ESE 650, SPRING 2023

HOMEWORK 2

ANIRUDH KAILAJE [KAILAJE@SEAS.UPENN.EDU]

Solution 1 (Time spent: 8 hours)**. Policy Iteration**

1. Map and Transition Matrix Creation

The map given as shown below. I created an array of the map size 10x10. I assigned values for

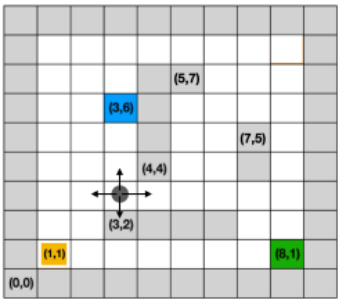

FIGURE 1. Given Map

free cells as -1, obstacles as -10, and the goal cell as 10. The plot of my map is below.
I have created a function to create the transition matrix based on the current policy.

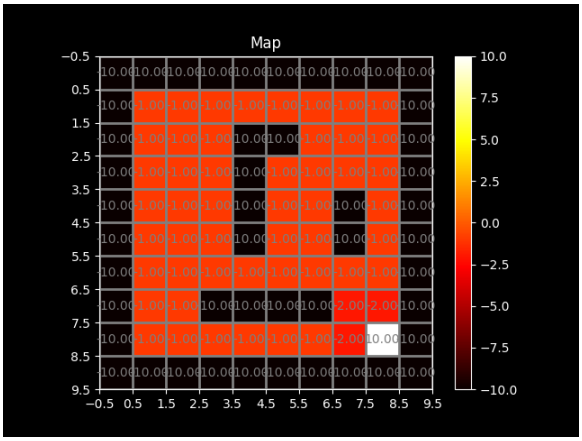

FIGURE 2. Created Map

2. Policy Evaluation

The policy evaluation is done by

$$J^{(i+1)}(x) = q(x, u^{(i)}(x)) + \gamma E_\epsilon[J^{(i)}(f(x, u^{(i)}(x) + \epsilon))]$$

Since this system is an MDP. This equation reduces to

$$J^\pi = q^u + \gamma T J^\pi$$

$$J^\pi = (I - \gamma T)^{-1} q^u$$

The plot after the first iteration is shown below.

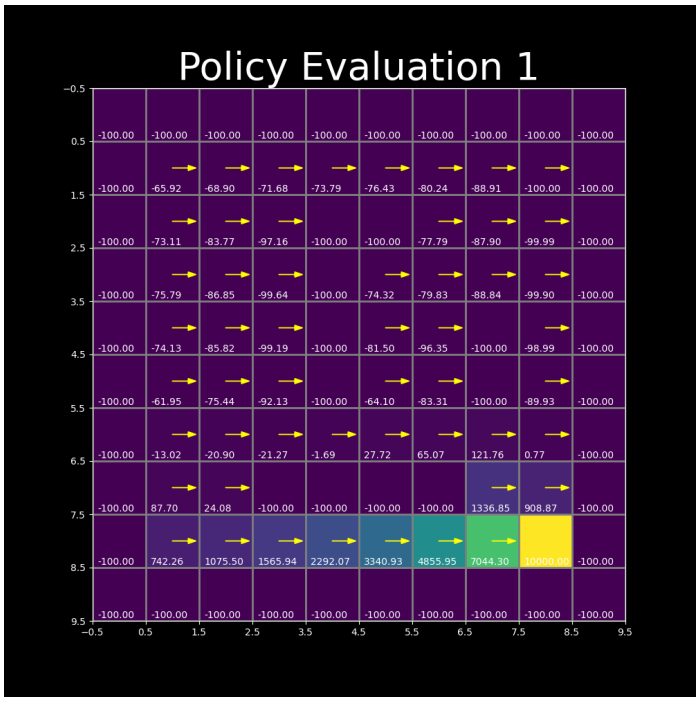

Figure 3. Caption

3. Policy Iteration

In each iteration, the policy improvement was done after the policy evaluation step. The policy improvement is done as:

$$u^{(k+1}(x) = argmin_{u \in U} E_\epsilon[q(x, u) + \gamma J^{\pi^{(k)}}(f(x, u) + \epsilon)]$$

I have attached the iterations for the cost and reward structure mentioned in the homework below.

