4F3 cribs

Question 1

(a.i) The motion of the robot is summarised by the following update function f:

- $f(i, j, \mathbf{n}) = (i, j)$ (standing still)
- $f(i, j, \mathbf{u}) = (i-1, j)$ if $i > 1$; $f(i, j, \mathbf{d}) = (i+1, j)$ if $i < 4$ (up and down);
- $f(i, j, \mathbf{r}) = (i, j+1)$ if $j < 4$; $f(i, j, \mathbf{l}) = (i, j-1)$ if $j > 1$ (left and right).

Stage cost:

- given the assumption that every motion take about 1 second, a minimum time path corresponds to a path that reaches the end position with a minimum number of actions. To achieve this, we associate a positive stage cost to every state and action: $c(i, j, \mathbf{a}) = 1$ for all $(i, j) \in X$ and all $\mathbf{a} \in U$;
- the robot will not move to a shaded box if the cost of such move is high. Thus, we change the stage cost as follows: $c(4, 1, \mathbf{r}) = c(4, 3, \mathbf{l}) = c(3, 1, \mathbf{r}) = c(3, 3, \mathbf{l}) = c(2, 2, \mathbf{d}) = c(2, 3, \mathbf{u}) = c(1, 2, \mathbf{r}) = c(1, 4, \mathbf{l}) = 10$.

Terminal cost:

- ending in the wrong position must be penalized therefore we take $J_h(i, j) = 100$ for all $(i, j) \in X$ but $J_h(1, 4) = 0$.

2026

(a.ii) The recursive updates of the cost-to-go are represented as the following sequence of checkboards

100	100	100	0
100	100	100	100
100	100	100	100
100	100	100	100

→

101	101	1	0
101	101	101	1
101	101	101	101
101	101	101	101

→

102	11	1	0
102	102	2	1
102	102	102	2
102	102	102	102

→

12	11	1	0
103	3	2	1
103	103	3	2
103	103	103	3

→

12	4	1	0
4	3	2	1
104	4	3	2
104	104	4	3

→

5	4	1	0
4	3	2	1
5	4	3	2
105	5	4	3

→

5	4	1	0
4	3	2	1
5	4	3	2
6	5	4	3

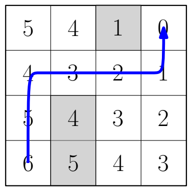
2026

(a.iii) Starting from S, the optimal input sequence is

$$\mathbf{u}, \mathbf{u}, \mathbf{r}, \mathbf{r}, \mathbf{r}, \mathbf{u}\ .$$

The optimal trajectory is

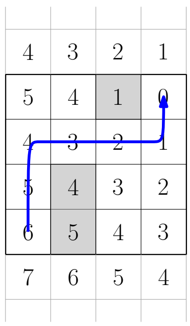
$$S = (4,1) \to (3,1) \to (2,1) \to (2,2) \to (2,3) \to (2,4) \to (1,4) = E\ .$$



The optimal moves are given by $\operatorname{argmin}_{\mathbf{a}\in U}\ c(x,\mathbf{a}) + V(f(x,\mathbf{a}))$. Specifically,

- $x = (4,1)$: $c(4,1,\mathbf{u}) + V(3,1) = 1+5 < c(4,1,\mathbf{r}) + V(4,2) = 10+5$;
- $x = (3,1)$: $c(3,1,\mathbf{u}) + V(2,1) = 1+4 < c(3,1,\mathbf{d}) + V(4,1) = 1+6 < c(3,1,\mathbf{r}) + V(3,2) = 10+4$;
- $x = (2,1)$: $c(2,1,\mathbf{r}) + V(2,2) = 1+3 < c(2,1,\mathbf{d}) + V(3,1) = 1+5 = c(2,1,\mathbf{u}) + V(1,1)$;
- ...

Adding rows at the top and at the bottom does not change the optimal trajectory since the cost-to-go remains unchanged. Any different move would be more expensive.



