\,\mathrm{s\,rad^{-1}}$, $R_t = 1\,\mathrm{s^{-1}}$, $R_{\tau,\theta} = 10\,\mathrm{N^{-1}\,m^{-1}}$ and $R_{\tau,\phi} = 100\,\mathrm{N^{-1}\,m^{-1}}$ are constants. The first term of the reward function is given as a reward for successfully completing a step. The second term penalizes the change in angle and angular velocity of the stance leg at the beginning of each step. This is to encourage a limit-cycle is reached where each step is the same. The third term penalized the change in time of step h from the time of the reference LC step $t_0 = 1.2180\,\mathrm{s}$. The fourth term penalizes the magnitude of the control torques to minimize the energy added to the system.

5. EVALUATION PROTOCOL

5.1 Implementation

The impedance control laws were implemented on a fully-actuated simple walking model for the three cases: the reference case on a slope, the slope-modified case on flat ground, and the mass-modified case, of the SWM with modified foot mass on flat ground, cf. Fig. 1.

For the least squares optimization, 50 RBFs were used. The center locations were determined using a grid step size of 0.05 rad for the ankle angle and 0.1 rad for the hip angle in the area of the ideal trajectory of the SWM.

For the policy search RL, a horizon of $H = 10$ was used corresponding to 10 steps of the robot. For the exploration strategy, a batch size of 100 iterations was used. A Gaussian exploration $\epsilon \sim \mathcal{N}(0, \sigma^2)$ was used which was decreasing linearly over episodes.

5.2 Experiment Setup

Initial unconstrained and PF-constrained impedance controllers were found using inverse dynamics for each of the three cases described above. For the Slope and Mass-modified cases, RL was used to attempt to improve the policy for both the unconstrained and the PF-constrained parameterizations. For the reference case, the performance of the controllers can not be improved further using RL based on the evaluation strategy since the control torques cannot decrease further.

5.3 Benchmarking Criteria

The unconstrained and PF-constrained impedance controllers were compared for each of the three cases based on the following benchmarking criteria.

Work and Energy: The energy of the LC of the ideal SWM (unactuated and on a slope) is bounded by the potential field of gravity. The energy bound can be measured as the maximum energy, E of the LC, defined $E = V + T$ where V is the potential energy and T is the kinetic energy. For the LC of the ideal SWM, the total energy is constant at 10.0108 J. At each step kinetic energy is dissipated at impact and an equivalent amount of potential energy is added by the slope. The energy added/dissipated at each step is equivalent to 0.0166 J.

Energy consumption can be measured for the actuated model as the work done by the actuators $W = \int_{\boldsymbol{q}_0}^{\boldsymbol{q}_1} \boldsymbol{\tau} \mathrm{d}\boldsymbol{\theta}$, where $\boldsymbol{q}_0$ is the configuration at the beginning of the step and $\boldsymbol{q}_1$ is the configuration at the end of the step.

Robustness: The robustness of an LC gait can be measured by its velocity disturbance rejection. An angular velocity disturbance is introduced to the stance leg at the beginning of the first step and the maximum disturbance that can be applied without causing the walker to fall is used as a measure for robustness.

RL Performance: The performance of the RL is assessed by plotting the mean performance over the episodes, for several trials, and observing how many episodes it takes to level off.

6. RESULTS

6.1 Reference case

The trajectories for the Unconstrained and PF-constrained policies were derived using inverse dynamics. The bench-

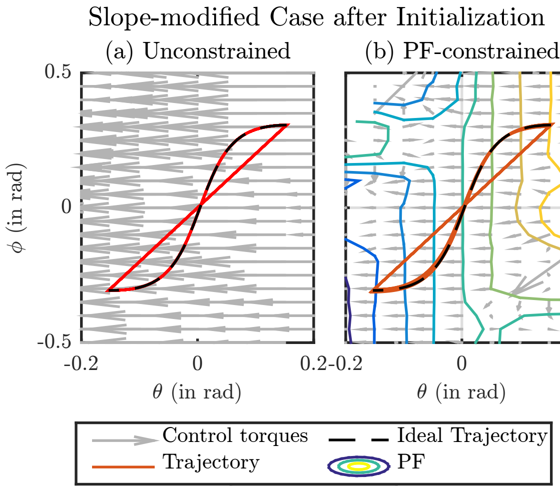


Fig. 2. Trajectory phase plot of the initial (a) Unconstrained and (b) PF-constrained policies for the Slope-modified Case.

