4F3 cribs - 2025

Question 1

(a)

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

The value function $V(x,k)$ is the minimum remaining cost from step k onwards, given that $x_k = x$.
Advantages: 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.
Disadvantages: computational cost increases exponentially with increasing number of states. [10%]

(b) $V(x,0)$ is the minimum cost for optimal control problem formulated with initial condition x.
$V(x,h) = J_h(x)$ is the terminal cost with $x_h = x$. [15%]

(c) (i) Using the candidate value function $V(x,k) = x^2 X_k$ where $X_k > 0$ the Dynamic Programming equation gives

$$\begin{aligned} x^2 X_k &= \min_u \left\{x^2 + u^2 + (\alpha x + u)^2 X_{k+1}\right\} \\ &= x^2 + x^2\alpha^2 X_{k+1} + \min_u \left\{ \left(\sqrt{1+X_{k+1}}u + \frac{\alpha x X_{k+1}}{\sqrt{1+X_{k+1}}}\right)^2 - \frac{(\alpha x X_{k+1})^2}{1+X_{k+1}} \right\} \\ &= x^2 \left(1 - \frac{(\alpha X_{k+1})^2}{1+X_{k+1}} + \alpha^2 X_{k+1}\right) \end{aligned}$$

So the iteration is

$$X_k = \left(1 - \frac{(\alpha X_{k+1})^2}{1+X_{k+1}} + \alpha^2 X_{k+1}\right)$$

with the terminal condition $X_h = \lambda$. Note that $X_k > 0$ for all k also holds with this iteration so $V(x,k) = x^2 X_k$ is a valid candidate value function. [25%]

(ii) We need to solve w.r.t. X the equation

$$X = \left(1 - \frac{(\alpha X)^2}{1+X} + \alpha^2 X\right)$$

which is equivalent to

$$X^2 - \alpha^2 X - 1 = 0$$

This has solutions

$$X = \frac{\alpha^2 \pm \sqrt{\alpha^4 + 4}}{2}$$

and we choose the positive solution

$$X = \frac{\alpha^2 + \sqrt{\alpha^4 + 4}}{2}$$

The optimal cost is $V(x_0, 0) = (x_0)^2 X$. [10%]

(iii) $X_1 = 1/2$, and from the iteration we have $X_0 = 13/12$.
The optimal cost is $J = (x_0)^2 X_0 = (1/2)^2 \times (13/12) = 13/48$ [10%]

(d) Note that $x^* = r$, $u^* = r(1-\alpha)$ is an equilibrium point of the system. Using the transformation $\tilde{u} = u - u^*$, $\tilde{x} = x - x^*$ the system becomes

$$\tilde{x}_{k+1} = \alpha \tilde{x}_k + \tilde{u}_k$$

and the cost is

$$\sum_{k=0}^{\infty} (\tilde{x}_k^2 + \tilde{u}_k^2)$$

i.e. the formulation is as in part (c)(ii) and hence the value of X is the same as in that part. The optimal control input is

$$\tilde{u}_k = -\alpha \frac{X}{1+X} \tilde{x}_k$$

which is equivalent to

$$u_k = u^* - \alpha \frac{X}{1+X}(x_k - x^*)$$

The cost function considered is relevant when x^* is a desired operating point for the system. [30%]

Question 2

(a) The $\mathcal{H}_2$ norm of G, denoted as $\|G\|_2$, is defined as

$$\|G\|_2^2 = \int_{-\infty}^{\infty} \text{trace}\left\{\overline{G(j\omega)}^T G(j\omega)\right\} \, \mathrm{d}\omega$$

The $\mathcal{H}_\infty$ norm of G, $\|G\|_\infty$, can be defined in two equivalent ways:

- $\|G\|_\infty = \sup_\omega \overline{\sigma}(G(jw))$
- $\|G\|_\infty = \sup_{\hat{u}\neq 0} \frac{\|G\hat{u}\|_2}{\|\hat{u}\|_2} = \sup_{u\neq 0} \frac{\|y\|_2}{\|u\|_2}$ where $\hat{y}(s) = G(s)\hat{u}(s)$

