
METHODS OF OPTIMAL CONTROL

ARASH FAHIM • •

Last Revision: October 21, 2025

Contents

1	Deterministic control	1
1.1	Optimization Versus Control	1
1.2	Solution methods for deterministic optimal control	3
1.2.1	Dynamic programming principle (DPP)	3
1.2.2	An introduction to viscosity solution	10
1.2.3	Summary	13
1.2.4	Pontryagin principle	14
1.3	Linear quadratic control problem	16
2	Stochastic Control	19
2.1	Noise and Brownian motion	19
2.2	Stochastic differential equation (SDE)	21
2.2.1	What do we need to know about Wiener process and Itô integral?	23
2.2.2	Itô's formula	24
2.2.3	Controlled SDE	26
2.2.4	Admissibility	27
2.2.5	Value function	27
2.2.6	Dynamic programming principle (DPP) for stochastic control	28
2.2.7	DPP with stopping time	31
2.2.8	Why is stopping time in DPP important?	32
3	Numerical optimal control	35
3.1	Approximating functions	35
3.1.1	Regression	35
3.1.2	Neural networks	36
3.1.3	Comparison between polynomial approximation and neural network	37
3.2	Policy gradient method for optimal control	38
3.2.1	Loss function for control problems and functional optimization	38
3.2.2	Parametrization of functions	38
3.2.3	Discretization	38
3.2.4	Monte Carlo sampling and empirical loss function	39
3.2.5	Design of PGM	39
4	Old old stuff Stochastic Control	41
4.0.1	Validity of HJB equation	43
4.1	Viscosity solutions to the HJB equations	44
4.2	Verification	47

4.2.1	Martingale approach	49
4.3	Some nonstandard HJB equations	51
4.3.1	Stochastic singular control problems	51
4.3.2	Optimal stopping problem	53
4.4	Backward stochastic differential equations	53
4.4.1	Linear BSDEs	57
4.4.2	Comparison principle	57
4.4.3	Maximum principle	58
5	Numerical evaluation of stochastic control problems	59
5.1	DPP based approximation	59
5.1.1	Evaluation of infimum	59
5.2	Multi-step evaluation	60
5.3	Numerical methods based on BSDEs	61
.1	Results from optimization	61
.1.1	Lagrange multiplier	61
.1.2	Gradient descent	64

Notations

$(\Omega, \mathcal{F}, \mathbb{P})$	a probability space with a σ -field \mathcal{F} and probability \mathcal{P}
$\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$	a filtration in a probability space
\mathbb{F}^X	filtration generated by process X
\mathbb{F}^{X+}	$\cup_{s>t} \mathcal{F}_s^X$
$\bar{\mathbb{F}}$	filtration augmented by adding all the \mathbb{P} -null events
$(\Omega, \bar{\mathbb{F}}, \mathbb{P})$	a filtered probability space with right-continuous augmented filtration (usual conditions) that hosts a Brownian motion adapted to the filtration $\bar{\mathbb{F}}$
$O \subseteq \mathbb{R}^d$	open set
$\text{USC}(O)$ ($\text{LSC}(O)$)	lower semicontinuous (upper semicontinuous) functions on O
$\mathcal{C}^k(O)$	k times continuously differentiable functions on O
$\bar{\mathcal{C}}^k(\bar{O})$	functions on \bar{O} k times continuously differentiable on all variables over on O with derivatives continuously extendable to \bar{O}
$\mathcal{C}^{k,l,\dots}(O)$	functions on O k times continuously differentiable on first variable, l times on second variable, ...
$\bar{\mathcal{C}}^{k,l,\dots}(\bar{O})$	functions on \bar{O} times continuously differentiable on first variable, l times on second variable, ... on O with derivatives continuously extendable to \bar{O}
$M(n, m)$	the set of all real n by m matrices
$A \cdot B := \text{Tr}[AB^T]$	the inner product of two matrices A and B in $M(n, m)$
$\ A\ := A \cdot A = \sum_{i=1}^n \sum_{j=1}^m a_{ij}^2$	l^2 -norm of matrices in $M(n, m)$
Ω	the sample space of a random event
ω	is reserved for the members of Ω
$\partial_t V, \partial_x V, \partial_{xx} V$	Partial derivatives of a function $V : [0, T] \times \mathbb{R}$ once wrt t , once wrt x and twice wrt x
$\partial_t V, \nabla V, D^2 V$	Partial derivative of a function $V : [0, T] \times \mathbb{R}^2$ once wrt t , gradient of V wrt x and Hessian of V wrt x

Chapter 1

Deterministic control

1.1 Optimization Versus Control

In this chapter, we provide a brief overview of the aspect in which optimization and control are different. We start the chapter with some examples.

Example 1. We start by a quadratic problem. Let $\alpha : [0, T] \rightarrow \mathbb{R}$ be given.

$$\inf \left\{ \int_0^T (x_t^2 - \alpha_t x_t) dt \right\} \quad (1.1.1)$$

where the infimum is over all functions $x : [0, T] \rightarrow \mathbb{R}$. As x can be any mapping, one can solve the following problem for each x_t separately to obtain $x_t^* = \alpha_t/2$.

$$\inf_{x \in \mathbb{R}^d} \left\{ x^2 - \alpha_t x \right\} \quad (1.1.2)$$

The above problem is a dynamic optimization problem.

Example 2. We make the above example more complicated by specifying a dynamics for x_t :

$$\inf \left\{ \int_0^T (x_t^2 - \alpha_t x_t) dt \right\} \quad (1.1.3)$$

where the infimum is over all functions $x : [0, T] \rightarrow \mathbb{R}$ such that for some (Borel measurable) function $u : [0, T] \rightarrow \mathbb{R}$

$$dx_t = (-\beta x_t + u_t) dt, \quad x_0 = x \quad (1.1.4)$$

Note that the existence of the term $u_t dt$ is necessary for the meaning of infimum. This problem is not a simple dynamic optimization problem because one can only choose x such that the dynamic equation (1.1.4) holds for some function u . **Therefore**, we can only choose u and indirectly modify x_t to minimize (1.1.3).

Exercise 1. Solve (1.1.4) for x_t in terms of u . Hint: $d(e^{\beta t} x_t) = e^{\beta t} u_t dt$.

Exercise 2. Show that if there exists a function u such that $\alpha_t = 2 \int_0^t e^{\beta(s-t)} u_s ds$, then u minimizes (1.1.3).

Exercise 3. If we modify an optimal control u in a countable number of points described in the exercise above, does it remain an optimal control? Does the initial value $x_0 = x$ play a role in the problem?

Exercise 4. Assume that there exists no function u such that $\alpha_t = 2 \int_0^t e^{-\beta(s-t)} u_s ds$. What is the minimum value of (1.1.3)?

The problem (1.1.3) is a simple control problem. However, after doing the above exercises, you note that it can simply be reduced to an optimization problem. Such a solution for control problems are called myopic solutions and do not necessarily exist for more interesting control problems. Here is an example which does not allow for a myopic solution.

Example 3. Consider the control problem:

$$\inf \int_0^T (x_t^2 - \alpha_t x_t + u_t^2) dt \quad (1.1.5)$$

where the infimum is over all (Riemann integrable) functions $u : [0, T] \rightarrow \mathbb{R}$ and $x : [0, T] \rightarrow \mathbb{R}$ and $u : [0, T] \rightarrow \mathbb{R}$ satisfy (1.1.4). Unlike the previous example, the choice of u induces a new cost, the term $\int_0^T u_t^2 dt$. Due to this new cost, we cannot freely choose u_t to make $x_t = \alpha_t/2$ optimal. It is possible that $x_t = \alpha_t/2$ is not even optimal.

In the above example, there is a tradeoff between the choice of u_t to minimize the dependence of cost on x_t and to minimize the dependence of cost on u_t itself. In such cases, the tradeoff prevents us from finding a myopic solution. Here is another example.

Example 4. Consider the control problem:

$$\inf \int_0^T (x_t^2 + u_t^2) dt \quad (1.1.6)$$

where the infimum is over all (Riemann integrable) functions $u : [0, T] \rightarrow \mathbb{R}$ and $x : [0, T] \rightarrow \mathbb{R}$ and $u : [0, T] \rightarrow \mathbb{R}$ satisfy $dx_t = (-\beta x_t + u_t)dt$, $x_0 = x$. On one hand, we like x_t to be zero to minimize the cost, but pushing x_t to zero requires application of u_t , which introduces another cost term.

A general control problem is described as

$$\inf_{u \in \mathcal{U}} \int_0^T C(t, x_t, u_t) dt + g(x_T) \quad (1.1.7)$$

where $dx_t = f(x_t, u_t)dt$. The function $C : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathbb{R}$ is called the running cost and $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is called the terminal cost.

Set \mathcal{U} is a set of functions $u : [0, T] \rightarrow \mathbb{R}^n$ called control variable, which is determined by the application or by the wellposedness of the problem. It is crucial to choose a set \mathcal{U} of control variables to fit the proper application. More over, it is also important to choose a set that makes the control problem wellposed. For instance, for the control problem

$$\inf_{u \in \mathcal{U}} \int_0^T (x_t - u_t^2) dt, \quad dx_t = (x_t - u_t) dt \quad (1.1.8)$$

if we choose \mathcal{U} to be the set of all functions $u : [0, T] \rightarrow \mathbb{R}$, by simply choosing u_t very large, the value of the infimum is $-\infty$. However, if we restrict \mathcal{U} to the set of functions $u : [0, T] \rightarrow [-1, \infty)$ (some lower bound on the value), then

$$\inf_{u \in \mathcal{U}} \int_0^T (x_t - u_t^2) dt > -\infty, \quad dx_t = (x_t - u_t) dt. \quad (1.1.9)$$

A suitable set of controls chosen for a specific problem is referred to as the set of all admissible control. We denote this set by \mathcal{U} .

An infinite horizon control problem is accommodated by setting $T = \infty$. For example,

$$\inf_{u \in \mathcal{U}} \int_0^\infty e^{-t} (x_t^2 + u_t^2) dt \quad (1.1.10)$$

The following exercise is an example of infinite horizon control problem.

Exercise 5. Write the following problem as a generic control problems by associating the horizon T , the running cost $C(t, x, u)$ and terminal cost $g(x)$ in (1.1.7):

(Shortest time to exit a bounded domain) Given a bounded domain $D \subset \mathbb{R}^d$, find

$$\inf_u \{t \geq 0 : x_t \notin D\} \quad (1.1.11)$$

where $dx_t = u_t dt$ with control $|u_t| \leq 1$ and $u_t \in \mathbb{R}^d$ and initial position $x_0 = x \in D$.

Exercise 6. In your area of study, find an optimal control problem. Then, write the cost functions and the control variable and determine a set of admissible controls.

1.2 Solution methods for deterministic optimal control

In this section, we propose methods for the optimal control problems. There are two groups of numerical methods, first group are based on the dynamic programming principle (DPP). DPP provides a backward recursive method to solve an optimal control problem. The second group completely avoids DPP and formulates problem as an optimization.

1.2.1 Dynamic programming principle (DPP)

Consider the optimal control problem (1.1.7):

$$\inf_{u \in \mathcal{U}} \int_0^T C(t, x_t, u_t) dt + g(x_T), \quad dx_t = f(x_t, u_t) dt \quad (1.2.1)$$

A key concept in DPP is the value function:

Definition 1.2.1. Let $x_t = x$. Then, the value function of (1.1.7) is defined by

$$V(t, x) := \inf_{u \in \mathcal{U}_t} \int_t^T C(s, x_s, u_s) ds + g(x_T), \quad dx_s = f(x_s, u_s) ds \quad (1.2.2)$$

where \mathcal{U}_t is the set of admissible controls restricted to time interval $[t, T]$.

The value function encodes the outcome of optimal control at time t at point x . The function $V(t, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ returns the optimal value of the control problem for any point x at time t .

Given the value function, the DPP is explained in the following result.

Theorem 1.2.1. [Dynamic Programming Principle] Let $0 \leq t < s \leq T$ and $x_t = x$. Then, the DPP for (1.1.7) is given by

$$V(t, x) = \inf_{u \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_r) dr + V(s, x_s), \quad dx_r = f(r, x_r, u_r) dr \quad (1.2.3)$$

where $\mathcal{U}_{t,s}$ is the set of admissible controls restricted to time interval $[t, s]$.

DPP can be interpreted as the tradeoff between two intervals $[t, s]$ and $[s, T]$. More precisely, to obtain the optimal value of the control problem at time t , DPP finds a balance between the running cost on $[t, s]$, $\int_t^s C(r, x_r, u_r)dr$, and the continuation cost on $[s, T]$, $V(t, x_s)$. Note that choosing a control u on $[t, s]$, determines the position of x_s . This interpretation reveals the proof of DPP.

Proof of DPP, Theorem 1.2.1. To show that DPP holds, we need to decompose a control u on $[t, T]$ into two pieces, u_1 on $[t, s]$ and u_2 on $[s, T]$. More formally,

$$u(r) = \begin{cases} u_1(r) & r \in [t, s] \\ u_2(r) & r \in [s, T] \end{cases}$$

We denote this decomposition by $u = u_1 \oplus u_2$. Note that $u_1 \in \mathcal{U}_{t,s}$ and $u_2 \in \mathcal{U}_s$. Therefore, by (1.2.2),

$$\begin{aligned} V(t, x) &= \inf_{u \in \mathcal{U}_t} \int_t^T C(r, x_r, u_r)dr + g(x_T) \\ &= \inf_{u_1 \in \mathcal{U}_{t,s}} \inf_{u_2 \in \mathcal{U}_s} \int_t^s C(r, x_r, u_1(r))dr + \int_s^T C(r, x_r, u_2(r))dr + g(x_T) \\ &= \inf_{u_1 \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_1(r))dr + \inf_{u_2 \in \mathcal{U}_s} \int_s^T C(r, x_r, u_2(r))dr + g(x_T) \end{aligned}$$

Note that $\inf_{u_2 \in \mathcal{U}_s} \int_s^T C(r, x_r, u_2(r))dr + g(x_T) = V(s, x_s)$. Therefore,

$$V(t, x) = \inf_{u_1 \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_1(r))dr + V(s, x_s)$$

□

Numerical solution based on DPP

Let $\Delta t = \frac{T}{N}$ and set $t_i = i\Delta t$ with $i = 0, \dots, N$. The DPP (1.2.3) for $t = t_i$ and $s = t_{i+1}$ is written by

$$V(t_i, x) = \inf_{u \in \mathcal{U}_{t_i, t_{i+1}}} \int_{t_i}^{t_{i+1}} C(r, x_r, u_r)dr + V(t_{i+1}, x_{t_{i+1}}) \quad (1.2.4)$$

We can approximate $\int_{t_i}^{t_{i+1}} C(r, x_r, u_r)dr \approx C(t_i, x_{t_i}, u_{t_i})\Delta t$ to write

$$V(t_i, x) \approx \inf_u C(t_i, x_{t_i}, u)\Delta t + V(t_{i+1}, x_{t_{i+1}}) \quad (1.2.5)$$

This suggest to approximate the value function and optimal control for the control problem (1.1.7) in Algorithm 1. Throughout this notes, \hat{f} represents an approximation of function f . Therefore, if $V(t_{i+1}, \cdot)$ is approximated by $\hat{V}(t_{i+1}, \cdot)$, then

$$\hat{V}(t_i, x) := \inf_u C(t_i, x_{t_i}, u)\Delta t + \hat{V}(t_{i+1}, x_{t_{i+1}}), \quad x_{t_{i+1}} = x + f(t_i, x, u)\Delta t \quad (1.2.6)$$

suggests the approximation for $V(t_i, \cdot)$. Note that unlike DPP, the infimum in the approximate DPP (1.2.6) is over a $u \in \mathbb{R}^m$ and not the functions $u : [t_i, t_{i+1}] \rightarrow \mathbb{R}^m$. Therefore, \hat{u}^* given by

$$\hat{u}^*(t_i, x) \in \operatorname{argmin}_u C(t_i, x_{t_i}, u)\Delta t + \hat{V}(t_{i+1}, x + f(t_i, x, u)\Delta t)$$

provides a constant control over the interval $[t_i, t_{i+1}]$.

We provide an algorithm derived from discrete DPP in Algorithm 1.

Algorithm 1: Numerical DPP

Parameter: $T, N, f(t, x, u), C(t, x, u)$, and $g(x)$;

$\Delta t = \frac{T}{N}$

Data: $\hat{V}(t_N, x) = g(x)$;

x_i^j for $j = 1, \dots, J$ and $i = 0, \dots, N - 1$;

(x_i^j means the j th discrete point at time t_i .)

1 **for** $i \leftarrow N - 1$ **to** 0 **do**

2 $\hat{x}_{i+1}^j = x_i^j + f(t_i, x_i^j, u) \Delta t$;
 3 $\tilde{V}(t_i, x_i^j) \leftarrow \inf_u C(t_i, x_i^j, u) \Delta t + \hat{V}(t_{i+1}, \hat{x}_{i+1}^j)$;
 4 $\hat{V}(t_i, x)$ obtained from interpolation on $\tilde{V}(t_i, x_i^j)$ for $j = 1, \dots, J$;
 5 $\hat{u}^*(t_i, x_i^j) \in \underset{u}{\operatorname{argmin}} C(t_i, x_i^j, u) \Delta t + \hat{V}(t_{i+1}, \hat{x}_{i+1}^j)$;

6 **return** $\hat{V}(t_i, \cdot)$ and $\hat{u}^*(t_i, \cdot)$ for $i = 0, \dots, N - 1$.

Exercise 7. Why interpolation is required in Algorithm 1? Can we perform the algorithm by only knowing $\hat{V}(t_{i+1}, x_{i+1}^j)$ for all $j = 1, \dots, J$? Note the difference between $\hat{V}(t_{i+1}, x_{i+1}^j)$ and $\hat{V}(t_{i+1}, \hat{x}_{i+1}^j)$ with $\hat{x}_{i+1}^j = x_i^j + f(t_i, x_i^j, u) \Delta t$.

Next example provides a special case where we can evaluate the approximate value function from Algorithm 1 without using interpolation.

Example 5. Consider

$$\inf_{u \in \mathcal{U}} \int_0^T (x_t^2 + u_t^2) dt + \frac{1}{2} x_T^2 - x_T, \quad dx_t = (x_t - u_t) dt \quad (1.2.7)$$

We write the value function by (1.2.2).

$$V(t, x) := \inf_{u \in \mathcal{U}_t} \int_t^T (x_s^2 + u_s^2) ds + \frac{1}{2} x_T^2 - x_T, \quad dx_s = (x_s - u_s) ds. \quad (1.2.8)$$

We cannot find value functions using a myopic argument. We will later find a way to find the value function V . However, In the next exercise, we apply Algorithm 1 to find the approximate value function.

Exercise 8. In example above, write the approximate DPP from time t_i to t_{i+1} . Then, assume that $\hat{V}(t_{i+1}, x) = a_{i+1}x^2 + b_{i+1}x + c_{i+1}$ for some known values a_{i+1} , b_{i+1} , and c_{i+1} . Use optimization of a quadratic function to find $\hat{V}(t_i, x)$. Note that you need to use $\hat{x}_{i+1} = x + (x^2 + u^2) \Delta t$. Does $\hat{V}(t_i, x)$ is of the form $a_i x^2 + b_i x + c_i$? What is the relation between (a_i, b_i, c_i) and $(a_{i+1}, b_{i+1}, c_{i+1})$?

Example (8) is a linear-quadratic optimal control problem, LQC. In LQC, the running cost is quadratic in x and u , the terminal cost is quadratic in x , and the ODE for state variable is linear in x and u .

Remark 1.2.1. Consider the discretized version of the control problem (1.1.7):

$$\inf \left\{ \sum_{i=0}^{N-1} C(t_i, \hat{x}_{t_i}, u_{t_i}) \Delta t + g(\hat{x}_T) \mid u_{t_i} : i = 0, \dots, N - 1 \right\} \quad (1.2.9)$$

with $\hat{x}_{t_{i+1}} = \hat{x}_{t_i} + f(t_i, \hat{x}_{t_i}, u_{t_i})\Delta t$. The DPP for this discrete-time optimal control problem is the same as the discretized DPP (1.2.6).

Hamilton-Jacobi equation

One of the side-products of DPP is a partial differential equation, Hamilton-Jacobi equation or HJ for short, that can be solved via different methods to provide us with the solution of the control problem. To derive the HJ, we first assume that the value function is sufficiently continuously differentiable, first order derivatives exist and are continuous. Then, we use the first Taylor polynomial with the remainder term for the value function $V(s, x_s)$ about (t, x) as follows:

$$V(s, x_s) = V(t, x) + V_t(t, x)(s - t) + V_x(t, x)(x_s - x) + R_2$$

Note that $x_s = x + \int_t^s f(x_r, u_r)dr$. Let's insert the above Taylor polynomial into the DPP:

$$\begin{aligned} V(t, x) &= \inf_{u \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_r)dr + V(s, x_s) \\ &= \inf_{u \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_r)dr + V(t, x) + V_t(t, x)(s - t) + V_x(t, x) \int_t^s f(x_r, u_r)dr + R_2 \\ &= V(t, x) + V_t(t, x)(s - t) + \inf_{u \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_r)dr + V_x(t, x) \int_t^s f(x_r, u_r)dr + R_2 \end{aligned}$$

$V(t, x)$ on both sides can be canceled. Note that $R_2 = o(s - t)$, which means $\lim_{s \rightarrow t} \frac{R_2}{s - t} = 0$. Dividing both sides by $s - t$ and sending $s \rightarrow t$, we obtain

$$0 = V_t(t, x) + \inf_u \{C(t, x, u) + V_x(t, x)f(t, x, u)\} \quad (1.2.10)$$

In the above, we used the fundamental theorem of calculus to write $\lim_{s \rightarrow t} \frac{\int_t^s f(x_r, u_r)dr}{s - t} = f(t, x, u)$ and $\lim_{s \rightarrow t} \frac{\int_t^s C(r, x_r, u_r)dr}{s - t} = C(t, x, u)$. A PDE requires proper boundary condition. For HJ, the boundary condition is the terminal cost at T :

$$V(T, x) = g(x) \quad (1.2.11)$$

The nonlinear function in the HJ equation

$$H(t, x, p) := \inf_u \{C(t, x, u) + pf(t, x, u)\} \quad (1.2.12)$$

is called Hamiltonian. We conclude that the value function of the control problem satisfies the HJ equation:

$$\begin{cases} 0 = V_t(t, x) + H(t, x, V_x(t, x)) \\ V(T, x) = g(x) \end{cases} \quad (1.2.13)$$

The optimal control $u^*(t, x)$ must satisfy

$$C(t, x, u^*(t, x)) + V_x(t, x)f(t, x, u^*(t, x)) = H(t, x, V_x(t, x)) \leq H(t, x, u), \quad \forall u. \quad (1.2.14)$$

Let's explore the HJ inside an example.

Example 6. Consider the control problem in Section 1.1:

$$\inf_u \left\{ \int_0^T (x_t^2 + u_t^2) dt \right\} \quad (1.2.15)$$

where $dx_t = (-\beta x_t + u_t)dt$. Here, $C(t, x, u) = x^2 + u^2$ and $f(t, x, u) = -\beta x + u$. The HJ is given by:

$$0 = V_t + x^2 - \beta x V_x + \inf_u \{u^2 + V_x u\} = V_t + x^2 - \beta x V_x - \frac{1}{4} V_x^2$$

Note that the infimum above is attained at $u^* = -\frac{1}{2} V_x$ and, therefore, we can write the HJ without using infimum:

$$0 = V_t + H(x, p) \quad \text{with} \quad H(x, p) = x^2 - \beta x p - \frac{1}{4} p^2$$

Since there is no terminal cost, the boundary condition for the HJ is $V(T, x) = 0$.

An important question is that how can the HJ equation help us find the optimal control. For the above example, if we solve the HJ and find the value function, then a candidate for the optimal control is $u_t^* = -\frac{1}{2} V_x(t, x_t)$ where $dx_t = (-\beta x_t + u_t^*)dt$. In general, we cannot guarantee a simple way to solve the HJ. However, for certain problems there is a simple way and for other problems numerical solutions to PDEs can be used. Next, exercise shows how to solve the HJ derived for the previous example.

