LECTURE NOTES 3

Stochastic Optimal Control[1]

Petter Kolm
NYU Courant

In addition to these notes, the required readings are (you can pick either one):

- Chapter 19.1-19.4 in the book "Arbitrage Theory in Continuous Time" by Björk (2009), or
- Chapter 5.1-5.3 in the book "Algorithmic and High-Frequency Trading" by Cartea, Jaimungal, and Penalva (2015)

1. Introduction

In these notes we provide an introduction to the dynamic programming method for solving stochastic control problems. While the technical details of stochastic control can be mathematically involved, this treatment will be kept on a more informal level. Our primary focus is on developing intuition and providing some of the main results from stochastic control that will allow us in future lectures to solve problems in algorithmic trading. If you are interested in reading a more detailed mathematical treatment of this material see, for example, Øksendal (2013), Pham (2009), and Touzi (2012).

2. Review: Results from Stochastic Calculus

2.1. Notation. We use the following notation:

[1]Version: January 31, 2023. These notes are copyrighted and intended only for students registered for the course *"Algorithmic Trading & Quantitative Strategies"* at NYU Courant. Please do not copy or post the notes on the web or pass them on to others by any means, whether digitally or otherwise. If you find any typos, I would much appreciate you email them to me.

- For a matrix A, we denote its transpose and trace by by A' and $\mathrm{Tr}(A)$, respectively.
- $Df\,(t, X_t)$ denotes the m-dimensional vector of first derivatives, that is $\left(Df\,(t, X_t)\right)_j := \partial_{x^j} f\,(t, X_t)$.
- $D^2 f\,(t, X_t)$ denotes the $(m \times m)$-dimensional matrix of mixed second derivatives, that is $\left(D^2 f\,(t, X_t)\right)_{jk} := \partial_{x^j x^k} f\,(t, X_t)$.
- We abbreviate *almost surely* by a.s.
- We define $C^{1,2}$ to be the space of real-valued functions $f : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$ such that the first-order partial derivative with respect to the first argument and all second-order partial derivatives with respect to the second argument exist and are continous.

2.2. Itô's Formula. We start with the one-dimensional diffusion process

$$dX_t = \mu\,(t, X_t)\,dt + \sigma\,(t, X_t)\,dW_t\,, \tag{1}$$

with $W_t \in \mathbb{R}$ a Brownian motion, and the drift and volatility are known functions of the form

$$\mu : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \tag{2}$$

$$\sigma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}\,. \tag{3}$$

Now consider the $C^{1,2}$-function $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and define the stochastic process

$$Y_t := f(t, X_t)\,. \tag{4}$$

Then we recall that Itô's formula say that Y has the stochastic differential

$$dY_t = \left\{\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right\}dt + \sigma\frac{\partial f}{\partial x}dW_t\,. \tag{5}$$

Frequently, I write this compactly as

$$dY_t = \left(\partial_t f + \mu\partial_x f + \frac{1}{2}\sigma^2\partial_{xx} f\right)dt + \sigma\partial_x f \cdot dW_t\,. \tag{6}$$

or

$$dY_t = \left(\partial_t f + \mu D f + \frac{1}{2}\sigma^2 D^2 f\right) dt + Df \cdot \sigma \cdot dW_t \,. \tag{7}$$

There is of course a multi-dimensional version Itô's formula. As we will need that for our purposes, let us summarize the result here. Let $W_t \in \mathbb{R}^n$ be a vector of independent Brownian motions and suppose the vector-valued processes $X_t \in \mathbb{R}^m$ satisfy the SDE

$$dX_t = \mu\left(t, X_t\right) dt + \sigma\left(t, X_t\right) dW_t \,, \tag{8}$$

where the drifts $\mu\left(t, X_t\right) \in \mathbb{R}^m$ and volatilities $\sigma\left(t, X_t\right) \in \mathbb{R}^{m \times n}$ are known functions of the form

$$\mu : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^n \tag{9}$$

$$\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^{n \times d} \,. \tag{10}$$

We state the multi-dimensional version of Itô's formula as the following theorem.

