13 Markov decision processes in continuous time: examples

13.1 Controlled diffusion

Since last week, we have been considering a controlled diffusion process with state $x \in \mathbb{R}^n$, action $u \in \mathbb{R}^m$, and initial conditions $X_0 = x_0, U_0 = u_0$. The diffusion process is described by the drift $b(x,u) \in \mathbb{R}^n$ and diffusion matrix $A(x,u) = \sigma(x,u)\sigma(x,u)^T$ ($\sigma(x,u) \in \mathbb{R}^{n\times n}$). The mean and covariance of the diffusion process at time h is expressed as

$$\mathbf{E}[X_h] = h \cdot b(x,u) + o(h) \tag{1}$$

$$\mathrm{Cov}(X_h) = h \cdot \sigma(x,u)\sigma(x,u)^T + o(h) \tag{2}$$

The state-feedback control is determined by the policy $g : \mathbb{R}^n \times [0,T] \to \mathcal{U}$, which gives the expression for the control

$$u_t = g(X_t, t) \tag{3}$$

Also, from the discussion in the previous lectures, we have the stochastic differential equation of the diffusion process controlled by the policy g

$$dX_t^g = b^g(X_t,t)dt + \sigma^g(X_t,t)dW_t \ (t \in [0,T],\ X_0^g = x) \tag{4}$$

, where

$$b^g(x,t) = b^g(x, g(x,t)) \tag{5}$$

$$\sigma^g(x,t) = \sigma^g(x, g(x,t)) \tag{6}$$

The stochastic differential equation can be rewritten using the Ito integral as follows

$$X_t^g = X_0^g + \int_0^t b^g(X_s,s)ds + \int_0^t \sigma^g(X_s,s)dW_s \tag{7}$$

Now, the optimal control problem can be formulated as the minimization of the total expected cost

$$J(g) := \mathbf{E}^g\left[\int_0^t c_t(X_t,U_t)dt + c_T(X_T)\right] \tag{8}$$

$$= \mathbf{E}^g\left[\int_0^t c_t(X_t,g(X_t,t))dt + c_T(X_T)\right] \tag{9}$$

The cost-to-go function for $t \in [0,T]$, $x \in \mathbb{R}^n$ can be expressed as

$$J_t(x;g) := \mathbf{E}^g\left[\int_t^T c_s(X_s,U_s)ds + c_T(X_T) \mid X_t = x\right] \tag{10}$$

The value function is given by

$$V_t(x) = \min_g J_t(x;g) \tag{11}$$

As discussed in the last lecture, the optimal control can be obtained by solving the **Hamilton-Jacobi-Bellman equation** stated as follows:
If $\exists V_t(x)$ for $x \in \mathbb{R}^n$, $t \in [0,T]$, which is C^2 in x and C^1 in t that solves

$$\frac{dV_t}{dt}(x) = -\min_{u \in \mathcal{U}} \{c_t(x,u) + \mathcal{A}^u V_t(x)\} \tag{12}$$

$$V_T(x) = c_T(x) \text{ for } \forall x \tag{13}$$

Then, $V_t(x)$ is a value function and

$$g^*(x,t) = \arg\min_{u \in \mathcal{U}} \{c_t(x,u) + \mathcal{A}^u V_t(x)\} \tag{14}$$

Here, recall that

$$\mathcal{A}^u V_t(x) = b(x,u)^T \nabla_x V_t(x) + \frac{1}{2}\text{tr}\left(\sigma(x,u)\sigma(x,u)^T \nabla_x^2 V_t(x)\right) \tag{15}$$

13.2 LQR in continuous time

Now, we consider an LQR problem in continuous time with state space $\mathcal{X} = \mathbb{R}^n$ and action space $\mathcal{U} = \mathbb{R}^m$. The drift of the corresponding diffusion process can be expressed as

$$b(x,u) = Ax + Bu \tag{16}$$

and

$$\sigma(x,u) = \Gamma \in \mathbb{R}^{n \times n} \tag{17}$$

The generator of the diffusion process is then expressed as

$$\mathcal{A}^u V_t(x) = b(x,u)^T \nabla V_t(x) + \frac{1}{2}\text{tr}\left\{\sigma(x,u)\sigma(x,u)^T \nabla^2 V_t(x)\right\} \tag{18}$$