Exercise 9. We guess that the solution for the HJ

$$\begin{cases} 0 = V_t + x^2 - \beta x V_x - \frac{1}{4} V_x^2 \\ V(T, x) = 0 \end{cases}$$

takes the form $V(t, x) = a(t)x^2 + b(t)x + c(t)$. Insert this guess into the HJ and find ODEs for $a(t)$, $b(t)$, and $c(t)$. Use terminal conditions to prescribe terminal conditions for a , b , and c . Can you solve the system of ODEs?

After solving the HJ equation and obtaining a solution, we often require a verification step. The reason for the necessity of verification is that in certain (rare) cases, there can be multiple solutions for HJ equation with the terminal condition. We want to make sure the solution matches the value function

Theorem 1.2.2. Assume that function $v(t, x)$ is once continuously differentiable with respect to t and twice continuously differentiable with respect to x and satisfies the HJ equation (1.2.13). Further more, assume that there exists a function $u^*(t, x)$ such that (1.2.14) holds and $u^*(t, x_t^*)$ with x_t^* satisfying $dx_t^* = f(t, x_t^*, u^*(t, x_t^*))$ constitutes an admissible control; $u^*(t, x_t^*) \in \mathcal{U}$. Then, $v = V$.

Exercise 10. Does $v(t, x) = a(t)x^2 + b(t)x + c(t)$ in Exercise 9 satisfies the condition of verification theorem, Theorem 1.2.2?

Exercise 11. In epidemiology, modeling a disease in a population is done by splitting the population into three groups: susceptible, infected, and recovered. Then, then number of individuals in each group at time t is denoted by S_t , I_t , and R_t which follow the following ODEs.

$$\begin{cases} dS_t = -\beta_t S_t I_t dt \\ dI_t = (\beta_t I_t S_t - \gamma_t I_t) dt \\ dR_t = \gamma_t I_t dt \end{cases}$$

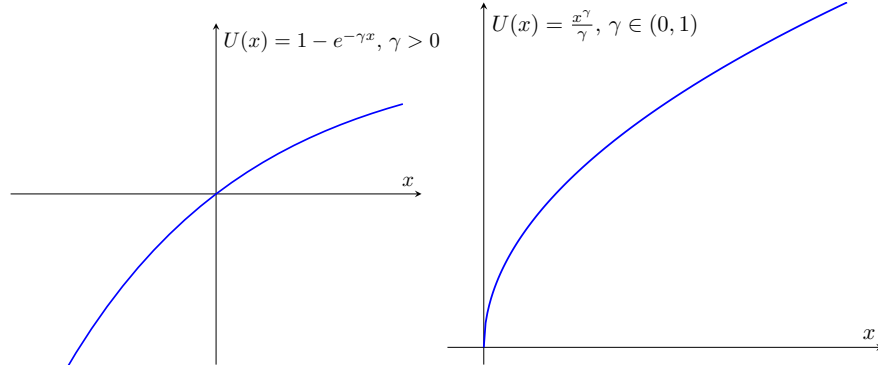


Figure 1.2.1: Utility functions: left $U(x) = 1 - e^{-\gamma x}$, right $U(x) = \frac{x^\gamma}{\gamma}$

Here $\beta_t \in [b_0, b_1]$ is the rate at which a susceptible individual gets infected and $\gamma_t \in [c_0, c_1]$ is the rate at which an infected individual recovers from the disease. We assume that investing in some public health measures decreases β_t and investing in a cure increases γ_t . Both investments come with a cost function, $C(\beta, \gamma) = (b_1 - \beta_t)^2 + \gamma_t^2$. Finally, at a finite time T , we like to have the number of infected individuals as low as possible, $g(I) = I^2$.

$$\inf_{\substack{\beta_t \in [b_0, b_1] \\ \gamma_t \in [c_0, c_1]}} \int_0^T ((b_1 - \beta_t)^2 + \gamma_t^2) dt + I_T^2$$

Write the HJ equation for this control problem.

Exercise 12. In Exercise 11, notice that $d(S_t + I_t + R_t) = 0$. This should allow us to reduce the number of state variables $x_t = (S_t, I_t, R_t)$ to two, in place of three. Assume that the population size is given by N , $S_t + I_t + R_t = N$. Remove the variable R_t and write the HJ equation in terms of (S_t, I_t) .

Exercise 13. The following problem studies the optimal consumption from a savings account. Let the balance of the savings account be given by:

$$dX_t = (rX_t - c_t)dt$$

$c_t \geq 0$ is the rate of consumption.

$$\sup_{c_t \geq 0} \int_0^T U(c_t)dt + U(X_T)$$

Function U , called utility function, is a concave function that represent our enjoyment from consumption or wealth. The concavity signifies the fact that if our consumption or wealth level is low, increasing one more unit grants more joy compared to when our consumption or wealth level is higher and we obtain one more unit. In general, a utility function is an increasing concave function. Example of utility function is given in Figure 1.2.1.

In the above problem, we are maximizing a combination running gain on consumption and terminal gain on wealth. Therefore, the Hamiltonian in the HJ equation has a supremum.

Derive the HJ equation for this problem.

Exercise 14. In the above exercise, if we ignore the terminal gain, we obtain

$$\sup_{c_t \geq 0} \int_0^T U(c_t)dt, \quad dX_t = (rX_t - c_t)dt$$

Is there something wrong with this problem? What stops us from consuming infinite amount and go negative on the wealth?

If we restrict the consumption to maintain a positive balance, $X_t \geq 0$ for all t , does the issue go away?

Exponential decay in control problems

We consider the case where, $C(t, x, u) = e^{-rt} \bar{C}(u, x)$ and $f(t, x, u) = f(x, u)$

$$\inf_u \int_0^T e^{-kt} \bar{C}(u_t, x_t) dt + e^{-kT} g(x_T), \quad dx_t = f(x_t, u_t) dt \quad (1.2.16)$$

The exponential decay term, e^{-kt} emphasized on the importance of cost functions relative to the starting time. As time grow, the contribution of a unit of cost decreases. This is a common practice in applications from economics. While we can write the value function and the dynamic programming principle the same way as (1.2.2) and (1.2.3), we can slightly modify the definition of value function and simplify the DPP and the HJ equation.

Definition 1.2.2. The value function of (1.2.16) for $x_t = x$ is defined by

$$V(t, x) := \inf_{u \in \mathcal{U}_t} \int_t^T e^{-k(s-t)} \bar{C}(x_s, u_s) ds + e^{-k(T-t)} g(x_T), \quad dx_s = f(x_s, u_s) ds \quad (1.2.17)$$

where \mathcal{U}_t is the set of admissible controls restricted to time interval $[t, T]$.

The dynamic programming principle is modified by:

Theorem 1.2.3. [Dynamic Programming Principle] Let $0 \leq t < s \leq T$ and $x_t = x$. Then, the DPP for (1.1.7) is given by

$$V(t, x) = \inf_{u \in \mathcal{U}_{t,s}} \int_t^s e^{-k(r-t)} \bar{C}(x_r, u_r) dr + e^{-k(s-t)} V(s, x_s), \quad dx_r = f(x_r, u_r) dr \quad (1.2.18)$$

where $\mathcal{U}_{t,s}$ is the set of admissible controls restricted to time interval $[t, s]$.

Proof. The proof is quite similar to Theorem 1.2.1. □

We apply Taylor series on $e^{-k(s-t)} V(s, x_s)$ around (t, x) .

$$\begin{aligned} e^{-k(s-t)} V(s, x_s) &= V(t, x) + e^{-k(s-t)} \left((V_t(t, x) - kV(t, x))(s - t) \right. \\ &\quad \left. + V_x(t, x) \cdot \int_t^s f(x_r, u_r) dr \right) + o((s - t)^2) \end{aligned}$$

We use Taylor series inside (1.2.18) to write

$$\begin{aligned} 0 &= \inf_u \int_t^s e^{-k(r-t)} \bar{C}(x_r, u_r) dr \\ &\quad + e^{-k(s-t)} \left((V_t(t, x) - kV(t, x))(s - t) \right. \\ &\quad \left. + V_x(t, x) \cdot \int_t^s f(x_r, u_r) dr \right) + o((s - t)^2) \end{aligned}$$

Dividing both sides by $s - t$ and sending $s \rightarrow t$ yields the HJ equation.

$$0 = V_t(t, x) - kV(t, x) + \inf_u \left\{ \bar{C}(x, u) + V_x(t, x) \cdot f(x_r, u_r) \right\} \quad (1.2.19)$$

The term $-kV$ in the HJ equation is coming from the discounting. The boundary (terminal) condition remains the same, $V(T, x) = g(x)$.

$$\begin{cases} 0 = V_t(t, x) - kV(t, x) + H(x, V_x) \\ V(T, x) = g(x) \end{cases} \quad (1.2.20)$$

with

$$H(x, p) = \inf_u \left\{ \bar{C}(x, u) + p \cdot f(x_r, u_r) \right\}$$

The relation between the HJ equation with decay term in (1.2.19) and the HJ equation in (1.2.10) can easily be verified by the change of variable $\bar{V}(t, x) = e^{k(T-t)}V(t, x)$. $V(t, x)$ satisfies (1.2.20) if and only if $\bar{V}(t, x)$ satisfies (1.2.13).

Exercise 15. Show that $V(t, x)$ satisfies (1.2.20) if and only if $\bar{V}(t, x)$ satisfies (1.2.13).

A good question to ask ourselves is that for Problem (1.2.16) if the the HJ equation with or without decay term can be obtained via a simple change of variable, why do we introduce a different value function and DPP for exponential decay? The answer lies in the case when $T = \infty$:

$$\inf_u \int_0^\infty e^{-kt} \bar{C}(u_t, x_t) dt, \quad dx_t = f(x_t, u_t) dt \quad (1.2.21)$$

Since $f(x_t, u_t)$ does not depend on t , the solution x_s of $dx_s = f(x_s, u_s)ds$ with $x_t = x$ is the same as a time shift of \bar{x}_r the solution of $dx_r = f(x_r, u_r)ds$ with $x_0 = x$ by t .

$$V(t, x) = \inf_{u \in \mathcal{U}_t} \int_t^\infty e^{-k(s-t)} \bar{C}(x_s, u_s) ds = \inf_{u \in \mathcal{U}} \int_0^\infty e^{-kr} \bar{C}(\bar{x}_r, u_r) dr$$

Therefore, value function is time homogeneous and is given by

$$V(x) = \inf_{u \in \mathcal{U}} \int_0^\infty e^{-kr} \bar{C}(\bar{x}_r, u_r) dr \quad (1.2.22)$$

and the HJ equation is given by

$$0 = -kV(t, x) + H(x, V_x) \quad (1.2.23)$$

Exercise 16. Consider the consumption problems, Exercise 13 and Exercise 14, in infinite horizon with decay term:

$$\sup_{c_t \geq 0} \int_0^\infty e^{-kt} U(c_t) dt, \quad dX_t = (rX_t - c_t) dt$$

Here, set of admissible consumption are those such that we maintain $X_t \geq 0$ for all $t \geq 0$. Write the value function by using (1.2.22) and the HJ equation by (1.2.23). Find proper boundary condition and solve the HJ equation with the boundary condition.

1.2.2 An introduction to viscosity solution

To motivate viscosity solutions, we introduce the following example.

Example 7. Recall from Exercise 5 that

$$\inf_u \int_0^\infty \mathbb{1}_{\{x_t \in D\}} dt, \quad dx_t = u_t dt \quad \text{with } |u_t| \leq 1$$

The value function for initial state $x_0 = x \in D$ is given by

$$V(x) = \inf_u \int_0^\infty \mathbb{1}_{\{x_t \in D\}} dt, \quad dx_t = u_t dt \quad \text{with } |u_t| \leq 1$$

Here $C(x, u) = C(x) = \mathbb{1}_{\{x \in D\}}$ and $f(x, u) = f(u) = u$. Notice that the value function does not depend on t , because the total cost is a function of duration of time until exit and not the function of current time.

The HJ equation is given by

$$H(V_x) = 0, \quad H(x, p) := \inf_{|u| \leq 1} \{ \mathbb{1}_{\{x \in D\}} + p \cdot u \}$$

We can restrict the Hamiltonian to points $x \in D$ and write

$$H(p) := \inf_{|u| \leq 1} \{ 1 + p \cdot u \} = 1 + \inf_{|u| \leq 1} \{ p \cdot u \}$$

The infimum above is attained for $u = -\frac{u}{|u|}$. Therefore,

$$H(p) = |p|$$

Outside (or on the boundary of) D , the exit time is zero. Therefore, we restrict the Hamiltonian to inside of the domain and impose the boundary condition $V(x) = 0$ for $x \in \partial D$. We conclude that the HJ equation with the boundary condition is given by

$$\begin{cases} 1 - |V_x| = 0 & x \in D \\ V(x) = 0 & x \in \partial D \end{cases}$$

Example 7 can shed a light on the notion of solution to the HJ equations. Take $D = [-1, 1] \subset \mathbb{R}$ and consider the following functions, shown in Figure 1.2.2:

$$v_1(x) = 1 - |x|, \quad v_2(x) = \begin{cases} \frac{1}{2} - |x - \frac{1}{2}| & 0 \leq x \leq 1 \\ \frac{1}{2} - |x + \frac{1}{2}| & -1 \leq x < 0 \end{cases}$$

The minimal exit time of $[-1, 1]$ is given by v_1 . Additionally, except at point $x = 0$, v_1 is differentiable, satisfies the HJ equation, and the optimal direction and speed is given by $V_x = \text{sgn}(x)$. v_2 is not the value function, but it also satisfies the HJ equation except at three points, $x = 0, \pm \frac{1}{2}$. In fact, with a little bit of effort, one can write infinitely many solutions for the Eikonal equations. All this solutions have points of non-differentiability. However, the right and the left derivatives at those points still satisfy the HJ equation. That raises an issue; if we solve an HJ equation, how do we know if we obtained the correct solution, hence, the value function? The answer to this problem is quite complicated. The theory that solves the issue is the theory of viscosity solutions, which we discuss in more details in a later chapter. Here, we explain a simpler version of this theory by focusing on the Eikonal equation, $0 = 1 - |v'(x)|$ with boundary condition $v(-1) = v(1) = 0$.

We introduce a new equation by adding a second order term.

$$0 = 1 - |v^\epsilon(x)| + \epsilon V^{\epsilon''}, \quad v^\epsilon(-1) = v^\epsilon(1) = 0$$

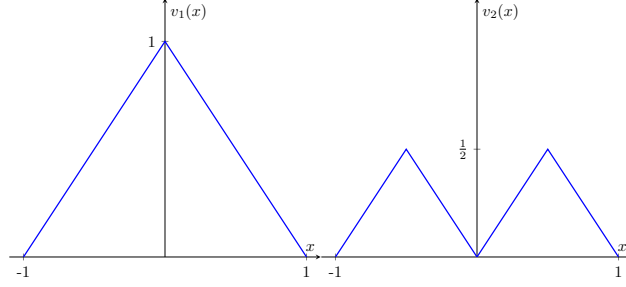


Figure 1.2.2: Functions v_1 and v_2 solve Eikonal equation, the HJ equation for 7. But, only v_1 is the same as the value function.

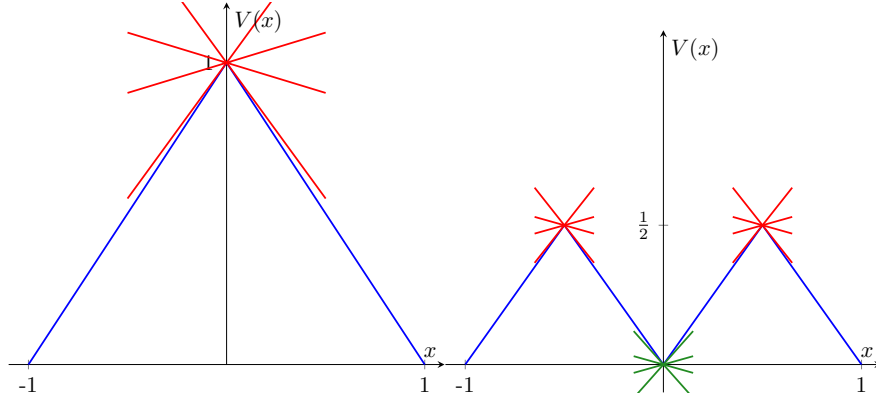


Figure 1.2.3: Superjets and subjets for functions v_1 and v_2 at the points of non-differentiability.

The above equation is a second order equation with two boundary conditions, which guarantees existence of a unique differentiable solution which can be found in closed form by solving

$$\begin{cases} 0 = 1 + v^{\epsilon'}(x) + \epsilon V^{\epsilon''} & x \in (0, 1) \\ 0 = 1 - v^{\epsilon'}(x) + \epsilon V^{\epsilon''} & x \in (-1, 0) \\ v^{\epsilon'}(0) = 0 \\ v^{\epsilon}(0-) = v^{\epsilon}(0+) \\ v^{\epsilon}(-1) = v^{\epsilon}(1) = 0 \end{cases}$$

Exercise 17. Find a closed form solution. above system of equations.

Then, if we evaluate $\lim_{\epsilon \rightarrow 0} v^{\epsilon}$, we obtain $v_1 = V$.

Crandall et al. [1992], codified the above construction into the theory of viscosity solutions, which perfectly fits the need of HJ equations in optimal control. For the points of non-differentiability, they introduces candidates for derivative from above and from below. More precisely, they defined the set of superjets as the slope of all lines that touch the function from above and the set of subjets as the slope of all lines that touch the function from below, as shown in Figure 1.2.3.

The set of superjets, \overline{J} , and subjets, \underline{J} , for all the points of non-differentiability are the closed interval $[-1, 1]$. According to the theory of viscosity solutions, the members of superjet must satisfy the equation with inequality:

$$\forall a \in \overline{J} = [-1, 1], 1 - |a| \geq 0$$

which is trivially correct for v_1 at $x = 0$ and for v_2 at $x = \frac{1}{2} = -\frac{1}{2}$. Similarly, the members of subject must satisfy the equation with inequality in the opposite direction of superjets:

$$\forall a \in \underline{J} = [-1, 1], 1 - |a| \leq 0.$$

However, for v_2 at $x = 0$, $a \in \underline{J} = (-1, 1)$ $1 - |a| \not\leq 0$. Therefore, v_2 is not a viscosity solution of the eikonal equation.

Interesting results from the theory of viscosity solution are as follows.

- 1) The value function of a control problem is a viscosity solution of the HJ equation.
- 2) HJ equation has a unique viscosity solution.
- 3) Numerical solutions to HJ equations with certain standard properties converge to the viscosity solution.

Even though there are more details into these results, we continue to use these results in this note until we require more clarifications.

1.2.3 Summary

The steps of solving deterministic control problems via HJ problem are described in Chart 1.2.4: It is

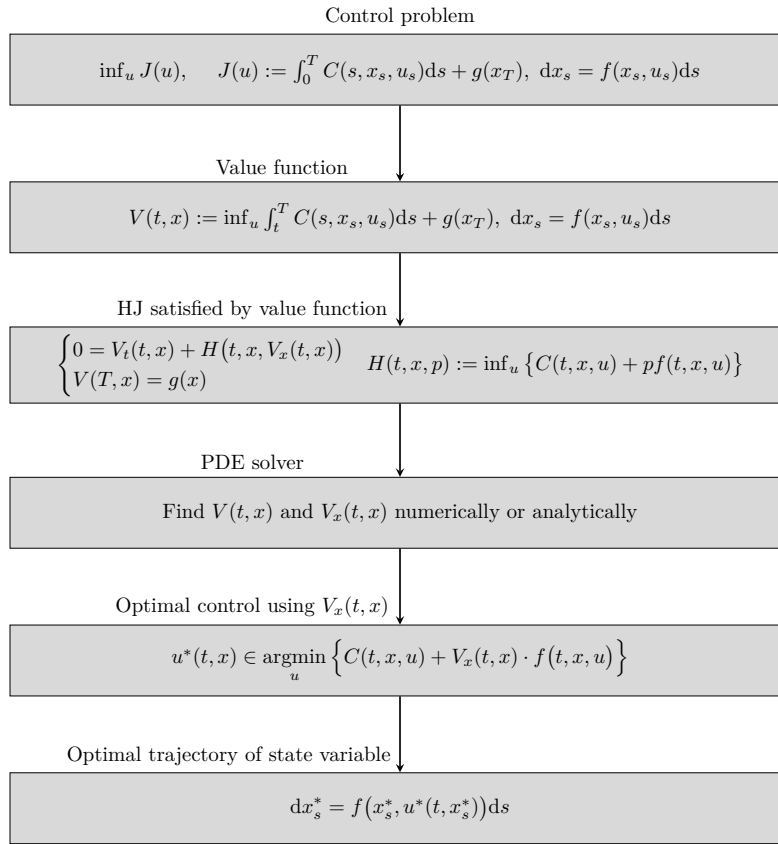


Figure 1.2.4: Summary of HJ solution method for control problems

important to note that when the HJ equation is derived, solving HJ equation is an standalone problem.

After solving the HJ equation, the solution and its derivatives are used to find an optimal control and the corresponding optimal trajectory.

1.2.4 Pontryagin principle

Consider the simple control problem in Section 1.1:

$$\inf_u \left\{ \int_0^T (x_t^2 - \alpha_t x_t) dt \right\} \quad (1.2.24)$$

where $dx_t = (-\beta x_t + u_t)dt$, $x_0 = x$. One can consider the dynamics of x as a constraint. Therefore, we can formally write Lagrangian by

$$\sup_{\lambda} \inf_{u, x} \left\{ \int_0^T (x_t^2 - \alpha_t x_t) dt - \int_0^T \lambda_t (dx_t + (\beta x_t - u_t)dt) \right\} \quad (1.2.25)$$

Here, there are two important remarks. First, since the constraint is given by a differential equation for each t , the dual variable λ is a function of t . Second, the problem is now unconstrained, that is, the state variable x and the control variable u are now both variables in an optimization problem.

The main trick to solve this optimization problem is integration by part formula

$$\int_0^T \lambda_t dx_t = \lambda_T x_T - \lambda_0 x_0 - \int_0^T x_t d\lambda_t \quad (1.2.26)$$

to write (1.2.25) as

$$\sup_{\lambda} \inf_{u, x} \left\{ \int_0^T (x_t d\lambda_t - \lambda_t (\beta x_t - u_t) + x_t^2 - \alpha_t x_t) dt - \lambda_T x_T + \lambda_0 x_0 \right\} \quad (1.2.27)$$

Note that optimization on x_T is independent of x_t for $t < T$. The KKT conditions for the strong duality in the saddle point problem (1.2.27) were discovered by Lev Pontryagin in 1952. Define the Hamiltonian by

$$H(t, x, \lambda, u) := -\lambda(\beta x - u) + x^2 - \alpha_t x \quad (1.2.28)$$

Thus, (1.2.27) is written as a saddle point problem with free variables x_t , u_t , and λ_t :

$$\sup_{\lambda} \inf_{u, x} \left\{ \int_0^T (x_t d\lambda_t + H(t, x_t, \lambda_t, u_t)) dt - \lambda_T x_T + \lambda_0 x_0 \right\} \quad (1.2.29)$$

$$\begin{cases} d\lambda_t^* + \partial_x H(t, x_t^*, \lambda_t^*, u_t^*) dt = d\lambda_t^* + (-\lambda_t^* \beta + 2x_t^* - \alpha_t) dt = 0 \text{ (minimize integrand wrt } x) \\ \lambda_T^* = 0 \text{ (minimize terminal wrt } x_T) \\ H(t, x_t^*, \lambda_t^*, u_t^*) \leq H(t, x_t^*, \lambda_t^*, u) \text{ for all } u \text{ (minimize integrand wrt } u) \\ dx_t^* = (-\beta x_t^* + u_t^*) dt \text{ (constraint)} \end{cases} \quad (1.2.30)$$

In the above, the first equation is obtained from taking derivative with respect to x_t , second equality corresponds to derivative with respect to x_T in $\lambda_T x_T$, the third line guarantees the optimality of u^* , and the last equation is the constraint of the problem which is the dynamic of the state variable in the control.

Exercise 18. Show that $\lambda_t^* = 0$, $x_t^* = \alpha_t/2$, and u^* with $\alpha_t = 2 \int_0^t e^{-\beta(s-t)} u_s^* ds$ satisfy Pontryagin

principle for the above problem.

