Behavioral Development of Ball Kicking Motion of a Two-wheeled Inverted Pendulum Mobile Robot

WCCI 2010 IEEE World Congress on Computational Intelligence
July, 18-23, 2010 - CCIB, Barcelona, Spain FUZZ-IEEE

Behavioral Development of Ball Kicking Motion of a Two-wheeled Inverted Pendulum Mobile Robot

Yasutake Takahashi, Hiroaki Nonoshita, Takayuki Nakamura, and Yoichiro Maeda

Abstract— **In this paper, we introduce a method for generating a dynamic motion such that a two-wheeled inverted pendulum robot kicks a ball far away utilizing its own body dynamics while it keeps standing. Such a dynamic motion can be acquired through trial and error based on a reinforcement learning scheme. We utilize a simple policy gradient method to acquire a kicking motion which is designed by defining the desired parameters such as body angle, wheel angular velocity and so on. To show the validity of our approach, we perform computer simulation experiments of behavior acquisition for the two-wheeled inverted pendulum robot. Based our approach, we succeeded in acquiring the kicking motion of the two-wheeled inverted pendulum robot. A very interesting finding is: each of the acquired motions deviates from the desired trajectory, which is given by the human designer while keeping evaluation value of the acquired motion high.**

I. Introduction

Many kinds of mobile robots working in a human living environment have been developed. Most of the existing mobile robots have three-wheeled or four-wheeled locomotion mechanism and low center of gravity for achieving static stability. Furthermore, such mobile robots tend to have a larger support surface and lower height than human's ones. As a result, the motion of such mobile robot becomes slow and any motion with a rapid acceleration and deceleration is impossible for such mobile robot. It is considered that such three-wheeled or four-wheeled mobile robot is not suitable as human partners robot.

Recently, many kinds of humanoid robots have been developed so far as human partners under the human-symbiotic environment, because it is reasonable for the mobile robot to have a similar shape and size of a human. However, a two-legged humanoid robot[1] still has a lot of difficulties. For example, it has problems of robustness against unpredictable disturbances, heavy weight, and dynamic motions. A pneumatic actuated biped robot[2][3] has a lightweight body and a potential to generate dynamic motions but still has big problems in terms of the stability for usage under the human living environment.

On the other hand, a two-wheeled inverted pendulum robot[4][5] has many advantages over the statically stable wheeled robots and any other biped robots. It requires a smaller amount of space to stand and stay upright than the other wheeled robots and smaller number of actuators than the conventional biped robots. Furthermore, it can easily generate more dynamic motions only by its body balancing control. The two-wheeled inverted pendulum robot has a possibility to be standard platform as a human partner robot. Most of researches regarding to the two-wheeled inverted pendulum robot focuses on the control theory for achieving stable navigation. There are few researches to realize more dynamic motions such as throwing, jumping, pushing, kicking, and so on.

For example, the two-wheeled inverted pendulum robot needs to break static stability while swinging the body back and forth, when it kicks a ball as fast as possible. This means that the robot has to utilize its own body dynamics as much as possible in order to realize such dynamic motions in the same way that a human does. Of course, it should achieve global stability so that it avoids falling down during the dynamic motion. The dynamic motion requires not only the conventional control theory to stabilize body balance but also a good trajectory of body posture and traveling. It is not trivial to find such good trajectory to generate the dynamic motion by hand.

In this paper, we introduce a method for generating a dynamic motion such that a two-wheeled inverted pendulum robot kicks a ball far away utilizing its own body dynamics while it keeps standing. Our method assumes that a kicking motion consists of two primitive motions. In this work, we utilize a simple policy gradient method[6] to acquire the two primitive motions each of which is designed by defining the desired parameters such as body angle, wheel angular velocity and so on. The kicking motions are evaluated with the ball velocity after the robot kicks and its stableness. Our policy gradient method just updates parameters of the kicking motions so that the evaluations become as higher as possible.