15%

(b.i) This is extensively discussed in Example 2 of the handout on Optimal Control. Define $\tilde{z} = z - z_E$ and $\tilde{u} = u$. The minimum energy problem can be solved as the limit of the quadratic cost

$$J(\tilde{z}(0), u(\cdot)) = \int_0^6 u(t)^2 dt + \frac{1}{\varepsilon}\tilde{z}(6)^2$$

for $\varepsilon \to 0$. Thus, we need to solve the Riccati equation

$$-\dot{X} = Q + XA + A^T X - XBR^{-1}B^T X$$

for $Q = A = 0$ and $R = B = 1$. Note that X is a scalar. This leads to the equation

$$\dot{X} = X^2$$

that we need to integrate backward from $X(6) = \frac{1}{\varepsilon}$. The minimum energy is then given by

$$J(\tilde{z}(0), u^*(\cdot)) = \tilde{z}(0)^2 X(0) = (z_S - z_E)^2 X(0) = 36X(0)$$

20%

(b.ii) As in Example 2 in the handout, to make the computation easy we take $Y = -X^{-1}$ and we use the identity $-\dot{X} = \frac{d}{dt}(Y^{-1}) = -Y^{-1}\dot{Y}Y^{-1}$ to rewrite the Riccati equation in (b.i) as

$$Y^{-1}\dot{Y}Y^{-1} = Y^{-1}Y^{-1} \quad \to \quad \dot{Y} = 1.$$

Its solution at time T satisfies

$$Y(T) = T + Y(0) \quad \to \quad Y(0) = Y(T) - T \quad \to \quad X(0) = \frac{1}{T - Y(T)} \ .$$

Taking $T = 6$ and $Y(T) = \varepsilon$, for $\varepsilon \to 0$ we get

$$Y(0) = -6 \quad \to \quad X(0) = \frac{1}{6} \ .$$

The optimal cost is thus

$$J(z_S - z_E, u^*(\cdot)) = (z_S - z_E)^2 X(0) = 6$$

for the optimal control

$$u^*(t) = -B^T X(t)\tilde{z} = B^T Y^{-1}(t)\tilde{z} = \frac{1}{(6-t)}(z_E - z(t)) \ .$$

25%

Question 2

(a.i) $\|T_{w\to y}\|_2 = \sqrt{2\pi \text{trace}(B^T L B)}$ where $L = L^T > 0$ is the solution to

$$A_{(k,c)}^T L + L A_{(k,c)} + C^T C = 0$$

for

$$A_{(k,c)} = \begin{bmatrix} 0 & 1 \\ -k & -c \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & 0 \end{bmatrix} .$$

For $L = \begin{bmatrix} \ell_1 & \ell_3 \\ \ell_3 & \ell_2 \end{bmatrix}$ we have

$$0 = A_{(k,c)}^T L + L A_{(k,c)} + C^T C = \begin{bmatrix} -2\ell_3 + 1 & \ell_1 - \ell_3 - \ell_2 \\ \ell_1 - \ell_3 - \ell_2 & 2(\ell_3 - \ell_2) \end{bmatrix}$$

from which

$$\ell_3 = \frac{1}{2} \qquad \ell_2 = \frac{1}{2} \qquad \ell_1 = 1 .$$

Thus,

$$\|T_{w\to y}\|_2 = \sqrt{2\pi \text{trace}(B^T L B)} = \sqrt{2\pi \ell_2} = \sqrt{\pi}$$

- In terms of the impulse response of the system, $g_{w\to y}(t)$, we have $\|g_{w\to y}(t)\| = \frac{1}{\sqrt{2\pi}}\|T_{w\to y}\|_2$ therefore the 2-norm provides a bound on the energy of the impulse response.
- In terms of $\|y\|_\infty$, we have $\|y\|_\infty \leq \frac{1}{\sqrt{2\pi}}\|T_{w\to y}\|_2 \|u\|_2$. Thus, the 2-norm provides a point-wise bound on the largest displacement of the shock absorber for input perturbations of finite energy, 20%

(a.ii) The solution corresponds to the state-feedback $\mathcal{H}_2$ optimal control. We compute the solution $X = X^T > 0$ of the CARE

$$0 = XA + A^T X + C^T C - X B B^T X$$

that also guarantees stability of $A - BB^T X$. Then, $u = -B^T X x$ is the optimal control, that is, optimal stiffness and damping correspond to $\begin{bmatrix} k & c \end{bmatrix} = B^T X$.

