Unconstrained Optimization

Denis Mulumba and Sandra Babyale

November 14, 2023

1 Introduction

In this project, the primary objective is to design and optimize a controller for the cart-pole system, which is widely recognized in the fields of control theory and reinforcement learning. The inverted pendulum system serves as an essential benchmark for the evaluation, testing, and comparison of a variety of control strategies. In addition to its inherent simplicity, its inherently unstable dynamics make it an ideal candidate for exploring the depths of unconstrained optimization techniques. This project focuses on employing unconstrained optimization methods to design a controller that can effectively stabilize the cart-pole system in its upright position.

The task involves stabilizing the cart-pole system's upright equilibrium using two methods: an optimal linear controller for a linear time-invariant (LTI) system and a neural network-based controller.

2 Designing the optimal controller

In order to derive the form of the optimal linear controller, we consider a generic LTI system in state-space form.

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{1}$$

where system state $\mathbf{x}$ an n-vector is given by $\mathbf{x} = \begin{bmatrix} q & \dot{q} \end{bmatrix}$ where $q = \begin{bmatrix} x & \theta \end{bmatrix}$, $\mathbf{u}$ is an m-vector, depicting the control input, $\mathbf{A}$ is the state-matrix and $\mathbf{B}$ is the input matrix with an infinite-horizon cost function is given by;

$$\mathcal{J} = \int_0^\infty \big(\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u}\big)dt, \ \ \mathbf{Q} = \mathbf{Q}^T \geq 0 \ \mathbf{R} = \mathbf{R}^T > 0 \tag{2}$$

Our goal is to find the optimal cost-to-go function $\mathcal{J}^*(x)$ which satisfies the Hamilton - Jacobi-Bellman equation:

$$\min_u \left[\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u} + \frac{\partial \mathcal{J}^*}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u})\right] = 0, \quad \forall \mathbf{x} \tag{3}$$

And it is well-known that for linear control systems such as (1), the optimal cost-to-go function is quadratic in the states and therefore $\mathcal{J}^*(x)$ is given by

$$\begin{aligned} &\mathcal{J}^*(x) = \mathbf{x}^T\mathbf{S}\mathbf{x}, \qquad \mathbf{S} = \mathbf{S}^T \geq 0 \\ &\Rightarrow \frac{\partial \mathcal{J}^*}{\partial \mathbf{x}} = 2\mathbf{x}^T\mathbf{S} \end{aligned} \tag{4}$$

Substituting for $\frac{\partial \mathcal{J}^*}{\partial \mathbf{x}}$ into equation (3), we obtain the Hamilton - Jacobi-Bellman equation as

$$\min_u \left[\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u} + 2\mathbf{x}^T\mathbf{S}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u})\right] = 0, \quad \forall \mathbf{x} \tag{5}$$

Assuming that $\mathbf{u}^* = \pi^*(x)$ is the the optimal policy for the form given in (3) for the optimal cost-to-go function, since u is unbounded, then $\mathbf{u}^*$ lies in the interior of the space $\Omega \in \mathrm{R}^n$. Therefore using the first order necessary condition, we equate the gradient of the form (3) for the optimal cost-to-go function with respect to $\mathbf{u}$ to zero as below

$$
\begin{aligned}
&\frac{\partial}{\partial \mathbf{u}}\big(\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u} + 2\mathbf{x}^T\mathbf{S}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u})\big) = 0 \\
&2\mathbf{u}^T\mathbf{R} + 2\mathbf{x}^T\mathbf{S}\mathbf{B} = 0 \\
&\mathbf{u}^* = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x} = \pi^*(\mathbf{x})
\end{aligned} \tag{6}
$$

Therefore the optimal policy is given as $\mathbf{u}^* = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x}$. Substituting the expression for $\mathbf{u}^*$ into the Hamilton-Jacobi-Bellman equation, we obtain

$$
\begin{aligned}
&\mathbf{x}^T\mathbf{Q}\mathbf{x} + (-\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x})^T\mathbf{R}(-\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x}) + 2\mathbf{x}^T\mathbf{S}(\mathbf{A}\mathbf{x} + \mathbf{B}(-\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x})) = 0 \\
&\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{x}^T\mathbf{S}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x} + 2\mathbf{x}^T\mathbf{S}\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{S}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x} = 0 \\
&\mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{x}^T\mathbf{S}\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{S}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x} = 0
\end{aligned} \tag{7}
$$

