

**Exam Part**

---

**C8.7 Optimal Control  
Mock Exam**

---

**June 8, 2026**

**Do not turn this page until you are told that you may do so**

1. In this question we will consider a stochastic optimal control problem, where a controlled process  $X$  takes values in  $\mathbb{R}$ , with controlled dynamics

$$dX_t = f(t, X_t, U_t)dt + \sigma(t, X_t, U_t)dW_t; \quad X_0 = x_0.$$

The control  $U$  is a progressively measurable process taking values in a topological space  $\mathcal{U}$ , chosen to minimize the expected cost

$$\mathbb{E} \left[ \int_0^T g(t, X_t, U_t)dt + \Phi(X_T) \middle| \mathcal{F}_t \right]$$

You may assume whenever needed that all functions are sufficiently integrable and continuous that the state dynamics and expected costs are well defined.

- (a) Write down the Hamilton–Jacobi–Bellman equation satisfied by the value function for the control problem, including appropriate boundary conditions.
- (b) (i) Suppose you are given a  $(C^{1,2}$ , polynomial growth) function  $v : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $v(\cdot, T) = \Phi(\cdot)$ , and a  $(C^{1,2})$  function  $u^* : [0, T] \times \mathbb{R} \rightarrow \mathcal{U}$ . If the running cost is written in the form

$$g(t, x, u) = \gamma(t, x, u) - \partial_t v(t, x) - f(t, x, u)\partial_x v(t, x) - \frac{1}{2}(\sigma(t, x, u))^2 \partial_{xx}^2 v(t, x)$$

for some function  $\gamma$ , give a necessary and sufficient condition on  $\gamma$  such that  $v$  is the value function of the control problem, and  $u^*$  is the optimal control.

- (ii) Write down a control problem where  $v(t, x) = \sin(x) + \frac{t}{1+t}$  and  $u^*(t, x) = \cos(tx)$ , when  $\mathcal{U} = \mathbb{R}$ ,  $f(t, x, u) = -u^2x$ , and  $\sigma(x) = 2x$ .
- (c) Now suppose that  $f(t, x, u) = g(t, x, u) = 0$ , that  $\Phi$  is  $C^2$  and convex, and that  $\sigma$  depends continuously on  $\mathcal{U}$ . Derive the optimal strategy.

Solution:

- (a) Let

$$v(t, x) := \inf_U \mathbb{E} \left[ \int_t^T g(s, X_s, U_s) ds + \Phi(X_T) \middle| X_t = x \right].$$

The Hamilton–Jacobi–Bellman equation is

$$\partial_t v(t, x) + \inf_{u \in \mathcal{U}} \left\{ f(t, x, u)\partial_x v(t, x) + \frac{1}{2}\sigma^2(t, x, u)\partial_{xx}^2 v(t, x) + g(t, x, u) \right\} = 0,$$

for  $(t, x) \in [0, T) \times \mathbb{R}$ , with terminal condition  $v(T, x) = \Phi(x)$ .

- (b) (i) Substituting

$$g(t, x, u) = \gamma(t, x, u) - \partial_t v(t, x) - f(t, x, u)\partial_x v(t, x) - \frac{1}{2}\sigma^2(t, x, u)\partial_{xx}^2 v(t, x)$$

into the HJB equation yields

$$\partial_t v + f\partial_x v + \frac{1}{2}\sigma^2\partial_{xx}^2 v + g = \gamma.$$

Hence the HJB equation for  $v$  becomes

$$\inf_{u \in \mathcal{U}} \gamma(t, x, u) = 0.$$

Therefore a necessary and sufficient condition for  $v$  to be the value function and  $u^*$  to be an optimal feedback control is

$$\gamma(t, x, u) \geq 0 \quad \forall(t, x, u),$$

and

$$\gamma(t, x, u^*(t, x)) = 0 \quad \forall(t, x).$$

Equivalently,

$$u^*(t, x) \in \arg \min_{u \in \mathcal{U}} \gamma(t, x, u)$$

and the minimum value is 0.