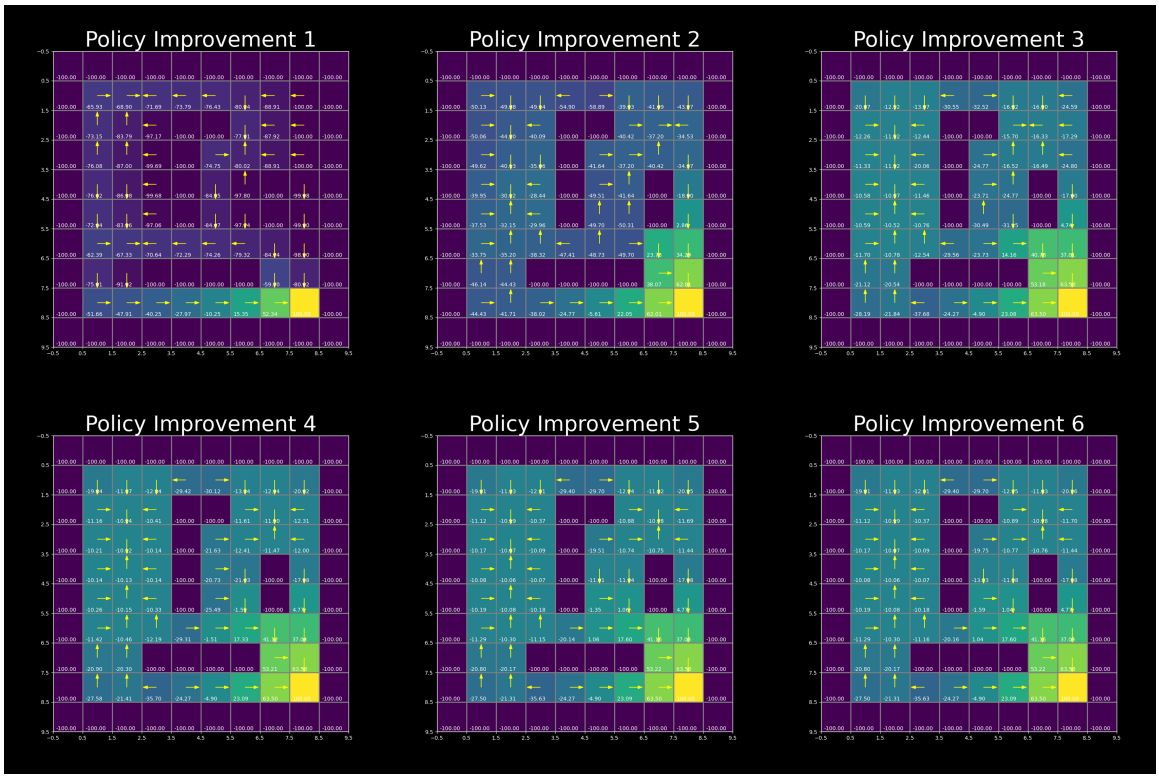

FIGURE 4. Policy Improvement with Goal Reward 10

I noticed that the policy doesn't necessarily ever converge. I increased the reward to the goal and observed the policy converged with the goal.

