16899: Adaptive Control and Reinforcement Learning-HW1

Akshit Mishra

February 8, 2020

1 Systems

Given system:

$$\dot{x} = \begin{bmatrix} \dot{p_x} \\ \dot{p_y} \\ \dot{v} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} v\cos\theta \\ v\sin\theta \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} u \tag{1}$$

1. As the system can be written in the following form where we can separate out control input u from the function g(x), the system can be deemed **control-affine**:

$$\dot{x} = f(x) + g(x)u \tag{2}$$

2. The system can be best defined in the following manner:

   (a) **Non-linear**: The system is non-linear because it can be written in the following form:

   $$\dot{x} = \vec{A}(x) + \vec{B}u \tag{3}$$

   i.e. the system can be separated in the form of the states and the inputs.

   (b) **Continuous-Time**: The system can be categorized as a continuous time system because the inputs and outputs for the given system have continuous values.

   (c) **Time-invariant**: The system can be classified as time-invariant as the dynamics of the system do not change with time. The same system as stated in equation 1 can be used to define the motion of system at all times.

   (d) **Deterministic**: The given system is considered deterministic because it does not have any noise elements. The system is only made up of known states and inputs. Moreover, the system always produces the same output, for a given input condition.

(e) **Continuous-State:** The system can be described as continuous-state as all the independent states of the systems can take continuous values.

3. **Linearizing the system** at arbitrary $\begin{bmatrix} p_x^r & p_y^r & v^r & \theta^r \end{bmatrix}$ yields the following:

$$\mathbf{A} = \frac{\delta f}{\delta \vec{x}} = \begin{bmatrix} 0 & 0 & \cos\theta^r & -v^r \cos\theta \\ 0 & 0 & \sin\theta^r & v^r \cos\theta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

$$\mathbf{B} = \frac{\delta f}{\delta \vec{u}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5}$$

The linearized system can be represented as:

$$\mathbf{A}\vec{x} + \mathbf{B}\vec{u} = \begin{bmatrix} 0 & 0 & \cos\theta^r & -v^r \cos\theta \\ 0 & 0 & \sin\theta^r & v^r \cos\theta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{x} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \vec{u} \tag{6}$$

4. **Controllability**: For checking controllability at the given point $\begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}$: At given point the system has the following dynamics:

$$\dot{\vec{\mathbf{x}}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{\mathbf{x}} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \vec{\mathbf{u}} \tag{7}$$

Computing the controllability matrix in MATLAB gives the following result:

$$C = \begin{bmatrix} B & AB & A^2B & A^3B \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

Since the controllability matrix is not full rank, the linearized system is **not controllable** at the given point.

5. The linearized system is controllable for the following conditions:

(a) For non-zero velocities:

$$v^r \neq 0 \tag{9}$$

At non-zero velocity the controllability matrix would be full rank for all θ values.

6. **Controllability-Non-linear system**: The non-linear system is controllable because all the states can be controlled to go from any initial state to a final state. As shows from the equations below:
$$\dot{p_x} = v\cos\theta \tag{10}$$
$$\dot{p_y} = v\sin\theta \tag{11}$$
Since $\dot{v}$ and $\dot{\theta}$ are control inputs, they can directly affect our states v , which in turn affect p_x and p_y as seen from equations 10 and 11.

The controllability of the non-linear system can also be evaluated by employing Lie Brackets, which showed full rank, indicating a controllable non-linear system.

2 Continuous Time Optimal Control

(a) **Co-to-go function**:
$$J_t(x_t : u_{t:T}) = \int_t^T l_\tau(x_\tau),(u_\tau)\ d\tau + l_T(x_T)) \tag{12}$$
where,
$$l_\tau(x_\tau),(u_\tau) = \frac{1}{2}(x_t - xg)^T Q(x_\tau - xg) \tag{13}$$
$$l_T(x_T),(u_T) = \frac{1}{2}(x_t - xg)^T S(x_\tau - xg) \tag{14}$$