(ii) We have

$$v(t, x) = \sin(x) + \frac{t}{1+t}, \quad u^*(t, x) = \cos(tx),$$

with

$$f(t, x, u) = -u^2x, \quad \sigma(t, x, u) = 2x.$$

The derivatives of  $v$  are

$$\partial_t v(t, x) = \frac{1}{(1+t)^2}, \quad \partial_x v(t, x) = \cos(x), \quad \partial_{xx} v(t, x) = -\sin(x).$$

Choose

$$\gamma(t, x, u) = (u - \cos(tx))^2.$$

Then

$$\begin{aligned} g(t, x, u) &= \gamma(t, x, u) - \partial_t v(t, x) - f(t, x, u) \partial_x v(t, x) - \frac{1}{2} \sigma^2(t, x, u) \partial_{xx} v(t, x) \\ &= (u - \cos(tx))^2 - \frac{1}{(1+t)^2} + u^2x \cos(x) + 2x^2 \sin(x). \end{aligned}$$

Thus one such control problem is

$$dX_t = -U_t^2 X_t dt + 2X_t dW_t,$$

with terminal cost

$$\Phi(x) = v(T, x) = \sin(x) + \frac{T}{1+T},$$

and running cost

$$g(t, x, u) = (u - \cos(tx))^2 - \frac{1}{(1+t)^2} + u^2x \cos(x) + 2x^2 \sin(x).$$

(c) Since  $f = g = 0$ , the HJB equation is

$$\partial_t V(t, x) + \inf_{u \in \mathcal{U}} \left\{ \frac{1}{2} \sigma^2(u) \partial_{xx} V(t, x) \right\} = 0, \quad V(T, x) = \Phi(x).$$

Let

$$a_* := \min_{u \in \mathcal{U}} \sigma^2(u),$$

and choose  $u^* \in \arg \min_{u \in \mathcal{U}} \sigma^2(u)$ .

We claim that  $u^*$  is optimal. If  $u^*$  is used, then the HJB equation becomes

$$\partial_t V + \frac{1}{2} a_* \partial_{xx} V = 0, \quad V(T, x) = \Phi(x).$$

To justify that this indeed corresponds to the minimizer in the HJB, we show that  $V(t, \cdot)$  is convex. Differentiating the above equation twice with respect to  $x$ , and writing

$$W(t, x) := \partial_{xx}V(t, x),$$

gives

$$\partial_t W + \frac{1}{2}a_* \partial_{xx}W = 0, \quad W(T, x) = \Phi''(x).$$

Since  $\Phi$  is convex,  $\Phi''(x) \geq 0$ . By the maximum principle for the heat equation,

$$W(t, x) \geq 0 \quad \text{for all } (t, x),$$

and hence  $\partial_{xx}V(t, x) \geq 0$ .

Therefore  $V(t, \cdot)$  is convex, so the quantity  $\frac{1}{2}\sigma^2(u) \partial_{xx}V(t, x)$  is minimized by choosing  $\sigma^2(u)$  as small as possible. Thus any control satisfying  $u_t^* \in \arg \min_{u \in \mathcal{U}} \sigma^2(u)$  for all  $t$  is optimal.

2. In this question we will consider an undiscounted discrete time, deterministic, finite horizon control problem, where the control takes values in an open set  $\mathcal{U}$ . We have a continuous state variable  $X$ , taking values in  $\mathcal{X} = \mathbb{R}$ , with initial state  $X_0 = x_0$ , and one-step controlled dynamics

$$X_{t+1} = X_t + f(t, x, U_t)$$

There is a continuous running cost function  $g : \mathbb{T} \times \mathcal{X} \times \mathcal{U}$ , and a terminal cost function  $\Phi$ .

- (a) Explain what is meant by a pasting of a control process.
- (b) Assuming that  $f$ ,  $g$  and  $\Phi$  are Lipschitz continuous (uniformly in  $u$ ), construct an  $\varepsilon$ -optimal control for this problem, and hence prove the the value function satisfies the Bellman equation.
- (c) Using the Bellman equation (which you may assume has a differentiable solution) or otherwise, show that the optimal control satisfies a discrete version of Pontryagin's principle, that is:

$$\begin{aligned} X_{t+1} &= X_t + f(t, X_t, U_t^*); & X_0 &= x_0; \\ q_t &= q_{t+1} (1 + \partial_x f(t, X_t, U_t^*)) + \partial_x g(t, X_t, U_t^*); & q_T &= \frac{d\Phi}{dx}(X_T); \\ 0 &= \partial_u g(t, X_t, U^*) + f(t, X_t, U^*) q_{t+1}. \end{aligned}$$

- (d) Explain the key advantages and disadvantages of using Pontryagin's principle, rather than the Bellman equation, to compute the optimal control.