$$= (Ax + Bu)^T \nabla V_t(x) + \frac{1}{2}\text{tr}\left\{\Gamma\Gamma^T \nabla^2 V_t(x)\right\} \tag{19}$$

with per-step and terminal costs

$$c_t(x,u) = x^T Q x + u^T R u, \qquad c_T(x) = x^T \tilde{Q} x \tag{20}$$

where Q and $\tilde{Q}$ are symmetric and positive semidefinite, while R is symmetric and positive definite. The HJB equation can be written down as

$$\frac{\partial V_t}{\partial t}(x) = -\min_{u \in \mathbb{R}^n} \left\{x^T Q x + u^T R u + (Ax + Bu)^T \nabla V_t(x) + \frac{1}{2}\text{tr}\left(\Gamma\Gamma^T \nabla^2 V_t(x)\right)\right\} \tag{21}$$

with

$$V_T(x) = x^T \tilde{Q} x \tag{22}$$

Now, we “guess" the solution by the quadratic form of x with an additional term as a function of t only:

$$V_t(x) = x^T K_t x + f(t) \tag{23}$$
$$K_t = K_t^T \geq 0, f(t) \geq 0 \tag{24}$$
$$K_T = \tilde{Q}, f(T) = 0 \tag{25}$$

This gives

$$\frac{\partial V_t}{\partial t} = x^T \frac{dK_t}{dt} x + \frac{df(t)}{dt} \tag{26}$$

The spatial derivatives of the value function can be expressed as

$$\nabla V_t(x) = 2K_t x \tag{27}$$
$$\nabla^2 V_t(x) = 2K_t \tag{28}$$

Using these expressions, the HJB equation for our case can be written as

$$\frac{\partial V_t}{\partial t} = x^T \frac{dK_t}{dt} x + \frac{df(t)}{dt} \tag{29}$$
$$= -x^T Q x - 2x^T A^T K_t x - \mathrm{tr}(\Gamma \Gamma^T K_t) - \min_{u \in \mathbb{R}^m} \left\{ u^T R u + 2u^T B^T K_t x \right\} \tag{30}$$

The completion-of-squares lemma ($R \succ 0$) gives the optimal control

$$u^* = g^*(x,t) = -R^{-1} B^T K_t x (:= G_t x) \tag{31}$$

Now, we can write the terms of the HJB equation as

$$\frac{df(t)}{dt} = -\mathrm{tr}(\Gamma \Gamma^T K_t) \tag{32}$$

with $f(T) = 0$, and

$$\min_{u \in \mathbb{R}^m} \left\{ u^T R u + 2u^T B^T K_t x \right\} = -x^T K_t B R^{-1} B^T K_t x \tag{33}$$

for the optimal control. Therefore,

$$x^T \frac{dK_t}{dt} x = -x^T (Q + A^T K_t + K_t A - K_t B R^{-1} B^T K_t) x \tag{34}$$

for $\forall x \in \mathbb{R}^n$, or equivalently,

$$-\frac{dK_t}{dt} = Q + A^T K_t + K_t A - K_t B R^{-1} B^T K_t \tag{35}$$

with $K_T = \tilde{Q}$.

Theorem 13.1. *The optimal cost of the continuous-time (stochastic) LQR problem can be expressed as*

$$\min_g \mathbf{E}^g \left\{ \int_0^T (X_t^T Q X_t + U_t^T R U_t) dt + X_T^T \tilde{Q} X_T \right\} = \mathbf{E}[X_0^T K_0 X_0 + f(0)] \tag{36}$$

where $\{K_t\}$, $\{f_t\}$ *solves*

$$-\frac{dK_t}{dt} = Q + A^T K_t + K_t A - K_t B R^{-1} B^T K_t \tag{37}$$

$$K_T = \tilde{Q} \tag{38}$$

and

$$\frac{df(t)}{dt} = -\mathrm{tr}(\Gamma \Gamma^T K_t) \tag{39}$$

$$f(T) = 0 \tag{40}$$

The optimal policy is then obtained by

$$g^*(x,t) = G_t x \tag{41}$$

$$G_t = -R^{-1} B^T K_t x \tag{42}$$

Using the optimal control derived in 13.1, the optimal controlled processed can be described as follows:

