Methods Of Optimal Control

Arash Fahim • •

Last Revision: December 2, 2025

Contents

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
$(\Omega, \mathbb{F}, \mathbb{P})$	a filtered probability space with right-continuous augmented filtration (usual conditions) that hosts a Brownian motion adapted to the filtration $\mathbb{F}$
$O \subseteq \mathbb{R}^d$	open set
USC(O) (LSC(O))	lower semicontinuous (upper semicontinuous) functions on O
$\mathrm{C}^k(O)$	k times continuously differentiable functions on O
$\mathrm{C}^k(\bar{O})$	functions on $\bar{O}$ k times continuously differentiable on all variables over on O with derivatives continuously extendable to $\bar{O}$
$\mathrm{C}^{k,l,\cdots}(O)$	functions on O k times continuously differentiable on first variable, l times on second variable, ...
$\mathrm{C}^{k,l,\cdots}(\bar{O})$	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
$A \cdot B := \mathrm{Tr}[AB^T]$	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
$V_t, V_x, \partial_{xx}V$	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 $\alpha : [0, T] \to\in \mathbb{R}$ be given.*

$$\inf \left\{ \int_0^T \left(x_t^2 - \alpha_t x_t \right) \mathrm{d}t \right\} \tag{1.1.1}$$

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

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

The above problem is a dynamic optimization problem.

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

$$\inf \left\{ \int_0^T \left(x_t^2 - \alpha_t x_t \right) \mathrm{d}t \right\} \tag{1.1.3}$$

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

$$\mathrm{d}x_t = (-\beta x_t + u_t)\mathrm{d}t, \quad x_0 = x \tag{1.1.4}$$

*Note that the existence of the term $u_t \mathrm{d}t$ 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: $\mathrm{d}(e^{\beta t} x_t) = e^{\beta t} u_t \mathrm{d}t$.*

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

Exercise 3. *If we modify an <u>optimal</u> 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 $\alpha_t = 2 \int_0^t e^{-\beta(s-t)} u_s \mathrm{d}s$. 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 <u>myopic</u> solution.

Example 3. *Consider the control problem:*

$$\inf \int_0^T \left(x_t^2 - \alpha_t x_t + u_t^2 \right) \mathrm{d}t \tag{1.1.5}$$

where the infimum is over all (Riemann integrable) functions $u : [0, T] \to \mathbb{R}$ and $x : [0, T] \to \mathbb{R}$ and $u : [0, T] \to \mathbb{R}$ satisfy (1.1.4). Unlike the previous example, the choice of u induces a new cost, the term $\int_0^T u_t^2 \mathrm{d}t$. 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:*

$$\inf \int_0^T (x_t^2 + u_t^2) \mathrm{d}t \tag{1.1.6}$$

where the infimum is over all (Riemann integrable) functions $u : [0, T] \to \mathbb{R}$ and $x : [0, T] \to \mathbb{R}$ and $u : [0, T] \to \mathbb{R}$ satisfy $\mathrm{d}x_t = (-\beta x_t + u_t)\mathrm{d}t, \ x_0 = x$. 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

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

where $\mathrm{d}x_t = f(x_t, u_t)\mathrm{d}t$. The function $C : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$ is called the <u>running cost</u> and $g : \mathbb{R}^d \to \mathbb{R}$ is called the <u>terminal cost</u>.

Set $\mathcal{U}$ is <u>a</u> set of functions $u : [0, T] \to \mathbb{R}^n$ called <u>control variable</u>, which is determined by the application or by the wellposedness of the problem. It is crucial to choose <u>a</u> 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

$$\inf_{u \in \mathcal{U}} \int_0^T (x_t - u_t^2)\mathrm{d}t, \quad \mathrm{d}x_t = (x_t - u_t)\mathrm{d}t \tag{1.1.8}$$

if we choose $\mathcal{U}$ to be the set of all functions $u : [0, T] \to \mathbb{R}$, 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 $u : [0, T] \to [-1, \infty)$ (some lower bound on the value), then

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

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

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

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

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

$$\inf_u \{t \geq 0 \ : \ x_t \notin D\} \tag{1.1.11}$$

where $\mathrm{d}x_t = u_t \mathrm{d}t$ 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):

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

A key concept in DPP is the value function:

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

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

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 $V(t, \cdot) : \mathbb{R}^d \to \mathbb{R}$ 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*

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

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 $[t, s]$, $\int_t^s C(r, x_r, u_r) \mathrm{d}r$, 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),

$$V(t, x) = \inf_{u \in \mathcal{U}_t} \int_t^T C(r, x_r, u_r) \mathrm{d}r + g(x_T)$$

$$= \inf_{u_1 \in \mathcal{U}_{t,s}} \inf_{u_2 \in \mathcal{U}_s} \int_t^s C(r, x_r, u_1(r)) \mathrm{d}r + \int_s^T C(r, x_r, u_2(r)) \mathrm{d}r + g(x_T)$$

$$= \inf_{u_1 \in \mathcal{U}_{t,s}} \int_t^s C(r, x_r, u_1(r)) \mathrm{d}r + \inf_{u_2 \in \mathcal{U}_s} \int_s^T C(r, x_r, u_2(r)) \mathrm{d}r + g(x_T)$$

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

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

$\square$

Numerical solution based on DPP

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

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

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

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

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

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

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 $u : [t_i, t_{i+1}] \to \mathbb{R}^m$. 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 $j = 1, ..., J$ and $i = 0, ..., N - 1$;
(x_i^j means the jth 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 $\tilde{V}(t_i, x_i^j) \leftarrow \inf_u C(t_i, x_i^j, u)\Delta t + \hat{V}(t_{i+1}, \hat{x}_{t_{i+1}}^j)$;

4 $\hat{V}(t_i, x)$ obtained from interpolation on $\tilde{V}(t_i, x_i^j)$ for $j = 1, ..., J$;

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 $i = 0, ..., N - 1$.

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 $j = 1, ..., J$? 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*

$$\inf_{u \in \mathcal{U}} \int_0^T \left(x_t^2 + u_t^2 \right) \mathrm{d}t + \frac{1}{2}x_T^2 - x_T, \quad \mathrm{d}x_t = (x_t - u_t)\mathrm{d}t \tag{1.2.7}$$

We write the value function by (1.2.2).

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

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 $\hat{x}_{t_{i+1}} = x + (x^2 + u^2)\Delta t$. 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 <u>linear-quadratic</u> 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):*

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

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, <u>Hamilton-Jacobi equation</u> 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 are 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 $x_s = x + \int_t^s f(x_r, u_r)\mathrm{d}r$. Let's insert the above Taylor polynomial into the DPP:

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

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

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

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

$$V(T, x) = g(x) \tag{1.2.11}$$

The nonlinear function in the HJ equation

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

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

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

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

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

Let's explore the HJ inside an example.

Example 6. *Consider the control problem in Section 1.1:*

$$\inf_u \left\{ \int_0^T \left(x_t^2 + u_t^2 \right) \mathrm{d}t \right\} \tag{1.2.15}$$

where $\mathrm{d}x_t = (-\beta x_t + u_t)\mathrm{d}t$. 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:

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

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 $\mathrm{d}x_t = (-\beta x_t + u_t^*)\mathrm{d}t$. 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 $\mathrm{d}x_t^* = f(t,x_t^*, u^*(t,x_t^*))$ 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.*

$$\begin{cases} \mathrm{d}S_t = -\beta_t S_t I_t \mathrm{d}t \\ \mathrm{d}I_t = (\beta_t I_t S_t - \gamma_t I_t)\mathrm{d}t \\ \mathrm{d}R_t = \gamma_t I_t \mathrm{d}t \end{cases}$$
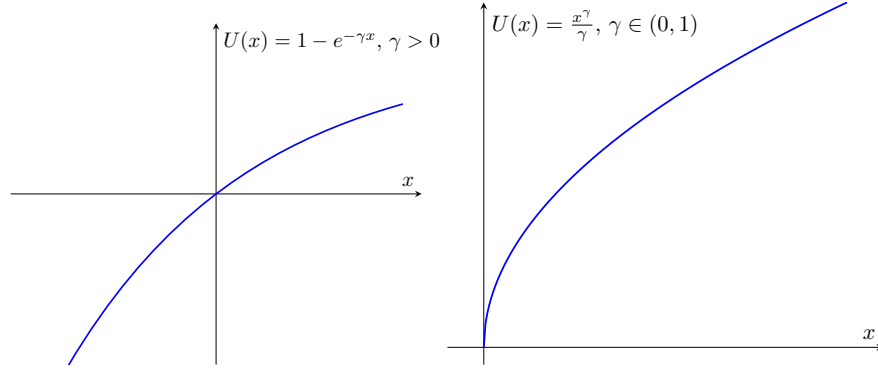
Figure 1.2.1: Utility functions: left $U(x) = 1 - e^{-\gamma x}$, right $U(x) = \frac{x^\gamma}{\gamma}$

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 and investing in a cure increases γ_t. Both 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, $g(I) = I^2$.

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

Write the HJ equation for this control problem.

Exercise 12. *In Exercise 11, notice that $\mathrm{d}(S_t + I_t + R_t) = 0$. 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:*

$$\mathrm{d}X_t = (rX_t - c_t)\mathrm{d}t$$

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

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

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 <u>gain</u> 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*

$$\sup_{c_t \geq 0} \int_0^T U(c_t)\mathrm{d}t, \quad \mathrm{d}X_t = (rX_t - c_t)\mathrm{d}t$$

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 $f(t, x, u) = f(x, u)$

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

The underlined 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*

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

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*

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

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. $\qquad\square$

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

$$e^{-k(s-t)}V(s, x_s) = V(t, x) + e^{-k(s-t)}\Big((V_t(t, x) - kV(t, x))(s - t)$$
$$+ V_x(t, x) \cdot \int_t^s f(x_r, u_r)\mathrm{d}r\Big) + o((s - t)^2)$$

We use Taylor series inside (1.2.18) to write

$$0 = \inf_u \int_t^s e^{-k(r-t)}\bar{C}(x_r, u_r)\mathrm{d}r$$
$$+ e^{-k(s-t)}\Big((V_t(t, x) - kV(t, x))(s - t)$$
$$+ V_x(t, x) \cdot \int_t^s f(x_r, u_r)\mathrm{d}r\Big) + o((s - t)^2)$$

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

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

9

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

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

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 verifies 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 $T = \infty$:

$$\inf_u \int_0^\infty e^{-kt} \bar{C}(u_t, x_t) \mathrm{d}t, \quad \mathrm{d}x_t = f(x_t, u_t) \mathrm{d}t \tag{1.2.21}$$

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

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

Therefore, value function is time homogeneous and is given by

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

and the HJ equation is given by

$$0 = -kV(t, x) + H(x, V_x) \tag{1.2.23}$$

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

$$\sup_{c_t \geq 0} \int_0^\infty e^{-kt} U(c_t) \mathrm{d}t, \quad \mathrm{d}X_t = (rX_t - c_t) \mathrm{d}t$$

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*

$$\inf_u \int_0^\infty \mathbb{1}_{\{x_t \in D\}} \mathrm{d}t, \quad \mathrm{d}x_t = u_t \mathrm{d}t \quad with \quad |u_t| \leq 1$$

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

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

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

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

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

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

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:

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

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.

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

The above equation is a second order equation with two boundary conditions, which guarantees existence
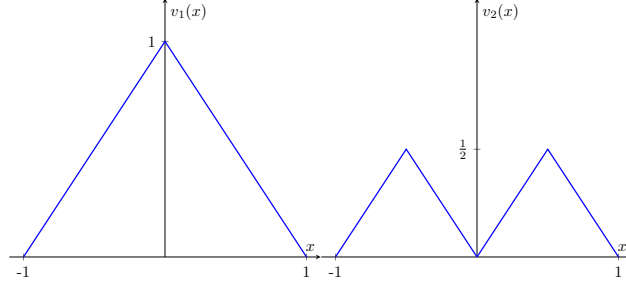
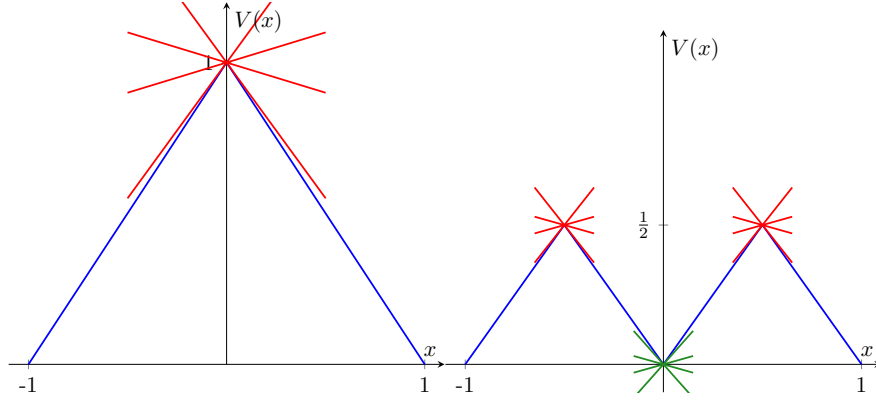**Figure 1.2.2:** Functions v_1 and v_2 solve Eikonal equation, the HJ equation for 7. But, only v_1 is the same as the value function.