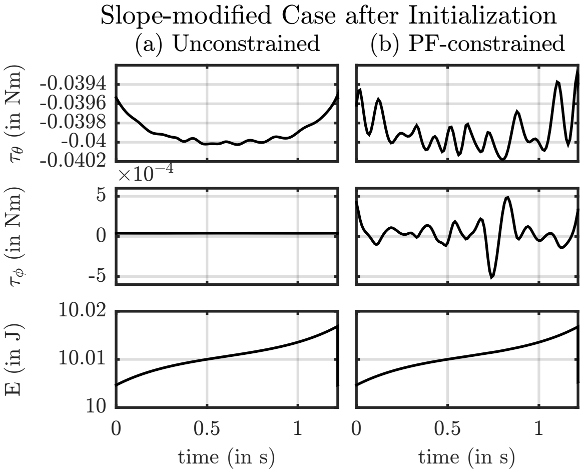


Fig. 3. Control torques and energy of one LC step of the initial (a) Unconstrained and (b) PF-constrained policies for the Slope-modified Case.

marking criteria for the energy, work and robustness of the reference case are specified in Table 1.

6.2 Slope-modified Case

Initialization The trajectory phase plots for the initial Unconstrained and PF-constrained policies for the Slope-modified case are shown in Fig. 2 (a) and (b) respectively. The control torques and total energy for the initial Unconstrained and PF-constrained policies for the Slope-modified case are shown in Fig. 3 (a) and (b) respectively.

Reinforcement Learning The mean performance of the RL for the Unconstrained and PF-constrained controllers are shown in Fig. 4. The resulting trajectory phase plot for the learned Unconstrained and PF-constrained policies for the Slope-modified case are shown in Fig. 5 (a) and (b) respectively. The resulting control torques and energy for the learned Unconstrained and PF-constrained policies for the Slope-modified case are shown in Fig. 6 (a) and (b) respectively.

The benchmarking criteria for the energy, work and robustness of the Slope-modified case are specified in Table 1.

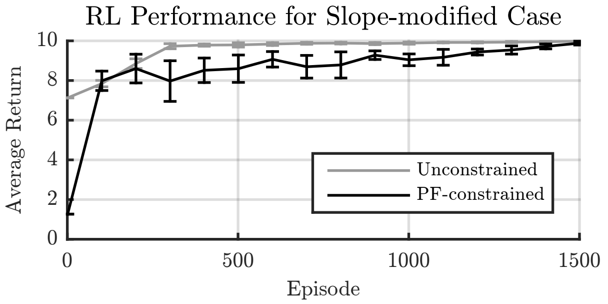


Fig. 4. Mean performance of the RL for the Unconstrained and PF-constrained policies for the Slope-modified case averaged over 10 runs with the error bars indicating the standard deviation. For both policies the exploration variance decreased linearly from 1e-6 to 1e-11 throughout the episodes.

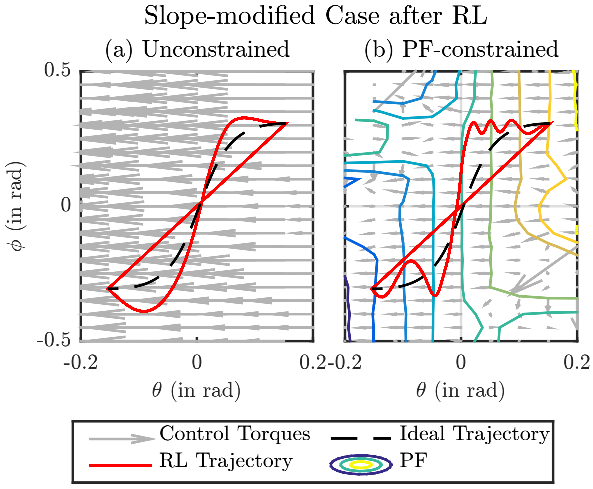


Fig. 5. Trajectory phase plot of the learned (a) Unconstrained and (b) PF-constrained policies for the Slope-modified Case.