(b) **Value function (optimal cost-to-go)**:
$$V_t(x_t : u_{t:T}) = \underbrace{min}_{u_{t:T}} J_t(x_t : u_{t:T}) = \underbrace{min}_{u_{t:T}} \int_t^T l_\tau(x_\tau),(u_\tau)\ d\tau + l_T(x_T)) \tag{15}$$

(c) **Hamilton-Jacobi-Bellman Equation**:
$$0 = \frac{\delta}{\delta t}V_t(t, x_t) + \underbrace{min}_{u_t} l_t(x_t),(u_t)+ \tag{16}$$

(d) **Hamiltonian and co-state:**
$$H(t, x_t, u_t, \lambda_t) = l_t(x_t, u_t) + \lambda^T f(x_t, u_t) \tag{17}$$

where, l_t is as shown in equation 12 and

$$f(x_t, u_t) = \dot{x} = \begin{pmatrix} v\cos\theta \\ v\sin\theta \\ \dot{v} \\ \dot{\theta} \end{pmatrix} \tag{18}$$

where,

$$\lambda_t^T = \nabla_x V(t, x_t) = \nabla_x min J_t(x_t : u_{t:T}) \tag{19}$$

$$\lambda_t^T = \underbrace{min}_{u_{t:T}} \int_t^T (x_\tau - x_g)Q d\tau + (x_T - x_g)S \tag{20}$$

(e) **Co-state Dynamics**:
Maximum Principle in Continuous time: If $x_{0:T}^*$ and $u_{0:T}^*$ and are the optimal state and control trajectories starting at , then there exits a co-state trajectory $\lambda_{0:T}^*$ with λ_T^* satisfying:

$$\dot{x_t^*} = \frac{\delta}{\delta\lambda} H(t, x_t^*, u_t^*, \lambda_t^*) = f(x_t^*, u_t^*) \tag{21}$$

$$-\dot{\lambda_t^*} = \frac{\delta}{\delta x} H(t, x_t^*, u_t^*, \lambda_t^*) = \frac{\delta}{\delta x} l_t(x_t^*, u_t^*) + [\frac{\delta}{\delta x} f(x_t^*, u_t^*)]^T \lambda_t^* \tag{22}$$

$$u_t^* = arg \underbrace{min}_{u} H(t, x_t^*, u, \lambda_t^*) \tag{23}$$

where,

$$f(x_t^*, u_t^*) = \begin{bmatrix} \dot{p_x^*} \\ \dot{p_y^*} \\ \dot{v^*} \\ \dot{\theta^*} \end{bmatrix} \tag{24}$$

$$\frac{\delta}{\delta x} f(x_t^*, u_t^*) = \begin{bmatrix} \ddot{p_x^*} \\ \ddot{p_y^*} \\ \ddot{v^*} \\ \ddot{\theta^*} \end{bmatrix} \tag{25}$$

Boundary Condition:

$$\lambda_{0:T}^* = \frac{\delta}{\delta x} l_T(x_T^*) \tag{26}$$

where,

$$\frac{\delta}{\delta x} l_T(x_T^*) = (x_T^* - x_g)^T S \tag{27}$$

(f) **Optimal Control Law:**
Maximum Principle states:

$$-\dot{\lambda}_t = \frac{\delta}{\delta x} H(t, x_t, u_t, \lambda_t) = Q_t x_t + A_t^T \lambda_t \tag{28}$$

$$0 = \frac{\delta}{\delta u} H(t, x_t, u_t, \lambda_t) = R_t u_t + B_t^T \lambda_t \tag{29}$$

Equations 26 and 27 yield the following optimal control law:

$$u_t = -R^{-1} B_t \lambda_t \tag{30}$$

(g) **Hamilton-Jacobi-Bellman Equation-Infinite Horizon**:

$$-\ln(\delta) V(x_t) = \underbrace{min}_{u_t}(l(x_t, u_t) + \nabla_x V(x_t) f(x_t, u_t)) \tag{31}$$

where,

$$V(x_t) = \underbrace{min}_{u_t:\infty} \int_{\tau=t}^{\infty} (\delta^{\tau-t} l(x,u)\delta\tau) = \underbrace{min}_{u_t:t'} \int_{\tau=t}^{t'} (\delta^{\tau-t} l(x,u)\delta\tau) + (\delta^{t'} V(x_{t'})) \tag{32}$$

Taking limit for $t' \rightarrow$t$^+$ yields HJB equation with discount as seen in equation 31.