[10%]

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

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

Denoting

$$X = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix}$$

and substituting in CARE we get

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

Hence

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

We therefore have

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

Then denoting

$$Y = \begin{bmatrix} \tilde{\alpha} & \tilde{\beta} \\ \tilde{\beta} & \tilde{\gamma} \end{bmatrix}$$

and substituting in FARE we get

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

This gives

$$\begin{aligned}
1 - \tilde{\beta}^2 = 0 \quad &\Longrightarrow \tilde{\beta} = 1 \\
\tilde{\gamma} - \tilde{\beta}\tilde{\alpha} = 0 \quad &\Longrightarrow \tilde{\gamma} = \sqrt{2} \\
2\tilde{\beta} - \tilde{\alpha}^2 = 0 \quad &\Longrightarrow \tilde{\alpha} = \sqrt{2}
\end{aligned}$$

The matrices A_K, B_K, C_K in the state space realization of the controller are then as specified in the data sheet, i.e.

$$\begin{aligned}
A_K &= A - B_2 F - HC_2 \\
B_K &= -H \\
C_K &= F
\end{aligned}$$

where $F = B_2^T X$, $H = YC_2^T$.

[40%]

(c) (i) This is a full information $\mathcal{H}_\infty$ control problem. Hence we need to solve the corresponding Riccati equation in the data sheet with $\gamma = \sqrt{2}$. Denoting

$$X = \begin{bmatrix} \hat{\alpha} & \hat{\beta} \\ \hat{\beta} & \hat{\gamma} \end{bmatrix}$$

we have

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

Hence

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

We therefore have

$$\begin{aligned}
4 - \frac{1}{2}\hat{\beta}^2 = 0 \quad &\Longrightarrow \hat{\beta} = \sqrt{8} \\
\hat{\alpha} - \frac{1}{2}\hat{\beta}\hat{\gamma} = 0 \quad &\Longrightarrow \hat{\alpha} = 8^{3/4} \\
2\hat{\beta} - \frac{1}{2}\hat{\gamma}^2 = 0 \quad &\Longrightarrow \hat{\gamma} = 2 \times 8^{1/4}
\end{aligned}$$

The controller is of the form $u = -Fx$ where $F = B_2^T X$

[30%]

(ii) Use a bisection algorithm to find the minimum γ for which the Riccati equation has a solution. [10%]

(iii) No, since the Riccati equaiton does not have a positive definite solution. [10%]

Question 3

(a) It is better to penalize Δu_i because this allows u_i to evolve towards a nonzero constant, which is not penalized. Notice u_i must be nonzero to keep y_i near the setpoint y_s. **[10%]**

(b) For $N = 2$, we have

$$M_x = \begin{bmatrix} C^\intercal Q C & 0 \\ 0 & C^\intercal P C \end{bmatrix}, \qquad f_x = \begin{bmatrix} -2y_s^\intercal Q C \\ -2y_s^\intercal P C \end{bmatrix}$$
$$M_u = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \qquad f_u = \begin{bmatrix} -2u(k-1) \\ 0 \end{bmatrix}$$
$$\text{constant} = y_s^\intercal Q y_s + y_s^\intercal P y_s + u(k-1)^\intercal u(k-1)$$

[25%]

(c) The system can be written as

$$x_1 - Bu_0 = Ax_0 = Ax(k)$$
$$x_2 - Ax_1 - Bu_1 = 0$$

or equivalently

$$\underbrace{\begin{bmatrix} I & 0 & -B & 0 \\ -A & I & 0 & -B \end{bmatrix}}_{\tilde{A}} \begin{bmatrix} x_1 \\ x_2 \\ u_0 \\ u_1 \end{bmatrix} = \underbrace{\begin{bmatrix} Ax(k) \\ 0 \end{bmatrix}}_{\tilde{B}}$$

while the constraints can be written as

$$-1 \leq u_0 - u(k-1) \leq 1$$
$$-1 \leq u_1 - u_0 \leq 1$$

or equivalently

$$\underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}}_{G} \begin{bmatrix} x_1 \\ x_2 \\ u_0 \\ u_1 \end{bmatrix} \leq \underbrace{\begin{bmatrix} 1 + u(k-1) \\ 1 - u(k-1) \\ 1 \\ 1 \end{bmatrix}}_{h}$$

where in the last equation the inequality is element-wise.