Following the steps of in the above example, the Pontryagin principle for a generic deterministic control problem, 1.1.7 is described as the following saddle point problem.

$$\sup_{\lambda} \inf_{u, x} \left\{ \int_0^T C(t, x_t, u_t) dt + g(x_T) - \int_0^T \lambda_t (dx_t + (\beta x_t - u_t) dt) \right\} \quad (1.2.31)$$

By applying (1.2.26), we obtain

$$\sup_{\lambda} \inf_{u, x} \left\{ \int_0^T \left(x_t d\lambda_t + H(t, x_t, \lambda_t, u_t) \right) dt + g(x_T) - \lambda_T x_T + \lambda_0 x_0 \right\} \quad (1.2.32)$$

where the Hamiltonian is given by

$$H(t, x, \lambda, u) := \lambda f(t, x, u) + C(t, x, u) \quad (1.2.33)$$

Theorem 1.2.4. [Pontryagin principle] Assume that there exists $(x^*, u^*, \lambda^*) : [0, T] \rightarrow \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^d$ such that

$$\begin{cases} d\lambda_t^* + \partial_x H(t, x_t^*, \lambda_t^*, u_t^*) dt = 0 & (\text{minimize integrand wrt } x) \\ \lambda_T^* = \nabla g(x_T^*) & (\text{minimize terminal wrt } x_T) \\ H(t, x_t^*, \lambda_t^*, u_t^*) \leq H(t, x_t^*, \lambda_t^*, u) & \text{for all } u \text{ (minimize integrand wrt } u) \\ dx_t^* = f(t, x_t^*, u_t^*) dt & (\text{constraint}) \end{cases} \quad (1.2.34)$$

Then, u^* is an optimal control for (1.1.7).

The function λ_t^* , described by 1.2.4 is called the adjoint process. It is important to distinguish between the two ODEs in (1.2.34), namely,

$$dx_t^* = f(t, x_t^*, u_t^*) dt \text{ and } d\lambda_t^* + \partial_x H(t, x_t^*, \lambda_t^*, u_t^*) dt = 0 \quad (1.2.35)$$

The first one has an initial condition $x_0^* = x_0$, the initial position of the state process. However, the adjoint equation comes with a terminal condition, $\lambda_T^* = \nabla g(x_T^*)$. This makes solving Theorem 1.2.4 challenging. We shall see later how Pontryagin principle is applied to some specific examples such as linear quadratic linear control problem, where solving (1.2.34) is simpler to solve.

Remark 1.2.2. If u_t^* is an interior minimizer of $H(t, x_t^*, \lambda_t^*, u)$, then we can write $\partial_u H(t, x_t^*, \lambda_t^*, u_t^*) = 0$. However, if there are constraint on u , e.g., $u \geq 0$, we shall stick to the inequality above.

We start with an example of a LQC problem.

Example 8. Consider the control problem with $C(x, u) = x^2 + u^2$, $g(x) = x^2 + x$, and $f(x, u) = x + u$:

$$\inf_u \int_0^T (x_t^2 + u_t^2) dt + x_T^2 + x_T, \text{ subject to } dx_t = (x_t + u_t) dt \quad (1.2.36)$$

LQC problems can be solved via Pontryagin principle. More precisely, for the above example, the Hamiltonian is given by

$$H(x, \lambda, u) = x^2 + u^2 + \lambda(x + u)$$

and (1.2.34) is

$$\begin{cases} d\lambda_t^* + (2x_t^* + \lambda_t^*)dt = 0 \\ \lambda_T^* = 2x_T^* + 1 \\ H(t, x_t^*, \lambda_t^*, u_t^*) \leq H(t, x_t^*, \lambda_t^*, u) \\ dx_t^* = (x_t^* + u_t^*)dt \end{cases} \quad (1.2.37)$$

Minimizing $H(x, \lambda, u)$ in u suggests that $H_u(x, \lambda_t^*, u_t^*) = 2u_t^* + \lambda_t^* = 0$, equivalently, $u_t^* = -\frac{1}{2}\lambda_t^*$.

Exercise 19. (1) Use $u_t^* = -\frac{1}{2}\lambda_t^*$ to write the system of ODEs (1.2.37) for λ^* and x^* by

$$\begin{cases} d\lambda_t^* = (-\lambda_t^* - 2x_t^*)dt \\ dx_t^* = (-\frac{1}{2}\lambda_t^* + x_t^*)dt \end{cases} \quad (1.2.38)$$

(2) Find the general solution the system of ODEs in terms of x_0^* and λ_0^* . Hint: Use denationalization of the matrix

$$\begin{bmatrix} -1 & -2 \\ -\frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 2(\sqrt{2}+1) & 1 \\ 1 & 2(\sqrt{2}-1) \end{bmatrix} \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{2}{3}(\sqrt{2}-1) & -1 \\ -1 & \frac{2}{3}(\sqrt{2}+1) \end{bmatrix} \quad (1.2.39)$$

and write the ODEs as

$$\begin{bmatrix} d\lambda^* \\ dx^* \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \lambda^* \\ x^* \end{bmatrix} = \begin{bmatrix} 2(\sqrt{2}+1) & 1 \\ 1 & 2(\sqrt{2}-1) \end{bmatrix} \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{2}{3}(\sqrt{2}-1) & -1 \\ -1 & \frac{2}{3}(\sqrt{2}+1) \end{bmatrix} \begin{bmatrix} \lambda^* \\ x^* \end{bmatrix} \quad (1.2.40)$$

(3) Consider x_0^* given. Use $\lambda_T^* = 2x_T^* + 1$ to find λ_0^* , hence a special solution for the system of ODEs from Pontryagin principle as a function x_0^* and x_T^* .

1.3 Linear quadratic control problem

A class of control problem with closed-form solution consists of linear-quadratic optimal control problems, LQC henceforth. In LQC, the running cost is quadratic in x and u , the terminal cost is quadratic in x , and the ODE for state variable is linear in x and u . We assume that x is a d -dimensional column vector and u is an m -dimensional column vector.

$$\begin{aligned} C(t, x, u) &= \begin{bmatrix} x^\top & u^\top \end{bmatrix} Q_t \begin{bmatrix} x \\ u \end{bmatrix} + a_t \cdot x + b_t \cdot u, \text{ and } Q_t = \begin{bmatrix} A_t & C_t \\ C_t^\top & B_t \end{bmatrix} \\ g(x) &= x^\top A_T x + a_T \cdot x \\ f(t, x, u) &= M_t x + N_t u \end{aligned} \quad (1.3.1)$$

Here, $A : [0, T] \rightarrow \mathbb{M}(d, d)$, $B : [0, T] \rightarrow \mathbb{M}(m, m)$, $C : [0, T] \rightarrow \mathbb{M}(d, m)$, $a : [0, T] \rightarrow \mathbb{R}^d$, $b : [0, T] \rightarrow \mathbb{R}^m$, $M : [0, T] \rightarrow \mathbb{M}(d, d)$, and $N : [0, T] \rightarrow \mathbb{M}(d, m)$.

For LQC problem to be well-posed, the value is not $-\infty$, it is sufficient to assume that Q_t and A_T are symmetric and positive definite. In this case C and g are convex functions of (x, u) and x , respectively. See exercise below.

Exercise 20. In Example 8, change the running cost function to $C(x, u) = x^2 - u^2$. Make some effort to find a solution to this modified problem. Of course, since the cost function is not convex, your effort fails. Explain why the optimal control problem is ill-defined.

Exercise 21. Write the HJB equation for the LQC problem.

Exercise 22. Assume that the cost functions of the general LQC problem are convex. Verify that there exists $G : [0, T] \rightarrow \mathbb{M}(d, d)$, $H : [0, T] \rightarrow \mathbb{R}^d$, and $K : [0, T] \rightarrow \mathbb{R}$ such that $V(t, x) = x^\top G_t x + H_t \cdot x + K_t$ satisfied the HJ equation.

Exercise 23. Assume that the cost functions of the general LQC problem are convex. Write Pontryagin principle for the general LQC and suggest a solution method.

Chapter 2

Stochastic Control

In this chapter, we start generalizing the material in the previous section to include stochastic control. The ingredients of a stochastic control problem is similar to the deterministic control: running cost function $C(t, x, u)$, terminal cost function $g(x)$, and dynamics of the state variable. However, the dynamics of the state variable comes with uncertainty modeled with white noise. In the next section, we describe the dynamics of the state variable with a noise component. Such dynamics is called stochastic differential equation, SDE henceforth.

2.1 Noise and Brownian motion

Recall from Section 1 that the dynamics of state variable is given by

$$dX_t = b(t, X_t, u_t)dt$$

When the dynamics is noisy, we simply need to write

$$dX_t = b(t, X_t, u_t)dt + \text{noise}$$

But, what is the noise. Simplest way to model noise is through white noise. White noise, denoted by dW_t is a stochastic process that evolves by time t and for two different times t_1 and t_2 , dW_{t_1} and dW_{t_2} are independent. To many's disappointment, it turns out such a stochastic process does not exist in a conventional sense and we have to introduce something called Brownian motion, a.k.a. Wiener process that describes the noise. A Wiener process is defined by certain properties:

Definition 2.1.1. A stochastic process $\{X_t\}_t$ is a set of random variables indexed by time $t \in [0, \infty)$. It represents evolution of an uncertain quantity over time. At each time t , X_t is a random variable with a certain distribution that depends on t . Furthermore, the joint distribution of X_{t_1}, \dots, X_{t_n} for different times t_1, \dots, t_n is important in shaping a stochastic process.

An observation, denoted by ω , is one realization of a stochastic process. The function $X(\omega) : [0, \infty) \rightarrow \mathbb{R}^d$ is called a sample path of X .

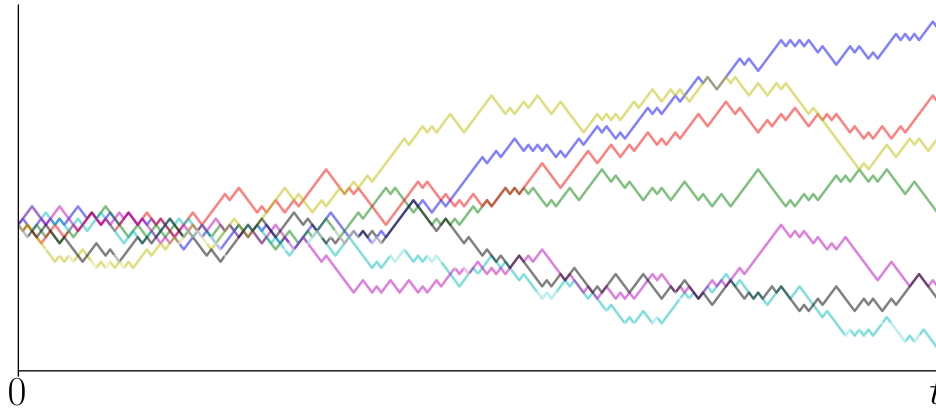


Figure 2.1.1: Eight different sample paths of a stochastic process. The randomness causes each sample paths to be different.

Definition 2.1.2. A Wiener process or Brownian motion is a stochastic process with the following property.

- 1) W_0 is distributed according to a known distribution, μ . μ can be a Dirac delta at a point, $W_0 = 0$.
- 2) For any $t_1 < \dots < t_n$, $W_{t_n} - W_{t_{n-1}}, \dots, W_{t_2} - W_{t_1}$ are independent and have Gaussian distribution with mean zero and variance $t_n - t_{n-1}, \dots, t_2 - t_1$.
- 3) Sample paths of W are continuous functions of t .

For practical purposes, we do not need the above definition. A Wiener process can be understood through the lens of random walk. Let's in dimension d , consider a sequence of independent random variables which take $2d$ values $S := \{\pm e_1, \dots, \pm e_d\}$, where $\{e_i = (0, \dots, 0, \underbrace{1}_{i^{\text{th}}}, 0, \dots, 0) : i = 1, \dots, d\}$ with equal

probabilities is the standard basis with for \mathbb{R}^d . In dimension 1, there are two directions. In dimension $d = 2$, there are four directions. The set S represents the steps of the walk. At each time, the random walk randomly chooses a step from set S and moves accordingly. A sample paths of a 2- d random walk is shown in Figure 2.1.2.

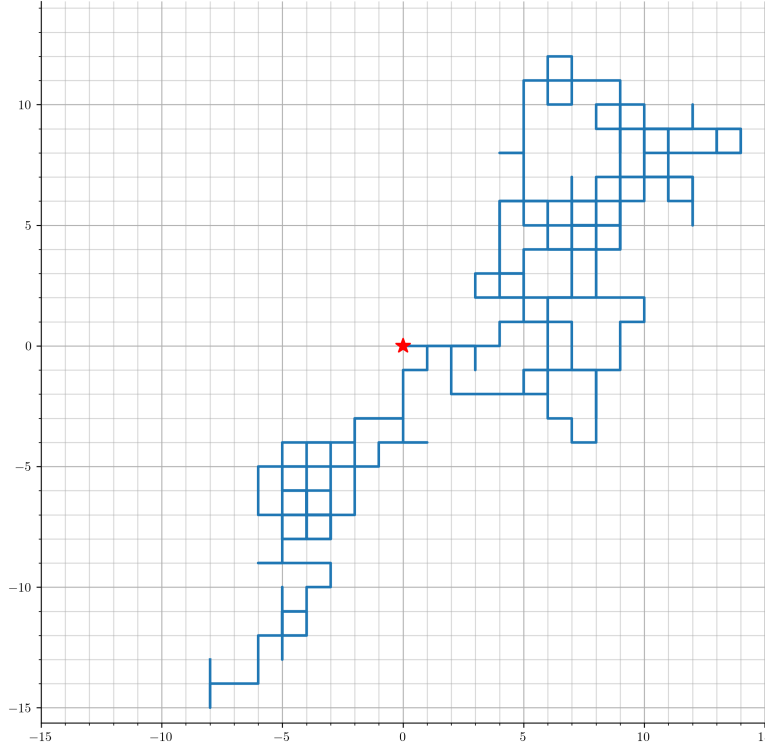


Figure 2.1.2: A sample paths of a 2-d random walk starting at $(0, 0)$ with 400 steps.

To obtain a Wiener process from a random walk, we modify the size of the steps and the timing between two steps. For $N \in \mathbb{N}$, we modify the step size to $\sqrt{1/N}$ and modify the time between two steps by $1/N$. As $N \rightarrow \infty$, we obtain a process that possesses the properties described in Definition 2.1.2. At a point in the random walk, the slope of the path is $\frac{\sqrt{1/N}}{1/N} = \sqrt{1/N} \rightarrow \infty$. Therefore, it is expected that the derivative of the Wiener process, dW_t , does not exist. This derivative is what we heuristically know as the white noise.

2.2 Stochastic differential equation (SDE)

To include noise in $dX_t = b(t, X_t, u_t)dt + \text{noise}$, we can still write $dX_t = f(t, X_t, u_t)dt + dW_t$. However, we should interpret the equation as in integral equation:

$$X_t = X_0 + \int_0^t b(s, X_s, u_s)ds + W_t$$

The magnitude of noise W_t is governed by a multiplicative constant, σW_t . In this case, the variance of the noise is $\sigma^2 t$.

$$X_t = X_0 + \int_0^t b(s, X_s, u_s)ds + \sigma W_t$$

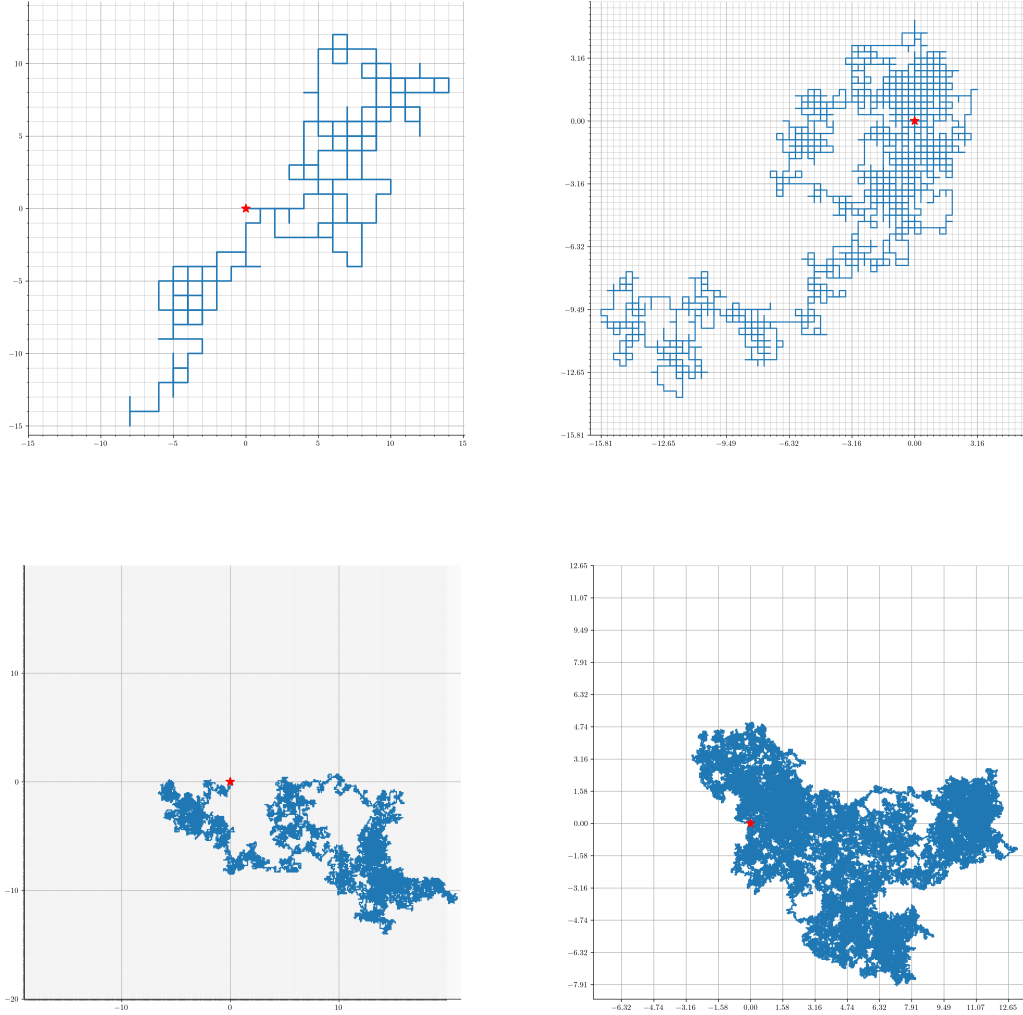


Figure 2.1.3: Top left: a random walk with step size 1 and pacing 1. Top right: a random walk with step size $\sqrt{0.1}$ and pacing 0.1. Bottom left: a random walk with step size 0.1 and pacing 0.01. Bottom right: a random walk with step size $\sqrt{0.001}$ and pacing 0.001. As $N \rightarrow \infty$, a random walk with step size $\sqrt{1/N}$ and pacing $1/N$ converges to a Wiener process given in Definition 2.1.2. Specifically, the continuous sample paths are preserved under the limit.

While σ can be a constant, we can think about a generalization where σ a function of time t and state X_t , i.e. $\sigma(t, X_t)dW_t$. This means for different (t, x) the magnitude of noise evolves in time and bases on the current value of state variable. For example, $X_t dW_t$ means when X_t is close to zero, the noise is small. If X_t is large, the noise is significant.

Since we are in the control context, we can talk about controlling the noise magnitude, $\sigma(t, X_t, u_t)dW_t$. For instance, for $\sigma(t, x, u) = ux$, $u_t X_t dW_t$, when we set $u = 0$, the noise vanishes.

To summarize, the general noise term added to an ODE is written by

$$dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dW_t$$

or in integral form

$$X_t = X_0 + \int_0^t b(s, X_s, u_s)ds + \int_0^t \sigma(s, X_s, u_s)dW_s$$

This is what we refer to by stochastic differential equation or SDE.

The integral $\int_0^t \sigma(s, X_s, u_s)dW_s$ is called Itô integral and is defines by

$$\int_0^t \sigma(s, X_s, u_s)dW_s = \lim_{|t_n - t_{n-1}| \rightarrow 0} \sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \quad (2.2.1)$$

We can call the above summation an Itô sum, which is different from a Riemann sum. The time chosen in the interval $[t_{n-1}, t_n]$ in the Itô sum is always the left point of the interval t_{n-1} . Any other choice of point leads to an integral different from Itô's. This is important because in the Itô sum (2.2.1) $\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})$ and $W_{t_n} - W_{t_{n-1}}$ are independent. Recall property (2) from the definition of Wiener process; Definition 2.1.2.