Solution:

- (a) Bookwork
- (b) We know that  $v(T, x) = \Phi(x)$  is Lipschitz. Hence

$$g(t, x, u) + v(t+1, x + f(t, x, u))$$

is also Lipschitz in  $x$ , uniformly in  $u$ . Taking the set  $x = k\delta$ , for fixed  $\delta > 0$  and  $k \in \mathbb{Z}$ , we can find an  $\varepsilon$ -optimal control  $u_{k\delta}$ , that is, a control such that

$$g(t, k\delta, u_{k\delta}) + v(t+1, k\delta + f(t, k\delta, u)) \leq \min_{u \in \mathcal{U}} \left\{ g(t, k\delta, u) + v(t+1, k\delta + f(t, k\delta, u)) \right\} + \varepsilon.$$

By Lipschitz continuity, there exists  $c > 0$  such that

$$g(t, x, u_{k\delta}) + v(t+1, x + f(t, x, u)) \leq \min_{u \in \mathcal{U}} \left\{ g(t, k\delta, u_{k\delta}) + v(t+1, k\delta + f(t, k\delta, u)) \right\} + (1+c)\varepsilon$$

for all  $x \in [k\delta, (k+1)\delta)$ . Defining the pasted control

$$u^\varepsilon(t, x) = u_{k\delta} \text{ for } x \in [k\delta, (k+1)\delta),$$

we have a control which is near-optimal at time  $t$ . If we do this for all  $t$ , we obtain a total error at most  $T(1+c)\varepsilon$ , that is,

$$J(t, x, u^\varepsilon) \leq v(t, x) + (T-t)(1+c)\varepsilon$$

showing that we have an  $\varepsilon$ -optimal control (after adjusting constants).

As the control we have constructed is feedback, we know by induction that the corresponding cost depends only on  $t$  and  $x$ . Therefore, taking  $\varepsilon \rightarrow 0$ , we obtain the Bellman equation

$$v(t, x) = \min_{u \in \mathcal{U}} \left\{ g(t, x, u) + v(t+1, x + f(t, x, u)) \right\}.$$

- (c) See notes
- (d) Advantage of Pontryagin: only have to solve a forward-backward pair of ODEs, rather than a PDE. Disadvantage: Not guaranteed to find an optimizer if the Hamiltonian isn't convex (only a necessary condition for optimality).

3. Consider an infinite-horizon, discounted, discrete-time, finite-state stochastic control problem. The state has transition dynamics

$$\mathbb{P}(X_{t+1} = x' | X_t = x, U_t = u) = p(x'; x, u)$$

and the goal is to minimize the cost

$$J(U, x) = \sum_{t \geq 0} e^{-\rho t} g(X_t, U_t)$$

for a discount rate  $\rho > 0$ , initial state  $X_0 = x$  and cost function  $g$ . The controls take values from a compact set, and  $p$  and  $g$  are both continuous with respect to the control.

- (a) Define the Bellman evaluation and optimality operators  $\mathcal{T}_u$  and  $\mathcal{T}$ . Explain how they relate to the cost-to-go  $J(U, x)$  and the value function.
- (b) (i) Define the standard policy iteration algorithm for solving the control problem.  
(ii) Consider the variation of policy iteration where the evaluation procedure remains unchanged, but at each improvement stage a state is chosen uniformly at random, and the action chosen in this state is optimized (and the actions chosen in other states are left unchanged). Show that the value function approximations resulting from this procedure are (componentwise) decreasing, and that they converge to the true value function.  
(You should prove any properties of  $\mathcal{T}$  and  $\mathcal{T}_u$  which you need to use.)
- (c) Consider now an agent with a state-dependent discount rate, that is, who wishes to optimize

$$J(U, x) = \sum_{t \geq 0} \exp\left(-\sum_{0 < s \leq t} \rho(X_s)\right) g(X_t, U_t)$$

for some given function  $\rho : \mathcal{X} \rightarrow (0, \infty)$ .