Theorem 1 (Itô's Formula)**.** *For $f \in C^{1,2}$ we define the stochastic process $Y_t := f\left(t, X_t\right)$. Then Y_t is an Itô process satisfying the SDE*

$$dY_t = \Big(\partial_t f\left(t, X_t\right) + \mu\left(t, X_t\right)' Df\left(t, X_t\right) \tag{11}$$

$$+ \frac{1}{2} \operatorname{Tr}\left\{\sigma\left(t, X_t\right) \sigma\left(t, X_t\right)' D^2 f\left(t, X_t\right)\right\}\Big) dt \tag{12}$$

$$+ Df\left(t, X_t\right)' \sigma\left(t, X_t\right) dW_t \,. \tag{13}$$

Suppose $f \in C^2$. Then we define the *infinitesimal generator* of a process X_t acting on f by

$$\mathcal{L}_t f(x) := \lim_{h \to 0} \frac{\mathbb{E}\left[f\left(X_{t+h}\right) | X_t = x\right] - f(x)}{h} \,. \tag{14}$$

Recall that the infinitesimal generator is the generalization of a derivative of a function to the derivative of function acting on a stochastic process. From Itô's formula above, the generator of an Itô process is given by

$$\mathcal{L}_t f(x) = \mu(t,x)' Df(x) + \frac{1}{2} \operatorname{Tr}\left\{\sigma(t,x)\sigma(t,x)' D^2 f(x)\right\}. \tag{15}$$

3. Controlled Diffusion Processes

We will consider an n-dimensional *stochastic state process* $X^u := (X_t^u)_{t=0,\dots,T}$ which we are trying to "steer" by choosing a k-dimensional *control process* u in some suitable way. Here we restrict ourselves to a state process described by the *controlled stochastic differential equation*

$$dX_t^u = \mu\left(t, X_t^u, u_t\right) dt + \sigma\left(t, X_t^u, u_t\right) dW_t \tag{16}$$

$$X_0^u = x_0 \,, \tag{17}$$

where $X_t^u \in \mathbb{R}^n$, $x_0 \in \mathbb{R}^n$ is the initial state, $u \in \mathbb{R}^k$ is the control, W_t is a d-dimensional vector of independent Brownian motions, and $\mu(t, x, u)$ and $\sigma(t, x, u)$ are known functions of the form

$$\mu : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n \tag{18}$$

$$\sigma : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^{n\times d} \,. \tag{19}$$

While we will not "fuzz about it" in these notes, technically speaking one has to make some assumptions about the regularity of the functions μ, σ. It is common to assume they satisfy a uniform Lipschitz condition. That is, there exists $K \geq 0$ such that

$$|\mu(t, x, u) - \mu(t, y, u)| + |\sigma(t, x, u) - \sigma(t, y, u)| \leq K|x - y| \,, \tag{20}$$

for all $t \in \mathbb{R}_+$, $x, y \in \mathbb{R}^n$ and $u \in \mathbb{R}^d$.

We say that the control process u is *adapted* to X if at time t the value of the control process, u_t, is only allowed to depend on past observed values of our state process, X. This simply means that the control process cannot "peek ahead" at future realization of the state process. One possibility to obtain an adapted control process is by choosing a deterministic function $g(t, x)$

$$g : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}^k \tag{21}$$

and defining the control process u by

$$u_t := g\left(t, X_t^u\right) \,. \tag{22}$$

Such function g is called a *feedback control law.* In the following, we will restrict ourselves to feedback control laws only. To ease the notation, we denote control laws by $u(t,x)$ rather than $g(t,x)$, and use the shorthand $u_t := u(t, X_t^u)$. Note that if we insert a given control law $u(t,x)$ into (16) we obtain the familiar SDE

$$dX_t^u = \mu^u\left(t, X_t^u\right) dt + \sigma^u\left(t, X_t^u\right) dW_t\,, \tag{23}$$

where

$$\mu^u\left(t, X_t^u\right) := \mu^u\left(t, X_t^u, u(t, X_t)\right) \tag{24}$$

$$\sigma^u\left(t, X_t^u\right) := \sigma^u\left(t, X_t^u, u(t, X_t)\right)\,. \tag{25}$$

Frequently, we want the control to satisfy some constraints. We will denote the "permissible" subset after the constraints have been taken into account by $U \subseteq \mathbb{R}^k$. We say that u is *admissible* if $u(t,x) \in U$ for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^n$ and the SDE

$$dX_t^u = \mu\left(t, X_t^u, u(t, X_t)\right) dt + \sigma\left(t, X_t^u, u(t, X_t)\right) dW_t \tag{26}$$

$$X_0 = x_0\,, \tag{27}$$

has a unique solution for some initial state x_0. We denote the set admissible control laws for $0 \leq t \leq T$ by $\mathcal{A}_{0,T}$.