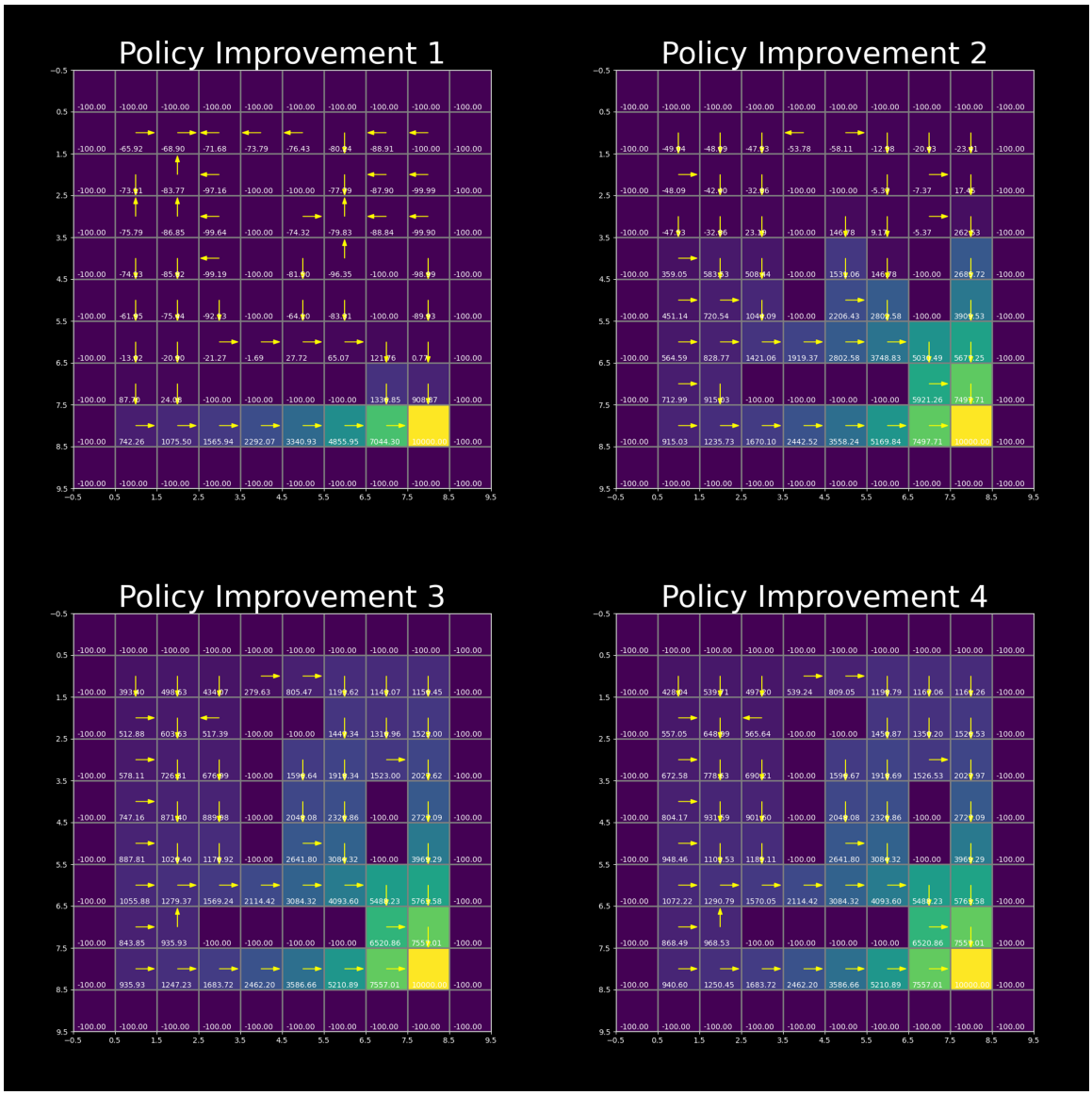

FIGURE 5. Policy Improvement with Goal Reward 100

Solution 2 (Time Spent: 12 hours)**.** .

4. SLAM

The code was written according to the instructions provided. A copy of the code has been submitted on Gradescope. Below are the results for the dynamics step test:

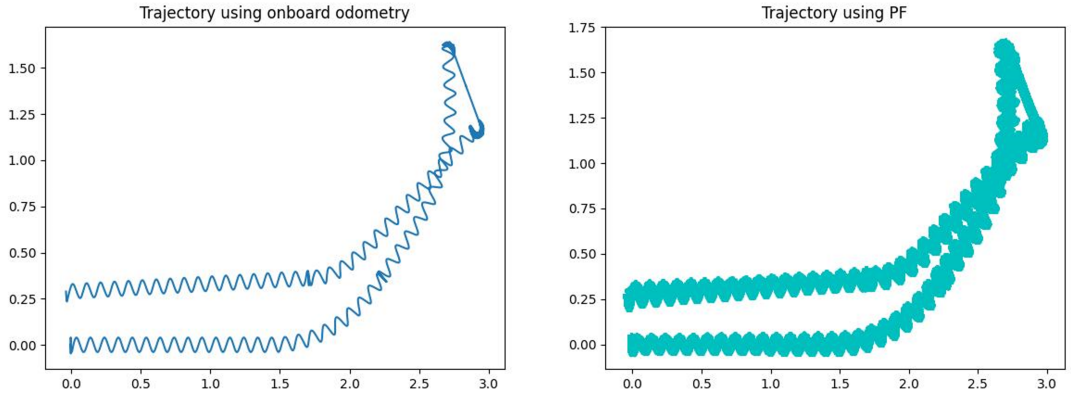

FIGURE 6. Dynamics Test

After running the full SLAM, The maps created with the estimated trajectories are shown int the next page.