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

of a unique differentiable solution which can be found in closed form by solving

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

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

Then, if we evaluate $\lim_{\epsilon \to 0} v^{\epsilon}$, 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 for v_1 at $x = 0$ and for v_2 at $x = \frac{1}{2} = -\frac{1}{2}$. Similarly, the members of subjet

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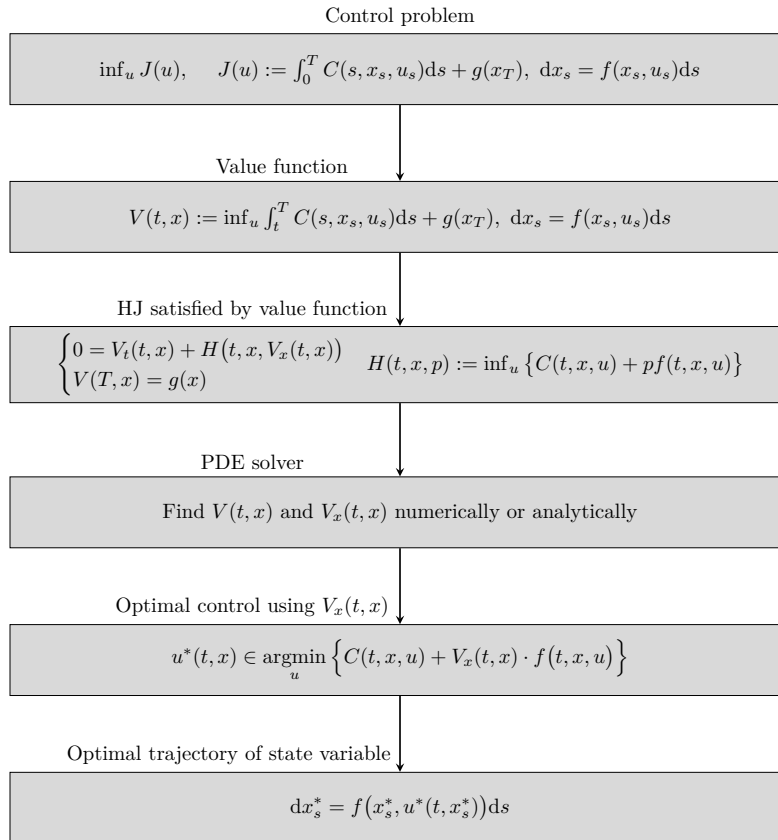
Interesting results from the theory of viscosity solution are as follows.

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

Control problem

$$\inf_u J(u), \quad J(u) := \int_0^T C(s, x_s, u_s) \mathrm{d}s + g(x_T), \ \mathrm{d}x_s = f(x_s, u_s) \mathrm{d}s$$

Value function

$$V(t, x) := \inf_u \int_t^T C(s, x_s, u_s) \mathrm{d}s + g(x_T), \ \mathrm{d}x_s = f(x_s, u_s) \mathrm{d}s$$

HJ satisfied by value function

$$\begin{cases} 0 = V_t(t, x) + H\big(t, x, V_x(t, x)\big) \\ V(T, x) = g(x) \end{cases} \quad H(t, x, p) := \inf_u \big\{ C(t, x, u) + pf(t, x, u) \big\}$$

PDE solver

Find $V(t, x)$ and $V_x(t, x)$ numerically or analytically

Optimal control using $V_x(t, x)$

$$u^*(t, x) \in \operatorname*{argmin}_u \Big\{ C(t, x, u) + V_x(t, x) \cdot f\big(t, x, u\big) \Big\}$$

Optimal trajectory of state variable

$$\mathrm{d}x_s^* = f\big(x_s^*, u^*(t, x_s^*)\big) \mathrm{d}s$$

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:

$$\inf_u \left\{ \int_0^T \left(x_t^2 - \alpha_t x_t \right) \mathrm{d}t \right\} \tag{1.2.24}$$

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

$$\sup_\lambda \inf_{u,x} \left\{ \int_0^T (x_t^2 - \alpha_t x_t)\mathrm{d}t - \int_0^T \lambda_t(\mathrm{d}x_t + (\beta x_t - u_t)\mathrm{d}t) \right\} \tag{1.2.25}$$

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

$$\int_0^T \lambda_t \mathrm{d}x_t = \lambda_T x_T - \lambda_0 x_0 - \int_0^T x_t \mathrm{d}\lambda_t \tag{1.2.26}$$

to write (1.2.25) as

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

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 <u>Hamiltonian</u> by

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

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

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

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

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 $\alpha_t = 2\int_0^t e^{-\beta(s-t)}u_s^* \mathrm{d}s$ 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.

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

By applying (1.2.26), we obtain

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

where the Hamiltonian is given by

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

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

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

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

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

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

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 $f(x, u) = x + u$:*

$$\inf_u \int_0^T (x_t^2 + u_t^2) \mathrm{d}t + x_T^2 + x_T, \ \ \textit{subject to } \ \mathrm{d}x_t = (x_t + u_t) \mathrm{d}t \tag{1.2.36}$$

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

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

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*

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

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

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

and write the ODEs as

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

(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 <u>linear-quadratic</u> 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.

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

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

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 $G : [0, T] \to \mathbb{M}(d, d)$, $H : [0, T] \to \mathbb{R}^d$, and $K : [0, T] \to \mathbb{R}$ 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

$$\mathrm{d}X_t = b(t, X_t, u_t)\mathrm{d}t$$

When the dynamics is noisy, we simply need to write

$$\mathrm{d}X_t = b(t, X_t, u_t)\mathrm{d}t + \text{noise}$$

But, what is the noise. Simplest way to model noise is through <u>white noise</u>. White noise, denoted by $\mathrm{d}W_t$ is a stochastic process that evolves by time t and for two different times t_1 and t_2, $\mathrm{d}W_{t_1}$ and $\mathrm{d}W_{t_2}$ 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 <u>Brownian motion</u>, a.k.a. <u>Wiener process</u> 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 $X_{t_1}, ..., X_{t_n}$ for different times $t_1, ..., t_n$ is important in shaping a stochastic process.*

An observation, denoted by ω, is one <u>realization</u> of a stochastic process. The function $X(\omega) : [0, \infty) \rightarrow \mathbb{R}^d$ is called a sample path of X.
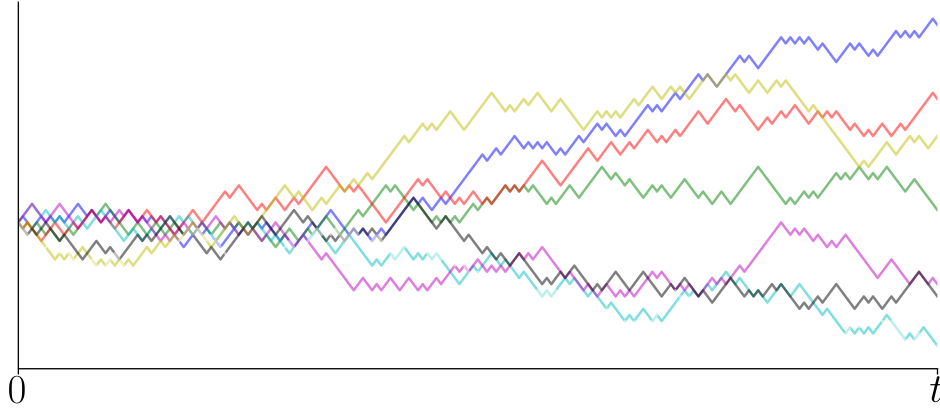
Figure 2.1.1: Eight different sample paths of a stochastic process. The randomness causes each sample paths to be different.

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 $t_1 < \cdots < t_n$, $W_{t_n} - W_{t_{n-1}}, ..., W_{t_2} - W_{t_1}$ are independent and have Gaussian distribution with mean zero and variance $t_n - t_{n-1}, ..., t_2 - t_1$.*

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 $S := \{\pm e_1, ..., \pm e_d\}$, where $\{e_i = (0, ..., 0, \underbrace{1}_{i^{\text{th}}}, 0, ..., 0) : i = 1, ..., d\}$ 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.

20
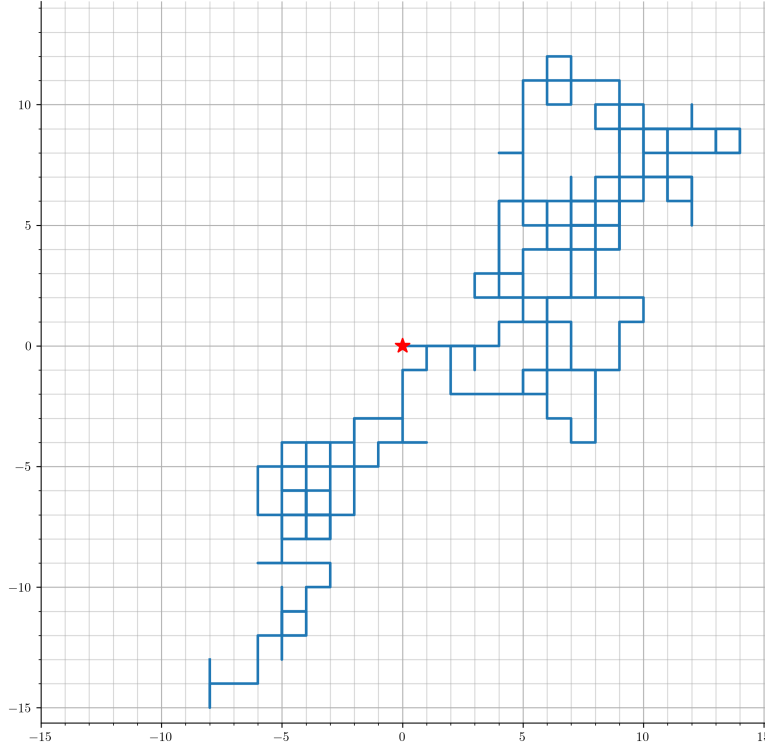
Figure 2.1.2: A sample paths of a 2-d random walk starting at $(0, 0)$ with 400 steps.

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 $N \to \infty$, 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 $\frac{\sqrt{1/N}}{1/N} = \sqrt{1/N} \to \infty$. Therefore, it is expected that the derivative of the Wiener process, $\mathrm{d}W_t$, does not exist. This derivative is what we heuristically know as the white noise.