2.2.1 What do we need to know about Wiener process and Itô integral?

For starters, we need to know that $W_{t_n} - W_{t_{n-1}}$ has mean zero. We also need to recall that by property (2) in Definition 2.1.2, $W_{t_n} - W_{t_{n-1}}$ is independent of the past including any term that has index t_{n-1} or smaller index. As a result of this,

$$\mathbb{E} [\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}})] = \mathbb{E} [\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})] \underbrace{\mathbb{E} [W_{t_n} - W_{t_{n-1}}]}_{=0} = 0$$

and therefore,

$$\mathbb{E} \left[\int_0^t \sigma(s, X_s, u_s)dW_s \right] = 0$$

On the other hand, we need to know

$$\mathbb{E} [(W_{t_n} - W_{t_{n-1}})^2] = t_n - t_{n-1}$$

We also need to remember that $\mathbb{E} [(W_{t_n} - W_{t_{n-1}})^2] = t_n - t_{n-1}$. Therefore,

$$\begin{aligned} & \mathbb{E} \left[(\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}))^2 \right] \\ &= \mathbb{E} [\sigma^2(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})] \underbrace{\mathbb{E} [(W_{t_n} - W_{t_{n-1}})^2]}_{=t_n - t_{n-1}} \\ &= \mathbb{E} [(\sigma^2(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}))(t_n - t_{n-1})] \end{aligned}$$

Further calculations:

$$\begin{aligned}
 & \left(\sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \right)^2 = \sum_{n=1}^N \sigma^2(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}})^2 \\
 & \quad + 2 \sum_{1 \leq n < m \leq N} \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \sigma(t_{m-1}, X_{t_{m-1}}, u_{t_{m-1}})(W_{t_m} - W_{t_{m-1}}) \\
 & \mathbb{E} \left[\sum_{n=1}^N \sigma^2(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}})^2 \right] \\
 & \quad = \sum_{n=1}^N \mathbb{E} [\sigma^2(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}})^2] \\
 & \quad = \sum_{n=1}^N \mathbb{E} [\sigma^2(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(t_n - t_{n-1})] \\
 & \quad \rightarrow \mathbb{E} \left[\int_0^t \sigma^2(s, X_s, u_s) ds \right]
 \end{aligned}$$

On the other hand, since $m > n$, $W_{t_m} - W_{t_{m-1}}$ is independent of

$$\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \sigma(t_{m-1}, X_{t_{m-1}}, u_{t_{m-1}})$$

and

$$\begin{aligned}
 & \mathbb{E} \left[\sum_{1 \leq n < m \leq N} \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \sigma(t_{m-1}, X_{t_{m-1}}, u_{t_{m-1}})(W_{t_m} - W_{t_{m-1}}) \right] \\
 & = \sum_{1 \leq n < m \leq N} \mathbb{E} [\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \sigma(t_{m-1}, X_{t_{m-1}}, u_{t_{m-1}})] \underbrace{\mathbb{E} [W_{t_m} - W_{t_{m-1}}]}_{=0} = 0
 \end{aligned}$$

Therefore,

$$\mathbb{E} \left[\left(\int_0^t \sigma(s, X_s, u_s) dW_s \right)^2 \right] = \mathbb{E} \left[\int_0^t \sigma^2(s, X_s, u_s) ds \right]$$

In other words;

<p>The noise term $\int_0^t \sigma(s, X_s, u_s) dW_s$ has mean zero and variance (magnitude)</p> $\mathbb{E} \left[\int_0^t \sigma^2(s, X_s, u_s) ds \right]$ <p>provided that the above expectation is finite.</p>

2.2.2 Itô's formula

We heuristically deduced that $W_{t_n} - W_{t_{n-1}} \sim O(\sqrt{t_n - t_{n-1}})$. This has implications in the use of Taylor series similar to what we did in derivation of HJ equation from DPP in Section 6. To clarify, Consider a

function f at its Taylor series about a point x .

$$f(y) = f(x) + f'(x)(y - x) + \frac{1}{2}f''(x)(y - x)^2 + \dots$$

If we replace $y = x + \int_t^s b(r, X_r, u_r)dr$, we obtain

$$\begin{aligned} f(y) &= f(x) + f'(x) \int_t^s b(r, X_r, u_r)dr + \frac{1}{2}f''(x) \left(\int_t^s b(r, X_r, u_r)dr \right)^2 + \dots \\ &\approx f(x) + f'(x)b(t, X_t, u_t)(s - t) + \frac{1}{2}f''(x)b^2(t, X_t, u_t)(s - t)^2 + \dots \\ &\approx f(x) + f'(x)b(t, X_t, u_t)(s - t) + O((\Delta t)^2) \end{aligned}$$

This is the simple chain rule:

$$f(X_s) = f(X_t) + \int_t^s f'(X_r)b(r, X_r, u_r)dr$$

When $y = x + \int_t^s \sigma(r, X_r, u_r)dW_r$,

$$\begin{aligned} f(y) &= f(x) + f'(x) \int_t^s \sigma(r, X_r, u_r)dW_r + \frac{1}{2}f''(x) \left(\int_t^s \sigma(r, X_r, u_r)dW_r \right)^2 + \dots \\ &\approx f(x) + f'(x)\sigma(t, X_t, u_t)(W_s - W_t) + \frac{1}{2}f''(x)\sigma^2(t, X_t, u_t)(W_s - W_t)^2 + \dots \\ &\approx f(x) + f'(x)\sigma(t, X_t, u_t)(W_s - W_t) + \frac{1}{2}f''(x)\sigma^2(t, X_t, u_t)(s - t) + o(\Delta t) \end{aligned}$$

Now with the stochastic term, this is a more complicated chain rule:

$$f(X_s) = f(X_t) + \int_t^s f'(X_r)\sigma(r, X_r, u_r)dW_r + \frac{1}{2} \int_t^s f''(X_r)\sigma^2(r, X_r, u_r)dr$$

Having a complete form for stochastic differential equation $y = x + \int_t^s b(r, X_r, u_r)dr + \int_t^s \sigma(r, X_r, u_r)dW_r$, we obtain the Itô's formula;

$$f(X_s) = f(X_t) + \int_t^s f'(X_r)\sigma(r, X_r, u_r)dW_r + \int_t^s \left(f'(X_r)b(r, X_r, u_r) + \frac{1}{2}f''(X_r)\sigma^2(r, X_r, u_r) \right) dr$$

When function f depends on (t, x) , Itô formula is given by

$$\begin{aligned} f(s, X_s) &= f(t, X_t) + \int_t^s f_x(r, X_r)\sigma(r, X_r, u_r)dW_r \\ &\quad + \int_t^s \left(f_t(r, X_r) + f_x(r, X_r)b(r, X_r, u_r) + \frac{1}{2}f_{xx}(r, X_r)\sigma^2(r, X_r, u_r) \right) dr \end{aligned}$$

The above Itô's formula is for 1-dimensional state variable x . For higher dimensions, f_x is the gradient vector and f_{xx} is the Hessian matrix. Then, the Itô formula is written by using the dot product of vectors

and Frobenius product of matrices.

$$f(s, X_s) = f(t, X_t) + \int_t^s f_x(r, X_r) \cdot \sigma(r, X_r, u_r) dW_r + \int_t^s \left(f_t(r, X_r) + f_x(r, X_r) \cdot b(r, X_r, u_r) + \frac{1}{2} f_{xx}(r, X_r) \cdot \sigma^\top \sigma(r, X_r, u_r) \right) dr$$

In the above, X is a d -dimensional stochastic process, b is a d -dimensional vector field, σ is a $d \times d$ matrix field, and X is given by

$$dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dW_t$$

where W is a d -dimensional Brownian motion. In the Itô's formula, $f_x(r, X_r) \cdot b(r, X_r, u_r)$ is the dot product of gradient and vector field b and $f_{xx} \cdot \sigma^\top \sigma$ is the Frobenius product of Hessian and matrix $\sigma^\top \sigma$ defined by

$$A \cdot B = \text{Tr}[AB^\top]$$

Frobenius product is simply dot product if you write matrices as d^2 vectors. If $A = [A_{ij}]_{i,j=1,\dots,d}$ and $B = [B_{ij}]_{i,j=1,\dots,d}$, then $A \cdot B = \sum_{i=1}^d \sum_{j=1}^d A_{i,j} B_{i,j}$, the sum of product of elements with the same indices.

2.2.3 Controlled SDE

The state process is a controlled SDE given by

$$\begin{cases} dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dB_t \\ X_0 = x \in \mathbb{R}^d \end{cases} \quad (2.2.2)$$

The objective function is similar to deterministic case; there is a running cost and a terminal cost. However, it takes the expected value of the randomness.

$$J(x; u) = \mathbb{E} \left[\int_0^T C(s, X_s^u, u_s) ds + g(X_T^u) \right]. \quad (2.2.3)$$

We also assume there is a set of admissible controls, \mathcal{U} . The goal is

$$\inf_{u \in \mathcal{U}} J(x; u)$$

Since the problem setting is stochastic, control variables are also stochastic processes. We can choose the control process to behaves randomly following the randomness in the state variable and/or the Wiener process.

Similar to deterministic case, we have myopic solutions.

Example 9. Let α_t be a given stochastic process, e.g. a Wiener process. Consider the stochastic control problem

$$\inf_{u \in \mathcal{U}} \mathbb{E} \left[\int_0^T (x_t^2 - \alpha_t x_t) dt \right] \quad (2.2.4)$$

with controlled state process is given by

$$dX_t = u_t dt + \sigma dW_t, \quad x_0 = x \quad (2.2.5)$$

If we can choose u such that $X_t = \alpha_t/2$ or $x + \int_0^t u_s ds + \sigma W_t = \alpha_t/2$. Under this condition, we have a myopic solution. However, this is very restrictive, even more restrictive to the deterministic case. For example, let assume α_t is deterministic and differentiable. Then, for $x + \int_0^t u_s ds + \sigma W_t = \alpha_t/2$ to hold, $\int_0^t u_s ds$ and σW_t should cancel each other. But, Brownian motion is never differentiable. This is a twist added by the stochastic term.

2.2.4 Admissibility

It is easy to model stochastic control problems from real applications. However, if not carefully done, the stochastic control problem becomes degenerate.

Exercise 24 (St. Petersburg game and doubling strategy). *In this example, we show that if the collection of admissible controls are not determined properly, the control problem becomes trivially useless. Consider a game with consecutive rounds. At each round the player makes a bet and a coin is flipped. Upon receiving favorable outcome, the player gains as much as his bet. Otherwise, he loses as much as his bet. The doubling strategy suggests to double the bet for the next round if you lose this round. As a result, if the player's bet in the first round is x and the first win happens at round n , he has lost $x(1 + \dots + 2^{n-1}) = x(2^n - 1)$. In the n th round, he wins $x2^n$, which makes his total wealth x more. This is an obvious arbitrage strategy. However, it requires the play to tolerate large amount of debt, i.e., $x(2^n - 1)$. Note that even if the coin is largely biased against the player, the doubling strategy works.*

2.2.5 Value function

For a stochastic control problem

$$\inf_{u \in \mathcal{U}} J(u), \quad J(u) := \mathbb{E} \left[\int_0^T C(t, X_t, u_t) dt + g(X_T) \right]. \quad (2.2.6)$$

where

$$dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dB_t$$

We define the value function as below

Definition 2.2.1 (Value function). *The value function of the control problem is defined by*

$$V(t, \omega) = \inf_{u \in \mathcal{U}_{t,T}} \mathbb{E} \left[\int_t^T C(s, X_s^u, u_s) ds + g(X_T^u) \middle| \mathcal{F}_t \right] \quad (2.2.7)$$

$\mathcal{U}_{t,T}$ is the set of admissible controls on $[t, T]$ and the condition, \mathcal{F}_t , inside the conditional expectation represents the path of X before or at time t .

There is a well-celebrated result, independently published by [Haussmann \[1986\]](#) and [El Karoui et al. \[1987\]](#), that shows we can restrict the set of controls to a simpler one.

Theorem 2.2.1. *For any well-defined (proper admissibility condition and finite value function) stochastic control problem, the value function does not change if we reduce the set of controls to **Markovian controls**, i.e., $u_t := \phi(t, X_t)$ for some function $\phi(t, X_t)$.*

Markovian controls are also called feedback controls, In deterministic control, the existence of optimal feedback control controls is a straightforward fact.

Corollary 2.2.1. *The value function takes the form $V(t, \omega) = V(t, X_t)$. Additionally,*

$$V(t, x) = \inf_{\phi(t, x) \in \mathcal{U}_{t, T}^m} \mathbb{E} \left[\int_t^T C(s, X_s^u, u_s) ds + g(X_T^u) \middle| X_t = x \right] \quad (2.2.8)$$

where $\mathcal{U}_{t, T}^m$ is the set of all admissible Markovian controls $u_t = \phi(t, X_t)$.

In other words, from the path of X until time t , \mathcal{F}_t , all that matters is the end point X_t and all the previous values that X_s took for $s < t$ are irrelevant.

It is worth mentioning there are three type of controls:

1. Open-loop controls which do not care about the current state of the system. In our case, an open-loop control is a deterministic control. In Example 1.1.3, if α_t is differentiable, then $u_t = \frac{\alpha'_t}{2}$ is a closed-loop control.
2. A closed-loop or feedback control is a control that depends on the path of the state process, i.e., $u_t = \phi(t, \mathcal{F}_t)$.
3. A Markovian control is a special case of feedback control that only depends on the latest value of the state variable, $u_t = \phi(t, X_t)$. Non-Markovian controls are only relevant if the cost function or the terminal cost depend on the path of X or X is not a controlled Markov process.

2.2.6 Dynamic programming principle (DPP) for stochastic control

Dynamic programming principle is not as simple as the deterministic case, Theorem 1.2.1. In deterministic case, we can break a control on $[t, T]$ into one on $[t, s]$ and another on $[s, T]$. Conversely, we can patch two controls on $[t, s]$ and $[s, T]$ to create one from $[t, T]$. In stochastic control, the splitting and patching does not work. Specifically, patching can create non-measurable control processes, which makes it mathematically inaccurate. However, one can use a measurable selection theorem to patch properly and avoid the issue. Other methods used quantization to solve the issue by avoid measurable selection theorems. For more information, see Touzi [2012] and references therein.

Theorem 2.2.2. *If the value function is continuous and have linear growth, then for $t < s < T$, we have*

$$V(t, x) = \inf_{u \in \mathcal{U}_{t, s}} \mathbb{E} \left[\int_t^s C(r, X_r, u_r) dr + V(s, X_s) \right], \quad X_t = x \quad (2.2.9)$$

To obtain an Hamilton-Jacobi-Bellman (HJB) equation from DPP, we require to take $s - t$ very small and expand $V(s, X_s)$ about (t, x) .

$$\begin{aligned} V(s, X_s) &= V(t, x) + V_t(t, x)(s - t) + V_x(t, x)(X_s - x) + \frac{1}{2} V_{xx}(t, x)(X_s - x)^2 + \dots \\ &= V(t, x) + V_t(t, x)(s - t) \\ &\quad + V_x(t, x) \left(\int_t^s b(r, X_r, u_r) dr + \int_t^s \sigma(r, X_r, u_r) dW_r \right) \\ &\quad + \frac{1}{2} V_{xx}(t, x) \left(\int_t^s b(r, X_r, u_r) dr + \int_t^s \sigma(r, X_r, u_r) dW_r \right)^2 + \dots \end{aligned}$$

In the expansion

$$\begin{aligned} \left(\int_t^s b(r, X_r, u_r) dr + \int_t^s \sigma(r, X_r, u_r) dW_r \right)^2 &= \left(\int_t^s b(r, X_r, u_r) dr \right)^2 \\ &\quad + 2 \int_t^s b(r, X_r, u_r) dr \int_t^s \sigma(r, X_r, u_r) dW_r \\ &\quad + \left(\int_t^s \sigma(r, X_r, u_r) dW_r \right)^2 \end{aligned}$$

$(\int_t^s b(r, X_r, u_r) dr)^2 = O((\tau-t)^2)$, $\int_t^s b(r, X_r, u_r) dr \int_t^s \sigma(r, X_r, u_r) dW_r = O((\tau-t)^{3/2})$, and $(\int_t^s \sigma(r, X_r, u_r) dW_r)^2 = O(\tau-t)$. Therefore, the first two terms are high order terms and we can ignore them.

$$\begin{aligned} V(s, X_s) &= V(t, x) + V_t(t, x)(s-t) + V_x(t, x)(X_s - x) + \frac{1}{2} V_{xx}(t, x)(X_s - x)^2 + \dots \\ &= V(t, x) + V_t(t, x)(s-t) \\ &\quad + V_x(t, x) \left(\int_t^s b(r, X_r, u_r) dr + \int_t^s \sigma(r, X_r, u_r) dW_r \right) \\ &\quad + \frac{1}{2} V_{xx}(t, x) \left(\int_t^s \sigma(r, X_r, u_r) dW_r \right)^2 + \dots \end{aligned}$$

In addition,

$$\mathbb{E} \left[\int_t^s \sigma(r, X_r, u_r) dW_r \right] = 0 \quad (2.2.10)$$

$$\mathbb{E} \left[\left(\int_t^s \sigma(r, X_r, u_r) dW_r \right)^2 \right] = \mathbb{E} \left[\int_t^s \sigma^2(r, X_r, u_r) dr \right] \quad (2.2.11)$$

Therefore,

$$\begin{aligned} \mathbb{E}[V(s, X_s)] &= V(t, x) + V_t(t, x)\mathbb{E}[s-t] + V_x(t, x)\mathbb{E} \left[\int_t^s b(r, X_r, u_r) dr \right] \\ &\quad + \frac{1}{2} V_{xx}(t, x)\mathbb{E} \left[\int_t^s \sigma^2(r, X_r, u_r) dr \right] + \dots \end{aligned}$$

Insert the Taylor expansion into DPP:

$$0 = \inf_{u \in \mathcal{U}_{t,s}} \mathbb{E} \left[\int_t^s C(r, X_r, u_r) dr + V_t(t, x)(s-t) + \int_t^s b(r, X_r, u_r) dr + \frac{1}{2} V_{xx}(t, x) \int_t^s \sigma^2(r, X_r, u_r) dr \right]$$

Take $\Delta t := s - t$, divide by Δt , and send $\Delta t \rightarrow 0$, we get the HJB equation:

$$\begin{cases} 0 = V_t(t, x) + \inf_{u \in U} \{ C(t, x, u) + V_x(t, x)b(t, x, u) + \frac{1}{2} V_{xx}(t, x)\sigma^2(t, x, u) \} \\ V(T, x) = g(x) \end{cases}$$

Using the notation of Hamiltonian, we write

$$\begin{cases} 0 = V_t(t, x) + H(t, x, V_x, V_{xx}) \\ V(T, x) = g(x) \end{cases} \quad (2.2.12)$$

$$H(t, x, p, \Gamma) := \inf_{u \in U} \left\{ C(t, x, u) + p \cdot b(t, x, u) + \frac{1}{2} \Gamma \sigma^2(t, x, u) \right\} \quad (2.2.13)$$

where we added terminal cost as a boundary condition.

Exercise 25. The wealth of an individual who splits his investment in the stock market and money market is given by

$$dX_t = (\theta_t \mu + r(X_t - \theta_t))dt + \theta_t \sigma dW_t$$

Here, $\mu \in \mathbb{R}$ is the rate of return on the stock, $r \geq 0$ is the interest rate of money market, and θ_t is the control variable that represents the amount invested in risky asset. For a given utility function, U , write the HJB for

$$\sup_{\theta_t} \mathbb{E}[U(X_T)]$$

To avoid doubling strategy, we define the set of all admissible investment strategies as all θ_t such that $X_t \geq 0$ for all t .

Exercise 26. In Exercise 25, we allow the consumptions to be withdrawn from the investment at any time.

$$dX_t = (\theta_t \mu + r(X_t - \theta_t) - c_t)dt + \theta_t \sigma dW_t$$

Here, $\mu \in \mathbb{R}$ is the rate of return on the stock, $r \geq 0$ is the interest rate of money market, and θ_t is the control variable that represents the amount invested in risky asset. For a given utility function, U , write the HJB for

$$\sup_{\theta_t, c_t \geq 0} \mathbb{E} \left[\int_0^T U(c_t)dt + U(X_T) \right]$$

To avoid doubling strategy, we define the set of all admissible investment as in Exercise 25.

Exercise 27. In Exercises 25 and 26, assume that $U(x) = \frac{x^\gamma}{\gamma}$ for $\gamma \in (0, 1)$. Find a solution of the form $V(t, x) = f(t)U(x)$ for the HJB equation.

If we consider higher dimensions, we should write Hamiltonian as

$$H(t, x, p, \Gamma) := \inf_{u \in U} \left\{ C(t, x, u) + p \cdot b(t, x, u) + \frac{1}{2} \Gamma \cdot (\sigma^\top \sigma)(t, x, u) \right\} \quad (2.2.14)$$

Exercise 28. The dynamics of portion of susceptible individuals and infected individuals in a population follows.

$$\begin{aligned} dS_t &= (\mu(1 - S_t) - \beta_t S_t I_t) dt + G_1(S_t) dW_t^1 - G_2(S_t, I_t) dW_t^2 \\ dI_t &= (\beta_t S_t I_t - (\gamma_t + \mu) I_t) dt + G_2(S_t, I_t) dW_t^2 - G_3(I_t) dW_t^3 \end{aligned}$$

where dW^1 , dW^2 , and dW^3 , are independent Wiener processes, $\beta_t > 0$ and $\gamma_t > 0$ are control variables, and

the diffusion coefficients are defined as:

$$\begin{aligned} G_1(s) &= \sqrt{\mu(1+s)} \\ G_2(s, i) &= \sqrt{\beta s i} \\ G_3(i) &= \sqrt{(\gamma + \mu)i} \end{aligned}$$

Write the HJB equation for the stochastic control problem

$$\inf_{\beta_t, \gamma_t} \mathbb{E} \left[\int_0^T (b - \beta_t)^2 + \gamma_t^2 dt + I_T^2 \right]$$

2.2.7 DPP with stopping time

The DPP in Theorem 2.2.2 is a special case of the general DPP for stochastic control problems. To explain the general DPP, we need the notion of stopping time in a probabilistic setting. When we deal with stochastic processes, the occurrence time of an event is random. For example, let τ be the first time a 2- d Wiener process exits a given circle. For two different sample paths, the exit time is different. A stopping time is a random time that can be determined whether it has happen or not based on the observed events. See Figure 2.2.1.

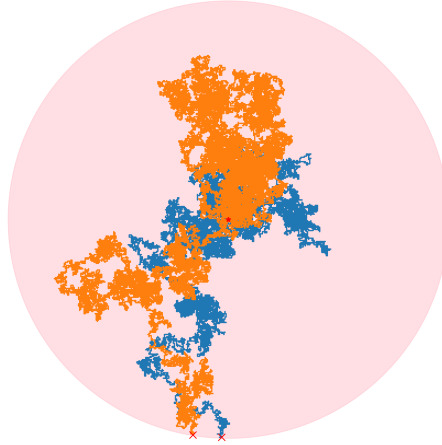


Figure 2.2.1: Two paths of random walks started at the center of the disk. One exits after 42188 steps and the other after 77374 steps.

Definition 2.2.2. Let \mathcal{F}_t represents all events that are revealed at or before time t . The a random variable $\tau : \Omega \rightarrow [0, \infty)$ is called a *stopping time* if for all $t \geq 0$, the event $\{\tau \leq t\}$ belong to \mathcal{F}_t . In simple words, a *stopping time* cannot depend on some secret information that are not observed.

The DPP with stopping time is presented next.

Theorem 2.2.3. *If the value function is continuous and have linear growth, then for any stopping time τ such that $t < \tau < T$, we have*

$$V(t, x) = \inf_{u \in \mathcal{U}_{t, \tau}} \mathbb{E} \left[\int_t^\tau C(r, X_r, u_r) dr + V(\tau, X_\tau) \right], \quad X_t = x \quad (2.2.15)$$

2.2.8 Why is stopping time in DPP important?

In Section 2.2.1, we show that for the Itô sum

$$\int_0^t \sigma(s, X_s, u_s) dW_s = \sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}) (W_{t_n} - W_{t_{n-1}})$$

we have

$$\mathbb{E} \left[\int_0^t \sigma(s, X_s, u_s) dW_s \right] = 0$$

However, to show this is true for a general stochastic integral, defined by the limit of Itô sums, i.e.,

$$\int_0^t \sigma(s, X_s, u_s) dW_s = \lim_{|t_n - t_{n-1}| \rightarrow 0} \sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}) (W_{t_n} - W_{t_{n-1}})$$

we require to be able to commute the limit and expected value:

$$\begin{aligned} & \mathbb{E} \left[\lim_{|t_n - t_{n-1}| \rightarrow 0} \sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}) (W_{t_n} - W_{t_{n-1}}) \right] \\ &= \lim_{|t_n - t_{n-1}| \rightarrow 0} \sum_{n=1}^N \mathbb{E} \left[\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}) (W_{t_n} - W_{t_{n-1}}) \right] = 0 \end{aligned}$$

This commutation of limit and expected value is not a straightforward operation. In fact, there are limit theorems, e.g., Lebesgue dominated convergence, monotone convergence, and Fatou's lemma, that come with restrictive conditions for such commutations. For example, if

$$\mathbb{E} \left[\int_0^t \sigma^2(s, X_s, u_s) ds \right] < \infty \quad (2.2.16)$$

then the commutation is possible and

$$\begin{aligned} & \mathbb{E} \left[\int_0^t \sigma(s, X_s, u_s) dW_s \right] \\ &= \mathbb{E} \left[\lim_{|t_n - t_{n-1}| \rightarrow 0} \sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}) (W_{t_n} - W_{t_{n-1}}) \right] \\ &= \lim_{|t_n - t_{n-1}| \rightarrow 0} \sum_{n=1}^N \mathbb{E} \left[\sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}}) (W_{t_n} - W_{t_{n-1}}) \right] = 0 \end{aligned}$$

Otherwise, one cannot show that the stochastic integral is mean-zero. In such cases, the mean of stochastic integral does not even exist. For example, in the derivation of HJB equation for stochastic control problems, specifically in (2.2.10), there is no guarantee that the expectation of the stochastic integral is zero.

$\sigma(r, X_r, u_r)$ is a general function and u_r is arbitrary. Therefore, we can not guaranty the square integrability in (2.2.16). However, if we write the DPP with a specific stopping time, then the issue is resolved. In (2.2.15), we can choose the stopping time to be

$$\tau = \min \left\{ s, \inf \{ r > t : |\sigma(r, X_r, u_r)| \geq M \} \right\} \quad (2.2.17)$$

for some $M > 0$. Therefore, we are sure that $\sigma(r, X_r, u_r)$ is bounded inside

$$\int_0^\tau \sigma(r, X_r, u_r) dW_r$$

As a result,

$$\mathbb{E} \left[\int_0^\tau \sigma^2(r, X_r, u_r) ds \right] < \infty$$

and the stochastic integral vanishes.

Remark 2.2.1. *The above discussion is not completely rigorous. In (2.2.17), we cannot assume that u_r behaves continuously. The control may simply jump to a high value and invalidate our argument. Usually, in many cases, we either assume $u \in U$ and U is bounded, or we assume $|\sigma(r, X_r, u_r)| \leq C(1 + |X_r|)$ and define*

$$\tau = \min \left\{ s, \inf \{ r > t : |X_r| \geq M \} \right\}$$

Chapter 3

Numerical optimal control

In this chapter, we present some of the most recent numerical methods to solve optimal control problems. The first numerical methods presented here are based on solving HJ or HJB equations. Since HJ or HJB equations are nonlinear PDEs, nonlinear PDE solvers can be potential solutions. However, some of the classical methods of solving nonlinear PDE, such as finite difference, finite elements, finite volume, and the likes, are only practical in low dimensions ($d \leq 3$). Moreover, learning such methods heavily relies on numerical analysis techniques, which is outside the interest of these notes. Therefore, we only present most recent methods that can be applied in higher dimensions.