$$dX_t = (AX_t + BG_t X_t)dt + \Gamma dW_t \tag{43}$$

$$= (A + BG_t)X_t dt + \Gamma dW_t (:= L_t X_t dt + \Gamma dW_t) \tag{44}$$

for $t \in [0, T]$, or equivalently,

$$X_t = X_0 + \int_0^t L_s X_s ds + \int_0^t \Gamma dW_s \tag{45}$$

13.3 Logarithmic transformation

Here is an example of a stochastic control problem that can be solved explicitly. Consider the simple case where

$$b(x,u) = b(x) + u \tag{46}$$

and

$$\sigma(x,u) = I_n \tag{47}$$

The uncontrolled process can then be expressed as

$$dX_t = b(X_t)dt + dW_t \tag{48}$$

for $t \in [0, T]$. The per-step and terminal cost functions are defined as

$$c_t(x, u) = \frac{1}{2}\|u\|^2 \tag{49}$$

$$c_T(x) = f(x) \tag{50}$$

where $t \in [0, T]$ and $f : \mathbb{R}^n \to \mathbb{R}$. Then, the optimal control problem can be formulated by the cost-minimization as follows:

$$\min_g J(g) = \mathbf{E}^g \left[\int_0^T \frac{1}{2}\|u\|^2 dt + f(X_T) \right] \tag{51}$$

with the stochastic differential equation of the diffusion process given by

$$dX_t^g = (b(X_t^g) + g(X_t^g, t))dt + dW_t \tag{52}$$

The infinitesimal generator $\mathcal{A}^u$ for this process is then written as

$$\mathcal{A}^u V_t(x) = b(x, u)^T \nabla V_t(x) + \frac{1}{2}\mathrm{tr}\left(\sigma(x, u)\sigma(x, u)^T \nabla^2 V_t(x)\right) \tag{53}$$

$$= b(x)^T \nabla V_t(x) + u^T \nabla V_t(x) + \frac{1}{2}\Delta V_t(x) \tag{54}$$

The HJB equation is

$$\frac{\partial V_t}{\partial t}(x) = -\min_{u \in \mathbb{R}^n} \left\{ \frac{1}{2}\|u\|^2 + b(x)^T \nabla V_t(x) + u^T \nabla V_t(x) + \frac{1}{2}\Delta V_t(x) \right\} \tag{55}$$

$$= -b(x)^T \nabla V_t(x) - \frac{1}{2}\Delta V_t(x) - \min_{u \in \mathbb{R}^n} \left\{ \frac{1}{2}\|u\|^2 + u^T \nabla V_t(x) \right\} \tag{56}$$

The optimal control is

$$u^*(x) = -\nabla V_t(x) \tag{57}$$

and the HJB equation then becomes

$$\frac{\partial V_t}{\partial t}(x) = -(b(x)^T \nabla V_t(x) + \frac{1}{2}\Delta V_t(x)) + \frac{1}{2}\|\nabla V_t(x)\|^2 \tag{58}$$

$$= -\mathcal{A}^0 V_t(x) + \frac{1}{2}\|\nabla V_t(x)\|^2 \tag{59}$$

with

$$V_T(\cdot) = f_T(\cdot) \tag{60}$$

Now, we rewrite eq.(59) as

$$\frac{\partial V_t}{\partial t}(x) + \mathcal{A}^0 V_t(x) = \frac{1}{2}\|\nabla V_t(x)\|^2 \tag{61}$$

$$V_T(\cdot) = f(\cdot) \tag{62}$$

To solve this nonlinear partial differential equation, we apply the **Cole-Hopf transformation**

$$h_t(x) = e^{-V_t(x)} (> 0) \tag{63}$$

Then,

$$\frac{\partial h_t}{\partial t}(x) = -h_t(x)\frac{\partial V_t}{\partial t}(x) \tag{64}$$

$$\nabla h_t(x) = -h_t(x)\nabla V_t(x) \tag{65}$$

$$\nabla^2 h_t(x) = -\nabla h_t(x)\nabla V_t(x)^T - h_t(x)\nabla^2 V_t(x) \tag{66}$$