To show the validity of our approach, we perform computer simulation experiments of behavior acquisition for the two-wheeled inverted pendulum robot. As a result of these learning process, we confirmed that a stepwise learning curve with two learning phases emerges. It means that the robot changes a learning strategy between the phases during the learning process. Based our approach, we succeeded in acquiring the kicking motion of the two-wheeled inverted pendulum robot. A very interesting finding is: each of the acquired motions deviates from the desired trajectory which is given by the human designer while keeping evaluation value of the acquired motion high. Based on the experiment results, we discuss the learning profile in detail.

Yasutake Takahashi, Hiroaki Nonoshita, and Yoichiro Maeda are with the Department of Human and Artificial Intelligent Systems, Graduate School of Engineering, University of Fukui, 3-9-1, Bunkyo, Fukui, Fukui, 910-8507, Japan (phone: +81 776 27 8540; email: {yasutake,hnonosita,maeda}@ir.his.u-fukui.ac.jp).

Takayuki Nakamura is with Faculty of Systems Engineering, Wakayama University, Sakaetani 930, Wakayama 640-8510, Japan (phone: +81 73 457 8071; email: ntakayuk@sys.wakayama-u.ac.jp).

978-1-4244-8126-2/10/$26.00

II. Two-wheeled Inverted Pendulum Robot

Fig. 1. Two-wheeled inverted pendulum robot with a ball

Fig. 1 shows a two-wheeled inverted pendulum robot which we designed and built and a ball. The robot body size is 20cm length, 30.5cm width, and 35cm height. The weight is 8kg including the body, wheels, batteries, motors, gears, encoders, an accelero-gyrometer unit, and control unit. The wheel radius is 8.5 cm and its weight is 500g each. The ball kicked by the robot is a soccer ball, roughly 22cm in diameter, and about 450g in weight. In order to perform comprehensive experiments for learning a kicking motion with the robot, we develop a computer simulation of the robot and the ball using the ODE(open dynamics engine) library.

III. Design for Learning Kicking Motion

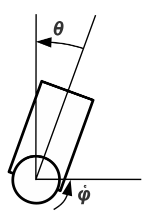


Fig. 2. Model of the two-wheeled inverted pendulum robot

The kicking motion is designed with two control layers, a low-level posture controller and a posture generator. The posture controller follows a conventional torque control theory. Torque for the wheels T is calculated as follows:

$$T = -k_1(\theta - \theta_d) - k_2\dot{\theta} - k_3\dot{\varphi} - k_4 \sum(\dot{\varphi} - \dot{\varphi}_d), \qquad (1)$$

where θ, θ_d, $\dot{\varphi}$, $\dot{\varphi}_d$ are body angle, desired body angle, wheel angular velocity, and desired wheel angular velocity, respectively. k_1, k_2, k_3, k_4 are gains for the body angle, body angular velocity, wheel angular velocity, and accumulated error of wheel angular velocity, respectively.

In this work, we design the kicking motion in combination with two primitive motions (Fig.3). Each primitive motion defines the desired body angle, desired wheel angular velocity, gain parameters, and a period of time. The posture generator sends the control parameters to the posture controller within the period defined by the primitive motion, one by one. The first primitive motion tries to lean forward to kick the ball and the second tries to make the robot kick the ball actually. Before and after the kicking motion, the robot controls itself to stay upright avoiding falling-down, autonomously.