Before we continue to learn the numerical methods, we shall emphasize on that we are trying to find the optimal control and/or the value function for each possible value of t and x . Therefore, we first study the approximation methods of functions.

3.1 Approximating functions

Let's assume that we want to approximate a function $f : [a, b]^d \rightarrow \mathbb{R}$ using only information at finite number of points, i.e., $f(x_i) = y_i$.

3.1.1 Regression

Simplest of all methods is regression. In regression, we choose a basis set of functions $\{\phi_1, \dots, \phi_K\}$ and find coefficients $a_1, \dots, a_K \in \mathbb{R}$ such that

$$f(x) \approx \hat{f}(x) := \sum_{k=1}^K a_k \cdot \phi_k(x)$$

Let $\{(x^j, y^j) : j = 1, \dots, J\}$ be such that $y^j = f(x^j)$, then we set values for a_1, \dots, a_K such that the mean square error is minimized:

$$\inf_{a_1, \dots, a_K} \sum_{j=1}^J \left(\sum_{k=1}^K a_k \phi_k(x^j) - y^j \right)^2 \quad (3.1.1)$$

Taking derivatives with respect to a_k , we obtain

$$\sum_{j=1}^J \phi_k(x^j) \sum_{\ell=1}^K a_\ell \phi_\ell(x^j) - y^j = 0$$

or

$$\sum_{\ell=1}^K \phi_{k\ell} a_\ell = b_k, \quad \text{for all } k = 1, \dots, K \quad (3.1.2)$$

where

$$\phi_{k\ell} := \sum_{j=1}^J \phi_k(x^j) \phi_\ell(x^j) \quad \text{and} \quad b_k := \sum_{j=1}^J \phi_k(x^j) y^j$$

Solving the linear system of equation (3.1.2) is often time-consuming. If the functions ϕ_k are not chosen properly, the solution may not be unique. However, if we choose an orthogonal system of functions, then the solution to (3.1.2) is simple. Let

$$\int_a^b \phi_k(x) \phi_\ell(x) dx = \begin{cases} 0 & k \neq \ell \\ \neq 0 & k = \ell \end{cases}$$

and x^j s are i.i.d. samples, then

$$\frac{1}{J} \phi_{k\ell} = \frac{1}{J} \sum_{j=1}^J \phi_k(x^j) \phi_\ell(x^j) \approx \int_a^b \phi_k(x) \phi_\ell(x) dx \begin{cases} = 0 & k \neq \ell \\ \int_a^b \phi_k^2(x) dx & k = \ell \end{cases}$$

and (3.1.2) is approximated by

$$a_k = \frac{\int_a^b \phi_k(x) f(x) dx}{\int_a^b \phi_k^2(x) dx} \approx \frac{\sum_{j=1}^J y^j \phi_k(x^j)}{\sum_{j=1}^J \phi_k^2(x^j)}$$

If we work with real data, we may not have the luxury to assume x^j s are i.i.d. In the context of optimal control problems, there is no prior real data and we do not have this issue. On a different point, the regression can be applied to approximate functions with values in \mathbb{R}^d . However, the complexity of orthogonal functions in higher dimensions can increase significantly.

3.1.2 Neural networks

A feed-forward neural network is a function that is constructed as follows. There are L layers. In the first layer, the input $x \in \mathbb{R}^d$ is multiplied by a $d_1 \times d$ matrix and the result is added to a d_1 vector.

$$\tilde{x}_1 = \Lambda_1 x := W_{d_1 \times d} x + b_1$$

Then, before the output of the first layer is fed into the next layer, a nonlinear function σ_1 is applied on x_1 ,

$$x_1 := \Sigma_1(\tilde{x}_1)$$

Similarly, at layer ℓ , the output $x_{\ell-1}$ of the previous layer is transformed by an affine function and inserted into the nonlinear function σ_ℓ :

$$x_\ell = \Sigma_\ell(\Lambda_{d_\ell} x_{\ell-1}), \quad \text{with } \Lambda_{d_\ell} := W_{d_\ell \times d_{\ell-1}} x_{\ell-1} + b_\ell$$

In the final layer of a feedforward neural network, the dimension of the output is determined by the dimension of the output of the function that needs to be approximated. If we like to approximate a function $f : [a, b]^d \rightarrow \mathbb{R}^m$, $d_L = m$. Then matrices $W_{d_{L-1} \times d_\ell}$ are called the weights and vectors b_ℓ are called the

bias of the neural network, or simply we can call them parameters of the neural network. We denote the neural network by

$$NN(x, \theta) := \Sigma_\ell \circ \Lambda_{d_\ell} \circ \cdots \circ \Sigma_1 \circ L_1(x)$$

where the parameters of the neural network are denoted by

$$\theta = (W_{d_1 \times d}, b_1, \dots, W_{d_{\ell-1} \times d_\ell}, b_{d_\ell})$$

What makes neural network approximation efficient is the simplicity of the functions involved. The affine functions are indeed simple. The nonlinear functions Σ_i are called activation functions. Some example of activation functions include

$$\Sigma(x) = (x)^+ \quad \text{ReLU} \quad (3.1.3)$$

$$\Sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{Sigmoid} \quad (3.1.4)$$

$$\Sigma(x) = \tanh(x) \quad \text{Tangent hyperbolic} \quad (3.1.5)$$

The activation functions on multi-dimensional data act elementwise. $\Sigma(x_1, \dots, x_d) = (\Sigma(x_1), \dots, \Sigma(x_d))$. For the goal of approximating a function with a neural network, one needs to minimize

$$\inf_{\theta} \mathcal{J}(\theta), \quad \mathcal{J}(\theta) := \sum_{j=1}^J (NN(x^j, \theta) - y^j)^2 \quad (3.1.6)$$

To evaluate the infimum above, we need to appeal to gradient descent algorithm. The most important step in the gradient descent is to evaluate the derivative of the loss function in (3.1.6). This is where simplicity of the activation functions in the neural network helps. One can evaluate $\mathcal{J}(\theta)$ once and use it via symbolic calculations to evaluate the gradient at each time the parameter is changed during gradient descent. Any

Algorithm 3: Gradient descent algorithm for neural networks

Data: θ_0 is randomly chosen.

Data: The gradient of $\mathcal{J}(\theta)$ is evaluated in a symbolic manner.

Parameter: Learning rate $\{\alpha_n\}_n$, number of iterations N , and tolerance $\epsilon > 0$.

1 **while** $|\nabla \mathcal{J}(\theta_n)| > \epsilon$ & $n \leq N$ **do**

2 $\theta_{n+1} \leftarrow \theta_n - \alpha_n \nabla \mathcal{J}(\theta_n);$

3 $n \leftarrow n + 1$

4 **return** θ_n

deep learning package has implemented automatic differentiation to evaluate the loss function and to build a tree using chain rule. This tree is saved in the memory and is used at any step of the algorithm to evaluate the gradient. This process is also referred to as forward propagation. After the tree is build and the loss function is evaluated, the algorithm goes backward in the tree to evaluate the gradient and to adjust the parameters using the chain rule. This step of the process is called back propagation.

3.1.3 Comparison between polynomial approximation and neural network

While polynomial approximation has been a state-of-art for many decades, neural networks claim a place in the approximation theory due to the capability so handle nonlinear problems. Especially, orthogonal polynomials are efficient if we are using regression method. However, they lack efficiency in more com-

plicated loss functions. The most important gain of the neural networks over polynomial approximation is where the problem is high-dimensional.

3.2 Policy gradient method for optimal control

Policy gradient methods (PGM) are one of the numerical methods to tackle complicated control problems. They are first introduced in the context of reinforcement learning (RL) and later were generalized to control problems, [Powell \[2019\]](#). Some literature, [Nüsken and Richter \[2021\]](#), refer to PGM as iterative diffusion optimization. In this section, we review the method and discuss the implementation details.

3.2.1 Loss function for control problems and functional optimization

The goal of approximation of $f : [a, b]^d \rightarrow \mathbb{R}^m$ is achieved by minimizing a loss function similar to (3.1.1). In the context of optimal control, we like to find a function $\phi(t, x)$ to minimize

$$\mathcal{J}(\phi) := \mathbb{E} \left[\int_0^T C(t, X_t, \phi(t, X_t)) dt + g(X_T) \right]. \quad (3.2.1)$$

where

$$dX_t = b(t, X_t, \phi(t, X_t))dt + \sigma(t, X_t, \phi(t, X_t))dB_t$$

In the above, $J(\phi)$ if the target of the minimization. The choice of the Markovian control $u_t = \phi(t, X_t)$ is allowed because of [El Karoui et al. \[1987\]](#), [Haussmann \[1986\]](#). An important consideration in the choice of ϕ is to obtain consistency with the space of admissible control. A control problem is simply an optimization over a set of functions.

3.2.2 Parametrization of functions

In the loss function (3.2.1), we replace ϕ with a parametrized function, such as a neural network.

$$\mathcal{J}(\phi(\cdot, \theta)) := \mathbb{E} \left[\int_0^T C(t, X_t, \phi(t, X_t; \theta)) dt + g(X_T) \right]. \quad (3.2.2)$$

where

$$dX_t = b(t, X_t, \phi(t, X_t; \theta))dt + \sigma(t, X_t, \phi(t, X_t; \theta))dB_t$$

Doing so, we restricted the set of all controls with the ones that can be represented by $\phi(t, x; \theta)$ for some value of θ . After this restriction, if the optimal control fits into such parametrized function, we don't lose anything. Additionally, we gain a great deal in simplifying the problem from optimization of functions to optimization of parameter θ .

$$\inf_{\theta} \mathcal{J}(\theta)$$

3.2.3 Discretization

For implementation, we cannot work with a continuous time. We surely require to discretize the SDE for state variable and the integral for the running cost. The simplest of such discretizations is Euler-Maruyama discretization given by $\Delta t = \frac{T}{N}$ and

$$\hat{\mathcal{J}}(\phi(\cdot, \theta)) := \mathbb{E} \left[\sum_{i=0}^{N-1} C(t_i, \hat{X}_{t_i}, \phi(t_i, \hat{X}_{t_i}; \theta)) \Delta t + g(\hat{X}_T) \right]. \quad (3.2.3)$$

where

$$\hat{X}_{t_{n+1}} = \hat{X}_{t_n} + b(t_n, \hat{X}_{t_n}, \phi(t_n, \hat{X}_{t_n}; \theta))\Delta t + \sigma(t_n, \hat{X}_{t_n}, \phi(t_n, \hat{X}_{t_n}; \theta))\Delta B_{t_{n+1}}$$

Here, $\Delta B_{t_{n+1}} = B_{t_{n+1}} - B_{t_n} \sim \mathcal{N}(0, \Delta t)$.

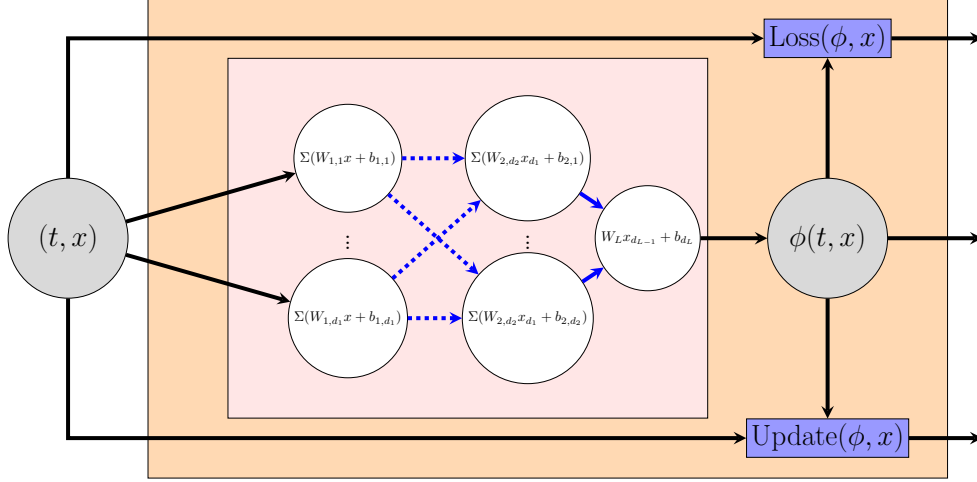


Figure 3.2.1: A design of a unit for solving control problems with deep learning. The neural network approximating control is evaluated in the inner rectangle with affine transformations $Wx + b$ and activation function Σ . The output of the neural network yields the control $\phi(t, x)$. This unit has two other functions, one is updating the state variable x after the control is evaluated and the other evaluates the contribution of the control and state variable in the loss function.

3.2.4 Monte Carlo sampling and empirical loss function

In practice, we cannot evaluate the expected value analytically. Therefore, we appeal to sampling from \hat{X} and empirical approximation of the expected value. To achieve this goal, we require to create sample path of \hat{X} . We start from sampling J i.i.d. samples of \hat{X}_0 via some arbitrary distribution \mathcal{L}_0 , \hat{X}_0^j . Then, we generate i.i.d. samples of the increments of the Wiener process.

$$\Delta B_{t_{n+1}}^j \sim \mathcal{N}(0, \Delta t), \text{ for } j = 1, \dots, J, n = 0, \dots, N-1$$

$$\hat{X}_{t_{n+1}}^j = \hat{X}_{t_n}^j + b(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta))\Delta t + \sigma(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta))\Delta B_{t_{n+1}}^j \quad (3.2.4)$$

Then, the empirical loss is given by

$$\bar{\mathcal{J}}(\phi(\cdot, \theta)) := \frac{1}{J} \sum_{j=1}^J \left[\sum_{i=0}^{N-1} C(t_i, \hat{X}_{t_i}^j, \phi(t_i, \hat{X}_{t_i}^j; \theta))\Delta t + g(\hat{X}_T^j) \right]. \quad (3.2.5)$$

3.2.5 Design of PGM

The key part of the policy gradient method is the evaluation in line 2 of Algorithm 4. In this step, we have an updated parameter θ_n , which yields control $\phi(t, x; \theta_n)$. First, we shall evaluate the sample paths in

Algorithm 5: Gradient descent algorithm for neural networks

Data: θ_0 is randomly chosen.

Data: $\Delta t = \frac{T}{N}$, \hat{X}_0^j are sampled from \mathcal{L}_0 and $\Delta B_{t_{n+1}}^j \sim \mathcal{N}(0, \Delta t)$ are sampled all independently.

Data: Sample paths \hat{X}^j are created.

Data: The gradient of $\bar{\mathcal{J}}(\theta)$ is evaluated in a symbolic manner.

Parameter: Learning rate $\{\alpha_n\}_n$, number of iterations N , and tolerance $\epsilon > 0$.

```

1 while  $|\nabla \mathcal{J}(\theta_n)| > \epsilon$  &  $n \leq N$  do
2   Sample paths  $\hat{X}^j$  are adjusted and  $\bar{\mathcal{J}}(\theta)$  is reevaluated.;
3    $\theta_{n+1} \leftarrow \theta_n - \alpha_n \nabla \mathcal{J}(\theta_n)$ ;
4    $n \leftarrow n + 1$ 
5 return  $\theta_n$ 
    
```

(3.2.4) based on θ_n . More specifically, at each time t_i , we evaluate

$$\hat{X}_{t_{n+1}}^j = \hat{X}_{t_n}^j + b(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta_n))\Delta t + \sigma(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta_n))\Delta B_{t_{n+1}}^j$$

We call this step Update(ϕ, x), as shown in Figure 3.2.1. This update corresponds to $\hat{X}_{t_{i+1}}^j$ is then used in the next time step, unless $t_i = T$. In the same step of the algorithm, we evaluate the contribution of the control $\phi(t, x; \theta_n)$ in the loss function by evaluating

$$C(t_i, \hat{X}_{t_i}^j, \phi(t_i, \hat{X}_{t_i}^j; \theta))\Delta t$$

We call this step Loss(ϕ, x), as shown in Figure 3.2.1. In the last step of the algorithm, when $t_i = T$, the contribution is $g(\hat{X}_T^j)$. This loss is then added to the accumulated loss.

Chapter 4

Old old stuff Stochastic Control

We explain the components of a stochastic control problem.

Exercise 29 (Merton problem). *In the same setting as 24, consider*

$$V(x) = \sup_u \mathbb{E}[U(X_T^u) | X_0^u = x] \quad (4.0.1)$$

where $U : \mathbb{R} \rightarrow \mathbb{R}$ is a differentiable strictly concave strictly increasing function such that $U'(\infty) = 0$, e.g., $U(x) = 1 - e^{-\alpha x}$, $U(x) = \frac{\ln \alpha x}{\alpha}$, or $U(x) = \frac{x^{1-\alpha}}{1-\alpha}$ for $\alpha \in (0, 1)$. Does the same paradox as in St. Petersburg occurs here? Does the set of admissible controls \mathcal{A}_C solves the issue?

Example 10. Consider the stochastic linear quadratic problem below:

$$\inf_u \mathbb{E} \left[\int_0^T (aX_t^2 + bX_t + Au_t^2 + Bu_t) dt + \alpha X_T^2 + \beta X_T \right] \quad (4.0.2)$$

with

$$dX_t = (cX_t + du_t)dt + \sigma dB_t \quad (4.0.3)$$

where where $a, A, \alpha > 0$ and b, B, c , and d are constants. The HJB is given by

$$\begin{cases} 0 = \partial_t V(t, x) + \frac{\sigma^2}{2} \partial_x^2 V(t, x) + cx \partial_x V(t, x) + ax^2 + bx + \inf_u \{ Au^2 + (B + d\partial_x V(t, x))u \} \\ V(T, x) = \alpha x^2 + \beta x \end{cases} \quad (4.0.4)$$

$$u = -\frac{B + d\partial_x V(t, x)}{2A}$$

$$\inf_u \{ Au^2 + (B + d\partial_x V(t, x))u \} = -\frac{(B + d\partial_x V(t, x))^2}{4A^2}$$

$$\begin{cases} 0 = \partial_t V(t, x) + \frac{\sigma^2}{2} \partial_x^2 V(t, x) + cx \partial_x V(t, x) + ax^2 + bx - \frac{(B + d\partial_x V(t, x))^2}{4A^2} \\ V(T, x) = \alpha x^2 + \beta x \end{cases} \quad (4.0.5)$$

We anticipate that $V(t, x) = f(t)x^2 + h(t)x + k(t)$, which is a separation of variable technique. If we plug in $V(t, x)$ into the HJB, we obtain

$$\begin{cases} f' = -2cf(t) + \frac{f^2(t)}{A^2} - a & f(T) = \alpha \\ h'(t) = (\frac{f(t)}{A^2} - c)h(t) + \frac{Bd}{A^2}f(t) - b & h(T) = \beta \\ k'(t) = \frac{PB^2}{4A} + \frac{bd}{2A^2}h(t) + \frac{1}{4A^2}h^2(t) + f(t) & k(T) = 0 \end{cases} \quad (4.0.6)$$

Note that the above system of ODEs is fully solvable. The first ODE is Riccati equation. We shall show in Section 4.2 that the solution of the PDE is indeed the value function of the linear quadratic optimal control problem.

Exercise 30. Solve modified version of Exercise 10 with modification

$$dX_t = (cX_t + du_t)dt + (eX_t + fu_t)dB_t \quad (4.0.7)$$

Exercise 31. Extend the stochastic linear quadratic problem in Exercise 30 to higher dimension and write the HJB.

Example 11 (Merton optimal investment problem). Remember a self-financing portfolio with a Black-Scholes risky asset, $dS_t = S_t(\mu dt + \sigma dB_t)$, under zero interest rate satisfies

$$dX_t = \theta_t(\mu dt + \sigma dB_t) \quad (4.0.8)$$

where θ is the amount of money invested in the risky asset. Merton problem is to maximize the expected utility of wealth at a time horizon T :

$$\sup_{\theta} \mathbb{E}[U(X_T^\theta)] \quad (4.0.9)$$

The HJB for this problem is given by

$$\begin{cases} 0 = \partial_t V + \sup_{\theta} \left\{ \frac{\sigma^2}{2} \partial_x^2 V + \mu \partial_x V \right\} \\ V(T, x) = U(x) \end{cases} \quad (4.0.10)$$

After simplification, we have

$$\begin{cases} 0 = \partial_t V - \frac{\mu^2 (\partial_x V)^2}{2\sigma^2 \partial_x^2 V} \\ V(T, x) = U(x) \end{cases} \quad (4.0.11)$$

We can find a closed form solution for the following cases of utility by using separation of variables, $V(t, x) = f(t)U(x)$:

1. Exponential utility $U(x) = 1 - e^{-\lambda x}$. The separation of variables is $V(t, x) = -f(t)e^{-\lambda x}$.
2. HARA utility $U(x) = \frac{x^{1-\lambda}}{1-\lambda}$, with $\lambda \in (0, 1)$.

Exercise 32 (Merton optimal consumption problem). It is similar to the Merton optimal investment problem except, the investor is consuming from the account and what matters is the utility of consumption.

$$dX_t = \theta_t(\mu dt + \sigma dB_t) - c_t dt \quad (4.0.12)$$

where θ is the amount of money invested in the risky asset and the consumption c satisfies $c_t \geq 0$. Merton problem is

$$\sup_{c \geq 0, \theta} \mathbb{E} \left[\int_0^\infty e^{-\gamma t} U(c_t) dt \right], \quad \gamma > 0 \quad (4.0.13)$$

Write the HJB for this problem.

Remark 4.0.1. $\gamma > 0$ in the above example represents preference of current consumption over future consumption.

4.0.1 Validity of HJB equation

In the previous section, we showed that if the value function is in $C^{1,2}$, then it satisfies (2.2.12). Therefore, HJB can be used to find the value function. If we know that the HJB has a unique $C^{1,2}$ solution in a suitable class of function which includes the value function, we are done. However, it is not always easy to obtain existence and uniqueness results for nonlinear PDEs, which HJB equations are.

Example 12. Recall from Exercise 5 the fastest exit problem:

$$\inf_{|u_t| \leq 1} \int_0^\infty 1_{\{X_t^u \in O\}} dt \quad (4.0.14)$$

with $dX_t^u = u_t dt$. The HJB for this problem is $0 = \inf_{|u| \leq 1} \{u \cdot \nabla V(x)\} + 1_{\{x \in O\}}$, which is equivalent to the following boundary value problem

$$\begin{cases} 0 = -|\nabla V(x)| + 1 & x \in O \\ 0 = V(x) & x \in \partial O \end{cases} \quad (4.0.15)$$

The optimal feedback (closed-loop) exit strategy is give by $u_t^*(x) = -\nabla V(x)$, which is equivalent to gradient descent.

When $O = [-1, 1] \subset \mathbb{R}$, the equation is $|V'(x)| = 1$. On the other hand, this one-dimensional simplification has an obvious answer: run as fast as you can to the nearest exit point. This yields $V(x) = 1 - |x|$. The value function satisfies the HJB except at $x = 0$, which is the sole point with two optimal exit strategies.

Assume that a robot only knows how to solve HJB equations and it ignores finite number of point of irregularities of the solution. For the robot the function $v(x)$ given below is as a good of a solution as $V(x) = 1 - |x|$.

$$v(x) = \begin{cases} 1 - |x| & \frac{1}{2} \leq |x| \leq 1 \\ |x| & |x| \leq \frac{1}{2} \end{cases} \quad (4.0.16)$$

However, if the robot comes of with v instead of V , it suggests to move towards 0 for $|x| \leq \frac{1}{2}$, which is obviously incorrect.

The above example questions the validity of HJB approach. There are two ways to address the validity issue for the solution of HJB equations. One is the notion of viscosity solution, which singles out $V(x) = 1 - |x|$, as the unique viscosity solution of the HJB. For more information of the viscosity solution approach, see Crandall et al. [1992]. There are simple criteria that the robot can check to see whether his approximate solution to the HJB equation converges to the unique viscosity solution. See for example, Barles and Souganidis [1991]. However, here we closed this discussion by giving a glimpse of what it means to be a viscosity solution in the following example.