2.2   Stochastic differential equation (SDE)

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

$$X_t = X_0 + \int_0^t b(s, X_s, u_s)\mathrm{d}s + W_t$$

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

$$X_t = X_0 + \int_0^t b(s, X_s, u_s)\mathrm{d}s + \sigma W_t$$

While σ can be a constant, we can think about a generalization where σ a function of time t and state X_t,
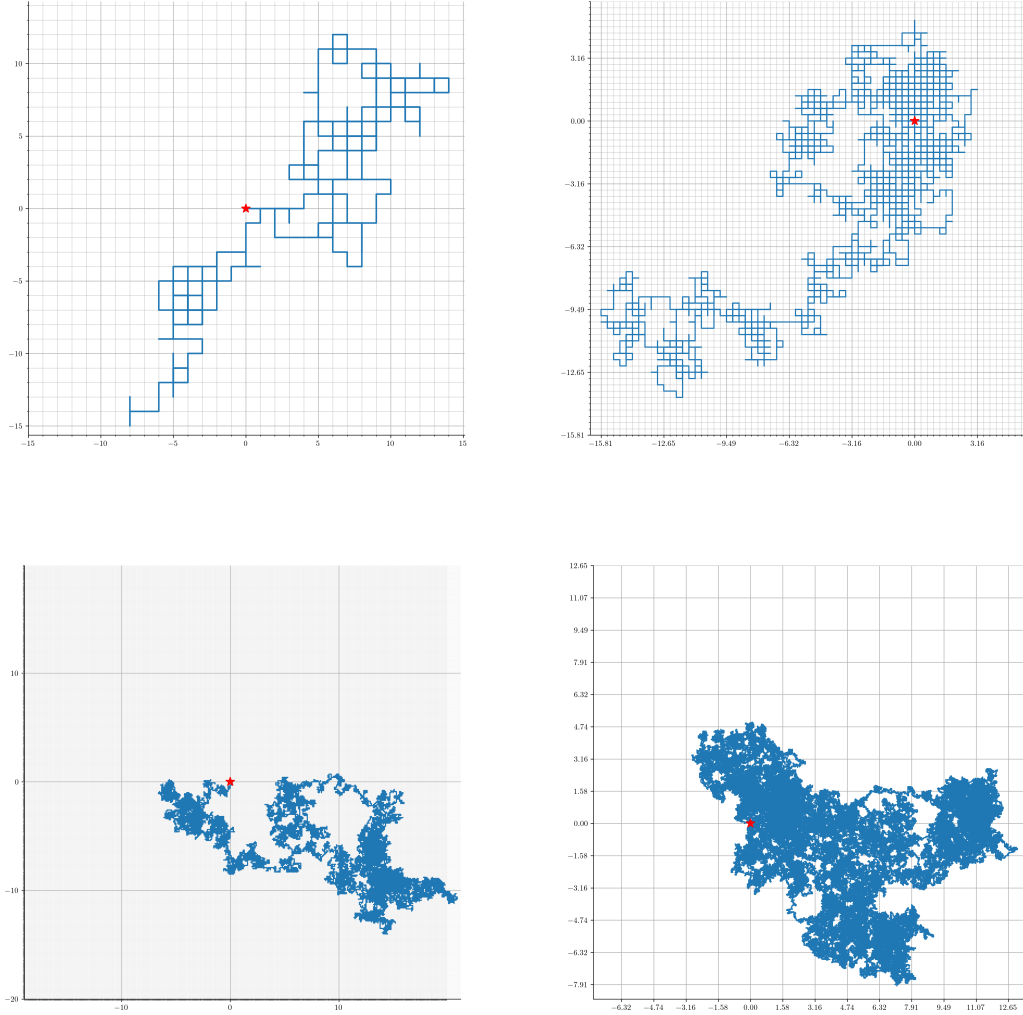
Figure 2.1.3: Top left: a random walk with step size 1 and pacing 1. Top right: a random walk with step size $\sqrt{0.1}$ and pacing 0.1. Bottom left: a random walk with step size 0.1 and pacing 0.01. Bottom right: a random walk with step size $\sqrt{0.001}$ and pacing 0.001. As $N \to \infty$, a random 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.

i.e. $\sigma(t, X_t)\mathrm{d}W_t$. This means for different (t, x) the magnitude of noise evolves in time and bases on the current value of state variable. For example, $X_t\mathrm{d}W_t$ 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, $\sigma(t, X_t, u_t)\mathrm{d}W_t$. For instance, for $\sigma(t, x, u) = ux$, $u_t X_t \mathrm{d}W_t$, when we set $u = 0$, the noise vanishes.

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

$$\mathrm{d}X_t = b(t, X_t, u_t)\mathrm{d}t + \sigma(t, X_t, u_t)\mathrm{d}W_t$$

or in integral form

$$X_t = X_0 + \int_0^t b(s, X_s, u_s)\mathrm{d}s + \int_0^t \sigma(s, X_s, u_s)\mathrm{d}W_s$$

This is what we refer to by <u>stochastic differential equation</u> or SDE.

The integral $\int_0^t \sigma(s, X_s, u_s)\mathrm{d}W_s$ is called Itô integral and is defines by

$$\int_0^t \sigma(s, X_s, u_s)\mathrm{d}W_s = \lim_{|t_n - t_{n-1}| \to 0} \sum_{n=1}^N \sigma(t_{n-1}, X_{t_{n-1}}, u_{t_{n-1}})(W_{t_n} - W_{t_{n-1}}) \tag{2.2.1}$$

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

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

and therefore,

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

On the other hand, we need to know

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

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

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

Further calculations:

$$\left(\sum_{n=1}^{N}\sigma(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(W_{t_n}-W_{t_{n-1}})\right)^2 = \sum_{n=1}^{N}\sigma^2(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(W_{t_n}-W_{t_{n-1}})^2$$

$$+ 2\sum_{1\le n<m\le N}\sigma(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(W_{t_n}-W_{t_{n-1}})\sigma(t_{m-1},X_{t_{m-1}},u_{t_{m-1}})(W_{t_m}-W_{t_{m-1}})$$

$$\mathbb{E}\left[\sum_{n=1}^{N}\sigma^2(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(W_{t_n}-W_{t_{n-1}})^2\right]$$

$$=\sum_{n=1}^{N}\mathbb{E}\left[\sigma^2(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(W_{t_n}-W_{t_{n-1}})^2\right]$$

$$=\sum_{n=1}^{N}\mathbb{E}\left[\sigma^2(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(t_n-t_{n-1})\right]$$

$$\to\mathbb{E}\left[\int_0^t\sigma^2(t_s,X_{t_s},u_{t_s})\mathrm{d}s\right]$$

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

$$\mathbb{E}\left[\sum_{1\le n<m\le N}\sigma(t_{n-1},X_{t_{n-1}},u_{t_{n-1}})(W_{t_n}-W_{t_{n-1}})\sigma(t_{m-1},X_{t_{m-1}},u_{t_{m-1}})(W_{t_m}-W_{t_{m-1}})\right]$$

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

Therefore,

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

In other words;

The noise term $\int_0^t\sigma(s,X_s,u_s)\mathrm{d}W_s$ has mean zero and variance (magnitude)

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

provided that the above expectation is finite.

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.

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

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

$$f(y) = f(x) + f'(x)\int_t^s b(r, X_r, u_r)\mathrm{d}r + \frac{1}{2}f''(x)\left(\int_t^s b(r, X_r, u_r)\mathrm{d}r\right)^2 + \cdots$$
$$\approx f(x) + f'(x)b(t, X_t, u_t)(s - t) + \frac{1}{2}f''(x)b^2(t, X_t, u_t)(s - t)^2 + \cdots$$
$$\approx f(x) + f'(x)b(t, X_t, u_t)(s - t) + O((\Delta t)^2)$$

This is the simple chair rule:

$$f(X_s) = f(X_t) + \int_t^s f'(X_r)b(r, X_r, u_r)\mathrm{d}r$$

When $y = x + \int_t^s \sigma(r, X_r, u_r)\mathrm{d}W_r$,

$$f(y) = f(x) + f'(x)\int_t^s \sigma(r, X_r, u_r)\mathrm{d}W_r + \frac{1}{2}f''(x)\left(\int_t^s \sigma(r, X_r, u_r)\mathrm{d}r\right)^2 + \cdots$$
$$\approx f(x) + f'(x)\sigma(r, X_r, u_r)(W_s - W_r) + \frac{1}{2}f''(x)\sigma^2(t, X_t, u_t)(W_s - W_r)^2 + \cdots$$
$$\approx f(x) + f'(x)\sigma(t, X_t, u_t)(W_s - W_r) + \frac{1}{2}\sigma''(x)f^2(t, X_t, u_t)(s - r) + o(\Delta t)$$

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

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

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

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

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

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

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.

2.3 Stochastic control

In this section, we consider the basic definitions of stochastic control.

2.3.1 Controlled SDE

The state process is a controlled SDE given by

$$\begin{cases} dX_t = b(t, X_t, u_t) dt + \sigma(t, X_t, u_t) dW_t \\ X_0 = x \in \mathbb{R}^d \end{cases} \tag{2.3.1}$$

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.

$$J(x; u) = \mathbb{E}\left[\int_0^T C(s, X_s^u, u_s) ds + g(X_T^u) \right]. \tag{2.3.2}$$

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*

$$\inf_{u \in \mathcal{U}} \mathbb{E}\left[\int_0^T \left(x_t^2 - \alpha_t x_t \right) dt \right] \tag{2.3.3}$$

with controlled state process is given by

$$\mathrm{d}X_t = u_t \mathrm{d}t + \sigma dW_t, \quad x_0 = x \tag{2.3.4}$$

If we can choose u such that $X_t = \alpha_t/2$ or $x + \int_0^t u_s \mathrm{d}s + \sigma W_t = \alpha_t/2$. 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 $x + \int_0^t u_s \mathrm{d}s + \sigma W_t = \alpha_t/2$ to hold, $\int_0^t u_s \mathrm{d}s$ and σW_t should cancel each other. But, Brownian motion is never differentiable. This is a twist added by the stochastic term.

2.3.2 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 $x(1 + \cdots + 2^{n-1}) = x(2^n - 1)$. In the nth 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.*

2.4 Dynamic programming principle and Hamilton-Jacobi-Bellman equations

2.4.1 Value function

Fo a stochastic control problem

$$\inf_{u \in \mathcal{U}} J(u), \quad J(u) := \mathbb{E}\left[\int_0^T C(t, X_t, u_t)\mathrm{d}t + g(X_T)\right]. \tag{2.4.1}$$

where

$$\mathrm{d}X_t = b(t, X_t, u_t)\mathrm{d}t + \sigma(t, X_t, u_t)\mathrm{d}W_t$$

We define the value function as below

Definition 2.4.1 (Value function)**.** *The value function of the control problem is defined by*

$$V(t, \omega) = \inf_{u \in \mathcal{U}_{t,T}} \mathbb{E}\left[\int_t^T C(s, X_s^u, u_s)\mathrm{d}s + g(X_T^u)\Big|\mathcal{F}_t\right] \tag{2.4.2}$$

$\mathcal{U}_{t,T}$ 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 2.4.1. *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 2.4.1. *The value function takes the form $V(t, \omega) = V(t, X_t)$. Additionally,*

$$V(t, x) = \inf_{\phi(t,x) \in \mathcal{U}_{t,T}^m} \mathbb{E}\left[\int_t^T C(s, X_s^u, u_s)\mathrm{d}s + g(X_T^u)\Big| X_t = x\right] \tag{2.4.3}$$

where $\mathcal{U}_{t,T}^m$ is he 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.

2.4.2 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 2.4.2. *If the value function is continuous and have linear growth, then for $t < s < T$, we have*

$$V(t, x) = \inf_{u \in \mathcal{U}_{t,s}} \mathbb{E}\left[\int_t^s C(r, X_r, u_r)\mathrm{d}r + V(s, X_s)\right], \quad X_t = x \tag{2.4.4}$$

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).

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

In the expansion

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

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

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

In addition,

$$
\mathbb{E}\left[\int_t^s \sigma(r, X_r, u_r)\mathrm{d}W_r \right] = 0 \tag{2.4.5}
$$

$$
\mathbb{E}\left[\left(\int_t^s \sigma(r, X_r, u_r)\mathrm{d}W_r \right)^2 \right] = \mathbb{E}\left[\int_t^s \sigma^2(r, X_r, u_r)\mathrm{d}r \right] \tag{2.4.6}
$$

Therefore,

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

Insert the Taylor expansion into DPP:

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

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

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

Using the notation of Hamiltonian, we write

$$
\begin{cases}
0 = V_t(t, x) + H(t, x, V_x, V_{xx}) \\
V(T, x) = g(x)
\end{cases} \tag{2.4.7}
$$

$$
H(t, x, p, \Gamma) := \inf_{u \in U} \left\{ C(t, x, u) + pb(t, x, u) + \frac{1}{2}\Gamma\sigma^2(t, x, u) \right\} \tag{2.4.8}
$$

where we added terminal cost as a boundary condition.

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

$$
\mathrm{d}X_t = (\theta_t \mu + r(X_t - \theta_t))\mathrm{d}t + \theta_t \sigma \mathrm{d}W_t
$$

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 $t \geq 0$; $\mathcal{U} = \{\theta_t : X_t \geq 0 \ \text{ for all } \ t \geq 0 \ a.s.\}$.

