Chapter 3 Notes

3.2 Controlled diffusion processes

We consider a control model where the state of the system is governed by a stochastic differential equation (SDE) valued in $\mathbb{R}^n$:

$$dX_s = b(X_s, a)ds + \sigma(X_s, \alpha_s)dW_s, \tag{3.1}$$

where W is a d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. We can consider coefficients $b(t, x, a)$ and $\sigma(t, x, a)$ depending on time t, but in the case of infinite horizon problems described below, it is important that coefficients do not depend on time in order to get the stationarity of the problem, and so a value function independent of time.

The control $\alpha = (\alpha_s)$ is a progressively measurable (w.r.t. $\mathbb{F}$) process, valued in A, subset of $\mathbb{R}^m$.

The measurable functions $b : \mathbb{R}^n \times A \to \mathbb{R}^n$ and $\mathbb{R}^n \times A \to \mathbb{R}^{n\times d}$ satisfy a uniform Lipschitz condition in $A : \exists K \geq 0, \forall x, y \in \mathbb{R}^n, \forall a \in A$,

$$|b(x, a) - b(y, a)| + |\sigma(x, a) - \sigma(y, a)| \leq K|x - y|. \tag{3.2}$$

In the sequel, for $0 \leq t \leq T \leq \infty$, we denote by $\mathcal{T}_{t,T}$ the set of stopping times valued in $[t, T]$. When $t = 0$ and $T = \infty$, we simply write $\mathcal{T} = \mathcal{T}_{t,T}$.

Finite horizon problem.

We fix a finite horizon $0 < T < \infty$. We denote by $\mathcal{A}$ the set of control processes α s.t.

$$E\left[\int_0^T |b(0, \alpha_t)|^2 + |\sigma(0, \alpha_t)|^2 dt\right] < \infty. \tag{3.3}$$

From Section 1.3, the condition (3.2) and (3.3) ensure for all $\alpha \in \mathcal{A}$ and for any initial condition $(t, x) \in [0, T] \times \mathbb{R}^n$, the existence and uniqueness of a strong solution to the SDE (with random coefficients) (3.1) starting from x at $s = t$. We then denote by $\{X_s^{t,x}, t \leq s \leq T\}$ this solution with a.s. continuous paths. Under these conditions on b, σ and α, we have

$$E\left[\sup_{t\leq s\leq T} |X_s^{t,x}|^2\right] < \infty. \tag{3.4}$$

$$\lim_{h \downarrow 0^+} E\left[\sup_{s\in[t,t+h]} |X_s^{t,x} - x|^2\right] < \infty. \tag{3.5}$$

Functional objective.

Let $f : [0,T] \times \mathbb{R}^n \times A \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ two measurable functions. We suppose that:

(**Hg**) (i) g is lower-bounded

or (ii) g satisfies a quadratic growth condition: $|g(x)| \leq C(1 + |x|^2)$, $\forall x \in \mathbb{R}^n$, for some C independent of x.

For $(t,x) \in [0,T] \times \mathbb{R}^n$, we denote by $\mathcal{A}(t,x)$ the subset of controls α in $\mathcal{A}$ s.t.

$$E\left[\int_t^T |f(s, X_s^{t,x}, \alpha_s)| ds\right] < \infty \tag{3.6}$$

and we assume that $\mathcal{A}(t,x)$ is not empty for all $(t,x) \in [0,T] \times \mathbb{R}^n$. We can then define under (**Hg**) the gain function:

$$J(t,x,a) = E\left[\int_t^T f(s, X_s^{t,x}, \alpha_s) ds + g(X_T^{t,x})\right],$$

for all $(t,x) \in [0,T] \times \mathbb{R}^n$ and $\alpha \in \mathcal{A}(t,x)$, and we introduce the associated value function:

$$v(t,x) = \sup_{\alpha\in\mathcal{A}(t,x)} J(t,x,a). \tag{3.7}$$

Given an initial condition $(t,x) \in [0,T) \times \mathbb{R}^n$, we say that $\hat{\alpha} \in \mathcal{A}(t,x)$ is an optimal control if $v(t,x) = J(t,x,\hat{\alpha})$.

A control process α in the form $\alpha_s = a(s, X_s^{t,x})$ for some measurable function a from $[0,T] \times \mathbb{R}^n$ into A, is called Markovian control.