(h) **Condition for co-state:**

$$-\dot{\lambda} = \frac{\delta}{\delta x} H(x, u, \lambda) = Qx + (A^T)\lambda \tag{33}$$

(i)

$$H(x, u, \lambda) = \frac{1}{2}(x - x_g)^T Q(x - x) + \frac{1}{2} u^T R u + (\lambda^T (Ax + Bu) \tag{34}$$

Maximum Principle:

$$-\dot{\lambda} = \frac{\delta}{\delta x} H(x, u, \lambda) = Qx + (A^T)\lambda \tag{35}$$

$$0 = \frac{\delta}{\delta u} H(x, u, \lambda) = Ru + B^T \lambda \tag{36}$$

Assuming $\lambda_k = Px_k$ and using equations 55 and 56

$$0 = Ru + B^T \lambda \tag{37}$$

$$\lambda = Qx + A^T \lambda \tag{38}$$

The above equations yield:

$$u = -[B^T PB + R]^{-1} B^T PAx \tag{39}$$

(j) **Applying infinite-time LQR** Using Linearized system from equation 6 yields the following output: Given Parameters:

- $Q = \boldsymbol{I}_{4\times4}$
- R $= \boldsymbol{I}_{2\times2}$
- Boundary Conditions:
  - Initial State: $x_0 = [0, 0, 0, 0]^T$
  - Goal State: $x_G = [1, 1, 0.00001, 0]^T$
- $N = 10$

The optimal control law for the infinite-time LQR was $u = K(x - x_G)$, where x was an output from ODE45 solver and was updated at each iteration. The 'K' here denotes the LQR gains for the linearized system and was determined using the *lqr* command in MATLAB. The response obtained with default cost and penalty matrices was as follows:

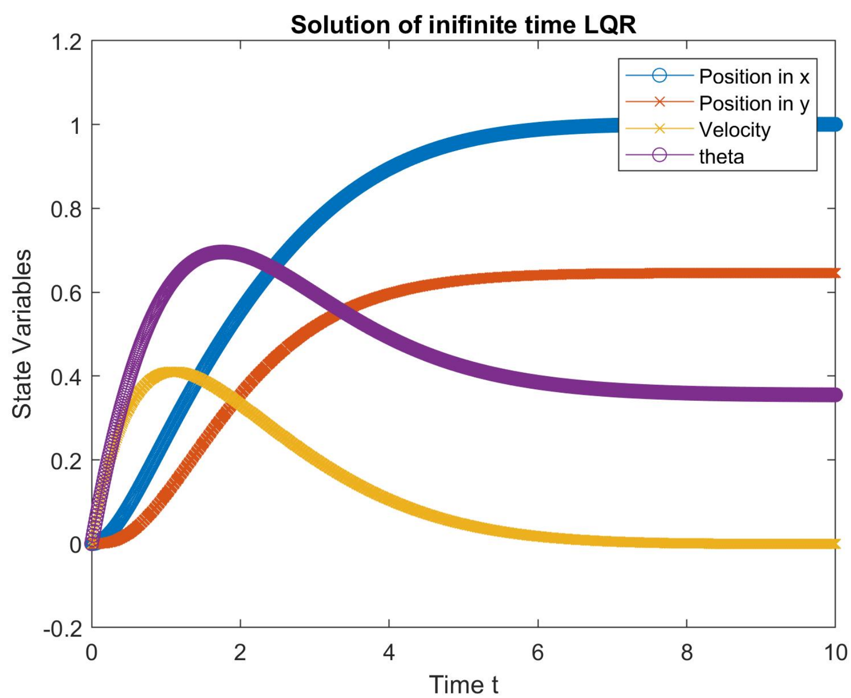

Figure 1: Output with default parameters

The plot above depicts that with the base parameters for the infinite-time LQR only position and velocity are able to converge close to the goal state.