It follows that the control problem is equivalent to

$$\min_\theta \quad \theta^\intercal \begin{bmatrix} M_x & 0 \\ 0 & M_u \end{bmatrix} \theta + \begin{bmatrix} f_x^\intercal & f_u^\intercal \end{bmatrix} \theta$$

subject to

$$\tilde{A}\theta = \tilde{B}$$
$$G\theta \leq h$$

which is in the standard QP form. **[25%]**

(d) The QP is strictly convex if M_x and M_u above are positive definite. It can be readily verified that M_u is positive definite. This can be done by computing its eigenvalues, or by checking that the principal minors of M_u ($[M_u]_{11} = 2$ and $\det M_u = 1$) are both positive. M_x is positive definite it $C^\mathsf{T}QC$ and $C^\mathsf{T}PC$ are both positive definite. This will be satisfied if Q and P are positive definite and C is full column rank (a rather restrictive assumption). If the QP is strictly convex, then the optimal MPC law $u(k) = u_0^*(x(k), u(k-1))$ is unique. **[15%]**

(e) We can eliminate x_1 and x_2 from the decision variables by writing

$$\begin{aligned}\theta_x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} Ax(k) & Bu_0 \\ A(A(x(k) + Bu_0) & Bu_1 \end{bmatrix} \\ &= \begin{bmatrix} B & 0 \\ AB & B \end{bmatrix} \theta_u + \begin{bmatrix} Ax(k) \\ A^2x(k) \end{bmatrix}\end{aligned}$$

Hence

$$\theta_x^\mathsf{T} M_x \theta_x = \theta_u^\mathsf{T} \begin{bmatrix} CB & 0 \\ CAB & CB \end{bmatrix}^\mathsf{T} \begin{bmatrix} Q & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} CB & 0 \\ CAB & CB \end{bmatrix} \theta_u + \text{non-quadratic terms}$$

The QP is now only in the decision variables θ_u. For it to be strictly convex, it is sufficient that Q and P be positive definite. **[25%]**

Question 4

(a) A simple cost function that satisfies the requirements is

$$J(s_0) = \int_0^\infty \underbrace{(x_f - x(t))^2 + (y_s - y(t))^2 + \frac{M}{2}v(t)^2 + \frac{J}{2}\omega(t)^2}_{r(s,a)} \, dt$$

The problem is episodic because even though it has an infinite time horizon, there is a stopping set S that can be reached in finite time and for which there is an action $a(t)$ that guarantees $r(s(t), a(t)) = 0$ and $s(t) \in S$ after the stopping set is reached. In our case,

$$S = \{(x, u, \theta) \in \mathbb{R}^3 \,|\, x = x_f, y = y_f\}$$

and for $a = (v, \omega) = (0, 0)$ we have

$$\begin{aligned} s(t_0) \in S, a = (0,0) &\implies \dot{s}(t) = 0, r(s(t), a(t)) = 0 \\ &\implies s(t) \in S, r(s(t), a(t)) = 0 \end{aligned}$$

for all $t \geq t_0$. **[20%]**

(b) The state and action spaces are continuous, and hence very large. We cannot apply tabular RL, and must instead rely on more advanced methods (e.g., actor-critic methods) that use a function approximation for the action-value function. Such methods can be difficult to fine tune due to their sensitivity to hyperparameter choices. **[10%]**

(c) Choosing $\mu = \sqrt{2}$ implies that, starting from $(x, y\theta) = (0, 0, 0)$, the robot can only move between integer values of x, y; furthermore, it can only face directions given by $\theta = k\pi/4$ for some k. This means that the state space is discrete (countable) and the action space is finite. We can apply tabular RL.