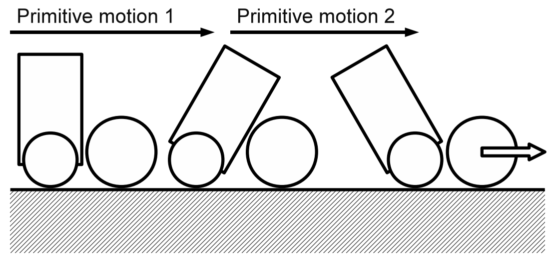


Fig. 3. Two primitive motions of Kicking motion

IV. Policy Gradient Method

In order to generate kicking motion of our robot, our method utilizes a simple policy gradient method introduced by Kohl and Stone[6]. The posture generator learns the control parameters based on the policy gradient method. A parameter set of the current motion is $\mathbf{\Theta}$. The learning system prepares T similar policies $R^1, R^2, ..., R^T$ by adding small disturbances ε_j, 0, or $-\varepsilon_j$ to the current motion parameter $\Theta_j(\Theta_j \in \mathbf{\Theta})$:

$$\Theta_j^i = \Theta_j + r\varepsilon_j \text{ where } r \in (-1, 0, 1) \qquad (2)$$

It evaluates the policies $R^i = \{\Theta_j^i\}$ one by one after the robot tries to generate the kicking motions based on the policies.

First of all, evaluation averages for the disturbances are estimated as follows:

- $Avg_{+\varepsilon,j}$ is average of evaluation of policies which parameter Θ_j^i is $\Theta_j + \varepsilon_j$
- $Avg_{0,j}$ is average of evaluation of policies which parameter Θ_j^i is Θ_j
- $Avg_{-\varepsilon,j}$ is average of evaluation of policies which parameter Θ_j^i is $\Theta_j - \varepsilon_j$

Then, A_j, that is, the gradient of evaluation to the policy parameters Θ_j is estimated approximately as follows: A_j is regarded as 0 if the $Avg_{0,j}$ is greater than $Avg_{+\varepsilon,j}$ and $Avg_{-\varepsilon,j}$. It is regarded as $Avg_{+\varepsilon,j} - Avg_{-\varepsilon,j}$, else:

$$A_j = \begin{cases} 0 & Avg_{+0,j} > Avg_{+\varepsilon,j} \\ & \& Avg_{+0,j} > Avg_{-\varepsilon,j} \\ Avg_{+\varepsilon,j} - Avg_{-\varepsilon,j} & \text{else} \end{cases} \qquad (3)$$

Finally, it updates the policy parameters $\mathbf{\Theta}$ according to the A as follows:

$$\Theta_j \leftarrow \Theta_j + \frac{A_j}{|A_j|} * \eta, \qquad (4)$$

where η is a certain learning step size. The learning system repeats this procedure and updates the motion parameters to reach the local maximum of evaluation. Fig.4 shows a flowchart of the learning procedure.