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

$$
\mathrm{d}X_t = (\theta_t \mu + r(X_t - \theta_t) - c_t)\mathrm{d}t + \theta_t \sigma \mathrm{d}W_t
$$

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

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

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

$$
H(t, x, p, \Gamma) := \inf_{u \in U} \left\{ C(t, x, u) + p \cdot b(t, x, u) + \frac{1}{2}\Gamma \cdot (\sigma^\top \sigma)(t, x, u) \right\} \tag{2.4.9}
$$

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

$$
\mathrm{d}S_t = (\mu(1 - S_t) - \beta_t S_t I_t)\,\mathrm{d}t + G_1(S_t)\mathrm{d}W_t^1 - G_2(S_t, I_t)\mathrm{d}W_t^2
$$

$$
\mathrm{d}I_t = (\beta_t S_t I_t - (\gamma_t + \mu)I_t)\,\mathrm{d}t + G_2(S_t, I_t)\mathrm{d}W_t^2 - G_3(I_t)\mathrm{d}W_t^3
$$

where $\mathrm{d}W^1$, $\mathrm{d}W^2$, *and* $\mathrm{d}W^3$, *are independent Wiener processes,* $\beta_t > 0$ *and* $\gamma_t > 0$ *are control variables, and the diffusion coefficients are defined as:*

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

Write the HJB equation for the stochastic control problem

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

Example 10. *Consider the stochastic linear quadratic problem below:*

$$\inf_u \mathbb{E}\left[\int_0^T \left(aX_t^2 + bX_t + Au_t^2 + Bu_t\right)\mathrm{d}t + \alpha X_T^2 + \beta X_T\right] \tag{2.4.10}$$

with

$$\mathrm{d}X_t = (\gamma X_t + \kappa u_t)\mathrm{d}t + \sigma \mathrm{d}W_t \tag{2.4.11}$$

where $a, A, \alpha > 0$ *and* b, B, β, γ *and* κ *are constants. The HJB is given by*

$$\begin{cases} 0 = V_t(t,x) + \frac{\sigma^2}{2}V_{xx}(t,x) + \gamma x V_x(t,x) + ax^2 + bx + \inf_u \left\{Au^2 + \left(B + \kappa V_x(t,x)\right)u\right\} \\ V(T,x) = \alpha x^2 + \beta x \end{cases} \tag{2.4.12}$$

We can evaluate $\inf_u\{Au^2 + \left(B + \kappa V_x(t,x)\right)u\} = -\frac{(B+\kappa V_x(t,x))^2}{4A}$ *with infimum attained at* $u = -\frac{B+\kappa V_x(t,x)}{2A}$.

$$\begin{cases} 0 = V_t(t,x) + \frac{\sigma^2}{2}V_{xx}(t,x) + \gamma x V_x(t,x) + ax^2 + bx - \frac{(B+\kappa V_x(t,x))^2}{4A} \\ V(T,x) = \alpha x^2 + \beta x \end{cases} \tag{2.4.13}$$

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

$$\begin{cases} f'(t) = -2\gamma f(t) + \frac{\kappa^2}{A}f^2(t) - a & f(T) = \alpha \\ h'(t) = (\frac{\kappa^2}{A}f(t) - \gamma)h(t) + \frac{B\kappa}{A}f(t) - b & h(T) = \beta \\ \ell'(t) = \frac{B^2}{4A} + \frac{B\kappa}{2A}h(t) + \frac{\kappa^2}{4A}h^2(t) - \sigma^2 f(t) & \ell(T) = 0 \end{cases} \tag{2.4.14}$$

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

2.4.3 DPP with stopping time

The DPP in Theorem 2.4.2 is a special case of the general DPP for stochastic control problems. To explain the general DPP, we need the notion of <u>stopping time</u> 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 2.4.1.
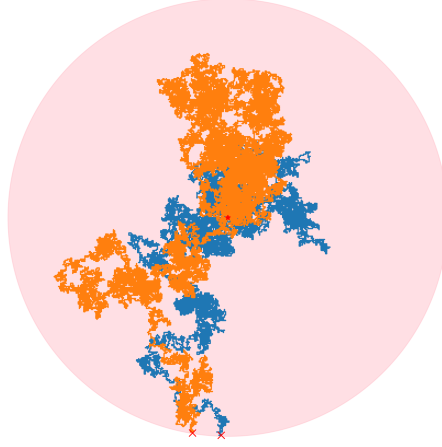
Figure 2.4.1: Two paths of random walks started at the center of the disk. One exits after 42188 steps and the other after 77374 steps.

Definition 2.4.2. *Let $\mathcal{F}_t$ represents all events that are revealed at or before time t. The a random variable $\tau : \Omega \to [0, \infty)$ 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 2.4.3. *If the value function is continuous and have linear growth, then for any stopping time τ such that $t < \tau < T$, we have*

$$V(t,x) = \inf_{u \in \mathcal{U}_{t,\tau}} \mathbb{E}\Big[\int_t^\tau C(r, X_r, u_r)\mathrm{d}r + V(\tau, X_\tau)\Big], \quad X_t = x \tag{2.4.15}$$

2.4.4 Why is stopping time in DPP important?

In Section 2.2.1, we show that for the Itô sum

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

we have

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

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

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

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

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

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

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

$$\mathbb{E}\left[\int_0^t\sigma^2(s,X_s,u_s)\mathrm{d}s\right]<\infty \tag{2.4.16}$$

then the commutation is possible and

$$\mathbb{E}\left[\int_0^t\sigma(s,X_s,u_s)\mathrm{d}W_s\right]$$

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

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

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 (2.4.5), 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 (2.4.16). However, if we write the DPP with a specific stopping time, then the issue is resolved. In (2.4.15). we can choose the stopping time to be

$$\tau=\min\left\{s,\inf\{r>t:|\sigma(r,X_r,u_r)|\geq M\}\right\} \tag{2.4.17}$$

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

$$\int_0^\tau\sigma(r,X_r,u_r)\mathrm{d}W_r$$

As a result,

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

and the stochastic integral vanishes.

Remark 2.4.1. *The above discussion is not completely rigorous. In (2.4.17), 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*

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

2.5   Verification step

In Exercise 27 and Example 10, we found closed-form solutions for the HJB. It is important to remember from Example 7 that the value function is only a viscosity solution of the HJB and finding classical solutions for the HJB is a delicacy. Nevertheless, when we have a classical solution for the HJB, there is still one more step before we can declare a victory; verification results show that the classical solution for the HJB is indeed the value function, under relatively straightforward assumptions, and it provides a way to derive an optimal control from the value function. Usually, for each control problem, the verification is slightly different from another. However, we can provide a generic verification theorem with sacrificing a little bit of rigor.

Theorem 2.5.1. *Let $v \in C^{1,2}$ be a classical solution of the HJB*

$$
\begin{cases}
0 = v_t(t,x) + \inf_{u \in U} \left\{ C(t,x,u) + v_x(t,x)b(t,x,u) \cdot \nabla v(t,x) + \frac{1}{2}(\sigma^\top \sigma)(t,x,u) \cdot D^2 v(t,x) \right\} \\
v(T,x) = g(x)
\end{cases}
$$

Then, we have $v \geq V$, where V is the value function of the control problem (2.4.1) *given by*

$$
V(t,x) = \inf_{u \in \mathcal{U}_{t,T}} \mathbb{E}\left[\int_t^T C(s, X_s^u, u_s)\mathrm{d}s + g(X_T^u) \Big| X_t = x \right] \tag{2.5.1}
$$

Furthermore, if there exists a control $u^* \in \mathcal{U}$ such that

$$
u^*(t,x) \in \mathrm{argmin}_{u \in U} \left\{ C(t,x,u) + v_x(t,x)b(t,x,u) \cdot \nabla v(t,x) + \frac{1}{2}(\sigma^\top \sigma)(t,x,u) \cdot v_{xx}(t,x) \right\}
$$

then $v = V$.

Some of the technical assumption that we skip are there to make sure the expectations and the stochastic integrals in the proof are defined. That is what needs to be modified for each specific problem.

Proof. For a fixed arbitrary control, $u \in \mathcal{U}$, we use Itô's formula to write

$$
\mathrm{d}v(t, X_t) = \left(v_t(t, X_t) + \frac{1}{2}(\sigma^\top \sigma)(t, X_t, u_t) \cdot D^2 v(t, X_t) + b(t, X_t, u_t) \cdot \nabla v(t, X_t) \right) \mathrm{d}t
$$
$$
+ \nabla v(t, X_t) \cdot \sigma(t, X_t, u_t)\mathrm{d}W_t
$$

Or equivalently,

$$
v(t, X_t) = g(X_T) - \int_0^T \Big(\left(v_t(s, X_s) + \frac{1}{2}(\sigma^\top \sigma)(s, X_s, u_s) \cdot D^2 v(s, X_s) + b(s, X_s, u_s) \cdot \nabla v(s, X_s) \right) \mathrm{d}s
$$
$$
+ \nabla v(s, X_s) \cdot \sigma(s, X_s, u_s)\mathrm{d}W_s \Big)
$$

By the HJB, we have

$$
v_t(s, X_s) + C(s, X_s, u) + \frac{1}{2}(\sigma^\top \sigma)(s, X_s, u_s) \cdot D^2 v(s, X_s) + b(s, X_s, u_s) \cdot \nabla v(s, X_s)
$$
$$
\geq v_t(s, X_s) + \inf_{u \in U} \left\{ C(s, X_s, u) + b(s, X_s, u) \cdot v_x(s, X_s) + \frac{1}{2}v_{xx}(t,x)\sigma^2(t,x,u) \right\} = 0
$$

Therefore,

$$v_t(s, X_s) + \frac{1}{2}(\sigma^\mathsf{T}\sigma)(s, X_s, u_s) \cdot D^2 v(s, X_s) + b(s, X_s, u_s) \cdot \nabla v(s, X_s) \geq -C(s, X_s, u)$$

$$v(t, X_t) \leq g(X_T) + \int_t^T \left(C(s, X_s, u_s)\mathrm{d}s + \nabla v(s, X_s) \cdot \sigma(s, X_s, u_s)\mathrm{d}W_s \right)$$

Taking expectation, we have

$$v(t, x) \leq \mathbb{E}_t \left[g(X_T) + \int_t^T C(s, X_s, u_s)\mathrm{d}s \Big| X_t = x \right]$$

Taking infimum over all $u \in \mathcal{U}$, we obtain

$$v(t, x) \leq \inf_{u \in \mathcal{U}} \mathbb{E}_t \left[g(X_T) + \int_t^T C(s, X_s, u_s)\mathrm{d}s \Big| X_t = x \right] =: V(t, x)$$

By repeating the above argument for $u^*(t, x)$ given in the assertion of the theorem, all the inequalities turn to equalities and, therefore, $v(t, x) = V(t, x)$. □

Example 11. *In Exercise 25, 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. We now perform a verification step on $v(t, x)$. The HJB for Exercises 25 is given by*

$$\begin{cases} 0 = v_t + rxv_x + \sup_\theta \left\{ \frac{\sigma^2}{2}v_{xx}\theta^2 + (\mu - r)v_x\theta \right\} \\ v(T, x) = U(x) \end{cases} \tag{2.5.2}$$

which is written in a simplified form by $0 = v_t + rxv_x - \frac{(\mu-r)^2 v_x^2}{2\sigma^2 v_{xx}}$. Note that

$$\theta^*(t, x) = \mathrm{argmax}\left\{ \frac{\sigma^2}{2}v_{xx}\theta^2 + (\mu - r)v_x\theta \right\} = -\frac{(\mu - r)v_x}{\sigma^2 v_{xx}} = -\frac{(\mu - r)f(t)x^{\gamma-1}}{\sigma^2 f(t)(\gamma - 1)x^{\gamma-2}} = \frac{(\mu - r)}{\sigma^2(1 - \gamma)}x$$

According to Theorem 2.5.1, we need to show that $\theta^* \in \mathcal{U}$, or, θ^* is an admissible control. Recall that $\mathcal{U} = \{\theta_t : X_t \geq 0 \text{ for all } t \text{ a.s.}\}$. Therefore, we need to show that $X_t^* \geq 0$ for all $t \geq 0$ a.s., where

$$\mathrm{d}X_t^* = (\theta^*(t, X_t^*)\mu + r(X_t^* - \theta^*(t, X_t^*)))\mathrm{d}t + \theta^*(t, X_t^*)\sigma\mathrm{d}W_t$$

For simplicity, we set $A = \frac{(\mu-r)}{\sigma^2(1-\gamma)}$. Then,

$$\mathrm{d}X_t^* = (A(\mu - r) + r)X_t^*\mathrm{d}t + AX_t^*\sigma\mathrm{d}W_t$$

Therefore, X_t^* is a geometric Brownian motion and is given by

$$X_t^* = X_0^* \exp\left(\left(A(\mu - r) + r - \frac{A^2}{2} \right)t + A\sigma W_t \right) \geq 0$$

which provides the desired result.

Exercise 29. *Establish verification result for Exercise 26.*

2.5.1 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 2.5.2. *Assume that there exists a function $v(t, x)$ such that for all $u \in \mathcal{U}$ the process $\{Y_t^u\}_{t \geq 0}$*

$$Y_t^u := v(t, X_t) + \int_0^t C(s, X_s, u_s) \mathrm{d}s \tag{2.5.3}$$

is a super martingale and that for some $u^* \in \mathcal{U}$, $\{Y_t^{u^*}\}_{t \geq 0}$ 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

$$Y_t^u \leq \mathbb{E}[Y_T^u | \mathcal{F}_t] = \mathbb{E}\Big[g(X_T) + \int_t^T C(s, X_s, u_s) \mathrm{d}s \Big| \mathcal{F}_t\Big] + \int_0^t C(s, X_s, u_s) \mathrm{d}s \tag{2.5.4}$$

Thus,

$$v(t, x) = Y_t^u - \int_0^t C(s, X_s, u_s) \mathrm{d}s \leq \mathbb{E}[Y_T^u | \mathcal{F}_t] = \mathbb{E}\Big[g(X_T) + \int_t^T C(s, X_s, u_s) \mathrm{d}s \Big| \mathcal{F}_t\Big], \tag{2.5.5}$$

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

$$v(t, x) \leq \sup_{u \in \mathcal{A}_{t,T}} \mathbb{E}\Big[g(X_T) + \int_t^T C(s, X_s, u_s) \mathrm{d}s \Big| X_t = x\Big] = V(t, x). \tag{2.5.6}$$

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

The above theorem can also be regarded as a verification without regularity condition. The function $v(t, x)$ in the martingale approach is the value function of the optimal control problem. The following example shows the use of this theorem.