The state transitions are easily found by integrating the states while each action is taken. We have

$$s(t+1) = \begin{cases} s(t) & \text{if } a(t) = a_0 \\ (x(t), y(t), \theta(t) + \pi/4) & \text{if } a(t) = a_1 \\ (x(t), y(t), \theta(t) - \pi/4) & \text{if } a(t) = a_2 \\ (x(t) + \cos\theta(t), y(t) + \sin\theta(t), \theta(t)) & \text{if } a(t) = a_3, \theta(t) = k\pi/2 \\ (x(t) + \sqrt{2}\cos\theta(t), y(t) + \sqrt{2}\sin\theta(t), \theta(t)) & \text{if } a(t) = a_3, \theta(t) = k\pi/2 + \pi/4 \end{cases}$$

(alternatively, the state transitions above can be explained with a diagram of the possible movements of the robot).

Hence we can reformulate the problem with the cost

$$\sum_{t=0}^{\infty} (x(t) - x_f)^2 + (y(t) - y_f)^2 + c(a(t), \theta(t))$$

with

$$c(a, \theta) = \begin{cases} 0, & a = a_0 \\ \int_0^1 \frac{J}{2}\left(\frac{\pi}{4}\right)^2 dt = J\pi^2/32, & a = a_1, a_2 \\ \int_0^1 \frac{M}{2} 1^2 dt = \frac{M}{2}, & a = a_3, \theta = k\pi/2 \\ \int_0^1 \frac{M}{2}(\sqrt{2})^2 dt = M, & a = a_3, \theta = k\pi/2 + \pi/4 \end{cases}$$

Notice the problem is still episodic since the target states are integer valued. **[35%]**

(d) Q-learning works by updating an initial guess of the action-value function $Q_0(s, a)$ according to

$$Q_{t+1}(s(t), a(t)) = (1 - \alpha)Q_t(s(t), a(t)) + \alpha(r(t) + \min_a Q_t(s(t + 1), a))$$

if $(s(t), a(t))$ is visited at time t during an observed trajectory of the robot, and

$$Q_{t+1}(s(t), a(t)) = Q_t(s(t), a(t))$$

if $(s(t), a(t))$ is not visited at that time. The action $a(t)$ is obtained by sampling actions from the ϵ-greedy policy.

The problem considers $\alpha = 0.5$, $s(0) = (0, 0, 0)$ and the action-value is initialized at zero for all a, s. Assuming $a(0) = a_1$, $a(1) = a_2$, and $a(2) = a$, the first three steps are given by:

Step 1:

$$\begin{aligned} s(0) &= (0, 0, 0) \\ a(0) &= a_1 \\ r(0) &= 5^2 + 4^2 + J\pi^2/32 \\ s(1) &= (0, 0, \pi/4) \end{aligned}$$

hence

$$Q_1((0, 0, 0), a_1) = 0.5(r(0) + 0)$$

Step 2:

$$\begin{aligned} s(1) &= (0, 0, \pi/4) \\ a(1) &= a_2 \\ r(1) &= 5^2 + 4^2 + J\pi^2/32 \\ s(2) &= (0, 0, 0) \end{aligned}$$

hence

$$Q_2((0, 0, 0), a_2) = 0.5(r(1) + 0)$$

Step 2:

$$\begin{aligned} s(2) &= (0, 0, 0) \\ a(2) &= a_1 \\ r(2) &= 5^2 + 4^2 + J\pi^2/32 \\ s(3) &= (0, 0, \pi/4) \end{aligned}$$

hence

$$Q_3((0, 0, 0), a_1) = 0.5r(0) + 0.5(r(2) + 0)$$

[35%]