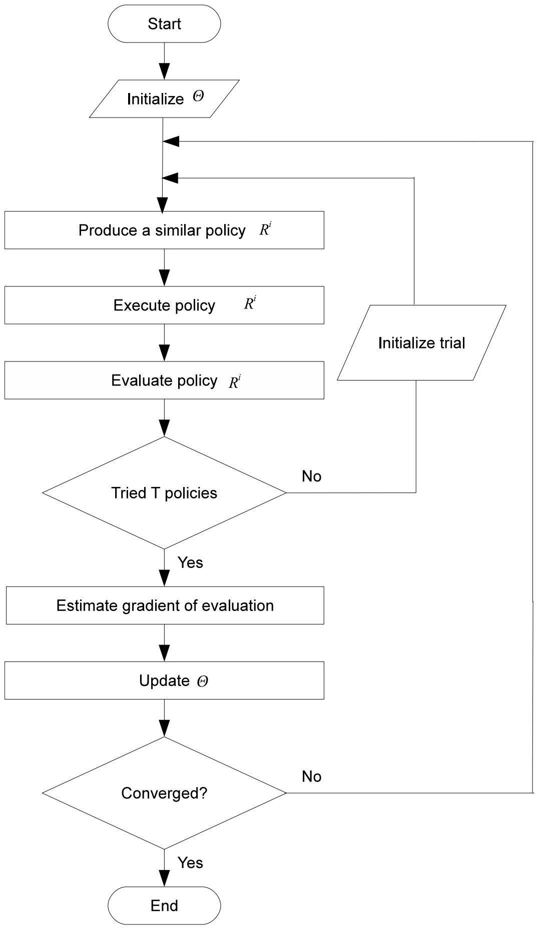


Fig. 4. Flowchart of learning procedure

V. Learning Kicking Motion

A. Experiment Setup

As mentioned in section III, each primitive motion is defined by the desired body angle θ_d, the desired wheel angular velocity $\dot{\varphi}_d$, the gain parameters k_1, k_2, k_3, and k_4, and a period of time t. The learning parameters are those primitive motion parameters for each, therefore, 14 parameters in all. We determine the evaluation criterion of the kicking motion while learning process in consideration of the following points:

- Traveling distance of the robot body should be small.
- Velocity of the kicked ball should be large.
- The robot should avoid falling down by kicking the ball.

The actual evaluation is defined as follows:

$$E = \begin{cases} \dfrac{w_1}{1+l_b} + w_2 v_b - w_f & \text{In case the robot fell down} \\ \\ \dfrac{w_1}{1+l_b} + w_2 v_b & \text{else,} \end{cases} \tag{5}$$

where l_b, v_b, w_1, w_2, and w_3 are the traveling distance of the robot body, the velocity of the kicked ball, weights for the traveling distance, the ball velocity, and cost of falling down, respectively. In this work, w_2 and w_f are fixed to 1000 and 100, respectively. w_1 is set to 1, 10, 100, and 1000 for analysis.

The parameters were initially set as shown in TABLE I. The learning step size parameters are shown in TABLE II.

TABLE I
Initial parameters of primitive motion 1 and 2

learning parameter	primitive motion 1	primitive motion 2
T [sec]	1.0	1.0
θ_d [rad]	-1.0	1.0
ϕ_d [rad/sec]	0	0
k_1	15.0	15.0
k_2	-0.3	-0.3
k_3	0.01	0.01
k_4	0.0	0.0

The number of deviated policies for estimating gradient of the evaluation T is set to 50 in the experiments.

TABLE II
Step size of learning parameters

learning parameter	step size
T [sec]	0.001
θ_d [rad]	0.01
ϕ_d [rad/sec]	0
k_1	0.01
k_2	0.001
k_3	0.0001
k_4	0.00001

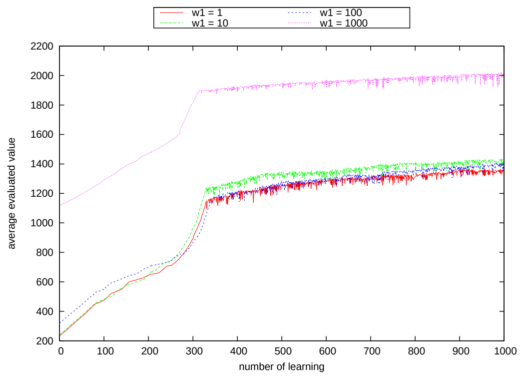
B. Learning Curves

Fig.5 shows the transition profile of evaluation, running distance, and ball velocity while learning kicking motion. All learning curves with varied values of w_1 show similar trends. The evaluations gradually increase from the beginning of the learning to around 250 updates of learning parameters. They increase rapidly from about 250 to 300 updates of parameters and still increases slowly after the 300th update.

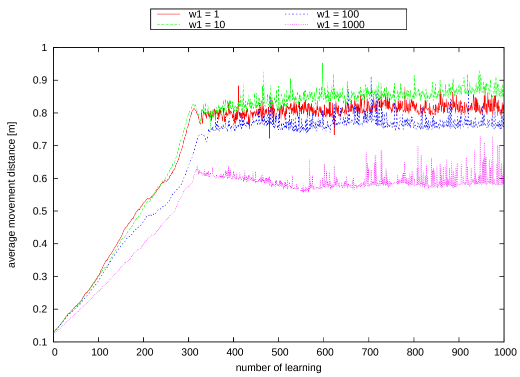
Fig.6 shows the transition profile of posture parameters, that is, desired body angles and periods of time for primitive motion 1 and 2. The periods of time are almost fixed and have constant trends from the beginning to the end of the learning. The desired body angles changes rapidly from the beginning to around the 15000 trials, 300 updates of learning parameters. After that, they increase or degrease very slowly.