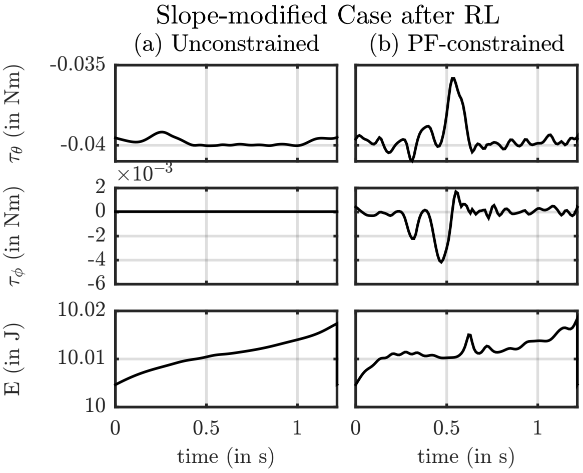


Fig. 6. Control torques and energy of one LC step of the learned (a) Unconstrained and (b) PF-constrained policies for the Slope-modified Case

6.3 Mass-modified Case

Initialization For the Mass-modified case neither the initial Unconstrained nor initial PF-constrained policy leads to a stable limit cycle so the corresponding plots are not shown.

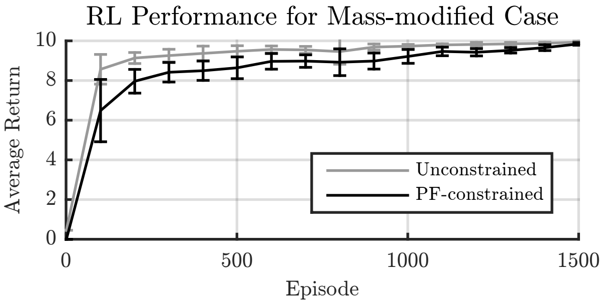


Fig. 7. RL mean performance of the Unconstrained and PF-constrained policies for the Mass-modified case averaged over 10 runs with the error bars indicating the standard deviation. For the Unconstrained policy the exploration variance decreased from 1e-5 to 1e-10 and for the PF-constrained policy the variance

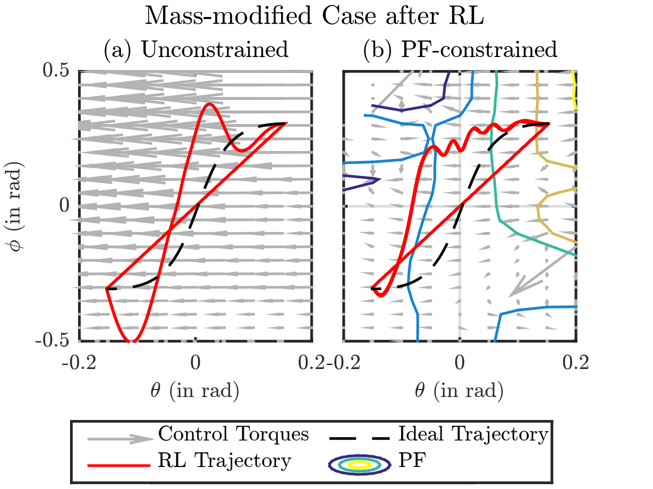


Fig. 8. Trajectory phase plot of the learned (a) Unconstrained and (b) PF-constrained policies for the Mass-modified Case.

Reinforcement Learning The mean performance of the RL for both the PF-constrained and unconstrained case are shown in Fig. 7. The resulting trajectory phase plots for the learned Unconstrained and PF-constrained policies for the Mass-modified case are shown in Fig. 8 (a) and (b) respectively. The resulting control torques and energy for the learned Unconstrained and PF-constrained policies for the Mass-modified case are shown in Fig. 9 (a) and (b) respectively.

The benchmarking criteria for the work, energy and robustness of the Mass-modified case are specified in Table 1.

7. DISCUSSION

For the reference case for both the unconstrained and PF-constrained parameterization the controlled trajectory perfectly follows the ideal trajectory. No actuator torques are generated and the total energy is equal to 10.011 J. It can be seen in Table 1 that both controllers have the same energy bound and maximum disturbance rejection as the unactuated ideal case. This serves as a validation for both the impedance controllers derived using inverse dynamics and least squares optimization.