Example 12. *In Example 26 with $U(x) = \frac{x^\gamma}{\gamma}$ for $\gamma \in (0, 1)$, recall that a solution to the HJB is given by $v(t, x) = f(t)U(x)$. Then,*

$$Y_t^{\theta, c} = v(t, X_t) + \int_0^t e^{-\gamma s} U(c_s) \mathrm{d}s. \tag{2.5.7}$$

By Itô formula, we have

$$\begin{aligned} \mathrm{d}Y_t^{\theta, c} =& \Big(v_t + \big(\theta_t(\mu - r) + rX_t - c_t\big)V_x + \frac{1}{2}\sigma^2\theta_t^2 V_{xx} + e^{-\gamma t}U(c_t)\Big)\mathrm{d}t \\ &+ \sigma\theta_t V_x \mathrm{d}W_t. \end{aligned} \tag{2.5.8}$$

By the supremum in the HJB, one can see that

$$v_t + \big(\theta_t(\mu - r) + rX_t - c_t\big)v_x + \frac{1}{2}\sigma^2\theta_t^2 v_{xx} + e^{-\gamma t}U(c_t) \leq 0, \ \ \mathbb{P}\text{-a.s.} \tag{2.5.9}$$

for all values of θ and c. In addition, for $c_t^*(x) = f^{-\frac{1}{(1-\gamma)}}(t)x$ and $\theta_t^*(x) = \frac{(\mu-r)\gamma}{\sigma^2(1-\gamma)}x$,

$$V_t + \big(\theta_t(\mu - r) + rX_t - c_t\big)V_x + \frac{1}{2}\sigma^2\theta_t^2\partial_{xx}V + e^{-\gamma t}U(c_t) = 0, \;\; \mathbb{P}\text{-}a.s. \tag{2.5.10}$$

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

$$dX_t^* = \Big(\theta_t(X_t^*)\big((\mu - r)\mathrm{d}t + \sigma\mathrm{d}W_t\big) - rX_t^*\mathrm{d}t\Big) - c_t^*(X_t^*)\mathrm{d}t, \tag{2.5.11}$$

has a non-negative solution X^*. We leave the details as an exercise.

Exercise 30. *In the above example, show that $X_t^* \geq 0$ a.s..*

2.6 Viscosity solutions to the HJB equations

In the previous section, we showed that if the value function is in $C^{1,2}$, then it satisfies (2.4.7). 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 13. *Recall from Exercise 5 the fastest exit problem:*

$$\inf_{|u_t|\leq 1}\int_0^\infty 1_{\{X_t^u\in O\}}dt \tag{2.6.1}$$

with $\mathrm{d}X_t^u = u_t dt$. 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

$$\begin{cases} 0 = -|\nabla V(x)| + 1 & x \in O \\ 0 = V(x) & x \in \partial O \end{cases} \tag{2.6.2}$$

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 $V(x) = 1 = |x|$.

$$v(x) = \begin{cases} 1 - |x| & \frac{1}{2} \leq |x| \leq 1 \\ |x| & |x| \leq \frac{1}{2} \end{cases} \tag{2.6.3}$$

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 14. *Consider the exit time problem with noise:*

$$\inf_{|u_t| \leq 1} \mathbb{E} \Big[\int_0^\infty 1_{\{X_t^u \in O\}} dt \Big] \tag{2.6.4}$$

with $dX_t^u = u_t dt + \epsilon dW_t$. *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 equivalent to the following boundary value problem*

$$\begin{cases} 0 = \frac{\epsilon^2}{2} V''(x) - |\nabla V(x)| + 1 & x \in O \\ 0 = V(x) & x \in \partial O \end{cases} \tag{2.6.5}$$

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* $\epsilon \to 0$, $V^\epsilon(x) \to 1 - |x|$. *This clarifies why* $v(x)$ *is not a solution.*

Exercise 31. *In the previous exercise, show that as* $\epsilon \to 0$, $V^\epsilon(x) \to 1 - |x|$.

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 be not exists. For instance, the backward heat equation given by

$$\begin{cases} 0 = V_t + \frac{\sigma^2}{2} \Delta V - rV + f(t, x) \\ V(0, x) = g(x) \end{cases} \tag{2.6.6}$$

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

$$\begin{cases} 0 = V_t + \frac{\sigma^2}{2} \Delta V - rV + f(t, x) \\ V(T, x) = g(x) \end{cases} \tag{2.6.7}$$

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, V_t and $\frac{\sigma^2}{2} \Delta V$ show up with the same sign. In the HJB equation (2.4.7), the coefficient of V_t 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

$$\begin{cases} 0 = \partial_t V(t, x) + F\big(t, x, V(t, x), \nabla V(t, x), D^2 V(t, x)\big) \\ V(T, x) = g(x) \end{cases} \tag{2.6.8}$$

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

$$F(t, x, \varrho, \Pi, \Gamma) := \sup_{u \in \mathbf{U}} \left\{ C(t, x, u) + \mu(t, x, u)\Pi + \frac{1}{2}\sigma^2(t, x, u)\Gamma - r(t, x, u)\varrho \right\} \tag{2.6.9}$$

is nondecreasing in Γ.

---

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

Example 15. *Show that the HJB equation* (**??**) *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 necessary 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 nondifferntiability. 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 2.6.1. Therefore, we have $\varphi(t, x) = a \cdot (x - x_0) + b(t - t_0) + V(t_0, x_0)$ such
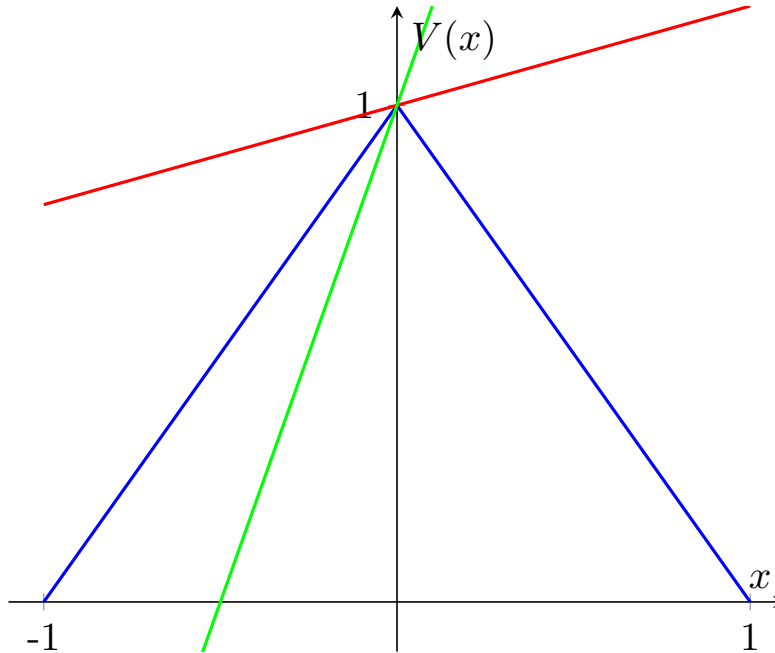
Figure 2.6.1: Red line is an acceptable tangent to the graph, while the green line is not. The green line does not keep the graph on one side locally.

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 2.6.2.

A function $\underline{V}$ (resp. $\overline{V}$) is called a viscosity subsolution (resp. super) of (2.6.8) if for any superderivative (resp. sub) (b, a, A) at point (t_0, x_0), we have

$$\begin{cases} 0 \geq b + F\big(t_0, x_0, \varphi(t_0, x_0), a, A\big), \quad \text{(resp. } 0 \leq) \\ V(T, x) \leq g(x) \quad \text{(resp. } 0 \geq) \end{cases} \tag{2.6.10}$$
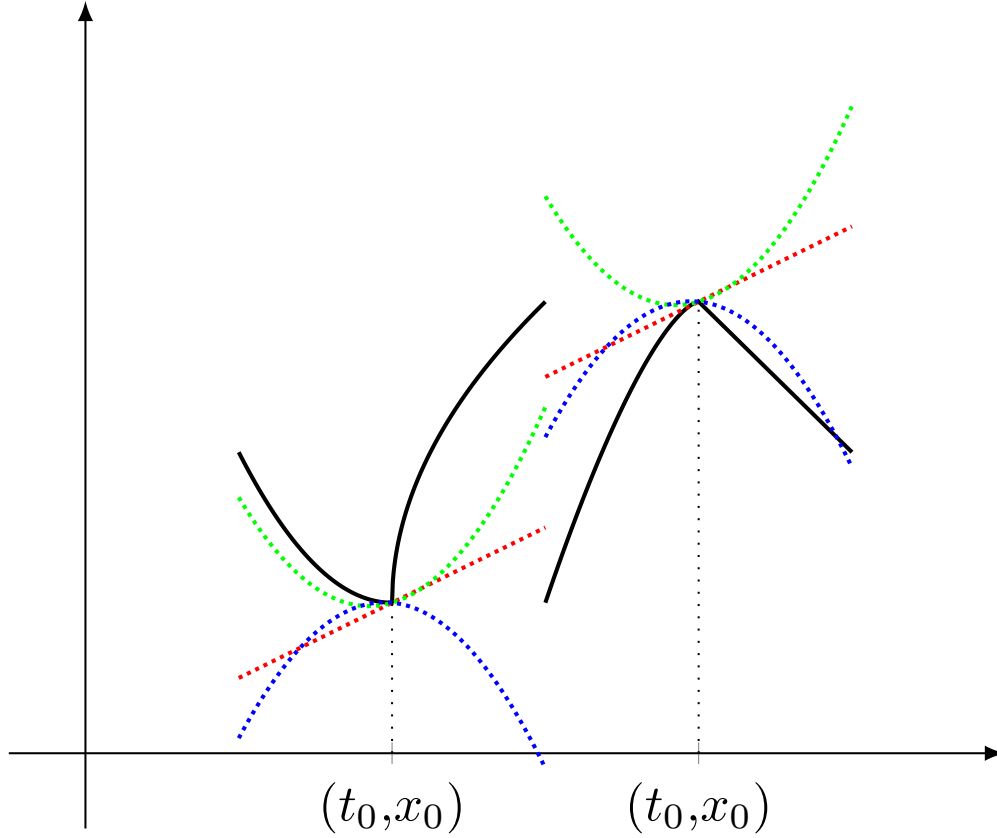
Figure 2.6.2: The test functions that represent the sub and superderivatives of a function at the points of nondifferentiability as well as other points that the function is differentiable.

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

Remark 2.6.1. *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*

$$0 = \partial_t V(t, x) + C(t, x) + \mu(t, x)\partial_x V(t, x) + \frac{1}{2}\sigma^2(t, x)\partial_{xx} V(t, x), \tag{2.6.11}$$

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

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t. \tag{2.6.12}$$

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 $\varepsilon \to 0$. For instance, in Exercise 5, if we set the state process $dX_t^u = u_t dt + \varepsilon dW_t$, we obtain the HJB

$$\begin{cases} 0 = 1 - |V'(x)| + \frac{\varepsilon^2}{2} V''(x) \\ V(\pm 1) = 0 \end{cases} \tag{2.6.13}$$

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

$$\begin{cases} 0 \geq 1 - |V'(x)| + \frac{\varepsilon^2}{2} V''(x) & (\text{resp. } 0 \leq) \\ V(\pm 1) \leq 0 & (\text{resp. } 0 \geq) \end{cases} \tag{2.6.14}$$

By sending $\varepsilon \to 0$, we obtain the following inequalities for the eikonal equation, (2.6.13).

$$\begin{cases} 0 \geq 1 - |V'(x)| & (\text{resp. } 0 \leq) \\ V(\pm 1) \leq 0 & (\text{resp. } 0 \geq) \end{cases} \tag{2.6.15}$$

We can easily check that the function $V(t,x) = 1 - |x|$ is a viscosity solution of the eikonal equation, (2.6.13). 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 32. *Show that $\tilde{V}$ is a viscosity subsolution to the eikonal equation (2.6.15), 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 2.5.1 shows that any classical solution is equal to the value function, and therefore, uniqueness is obtained.