Example 13. Consider the exit time problem with noise:

$$\inf_{|u_t| \leq 1} \mathbb{E} \left[\int_0^\infty 1_{\{X_t^u \in O\}} dt \right] \quad (4.0.17)$$

with $dX_t^u = u_t dt + \epsilon dB_t$. The HJB is given by $0 = \frac{\epsilon^2}{2} V''(x) + \inf_{|u| \leq 1} \{u \cdot \nabla V(x)\} + 1_{\{x \in O\}}$ ¹, which is

¹My best guess is that the term viscosity comes a fluid mechanics term for the second derivative $\frac{\epsilon^2}{2} V''$.

equivalent to the following boundary value problem

$$\begin{cases} 0 = \frac{\epsilon^2}{2} V''(x) - |\nabla V(x)| + 1 & x \in O \\ 0 = V(x) & x \in \partial O \end{cases} \quad (4.0.18)$$

Back to the simple case where $O = [-1, 1]$, there exists a unique bounded solution, $V^\epsilon(x)$, in C^2 . One can find this solution in closed form and verify that as $\epsilon \rightarrow 0$, $V^\epsilon(x) \rightarrow 1 - |x|$. This clarifies why $v(x)$ is not a solution.

Exercise 33. In the previous exercise, show that as $\epsilon \rightarrow 0$, $V^\epsilon(x) \rightarrow 1 - |x|$.

4.1 Viscosity solutions to the HJB equations

Before moving to the notion of viscosity solutions, we point out an important remark about partial differential equations (PDE) such as the HJBs. The solution to the simplest equations may not exist. For instance, the backward heat equation given by

$$\begin{cases} 0 = \partial_t V + \frac{\sigma^2}{2} \Delta V - rV + f(t, x) \\ V(0, x) = g(x) \end{cases} \quad (4.1.1)$$

does not have any solution unless f and g have sufficient regularity. However, the forward (regular) heat equation

$$\begin{cases} 0 = \partial_t V + \frac{\sigma^2}{2} \Delta V - rV + f(t, x) \\ V(T, x) = g(x) \end{cases} \quad (4.1.2)$$

has a classical solution regardless of the choice of g and f . The forward heat equation is a parabolic equation while the backward equation is not. Roughly speaking, in a parabolic equation, $\partial_t V$ and $\frac{\sigma^2}{2} \Delta V$ show up with the same sign. In the HJB equation (2.2.12), the coefficient of $\partial_t V$ and $\frac{\sigma^2}{2} \Delta V$, 1 and $\frac{\sigma^2}{2}$ respectively, have the same sign. More rigorously, a nonlinear equation of the form

$$\begin{cases} 0 = \partial_t V(t, x) + F(t, x, V(t, x), \nabla V(t, x), D^2 V(t, x)) \\ V(T, x) = g(x) \end{cases} \quad (4.1.3)$$

is parabolic if the nonlinearity, $F(t, x, \varrho, \Pi, \Gamma)$, is nondecreasing in the component Γ , which represents the Hessian matrix in the order induced by nonnegative-definite matrices. I.e., if for two matrices Γ_1 and Γ_2 , $\Gamma_2 - \Gamma_1$ is a nonnegative-definite matrix, then $F(t, x, \varrho, \Pi, \Gamma_2) \geq F(t, x, \varrho, \Pi, \Gamma_1)$. One can check that

$$F(t, x, \varrho, \Pi, \Gamma) := \sup_{u \in \mathbf{U}} \left\{ C(t, x, u) + \mu(t, x, u) \Pi + \frac{1}{2} \sigma^2(t, x, u) \Gamma - r(t, x, u) \varrho \right\} \quad (4.1.4)$$

is nondecreasing in Γ .

Example 14. Show that the HJB equation (4.2.16) is a parabolic equation.

On the other hand, even an equation as simple as the heat equation may have multiple solution if we do not restrict the solution to a proper class of functions. For instance, if $g(x)$ and $f(t, x)$ have linear growth in x , then the heat equation has a unique solution in the class of function with linear growth, but not in the class of functions with super exponential growth.

A linear partial differential equation such as the heat equation has classical solution. Nonlinear equations, however, do not necessarily have a classical solution. A classical solution is a solution that is continuously

differentiable as much as the equation requires. For instance, if in the heat equation above, $f(t, x)$ is a continuous function, then a solution $V(t, x)$ is a classical solution if $V \in C^{1,2}$. But, as seen in Exercise 5, $V(x) = 1 - |x|$ is the value function that we expect to satisfy the HJB equation. However, it is not differentiable as $x = 0$.

A suitable notion of weak solution for the HJB equations, which are nonlinear equations, is viscosity solutions. To understand the viscosity solutions, we first need to provide candidates for the weak derivatives of a function at the points of nondifferentiability. The first derivative can always be represented by the slope of a tangent plane. Locally speaking, the tangent plane is a plane that hits the graph of the function at at least one point locally and keeps the graph on one side. For example, for the function $V(x) = 1 - |x|$ at point $x = 0$, the line $y = 1 + mx$ is a tangent line by $x = 0$ for $m \in [-1, 1]$ (red line) but $y = 1 + mx$ for $m \notin [-1, 1]$ is not (green line). See Figure 4.1.1. Therefore, we have

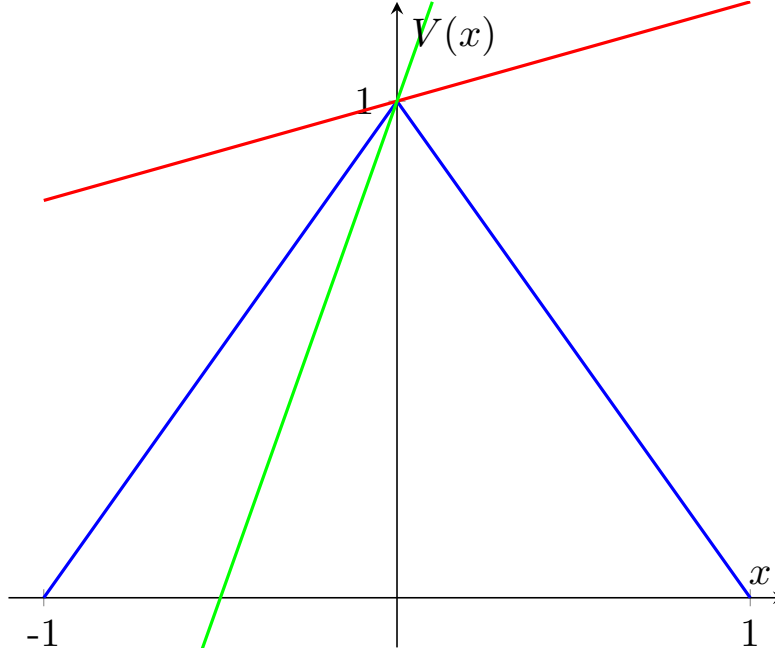


Figure 4.1.1: Red line is an acceptable tangent to the graph, while the green line is not. The green line does not keep the graph on one side locally.

$\varphi(t, x) = a \cdot (x - x_0) + b(t - t_0) + V(t_0, x_0)$ such that $\varphi(t, x) - f(t, x)$ has a local extrema at (t_0, x_0) .

For the second derivative tangent planes are not sufficient, because the second derivatives of the tangent planes are always zero. Therefore, we need to appeal to the quadratic functions of the form $\varphi(t, x) = (x - x_0) \cdot A(x - x_0) + a \cdot (x - x_0) + b(t - t_0) + V(t_0, x_0)$. Note that since we have first derivative on t , we only use first order term on t . If a function $\varphi(t, x)$ touches $V(t, x)$ at point (t_0, x_0) from above (resp. below), i.e., (t_0, x_0) is a local minimum (maximum) point for $\varphi - V$, then we call $(b, a, A) = (\partial_t \varphi(t_0, x_0), \nabla \varphi(t_0, x_0), D^2 \varphi(t_0, x_0))$ a superderivative (resp. sub) of V at point (t_0, x_0) . Such functions $\varphi(t, x)$ are called test functions. See Figure 4.1.2.

A function \underline{V} (resp. \bar{V}) is called a viscosity subsolution (resp. super) of (4.1.3) if for any superderivative (resp. sub) (b, a, A) at point (t_0, x_0) , we have

$$\begin{cases} 0 \geq b + F(t_0, x_0, \varphi(t_0, x_0), a, A), & (\text{resp. } 0 \leq) \\ V(T, x) \leq g(x) & (\text{resp. } 0 \geq) \end{cases} \quad (4.1.5)$$

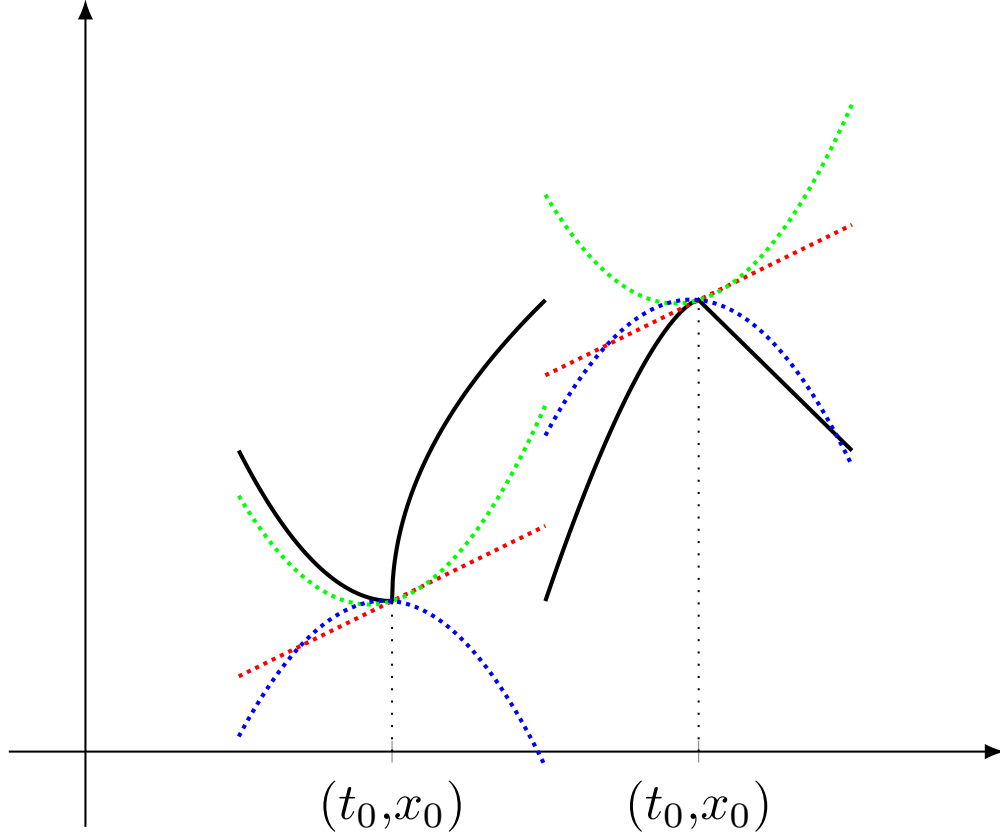


Figure 4.1.2: The test functions that represent the sub and superderivatives of a function at the points of nondifferentiability as well as other points that the function is differentiable.

A function is called a viscosity solution if it is a sub and a super solution.

Remark 4.1.1. The subsolution (resp. super) is named in accordance with submartingale (resp. super). If $U(t, x) \in C^{1,2}$ is a subsolution (resp. super) to the equation

$$0 = \partial_t V(t, x) + C(t, x) + \mu(t, x)\partial_x V(t, x) + \frac{1}{2}\sigma^2(t, x)\partial_{xx} V(t, x), \quad (4.1.6)$$

then $Y_t := U(t, X_t)$ is a submartingale (resp. super) martingale, where

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (4.1.7)$$

For the first order HJB equation, finding the direction of the inequality for the subsolution or super solutions is not particularly obvious. It can be determined while we add a singular perturbation term εB_t to the deterministic state process and send $\varepsilon \rightarrow 0$. For instance, in Exercise 5, if we set the state process $dX_t^u = u_t dt + \varepsilon dB_t$, we obtain the HJB

$$\begin{cases} 0 = 1 - |V'(x)| + \frac{\varepsilon^2}{2} V''(x) \\ V(\pm 1) = 0 \end{cases} \quad (4.1.8)$$

In this case, the subsolution (resp. super) can satisfies

$$\begin{cases} 0 \geq 1 - |V'(x)| + \frac{\varepsilon^2}{2} V''(x) & (\text{resp. } 0 \leq) \\ V(\pm 1) \leq 0 & (\text{resp. } 0 \geq) \end{cases} \quad (4.1.9)$$

By sending $\varepsilon \rightarrow 0$, we obtain the following inequalities for the eikonal equation, (4.1.8).

$$\begin{cases} 0 \geq 1 - |V'(x)| & (\text{resp. } 0 \leq) \\ V(\pm 1) \leq 0 & (\text{resp. } 0 \geq) \end{cases} \quad (4.1.10)$$

We can easily check that the function $V(t, x) = 1 - |x|$ is a viscosity solution of the eikonal equation, (4.1.8). For example, at all points $x \neq 0$, the sub and superderivatives are equal to $-\text{sgn}(x)$, which obviously satisfy the equation with equality. At $x = 0$, we only have super derivatives and the set of subderivatives is empty set. This makes the supersolution property to hold obviously by the false antecedent. The set of superderivatives contains all $m \in [-1, 1]$. Since $0 \leq 1 - |m|$, then $V(t, x) = 1 - |x|$ is also a subsolution. All the above arguments are valid since the boundary condition $V(t, \pm 1) = 0$ validates the sub and supersolution properties.

On the other hand, the function $\tilde{V} = \min\{|x|, 1 - |x|\}$ only satisfies the subsolution property.

Exercise 34. Show that \tilde{V} is a viscosity subsolution to the eikonal equation (4.1.10), but not a super solution. Explore the points $x = \pm \frac{1}{2}$.

The existence and uniqueness of the solution to the nonlinear parabolic equations in the class of functions with linear growth is studies in several papers. For instance, see Crandall et al. [1992]. For the HJB equations derived from the optimal control problems, the value function of the control problem is usually the viscosity solution of the HJB. The uniqueness is due to a technical lemma, Ishii's lemma. However, in most cases, if we manage to show that any viscosity solution is indeed a classical solution, $C^{1,2}$, then the verification theorem, Theorem 4.2.1 shows that any classical solution is equal to the value function, and therefore, uniqueness is obtained.

4.2 Verification

Instead of going through viscosity solutions, we consider the cases where the HJB equation has a $C^{1,2}$ solution, which is a candidate for the value function. Then, we introduce a verification theorem, which approves that the solution is indeed the value function.

Theorem 4.2.1 (Verification). *Let HJB equation, (2.2.12) has a $C^{1,2}$ solution, $v(t, x)$. Then, $v(t, x) \geq V(t, x)$. In addition, assume that there exists $u^*(t, x)$ such that*

$$u^*(t, x) \in \operatorname{argmin}_u \left\{ C(t, x, u) + \frac{1}{2} D^2 v(t, x) \cdot a(t, x, u) + \nabla v(t, x) \cdot \mu(t, x, u) \right\} \quad (4.2.1)$$

and $u^(t, x)$ is an admissible Markovian (feedback) control, i.e., (??) has a (strong) solution, X_t^* , and $u_t^* = u^*(t, X_t^*)$ is admissible, $u^* \in \mathcal{A}$. Then, $v(t, x) = V(t, x)$.*

Proof. Since v satisfies (2.2.12) in classical sense, we conclude that for any value of $u \in U$, we have

$$0 \geq \partial_t v(t, x) + C(t, x, u) + \mu(t, x, u) \nabla v(t, x) + \frac{1}{2} a(t, x, u) \cdot D^2 v(t, x) \quad (4.2.2)$$

Since $v(t, x) \in C^{1,2}$, it follows from the Itô formula that for any $u \in \mathcal{A}_{t,T}$,

$$\begin{aligned} v(T, X_T^{t,x,u}) &= v(t, x) \\ &+ \int_0^T \left(\partial_t v(s, X_s^{t,x,u}) + \mu(s, X_s^{t,x,u}, u_s) \cdot \nabla v(s, X_s^{t,x,u}) + \frac{1}{2} a(s, X_s^{t,x,u}, u_s) \cdot D^2 v(s, X_s^{t,x,u}) \right) ds \\ &+ \int_t^T \sigma(s, X_s^{t,x,u}, u_s) dB_s \end{aligned} \quad (4.2.3)$$

By (4.2.2) and terminal condition $v(T, x) = g(x)$, we obtain

$$g(X_T^u) \geq v(t, x) - \int_0^T C(s, X_s^{t,x,u}, u_s) ds + \int_t^T \sigma(s, X_s^{t,x,u}, u_s) dB_s \quad (4.2.4)$$

If we take conditional expectation, the martingale property of the stochastic integral implies that

$$\mathbb{E}_{t,x} \left[g(X_T^u) + \int_0^T C(s, X_s^{t,x,u}, u_s) ds \right] \geq v(t, x) \quad (4.2.5)$$

Now, by taking supremum over $u \in \mathcal{A}_{t,T}$, we obtain

$$V(t, x) = \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}_{t,x} \left[g(X_T^u) + \int_0^T C(s, X_s^{t,x,u}, u_s) ds \right] \geq v(t, x). \quad (4.2.6)$$

Now, assume that for some $u^*(t, X_t^*) \in \mathcal{A}$, $Y_t^{u^*} = v(t, X_t^{u^*}) + \int_0^t C(s, X_s^{u^*}, u_s^*) ds$ is a martingale. By applying the Itô formula, we have

$$\begin{aligned} Y_t^{u^*} &= v(t, x) + \int_t^T \sigma(s, X_s^{t,x,u^*}, u_s^*) dB_s + \int_0^T \left(\partial_t v(s, X_s^{t,x,u^*}) + \mu(s, X_s^{t,x,u^*}, u_s^*) \cdot \nabla v(s, X_s^{t,x,u^*}) \right. \\ &\quad \left. + \frac{1}{2} a(s, X_s^{t,x,u^*}, u_s^*) \cdot D^2 v(s, X_s^{t,x,u^*}) + C(s, X_s^{u^*}, u_s^*) \right) ds \\ &= v(t, x) + \int_t^T \sigma(s, X_s^{t,x,u^*}, u_s^*) dB_s \end{aligned} \quad (4.2.7)$$

Therefore, the martingale property of Y^{u^*} , implies that

$$\begin{aligned} C(s, X_s^{u^*}, u_s^*) + \partial_t v(s, X_s^{t,x,u^*}) + \mu(s, X_s^{t,x,u^*}, u_s^*) \cdot \nabla v(s, X_s^{t,x,u^*}) \\ + \frac{1}{2} a(s, X_s^{t,x,u^*}, u_s^*) \cdot D^2 v(s, X_s^{t,x,u^*}) = 0, \quad \text{a.s.} \end{aligned} \quad (4.2.8)$$

Therefore, for u^* , all the inequalities in (4.2.4)-(4.2) holds as equality and we obtain $v = V$. \square

We apply Theorem (4.2.1) to the following examples.

Exercise 35 (Merton optimal investment problem). *Remember a self-financing portfolio with a Black-Scholes risky asset, $dS_t = S_t(\mu dt + \sigma dB_t)$, under zero interest rate satisfies*

$$dX_t = \theta_t(\mu dt + \sigma dB_t) \quad (4.2.9)$$

where θ is the amount of money invested in the risky asset. Merton problem is

$$\sup_{\theta} \mathbb{E}[U(X_T^{\theta})] \quad (4.2.10)$$

The HJB is given by

$$\begin{cases} 0 = \partial_t v + \sup_{\theta} \left\{ \frac{\theta^2 \sigma^2}{2} \partial_x^2 v + \theta \mu \partial_x v \right\} \\ v(T, x) = U(x) \end{cases} \quad (4.2.11)$$

This simplifies to

$$\begin{cases} 0 = \partial_t v - \frac{(\mu \partial_x v)^2}{\sigma^2 \partial_x^2 v} \\ v(T, x) = U(x) \end{cases} \quad (4.2.12)$$

For $U(x) = -e^{-\lambda x}$, with $\lambda > 0$, use separation of variables $u(t, x) = -f(t)e^{-\lambda x}$ to find a closed form solution for the HJB.

Exercise 36. For $U(x) = x^{\lambda}$, use separation of variables $u(t, x) = f(t)x^{\lambda}$ to find a closed form solution for the HJB. Is there any restriction on the value of λ ?

Exercise 37. For $U(x) = \ln x$, can you suggest a separation of variables?

Exercise 38 (Merton optimal consumption problem). It is similar to the Merton optimal investment problem except, the investor is consuming from the account and what matters is the utility of consumption.

$$dX_t = \theta_t(\mu dt + \sigma dB_t) - c_t dt \quad (4.2.13)$$

where θ is the amount of money invested in the risky asset and c is the consumption with $c_t \geq 0$. Merton problem is

$$\sup_{\theta, c} \mathbb{E} \left[\int_0^{\infty} e^{-\gamma t} U(c_t) dt \right], \gamma > 0 \quad (4.2.14)$$

Show that the HJB is given by

$$0 = \sup_{\theta} \left\{ \frac{\theta^2 \sigma^2}{2} v'' + \theta \mu v' \right\} + \sup_{c \geq 0} \left\{ U(c) - cv' \right\} - \gamma v \quad (4.2.15)$$

or

$$0 = \frac{(\mu v')^2}{\sigma^2 v''} + \sup_{c \geq 0} \left\{ U(c) - cv' \right\} - \gamma v \quad (4.2.16)$$

For $U(c) = \frac{c^{1-\lambda}}{1-\lambda}$, with $\lambda \in (0, 1)$, we have

$$\sup_{c \geq 0} \left\{ U(c) - cv' \right\} = U(c^*) - c^* v' \quad (4.2.17)$$

with $U'(c^*) = v'$ or c^*

4.2.1 Martingale approach

The martingale principle for optimal control (Davis and Varaiya [1973]) is another verification result which does not require the differentiability of the value function.

Theorem 4.2.2. Assume that there exists a function $v(t, x)$ such that for all $u \in \mathcal{A}$ the process $\{Y_t^u\}_{t \geq 0}$

$$Y_t^u := v(t, X_t^u) + \int_0^t C(s, X_s^u, u_s) ds \quad (4.2.18)$$

is a super martingale and that for some $u^* \in \mathcal{A}$, $\{Y_t^{u^*}\}_{t \geq 0}$ is a martingale. Then, u^* is an optimal control and the value function is equal to $v(t, x)$.

Proof. By the supermartingale property, we have

$$Y_t^u \leq \mathbb{E}[Y_T^u | \mathcal{F}_t] = \mathbb{E}\left[g(X_T^u) + \int_t^T C(s, X_s^u, u_s) ds \middle| \mathcal{F}_t\right] + \int_0^t C(s, X_s^u, u_s) ds \quad (4.2.19)$$

Thus,

$$v(t, x) = Y_t^u - \int_0^t C(s, X_s^u, u_s) ds \leq \mathbb{E}[Y_T^u | \mathcal{F}_t] = \mathbb{E}\left[g(X_T^u) + \int_t^T C(s, X_s^u, u_s) ds \middle| \mathcal{F}_t\right], \quad (4.2.20)$$

for all $u \in \mathcal{A}_{t,T}$. Given $X_t^u = x$, we obtain

$$v(t, x) \leq \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}\left[g(X_T^u) + \int_t^T C(s, X_s^u, u_s) ds \middle| X_t^u = x\right] = V(t, x). \quad (4.2.21)$$

If for some u^* , Y^{u^*} is a martingale, then in all the above, the inequality turns into equality and we have $v(t, x) = V(t, x)$. \square

The above theorem can also be regarded as a verification that a possible candidate $V(t, x)$ is a value function of the optimal control problem. The following example shows the use of this theorem.

