4F3 cribs

Question 1

(a)

$$
\begin{aligned}
V(x,k) &= \min_u \{c(x,u) + V(x_{k+1}, k+1)\} \\
&= \min_u \{c(x,u) + V(f(x) + g(u), k+1)\} \\
V(x,h) &= J_h(x_h)
\end{aligned}
$$

$V(x,k)$ is the value function, i.e. it is the minimum remaining cost from step k onwards, given that $x_k = x$.
Significance: Can be used to derive analytical expressions for the optimal policy (e.g. LQR problem). Can reduce the computational complexity when x_k, u_k take discrete, finite values.

(b)

$$
\begin{aligned}
V(x,t) &= \min_u \{c(x,u)\delta t + V(x+\delta x, t+\delta t)\} + \mathcal{O}((\delta t)^2) \\
&= \min_u \left\{ c(x,u)\delta t + V(x,t) + \frac{\partial V}{\partial x}\delta x + \frac{\partial V}{\partial t}\delta t \right\} + \mathcal{O}((\delta t)^2)
\end{aligned}
$$

Hence

$$
\begin{aligned}
0 &= \min_u \left\{ c(x,u)\delta t + \frac{\partial V}{\partial x}\delta x + \frac{\partial V}{\partial t}\delta t \right\} + \mathcal{O}((\delta t)^2) \\
&= \min_u \left\{ c(x,u) + \frac{\partial V}{\partial x}\dot{x} + \frac{\partial V}{\partial t} \right\} + \frac{\mathcal{O}((\delta t)^2)}{\delta t}
\end{aligned}
$$

Taking the limit $\delta t \to 0$ we get

$$
\begin{aligned}
-\frac{\partial V}{\partial t} &= \min_u \left\{ c(x,u) + \frac{\partial V}{\partial x}(f(x) + g(u)) \right\}, \\
V(T,x) &= J_T(x)
\end{aligned}
$$

(c) (i)

$$
\begin{aligned}
\dot{\tilde{x}}(t) &= \alpha e^{\alpha t} x(t) + e^{\alpha t}\dot{x}(t) \\
&= \alpha \tilde{x}(t) + e^{\alpha t}\dot{x}(t)
\end{aligned}
$$

$$\left[\dot{\tilde{x}}(t) - \alpha\tilde{x}(t)\right] e^{-\alpha t} = -x + u$$

Hence

$$\dot{\tilde{x}}(t) = (\alpha - 1)\tilde{x}(t) + \tilde{u}(t)$$

i.e. this is linear in $\tilde{x}$, $\tilde{u}$. Also $c(x, u) = \tilde{x}^2 + \tilde{u}^2$. Hence the transformed problem is an infinite horizon LQR problem.

(ii) The CARE for this problem is

$$1 + 2X(\alpha - 1) - X^2 = X^2 - 2X(\alpha - 1) - 1 = 0$$

Hence

$$\begin{aligned} X &= \frac{2(\alpha - 1) \pm \sqrt{4(\alpha - 1)^2 + 4}}{2} \\ &= \alpha - 1 \pm \sqrt{(\alpha - 1)^2 + 1} \end{aligned}$$

$X > 0$ for a stabilising controller so choose

$$X = \alpha - 1 + \sqrt{(\alpha - 1)^2 + 1}$$

The controller is given by $\tilde{u}(t) = -X\tilde{x}(t)$, or $u(t) = -Xx(t)$.