Fig.7 shows the transition profile of gain parameters for primitive motion 1 and 2 during the learning. Gains for body angular velocity k_2 and wheel angular velocity k_3 of primitive motion 2 produces interesting learning curves. They decrease before around 300 updates of parameters and increase after that. This means, the robot explores posture parameters of the primitive motions while it relaxes the feedback gains for angular velocities. Once it finds good posture parameters, then, it recover the feedbacks again for the stability. The other curves increase or decrease monotonically and converge after the 300th update of parameters.

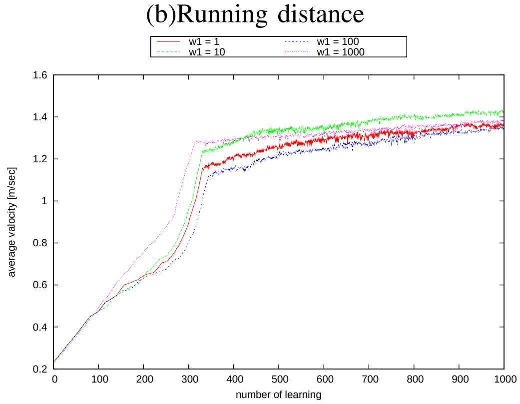
Fig.8 shows trajectories of the body angle while the robot is kicking the ball before the learning, after the learning with 350 and 1000 updates of parameters.



(a)Motion evaluation

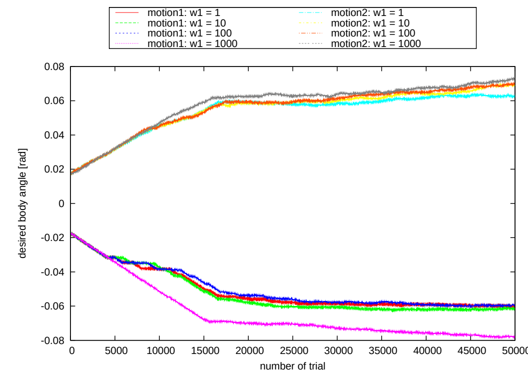


(b)Running distance

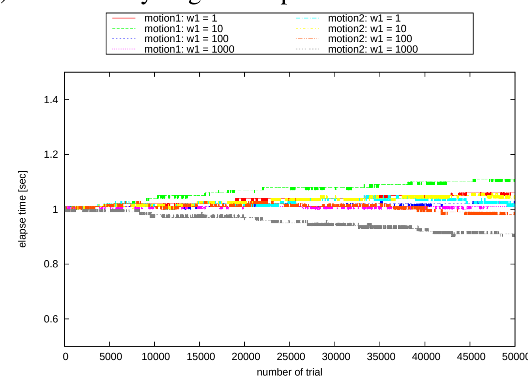


(c)Ball velocity

Fig. 5. Evaluation profiles



(d)Desired body angles for primitive motion 1 and 2



(e)Periods of time for primitive motion 1 and 2

Fig. 6. Profile of posture parameters

It also shows the desired trajectories of the kicking motion. Before the learning, the motion is small and the posture controller follows the desired trajectory. After the learning, the actual trajectory does not follow the desired trajectory well but it does not matter because the objective of the motion is kicking a ball as fast as possible without falling down. The gains for wheel angular velocity deviation k_4 of primitive motion 1 always becomes negative in comparison with Fig.7. The negative gain leads the robot body to lean forward to the ball more than the desired body angle. As a result of adding positive feedback of posture controller, the acquired motion deviates from a stepwise trajectory designed by hand.