Example 15. In Example 24, let $U(c) = \frac{c^{1-\alpha}}{1-\alpha}$ when $c \geq 0$ and $-\infty$ otherwise. Consider the function $V(t, x) = e^{-\gamma(T-t)}(T - t + e^{A/\alpha t})^\alpha e^{-AT} U(x)$ with $A = r(1 - \alpha) - \gamma + \frac{(\mu-r)(1-\alpha)}{2\sigma^2\alpha}$.

$$Y_t^u = V(t, X_t^u) + \int_0^t e^{-\gamma s} U(c_s) ds. \quad (4.2.22)$$

By Itô formula, we have

$$dY_t^u = \left(\partial_t V + (\theta_t(\mu - r) + rX_t^u - c_t) \partial_x V + \frac{1}{2} \sigma^2 \theta_t^2 \partial_{xx} V + e^{-\gamma t} U(c_t) \right) dt + \sigma \theta_t \partial_x V dB_t. \quad (4.2.23)$$

By direct calculation, one can see that

$$\partial_t V + (\theta_t(\mu - r) + rX_t^u - c_t) \partial_x V + \frac{1}{2} \sigma^2 \theta_t^2 \partial_{xx} V + e^{-\gamma t} U(c_t) \leq 0, \quad \mathbb{P}\text{-a.s.} \quad (4.2.24)$$

for all values of θ and c . In addition, for $c_t^*(x) = (\partial_x V(t, x))^{-1/\alpha}$ and $\theta_t^*(x) = -\frac{(\mu-r)\partial_x V(t, x)}{\sigma^2 \partial_{xx} V(t, x)}$,

$$\partial_t V + (\theta_t(\mu - r) + rX_t^u - c_t) \partial_x V + \frac{1}{2} \sigma^2 \theta_t^2 \partial_{xx} V + e^{-\gamma t} U(c_t) = 0, \quad \mathbb{P}\text{-a.s.} \quad (4.2.25)$$

It remains to show that c^* and θ^* are admissible Markov controls, i.e., to show that

$$dX_t^* = \left(\theta_t(X_t^*)((\mu - r)dt + \sigma dB_t) - rX_t^*dt \right) - c_t^*(X_t^*)dt, \quad (4.2.26)$$

has a strong solution. We leave the details as an exercise.

Remark 4.2.1. If the control problem is with infimum instead of supremum,

$$V(t, x) = \inf_{u \in \mathcal{A}_{t,T}} \mathbb{E} \left[\int_t^T C(s, X_s^{t,x,u}, u_s) ds + g(X_T^{t,x,u}) \right], \quad (4.2.27)$$

Theorem 4.2.2 is should be modified. More precisely, Y_t^u is a submartingale.

For deterministic cases, supermartingale (resp. submartingale) means nonincreasing (resp. nondecreasing).

Finding a candidate for a value function and an optimal control is the subject of the future sections.

4.3 Some nonstandard HJB equations

In this section, we provide a review of some HJB equations that come from stochastic singular control, optimal stopping time, stochastic impulse control, and switching problems. The treatment of such problems via HJB equations is similar, however, we first need to derive HJB equations.

4.3.1 Stochastic singular control problems

Consider the stochastic control problem below:

Example 16.

$$\inf_{u_t} \mathbb{E} \left[\int_0^T \left(\frac{1}{2} X_t^2 - u_t \right) dt + X_T^2 - X_T \right] \quad (4.3.1)$$

with $u_t \geq 0$ (takes nonnegative values) and

$$dX_t = (X_t + u_t)dt + \sigma dB_t \quad (4.3.2)$$

The HJB is formally written as

$$\begin{cases} \partial_t V + x \partial_x V + \frac{\sigma^2}{2} \partial_x^2 V + \frac{1}{2} x^2 + \inf_{u \geq 0} u(\partial_x V - 1) = 0 \\ V(T, x) = x^2 - x \end{cases} \quad (4.3.3)$$

Note that

$$\inf_{u \geq 0} u(\partial_x V - 1) = \begin{cases} -\infty & \partial_x V - 1 < 0 \\ 0 & \partial_x V - 1 \geq 0 \end{cases} \quad (4.3.4)$$

Therefore, it is natural to assume that the value function satisfies $\partial_x V - 1 \geq 0$. In fact, we expect so find the value functions such that if $\partial_x V > 1$, then $\partial_t V + x \partial_x V + \frac{\sigma^2}{2} \partial_x^2 V + \frac{1}{2} x^2 = 0$ and when $\partial_x V = 1$, the we only have $\partial_t V + x \partial_x V + \frac{\sigma^2}{2} \partial_x^2 V + \frac{1}{2} x^2 \leq 0$. More precisely, inequality $\partial_x V - 1 \geq 0$ divides $[0, T] \times \mathbb{R}$ into two regions:

- $\mathbf{N} = \{(t, x) : \partial_x V(t, x) > 1\}$
- $\mathbf{R} = \{(t, x) : \partial_x V(t, x) = 1\}$

Inside \mathbf{N} , we expect that $\partial_t V + x\partial_x V + \frac{\sigma^2}{2}\partial_x^2 V + \frac{1}{2}x^2 = 0$ holds. This equations has a solution of the form $v(t, x) = a(t)x^2 + b(t)x + c(t)$.

We denote the boundary of \mathbf{N} by $\partial\mathbf{N}$ and we define $R(t)$ such that $(t, R(t)) \in \partial\mathbf{N}$, assuming that $R(t)$ is uniquely determined. In addition, we expect to see that the value function and its first derivative are the same in the both sides of $\partial\mathbf{N}$, particularly, $\partial_x V(t, R(t)) = 1$. Therefore, if v is the solution to

$$\begin{cases} \partial_t v + x\partial_x v + \frac{\sigma^2}{2}\partial_x^2 v + \frac{1}{2}x^2 = 0 \\ v(T, x) = x^2 - x \end{cases} \quad (4.3.5)$$

then, we anticipate to write

$$V(t, x) = \begin{cases} v(t, x) & x \geq R(t) \\ x - R(t) + v(t, R(t)) & x < R(t) \end{cases} \quad (4.3.6)$$

Solving (4.3.5), we obtain that $v(t, x) = \frac{5e^{2(T-t)}-1}{4}x^2 - e^{T-t}x - \frac{\sigma^2}{4}\left(T - t - \frac{5e^{2(T-t)}-1}{2}\right)$ and, therefore, $R(t) = 2\frac{e^{T-t}+1}{5e^{2(T-t)}-1}$.

Exercise 39.

$$\inf_{u_t \geq 0} \mathbb{E} \left[\int_0^\tau e^{-rt} (\mu - u_t) dt \right] \quad (4.3.7)$$

where $\tau = \inf\{s \geq 0 : X_t = 0\}$ with

$$dX_t = (\gamma X_t - u_t)dt + \sigma dB_t, \quad X_0 = x \geq 0 \quad (4.3.8)$$

Write the HJB and obtain the nonlinear term by evaluating the infimum.

Singular control problems are easy to detect; the running cost and drift of the SDE are both linear in the control. To see this, consider the control problem

$$\inf_u \mathbb{E} \left[\int_0^T (C(s, X_s^u) + au_s) ds + g(X_T^u) \right] \quad (4.3.9)$$

where u_t takes values in a closed convex cone \mathcal{C} and

$$\begin{cases} dX_t^u = (\mu(t, X_t^u) + Au_t)dt + \sigma(t, X_t^u)dB_t \\ X_0^u = x \in \mathbb{R}^d \end{cases} \quad (4.3.10)$$

The heuristic derivation of the HJB yields

$$\begin{cases} 0 = \partial_t V(t, x) + \frac{1}{2}a(t, x, u) \cdot D^2 V(t, x) + \nabla V(t, x) \cdot \mu(t, x) + C(t, x) + \inf_u \{au + Au \cdot \nabla V(t, x)\} \\ V(T, x) = g(x) \end{cases} \quad (4.3.11)$$

$a = \sigma^\top \sigma$. The infimum above is given by

$$\inf_{\hat{u} \in \mathcal{C}, |\hat{u}|=1, \lambda \geq 0} \lambda \hat{u} \cdot (a + A\nabla V(t, x)) = \inf_{\lambda \geq 0} \lambda \inf_{|\hat{u}|=1} \hat{u} \cdot (a + A\nabla V(t, x)) = \inf_{\lambda \geq 0} \lambda \mathcal{H}(\nabla V(t, x)) \quad (4.3.12)$$

where $\mathcal{H}(\nabla V(t, x)) := \inf_{|\hat{u}|=1} \hat{u} \cdot (a + A\nabla V(t, x))$. If $\mathcal{H}(\nabla V(t, x)) < 0$, then the infimum is $-\infty$ and the problem becomes degenerate. To avoid degeneracy, we must have $\mathcal{H}(\nabla V(t, x)) \geq 0$, in which case

infimum is attained in $\lambda = 0$ and we have

$$\begin{cases} 0 = \partial_t V(t, x) + \frac{1}{2}a(t, x, u) \cdot D^2V(t, x) + \nabla V(t, x) \cdot \mu(t, x) + C(t, x) \\ V(T, x) = g(x) \end{cases} \quad (4.3.13)$$

For simplicity of notation, we set

$$\mathcal{L}V(t, x) := \partial_t V(t, x) + \frac{1}{2}a(t, x, u) \cdot D^2V(t, x) + \nabla V(t, x) \cdot \mu(t, x) + C(t, x) \quad (4.3.14)$$

The interpretation of the HJB requires some *variational inequalities*:²

1. for all (t, x) , $-\mathcal{L}V(t, x) \geq 0$ and $\mathcal{H}\nabla V(t, x) \geq 0$.
2. If $\mathcal{H}(\nabla V(t, x)) > 0$, then $-\mathcal{L}V(t, x) = 0$.

The description of the optimal control in singular control problems need the notion of local time of SDEs. For instance, in Exercise 39, the optimal control is the local time of the process, $dX_t = \gamma X_t dt + \sigma dB_t$ at the point \hat{x} , existence of which is obtained through solving the HJB and going through verification step.

4.3.2 Optimal stopping problem

Let X be given by the SDE

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \quad (4.3.15)$$

in a filtered probability space and consider the following problem

$$\sup \mathbb{E} \left[\int_0^{\tau \wedge T} e^{-rt} C(t, X_t) dt + g(X_{\tau \wedge T}) \right] \quad (4.3.16)$$

where the supremum is over all stopping time τ adapted to the filtration. These class of problems, which are not stochastic control problems, called optimal stopping problems. To solve optimal stopping problems, we can write the following DPP:

$$V(t, x) = \mathbb{E}_{t, x} \left[\int_0^\eta e^{-r(s-t)} C(t, X_s) ds + e^{-r(\eta-t)} g(X_\eta) \right] \quad (4.3.17)$$

for any stopping time η with values in $[t, T]$. The DPP above leads to the following variational HJB equation:

1. $V(t, x) \geq g(x)$ and $-\mathcal{L}V(t, x) \geq 0$
2. if $V(t, x) > g(x)$, then $-\mathcal{L}V(t, x) = 0$

where \mathcal{L} is given in (4.3.14).

4.4 Backward stochastic differential equations

The backward stochastic differential equations (BSDE) is an alternative way to tackle the optimal control problems. One advantage of the BSDEs is that they can cover the nonMarkovian optimal control problems.

²For more details on why we write the variational inequality this way, see [Fleming and Soner, 2006, Chapter 11] and the notion of sub and super solutions.

The derivation of the BSDEs from an optimal control problem follows from the Itô martingale representation theorem.

Theorem 4.4.1. *In a probability space which hosts a Brownian motion B , let $\mathbb{F}^B := \{\mathcal{F}_t^B\}_t$ be the right-continuous augmented filtration generated by a Brownian motion B . Let X be a \mathcal{F}_T^B -measurable random variable with finite expectation. Then, there exists a \mathbb{F}^B -progressively measurable process $Z_t(\omega)$ such that*

$$X = \mathbb{E}[X] + \int_0^T Z_t dB_t \quad (4.4.1)$$

In addition, if M is a continuous martingale with respect to \mathbb{F} , then, there exists a \mathbb{F}^B -progressively measurable process $\phi(\omega)$ such that

$$M_t = M_0 + \int_0^t Z_s dB_s, \quad \text{for } t \geq 0. \quad (4.4.2)$$

In the above theorem, Z can interpreted as the sensitivity of the martingale M with respect to the Brownian noise B .

Consider for a given stochastic process $C_t(\omega)$

$$Y_t := \mathbb{E} \left[\int_t^T L_s(\omega) ds + \xi \middle| \mathcal{F}_t^B \right]. \quad (4.4.3)$$

We simply assume that $r_s(\omega)$ and $L_s(\omega)$ are \mathbb{F}^B -progressively measurable processes and ξ is a \mathcal{F}_T^X -measurable and integrable random variable. For the martingale define by

$$M_t := \mathbb{E} \left[\int_0^T L_s(\omega) ds + \xi \middle| \mathcal{F}_t^B \right], \quad (4.4.4)$$

and by the Itô martingale representation theorem, we have

$$M_T = \int_0^T L_s(\omega) ds + \xi = M_t + \int_t^T Z_s dB_s, \quad (4.4.5)$$

for some \mathbb{F}^B -progressively measurable process $Z_t(\omega)$. Since $Y_T = \xi$, we have

$$M_T = \int_0^T L_s(\omega) ds + Y_T = M_t + \int_t^T Z_s dB_s, \quad (4.4.6)$$

On the other hand, since $\int_0^t L_s(\omega) ds$ is \mathcal{F}_t^B -measurable, we can write

$$M_t = \int_0^t L_s(\omega) ds + \mathbb{E} \left[\int_t^T L_s(\omega) ds + \xi \middle| \mathcal{F}_t^B \right] = \int_0^t L_s(\omega) ds + Y_t. \quad (4.4.7)$$

Therefore,

$$\int_0^T L_s(\omega) ds + Y_T = \int_0^t L_s(\omega) ds + Y_t + \int_t^T Z_s dB_s, \quad (4.4.8)$$

or,

$$Y_t = Y_T + \int_t^T L_s(\omega) ds - \int_t^T Z_s dB_s. \quad (4.4.9)$$

The BSDE (4.4.9) can be written formally by

$$\begin{cases} dY_t = -L(t, \omega)dt + Z_t dB_t \\ Y_T = g(X_T) \end{cases} \quad (4.4.10)$$

Note that we could have written the above forwardly, i.e., $Y_t = Y_0 + \int_0^t L_s(\omega)ds - \int_0^t Z_s dB_s$. However, this is not very useful, because Y_0 is not known. See example below.

Example 17. Recall that the solution to the linear equation

$$\begin{cases} 0 = \partial_t v(t, x) + C(t, x) + [\mu \cdot \nabla v](t, x) + \frac{1}{2}[\sigma^\top \sigma \cdot D^2 v](t, x) \\ v(T, x) = g(x) \end{cases} \quad (4.4.11)$$

is given by the Feynmann-Kac formula:

$$V(t, x) = \mathbb{E} \left[\int_t^T C(s, X_s)ds + g(X_T) \middle| X_t = x \right], \quad (4.4.12)$$

where

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (4.4.13)$$

If we define $Y_t = V(t, X_t)$, then (4.4.9) is given by

$$Y_t = g(X_T) + \int_t^T C(s, X_s)ds - \int_t^T Z_s dB_s. \quad (4.4.14)$$

In the above, g and L are known. Given that we can find Z , Y_t is known. However, if we write the equation forward

$$Y_t = Y_0 + \int_0^t C(s, X_s)ds - \int_0^t Z_s dB_s. \quad (4.4.15)$$

Here, $Y_0 = V(0, X_0)$ is not known.

The following theorem shows that a BSDE can generalize the Feynmann-Kac formula to nonlinear equations.

Theorem 4.4.2. Assume that the following semilinear³ PDE has a solution $V(t, x) \in C^{1,2}$.

$$\begin{cases} 0 = \partial_t v(t, x) + [\mu \cdot \nabla v](t, x) + \frac{1}{2}[\sigma^\top \sigma \cdot D^2 V](t, x) + C(t, x, v(t, x), \nabla v(t, x)) \\ V(T, x) = g(x). \end{cases} \quad (4.4.16)$$

Then, $Y_t = V(t, X_t)$ and $Z_t = \sigma(t, X_t)\nabla V(t, X_t)$ satisfy the BSDE

$$\begin{cases} dY_t = -C(t, X_t, Y_t, \sigma^{-1}(t, X_t)Z_t)dt + Z_t dB_t \\ Y_T = g(X_T) \end{cases} \quad (4.4.17)$$

where

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (4.4.18)$$

³The an equation with linear second order term and possibly nonlinear first order term is called semilinear.

Proof. By applying Itô lemma on $Y_t = V(t, X_t)$, we obtain

$$\begin{aligned} dY_t &= (\partial_t V(t, X_t) + \mu(t, X_t) \nabla V(t, X_t) + \frac{1}{2} \sigma^2(t, X_t) \partial_{xx} V(t, X_t)) dt + \sigma(t, X_t) \nabla V(t, X_t) dB_t \\ &= -C(t, X_t, V(t, X_t), \nabla V(t, X_t)) dt + \sigma(t, X_t) \nabla V(t, X_t) dB_t \\ &= -C(t, X_t, Y_t, \sigma^{-1}(t, X_t) Z_t) dt + Z_t dB_t \end{aligned} \quad (4.4.19)$$

□

In the above, the second equality is obtained from the PDE and the third equality is from the definition of Y and Z . In addition, by the terminal condition we have, $Y_T = V(T, X_T) = g(X_T)$. In general, the use of BSDEs over the HJB is preferable when the regularity of the function $V(t, x)$ is not established. Because by the existence theorem for the BSDEs, we already know that Z_t exists as stochastic process even if $\sigma(t, X_t) \nabla V(t, X_t)$ does not make sense in cases when ∇V does not exist as specific points.

Motivated by Theorem 4.4.2, we can define a general BSDE as

$$\begin{cases} dY_t = -C(t, Y_t, Z_t, \omega) dt + Z_s dB_s \\ Y_T = \xi \end{cases}, \quad (4.4.20)$$

The argument ω inside L represents a general dependence on the randomness. In particular, it can represent solution X_t of a possible path-dependent SDE

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dB_t, \quad (4.4.21)$$

a possible control process u_t , or both X_t and u_t . We discuss such dependencies in further details in Section 4.4.3.

Remark 4.4.1. In (4.4.20), we blended σ^{-1} inside the Lagrangian $L(s, X_s, Y_s, \sigma^{-1} Z_s, \omega)$ into it and simply write $L(s, X_s, Y_s, Z_s, \omega)$.

Theorem 4.4.3 (Existence and uniqueness theorem for BSDEs). *Assume that*

- i) $\mu(t, x)$ and $\sigma(t, x)$ are Lipschitz in t and x
- ii) for all (t, x) with $0 < \lambda |x|^2 \leq x^\top \sigma^\top \sigma x$ for all x and for some λ (in particular σ is invertible)
- iii) $L(t, x, \varrho, \Pi, \omega)$ is progressively measurable in ω and Lipschitz in other variables (t, x, ϱ, Π) and is decreasing in ϱ a.s.

Then, for any \mathcal{F}_T -measurable square-integrable random variable ξ , there exists a couple (Y_t, Z_t) such that (4.4.20) holds, i.e.,

$$Y_t = \xi + \int_t^T L(s, X_s, Y_s, Z_s, \omega) ds - Z_s dB_s, \quad (4.4.22)$$

and there exists a constant C that only depends on the Lipschitz constant of μ , σ , and L and the constant λ that

$$\mathbb{E} \left[\sup_{t \in [0, T]} |Y_t|^2 + \int_0^T Z_s^2 ds \right] \leq C (\mathbb{E}[\xi^2]) \quad (4.4.23)$$

4.4.1 Linear BSDEs

Linear BSDEs take the form

$$dY_t = -(\alpha_t Y_t + \beta_t Z_t + L_t)dt + Z_t dB_t, \quad (4.4.24)$$

where α , β , and L are arbitrary progressively measurable processes. One can write a closed-form solution for a linear BSDE. To do so, note that

$$dY_t = -(\alpha_t Y_t + L_t)dt + Z_t dB_t^\beta, \quad (4.4.25)$$

where by the Girsanov theorem, $dB_t^\beta := \beta_t dt + dB_t$ is a Brownian motion under the probability \mathbb{P}^β given in

$$\frac{d\mathbb{P}^\beta}{d\mathbb{P}} \Big|_{\mathcal{F}_t} := \exp \left(\int_0^t \beta_s dB_s - \frac{1}{2} \int_0^t \beta_s^2 ds \right) \quad (4.4.26)$$

If we define $\tilde{Y}_t := e^{\int_0^t \alpha_s ds} Y_t$, then

$$d\tilde{Y}_t = e^{\int_0^t \alpha_s ds} (L_t dt + Z_t dB_t^\beta). \quad (4.4.27)$$

In other words,

$$\tilde{Y}_t = \tilde{Y}_T - \int_t^T e^{\int_0^s \alpha_z dz} (L_s dt + Z_s dB_s^\beta), \quad (4.4.28)$$

or,

$$Y_t = e^{\int_t^T \alpha_s ds} \tilde{Y}_T - \int_t^T e^{\int_t^s \alpha_z dz} (L_s dt + Z_s dB_s^\beta) = e^{\int_t^T \alpha_s ds} \xi - \int_t^T e^{\int_t^s \alpha_z dz} (L_s dt + Z_s dB_s^\beta). \quad (4.4.29)$$

After taking conditional expectation with respect to \mathbb{P}^β , we obtain

$$Y_t = \mathbb{E}^\beta \left[e^{\int_t^T \alpha_s ds} \xi - \int_t^T e^{\int_t^s \alpha_z dz} L_s dt \Big| \mathcal{F}_t \right]. \quad (4.4.30)$$

If we define

$$\Gamma_t = e^{\int_t^T \alpha_s ds} \frac{d\mathbb{P}^\beta}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \exp \left(\int_0^t \beta_s dB_s + \int_0^t (\alpha_s - \frac{1}{2} \beta_s^2) ds \right), \quad (4.4.31)$$

by changing the measure back to \mathbb{P}

$$Y_t = \Gamma_t^{-1} \mathbb{E} \left[\Gamma_T \xi - \int_t^T \Gamma_s L_s dt \Big| \mathcal{F}_t \right]. \quad (4.4.32)$$

4.4.2 Comparison principle

Consider

$$\begin{cases} dY_t^{(i)} = -L^{(i)}(t, Y_t^{(i)}, Z_t^{(i)}, \omega)dt + Z_s^{(i)} dB_s \\ Y_T^{(i)} = \xi^{(i)} \end{cases}, \quad (4.4.33)$$

for $i = 1, 2$. The following theorem provides a sufficient condition for comparing $Y^{(1)}$ and $Y^{(2)}$, and hence, it is called comparison principle.

Theorem 4.4.4. Assume that for $i = 1, 2$, $(Y^{(i)}, Z^{(i)})$ is the solution for (4.4.33) and $L(t, y, z, \omega)$ is Lipschitz

in (y, z) uniformly in ω and t . Further assume that $\xi^{(1)} \geq \xi^{(2)}$ and, for each value (t, y, z) , $L^{(1)}(t, y, z, \omega) \leq L^{(2)}(t, y, z, \omega)$, a.s. Then, $Y^{(1)} \geq Y^{(2)}$, a.s.

Proof. □

4.4.3 Maximum principle

For the purpose of control theory, we assume that L takes the form $L(t, x, \varrho, \Pi, u_t)$, where u_t is a progressively measurable process that represents a control at time t . In this case, we denote the solution to the BSDE by (Y^u, Z^u) , where

$$\begin{cases} Y_t^u = \xi + \int_t^T F(s, X_s, Y_s^u, Z_s^u, u_s) ds - \int_t^T Z_s^u dB_s \\ Y_T = \xi \in \mathcal{F}_T^X \\ dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \end{cases} \quad (4.4.34)$$

and $F(s, X_s^u, Y_s^u, Z_s^u, u_s) = L(s, X_s, Y_s, \sigma^{-1}(s, X_s^u, u_s)Z_s, u_s)$. Then, a control problem can be written as

$$Y_t = \text{esssup}_{u \in \mathcal{A}_{t,T}} Y_t^u. \quad (4.4.35)$$

Note that the set of admissible controls are defined specific to a particular problem. In addition, if the terminal condition $\xi = g(X_t^u)$, the problem is a Markovian control problem that was studied in previous sections. Otherwise, it is not Markovian, and therefore, the value cannot be written as a function $V(t, x)$. In fact, value function takes a more general form of

$$Y_t = \text{esssup}_{u \in \mathcal{A}_{t,T}} \mathbb{E} \left[\int_t^T F(s, X_s^u, Y_s^u, Z_s^u, u_s) ds + \xi \middle| \mathcal{F}_t^X \right]. \quad (4.4.36)$$

In the above, the Lagrangian F not only depends on the state of the system X_s^u and control u_s , but also depends on the history of the value Y_s^u and the sensitivity of the value with respect to the Brownian noise Z_t^u .