3.1. Stochastic Control Problem. Having defined the controlled process we are interested in, next we formulate our objective function. In this context the objective function is often referred to as the *gain function.*

For $u \in \mathcal{A}_{0,T}$ we consider *gain functions* of the form

$$H^u(x) := \mathbb{E}_x\Big[\underbrace{G\left(X_T^u\right)}_{\text{terminal reward}} + \underbrace{\int_0^T F\left(s, X_s^u, u_s\right) ds}_{\text{running reward/penalty}}\Big]\,, \tag{28}$$

where $\mathbb{E}_x[\cdot]$ is the expectation conditional on the initial state $X_0^u = x$, $G : \mathbb{R}^n \to \mathbb{R}$ is the *terminal reward function* and $F : \mathbb{R}_+ \times \mathbb{R}^{n+p} \to \mathbb{R}$ is the *running penalty/reward function.* The numerical values of F, G are referred to as the running penalty/reward and terminal value, respectively. In general, the running penalty/reward may depend on time,

t, the current position of the controlled process, X_t^u, and the control itself, u_t. The terminal reward depends solely on the terminal value of the controlled process, X_T^u. To avoid mathematical technicalities, we assume the functions G and F are uniformly bounded. Our objective is to maximize the gain function H^u over all admissible controls. We define the *value function* H as

$$H(x) := \sup_{u \in \mathcal{A}_{0,T}} H^u(x)\,. \tag{29}$$

Given an initial condition $x \in \mathbb{R}^n$, we say that $\hat{u} \in \mathcal{A}_{0,T}$ is *an optimal control* if $H(x) = H^{\hat{u}}(x)$.

4. Examples

4.1. The Merton Problem. In this classical problem by Merton (1975) and Merton (1992), we seek to maximize expected discounted wealth by continuously allocating one portion of our wealth to the market portfolio and another to the risk-free bank account.

The problem is typically set up in the following way. We assume the state variables evolve according to

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma S_t dW_t, \quad S_0 = S, \\ dX_t^\pi &= (\pi_t \mu + (1 - \pi_t) r)\, X_t^\pi dt + \pi_t \sigma dW_t, \quad X_0^\pi = x\,, \end{aligned}$$

where

- S_t is the price of the market portfolio at time t,
- X_t is our wealth (in dollars) at time t,
- μ is the continuously compounded rate of growth of the traded asset,
- r is the continuously compounded rate of return of the risk-free bank account,
- $\pi = (\pi_t)_{\{0 \le t \le T\}}$ is a self-financing trading strategy corresponding to having π_t invested in the risky asset at time t and the remaining funds in the risk-free bank account, and
- $X^\pi = (X_t^\pi)_{\{0 \le t \le T\}}$ is the agent's discounted wealth process subject to us following the self-financing strategy π.

Our objective is to find the optimal trading strategy π^* such that

$$\pi^* = \text{argsup}_{\pi \in \mathcal{A}_{0,T}} \mathbb{E}_{S,x} \left[U \left(X_T^\pi \right) \right] , \tag{30}$$

where

- $U(x)$ is our utility function (such as power utility x^γ or exponential utility $-e^{-\gamma x}$), and
- $\mathcal{A}_{0,T}$ is the admissible set of $\mathcal{F}$-predictable self-financing strategies such that $\int_0^T \pi_s^2 ds < \infty$.

4.2. The Optimal Liquidation Problem. In the optimal execution problem, we have a large number of shares q of a stock that we want to liquidate by time T. We assume the state variables evolve according to

$$dQ_t^\nu = -\nu_t dt, \quad Q_0^\nu = q \tag{31}$$

$$dS_t^\nu = -g\left(\nu_t\right) dt + \sigma dW_t, \quad S_0^\nu = S, \tag{32}$$