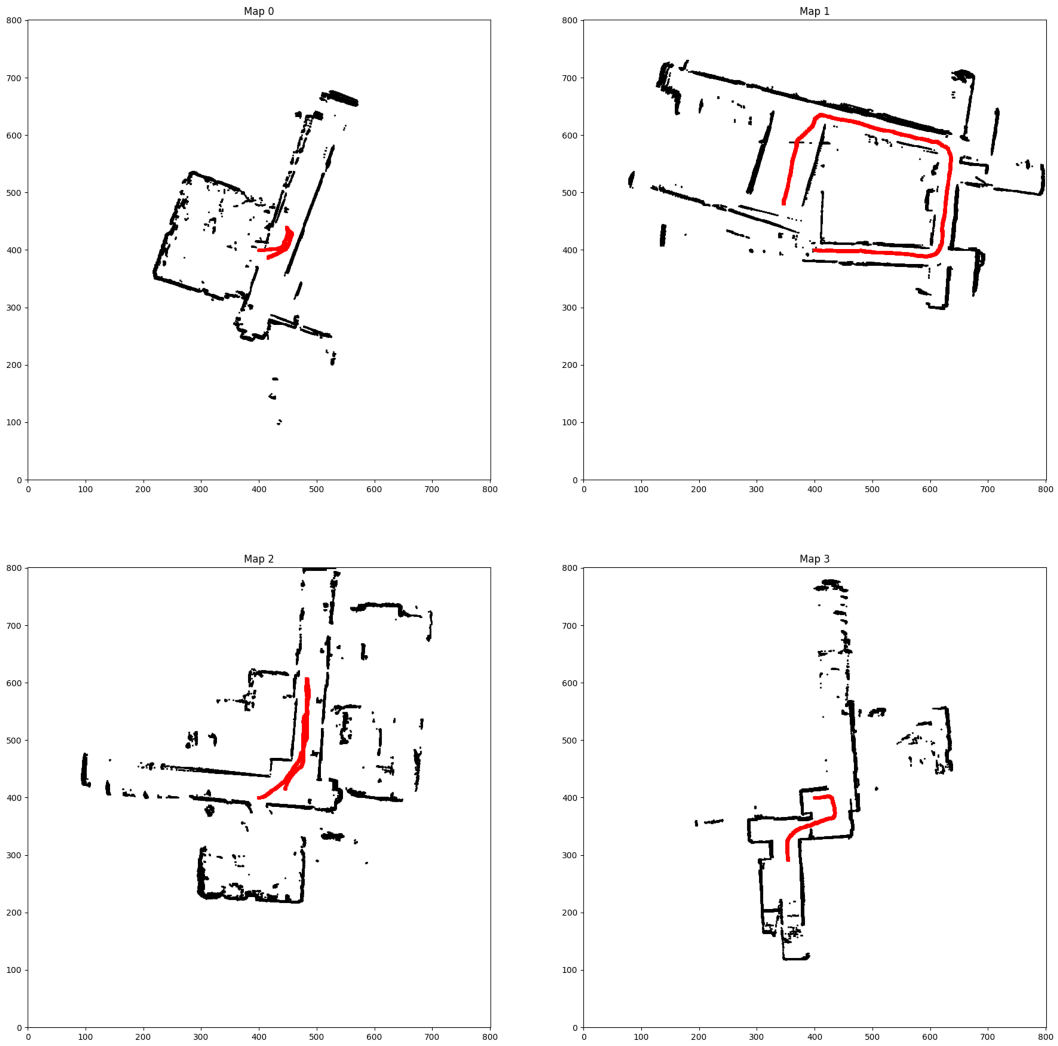

FIGURE 7. Maps

Solution 3 (Time Spent: 4 hours)**.** .