All of the terms here are symmetric except for the $2\mathbf{x}^T\mathbf{S}\mathbf{A}\mathbf{x}$, but since we know that $\mathbf{S}\mathbf{A} = \frac{1}{2}\big(\mathbf{S}\mathbf{A} + \mathbf{A}^T\mathbf{S}\big) + \frac{1}{2}\big(\mathbf{S}\mathbf{A} - \mathbf{A}^T\mathbf{S}\big)$, then

$$
\begin{aligned}
2\mathbf{x}^T\mathbf{S}\mathbf{A}\mathbf{x} &= \mathbf{x}^T\Big(\big(\mathbf{S}\mathbf{A} + \mathbf{A}^T\mathbf{S}\big) + \big(\mathbf{S}\mathbf{A} - \mathbf{A}^T\mathbf{S}\big)\Big)\mathbf{x}^T \\
&= \mathbf{x}^T\big(\mathbf{S}\mathbf{A} + \mathbf{A}^T\mathbf{S}\big)\mathbf{x}^T
\end{aligned}
$$

because $\mathbf{x}^T\big(\mathbf{S}\mathbf{A} - \mathbf{A}^T\mathbf{S}\big)\mathbf{x}^T = 0$ making $2\mathbf{x}^T\mathbf{S}\mathbf{A}\mathbf{x}$ symmetric. Therefore substituting for $2\mathbf{x}^T\mathbf{S}\mathbf{A}\mathbf{x}$ in equation (7) we get

$$
\mathbf{x}^T[\mathbf{Q} + \mathbf{S}\mathbf{A} + \mathbf{A}^T\mathbf{S} - \mathbf{S}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}]\mathbf{x} = 0, \quad \forall x \tag{8}
$$

Since the above condition must hold for all $\mathbf{x}$, it is sufficient to consider the matrix equation below

$$
\mathbf{Q} + \mathbf{S}\mathbf{A} + \mathbf{A}^T\mathbf{S} - \mathbf{S}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S} = 0 \tag{9}
$$

We observe that the matrix equation (9) is quadratic in $\mathbf{S}$, making its solution non-trivial. The equation has a single positive-definite solution if and only if the system is controllable.

2.1 Example

Given that $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\mathbf{R} = 1/2$, substituting in (9), we obtain

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mathbf{S}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\mathbf{S} - 2\mathbf{S}\begin{bmatrix} 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \end{bmatrix}\mathbf{S} = 0 \tag{10}
$$

Using python's LQR package, we obtain that

$\mathbf{S} = \begin{bmatrix} 1.55377397 & 0.70710678 \\ 0.70710678 & 1.09868411 \end{bmatrix}$ and the optimal controller is given by $K = \begin{bmatrix} 1.41421356 & 2.19736823 \end{bmatrix}$.

3 Question 2

Using the Lagrangian formulation, the equations of motion of the cart-pole system may be derived to obtain:

$$\mathbf{M(q)\ddot{q}} + \mathbf{C(q,\dot{q})\dot{q}} = \tau_{\mathbf{g}}(\mathbf{q}) + \tilde{\mathbf{B}}\mathbf{u},$$

where

$$\mathbf{M(q)} = \begin{bmatrix} m_c + m + p & -m_p l\cos(\theta) \\ -m_p l\cos(\theta) & m_p l^2 \end{bmatrix}, \quad \mathbf{C(q,\dot{q})} = \begin{bmatrix} 0 & m_p l\sin(\theta)\dot{\theta} \\ 0 & 0 \end{bmatrix}, \tau(\mathbf{q}) = \begin{bmatrix} 0 \\ m_p gl\sin(\theta) \end{bmatrix}, \quad \tilde{\mathbf{B}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{11}$$

where $\mathbf{u}$ is identified as f, the external (control) force applied on the cart. Suppose that the particular system we are interested in has $m_c = 0.2kg$, $m_p = 0.5kg$, $l = 0.5m$. At Boise State University, the gravitational acceleration is $g = 9.80364ms^{-2}$. The resulting system is nonlinear, linearizing this system gives

$$\begin{bmatrix} m_c + m + p & -m_p l \\ -m_p l & m_p l^2 \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 \\ m_p gl\theta \end{bmatrix} + \begin{bmatrix} f \\ 0 \end{bmatrix}$$