For $X = \begin{bmatrix} X_1 & X_3 \\ X_3 & X_2 \end{bmatrix}$ the Riccati equation reads

$$0 = \begin{bmatrix} 1 - X_3^2 & X_1 - X_3 X_2 \\ X_1 - X_3 X_2 & 2X_3 - X_2^2 \end{bmatrix}$$

from which

- $X_3 = \pm 1$;
- $X_2 = \sqrt{2X_3}$. This must be real and positive otherwise X is not positive definite. Thus, $X_3 = 1$ and $X_2 = \sqrt{2}$.
- $X_1 = X_3 X_2 = \sqrt{2}$.

This solution of the Riccati equation gurantees that $A - BB^T X$ is stable. Thus,

$$\left[\begin{array}{cc} k & c \end{array}\right] = B^T X = \left[\begin{array}{cc} 1 & \sqrt{2} \end{array}\right] .$$

For the optimal parameters (not required, provided for completeness),

$$\|T_{w\to z}\|_2 = \sqrt{2\pi \mathrm{trace}(B^T X B)} = \sqrt{2\pi X_3} = \sqrt{2\pi}$$

25%

(a.iii)

$$\|T_{w\to y}\|_2 = \left\| \left[\begin{array}{c} T_{w\to y} \\ 0 \end{array}\right] \right\| \leq \left\| \left[\begin{array}{c} T_{w\to y} \\ 0 \end{array}\right] + \left[\begin{array}{c} 0 \\ T_{w\to u} \end{array}\right] \right\| = \left\| \left[\begin{array}{c} T_{w\to y} \\ T_{w\to u} \end{array}\right] \right\| = \|T_{w\to z}\|_2 .$$

Repeating the computation in (a.i) for $k = 1$ and $c = \sqrt{2}$ we get

$$L = \left[\begin{array}{cc} \ell_1 & \ell_3 \\ \ell_3 & \ell_2 \end{array}\right] = \left[\begin{array}{cc} \frac{3}{2\sqrt{2}} & \frac{1}{2} \\ \frac{1}{2}. & \frac{1}{2\sqrt{2}} \end{array}\right] .$$

Thus,

$$\|T_{w\to y}\|_2 = \sqrt{2\pi \ell_2} = \sqrt{\frac{\pi}{\sqrt{2}}} < \sqrt{\pi} .$$

20%

(b.i) For $k = 1$ and for any $c \geq 2$ we have

$$T_{w\to y}(s) = C(sI - A)^{-1}B = \frac{1}{s(s+c)+1} = \frac{1}{s^2 + cs + 1} .$$

For $c \geq 2$ its poles are real and with negative real part since $s = \frac{-c\pm\sqrt{c^2-4}}{2}$. It follows that

$$|T_{w\to y}(j\omega)| \leq |T_{w\to y}(j0)| = 1 \text{ for all } c \geq 2 .$$

15%

(b.ii) Define $K = \left[\begin{array}{cc} -k & -c \end{array}\right]$. We want to find the matrix K that minimizes the gain from the input w to the output z of the system

$$\dot{x} = (A + BK) + Bw \qquad z = \left[\begin{array}{c} C \\ K \end{array}\right] x .$$

This is achieved by solving the Lyapunov inequality $\dot{V} \leq -z^2 + \gamma^2 w^2$ for $V = x^T X x$ while minimizing $\gamma > 0$. The Lyapunov inequality leads to the matrix inequality

$$\left[\begin{array}{cc} X(A+BK) + (A+BK)^T X + \left[\begin{array}{cc} C^T & K^T \end{array}\right] \left[\begin{array}{c} C \\ K \end{array}\right] & XB \\ B^T X & -\gamma^2 I \end{array}\right] \leq 0 .$$