For the slope-modified case, the initial impedance controllers, for both PF-constrained and unconstrained parameterization, allow the biped to achieve an LC gait on a flat surface ($\gamma = 0\,\text{rad}$) as can be seen in the trajectory phase plots in Fig. 2. It can be seen in Table 1 that the velocity disturbance rejections are comparable to the ideal SWM, however, the energy bound is higher than the ideal case for both controllers. The work done by the actuators is similar for both controllers, however, it is almost 100 times the work done by gravity in the ideal case.

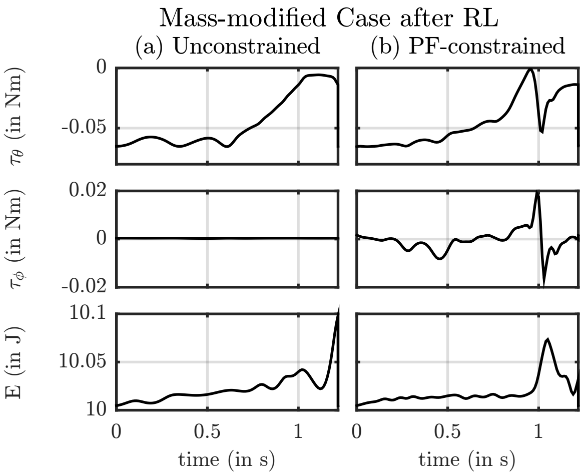


Fig. 9. Control torques and energy of one LC step of the learned (a) Unconstrained and (b) PF-constrained policies for the Mass-modified Case.

Table 1. Summary of Results. In the table “EB” stands for “Energy bound”, and “MVD” stands for “Maximum velocity disturbance”.

Case	Parameterization	Benchmarking Criteria	Initial Policy	Learned Policy
Reference	Unconstrained	EB (J)	10.011	-
		Work (J)	0.000	-
		MVD (rad/s)	-0.050	-
	PF-constrained	EB (J)	10.011	-
		Work (J)	0.000	-
		MVD (rad/s)	-0.050	-
Slope-modified	Unconstrained	EB (J)	10.017	10.026
		Work (J)	1.507	1.488
		MVD (rad/s)	-0.050	-0.030
	PF-constrained	EB (J)	10.019	10.019
		Work (J)	1.495	1.332
		MVD (rad/s)	-0.060	0.000
Mass-modified	Unconstrained	EB (J)	✗	10.215
		Work (J)	✗	1.311
		MVD (rad/s)	✗	-0.050
	PF-constrained	EB (J)	✗	10.062
		Work (J)	✗	1.481
		MVD (rad/s)	✗	-0.020

As can be seen in Table 1, RL of the initial impedance controllers for the slope-modified case increases the energy bound for both controllers, while decreasing the work done by the actuators. RL also leads to decreased disturbance rejection. As can be seen in Fig. 4, the performance of the unconstrained parameterization levels off before the PF-constrained parameterization, indicating the unconstrained parameterization achieves a higher performance with less episodes compared to the PF-constrained parameterization.

For the mass-modified case, the initial impedance controllers, for both PF-constrained and unconstrained parameterizations, do not allow the biped to achieve an LC gait. The impedance controllers derived from inverse dynamics appear not to be able to compensate for the modified dynamics of the model. However, RL of these initial policies allows the biped to achieve an LC gait as shown in Fig. 8. This validates the use of RL for achieving an LC gait. As can be seen in Table 1, for both controllers the energy bound and work done is greater than the ideal case. While the robustness of the unconstrained controller is comparable to the ideal case, it is reduced for the PF-constrained controller. As can be seen in Fig. 7, the performance of the unconstrained parameterization levels off before the PF-constrained parameterization.

For all cases, the energy bound and work done by the actuators was similar for both the PF-constrained and unconstrained controllers. As the implementation of the RL did not converge to a single optimal solution, the variance in the resulting energy and work was too large to draw an accurate comparison. For all cases, there are no improvements to the robustness of the limit-cycle against velocity disturbances. The reason for this is that the episode (consisting of H steps of the limit-cycle) is a black-box from the perspective of the episode-based RL. Learning is based only on the inputs and outputs of the episode, therefore any unknown disturbances throughout the episode are not accounted for, and consequently the robustness is not improved by the RL. Exploring and learning throughout the episode may be one way to improve the robustness. Additionally, learning could take place in an unknown environment with unknown disturbances.