Using these elements, we can get

$$\frac{\partial h_t}{\partial t}(x) = -\mathcal{A}^0 h_t(x) \tag{67}$$

with

$$h_T(x) = e^{-f(x)} \tag{68}$$

The solution to the transformed equation for the uncontrolled diffusion is given by **Feynman-Kac formula**:

$$h_t(x) = \mathbf{E}\left[e^{-f(X_T)} \mid X_t = x\right] \tag{69}$$

With this solution, we can get the value function and the optimal policy as follows:

$$V_t(x) = -\log h_t(x) \tag{70}$$

$$= -\log \mathbf{E}\left[e^{-f(X_T)} \mid X_t = x\right] \tag{71}$$

$$g_t^*(x) = -\nabla V_t(x) = \nabla \log \mathbf{E}\left[e^{-f(X_T)} \mid X_t = x\right] \tag{72}$$

$$V_0(x) = -\log \mathbf{E}\left[e^{-f(X_T)} \mid X_0 = x\right] \tag{73}$$

Proof of the Feynman-Kac formula starts with the stochastic differential equation of the uncontrolled diffusion process

$$dX_t = b(X_t)dt + dW_t \tag{74}$$

with $t \in [0, T]$. By defining

$$f(x,t) := V_t(x) (\in C^2) \tag{75}$$

and

$$Y_t := h_t(X_t) \tag{76}$$

, we can apply the Ito's lemma to get

$$dY_t = \underbrace{\left(\frac{\partial h_t}{\partial t}(X_t) + \mathcal{A}^0 h_t(X_t)\right)}_{=0} dt + \nabla h_t(X_t)^T dW_t \tag{77}$$

$$= \nabla h_t(X_t)^T dW_t \tag{78}$$

or equivalently

$$Y_T = Y_t + \int_t^T \nabla h_s(X_s)^T dW_s \tag{79}$$

Therefore, using $Y_t := h_t(X_t)$,

$$h_T(X_T) = e^{-f(X_T)} \tag{80}$$

$$= h_t(X_t) + M_t \tag{81}$$

where M_t vanishes when taking the expectation (Martingale process). Therefore, $h_t(x)$ expressed in eq.(69) is the solution to the diffusion process.

13.4 Schrödinger bridge

Now, we consider a standard Wiener process $\{W_t\}_{t\geq 0}$, where $W_t \sim \gamma_n$. Our problem is that, starting from the initial state 0 at time 0, we want to reach a state with the density

$$p(x) = e^{-f(x)}\gamma_n(x) \tag{82}$$

at time 1. Conditional probabilities can be written as

$$\mathbf{E}\left[e^{-f(W_1)} \mid W_0 = 0\right] = \int e^{-f(x)}\gamma_n(x)dx \tag{83}$$

$$\mathbf{E}\left[e^{-f(X_1)} \mid X_t = x\right] = \mathbf{E}\left[e^{-f(x+\sqrt{1-t}Z)}\right] \tag{84}$$

where $Z \sim \gamma_n$ and $0 \leq t \leq 1$. The optimal control is

$$u_t^* = \nabla \log \mathbf{E}_Z\left[e^{-f(x+\sqrt{1-t}Z)}\right] (= g^*(X_t, t)) \tag{85}$$

Therefore, the evolution of the state can be expressed by

$$dX_t = g^*(X_t, t)dt + \sigma W_t \tag{86}$$

with $X_0 \sim 0$ and $X_1 \sim p$. We then get

$$e^{-f(x)} = \frac{p(x)}{\gamma_n(x)} \tag{87}$$

$$f(x) = -\log \frac{p(x)}{\gamma_n(x)} \tag{88}$$

and it is possible to show that, indeed, the optimal control g^* results in $X_1^{g^*}$ having density p.

13.5 Describing a continuous-time Markov process by its generator

We have previously seen diffusion processes described by an associated time-invariant generator $\mathcal{A}$. We introduced control to diffusion processes by defining a family of generators $(\mathcal{A}^u)_{u\in\mathcal{U}}$ parameterized by the control u. Solving the optimal control problem for controlled diffusions gave rise to the Hamilton-Jacobi-Bellman equation, which incorporates this generator $\mathcal{A}^u$. This same idea of describing control processes by their generators can be applied to continuous-time Markov processes as well, and we will see that the HJB equation still applies for finding the optimal control policy for finite-state MDPs.