3.3 Dynamic Programming Principle

Theorem 3.3.1 (Dynamic programming principle)
(1) Finite horizon: let $(t,x) \in [0,T] \times \mathbb{R}^n$. Then we have

$$v(t,x) = \sup_{\alpha\in\mathcal{A}(t,x)} \sup_{\theta\in\mathcal{T}_{t,T}} E\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s) ds + v(\theta, X_\theta^{t,x})\right] = \tag{3.14}$$

$$= \sup_{\alpha\in\mathcal{A}(t,x)} \inf_{\theta\in\mathcal{T}_{t,T}} E\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s) ds + v(\theta, X_\theta^{t,x})\right]. \tag{3.15}$$

(1) Infinite horizon: let $x \in \mathbb{R}^n$. Then we have

$$v(x) = \sup_{\alpha \in \mathcal{A}(x)} \sup_{\theta \in \mathcal{T}} E\left[\int_0^\theta e^{-\beta s} f(X_s^x, \alpha_s) ds + e^{-\beta\theta} v(X_\theta^x)\right] = \tag{3.16}$$

$$= \sup_{\alpha \in \mathcal{A}(x)} \inf_{\theta \in \mathcal{T}} E\left[\int_0^\theta e^{-\beta s} f(X_s^x, \alpha_s) ds + e^{-\beta\theta} v(X_\theta^x)\right]. \tag{3.17}$$

The interpretation of the DPP is that the optimization problem can be split in two parts: optimal control on the whole time interval $[t, T]$ may be obtained by first searching for an optimal control from time θ given the state value $X_s^{t,x}$, i.e., compute $v(\theta, x_\theta^{t,x})$, and then maximizing over controls on $[t, \theta]$ the quantity

$$E\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s) ds + v(\theta, X_\theta^{t,x})\right].$$

Remark 3.3.3 We shall often use the following equivalent formulation (in the finite horizon case) of the dynamic programming principle:
(i) For all $\alpha \in \mathcal{A}(t, x)$ and $\theta \in \mathcal{T}_{t,T}$:

$$v(t, x) \geq E\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s) ds + v(\theta, X_\theta^{t,x})\right]. \quad (v \text{ is an upper bound}) \tag{3.18}$$

(ii) For all $\varepsilon > 0$, there exists $\alpha \in \mathcal{A}(t, x)$ such that for all $\theta \in \mathcal{T}_{t,T}$:

$$v(t, x) - \varepsilon \leq E\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s) ds + v(\theta, X_\theta^{t,x})\right]. \quad (v \text{ is the least upper bound}) \tag{3.19}$$

This is a stronger version than the usual version of the DPP, which is written in the finite horizon case as

$$v(t, x) = \sup_{\alpha \in \mathcal{A}(t,x)} E\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s) ds + v(\theta, X_\theta^{t,x})\right] \tag{3.20}$$

for any stopping time $\theta \in \mathcal{T}_{t,T}$.

Proof of the DPP.
We consider the finite horizon case.
1. Given an admissible control $\alpha \in \mathcal{A}(t, x)$, we have by pathwise uniqueness of the flow of the SDE for X, the Markovian structure

$$X_s^{t,x} = X_s^{\theta, X_\theta^{t,x}}, \quad s \geq \theta,$$

3.4 Hamilton-Jacobi-Bellman equation

The Hamilton-Jacobi-Bellman equation (HJB) is the infinitesimal version of the dynamic programming principle: it describes the local behavior of the value function when we send the stopping time θ to t. The HJB equation is also called the dynamic programming equation.

3.4.1 Formal derivation of HJB

Finite horizon problem.
Let us consider the time $\theta = t + h$ and a constant control $\alpha_s = a$, for some arbitrary $a \in A$ in the equivalent formulation of the dynamic programming principle:

$$v(t,x) \geq E\left[\int_t^{t+h} f(s, X_s^{t,x}, a)ds + v(t+h, X_{t+h}^{t,x})\right]. \tag{3.24}$$

By assuming that v is smooth enough, we may apply Ito's formula between t and $t+h$:

$$v(t+h, X_{t+h}^{t,x}) = v(t,x) + \int_t^{t+h} \left(\frac{\partial v}{\partial t} + \mathcal{L}^a v\right)(s, X_s^{t,x})ds + \text{(local) martingale}$$

where $\mathcal{L}^a$ is the operator associated to the diffusion for the constant control a, and defined by

$$\mathcal{L}^a = b(x,a)D_x v + \frac{1}{2}tr(\sigma(x,a)\sigma^{'}(x,a)D_x^2 v)$$

By substituting into (3.24), we then get

$$0 \geq E\left[\int_t^{t+h} \left(\frac{\partial v}{\partial t} + \mathcal{L}^a v\right)(s, X_s^{t,x}) + f(s, X_s^{t,x}, a)ds\right].$$