Figure 8 above shows that with some preliminary tuning positions in x and y, velocity and theta were able to converge close to their desired goal state. However, theta still has a considerable error at its steady state value. Moreover, an extensive

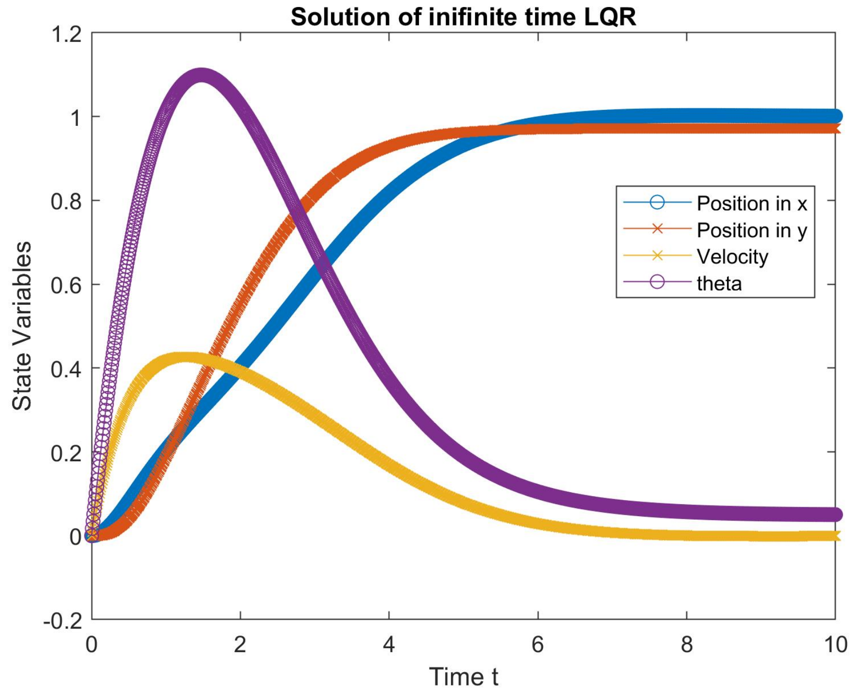

Figure 2: Output with tuned penalty matrices

tuning exercise did not seem to reduce the overall error for all the final states any further and such behavior could be attributed to the fact that the linearized system is **uncontrollable**. But the system characteristics such as the overshoot and settling time can be improved further.

3 Discrete Time Optimal Control

(a) Discretized system with $\Delta t = 0.01s$: Using difference equation $\dot{x}_t \approx \frac{x_{t+\Delta t} - x_t}{\Delta t}$,

$$x_{k+1} = \dot{x}_t * \Delta t + x_k \tag{40}$$

Substituting the dynamics($\dot{x}$) given in 1, we get:

$$x_{k+1} = \begin{bmatrix} v_k \cos\theta_k \\ v_k \sin\theta_k \\ 0 \\ 0 \end{bmatrix} \Delta t + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} u_k + \begin{bmatrix} p_{xk} \\ p_{yk} \\ v_k \\ \theta_k \end{bmatrix} \tag{41}$$

(b) • **Cost-to-go Function:**

$$\underbrace{min}_{u_0, u_1, \ldots, u_{N-1}} \quad J = \sum_{k=0}^{N-1} l_k(x_k, u_k) + l_N(x_N) \tag{42}$$

where the run-time cost is defined as,

$$l_k(x_k, u_k) = \frac{1}{2}(x_k - x_G)^T Q(x_k - x_G) + \frac{1}{2}{u_t}^T R u_t \tag{43}$$

and the terminal cost is:

$$l_N(x_k, u_k) = \frac{1}{2}(x_k - x_G)^T S(x_k - x_G) \tag{44}$$

- **Value function:** The optimal cost-to-go.