$$dX_t^\nu = \nu_t \hat{S}_t^\nu dt, \quad X_0^\nu = x\,, \tag{33}$$

where

- $\nu = (\nu_t)_{\{0 \le t \le T\}}$ is the rate of trading,
- $Q^\nu = (Q_t^\nu)_{\{0 \le t \le T\}}$ is our inventory,
- $X^\nu = (X_t^\nu)_{\{0 \le t \le T\}}$ is the cash process,
- $S^\nu = (S_t^\nu)_{\{0 \le t \le T\}}$ is the price process,
- $g : \mathbb{R}_+ \to \mathbb{R}_+$ is the permanent impact on the price from trading,
- The price we can sell stock for is given by

$$\hat{S}_t^\nu = S_t^\nu - f\left(\nu_t\right), \quad \hat{S}_0^\nu = \hat{S} \tag{34}$$

  where

$$f : \mathbb{R}_+ \to \mathbb{R}_+ \tag{35}$$

  is the temporary impact from trading. One can add bid-ask spread here as well, if desired.

Our objective is to find the optimal trading strategy ν^* such that

$$\nu^* = \operatorname{argsup}_{\nu \in \mathcal{A}_{0,T}} \mathbb{E}\left[X_T^\nu + Q_T^\nu\left(S_T^\nu - \alpha Q_T^\nu\right) - \phi \int_0^T \left(Q_s^\nu\right)^2 ds\right]$$

where $\mathcal{A}_{0,T}$ is the admissible set of $\mathcal{F}$-predictable non-negative and bounded strategies.

5. The Dynamic Programming Principle

The key idea in solving these problems is to introduce a larger class of problems that we index by time, t. Specifically, let us define

$$H^u(t,x) := \mathbb{E}_{t,x}\left[G\left(X_T^u\right) + \int_t^T F\left(s, X_s^u, u_s\right) ds\right] \tag{36}$$

$$H(t,x) := \sup_{u \in \mathcal{A}_{t,T}} H^u(t,x)\,, \tag{37}$$

where $\mathbb{E}_{t,x}$ denotes the expectation conditional on $X_t^u = x$. Observe that $H(0,x)$ coincides with our control problem (29) and $H^u(0,x)$ with the gain function (28).

Then we apply the *dynamic programming principle* (DPP) to this class of problem. DPP is a fundamental principle for deterministic and stochastic control problems that in the context of controlled diffusion processes takes the following form.

Theorem 2 (The Dynamic Programming Principle)**.** *Consider the controlled diffusion process defined above. For all points* $(t,x) \in [0,T] \times \mathbb{R}^n$ *we have*

$$H(t,x) = \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x}\left[H\left(\tau, X_\tau^u\right) + \int_t^\tau F\left(s, X_s^u, u_s\right) ds\right]. \tag{38}$$

Remark. Let us make a few remarks here

(1) The DPP is really a *sequence of equations* that connects the value function to its future expected value, plus a running reward or cost.
(2) The interpretation of the DPP is that the optimization problem can be split in two parts. An optimal control on the whole time interval $[t,T]$ may be obtained by first searching for an

optimal control from time τ given the state value X^u_τ, i.e. compute $H\left(\tau, X^u_\tau\right)$, and then maximizing over controls on $[t,\tau]$ the quantity

$$\mathbb{E}\left[\int_t^\tau F\left(s, X^u_s, u_s\right) ds + H\left(\tau, X^u_\tau\right)\right] \tag{39}$$

Proof. [Proof of the DPP]

Step 1. For an arbitrary admissible control $u \in \mathcal{A}_{t,T}$ we have

$$\begin{aligned}
H^u(t,x) &= \mathbb{E}_{t,x}\left[G\left(X^u_T\right) + \int_t^T F\left(s, X^u_s, u_s\right) ds\right] \qquad (40)\\
&= \mathbb{E}_{t,x}\left[G\left(X^u_T\right) + \int_\tau^T F\left(s, X^u_s, u_s\right) ds + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right]\\
&= \mathbb{E}_{t,x}\Bigg[\mathbb{E}_{\tau, X^u_\tau}\left[G\left(X^u_T\right) + \int_\tau^T F\left(s, X^u_s, u_s\right) ds\right]\\
&\qquad + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\Bigg] \quad \text{(tower rule)}\\
&= \mathbb{E}_{t,x}\left[H^u\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right] \quad (\text{def. of } H^u(t,x)) \qquad (41)
\end{aligned}$$

Let us denote by $\hat{u}$ the optimal control. Then, by definition $H^u(t,x) \leq H(t,x)$, with equality if $u \equiv \hat{u}$. Therefore, (41) becomes

$$\begin{aligned}
H^u(t,x) &\leq \mathbb{E}_{t,x}\left[H\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right] \qquad (42)\\
&\leq \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x}\left[H\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right], \qquad (43)
\end{aligned}$$

from which we conclude that

$$H(t,x) \leq \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x}\left[H\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right]. \tag{44}$$

Observe that on the right-hand sides of (42)–(44), the arbitrary control u only acts over the interval $[t,\tau]$ while the optimal control is implicitly incorporated in the value function $H\left(\tau, X^u_\tau\right)$ starting from the point, X^u_τ, at which u caused the process X^u to flow.