5. ACROBOT

5.1. **Linearizing the dynamics.** The dynamics of the system is given by

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = Bu$$

where

$$M(q) = \begin{bmatrix} I_1 + I_2 + m_2 l_1^2 + m_2 l_1 l_2 \cos q_2 & \frac{1}{2} I_2 + m_2 l_1 l_2 \cos q_2 \\ I_2 + \frac{1}{2} m_2 l_1 l_2 \cos q_2 & I_2 \end{bmatrix}$$

$$C(q,\dot{q}) = \begin{bmatrix} -m_2 l_1 l_2 \sin q_2 \dot{q}_2 & -\frac{1}{2} m_2 l_1 l_2 \sin q_2 \dot{q}_2 \\ \frac{1}{2} m_2 l_1 l_2 \sin q_2 \dot{q}_1 & 0 \end{bmatrix}$$

$$G(q) = \begin{bmatrix} \left(\frac{1}{2} m_1 l_1 g + m_2 l_1 g\right) \sin q_1 + \frac{1}{2} m_2 l_2 g \sin\left(q_1 + q_2\right) \\ \frac{1}{2} m_2 l_2 g \sin\left(q_1 + q_2\right) \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Now Consider the Taylor series expansion at x_f.

$$\begin{aligned} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{u}) &\approx \mathbf{f}\left(\mathbf{x_f},\mathbf{u_f}\right) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_{\mathbf{x=x_f},\mathbf{u=u_f}} \left(\mathbf{x}-\mathbf{x_f}\right) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right]_{\mathbf{x=x_f},\mathbf{u=u_f}} \left(\mathbf{u}-\mathbf{u_f}\right) \\ &\approx \mathbf{f}\left(\mathbf{x_f},\mathbf{u_f}\right) + [A]_{\mathbf{x=x_f},\mathbf{u=u_f}} \left(\mathbf{x}-\mathbf{x_f}\right) + [B]_{\mathbf{x=x_f},\mathbf{u=u_f}} \left(\mathbf{u}-\mathbf{u_f}\right) \end{aligned} \tag{1}$$

Where A and B are the Jacobians of f(x, u) with respect to x and u, respectively. Now,

$$\begin{aligned} f(x,u) = \begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} &= \begin{bmatrix} \dot{q} \\ M^{-1}(Bu - G - C\dot{q}) \end{bmatrix} \\ A = \begin{bmatrix} \frac{\partial f}{\partial q_1} & \frac{\partial f}{\partial q_2} \end{bmatrix} &= \begin{bmatrix} \ddot{q} & I \\ M^{-1}(-\frac{\partial G}{\partial q} - (\frac{\partial C}{\partial q}\dot{q} + C\ddot{q})) & 0 \end{bmatrix} \\ &\text{Here at } x_f,\ \dot{q}, \ddot{q} = 0 \\ &= \begin{bmatrix} 0 & I \\ -M^{-1}(\frac{\partial G}{\partial q}) & 0 \end{bmatrix} \\ B = \begin{bmatrix} 0 \\ -M^{-1}B \end{bmatrix} & \end{aligned} \tag{2}$$

The lower left quarter of A simplifies to:

$$\begin{bmatrix} (0.5 * m_1 l_1 g + m_2 l_1 g) cos(q_1) & 0.5 * m_2 l_2 g * cos(q_1 + q_2) \\ 0.5 * m_2 l_2 g * cos(q_1 + q_2) & 0.5 * m_2 l_2 g * cos(q_1 + q_2) \end{bmatrix} \tag{3}$$

5.2. **LQR Zones.** The P matrix obtained was the following:

$$P = \begin{bmatrix} 76890.12 & 36733.25 & 30950.15 & 18215.27 \\ 36733.25 & 17886.96 & 14858.95 & 8835.37 \\ 30950.15 & 14858.95 & 12474.13 & 7360.93 \\ 18215.27 & 8835.37 & 7360.93 & 4367.95 \end{bmatrix} \tag{4}$$

Whose Eigenvalues are [1.1131e+05, 3.0492e+02, 2.5433e-01, 1.1161e-01]