$$J_k{}^*(x_k) = \underset{u_k,\dots,u_{N-1}}{min} J_k(x_k; u_k, \dots, u_{N-1}) \tag{45}$$

- **Bellman Equation:** An update rule for value functions under the principle of optimality.

$$J_k{}^*(x_k) = \underset{u_k}{min}\ l_k(x_k, u_k) + J_{k+1}{}^*(x_{k+1}) \tag{46}$$

- **Hamiltonian:** Defined as equivalent to cost-to-go

$$H_k(x_k, u_k, \lambda_{k+1}) = l_k(x_k, u_k) + {\lambda_{k+1}}^T f(x_k, u_k) \tag{47}$$

where, l_k is the run-time cost defined in equation 43 and $f(x_k, u_k,)$ is the non-linear system defined in equation 40, and λ_{k+1} is the co-state defined in equation 48

- **Co-state:** Linear Parameterization with co-state λ_k:

$$J_k^*(x_k) = {\lambda_k}^T x_k + \gamma_k \tag{48}$$

here γ is a constant.

- **Dynamics of the co-state:** The Discrete Algebraic Riccati Equation (DARE) equation describes the dynamics of the co-state and is described below:

$$P_k = Q_k + {A_k}^T P_{k+1} A_k - {A_k}^T P_{k+1} B_k [R_k + {B_k}^T P_{k+1} B_k]^{-1} {B_k}^T P_{k+1} A_k \tag{49}$$

- **Maximum Principle and Boundary Condition:** The maximum principle can be defined as follows

$$\lambda_k^* = \frac{\delta}{\delta x} H_k(x_k^*, u_k^*, \lambda_{k+1}^*) = \frac{\delta}{\delta x} l_k(x_k^*, u_k^*) + (\frac{\delta}{\delta x} f(x_k^*, u_k^*))^T \lambda_{k+1}^* \tag{50}$$

where $f(x_k^*, u_k^*)$ is defined in equation 41.

$$x_{k+1}^* = \frac{\delta}{\delta \lambda} H_k(x_k^*, u_k^*, \lambda_{k+1}^*) = f(x_k^*, u_k^*) \tag{51}$$

$$u_k^* = arg\ \underset{u}{min} H_k(x_k^*, u_k^*, \lambda_{k+1}^*) \tag{52}$$

Boundary Condition:

$$\lambda_N^* = \frac{\delta}{\delta x_N} J_N^*(x_N^*) \tag{53}$$

(c) **Optimal Control Law**:

$$H(x_k, u_k, \lambda_{k+1}) = \frac{1}{2}(x_k - x_g)^T Q_k (x_k - x_g) + \frac{1}{2} u_{k^T} R u_k + (\lambda_{{k+1}^T}(Ax_k + Bu_k) \quad (54)$$

Maximum Principle:

$$\dot{\lambda}_k = \frac{\delta}{\delta x} H(x_k, u_k, \lambda_{k+1}) = Qx_k + A_t^T \lambda_{k+1} \quad (55)$$

$$0 = \frac{\delta}{\delta u} H(x_k, u_k, \lambda_{k+1}) = Ru_k + B_t^T \lambda_{k+1} \quad (56)$$

Assuming $\lambda_k = Px_k$ and using equations 55 and 56

$$0 = Ru_k + B^T \lambda_{k+1} \quad (57)$$

$$\lambda_k = Qx_k + A^T \lambda_{k+1} \quad (58)$$

The above equations yield the following optimal control law where P can be define as shows in equation 49

$$u = -[B^T PB + R]^{-1} B^T PAx \quad (59)$$

(d) **Solve for control trajectories**: At time N, the co-state function can be represented as:

$$\lambda_N = S(f(x_9, u_9)) - x_g \quad (60)$$

where S is the terminal cost.