For $Y = X^{-1}$ we can rewrite

$$\left[\begin{array}{cc} (A+BK)Y + Y(A+BK)^T + \left[\begin{array}{cc} YC^T & YK^T \end{array}\right] \left[\begin{array}{c} CY \\ KY \end{array}\right] & B \\ B^T & -\gamma^2 I \end{array}\right] \leq 0 ,$$

and using $Z = KY$ we get

$$\left[\begin{array}{cc} AY + BZ + YA^T + Z^T B^T + \left[\begin{array}{cc} YC^T & Z^T \end{array} \right] \left[\begin{array}{c} CY \\ Z \end{array} \right] & B \\ B^T & -\gamma^2 I \end{array} \right] \leq 0 \ .$$

Finally, using the Schur complement,

$$\left[\begin{array}{ccc} AY + BZ + YA^T + Z^T B^T & B & \left[\begin{array}{cc} YC^T & Z^T \end{array} \right] \\ B^T & -\gamma^2 I & 0 \\ \left[\begin{array}{c} CY \\ Z \end{array} \right] & 0 & -I \end{array} \right] \leq 0 \ .$$

which is a linear matrix inequality in the unknowns $Y = Y^T > 0$, Z, and $\gamma > 0$. Minimizing over γ returns Y and Z from which the minimizing $\mathcal{H}_\infty$ state-feedback controller reads

$$u = Kx = ZY^{-1}x \ .$$

20%

3) $J_1 = |2x+u_0| + |4x+2u_0+u_1| + \alpha|u_0| + \alpha|u_1|$

If $\alpha > 1 \Rightarrow u_1^* = 0$ else $u_1^* = -4x - 2u_0$

$\alpha < 1$ $\quad |2x+u_0| + \alpha|u_0| + \alpha|4x+2u_0|$

$= |(2+4\alpha)x + (1+2\alpha)u_0| + \alpha|u_0|$

$1+2\alpha > \alpha \Rightarrow \quad u_0^* = \frac{-(2+4\alpha)}{(1+2\alpha)} x = -2x$

$\alpha > 1 \quad |6x+3u_0| + \alpha|u_0| \qquad \Rightarrow u_0^* = \begin{cases} -2x & \alpha < 3 \\ 0 & \alpha > 3 \end{cases}$

$\alpha > 3, \; u_0^* = 0$

$\alpha < 3, \; u_0^* = -2x$

$J_2 = (2x+u_0)^2 + (4x+2u_0+u_1)^2 + \alpha^2 u_0^2 + \alpha^2 u_1^2$

$\frac{\partial J_2}{\partial u_0} = 2(2x+u_0) + 4(4x+2u_0+u_1) + 2\alpha^2 u_0$

$= 20x + (10+2\alpha^2)u_0 + 4u_1 = 0$

$\frac{\partial J_2}{\partial u_1} = 2(4x+2u_0+u_1) + 2\alpha^2 u_1$

$= 8x + 4u_0 + (2+2\alpha^2)u_1 = 0$

$\frac{8 \cdot 4}{2+2\alpha^2} x + \frac{16}{2+2\alpha^2} u_0 + 4u_1 = 0$

$\left(20 - \frac{32}{2+2\alpha^2}\right)x + \left(10+2\alpha^2 - \frac{16}{2+2\alpha^2}\right)u_0^* = 0$

$u_0^* = -\frac{\left(20 - \frac{32}{2+2\alpha^2}\right)}{10+2\alpha^2 - \frac{16}{2+2\alpha^2}} x$

$= -\frac{(8+40\alpha^2)}{4+24\alpha^2+4\alpha^2} x$

$\alpha = 0 \Rightarrow -2x \qquad \alpha = 1 \Rightarrow \frac{-48}{32}$

$J_\infty = \max\left(|2x_0+u_0|, |4x_0+2u_0+u_1|, \alpha|u_0|, \alpha|u_1|\right)$

$$4x_0 + 2u_0 + u_1 = -\alpha u_1$$

$$u_1 = -\frac{4x_0+2u_0}{1+\alpha} = -\frac{2}{1+\alpha}(2x_0+u_0)$$

$\alpha > 1 \quad -\alpha u_0^* = 2x_0 + u_0^* \Rightarrow u_0^* = -\frac{2}{1+\alpha} x_0$