Step 2. Now we show that the inequality (44) can be reversed. For $\epsilon > 0$, let us denote by u^ϵ a control that performs better than $H(t,x) - \epsilon$, but not as good as $H(t,x)$. In other words, for this control we have that

$$H(t,x) \geq H^{u^\epsilon}(t,x) \geq H(t,x) - \epsilon\,. \tag{45}$$

u^ϵ is referred to as an ϵ-*optimal control*. As a minor technical comment, we mention that such a control exists assuming that the value function is continuous in the space of controls.

Now let us define $\tilde{u}^\epsilon_t$ to be a modified control of u^ϵ between t and τ by an arbitrary control u, that is

$$\tilde{u}^\epsilon_t := u_t \cdot \mathbb{1}_{t \leq \tau} + u^\epsilon_t \cdot \mathbb{1}_{t > \tau}\,. \tag{46}$$

Then,

$$\begin{aligned} H(t,x) &\geq H^{\tilde{u}^\epsilon}(t,x) \\ &= \mathbb{E}_{t,x}\left[H^{\tilde{u}^\epsilon}\left(\tau, X^{\tilde{u}^\epsilon}_\tau\right) + \int_t^\tau F\left(s, X^{\tilde{u}^\epsilon}_s, \tilde{v}^\epsilon_s\right) ds\right] \qquad \text{(tower rule)} \\ &= \mathbb{E}_{t,x}\left[H^{u^\epsilon}\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right] \quad \text{(modified control)} \\ &\geq \mathbb{E}_{t,x}\left[H\left(\tau, X^u_\tau\right) - \epsilon + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right] \end{aligned}$$

Hence, letting $\epsilon \to 0$ and then taking the supremum we obtain

$$H(t,x) \geq \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x}\left[H\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right] \tag{47}$$

since u is arbitrary. Combining the two inequalities (44) and (47) we obtain

$$H(t,x) = \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x}\left[H\left(\tau, X^u_\tau\right) + \int_t^\tau F\left(s, X^u_s, u_s\right) ds\right], \tag{48}$$

which completes the proof. □

6. The Dynamic Programming and Hamilton-Jacobi-Bellman Equations

The *dynamic programming equation* (DPE), also referred to as the *Hamilton-Jacobi-Bellman equation* (HJB), is the infinitesimal version of the DPP we derived in the previous section. The theorem below states the result.

We consider the controlled diffusion process defined above and assume the following properties hold:

(1) Existence: There exists an optimal control law $\hat{u}$.
(2) Regularity: The optimal value function H is $C^{1,2}$.
(3) Some additional technical details (limiting procedures) can be justified (see, for example, Øksendal (2013) and Pham (2009)) for details.

Theorem 3 (The Dynamic Programming Equation / The Hamilton–Jacobi-Bellman Equation)**.** *Given the previous assumptions, the following results hold:*

(1) H satisfies the HJB

$$\partial_t H(t,x) + \sup_{u \in \mathcal{A}_{t,T}} \left(\mathcal{L}_t^u H(t,x) + F(t,x,u) \right) = 0 \tag{49}$$

$$H(T,x) = G(x) \tag{50}$$

for all $(t,x) \in (0,T) \times \mathbb{R}^n$ and all $x \in \mathbb{R}^n$.

(2) For each $(t,x) \in [0,T] \times \mathbb{R}^n$ the supremum in the HJB equation above is attained by $u \equiv \hat{u}(t,x)$.

Observe that the optimization of the control in the theorem above is only over its value at time t, and not over the whole path of the control. Hence, the optimal control can be obtained pointwise. A common way to solve these problems is to first assume the value function is known and then solve for the optimal control in feedback form in terms of the value function itself. After the feedback control has been obtained, it can be substituted back into (49) resulting in a non-linear PDE. See, for example, Björk (2009) for more details. In later lectures we will apply this technique in solving practical trading problems.