Question 2

(a) Comparing with the formulation in the data sheet for an $\mathcal{H}_2$ optimal control problem we have

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$
$$C_1 = \begin{bmatrix} 2 & 2 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Substituting X in CARE we get

$$\begin{bmatrix} \alpha & \beta \\ \beta & \beta \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}+\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \beta & \beta \end{bmatrix}+4\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}-\begin{bmatrix} \alpha & \beta \\ \beta & \beta \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \alpha & \beta \\ \beta & \beta \end{bmatrix}=0$$

Hence

$$\begin{bmatrix} 2\alpha & \alpha+2\beta \\ \alpha+2\beta & 4\beta \end{bmatrix} + 4\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \beta^2\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 0$$

We therefore have

$$\begin{aligned} 2\alpha + 4 - \beta^2 &= 0 \\ \alpha + 2\beta + 4 - \beta^2 &= 0 \\ 4\beta + 4 - \beta^2 &= 0 \end{aligned}$$

Hence from the third equation we have

$$\beta = \frac{4 \pm \sqrt{16+16}}{2} = 2 \pm 2\sqrt{2}$$

and from the first equation we have

$$\alpha = \frac{1}{2}(\beta^2 - 4) = 4(1 \pm \sqrt{2}) = 2\beta$$

(b) Substituting in FARE we get

$$\begin{bmatrix} \gamma & \gamma \\ \gamma & \delta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}+\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \gamma & \gamma \\ \gamma & \delta \end{bmatrix}+4\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}-\begin{bmatrix} \gamma & \gamma \\ \gamma & \delta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \gamma & \gamma \\ \gamma & \delta \end{bmatrix}=0$$

Hence

$$\begin{bmatrix} 4\gamma & \delta+2\gamma \\ \delta+2\gamma & 2\delta \end{bmatrix} + 4\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \gamma^2\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 0$$

Same equations as with CARE. Hence we have $\delta = 2\gamma = 2\beta = \alpha$.

(c) The matrices A_K, B_K, C_K in the state space realization of the controller are (from the data sheet),

$$\begin{aligned} A_K &= A - B_2F - HC_2 \\ B_K &= -H \\ C_K &= F \end{aligned}$$

where $F = B_2^T X$, $H = Y C_2^T$. Substituting the expressions derived for X, Y we have

$$
\begin{aligned}
A_K &= \begin{bmatrix} 1-\beta & 1 \\ -2\beta & 1-\beta \end{bmatrix}, \\
B_K &= -H = -\begin{bmatrix} 1 \\ 1 \end{bmatrix}\beta, \\
C_K &= F = \begin{bmatrix} 1 & 1 \end{bmatrix}\beta
\end{aligned}
$$

The transfer function is

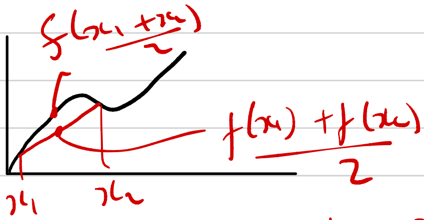
$$
\begin{aligned}
K(s) &= C_K(sI - A_K)^{-1}B_K \\
&= \beta \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} s-(1-\beta) & -1 \\ 2\beta & s-(1-\beta) \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} (-\beta) \\
&= \frac{-\beta^2}{[s-(1-\beta)]^2 + 2\beta} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} s-(1-\beta) & -2\beta \\ 1 & s-(1-\beta) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= \frac{(-\beta^2)(2s-1)}{[s-(1-\beta)]^2 + 2\beta} \\
&= \frac{\beta^2(1-2s)}{s - 2(1-\beta)^2 + 1 + \beta^2}
\end{aligned}
$$

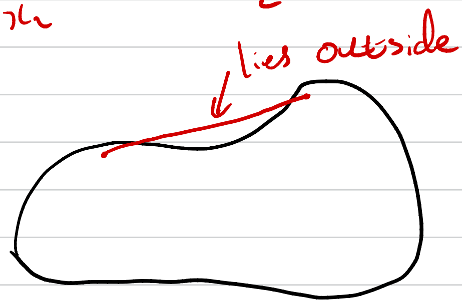
(d) Solve the Riccati equations specified in the data sheet for given γ to find a controller such that $\|T_{w\to\gamma}\|_\infty \le \gamma$. Use a bisection algorithm to find the minimum γ for which there exists a controller such that this $\mathcal{H}_\infty$ bound holds.