The scope of these results is limited by the variables of the simple walking model used. The only modifications tested were the ratio of the hip mass to foot mass, and the slope γ. An interesting observation is the learned behavior of “swing-leg retraction” seen in the learned policy for both cases, as shown in Fig. 5 and 8 . This is when the swing leg retracts at the end of a step until it hits the ground. It has been shown in Hobbelen and Wisse (2008) that swing-leg retraction can improve disturbance rejection.

8. CONCLUSION AND FUTURE WORK

In this work we successfully combined potential field control and reinforcement learning to achieve limit-cycle walking for a simple walking model. A limit-cycle was achieved on flat ground, and for a modified hip to foot mass ratio. The results demonstrate that a potential field controller can not only “emulate” the effect of gravity on the simple walking model, but also improve its performance if reinforcement learning is applied. The potential field-constrained controller provides safety by bounding the energy while performing equally well compared to an unconstrained controller. The performance of the RL leveled off faster for the unconstrained case.

Achieving a limit cycle gait on a SWM is trivial compared to more complex models. In future work the method presented in this paper could be applied to higher degree of freedom models. A strength of this method is the ability to bound the energy of the controlled system. In future work it could be explored how to enforce a desired energy bound. Improved tuning of the RL exploration and evaluation strategy could lead to improved policies and more conclusive results for the comparison of the unconstrained and PF-constrained parameterizations. More advanced RL methods could lead to potential fields that further improve performance and even increase robustness.

ACKNOWLEDGEMENTS

I. Koryakovskiy and H. Vallery were supported by the European project KOROIBOT FP7-ICT-2013-10/611909.

REFERENCES

Asano, F. and Yamakita, M. (2001). Virtual gravity and coupling control for robotic gait synthesis. *IEEE Trans. Systems, Man, and Cybernetics Part A: Systems and Humans*, 31(6), 737–745.

Deisenroth, M.P., Neumann, G., and Peters, J. (2011). A Survey on Policy Search for Robotics. *Foundations and Trends in Robotics*, 2, 1–142.

Garcia, M., Chatterjee, A., Ruina, A., and Coleman, M. (1998). The Simplest Walking Model: Stability, Complexity, and Scaling. *Journal of Biomechanical Engineering*, 120(2), 281–288.

Hobbelen, D.G.E. and Wisse, M. (2007). Limit Cycle Walking. *Humanoid Robots: Human-like Machines*, 642–659.

Hobbelen, D.G.E. and Wisse, M. (2008). Swing-leg retraction for limit cycle walkers improves disturbance rejection. *Trans. Robotics*, 24(2), 377–389.

Hogan, N. (1984). Impedance control: An approach to manipulation. In *American Control Conf.*.

Hyon, S.H. and Cheng, G. (2006). Passivity-based full-body force control for humanoids and application to dynamic balancing and locomotion. In *Int. Conf. Intelligent Robots and Systems*.

Kober, J., Bagnell, J.A., and Peters, J. (2013). Reinforcement learning in robotics: A survey. *Int. Journal of Robotics Research*, 32, 1238–1274.

Kober, J. and Peters, J. (2011). Policy search for motor primitives in robotics. *Machine Learning*, 84(1-2), 171–203.

Koditschek, D.E. (1987). Exact robot navigation by means of potential functions: Some topological considerations. In *Int. Conf. Robotics and Automation*.

McGeer, T. (1990). Passive Dynamic Walking. *Int. Journal of Robotics Research*, 9(2), 62–82.

Papageorgiou, M. (2012). *Optimierung: statische, dynamische, stochastische Verfahren*. Springer-Verlag.

Tedrake, R., Zhang, T., and Seung, H. (2004). Stochastic policy gradient reinforcement learning on a simple 3D biped. In *Int. Conf. Intelligent Robots and Systems*.

Vallery, H., Duschau-Wicke, A., and Riener, R. (2009a). Generalized elasticities improve patient-cooperative control of rehabilitation robots. In *Int. Conf. on Rehabilitation Robotics*.

Vallery, H., Duschau-Wicke, A., and Riener, R. (2009b). Optimized passive dynamics improve transparency of haptic devices. In *Int. Conf. Robotics and Automation*.