PROOF. [Proof of the DPE/HJB] There are two key elements to the proof:

(1) We will choose the stopping time $\tau < T$ in the DPP to be the minimum between the following:
   - the time it takes the process X_t^u to exit a neighborhood of size ϵ around its starting point, and
   - a fixed (small) time h.

   We state this mathematically by defining τ as

$$\tau := T \wedge \inf\left\{s > t : (s - t, |X_s^u - x|) \notin [0, h) \times [0, \epsilon)\right\} . \tag{51}$$

   From the definition of the stopping time τ we observe that as $h \to 0$ we have that $\tau \to t$ a.s. Therefore, we can always choose h small enough such that $\tau = t + h$.

(2) We will apply Itô's Formula to $H(\tau, X_\tau^u)$.

Step 1. From the DPP we have that

$$H(t, x) \geq \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x}\left[H\left(\tau, X_\tau^u\right) + \int_t^\tau F\left(s, X_s^u, u_s\right) ds\right] \tag{52}$$

Therefore, for an arbitrary admissible strategy v on the interval $[t, \tau]$ we obtain

$$H(t, x) \geq \mathbb{E}_{t,x}\left[H\left(\tau, X_\tau^v\right) + \int_t^\tau F\left(s, X_s^v, v\right) ds\right] . \tag{53}$$

Applying Itô's Formula to the first term of the right hand side of (53), we obtain

$$\begin{aligned} H\left(\tau, X_\tau^v\right) = H\left(t, X_t\right) &+ \int_t^\tau \left(\partial_t + \mathcal{L}_s^v\right) H\left(s, X_s^v\right) ds \\ &+ \int_t^\tau \partial_x H\left(s, X_s^v\right) \sigma\left(s, X_s^v, v\right) dW_s \end{aligned} \tag{54}$$

where $\mathcal{L}_t^v$ is the infinitesimal generator. Inserting (54) into (53), we have

$$\begin{aligned} H(t, x) \geq \mathbb{E}_{t,x}\Big[&H\left(t, X_t\right) + \int_t^\tau \left(\partial_t + \mathcal{L}_s^v\right) H\left(s, X_s^v\right) ds \\ &+ \int_t^\tau \partial_x H\left(s, X_s^v\right) \sigma\left(s, X_s^v, v\right) dW_s + \int_t^\tau F\left(s, X_s^v, v\right) ds\Big] \end{aligned} \tag{55}$$

Since $|X_t^v - x| < \epsilon$ on $[t, \tau]$, the stochastic integral in (55) is a martingale so its expectation vanishes, giving us

$$H(t,x) \geq \mathbb{E}_{t,x}\left[H(t,x) + \int_t^\tau \Big\{ \left(\partial_t + \mathcal{L}_s^v\right) H\left(s, X_s^v\right) + F\left(s, X_s^v, v\right) \Big\} ds\right] . \tag{56}$$

Canceling the $H(t,x)$ terms and dividing by h, we obtain

$$\begin{aligned} 0 &\geq \lim_{h \to 0} \mathbb{E}_{t,x}\left[\frac{1}{h}\int_t^\tau \Big\{ \left(\partial_t + \mathcal{L}_s^v\right) H\left(s, X_s^v\right) + F\left(s, X_s^v, v\right) \Big\} ds\right] \\ &= \left(\partial_t + \mathcal{L}_t^v\right) H(t,x) + F(t,x,v) . \end{aligned} \tag{57}$$

The equality in (57) follows from the following observations:

(1) As $h \to 0, \tau = t + h$ a.s. since the process will not hit the barrier of ϵ in very small time periods.
(2) $|X_\tau^u - x| \leq \epsilon$ implies that if the process does hit the barrier it is bounded.
(3) The mean-value theorem allows us to write

$$\lim_{h \to 0} \frac{1}{h} \int_t^{t+h} \omega_s ds = \omega_t . \tag{58}$$

(4) The process starts at $X_t^v = x$.

