Problem 1 (4pts):

Piezoelectric nozzles are commonly used in printers for dispensing inks. To eject an ink droplet, a voltage pulse is delivered to a piezoelectric element on the nozzle. The pulse deforms the element and creates pressure waves. If the pressure waveform is suitable, a droplet is fired. The attached file `piezoNozzle.p` simulates the pressure wave $P(t)$ generated when a voltage pulse $V(t)$ is inputed to the nozzle system $\mathcal{G}(V(t))$. For example, in Figure 1, the voltage signal shown generates the corresponding pressure wave. (This figure was generated by running the attached script `piezoNozzleExample.m`).

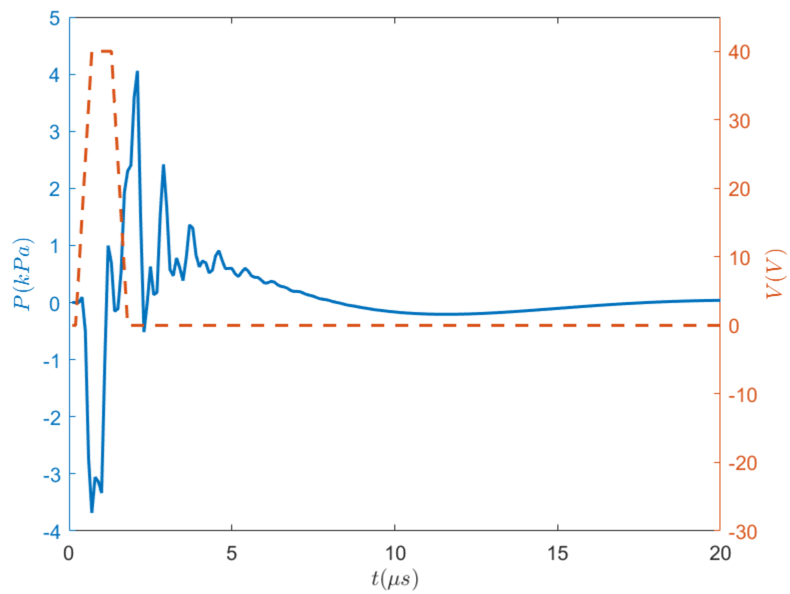


Figure 1: Actuation pulse and nozzle pressure over time.

(a) Is the system input-output operator $\mathcal{G}(V(t))$ linear? Why/why not?

(b) Find what voltage signal $V^*(t)$ we should send in to generate the desired pressure wave P_{ref} given in the file `P_ref.mat`. Plot $V^*(t)$ against time. Also plot the signal $\mathcal{G}(V^*(t))$, superimposed on a plot of P_{ref}.

NOTE: `piezoNozzle.p` is an obfuscated, execute-only file, so you cannot access the content but can run your input through it as demonstrated in `piezoNozzleExample.m`.

Problem 2 (2pts):

Suppose that you are given a discrete-time system

$$x(k+1) = F(x(k), u(k)), \tag{1}$$

and you are asked to use dynamic programming to compute the optimal control law

$$u^*(k) = \phi_k(x(k)),$$

that minimizes the objective

$$J = H(x(2N)) + \sum_{k=0}^{2N} G(x(k), u(k)).$$

Here $2N$ is the length of the control horizon, and $\phi_k(\cdot)$ is the optimal feedback function at time k. Now, suppose that you have two computers that you can use to solve this problem. Your colleague has an idea to break this problem into two sub-problems that can be solved in parallel.
Sub-problem 1: Use dynamic programming to compute the optimal control law

$$u^*(k) = A_k(x(k)), k = 0, \cdots, N-1,$$

that minimizes the objective

$$J_1 = \sum_{k=0}^{N-1} G(x(k), u(k)),$$

under the dynamics given in (1). Here $A_k(\cdot)$ is the optimal feedback function at time k that minimizes J_1.

Sub-problem 2: Use dynamic programming to compute the optimal control law

$$u^*(k) = B_k(x(k)), k = N, \cdots, 2N-1,$$

that minimizes the objective

$$J_2 = H(x(2N)) + \sum_{k=N}^{2N} G(x(k), u(k))$$

under the dynamics given in (1). Here $B_k(\cdot)$ is the optimal feedback function at time k that minimizes J_2.

Noting that $J = J_1 + J_2$ and J_1 only depends on $x(0), \cdots, x(N-1)$ and $u(0), \cdots, u(N-1)$, and J_2 only depends on $x(N), \cdots, x(2N)$ and $u(N), \cdots, u(2N)$, your colleague proposes to concatenate the optimal control laws. That is,

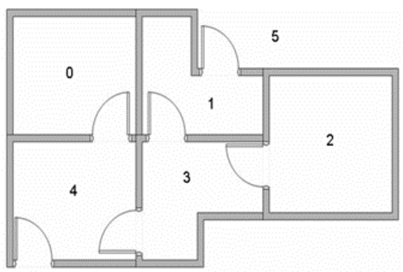
$$\phi_k(\cdot) = \begin{cases} A_k(\cdot), & k = 0, \cdots, N-1, \\ B_k(\cdot), & k = N, \cdots, 2N-1. \end{cases}$$

Question: Will this approach result in an optimal control law that minimizes J ? Explain why/why not?

Problem 3 (4pts):

Consider the following classical reinforcement learning problem. An agent is located in a building with 5 rooms connected by doors as shown in Figure 2a below. The rooms are numbered 0 through 4. The outside of the building can be thought of as one big room and is numbered 5. Only doors in Rooms 1 and 4 lead into the building from Room 5 (outside). For this problem, we want to find the optimal policy that the agent, initially located in any of the rooms (states), should follow to get outside the building (our target room). Suppose that

- the agent cannot stay in the same state from time step to time step, except once it is outside;
- there is a 1-point reward associated with transitioning outside from Rooms 1 or 4 outside, or remaining outside; a -1-point reward associated with attempted transitions between rooms that have no connecting doors; other transitions between rooms have zero reward;
- an action of the agent, would always lead result in the same outcome, that is, the transition probability is unity.



(a) Floor layout. (Source now defunct)

State	Action					
	0	1	2	3	4	5
0					0	
1				0		1
2				0		
3		0	0		0	
4	0			0		1
5		0			0	1

(b) Rewards associated with states-action pairs.

Figure 2: Floor layout and associated rewards.

(a) Find the optimal policy, that is, a table of Q-values associated with the states and actions. Assume a discount factor of 0.8.

(b) What is the optimal policy if the exit door in Room 4 is shut and there is a reward of -0.9 point associated with opening it and taking this exit? Draw the optimal path from room 0 outside based on the policy? Does the optimal path change if the associated reward with taking the exit in room 4 is -0.5 instead?