So the threshold c I chose was the first Eigenvalue. For the controller to switch to LQR only when

$$V(\delta x) = \delta x^\top P \delta x \leq c$$

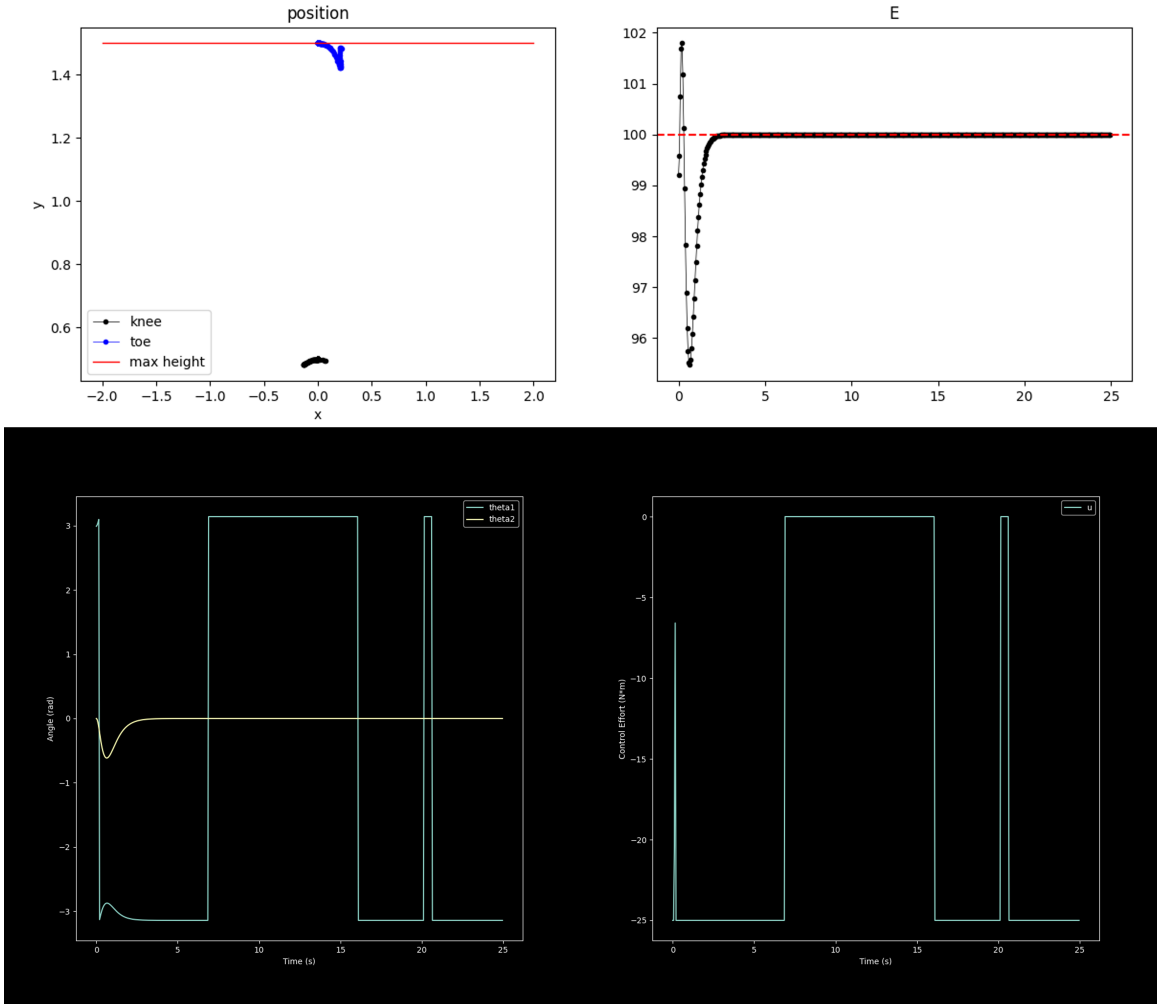

FIGURE 8. Controller behavior in the LQR zone

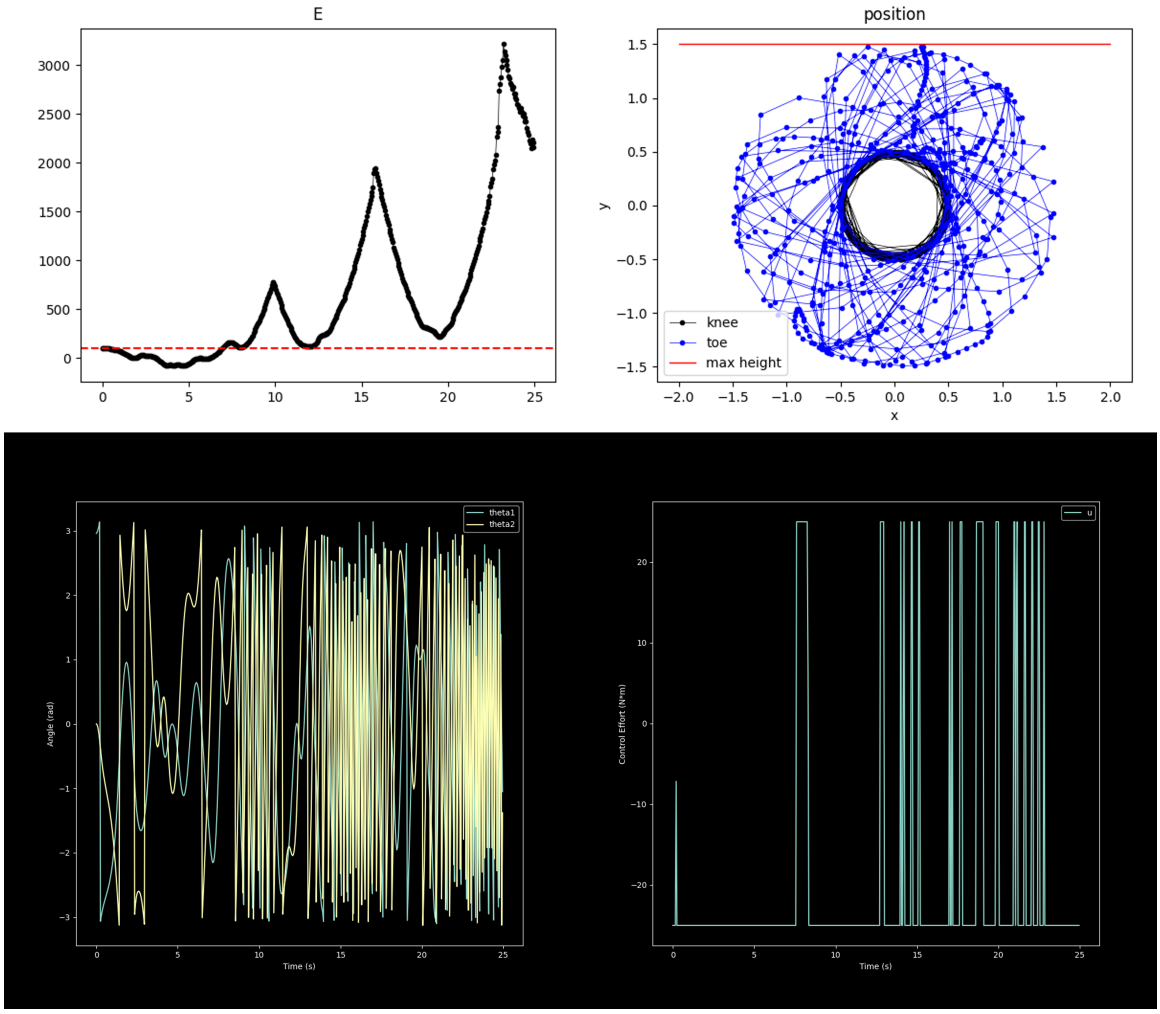

FIGURE 9. Controller behavior outside of the LQR zone

5.3. **Energy Shaping Controller.** The derivation for u is as below:

$$M_{11}\ddot{q}_1 + M_{12}\ddot{q}_2 = \tau_1$$
$$M_{21}\ddot{q}_1 + M_{22}\ddot{q}_2 = \tau_2 + u$$
$$\ddot{q}_1 = \frac{\tau_1 - M_{12}\ddot{q}_2}{M_{11}}$$

Substituting in equation (2), we get,

$$(M_{22} - \frac{M_{21}M_{12}}{M_{11}})\ddot{q}_2 + \begin{bmatrix} \frac{M_{21}}{M_{11}} & -1 \end{bmatrix} \begin{bmatrix} -C\dot{q} - G \end{bmatrix} = u$$

This was used to form the energy shaping controller, and the controllers switched based on the threshold mentioned above. The graphs for the initial position at the bottom are shown here.