Suppose $\{X_t\}_{t\geq 0}$ is a continuous-time Markov process with state space $\mathcal{X}$ and transition kernel $P_t(x, B) = \mathbf{P}(X_t \in B \mid X_0 = t)$. For any function f, we can treat P_t as an operator defined as follows.

$$P_t f(x) = \int_{\mathcal{X}} f(x') P_t(x, dx') = \mathbf{E}[f(X_t) \mid X_0 = x] \tag{89}$$

Whether P_t represents the transition kernel or the corresponding operator should be clear from context. If this operator is differentiable at $t = 0$, we can define the generator $\mathcal{A}$ for this process via

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t f(x)\Big|_{t=0} = \lim_{h\downarrow 0} \frac{1}{h}(P_h f(x) - f(x)) = \mathcal{A}f(x) \tag{90}$$

This gives the forward Kolmogorov equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t f = \mathcal{A} P_t f = P_t \mathcal{A} f \tag{91}$$

If $|\mathcal{X}| = n < \infty$, then the initial distribution μ_0 is an n-dimensional vector, and the transition kernels $(P_t)_{t\geq 0}$ are $n \times n$ matrices. Let $\mu_t(\cdot) = \mathbf{P}(X_t = \cdot)$ be the distribution of the state at time t. Then the forward Kolmogorov equation is given by the matrix differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \mu_t = \mu_t \Lambda, \mu_0 = \mu_0 \tag{92}$$

where $\Lambda = [\lambda_{x,x'}]_{x,x'\in\mathcal{X}}$ is the transition intensity matrix satisfying $\lambda_{x,x'} \geq 0$ for all $x \neq x'$ and $\sum_{x'\in\mathcal{X}} \lambda_{x,x'} = 0$. This matrix is defined via

$$\Lambda = \lim_{h\downarrow 0} \frac{1}{h}(P_h - I_n) \tag{93}$$

Notice that by definition, the transition intensity matrix Λ is equal to the generator $\mathcal{A}$. We can see that the null space of $\mathcal{A}$ includes constants. Now consider a function $f : \mathcal{X} \times [0, \infty) \to \mathbb{R}$. Let $Y_t = f(X_t, t)$, and suppose we want to say something about Y_t. From the definitions, we have that the expectation of Y_t given the initial state is

$$\mathbf{E}[Y_t \mid X_0 = x] = f(X_0, 0) + \mathbf{E}[\int_0^t \frac{\partial}{\partial s} f(X_s, s) + \mathcal{A}f(X_s, s)\mathrm{d}s \mid X_0 = x] \tag{94}$$

This serves as a starting point for the optimality equation for several MDPs.

13.6 Controlled continuous-time finite-state Markov processes

We will consider a continuous-time Markov process $\{X_t\}_{t\geq 0}$ with finite state space $\mathcal{X}$ and a family of generators $(\mathcal{A}_u)_{u\in\mathcal{U}}$ which are parameterized by the control.

Example 13.1 (Two-state controlled Markov process)**.** Suppose $\mathcal{X} = \{0, 1\}$ and the transition intensity matrix is given by

$$\Lambda = \begin{bmatrix} -\lambda_{01} & \lambda_{01} \\ \lambda_{10} & -\lambda_{10} \end{bmatrix}, \quad \lambda_{01}, \lambda_{10} \geq 0 \tag{95}$$

For $u \in [0, \infty)$, define the controlled transition intensity matrix

$$\Lambda^u = \begin{bmatrix} -\lambda_{01} - b_{01}u & \lambda_{01} + b_{01}u \\ \lambda_{10} + b_{10}u & -\lambda_{10} - b_{10}u \end{bmatrix}, \quad \lambda_{01}, \lambda_{10} \geq 0 \tag{96}$$

Let $\mu_t^u := (\mu_t^u(0), \mu_t^u(1))$ be the distribution of X_t under control u. Then the forward Kolmogorov equations are given by

$$\frac{d}{dt}\mu_t^u = (\frac{d}{dt}\mu_t^u(0), \frac{d}{dt}\mu_t^u(1)) \tag{97}$$

$$= (\mu_t^u(0), \mu_t^u(1)) \begin{bmatrix} -\lambda_{01} - b_{01}u & \lambda_{01} + b_{01}u \\ \lambda_{10} + b_{10}u & -\lambda_{10} - b_{10}u \end{bmatrix} \tag{98}$$

$$\frac{d}{dt}\mu_t^u(0) = -(\lambda_{01} - b_{01}u)\mu_t^u(0) + (\lambda_{10} + b_{10}u)\mu_t^u(1) \tag{99}$$

$$\frac{d}{dt}\mu_t^u(1) = (\lambda_{01} + b_{01}u)\mu_t^u(0) - (\lambda_{10} - b_{10}u)\mu_t^u(1) \tag{100}$$

$$\tag{101}$$

The solution to these equations specifies a controlled process $\{X_t^u\}_{t\geq 0}$.