The inequality (55) holds for any arbitrary admissible v and consequently

$$\partial_t H(t,x) + \sup_{u \in \mathcal{A}} \left(\mathcal{L}_t^u H(t,x) + F(t,x,u)\right) \leq 0 . \tag{59}$$

Step 2. It remains to show that the inequality is in fact an equality. Suppose that $\hat{u}$ is an optimal control, then from the DPP we have

$$H(t,x) = \mathbb{E}_{t,x}\left[H\left(\tau, x_\tau^{\hat{u}}\right) + \int_t^\tau F\left(s, x_s^{\hat{u}}, \hat{u}\right) ds\right] \tag{60}$$

As before, we apply Itô's Formula to express $H\left(\tau, x_\tau^{\hat{u}}\right)$ in terms of $H(t,x)$ plus the integral of its increments, taking expectations, and then the limit as $h \to 0$ we find that

$$\partial_t H(t,x) + \mathcal{L}_t^{\hat{u}} H(t,x) + F\left(t, x, \hat{u}\right) = 0 , \tag{61}$$

which completes the proof. $\square$

7. The Verification Theorem

The DPE/HJB from the previous section provides a necessary condition for the value function. But is it sufficient? Yes, under certain conditions it is.

The so-called *verification theorem*, stated below, asserts that:

(1) given a smooth enough solution to the DPE/HJB equation, and
(2) the resulting control being admissible, then this candidate coincides with the value function.

Theorem 4 (The Verification Theorem)**.**

(1) Suppose $\psi \in C^{1,2}\left([0,T]\times\mathbb{R}^n\right)$ and that it satisfies, for all $u \in \mathcal{A}$,

$$\partial_t\psi(t,x) + \left(\mathcal{L}_t^u\psi(t,x) + F(t,x,u)\right) \le 0, \text{ for all } (t,x) \in [0,T]\times\mathbb{R}^n$$
$$G(x) - \psi(T,x) \le 0\,.$$

Then

$$\psi(t,x) \ge H^u(t,x), \text{ for all } (t,x) \in [0,T]\times\mathbb{R}^n \tag{62}$$

for all Markov controls $u \in \mathcal{A}$.

(2) Furthermore, if for every $(t,x) \in [0,T]\times\mathbb{R}^n$, there exists (measurable) $\hat{u}(t,x)$ such that

$$0 = \partial_t\psi(t,x) + \left(\mathcal{L}_t^{\hat{u}(t,x)}\psi(t,x) + F\left(t,x,\hat{u}(t,x)\right)\right) \tag{63}$$
$$= \partial_t\psi(t,x) + \sup_{u\in\mathcal{A}_{t,T}}\left(\mathcal{L}_t^u\psi(t,x) + F(t,x,u)\right)\,, \tag{64}$$

for all $(t,x) \in [0,T]\times\mathbb{R}^n$ with $\psi(T,x) = G(x)$. And if the SDE

$$dX_s^{\hat{u}} = \mu\left(t, X_s^{\hat{u}}, \hat{u}\left(t, X_s^{\hat{u}}\right)\right)dt + \sigma\left(t, X_s^{\hat{u}}, \hat{u}\left(t, X_s^{\hat{u}}\right)\right)dW_s$$
$$X_t^{\hat{u}} = x$$

admits a unique solution and $\left\{\hat{u}\left(s, X_s^{\hat{u}}\right)\right\}_{t\le s\le T} \in \mathcal{A}$, then

$$H(t,x) = \psi(t,x), \text{ for all } (t,x) \in [0,T]\times\mathbb{R}^n \tag{65}$$

and $\hat{u}$ is an optimal Markov control.

The proof of this theorem can be found in, for example, Øksendal (2013) and Pham (2009).

References

Björk, Tomas (2009). *Arbitrage theory in continuous time*. Oxford university press.

Cartea, Álvaro, Sebastian Jaimungal, and José Penalva (2015). *Algorithmic and high-frequency trading*. Cambridge University Press.

Merton, Robert C (1975). "Optimum consumption and portfolio rules in a continuous-time model". In: *Stochastic optimization models in finance*. Elsevier, pp. 621–661.

Merton, Robert C. (1992). "Continuous-Time Finance". In.

Øksendal, Bernt (2013). *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media.

Pham, Huyên (2009). *Continuous-time stochastic control and optimization with financial applications*. Vol. 61. Springer Science & Business Media.

Touzi, Nizar (2012). *Optimal stochastic control, stochastic target problems, and backward SDE*. Vol. 29. Springer Science & Business Media.