$\alpha < 1 \quad -\alpha u_0^* = \frac{2}{1+\alpha}(2x_0+u_0^*) \Rightarrow$

$$-(\alpha^2+\alpha+2)u_0^* = 4x_0 \Rightarrow \quad u_0^* = -\frac{4}{\alpha^2+\alpha+2} x_0$$

50%

b) For each k use above to minimise cost for $x(k) = x_0$, $x(k+1) = x_1$, $x(k+2) = x_2$ and then apply controller $u(k) = u_0^*$ ignoring the u_1^*. Then move timestep on by 1.

if $u(k) = -K x(k)$ then stable for $1 < K < 3$
Since $x(k+1) = (2-K)\, x(k)$ & need $-1 < 2-K < 1$

J_1 : need $\alpha < 3$

J_2 : need $\frac{8+40\alpha^2}{4+24\alpha^2+4\alpha^4} > 1 \iff 4 + 16\alpha^2 > 4\alpha^4$

$\iff \alpha^4 - 4\alpha^2 - 4 < 0$

$\alpha^2 < 4.82$

$\alpha < 2.2$

40%

J_∞ : need $\alpha < 1$

c) The system is open-loop unstable and so requires a sufficiently large input to be stabilized. Receding horizon control comes with no guarantees of stability. Over penalizing the size of the input results in this not being achieved.

10%

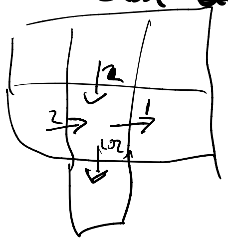
4 a) Sample $s, a \quad r, s', a'$

$Q(s,a) \leftarrow c + Q(s',a')$ if s' not terminal

$\quad\quad\quad\;\; \leftarrow c$ if s' terminal 10%

& repeat, take samples from episodes starting at 4.

(just showing where changes occur – a full answer would be longer than the below



b) i) 1) $Q(4,\rightarrow 5) \leftarrow 1 + Q(5,\rightarrow 6) = \infty$

$Q(5,\rightarrow 6) \leftarrow 1 + 0 = 1$

2) $Q(2,\rightarrow 5) \leftarrow 1 + Q(5,\rightarrow 6) = 2$

3) $Q(4,\rightarrow 5) \leftarrow 1 + Q(5,\rightarrow 6) = 2$

4) $Q(5,\rightarrow 7) \leftarrow 100 + Q(4,\rightarrow 5) = 102$

$Q(4,\rightarrow 5) \leftarrow 1 + Q(5,\rightarrow 6) = 2$ (again)

5) $Q(4,\rightarrow 5) \leftarrow 1 + Q(5,\rightarrow 7) = 102 \Rightarrow$ average of $2, 2, 102 \approx 35$

$Q(5,\rightarrow 7) \leftarrow 100 + 35 = 135$

$Q(4,\rightarrow 5) \leftarrow 2$ again $\Rightarrow$ average of $2, 2, 102, 2 \approx 27$ 30%

ii) All actions are greedy except for all $5, \rightarrow 7$ 10%

iii) Small ϵ, $4 \rightarrow 5 \rightarrow 6 = 2(1-\epsilon-\epsilon^2 \ldots) + 103.\epsilon + 204.\epsilon^2 + \ldots 305\epsilon^3$

$\approx 2 + 101\epsilon + 202\epsilon^2 + 303\epsilon^3$

large ϵ, $4 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 6 = 4$

$101\epsilon \approx 2 \quad \underline{\underline{\epsilon \approx 0.02}} \Rightarrow \epsilon^2$ etc small 30%

(actually equal to 0.0191)

c) Q-learning would find the optimal path $4 \rightarrow 5 \rightarrow 6$ for all ϵ but average episodic cost would be $2 + 101\epsilon + 202\epsilon^2 \ldots$ for all ϵ. 20%