Putting the above linear time-invariant system into standard form gives;

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \\ \dot{x_4} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{m_p g}{m_c} & 0 & 0 \\ 0 & \frac{(m_c+m_p)g}{m_c l} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m_c} \\ \frac{1}{m_c l} \end{bmatrix} f = \mathbf{Ax} + \mathbf{Bu} \tag{12}$$

Using the result in question 1 above together with the loss function

$$l(x) \approx \frac{1}{20}x^2 + \frac{1}{2}m_p gl\theta^2 + \frac{1}{2}(m_c + m_p)\dot{x}^2 + m_p l\dot{x}\dot{\theta} + \frac{a}{2}m_p l^2\dot{\theta}^2 + \frac{1}{10}f^2,$$

we deduce $\mathbf{Q}$ and $\mathbf{R}$. From equation 12, we notice that

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{m_p g}{m_c} & 0 & 0 \\ 0 & \frac{(m_c+m_p)g}{m_c l} & 0 & 0 \end{bmatrix} \quad and \quad B = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m_c} \\ \frac{1}{m_c l} \end{bmatrix}$$

From question 1, we have concluded that

$$\mathbf{u}^* = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{Sx} \quad and \quad \mathbf{Q} + \mathbf{SA} + \mathbf{A}^T\mathbf{S} - \mathbf{SBR}^{-1}\mathbf{B}^T\mathbf{S} = 0 \tag{13}$$

To deduce matrix Q, we let

$$Q = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix},$$

and

$$\begin{bmatrix} x & \theta & \dot{x} & \dot{\theta} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} x \\ \theta \\ \dot{x} \\ \dot{\theta} \end{bmatrix} = x^2 a_{11} + \theta^2 a_{22} + \dot{x}^2 a_{33} + a_{43}\dot{\theta}\dot{x} + \dot{\theta}^2 a_{44} + \dot{x}\dot{\theta}a_{34}$$

Comparing this with the loss function, we deduce Q as

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{20} & 0 & 0 & 0 \\ 0 & \frac{1}{2}m_p gl & 0 & 0 \\ 0 & 0 & \frac{1}{2}(m_c + m_p) & m_p l \\ 0 & 0 & m_p l & \frac{1}{2}m_p l^2 \end{bmatrix}.$$

Also, we conclude from the loss function, we conclude that $R = \frac{1}{10}$

Using the *Python Control System Library*'s **LQR** function, the optimal control matrix is

$$\mathbf{K} = \begin{bmatrix} -0.70710678 \\ 24.04259064 \\ -2.44505391 \\ 4.17850521 \end{bmatrix}^T,$$

and

$$\mathbf{S} = \begin{bmatrix} 0.17289142 & -0.29546494 & 0.12391443 & -0.06902828 \\ -0.29546494 & 4.98641143 & -1.20263877 & a0.84174529 \\ 0.12391443 & -1.20263877 & 0.39833655 & -0.22361881 \\ -0.06902828 & 0.84174529 & -0.22361881 & 0.15359446 \end{bmatrix}$$

S, **Q**,**A**,**B** and **R** satisfy the equation derived in question 1.

$$\mathbf{Q} + \mathbf{SA} + \mathbf{A}^T\mathbf{S} - \mathbf{SBR}^{-1}\mathbf{B}^T\mathbf{S} = \begin{bmatrix} 1.44e-15 & -2.29e-14 & 3.03e-15 & -4.94e-15 \\ -2.31e-14 & 1.78e-13 & -2.84e-14 & 4.09e-14 \\ 3.08e-15 & -3.46e-14 & 4.00e-15 & -8.44e-15 \\ -4.712e-15 & 3.38e-14 & -7.33e-15 & 6.88e-15 \end{bmatrix} = \mathbf{0}$$

3.1 Neural Network

In this section, we find a stabilizing controller u as a neural network using data-driven techniques.

$$\mathbf{K} = \begin{bmatrix} -0.707107 \\ 23.456890 \\ -2.427715 \\ 4.041519 \end{bmatrix}^T$$

Comparison with the result in (a): We observe that the optimal controller obtained using the Neural network is nearly identical to that reported in (a) above. This implies that the equation derived in (1) is still satisfied.

END