2.7 Backward stochastic differential equations

In this section, we introduce a tool related to stochastic optimal control, i.e., backward stochastic differential equations, BSDE henceforth.

2.7.1 Itô representation theorem and BSDE for linear PDEs

To understand where a BSDE comes from, we present a result from stochastic calculus.

Theorem 2.7.1 (Itô's martingale representation theorem)**.** *Let Y be a random variable dependent on the paths of a Wiener process $\{W_t\}$ and $M_t := \mathbb{E}[Y|\mathcal{F}_t]$ be the conditional expectation of Y given the path of W*

until time t. Then, there exist a stochastic process Z_t such that

$$M_s = M_t + \int_t^s Z_r dW_r$$

In particular,

$$M_t = M_0 + \int_0^t Z_r dW_r = \mathbb{E}[Y] + \int_0^t Z_r dW_r$$

For the moment, we forget about stochastic control and focus on the random variable given by

$$Y := \int_0^T C(s, X_s)ds + g(X_T) \tag{2.7.1}$$

where

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

Note that the control is absent in the running cost function $C(t, x)$, drift $b(t, x)$, and diffusion $\sigma(t, x)$. By Itô representation theorem for Y, we obtain

$$Y = \mathbb{E}[Y|\mathcal{F}_t] + \int_t^T Z_r dW_r$$

In other words,

$$\int_0^T C(r, X_r)dr + g(X_T) = \mathbb{E}\left[\int_0^T C(r, X_r)dr + g(X_T)\Big|\mathcal{F}_t\right] + \int_t^T Z_r dW_r$$

We can use the properties of conditional expectation to write

$$\mathbb{E}\left[\int_0^T C(r, X_r)dr + g(X_T)\Big|\mathcal{F}_t\right] = \int_0^t C(r, X_r)dr + \mathbb{E}\left[\int_t^T C(r, X_r)dr + g(X_T)\Big|\mathcal{F}_t\right]$$

This implies

$$\int_0^T C(r, X_r)dr + g(X_T) = \int_0^t C(r, X_r)dr + \mathbb{E}\left[\int_t^T C(r, X_r)dr)\Big|\mathcal{F}_t\right] + \int_t^T Z_r dW_r$$

or

$$\int_t^T C(r, X_r)dr + g(X_T) = \mathbb{E}\left[\int_t^T C(r, X_r)dr)\Big|\mathcal{F}_t\right] + \int_t^T Z_r dW_r$$

We set

$$Y_t := \mathbb{E}\left[\int_t^T C(r, X_r)dr + g(X_T)\Big|\mathcal{F}_t\right]$$

Therefore, we derive the simple BSDE

$$Y_t = g(X_T) + \int_t^T C(r, X_r)dr - \int_t^T Z_r dW_r$$

Or in differential form, we write

$$\begin{cases} \mathrm{d}Y_t = -C(t, X_t)\mathrm{d}t + Z_t \cdot \mathrm{d}W_t \\ Y_T = g(X_T) \end{cases} \tag{2.7.2}$$

In order to see what Z and Y represent, we review the Feynman-Kac formula for linear parabolic PDEs. By the Markovian property of X, Y_t can be written as a function of (t, X_t)

$$Y_t = V(t, X_t) = \mathbb{E}\left[\int_t^T C(r, X_r)\mathrm{d}r + g(X_T)\Big|\mathcal{F}_t\right] \tag{2.7.3}$$

We assume that $V(t, x)$ is first order continuously differentiable with respect to t and second order continuously differentiable with respect to x. Then, we can apply Itô's formula to write

$$\mathrm{d}Y_t = \mathrm{d}V(t, X_t) = \Big(V_t(t, X_t) + \frac{1}{2}(\sigma^\mathsf{T}\sigma)(t, X_t) \cdot D^2 V(t, X_t) + b(t, X_t) \cdot \nabla V(t, X_t)\Big)\mathrm{d}t$$
$$+ \nabla V(t, X_t) \cdot \sigma(t, X_t)\mathrm{d}W_t$$

We compare the above with (2.7.2) to conclude that $Z_t^\top = \nabla V(t, X_t)^\top \sigma(t, X_t)$ and

$$\begin{cases} V_t(t, x) + \frac{1}{2}(\sigma^\mathsf{T}\sigma)(t, x) \cdot D^2 V(t, x) + b(t, x) \cdot \nabla V(t, x) + C(t, x) = 0 \\ V(T, x) = g(X_T) \end{cases} \tag{2.7.4}$$

Convention. For aesthetic reasons, we switch Z to $\sigma^{-1\top}Z$ to write the BSDE as

$$\begin{cases} \mathrm{d}Y_t = -C(t, X_t)\mathrm{d}t + Z_t \cdot \sigma(t, x)\mathrm{d}W_t \\ Y_T = g(X_T) \end{cases} \tag{2.7.5}$$

or equivalently

$$\begin{cases} \mathrm{d}Y_t = -\big(C(t, X_t) + Z_t \cdot b(t, X_t)\big)\mathrm{d}t + Z_t \cdot \mathrm{d}X_t \\ Y_T = g(X_T) \end{cases} \tag{2.7.6}$$

Formula (2.7.3) provides a stochastic representation for the solution of (2.7.4) known as <u>Feynman-Kac</u> formula, which is used to approximate the solution of parabolic equations of the form (2.7.4). This approximation method is discussed in Section 3.4. If $X_t = W_t$ is a Wiener process, $b \equiv 0$ and $\sigma \equiv \mathrm{I}_d$ (the identity matrix) and we have the heat equation with a source term

$$\begin{cases} V_t(t, x) + \frac{1}{2}\Delta V(t, x) + C(t, x) = 0 \\ V(T, x) = g(X_T) \end{cases}$$

and

$$V(t, X_t) = \mathbb{E}\left[\int_t^T C(r, W_r)\mathrm{d}r + g(W_T)\Big|\mathcal{F}_t\right]$$

However, the BSDE (2.7.5) provide more than an approximation for Y_t or $V(t, x)$. The solution to the BSDE consists of the pair (Y_t, Z_t) which yields simultaneously, $Y_t = V(t, X_t)$ and $Z_t = \nabla V(t, X_t)$. The BSDE approximation methods for control problems can also be applied for linear parabolic PDEs to obtain solution and first derivative at the same time.

2.7.2 BSDE representation for optimal control problems

In this section, we derive the BSDE for control problems in two ways. First, we use the HJB equation. Then, we derive it without using the HJB equation. In both cases, we only consider the case where the control variable is present in the drift of X,

$$\mathrm{d}X_t = b(t, X_t, u_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

When σ depends on the control, the BSDE is more involved and is refereed to as second order BSDE, 2BSDE. To know more about 2BSDEs, see Cheridito et al. [2007], Soner et al. [2012].

2.7.3 BSDE via HJB

Recall that the value function of the control problem

$$V(t, x) := \inf_{u \in \mathcal{U}_{t,T}} \mathbb{E}\left[\int_t^T C(r, X_r, u_r)\mathrm{d}r + g(X_T)\Big| X_t = x\right]$$

satisfies the HJB equation

$$\begin{cases} V_t(t, x) + \frac{1}{2}(\sigma^\mathsf{T}\sigma)(t, x) \cdot D^2 V(t, x) + H(t, x, \nabla V(t, x)) = 0 \\ V(T, x) = g(X_T) \end{cases} \tag{2.7.7}$$

with

$$H(t, x, p) = \sup_{u \in U}\{b(t, x, u) \cdot p + C(t, x, u)\}$$

Define

$$\mathrm{d}\tilde{X}_t = \sigma(t, \tilde{X}_t)\mathrm{d}W_t$$

$\tilde{X}$ is an SDE derived from X by dropping the drift $b(t, x, u)$ and it does not depend on the control anymore. We apply Ito's formula on $V(t, \tilde{X}_t)$ to write

$$\mathrm{d}V(t, \tilde{X}_t) = \left(V_t(t, \tilde{X}_t) + \frac{1}{2}(\sigma^\mathsf{T}\sigma)(t, \tilde{X}_t) \cdot D^2 V(t, \tilde{X}_t)\right)\mathrm{d}t + \nabla V(t, \tilde{X}_t) \cdot \sigma(t, \tilde{X}_t)\mathrm{d}W_t$$

By (2.7.7), we have

$$\begin{aligned} \mathrm{d}V(t, \tilde{X}_t) &= -H(t, \tilde{X}_t, \nabla V(t, X_t))\mathrm{d}t + \nabla V(t, \tilde{X}_t) \cdot \sigma(t, \tilde{X}_t)\mathrm{d}W_t \\ &= -H(t, \tilde{X}_t, \nabla V(t, X_t))\mathrm{d}t + \nabla V(t, \tilde{X}_t) \cdot \mathrm{d}\tilde{X}_t \end{aligned}$$

We set $Y_t := V(t, \tilde{X}_t)$ and $Z_t = \nabla V(t, \tilde{X}_t)$ to write

$$\begin{cases} \mathrm{d}Y_t = -H(t, \tilde{X}_t, Z_t)\mathrm{d}t + Z_t \cdot \mathrm{d}\tilde{X}_t \\ Y_T = g(\tilde{X}_T) \end{cases} \tag{2.7.8}$$

The above discussion leads to the following theorem.

Theorem 2.7.2. *Let $V(t, x)$ be the classical solution (once continuously differentiable in t and twice continuously differentiable in x) to (2.7.7) and (Y_t, Z_t) satisfies (2.7.8). Then, $V(t, \tilde{X}_t) = Y_t$ and $\nabla V(t, \tilde{X}_t) = Z_t$.*

2.7.4   BSDE via maximum principle

We present a formal process for writing the BSDE for a generic control problem. We fix a control u_t and denote

$$Y_t^u := \mathbb{E}\left[\int_t^T C(r, X_r, u_r)\mathrm{d}r + g(X_T)\right]$$

where

$$\mathrm{d}X_t = b(t, X_t, u_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t$$

Inspired by the martingale representation theorem, Theorem 2.7.1 and (2.7.6),

$$\mathrm{d}Y_t^u = -\big(C(t, X_t, u_t) + Z_t^u \cdot b(t, X_t, u_t)\big)\mathrm{d}t + Z_t^u \cdot \mathrm{d}X_t$$

We define

$$H(t, x, z) := \inf_{u \in U} \{C(t, x, u) + z \cdot b(t, x, u)\}$$

and consider the <u>nonlinear</u> BSDE

$$\mathrm{d}Y_t = -H(t, X_t, Z_t)\mathrm{d}t + Z_t \cdot \mathrm{d}X_t \tag{2.7.9}$$

Now, we appeal to the following maximum principle for BSDEs

Theorem 2.7.3. *For $i = 1, 2$, let (Y^i, Z^i) be the solution to the nonlinear BSDEs*

$$\mathrm{d}Y_t^i = -H_i(t, X_t, Z_t^i)\mathrm{d}t + Z_t^i \cdot \mathrm{d}X_t, \quad Y_T^i = g(X_T)$$

If $H_1(t, x, z) \leq H_2(t, x, z)$, then $Y_t^1 \leq Y_t^2$ for all t.

To use the above theorem, if we take $H_1(t, x, z) = C(t, x, u) + z \cdot b(t, x, u)$ and $H_2(t, x, z) = H(t, x, z)$, then we argue that

$$Y_t^u \leq Y_t$$

Furthermore, the equality is attained at

$$u^*(t, x, z) := \mathrm{argmin}_{u \in U}\{z \cdot b(t, x, u) + C(t, x, u)\}$$

Exercise 33. *Consider the stochastic linear quadratic problem below:*

$$\inf_u \mathbb{E}\left[\int_0^T \left(aX_t^2 + bX_t + Au_t^2 + Bu_t\right)\mathrm{d}t + \alpha X_T^2 + \beta X_T\right] \tag{2.7.10}$$

with

$$\mathrm{d}X_t = (\gamma X_t + \kappa u_t)\mathrm{d}t + \sigma\mathrm{d}W_t \tag{2.7.11}$$

where $a, A, \alpha > 0$ and b, B, β, γ and κ are constants. Use both methods discussed in this section to write the BSDE for this control problem.

2.7.5 Beyond Markovian control

A generic BSDE is usually independent of a stochastic control problem and is given by

$$\begin{cases} \mathrm{d}Y_t = -H(t, X_t, Y_t, Z_t)\mathrm{d}t + Z_t dW_t \\ Y_T = \xi \end{cases} \tag{2.7.12}$$

where ξ is a random variable dependent on $\mathcal{F}_T$. If H is a Hamiltonian derived from stochastic control problem, then Y_t and Z_t represent the value function and its gradient.