VI. Conclusion

In this paper, we introduced a method for generating a dynamic motion such that a two-wheeled inverted pendulum robot kicks a ball far away utilizing its own body dynamics while it keeps standing. We used the simple policy gradient method to acquire the primitive motions of the kicking motion. Based our approach, we succeeded in acquiring the kicking motion of the two-wheeled inverted pendulum robot.

The policy gradient method just updates parameters of the kicking motion so that the evaluation becomes as higher as possible. As a result of these learning process, we confirmed that a stepwise learning curve with two learning phases emerged. From the profile of the learning parameters, we confirmed that the robot decreased feedback gains for angular velocities and updated posture parameters rapidly in the earlier learning phase while it increased the feedback gains once it found good posture parameters in latter learning phase. It means that the robot changes a learning strategy between the phases during the learning process. A very interesting finding is: as a result of adding positive feedback of posture controller, the trajectory of the acquired motions deviated from the desired one given by the human designer while keeping evaluation value of the acquired motion high. In the near future, we will perform real robot experiments.

References

[1] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, “The development of honda humanoid robot,” in *Proceedings of 1998 {IEEE} International*

Conference on Robotics and Automation, vol. 2, 1998, pp. 1321 – 1326.

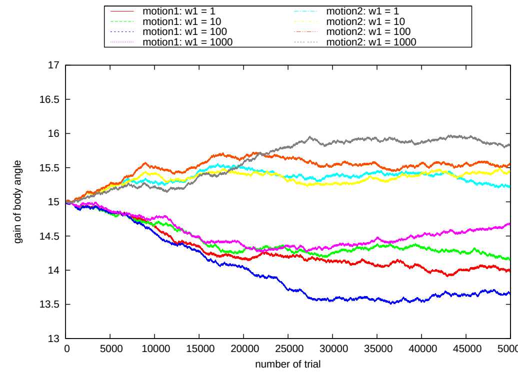
[2] M. Wisse and J. van Frankenhuyzen, "Design and construction of mike and -D autonomous biped based on passive dynamic walking," in *Adaptive Motion of Animals and Machines*, H. Kimura, K. Tsuchiya, A. Ishiguro, and H. Witte, Eds. Springer Tokyo, 2006, ch. 4, pp. 143–154.

[3] K. Narioka and K. Hosoda, "Designing synergistic walking of a whole-body humanoid driven by pneumatic artificial muscles: An empirical study," *Advanced Robotics*, vol. 22, no. 10, pp. 1107–1123, 2008.

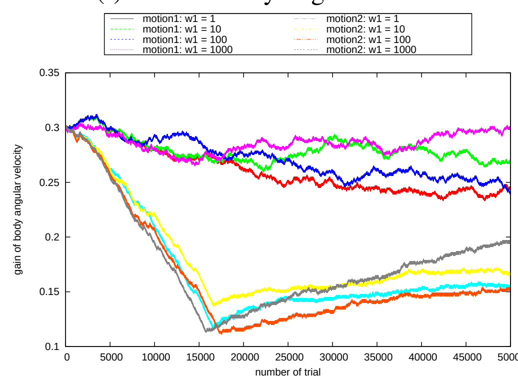
[4] S. Ha, Y. Yuta, "Trajectory tracking control for navigation of self-contained mobileinverse pendulum," in *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems '94*, vol. 3, 1994, pp. 1875–1882.

[5] B. Browning, J. Searock, P. E. Rybski, and M. Veloso, "Turning segways into soccer robots," *Industrial Robot: An International Journal*, vol. 32, no. 2, pp. 149 – 156, 2005.