Moving backwards for $k = N-1, \ldots, 1, 0$, I solved the following equations using the Symbolic toolbox in MATLAB:

$$\frac{\delta}{\delta x} H_k(x_k, u_k, \lambda_{k+1}) = \frac{\delta}{\delta x} l_k(x_k, u_k) + [\frac{\delta}{\delta x} f(x_k, u_k)]^T \lambda_{k+1} \quad (61)$$

where $l_k(x_k, u_k)$ is as defined in equation 43 and $\frac{\delta}{\delta x} f(x_k, u_k)$ is defined as:

$$\frac{\delta}{\delta x} f(x_k, u_k) = \begin{bmatrix} 1 & 0 & \cos\theta_k \Delta t & -v_k \sin\theta_k \Delta t \\ 0 & 1 & \sin\theta_k \Delta t & v_k \cos\theta_k \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (62)$$

Moreover, for the backward pass, using the optimal control law derived earlier in equation 59, we get:

$$u_{N-1} = -R^{-1} B^T \lambda_N \quad (63)$$

where λ_N is defined in equation 60. λ_N and equation 63 can the be used to obtain the following:

$$\lambda_{N-1} = Q(x_{N-1} - x_G) + [\frac{\delta}{\delta x} f(x_k, u_k)]^T \lambda_N \tag{64}$$

The dynamics for the non-linear system can be defined by using equation 41 The next move was to use the given initial state $x_0 = [0, 0, 0, 0]^T$ to commence a forwards pass to determine the optimal control policy as shows below:

$$x_0^* = x_0 \tag{65}$$

$$u_k^* = q_k^*(x_k^*), \forall k = 0, 1, \ldots, N-1 \tag{66}$$

$$\lambda_k^* = \lambda_k(x_k^*), \forall k = 0, 1, \ldots, N-1 \tag{67}$$

$$x_{k+1}^* = f(x_k^*, u_k^*) \forall k = 0, 1, \ldots, N-1 \tag{68}$$

The passes were carried out for $N = 1000$. The four above equations for the backward pass yield the following unknown terms: $\lambda_N, \lambda_{N-1}, u_{N-1}, x_{k+1}, x_k$. Now, since we have four equations and five unknowns in the backward pass, we obtain $\lambda_{N-1} intermsof x_k$ and using this information along with the given initial conditions and goal state to obtain the optimal state, co-state, and control trajectories.

(e) **Finite-time LQR:** Linearized discrete-time System Initial and goal state are the same as described in the previous part. The equations used to find the solution to the Finite-time LQR were equations 49 and equation 69 in the backward pass to determine the required gains for the LQR system.

$$K = [R_k + {B_k}^T P_{k+1} B_k]^{-1} {B_k}^T P_{k+1} A_k \tag{69}$$

After determining the gains, a forward pass was conducted to determine the optimal control law from and also find the dynamics of the system using equations 71, 72, and 73

$$u_k = K(x_k - x_G) \tag{70}$$

$$x_{k+1} = Ax_k + Bu_k \tag{71}$$

where linearized matrices A and B are:

$$A = \begin{bmatrix} 1 & 0 & 0.01 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{72}$$