To solve such optimal control problems, we require the following comparison principle.

Theorem 4.4.5 (Maximum principle for BSDEs). *The process*

$$Y_t = \text{esssup}_{u \in \mathcal{A}_{t,T}} \mathbb{E} \left[\int_t^T F(s, X_s^u, Y_s^u, Z_s^u, u_s) ds + \xi \middle| \mathcal{F}_t^X \right] = \text{esssup}_{u \in \mathcal{A}_{t,T}} Y_t^u \quad (4.4.37)$$

there exists a Z such that (Y, Z) satisfies the BSDE

$$\begin{cases} Y_t = \xi + \int_t^T F^*(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s dB_s \\ Y_T = \xi \in \mathcal{F}_T^X \\ dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \end{cases} \quad (4.4.38)$$

where

$$F^*(s, x, y, z) := \sup_{u \in U} F(s, x, y, z, u), \quad \text{and} \quad u^*(s, x, y, z) := \operatorname{argmax}_{u \in U} F(s, x, y, z, u) \quad (4.4.39)$$

Chapter 5

Numerical evaluation of stochastic control problems

5.1 DPP based approximation

Recall from (??), the value function and the optimal control in discrete stochastic control problem (??) can be evaluated through.

$$\begin{cases} V(t, x) = \inf_u C(t, x, u) + \mathbb{E}[V(t+1, X_{t+1}^u) | X_t = x] \\ V(T, x) = g(x) \\ u_t^*(x) \in A(x) := \operatorname{argmax}_u C(t, x, u) + \mathbb{E}[V(t+1, X_{t+1}^u) | X_t = x] \end{cases} \quad (5.1.1)$$

where

$$X_{t+1}^u = x + \mu(t, x, u) + \sigma(t, x, u)\xi_{t+1} \quad (5.1.2)$$

and $\{\xi_t\}_{t=1}^T$ is a sequence of i.i.d. random variables with mean 0 and variance 1. In the above, we need to evaluate (1) conditional expectation $\mathbb{E}[\cdot | X_t = x]$ and (2) the infimum over control u . Given $V(t+1, \cdot)$, these calculations can be done separately. However, the separate evaluation creates inefficiency in the calculations. Therefore, we propose the one-shot approximation of $\inf_u C(t, x, u) + \mathbb{E}[V(t+1, X_{t+1}^u) | X_t = x]$ through the following methods.

5.1.1 Evaluation of infimum

First note that

$$\begin{aligned} C(t, x, u(x)) + \mathbb{E}[V(t+1, X_{t+1}^u) | X_t = x] \\ = \mathbb{E}[C(t, X, u(X)) + V(t+1, X + \mu(t, X, u(X)) + \sigma(t, X, u(X))\xi_{t+1}) | X = x] \end{aligned} \quad (5.1.3)$$

Denoting the right-hand side above by $\Phi(x, u)$, then we seek a function $\hat{u}^*(x)$ such that

$$\hat{u}^*(\cdot) \in A(X) := \operatorname{argmin}_{u(\cdot)} \mathbb{E}[\Phi(X, u(X))] \quad (5.1.4)$$

The following Lemma guarantees that \hat{u}^* solves (5.1.1).

Lemma 5.1.1. *Assume that $\hat{u}^* \in A(X)$ defined above. Then, $\hat{u}^*(x) \in A(x)$ for all x in the set of values of X .*

Proof. For a complete proof, we need a measurable selection theorem and some other conditions. For simplicity, we only provide the sketch of the proof. Assume that $u^*(x) \notin A(x)$ with a positive probability on the set of values of X . We denote the set by B . Then, $x \in B$, we define $\tilde{u}^*(x)$ such that

$$\tilde{u}^*(x) \in \operatorname{argmax}_u C(t, x, u) + \mathbb{E}[V(t+1, X_{t+1}^u) | X_t = x]$$

and $\tilde{u}^*(x) = \hat{u}^*(x)$. Therefore,

$$\mathbb{E}[\Phi(X, \tilde{u}^*(X))] < \mathbb{E}[\Phi(X, \hat{u}^*(X))]$$

which contradicts the definition of \hat{u}^* . \square

From the nonparametric point of view, we can approximate $\hat{u}^*(\cdot)$ in (5.1.4) by a nonparametric model $u(\cdot; \theta)$ via the following optimization

$$\theta^* \in \operatorname{argmin}_{\theta} \mathbb{E}[\Phi(X, u(X; \theta))] \quad (5.1.5)$$

For instance, $u(\cdot; \theta)$ can be a neural network with parameter θ .

5.2 Multi-step evaluation

While (5.1.1) is a classic way to evaluate optimal control problems, it is not very efficient. There are two reasons for the lack of efficiency. First, we have to run a loop over the number of time steps. Second, after evaluation of each nonparametric, we need to evaluate the value function. The total number of parameters is the number of parameters in each step times the number of steps, which for nonparametric models is massive and potentially needs a lot of memory. Therefore, we propose the following optimization in place of (5.1.1).

$$\inf_{u(\cdot, \cdot)} \mathbb{E} \left[\sum_{t=0}^{T-1} C(t, X_t^u, u(t, X_t^u)) + g(X_T^u) \right] \quad (5.2.1)$$

where

$$\begin{cases} X_{t+1}^u = X_t^u + \mu(t, X_t^u, u(t, X_t^u)) + \sigma(t, X_t^u, u(t, X_t^u))\xi_{t+1} \\ X_0 \text{ is a random variable.} \end{cases} \quad (5.2.2)$$

The justification for the above problem is the same as Lemma 5.1.1 for one-step DPP method. Note that we can use a nonparametric model, $u(t, x; \theta)$, to approximate the minimization problem by

$$\theta^* \in \operatorname{argmin}_{\theta} \mathbb{E} \left[\sum_{t=0}^{T-1} C(t, X_t^u, u(t, X_t^u; \theta)) + g(X_T^u) \right] \quad (5.2.3)$$

with

$$\begin{cases} X_{t+1}^u = X_t^u + \mu(t, X_t^u, u(t, X_t^u; \theta)) + \sigma(t, X_t^u, u(t, X_t^u; \theta))\xi_{t+1} \\ X_0 \text{ is a random variable.} \end{cases} \quad (5.2.4)$$

When an approximate optimal strategy, $u(t, x; \theta^*)$, is found, then one can find the value function through evaluating

$$\mathbb{E}_{t,x} \left[\sum_{s=t}^{T-1} C(s, X_s^u, u(s, X_s^u; \theta^*)) + g(X_T^u) \right] \quad (5.2.5)$$

with

$$\begin{cases} X_{s+1}^u = X_s^u + \mu(s, X_s^u, u(s, X_s^u; \theta^*)) + \sigma(s, X_s^u, u(s, X_s^u; \theta^*))\xi_{s+1} \\ X_t = x \end{cases} \quad (5.2.6)$$

Of course, the evaluation of the conditional expectation above is a different problem.

5.3 Numerical methods based on BSDEs

Recall from Section 4.4 that the solution $V(t, x)$ to the semilinear equation

$$\begin{cases} 0 = \partial_t v(t, x) + [\mu \cdot \nabla v](t, x) + \frac{1}{2}[\sigma^\top \sigma \cdot D^2 V](t, x) + C(t, x, v(t, x), \nabla v(t, x)) \\ V(T, x) = g(x). \end{cases} \quad (5.3.1)$$

is related to the BSDE

$$\begin{cases} dY_t = -C(t, X_t, Y_t, \sigma^{-1}(t, X_t)Z_t)dt + Z_t dB_t \\ Y_T = g(X_T) \end{cases} \quad (5.3.2)$$

by

$$Y_t = V(t, X_t) \text{ and } Z_t = [\sigma \nabla V](t, X_t) \quad (5.3.3)$$

where

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t. \quad (5.3.4)$$

One way to interpret the solution to the PDE is to find functions $\mathbf{Y}(t, x)$ and $\mathbf{Z}(t, x)$ such that

$$\mathbf{Y}(t, X_t) = g(X_T) - \int_t^T \left(C(s, X_s, \mathbf{Y}(s, X_s), \mathbf{Z}(s, X_s))ds + \mathbf{Z}(s, X_s)dB_s \right) \quad (5.3.5)$$

Functions \mathbf{Y} and \mathbf{Z} can be approximated by neural networks $\hat{\mathbf{Y}}(t, x, \alpha)$ and $\hat{\mathbf{Z}}(t, x; \beta)$ and the equations (5.3.4) and (5.3.5) can be approximated discretely by

$$\begin{cases} \hat{\mathbf{Y}}(t_n, \hat{X}_{t_n}) = \hat{\mathbf{Y}}(t_{n+1}, \hat{X}_{t_{n+1}}) - C(t_n, \hat{X}_{t_n}, \hat{\mathbf{Y}}(t_n, \hat{X}_{t_n}), \hat{\mathbf{Z}}(t_n, \hat{X}_{t_n}))\Delta t + \hat{\mathbf{Z}}(t_n, \hat{X}_{t_n})\Delta B_{t_{n+1}} \\ \hat{X}_{t_{n+1}} = \hat{X}_{t_n} + \mu(t_n, \hat{X}_{t_n})\Delta t + \sigma(t_n, \hat{X}_{t_n})\Delta B_{t_{n+1}}. \end{cases} \quad (5.3.6)$$

Specifically, we shall determine $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}$ such that $g(\hat{X}_T)$ is as close to $\hat{\mathbf{Y}}(T, \hat{X}_T)$ as possible. In other words, we can find the approximate solution $\hat{\mathbf{Y}}(t, x, \alpha^*)$ and $\hat{\mathbf{Z}}(t, x; \beta^*)$ with

$$(\alpha^*, \beta^*) \in \underset{\alpha, \beta}{\operatorname{argmin}} \mathbb{E} \left[\left(\hat{\mathbf{Y}}(T, \hat{X}_T) - g(\hat{X}_T) \right)^2 \right] \quad (5.3.7)$$

.1 Results from optimization

In this sections, we quickly review two of the most influential results from optimization, namely the Lagrange multiplier and gradient descent, which later be used in the context of control problems.

.1.1 Lagrange multiplier

If we add a constraint to a simple minimization problem such as $\min_x f(x)$, the Lagrange multiplier method is the way to proceed. In a nutshell, the Lagrange multiplier method turns a constrained optimization problem into a saddle point problem without constraints by adding more variables.

Consider the constrained problem below:

$$\inf_x f(x) \quad \text{subject to} \quad g(x) = 0 \quad (.1.1)$$

Define the Lagrangian:

$$L(x, \lambda) := f(x) - \lambda \cdot g(x) \quad (.1.2)$$

Then, under proper conditions, the following saddle point problem yields the solution to (.1.1).

$$\sup_{\lambda} \inf_x L(x, \lambda) \quad (.1.3)$$

The function $H(\lambda) = \inf_x L(x, \lambda)$ is called Hamiltonian and the Lagrange multiplier λ is called the dual variable. The unconstrained problem

$$\sup_{\lambda} H(\lambda) \quad (.1.4)$$

is called the dual problem for the primal problem (.1.1).

The key to the success of the Lagrange multiplier method is the strong duality.

$$\sup_{\lambda} \inf_x L(x, \lambda) = \inf_x \sup_{\lambda} L(x, \lambda) \quad (.1.5)$$

With strong duality, if for some x , $g(x) \neq (0)$, then

$$\sup_{\lambda} L(x, \lambda) = \sup_{\lambda} \{f(x) - \lambda \cdot g(x)\} = \infty \quad (.1.6)$$

Therefore, $\inf_x \sup_{\lambda} L(x, \lambda)$ is restricted to x with $g(x) = 0$ and

$$\inf_{x: g(x)=0} \sup_{\lambda} L(x, \lambda) = \inf_{x: g(x)=0} f(x) \quad (.1.7)$$

It is not always easy to check if strong duality holds. However, Karush-Kuhn-Tucker conditions (KKT) provide some necessary and sufficient conditions for strong duality.

Theorem .1.1. Assume the differentiability of f and g . If x^* solves (.1.1) and λ^* solves (.1.4) such that

$$\begin{cases} \nabla f(x^*) - \lambda^* \cdot \nabla g(x^*) = 0 \\ g(x^*) = 0 \end{cases} \quad (.1.8)$$

then, strong duality holds and (x^*, λ^*) is a saddle point for $\sup_{\lambda} \inf_x L(x, \lambda)$.

Conversely, if strong duality holds, then any saddle point (x^*, λ^*) satisfies (.1.8). In particular, x^* solve (.1.1).

A geometric interpretation of λ^* in KKT conditions is explained in See Figure .1.1.

When the constraint is given by some inequalities and equalities, i.e.,

$$\inf_x f(x) \quad \text{subject to} \quad g(x) = 0, \quad h(x) \geq 0 \quad (.1.9)$$

the range of dual variable in Hamiltonian changes:

$$L(x, \lambda, \mu) := f(x) - \lambda \cdot g(x) - \mu \cdot h(x) \quad (.1.10)$$

$$\sup_{\mu \geq 0} \sup_{\lambda} \inf_x L(x, \lambda, \mu) \quad (.1.11)$$

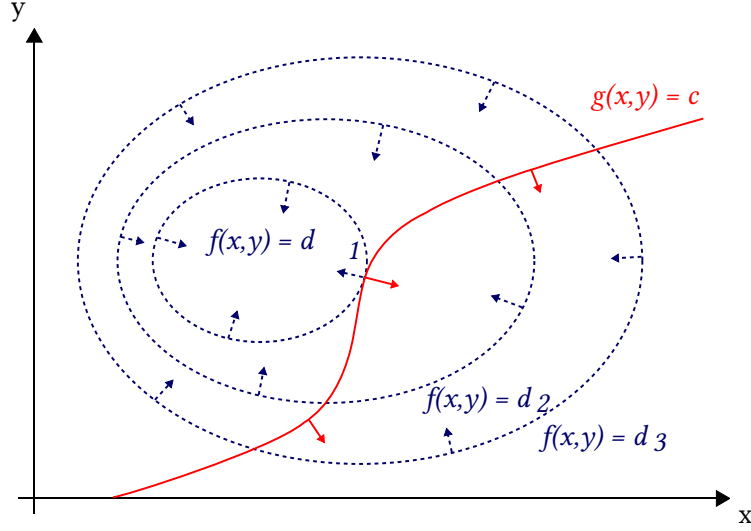


Figure .1.1: KKT conditions for $g(x) = 0$. Source WIKIPEDIA

The reason for such modification can formally be seen after switching \inf_x and $\sup_{\mu \geq 0}$ in the above

$$\sup_{\mu \geq 0} \inf_x L(x, \lambda, \mu) = \inf_x \sup_{\mu \geq 0} L(x, \lambda, \mu) \quad (.1.12)$$

If x_1 is such that $h(x_1) < 0$, then

$$\sup_{\mu \geq 0} L(x_1, \lambda, \mu) = f(x_1) - \lambda g(x_1) - h(x_1) \sup_{\mu \geq 0} \mu = \infty$$

Therefore, $\inf_x \sup_{\mu \geq 0} L(x, \lambda, \mu)$ is not attained at x_1 . Thus, any saddle point (x^*, λ^*, μ^*) with $\mu^* > 0$ must satisfy $h(x^*) \geq 0$. More general KKT conditions guarantee the strong duality in this case:

$$\begin{cases} \nabla f(x^*) - \lambda^* \nabla g(x^*) - \mu^* \nabla h(x^*) = 0 \\ g(x^*) = 0 \\ h(x^*) \geq 0 \\ \mu^* \geq 0 \\ \mu^* \cdot h(x^*) = 0 \end{cases} \quad (.1.13)$$

The last equality emphasizes that either $h(x^*) > 0$ holds, in which case $\mu^* = 0$, or $h(x^*) = 0$, in which case μ^* is irrelevant. See Figure .1.2.

Lagrange multiplier and constrained dynamic optimization

Let's first put the Lagrange multiplier in the context of an optimization problem:

$$\inf \left\{ \int_0^T (x_t^2 - \alpha_t x_t) dt \right\} \quad (.1.14)$$

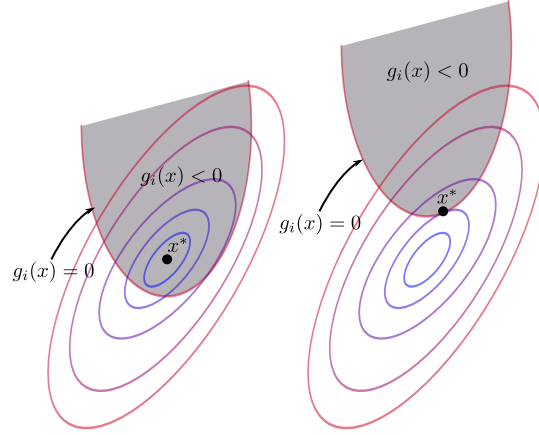


Figure .1.2: KKT conditions for $h(x) \geq 0$. The figure uses g for h . Source WIKIPEDIA

subject to $x_T = 0$. The Lagrangian is

$$\int_0^T (x_t^2 - \alpha_t x_t) dt + \lambda x_T \quad (.1.15)$$

and KKT condition suggests that $x_t^* = 2\alpha_t$ for $t < T$ and $x_T^* = 0$ solves the problem. However, this is not an interesting problem. One can simply argue that changing x at T does not change the value of the integral, and therefore it is not really a constraint. Even if we impose a more restricted constraint such as $x_t \geq 0$ on all t , the myopic solution is simply $x_t^* = \max\{\alpha_t, 0\}/2$.

However, for the constraint $\int_0^T x_t dt = 0$, the optimization problem becomes more interesting.

$$\sup_{\lambda} \inf_x \int_0^T (x_t^2 - (\alpha_t + \lambda)x_t) dt \quad (.1.16)$$

KKT condition becomes $x_t^* = (\alpha_t + \lambda^*)/2$ and $\int_0^T x_t^* dt = 0$. This implies that $\lambda^* = -\frac{1}{T} \int_0^T \alpha_t dt$, and therefore $x_t^* = (\alpha_t - \frac{1}{T} \int_0^T \alpha_t dt)/2$.

Adding restrictions to a control problem is slightly more subtle due to the dynamics of x . We postpone the study of such problems to the future endeavors. However, a simple control problem can be described as a constrained optimization and can be solved via the Lagrange multiplier. We discuss this approach in the next section.

.1.2 Gradient descent

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function. We know that if x_* is a local minimum of f , then $\nabla f(x_*) = 0$. Therefore, solving $\nabla f(x) = 0$ yields critical points of the function f including the local and global minimums, if there is any. In very high-dimensional problems such as those in deep learning, f is a highly nonlinear function. The set of critical points of f can be very irregular and hard to find via solving $\nabla f(x) = 0$, [Petersen et al. \[2021\]](#). In such situation, finding a global minimum requires some luck. However, a local minimum can be found via the method of gradient descent, GD henceforth. GD is based on a simple fact. At any point x , direction of maximum descent (ascent) of f is parallel to $-\nabla f(x)$ ($\nabla f(x)$). Therefore, if we move from point x toward $-\nabla f(x)$, the value of f decreases and we are closer to a local minimum than we were at x . In other words, for some small $\alpha > 0$ $x - \alpha \nabla f(x)$ is closer to a local minimum than x was. The most important parameter in GD is α , learning rate. Ideally, when we are further from a local minimum,

Algorithm 7: Gradient descent algorithm

Data: x randomly chosen from the domain.

Parameter: Learning rate $\{\alpha_n\}_n$, number of iterations N , and tolerance $\epsilon > 0$.

1 **while** $|\nabla f(x_n)| > \epsilon$ & $n \leq N$ **do**

2 $x \leftarrow x - \alpha_n \nabla f(x);$

3 $n \leftarrow n + 1$

4 **return** x

we like to move faster, therefore, we choose a larger learning rate. When we are very close to the minimum, we like the learning rate to be smaller. A general rule of thumb suggests α_n should be such that $\alpha_n \rightarrow 0$ and $\sum_n \alpha_n = \infty$. Decrease of α_n to 0, allows DG to slow down when it gets closer to the minimum. If α_n is constant, it can potentially jump over the minimum. The justification for $\sum_n \alpha_n = \infty$ can be rigorously explained. The following theorem shows why this condition holds under some strong assumption. The theorem shows the continuous gradient descent to the global minimum of a convex function. In practice, GD is a discrete algorithm.

Theorem .1.2. Let $\int_0^\infty \alpha_t dt = \infty$ and f be a convex function with a global minimum at x_* such that for all x

$$(x - x_*) \cdot \nabla f(x) \geq \lambda |x - x_*|^2$$

Then, the solution of the ODE given by $dx_t = -\alpha_t \nabla f(x_t) dt$ converges to x_* .

Proof. We use chain rule to evaluate

$$d|x_t - x_*|^2 = 2(x_t - x_*) \cdot dx_t = -2\alpha(x_t - x_*) \cdot \nabla f(x_t) dt$$

By the assumption of the theorem, we have

$$d|x_t - x_*|^2 = -2\alpha_t(x_t - x_*) \cdot \nabla f(x_t) dt \leq -2\lambda\alpha_t|x_t - x_*|^2 dt$$

or

$$d|x_t - x_*|^2 + 2\lambda\alpha_t|x_t - x_*|^2 dt \leq 0$$

By multiplying the above by $\exp(2\lambda \int_0^t \alpha_s ds)$ and integrating, we obtain

$$\exp(2\lambda \int_0^t \alpha_s ds) |x_t - x_*|^2 \leq |x_0 - x_*|^2$$

or

$$|x_t - x_*|^2 \leq |x_0 - x_*|^2 \exp(-2\lambda \int_0^t \alpha_s ds)$$

The right-hand above goes to zero as $t \rightarrow \infty$. □

Plain GD is not working for many applications and several versions of GD are introduced overcome the issues arising in applications. For example, if dimension is very large, $d > 1e6$, evaluation of gradient of a function is very time consuming. In such cases, stochastic gradient descent, SGD henceforth, is used. SGD randomly chooses small number of directions from $\{1, \dots, d\}$ e.g. 10 out of $1e6$ and gradient is only evaluated in those directions. To stabilize SGD, the introduced momentum into gradient descent. A popular SGD algorithm is ADAMS algorithm.

Bibliography

- Guy Barles and Panagiotis E Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. In Asymptotic analysis, number 4, pages 2347–2349, 1991.
- Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. Bulletin of the American Mathematical Society, 27(1):1–67, 1992.
- MHA Davis and Pravin Varaiya. Dynamic programming conditions for partially observable stochastic systems. SIAM Journal on Control, 11(2):226–261, 1973.
- Nicole El Karoui, Nguyen Du Huu, and Monique Jeanblanc-Picqué. Compactification methods in the control of degenerate diffusions: existence of an optimal control. Stochastics, 20(3):169–219, 1987.
- Wendell Helms Fleming and Halil Mete Soner. Controlled Markov processes and viscosity solutions, volume 25. Springer, 2006.
- UG Haussmann. Existence of optimal markovian controls for degenerate diffusions. In Stochastic differential systems, pages 171–186. Springer, 1986.
- Nikolas Nüsken and Lorenz Richter. Solving high-dimensional hamilton–jacobi–bellman pdes using neural networks: perspectives from the theory of controlled diffusions and measures on path space. Partial Differential Equations and Applications, 2(4):48, 2021. doi: 10.1007/s42985-021-00102-x. URL <https://doi.org/10.1007/s42985-021-00102-x>.
- Philipp Petersen, Mones Raslan, and Felix Voigtlaender. Topological properties of the set of functions generated by neural networks of fixed size. Foundations of computational mathematics, 21(2):375–444, 2021.
- Warren B Powell. From reinforcement learning to optimal control: A unified framework for sequential decisions, 2019. URL <https://arxiv.org/abs/1912.03513>.
- Nizar Touzi. Optimal stochastic control, stochastic target problems, and backward SDE, volume 29. Springer, 2012.