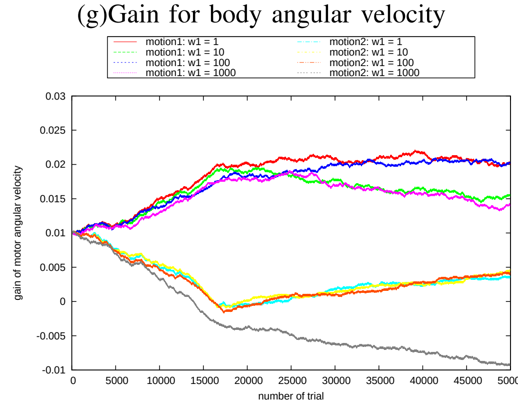
[6] N. Kohl and P. Stone, "Policy gradient reinforcement learning for fast quadrupedal locomotion," in *In Proceedings of the IEEE International Conference on Robotics and Automation*, May 2004, pp. 2619–2624.



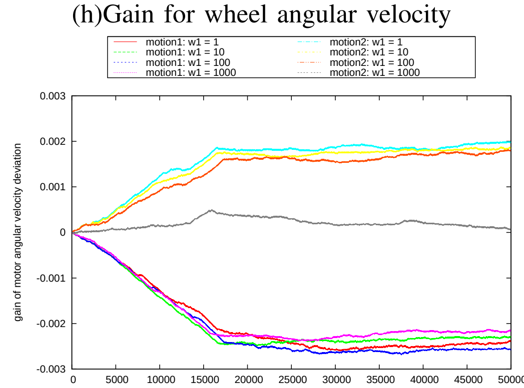
(f)Gain for body angular error



(g)Gain for body angular velocity

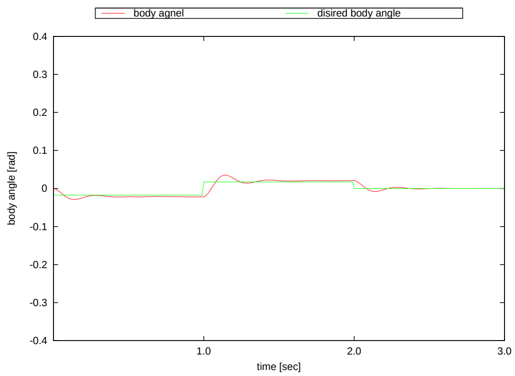


(h)Gain for wheel angular velocity

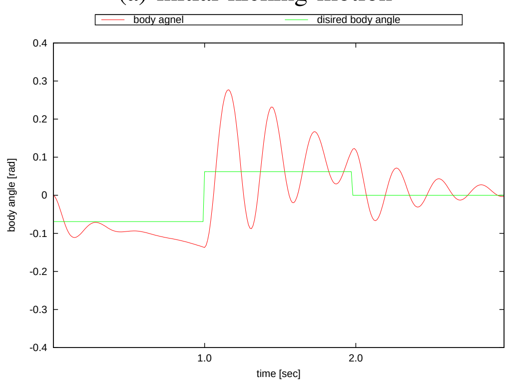


(i)Gain for wheel angular velocity deviation

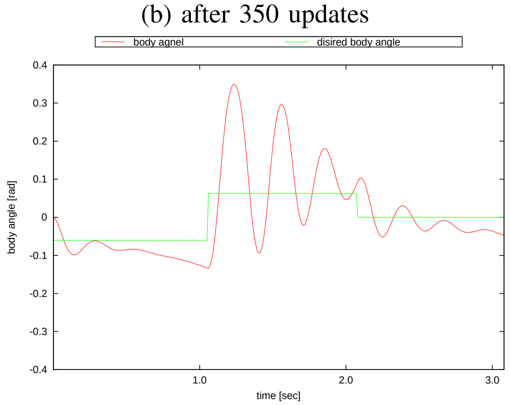
Fig. 7. Profiles of gain parameters for primitive motions



(a) initial kicking motion



(b) after 350 updates



(c) after 1000 updates

Fig. 8. Trajectories of body angle while kicking the ball after updates of learning parameters