Dependence of H on X_t can be generalized to dependence of H to the path of X until time t, i.e., $H(t, X_{\cdot \wedge t}, Y_t, Z_t)$ where $X_{\cdot \wedge t} = (X_s : s \leq t)$. Furthermore, we can relax dependence of H on a specific state variable X and write $H(t, \omega, Y_t, Z_t)$, where ω is a random outcome in the sample space. for instance, ω can be the path of Brownian motion.

In this case, we write a non-Markovian (path-dependent) BSDE by

$$\begin{cases} \mathrm{d}Y_t = -H(t, \omega, Y_t, Z_t)\mathrm{d}t + Z_t dW_t \\ Y_T = \xi \end{cases} \tag{2.7.13}$$

For a generic BSDE to have a solution, it is sufficient that

1) $\mathbb{E}[\xi^2] < \infty$

2) for all y and z, $H(t, \omega, y, z)$ is progressively measurable, i.e., measurable with respect to t and ω simultaneously with respect to the information until time t

3) $\mathbb{E}[\int_0^T H^2(t, \omega, 0, 0)\mathrm{d}t] < \infty$

4) $H(t, \omega, y, z)$ is Lipschitz on y and z uniformly in t, ω, i.e.

$$|H(t, \omega, y, z) - H(t, \omega, y', z')| \leq C|y - y'| + |z - z'|)$$

for some constant C independent of (t, ω).

We have the above theorem from [Pardoux, Proposition 1.1].

Theorem 2.7.4 (Existence and uniqueness theorem for BSDEs). *Under the above conditions, there exists a couple (Y_t, Z_t) such that (2.7.13) holds, i.e.,*

$$Y_t = \xi + \int_t^T \left(H(s, \omega, Y_s, Z_s)\mathrm{d}s - Z_s \mathrm{d}W_s \right), \tag{2.7.14}$$

and there exists a constant $C > 0$ that only depends on the Lipschitz constant H such that that

$$\mathbb{E}\left[\sup_{t \in [0,T]} |Y_t|^2 + \int_0^T Z_s^2 \mathrm{d}s \right] \leq C \left(\mathbb{E}[\int_0^T H^2(t, \omega, 0, 0)\mathrm{d}t + \xi^2] \right) \tag{2.7.15}$$

2.7.6 Linear BSDEs

Linear BSDEs take the form

$$\mathrm{d}Y_t = -(\alpha_t Y_t + \beta_t Z_t + L_t)\mathrm{d}t + Z_t \mathrm{d}W_t, \tag{2.7.16}$$

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

$$\mathrm{d}Y_t = -(\alpha_t Y_t + L_t)\mathrm{d}t + Z_t dB_t^\beta, \tag{2.7.17}$$

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

$$\left.\frac{d\mathbb{P}^\beta}{d\mathbb{P}}\right|_{\mathcal{F}_t} := \exp\left(\int_0^t \beta_s \mathrm{d}W_s - \frac{1}{2}\int_0^t \beta_s^2 \mathrm{d}s\right) \tag{2.7.18}$$

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

$$d\tilde{Y}_t = e^{\int_0^t \alpha_s \mathrm{d}s}\left(L_t \mathrm{d}t + Z_t dB_t^\beta\right). \tag{2.7.19}$$

In other words,

$$\tilde{Y}_t = \tilde{Y}_T - \int_t^T e^{\int_0^s \alpha_z dz}\left(L_s \mathrm{d}t + Z_s dB_s^\beta\right), \tag{2.7.20}$$

or,

$$Y_t = e^{\int_t^T \alpha_s \mathrm{d}s}\tilde{Y}_T - \int_t^T e^{\int_t^s \alpha_z dz}\left(L_s \mathrm{d}t + Z_s dB_s^\beta\right) = e^{\int_t^T \alpha_s \mathrm{d}s}\xi - \int_t^T e^{\int_t^s \alpha_z dz}\left(L_s \mathrm{d}t + Z_s dB_s^\beta\right). \tag{2.7.21}$$

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

$$Y_t = \mathbb{E}^\beta\left[e^{\int_t^T \alpha_s \mathrm{d}s}\xi - \int_t^T e^{\int_t^s \alpha_z dz}L_s \mathrm{d}t\,\Big|\,\mathcal{F}_t\right]. \tag{2.7.22}$$

If we define

$$\Gamma_t = e^{\int_t^T \alpha_s \mathrm{d}s}\left.\frac{d\mathbb{P}^\beta}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \exp\left(\int_0^t \beta_s \mathrm{d}W_s + \int_0^t (\alpha_s - \frac{1}{2}\beta_s^2)\mathrm{d}s\right), \tag{2.7.23}$$

by changing the measure back to $\mathbb{P}$

$$Y_t = \Gamma_t^{-1}\mathbb{E}\left[\Gamma_T \xi - \int_t^T \Gamma_s L_s \mathrm{d}t\,\Big|\,\mathcal{F}_t\right] = \mathbb{E}^\beta\left[e^{\int_t^T \alpha_z dz}\xi - \int_t^T e^{\int_t^s \alpha_z dz}L_s \mathrm{d}t\,\Big|\,\mathcal{F}_t\right] \tag{2.7.24}$$

2.8 Some nonstandard control problems and their 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.

2.8.1 Stochastic singular control problems

Consider the stochastic control problem below:

Example 16.

$$\inf_{u_t} \mathbb{E}\left[\int_0^T \left(\frac{1}{2}X_t^2 - u_t\right)\mathrm{d}t + X_T^2 - X_T\right] \tag{2.8.1}$$

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

$$dX_t = (X_t + u_t)dt + \sigma dW_t \tag{2.8.2}$$

The HJB is formally written as

$$
\begin{cases}
V_t + xV_x + \frac{\sigma^2}{2}\partial_x^2 V + \frac{1}{2}x^2 + \inf_{u \geq 0} u(V_x - 1) = 0 \\
V(T, x) = x^2 - x
\end{cases} \tag{2.8.3}
$$

Note that

$$
\inf_{u \geq 0} u(V_x - 1) = \begin{cases}
-\infty & V_x - 1 < 0 \\
0 & V_x - 1 \geq 0
\end{cases} \tag{2.8.4}
$$

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

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

- $\mathbf{R} = \{(t, x) : V_x(t, x) = 1\}$

Inside $\mathbf{N}$, we expect that $V_t + xV_x + \frac{\sigma^2}{2}\partial_x^2 V + \frac{1}{2}x^2 = 0$ 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, $V_x(t, R(t)) = 1$. Therefore, if v is the solution to

$$
\begin{cases}
V_t + xV_x + \frac{\sigma^2}{2}\partial_x^2 v + \frac{1}{2}x^2 = 0 \\
v(T, x) = x^2 - x
\end{cases} \tag{2.8.5}
$$

then, we anticipate to write

$$
V(t, x) = \begin{cases}
v(t, x) & x \geq R(t) \\
x - R(t) + v(t, R(t)) & x < R(t)
\end{cases} \tag{2.8.6}
$$

Solving (2.8.5), 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 34.

$$\inf_{u_t \geq 0} \mathbb{E}\left[\int_0^\tau e^{-rt}\left(\mu - u_t\right)dt\right] \tag{2.8.7}$$

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

$$dX_t = (\gamma X_t - u_t)dt + \sigma dW_t, \quad X_0 = x \geq 0 \tag{2.8.8}$$

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

$$\inf_u \mathbb{E}\Big[\int_0^T \big(C(s, X_s^u) + a u_s\big) \mathrm{d}s + g(X_T^u)\Big] \tag{2.8.9}$$

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

$$\begin{cases} \mathrm{d}X_t^u = (\mu(t, X_t^u) + A u_t)\mathrm{d}t + \sigma(t, X_t^u)\mathrm{d}W_t \\ X_0^u = x \in \mathbb{R}^d \end{cases} \tag{2.8.10}$$

The heuristic derivation of the HJB yields

$$\begin{cases} 0 = V_t(t, x) + \frac{1}{2}a(t, x, u) \cdot D^2 V(t, x) + \nabla V(t, x) \cdot \mu(t, x) + C(t, x) + \inf_u \big\{ au + Au \cdot \nabla V(t, x)\big\} \\ V(T, x) = g(x) \end{cases}$$
$$\tag{2.8.11}$$

$a = \sigma^\mathsf{T}\sigma$. The infimum above is given by

$$\inf_{\hat{u} \in \mathcal{C}, \, |\hat{u}|=1, \lambda \geq 0} \lambda \hat{u} \cdot \big(a + A\nabla V(t, x)\big) = \inf_{\lambda \geq 0} \lambda \inf_{|\hat{u}|=1} \hat{u} \cdot \big(a + A\nabla V(t, x)\big) = \inf_{\lambda \geq 0} \lambda \mathcal{H}(\nabla V(t, x)) \quad \tag{2.8.12}$$

where $\mathcal{H}(\nabla V(t, x)) := \inf_{|\hat{u}|=1} \hat{u} \cdot \big(a + A\nabla V(t, x)\big)$. 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

$$\begin{cases} 0 = V_t(t, x) + \frac{1}{2}a(t, x, u) \cdot D^2 V(t, x) + \nabla V(t, x) \cdot \mu(t, x) + C(t, x) \\ V(T, x) = g(x) \end{cases} \tag{2.8.13}$$

For simplicity of notation, we set

$$\mathcal{L}V(t, x) := V_t(t, x) + \frac{1}{2}a(t, x, u) \cdot D^2 V(t, x) + \nabla V(t, x) \cdot \mu(t, x) + C(t, x) \tag{2.8.14}$$

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 34, the optimal control is the local time of the process, $\mathrm{d}X_t = \gamma X_t \mathrm{d}t + \sigma \mathrm{d}W_t$ at the point $\hat{x}$, existence of which is obtained through solving the HJB and going through verification step.

2.8.2 Optimal stopping problem

Let X be given by the SDE

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t \tag{2.8.15}$$

in a filtered probability space and consider the following problem

$$\sup \mathbb{E}\Big[\int_0^{\tau \wedge T} e^{-rt} C(t, x_t)\mathrm{d}t + g(X_{\tau \wedge T})\Big] \tag{2.8.16}$$

---

[2]For 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.

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

$$V(t,x) = \mathbb{E}_{t,x}\left[\int_0^\eta e^{-r(s-t)}C(t,X_s)\mathrm{d}s + e^{-r(\eta-t)}g(X_\eta)\right] \qquad (2.8.17)$$

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 (2.8.14).

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, are 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 present the most recent methods that can be applied in higher dimensions alongside some general conditions for the numerical PDE schemes to converge to the value function of the control problem.

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 $\{\phi_1, \cdots, \phi_K\}$ and find coefficients $a_1, ..., a_K \in \mathbb{R}^d$ such that

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

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

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

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

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

where

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

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 <u>orthogonal</u> 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^js 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

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

If we work with real data, we may not have the luxury to assume x^js 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 <u>feed-forward neural network</u> 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:

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

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 $f : [a,b]^d \to \mathbb{R}^m$, $d_L = m$. Then matrices $W_{d_{\ell-1} \times d_\ell}$ 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

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

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

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

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

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

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

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

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 were 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 $\quad$ $\theta_{n+1} \leftarrow \theta_n - \alpha_n \nabla \mathcal{J}(\theta_n)$;
3 $\quad$ $n \leftarrow n + 1$
4 **return** $\underline{\theta_n}$

---

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 evaluates 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

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

3.2 Convergence of numerical scheme for PDEs

Numerical schemes for PDEs are studied extensively. Examples of such numerical schemes are finite-difference methods, finite-element methods, and Monte-Carlo methods. In either cases, there are some general conditions that guarantees the convergence of the approximate solution to the solution of the PDE. In the context of control, the notion of solution to the HJB equation is viscosity solution, which is not necessarily a smooth function. Therefore, we study a general framework that guarantees the convergence of approximation schemes to the viscosity solutions for the HJB equations. We note that there are some prerequisites to this framework. First, we need to show that the HJB has a unique viscosity solution. This is studied in several literature including but not limited to Crandall et al. [1992], Fleming and Soner [2006], Touzi [2012]. Second, we need to show that the value function satisfies HJB equation in viscosity sense. This is also a classical result obtained in many literature, e.g., Fleming and Soner [2006], Touzi [2012]. Here we assume that both of these results are established. Consider the HJB equation

$$\begin{cases} 0 = V_t(t,x) + H(t,x,V_x,V_{xx}) \\ V(T,x) = g(x) \end{cases} \tag{3.2.1}$$

$$H(t,x,p,\Gamma) := \inf_{u \in U} \left\{ C(t,x,u) + p \cdot b(t,x,u) + \frac{1}{2}\Gamma \cdot (\sigma^\top \sigma)(t,x,u) \right\} \tag{3.2.2}$$

Assumption 1. *The HJB equation* (3.2.1) *has a unique viscosity solution which is equal to the value function of the control problem*

$$\inf_{u \in \mathcal{U}} \left\{ \int_0^T C(t,X_t,u_t)\mathrm{d}t + g(X_T) \right\}$$
$$\mathrm{d}X_t = b(t,X_t,u_t)\mathrm{d}t + \sigma(t,X_t,u_t)\mathrm{d}W_t$$

Note that if we set $\sigma \equiv 0$, we have a deterministic control problem. We consider a numerical method for an HJB equation as an equation of discrete points. For instance, in the finite-difference method, the evaluations are performed on a grid on the (t,x) variables.