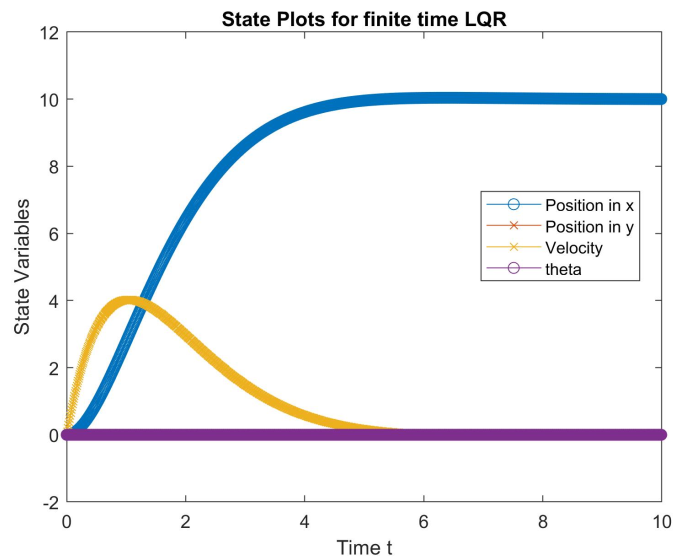

Figure 3: Output with tuned penalty matrices

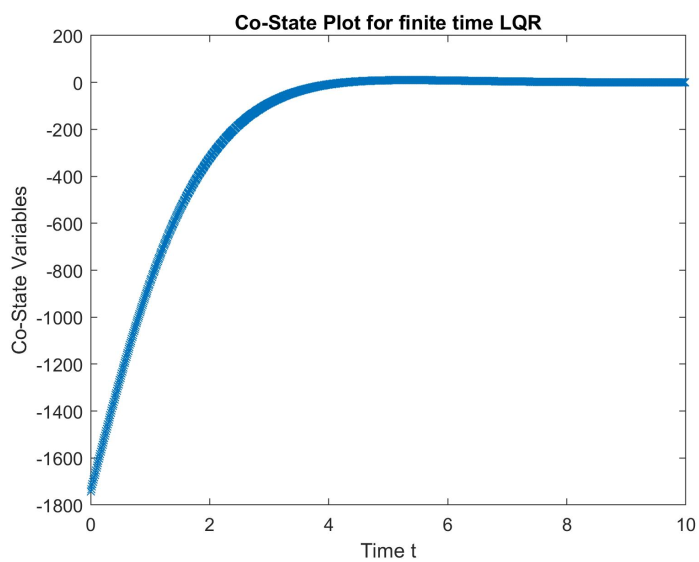

Figure 4: Output with tuned penalty matrices

$$B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} \tag{73}$$

The plots for states, co-states and controls are provided below.
The plots above indicate the following:

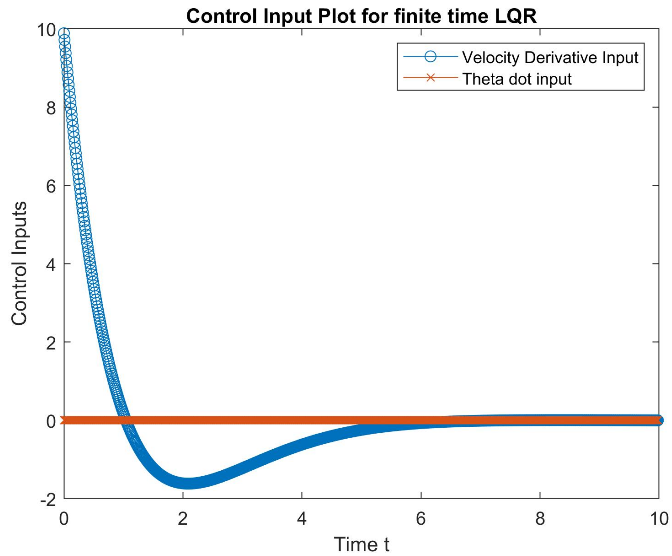

Figure 5: Output with tuned penalty matrices

- The states' plot shows that the states position in x, velocity, and θ all converge to the goal state over time. However, position in y does not reach its goal value of 10 and this aberration can be attributed to the uncontrollability of the linearized system highlighted in previous problem.
- The co-state plot indicates that the co-states' line-up with the state plot.
- The plot for the control trajectories indicate that theta does not change over the horizon at all as there is no input for theta change. For velocity, the change in velocity is negative at first and then the change becomes zero as the velocity reaches a steady state value.

(f) **Finite-time LQR:** Nonlinear discrete-time System Initial and goal state are the same as described in the previous parts. The equations used to find the solution to the Finite-time LQR were equations 49 and equation 69 in the backward pass to determine the required gains for the LQR system.

After determining the gains, a forward pass was conducted to determine the optimal control law from and also find the dynamics of the system using equations 41

$$x_{k+1} = f(x_k, u_k) \tag{74}$$

The plots for states, co-states and controls are provided below.

The plots above indicate the following:

- The state, co-state and the control trajectories plots are same for linearized and discretized systems.