- (i) Write down a Bellman equation satisfied by the corresponding value function.  
(ii) Show that the Bellman equation admits a bounded solution, which can be approximated by finite-horizon problems.

Solution:

- (a) Let  $\beta := e^{-\rho} \in (0, 1)$ . For each stationary deterministic policy  $u : \mathcal{X} \rightarrow \mathcal{U}$ , define the *Bellman evaluation operator*

$$(\mathcal{T}_u v)(x) := g(x, u(x)) + \beta \sum_{x' \in \mathcal{X}} p(x'; x, u(x)) v(x'), \quad x \in \mathcal{X},$$

for any bounded function  $v : \mathcal{X} \rightarrow \mathbb{R}$ .

Define the *Bellman optimality operator*

$$(\mathcal{T}v)(x) := \min_{u \in \mathcal{U}} \left\{ g(x, u) + \beta \sum_{x' \in \mathcal{X}} p(x'; x, u) v(x') \right\}, \quad x \in \mathcal{X}.$$

The minimum is attained because  $\mathcal{U}$  is compact and  $g, p$  are continuous in  $u$ .

For a fixed policy  $u$ , the cost-to-go  $J(u, \cdot)$  is the unique fixed point of  $\mathcal{T}_u$ :

$$J(u, \cdot) = \mathcal{T}_u J(u, \cdot).$$

Indeed,  $J(u, x)$  satisfies the one-step dynamic programming identity

$$J(u, x) = g(x, u(x)) + \beta \sum_{x'} p(x'; x, u(x)) J(u, x').$$

The value function

$$V(x) := \inf_U J(U, x)$$

is the unique fixed point of  $\mathcal{T}$ :

$$V = \mathcal{T}V.$$

Moreover, both  $\mathcal{T}_u$  and  $\mathcal{T}$  are monotone and contractions in the sup norm with modulus  $\beta$ :

$$\|\mathcal{T}_u - \mathcal{T}_u w\|_\infty \leq \beta \|v - w\|_\infty, \quad \|\mathcal{T}v - \mathcal{T}w\|_\infty \leq \beta \|v - w\|_\infty.$$

(b) (i) The standard policy iteration algorithm is as follows.

1. Choose an initial stationary policy  $u_0 : \mathcal{X} \rightarrow \mathcal{U}$ .
2. Given  $u_n$ , evaluate it by solving

$$v_n = \mathcal{T}_{u_n} v_n,$$

so  $v_n = J(u_n, \cdot)$ .

3. Improve the policy state by state:

$$u_{n+1}(x) \in \arg \min_{u \in \mathcal{U}} \left\{ g(x, u) + \beta \sum_{x'} p(x'; x, u) v_n(x') \right\}, \quad x \in \mathcal{X}.$$

4. Repeat until  $u_{n+1} = u_n$ , in which case  $u_n$  is optimal.

(ii) Let  $u_n$  be the current policy and  $v_n = J(u_n, \cdot)$  its value function. In the randomized variant, choose a state  $x_n \in \mathcal{X}$  uniformly at random, and define the new policy  $u_{n+1}$  by

$$u_{n+1}(x_n) \in \arg \min_{u \in \mathcal{U}} \left\{ g(x_n, u) + \beta \sum_{x'} p(x'; x_n, u) v_n(x') \right\},$$

while

$$u_{n+1}(x) = u_n(x), \quad x \neq x_n.$$

We first prove the key operator facts.

**Monotonicity.** If  $v \leq w$  componentwise, then

$$\mathcal{T}_u v \leq \mathcal{T}_u w, \quad \mathcal{T}v \leq \mathcal{T}w.$$

**Contraction.** For every bounded  $v, w$ ,

$$\|\mathcal{T}_u v - \mathcal{T}_u w\|_\infty \leq \beta \|v - w\|_\infty, \quad \|\mathcal{T}v - \mathcal{T}w\|_\infty \leq \beta \|v - w\|_\infty.$$

Hence each  $\mathcal{T}_u$  and  $\mathcal{T}$  has a unique fixed point.

