Exercise 7: Dynamic Programming

If you wish to receive feedback, please hand in before July 3, 2020,
by sending as email to florian.messerer@imtek.de (voluntary)

Prof. Dr. Moritz Diehl, Andrea Zanelli, Dimitris Kouzoupis, Florian Messerer

In this exercise, we will use dynamic programming (DP) to implement a controller for the inverted pendulum from Exercise 6,

$$
\begin{aligned}
\dot{\theta} &= \omega \\
\dot{\omega} &= \sin(\phi) + \tau,
\end{aligned}
\tag{1}
$$

where θ is the angle describing the orientation of the pendulum, ω is its angular velocity and τ is the input torque. The goal is to design a feedback policy capable of swinging up the pendulum starting from $\phi = \pi$. Moreover, we will prove the Schur Complement Lemma, which can be used to derive the formulation for the LQR controller.

1. **Dynamic programming:** Consider the following optimal control problem,

$$
\begin{aligned}
\min_{\substack{x_0,\dots,x_N, \\ u_0,\dots,u_{N-1}}} \quad & \sum_{i=0}^{N-1} \left(x_i^T Q x_i + u_i^T R u_i \right) + x_N^T Q_N x_N \\
\text{s.t.} \quad & x_0 = \bar{x}_0 \\
& x_{i+1} = F(x_i, u_i), \ i = 0, \dots, N-1, \\
& -10 \le u_i \le 10, \ i = 0, \dots, N-1,
\end{aligned}
\tag{2}
$$

where $F(x, u)$ describes the discretized dynamics obtained by applying one step of the explicit RK4 integrator with step-size $h = 0.1$ to (1).

(a) Consider the unconstrained linear quadratic infinite horizon problem that is obtained from (3) by linearizing the dynamics at $x_{\text{lin}} = (0, 0)$, $u_{\text{lin}} = 0$, and dropping the control constraints,

$$
\begin{aligned}
\min_{\substack{x_0, x_1, \dots, \\ u_0, u_1, \dots,}} \quad & \sum_{i=0}^{\infty} \left(x_i^T Q x_i + u_i^T R u_i \right) \\
\text{s.t.} \quad & x_0 = \bar{x}_0 \\
& x_{i+1} = A x_i + B u_i, \ i = 0, 1, \dots,
\end{aligned}
\tag{3}
$$

where $A := \frac{\partial F(x,u)}{\partial x}\big|_{\substack{x=x_{\text{lin}} \\ u=u_{\text{lin}}}}$ and $B := \frac{\partial F(x,u)}{\partial u}\big|_{\substack{x=x_{\text{lin}} \\ u=u_{\text{lin}}}}$.

Complete the template `LQR_design.m` to obtain the LQR gain matrix K, which defines the optimal control at each stage as the time-independent linear feedback law $u_i^*(x) = -Kx$.

Hint: MATLAB function `dlqr()`

(b) Complete the template `dynamic_programming.m` to implement the DP algorithm and use it to compute the cost-to-go associated with the initial state of (3). Choose $N = 20$, $Q = \text{diag}(100, 0.01)$, $R = 0.001$ and Q_N equal to the cost matrix associated with the LQR controller. Discretize the angle θ into 200 values between $-\frac{\pi}{2}$ and 2π. Analogously, discretize the angular velocity into 40 values between -10 and 10 and the torque τ into 20 values between 10 and -10.

Remark: in order to compute the cost-to-go you will have to project the state obtained by simulating the dynamics forward onto the defined discretization grid. To this end, use in your code the MATLAB function `project` *provided with this exercise.*

(c) Consider the plots provided by the previous template, showing the cost of DP and LQR as well as their control policies. Where is the LQR policy similar to the one obtained with DP? Where is it different? Why?

(d) Complete the template `closed_loop.m` to obtain a closed loop simulation of the system, for both LQR and DP. For LQR keep in mind that it was obtained without considering the control constraints, so you have to clip the controls obtained from the LQR feedback law to the feasible control interval $[-10, 10]$. For the DP controller we will always consider the current state of the system as initial state of (3), so you can choose the control according to the cost-to-go function obtained as result of the recursion in (b).

Which of the two controllers achieves the better performance? Why?

2. **Schur Complement Lemma:** Consider the following lemma:

Lemma 1 *(Schur Complement Lemma) Let R be a positive-definite matrix. Then, the following holds:*

$$\min_{u} \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} Q & S^T \\ S & R \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = x^T(Q - S^T R^{-1} S)x \tag{4}$$

and the minimizer $u^*(x)$ is given by $u^*(x) = -R^{-1}Sx$.

(a) Prove Lemma 1.

Abbreviate the objective function as $f(u)$. If we compute the vector-matrix-vector product, we can write it as

$$f(u) = x^\top Q x + 2x^\top S^\top u + u^\top R u$$

with the gradient given by
$$\nabla f(u) = 2Sx + 2Ru.$$

Since $R \succ 0$ (i.e., $f(u)$ is strictly convex in u), the first-order condition $\nabla f(u^*) = 0$ will be sufficient for optimality of u^* and u^* will also be the unique minimizer. From

$$2Sx + 2Ru^* = 0 \quad \Leftrightarrow \quad u^* = -R^{-1}Sx$$

it therefore follows that $u^* = -R^{-1}Sx$ is a (the) minimizer of $f(u)$. Evaluating the corresponding objective value we see that

$$\begin{aligned} f(u^*) &= x^\top Q x + 2x^\top S^\top (R^{-1}Sx) + (x^\top S^\top R^{-1})R(-R^{-1}Sx) \\ &= x^\top Q x - 2x^\top S^\top R^{-1}Sx + x^\top S^\top - R^{-1}Sx \\ &= x^\top(Q - S^\top R^{-1}S)x, \end{aligned}$$

which is identical to the right-hand side of (4), concluding the proof. $\blacksquare$