Therefore, by the mean-value-theorem, we have

$$0 \geq \lim_{h\to 0} \frac{1}{h} E\left[\int_t^{t+h} \left(\frac{\partial v}{\partial t} + \mathcal{L}^a v\right)(s, X_s^{t,x}) + f(s, X_s^{t,x}, a)ds\right] =$$

$$= \lim_{h\to 0} \left(\frac{\partial v}{\partial t} + \mathcal{L}^a v\right)(s^*, X_{s^*}^{t,x}) + f(s^*, X_{s^*}^{t,x}, a) = \frac{\partial v}{\partial t}(t,x) + \mathcal{L}^a v(t,x) + f(t,x,a).$$

where $s^* \in (t, t+h)$.
Since this holds true for any $a \in A$, we obtain the inequality

$$-\frac{\partial v}{\partial t}(t,x) - \sup_{a\in A}[\mathcal{L}^a v(t,x) + f(t,x,a)] \geq 0. \tag{3.25}$$

On the other hand, suppose that α^* is an optimal control. Then, in (3.20), we have

$$v(t,x) = E\left[\int_t^{t+h} f(s, X_s^*, \alpha^*)ds + v(t+h, X_{t+h}^*)\right],$$

where X^* is the state system solution to (3.1) starting from x at t, with the control α^*.
By similar arguments as above starting from (3.24), but with equality instead of inequality, we have

$$-\frac{\partial v}{\partial t}(t,x) - \mathcal{L}^{\alpha^*} v(t,x) - f(t,x,\alpha^*) = 0, \tag{3.26}$$

which combined with (3.25) suggests that v should satisfy

$$-\frac{\partial v}{\partial t}(t,x) - \sup_{a \in A}[\mathcal{L}^a v(t,x) + f(t,x,a)] = 0. \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n, \tag{3.27}$$

if the above supremum in a is finite.
We often rewrite this PDE in the form

$$-\frac{\partial v}{\partial t}(t,x) - H(t,x,D_x v(t,x), D_x^2 v(t,x)) = 0. \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n, \tag{3.28}$$

where for $(t,x,p,M) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{S}_n$

$$H(t,x,p,M) = \sup_{a \in A}[b(x,a)p + \frac{1}{2}\mathrm{tr}(\sigma\sigma^{'}(x,a)M) + f(t,x,a)].$$

This function H is called the Hamiltonian of the associated control problem. The equation (3.28) is called the dynamic programming equation or Hamilton-Jacobi-Bellman (HJB) equation. The regular terminal condition associated to this PDE is

$$v(T,x) = g(x), \quad \forall x \in \mathbb{R}^n, \tag{3.29}$$

3.5 Verification Theorem

(Finite horizon) Let w be a function in $C^{1,2}([0,T) \times \mathbb{R}^n) \cap C^0([0,T] \times \mathbb{R}^n)$, and satisfying a quadratic growth condition, i.e., there exists a constant C s.t.

$$|w(t,x)| \leq C(1 + |x|^2), \quad \forall (t,x) \in [0,T] \times \mathbb{R}^n.$$

(i) Suppose that

$$-\frac{\partial w}{\partial t}(t,x) - \sup_{a \in A}[\mathcal{L}^a w(t,x) + f(t,x,a)] \geq 0. \quad \forall (t,x) \in [0,T) \times \mathbb{R}^n, \tag{3.35}$$

$$w(T,x) \geq g(x), \quad x \in \mathbb{R}^n. \tag{3.36}$$

Then $w \geq v$ on $[0, T] \times \mathbb{R}^n$.
(ii) Suppose further that $w(T, \cdot) = g$, and there exists a measurable function $\hat{\alpha}(t, x)$, $(t, x) \in [0, T) \times \mathbb{R}^n$, valued in A s.t.

$$-\frac{\partial w}{\partial t}(t,x) - \sup_{a \in A}[\mathcal{L}^a w(t,x) + f(t,x,a)] = -\frac{\partial w}{\partial t}(t,x) - \mathcal{L}^{\hat{\alpha}(t,x)} w(t,x) - f(t,x,\hat{\alpha}(t,x)) = 0,$$

the SDE

$$dX_s = b(X_s, \hat{\alpha}(s, X_s))ds + \sigma(X_s, \hat{\alpha}(s, X_s))dW_s$$

admits a unique solution, denoted by $\hat{X}_s^{t,x}$, given an initial condition $X_t = x$, and the process $\{\hat{\alpha}(s, \hat{X}_s), t \leq s \leq T\}$ lies in $\mathcal{A}(t, x)$. Then

$$w = v \quad \text{on } [0, T] \times \mathbb{R}^n,$$

and $\hat{\alpha}$ is an optimal Markovian control.