Now let  $u_{n+1}$  be the policy obtained from  $u_n$  by improving only the chosen state  $x_n$ . Then

$$\mathcal{T}_{u_{n+1}} v_n \leq v_n$$

componentwise, because at  $x_n$  the action is chosen to minimize the right-hand side, and at all other states the policy is unchanged. Since  $\mathcal{T}_{u_{n+1}}$  is monotone,

$$\mathcal{T}_{u_{n+1}}^k v_n$$

is a decreasing sequence in  $k$ , and its limit is the fixed point  $v_{n+1} = J(u_{n+1}, \cdot)$ . Therefore

$$v_{n+1} \leq v_n \quad \text{componentwise.}$$

So the value iterates are componentwise decreasing.

Also, since  $V$  is the optimal value function, every policy value dominates it:

$$V \leq v_n \quad \text{for all } n.$$

Thus, for each state  $x$ , the decreasing sequence  $v_n(x)$  is bounded below by  $V(x)$ , so it converges. Let

$$v_n \downarrow v_\infty$$

componentwise.

It remains to show  $v_\infty = V$ . Suppose not. Then, since  $V = \mathcal{T}V$  and  $\mathcal{T}$  is continuous, there exists a state  $x^*$  and  $\varepsilon > 0$  such that

$$(\mathcal{T}v_\infty)(x^*) \leq v_\infty(x^*) - 2\varepsilon.$$

By continuity of  $\mathcal{T}$ , for all sufficiently large  $n$ ,

$$(\mathcal{T}v_n)(x^*) \leq v_n(x^*) - \varepsilon.$$

Because the states are chosen uniformly at random,  $x^*$  is selected infinitely often almost surely. Hence, almost surely, for infinitely many  $n$  the improvement step is applied at  $x^*$ , and then the new policy satisfies

$$v_{n+1}(x^*) \leq (\mathcal{T}v_n)(x^*) \leq v_n(x^*) - \varepsilon.$$

This contradicts the convergence  $v_n(x^*) \downarrow v_\infty(x^*)$ . Therefore

$$v_\infty = V.$$

So the randomized single-state policy iteration produces a componentwise decreasing sequence of value functions converging almost surely to the true value function.

(c) (i) Write

$$\beta(x) := e^{-\rho(x)} \in (0, 1).$$

Since  $\mathcal{X}$  is finite, we have

$$\beta_* := \max_{x \in \mathcal{X}} \beta(x) < 1.$$

The Bellman equation for the value function  $V$  is

$$V(x) = \min_{u \in \mathcal{U}} \left\{ g(x, u) + \sum_{x' \in \mathcal{X}} p(x'; x, u) \beta(x') V(x') \right\}, \quad x \in \mathcal{X}.$$

Equivalently, with the operator

$$(\mathcal{T}v)(x) := \min_{u \in \mathcal{U}} \left\{ g(x, u) + \sum_{x'} p(x'; x, u) \beta(x') v(x') \right\},$$

the Bellman equation is  $V = \mathcal{T}V$ .

(ii) The operator  $\mathcal{T}$  is a contraction in the sup norm. Indeed, for bounded  $v, w$ ,

$$\begin{aligned} |(\mathcal{T}v)(x) - (\mathcal{T}w)(x)| &\leq \max_{u \in \mathcal{U}} \sum_{x'} p(x'; x, u) \beta(x') |v(x') - w(x')| \\ &\leq \beta_* \|v - w\|_\infty. \end{aligned}$$

Hence

$$\|\mathcal{T}v - \mathcal{T}w\|_\infty \leq \beta_* \|v - w\|_\infty.$$

By the Banach fixed-point theorem,  $\mathcal{T}$  has a unique bounded fixed point  $V$ , which is the value function.

To approximate  $V$  by finite-horizon problems, define

$$V_0(x) := 0, \quad V_{n+1} := \mathcal{T}V_n, \quad n \geq 0.$$

Then  $V_n$  is exactly the  $n$ -stage truncated value function with terminal cost 0, and by contraction,

$$\|V_n - V\|_\infty = \|\mathcal{T}^n V_0 - \mathcal{T}^n V\|_\infty \leq \beta_*^n \|V_0 - V\|_\infty.$$

Therefore

$$V_n \rightarrow V \quad \text{uniformly as } n \rightarrow \infty.$$

So the infinite-horizon discounted problem has a bounded solution, and this solution is obtained as the limit of the finite-horizon value functions.