We want to introduce state feedback in the form of a policy $g : \mathcal{X} \times [0, T] \to \mathcal{U}$. The generator corresponding to policy g is defined by $\mathcal{A}_t^g f(x) = \mathcal{A}^u f(x)$ if $g(x, t = u$, so we have

$$\frac{1}{h}[\mathbf{E}[f(X_{t+h}) \mid X_t = x, U_t = g(x, t)] - f(x, t)] \to \mathcal{A}_t^g f(x) \text{ as } \downarrow 0 \tag{102}$$

The generator $\mathcal{A}_t^g$ specifies a time-inhomogeneous Markov process $\{X_t^g\}_{t\geq 0}$. To control this process over a finite horizon, we introduce the cost

$$J(g) = \mathbf{E}^g[\int_0^T c_t(X_t, U_t)dt + \tilde{(c)}_T(X_T)] \tag{103}$$

which must be minimized over policies g. As usual, define the cost-to-go

$$J_t(x; g) = \mathbf{E}^g[\int_t^T c_s(X_s, U_s)ds + \tilde{(c)}_T(X_T) \mid X_t = x] \tag{104}$$

and the value function

$$V_t(x) = \min_g J_t(x; g) \tag{105}$$

Proposition 13.1. *(Hamilton-Jacobi-Bellman equation for controlled Markov processes) If a solution to the HJB equation*

$$\frac{\partial V_t}{\partial t}(x) + \min_{u \in \mathcal{U}}(c_t(x,u) + \mathcal{A}^u V_t(x)) = 0, \quad V_T(\cdot) = \tilde{(c)}_T(\cdot) \tag{106}$$

exists, then the optimal control is given by

$$g^*(x,t) = \arg\min_{u \in \mathcal{U}}(c_t(x,u) + \mathcal{A}^u V_t(x)) \tag{107}$$

That is, the HJB equation holds in larger generality than just diffusions, including finite-state Markov processes.

Proof: By the Markov property, we have that for all policies g and all times $0 \le s < t \le T$,

$$J_s(x;g) = \mathbf{E}^g[\int_s^t c_r(X_r, U_r)\mathrm{d}r \mid X_s = x] + \mathbf{E}^g[J_t(X_t;g) \mid X_s = x] \tag{108}$$

Assume that $V_t(x)$ exists, is differentiable in t, and that $\mathcal{A}^u V_t$ exists. Then we have

$$\mathbf{E}^g[V_t(X_t) \mid X_s = x] = V_s(x) + \mathbf{E}[\int_s^t \frac{\partial}{\partial r} V_r(X_r) + \mathcal{A}_r^g V_r(X_r)\mathrm{d}r \mid X_s = x] \tag{109}$$

$$V_s(x) = \mathbf{E}^g[V_t(X_t) \mid X_s = x] - \mathbf{E}[\int_s^t \frac{\partial}{\partial r} V_r(X_r) + \mathcal{A}_r^g V_r(X_r)\mathrm{d}r \mid X_s = x] \tag{110}$$

Assume that g^* is an optimal policy, and construct a policy $\bar{g}$ to be equal to some arbitrary (non-optimal) policy g for times in $[0,t)$ and equal to the optimal policy g^* for times in $[t,T]$. Then

$$V_s(x) \le J_s(x;\bar{g}) = \mathbf{E}[\int_s^t c_r(X_r, U_r)\mathrm{d}r \mid X_s = x] + \mathbf{E}^g[J_t(X_t;\bar{g}) \mid X_s = x] \tag{111}$$

$$= \mathbf{E}[\int_s^t c_r(X_r, U_r)\mathrm{d}r \mid X_s = x] + V_t(X_t) \tag{112}$$

where we have equality if $\bar{g} = g^*$. This equation implies

$$\frac{\partial}{\partial t} V_t(x) + c_t(x,u) + \mathcal{A}^u V_t(x) = 0, \quad V_T(\cdot) = \tilde{c}_T(\cdot) \tag{113}$$