3 a) i) $f()$ is convex if $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$

for all x_1, x_2 & $\lambda \in [0,1]$

ii) A set is convex if a line joining any two points in the set always lies within the set





Useful, as a feasible solution (ie one that satisfies the constraints) can be constructed from the average, and will be an improvement.

b) Need $P, Q, R \ge 0$

Since $(\lambda x + (1-\lambda)y)^T Q (\lambda x + (1-\lambda)y) - \lambda x^T Q x - (1-\lambda) y^T Q y$

$= -\lambda(1-\lambda)(x-y)^T Q(x-y) \le 0$

Clearly $M(x_1 + x_2) = Mx_1 + Mx_2$ etc

So no other conditions.

$$H = \begin{bmatrix} R & & & & & \\ & \ddots & & & & \\ & & R & & & \\ & & & Q & & \\ & & & & \ddots & \\ & & & & & Q \\ & & & & & & P \end{bmatrix}$$ etc

& Need to add extra constraints $x_{k+1} = Ax_k + u_k$

So $\begin{bmatrix} I & & A & -I & \cdots \\ & I & & 0 & A & I & \cdots \\ & & \vdots \end{bmatrix} \begin{bmatrix} u_0 \\ \vdots \\ u_{N-1} \\ x_1 \\ \vdots \\ x_N \end{bmatrix} = 0$ etc

c) Need P to satisfy Discrete time ARE for Q & R and find terminal set which is invariant under K & constraint admissible, and then add this as a terminal constraint.

4) a) Q-learning : Initialise $Q(s,a) = 0$, all s,a

Pick s,a and set $Q(s,a) = r(s,a) + \gamma \max_{a'} Q(s',a')$

(ie learning rate = 1)

& then repeat with randomly chosen s,a

MCES: Pick s,a & put $Q(s,a) = r(s,a) + \gamma r(s',a') + \gamma^2 r(s'',a'')$

where $s' = f(s,a)$ and a', s'', a'' etc

are results of following current greedy policy. Also put $Q(s',a') = r(s',a') + \gamma r(s'',a'')$ etc

updating first visit only to each state action pair

b) optimal policy is clearly to take action 1 in state 1 with reward $2(1+\gamma+\gamma^2+\dots) = \frac{2}{1-\gamma} = 10 = V^*(s_1)$

and action 2 in state 2 with reward $-1 + \gamma V^*(s_1) = 7$

$V^*(s_1) = 10 = \max\left(2 + 0.8\,V^*(s_1), 3 + 0.8\,V^*(s_2)\right)$

$V^*(s_2) = 7 = \max\left(-1 + 0.8\,V^*(s_1), 1 + 0.8\,V^*(s_2)\right)$

$\Rightarrow$ optimal.

c)

$s_1, a_1, 2, s_1, a_2, 3, s_2, a_1, 1, s_2, a_1 \dots$ etc

i) $Q(s_1,a_1) = 2 + 0 = 2$

$Q(s_1,a_2) = 3 + 0 = 3$

$Q(s_2,a_1) = 1 + 0 = 1$

$Q(s_2,a_1) = 1 + 0.8 \times 1 = 1.8$ etc

eventually $Q(s_2,a_1) = 5$

ii) $Q(s_1,a_1) = 2 + 0.8 \times 3 + 0.8^2 + 0.8^3 = 2 + 2.4 + 0.64 \times 5 = 7.6$

$Q(s_1,a_2) = 3 + 0.8 + 0.8^2 + \dots = 3 + 0.8 \times 5 = 7$

$Q(s_2, a_1) = 1 + 0.8 + 0.8^2 + \dots = 5$

e) Q-learning a_2 greedy in s_1 $\Rightarrow$
a_1 greedy in s_2

$s_1, a_2, -1, s_2, a_1, 1, s_2, a_1, 1$ etc

MCES a_1 greedy in s_1

$s_1, a_1, 2, s_1, a_1, 2$ etc

$\Rightarrow Q(s_1, a_1) = 10$ ✓

f) MCES converges fastest here, but in general Q-learning is better at using off-policy data