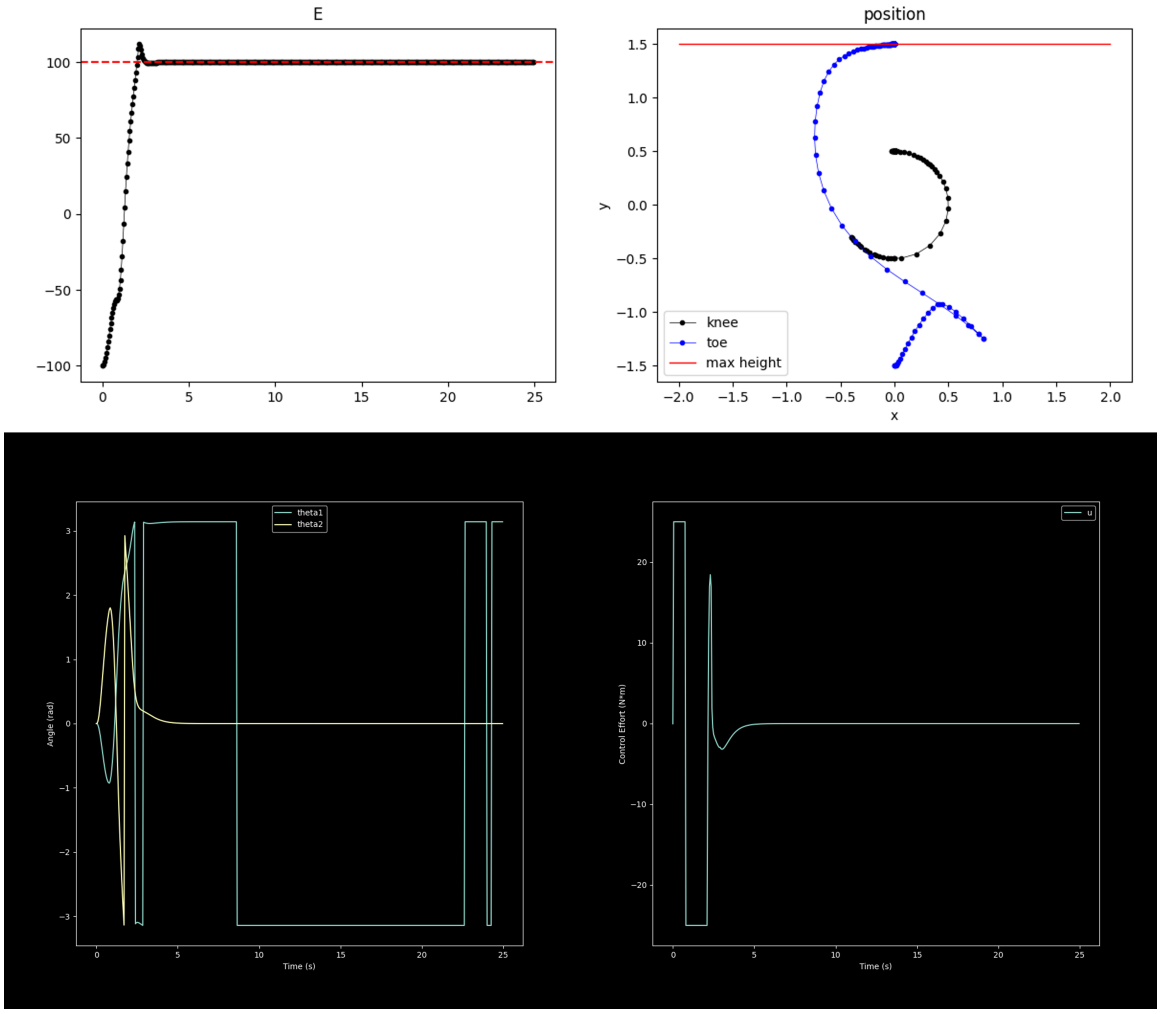

FIGURE 10. Controller behavior outside of the LQR zone