when $u = g^*(x,t)$, which gives the HJB equation, as desired. ■

Example 13.2 (Two-state controlled Markov process with quadratic costs)**.** Consider the setting of the previous example about the two-state controlled Markov process. We introduce the following quadratic cost function

$$c_t(x,u) = ku^2, \ k > 0, \quad \tilde{c}_t(\cdot) = f(\cdot) \tag{114}$$

The HJB equation for this example is

$$\frac{\partial}{\partial t} V_t(x) = -\min_{u \ge 0}(ku^2 + \Lambda^u V_t(x)), \quad V_T(\cdot) = f(\cdot) \tag{115}$$

where the controlled transition intensity matrix operating on the value function gives

$$\Lambda^u V_t(x) = (\Lambda^u V_t(0), \Lambda^u V_t(1))^\mathsf{T} \tag{116}$$

$$= \begin{bmatrix} -\lambda_{01} - b_{01}u & \lambda_{01} + b_{01}u \\ \lambda_{10} + b_{10}u & -\lambda_{10} - b_{10}u \end{bmatrix} \begin{bmatrix} V_t(0) \\ V_t(1) \end{bmatrix} \tag{117}$$

$$= ((\lambda_{01} + b_{01}u)(V_t(1) - V_t(0)), (\lambda_{10} - b_{10}u)(V_t(0) - V_t(1)))^\mathsf{T} \tag{118}$$

Substituting this into the HJB equation gives

$$\frac{\partial}{\partial t} V_t(0) = -\lambda_{01}(V_t(1) - V_t(0)) - \min_{u \geq 0}(ku^2 + b_{01}u(V_t(0) - V_t(1))), \quad V_T(0) = f(0) \tag{119}$$

$$\frac{\partial}{\partial t} V_t(1) = -\lambda_{10}(V_t(0) - V_t(1)) - \min_{u \geq 0}(ku^2 + b_{10}u(V_t(1) - V_t(0))), \quad V_T(1) = f(1) \tag{120}$$

Recall that for any arbitrary positive constant c,

$$\min_{u \geq 0} ku^2 + cu = k(u + \frac{c}{2k})^2 - \frac{c^2}{4k} \tag{121}$$

Then we have that for the first equation, the minimum-cost attaining control is

$$u^* = \frac{b_{01}(V_t(0) - V_t(1))}{2k} \mathbf{1}_{\{V_t(1) \leq V_t(0)\}} \tag{122}$$

and for the second equation,

$$u^* = \frac{b_{10}(V_t(1) - V_t(0))}{2k} \mathbf{1}_{\{V_t(0) \leq V_t(1)\}} \tag{123}$$

This simplifies the HJB equation to

$$\frac{\partial}{\partial t} V_t(0) = -\lambda_{01}(V_t(1) - V_t(0)) + \frac{b_{01}^2(V_t(0) - V_t(1))^2}{4k} \mathbf{1}_{\{V_t(1) \leq V_t(0)\}}, \quad V_T(0) = f(0) \tag{124}$$

$$\frac{\partial}{\partial t} V_t(1) = -\lambda_{10}(V_t(0) - V_t(1)) + \frac{b_{10}^2(V_t(1) - V_t(0))^2}{4k} \mathbf{1}_{\{V_t(0) \leq V_t(1)\}}, \quad V_T(1) = f(1) \tag{125}$$

$$\tag{126}$$

These equations are difficult to solve analytically becuase of the constraints on the value functions (however, they can be solved numerically). To attempt to remove the constraints, we can introduce the transformation $u \mapsto e^u$, which takes the HJB equation into the following form

$$\frac{\partial}{\partial t} V_t(0) = -\min_{u \in \mathbb{R}}(ke^{2u} + (\lambda_{01} + b_{01}e^u)(V_t(1) - V_t(0))), \quad V_T(0) = f(0) \tag{127}$$

$$\frac{\partial}{\partial t} V_t(1) = -\min_{u \in \mathbb{R}}(ke^{2u} + (\lambda_{10} - b_{10}e^u)(V_t(0) - V_t(1))), \quad V_T(1) = f(1) \tag{128}$$

but unfortunately this still creates a discontinuity in the control law (specifically, we still need an ordering on $V_t(0)$ and $V_t(1)$ for a minimum u^* to exist, which creates the same discontinuity as before).