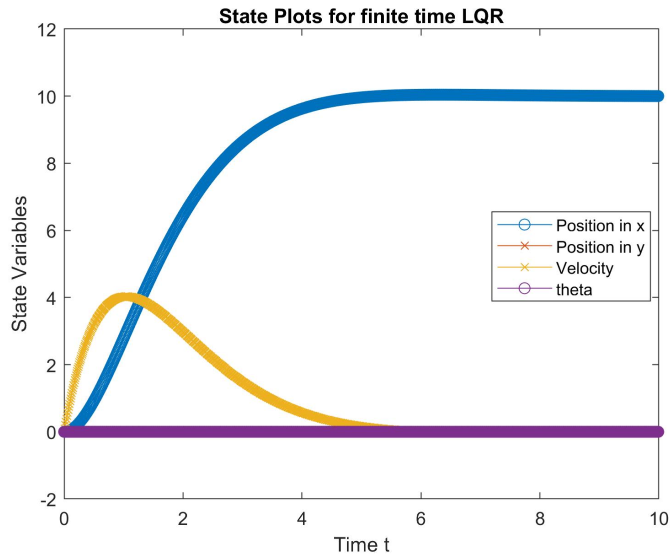

Figure 6: Output with tuned penalty matrices

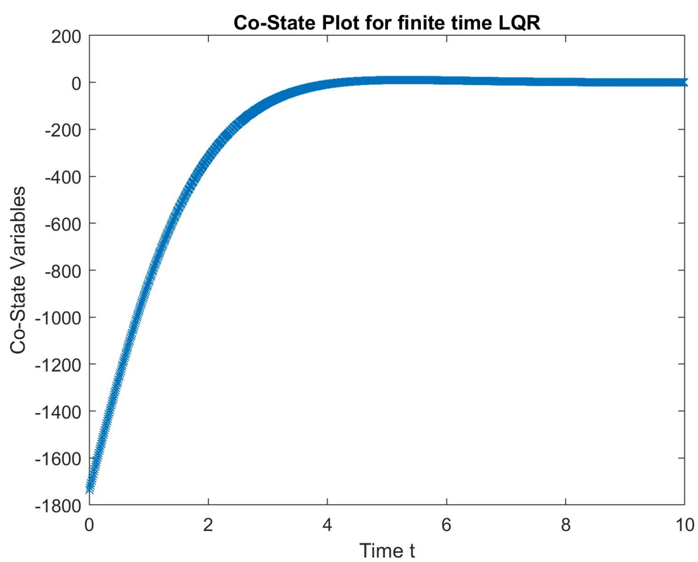

Figure 7: Output with tuned penalty matrices

- The similarities indicate the overall dynamics are the same. This assumption is based on the fact that the linearized system is an approximation of the non-linear system.
- It is possible that if we had linearized the system at some other arbitrary point then the plots might have turned to be different.
- The values of theta and velocity are zero in both cases because those corre-

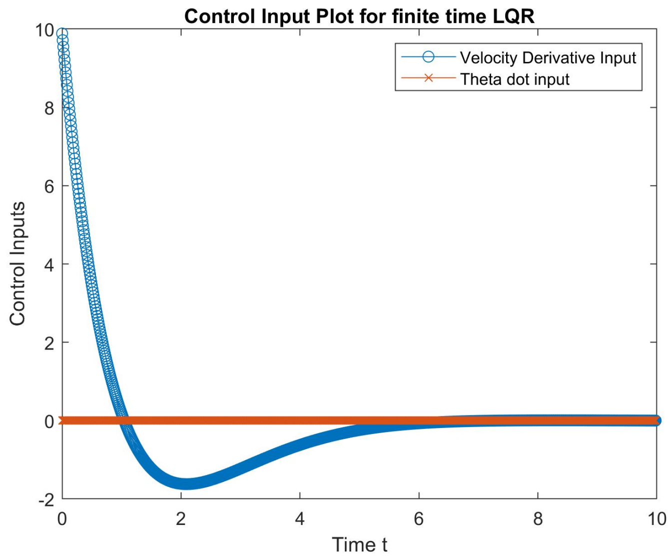

Figure 8: Output with tuned penalty matrices

spond to the linearization point.