Example 17. *In the simplest finite-difference method in dimension* 1, *explicit centered finite-difference method, we work on the grid* $\mathcal{G}_h = \{(t_i,x_j) : t_i = ih, \ x_j = j\sqrt{h}, \ i = 0,...,N-1, \ j = -J,...,J\}$, *where* $h = \frac{T}{N}$. *The approximate partial derivatives are given by*

$$V_t(t_i,x_j) \approx \frac{V(t_{i+1},x_j) - V(t_i,x_j)}{h}$$
$$V_x(t_i,x_j) \approx \frac{V(t_i,x_{j+1}) - V(t_i,x_{j-1})}{2h}$$
$$V_{xx}(t_i,x_j) \approx \frac{V(t_i,x_{j+1}) + V(t_i,x_{j-1}) - 2V(t_i,x_j)}{h^2}$$

Then, we find $V^h(t_i,x_i)$ *such that*

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

3.3 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.

3.3.1 Loss function for control problems and functional optimization

The goal of approximation of $f : [a, b]^d \to \mathbb{R}^m$ 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

$$\mathcal{J}(\phi) := \mathbb{E}\left[\int_0^T C(t, X_t, \phi(t, X_t))\mathrm{d}t + g(X_T)\right]. \tag{3.3.1}$$

where

$$\mathrm{d}X_t = b(t, X_t, \phi(t, X_t))\mathrm{d}t + \sigma(t, X_t, \phi(t, X_t))\mathrm{d}W_t$$

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.

3.3.2 Parametrization of functions

In the loss function (3.3.1), we replace ϕ with a parametrized function, such as a neural network.

$$\mathcal{J}(\phi(\cdot, \theta)) := \mathbb{E}\left[\int_0^T C(t, X_t, \phi(t, X_t; \theta))\mathrm{d}t + g(X_T)\right]. \tag{3.3.2}$$

where

$$\mathrm{d}X_t = b(t, X_t, \phi(t, X_t; \theta))\mathrm{d}t + \sigma(t, X_t, \phi(t, X_t; \theta))\mathrm{d}W_t$$

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.

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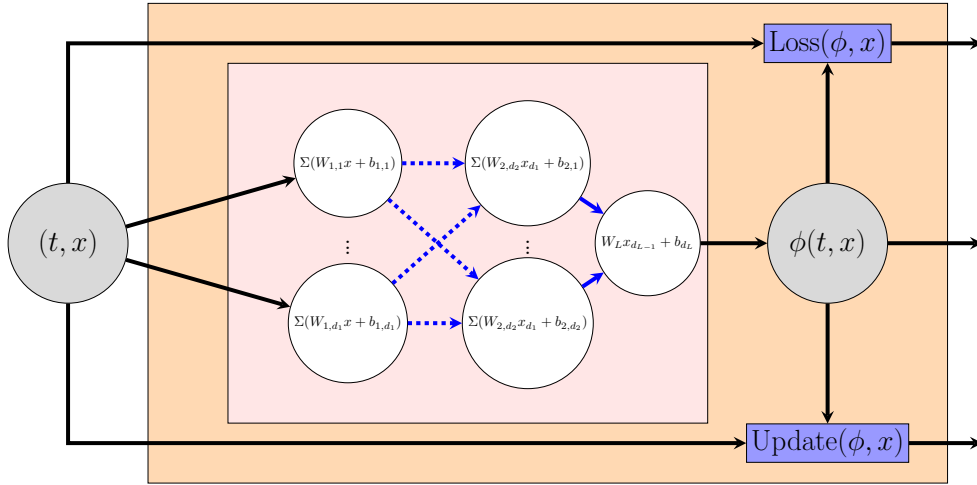
3.3.3 Discretization

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

$$\hat{\mathcal{J}}(\phi(\cdot, \theta)) := \mathbb{E}\left[\sum_{i=0}^{N-1} C(t_i, \hat{X}_{t_i}, \phi(t_i, \hat{X}_{t_i}; \theta))\Delta t + g(\hat{X}_T)\right]. \tag{3.3.3}$$

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

Figure 3.3.1: A design of a unit for solving control problems with deep learning. The neural network approximating control in 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.

3.3.4 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.

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

$$\hat{X}_{t_{n+1}}^j = \hat{X}_{t_n}^j + b(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta))\Delta t + \sigma(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta))\Delta B_{t_{n+1}}^j \tag{3.3.4}$$

Then, the empirical loss is given by

$$\overline{\mathcal{J}}(\phi(\cdot, \theta)) := \frac{1}{J} \sum_{j=1}^{J} \left[\sum_{n=0}^{N-1} C(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta))\Delta t + g(\hat{X}_T^j) \right]. \tag{3.3.5}$$

The gradient of $\bar{\mathcal{J}}$ with respect to θ can be evaluated by using chain rule.

3.3.5 Design of PGM

A key part of the policy gradient method is the evaluation in lines 4, 5, and 6 in Algorithm 4. After θ_i and consequently $\phi(t, x; \theta_i)$ are updated, in line 4, we re-evaluate the sample paths in (3.3.4) based on θ_i. We call this step $\underline{\text{Update}(\phi, x)}$ in Figure 3.3.1. In line 5, we evaluate the contribution of the control $\phi(t, x; \theta_i)$ to the running cost in the loss function by evaluating

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

Algorithm 5: Gradient descent algorithm for neural networks

Data: $i = 0$, $\phi(t, x; \theta)$ is constructed and $\theta = \theta_i$ 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, $n = 0, ..., N - 1$ & $j = 1, ..., J$.

Parameter: Learning rate $\{\alpha_i\}_i$ and number of iterations I

1 **while** $i \leq I$ **do**

2    $\overline{\mathcal{J}}(\phi(\cdot; \theta_i) = 0$;

3   **for** $n = 0, ..., N - 1$ & $j = 1, ..., J$ **do**

4   (3.3.4) is used to create J sample paths of $\hat{X}$.;

5       $\overline{\mathcal{J}}(\phi(\cdot; \theta_i) \leftarrow \overline{\mathcal{J}}(\phi(\cdot; \theta_i) + \frac{1}{J} \sum_{j=1}^{J} C(t_n, \hat{X}_{t_n}^j, \phi(t_n, \hat{X}_{t_n}^j; \theta_i)) \Delta t$;

6       $\overline{\mathcal{J}}(\phi(\cdot; \theta_i) \leftarrow \overline{\mathcal{J}}(\phi(\cdot; \theta_i) + \frac{1}{J} \sum_{j=1}^{J} g(\hat{X}_T^j)$

7   The gradient of empirical loss, $\nabla_\theta \bar{\mathcal{J}}(\theta_i)$, is evaluated.;

8    $\theta_{i+1} \leftarrow \theta_i - \alpha_i \nabla_\theta \overline{\mathcal{J}}(\theta_i)$;

9    $i \leftarrow i + 1$

10 **return** θ_i

We call this step $\underline{\text{Loss}(\phi, x)}$, as shown in Figure 3.3.1. In line 6, when $t_i = T$, we add the contribution is $g(\hat{X}_T^j)$ to the total loss. Next steps of the algorithm are computing $\nabla_\theta \overline{\mathcal{J}}(\theta_i)$ and adjusting the value of $theta_i$ accordingly.

It is important to note that deep learning packages such as `torch` or `tensorflow`, construct $\nabla_\theta \overline{\mathcal{J}}(\theta_i)$ using chain rule only once and update it at each iteration of the algorithm.

Exercise 35. *Solve the following linear quadratic problem via PGM and compare the PGM to the closed-form solution.*

$$\inf_u \mathbb{E}\left[\int_0^1 \left(X_t^2 + X_t + u_t^2\right) \mathrm{d}t + X_1^2 - X_1\right]$$

with

$$dX_t = (X_t - u_t)\mathrm{d}t + \mathrm{d}W_t$$

Recall that the closed-form solution is obtained by writing the HJB equation, similar to Example 10, and verifying that the value function takes the form

$$V(t, x) = f(t)x^2 + h(t)x + \ell(t)$$

In this process we find ODEs for the coefficients f, h, and ℓ which we need to solve.

3.4 Feynman-Kac approximation of linear equations

While not related to control, it is important to draw an analogy between the BSDE approximation of value function in a control problem and the Feynman-Kac representation of the value of linear parabolic equations. Since the value function of a control problem satisfies an HJB equation, the nonlinear version of a parabolic equation, BSDE serves as a generalization of Feynman-Kac formula to nonlinear problems.

One can use the expected value in (2.7.3) to approximate the solution to (2.7.4). We generate the samples of X by the $\underline{\text{Euler-Maruyama}}$ discretization. Set $\Delta t = T/N$ and $t_n = n\Delta t$ and generate J samples of

increments of Wiener process $\Delta W^j_{t_{n+1}} = W^j_{t_{n+1}} - W^j_{t_n} \sim N(0, \Delta t)$

$$\hat{X}^j_{t_{n+1}} = \hat{X}^j_{t_n} + b(t_n, \hat{X}^j_{t_n})\Delta t + \sigma(t_n, \hat{X}^j_{t_n})\Delta W^j_{t_{n+1}} \tag{3.4.1}$$

Then, we can approximate $V(0, x)$ by

$$\mathbb{E}\left[\int_0^T C(r, X_r)\mathrm{d}r + g(X_T)\right] \approx \frac{1}{J}\left(\sum_{n=0}^{N-1} C(t_n, \hat{X}^j_{t_n})\Delta t + g(\hat{X}^j_T)\right)$$

The initial value of X^j_0 can be taken as a given point x or be generated according to an arbitrary distribution,

---

Algorithm 6: Feynman-Kac approximation for linear parabolic equations

---

Data: T, N, $X^j_0 \sim \mathcal{L}_0$, $W^j_{t_{n+1}} \sim N(0, \Delta t)$, $\phi(\cdot, \theta)$

1 $\Delta t \leftarrow T/N$;

2 **for** $n = 0, ..., N-1, j = 0, ..., J$ **do**

3 $\quad$ Evaluate $\hat{X}^j_{t_{n+1}}$ using (3.4.1);

4 $\hat{Y}^j := \sum_{n=0}^{N-1} C(t_n, \hat{X}^j_{t_n})\Delta t + g(\hat{X}^j_T)$;

5 Mean-square minimization: $\theta^* = \mathrm{argmin}\frac{1}{J}\left(\phi(X^j_0; \theta) - \hat{Y}^j\right)^2$;

6 **return** $\underline{\phi(x; \theta^*)}$

$\mathcal{L}_0$. If X^j_0 is generated by a distribution, in Algorithm 6, we need approximate $V(0, x)$ by a parametrized function $\phi(\cdot; \theta)$ by minimizing a mean-square error. This minimization is performed in Step 5 in Algorithm 6.

Algorithm 6 only approximates the value function. If we write the BSDE corresponding to (2.7.4), we can approximate both the value function and its gradient simultaneously.

3.5 Numerical methods based on BSDE

Consider the HJB

$$\begin{cases} V_t(t, x) + \frac{1}{2}(\sigma^\mathsf{T}\sigma)(t, x) \cdot D^2V(t, x) + H(t, x, \nabla V(t, x)) = 0 \\ V(T, x) = g(X_T) \end{cases} \tag{3.5.1}$$

and the corresponding BSDE

$$\begin{cases} \mathrm{d}Y_t = -H(t, X_t, Z_t)\mathrm{d}t + Z_t \cdot \sigma(t, X_t)\mathrm{d}W_t \\ Y_T = g(X_T) \end{cases} \tag{3.5.2}$$

By Theorem (2.7.2), we know that $V(t, X_t) = Y_t$ and $\nabla V(t, X_t) = Z_t\sigma^{-1}(t, X_t)$, where $\mathrm{d}X_t = \sigma(t, X_t)\mathrm{d}W_t$. Consider the discretization of X given by the Euler-Maruyama scheme:

$$\hat{X}_{t_{n+1}} = \hat{X}_{t_n} + \sigma(t_n, \hat{X}_{t_n})\Delta W_{t_{n+1}}$$

where $\Delta W_{t_{n+1}} = W_{t_{n+1}} - W_{t_n}$. We also discretize the BSDE:

$$\begin{cases} \hat{Y}_{t_n} = \hat{Y}_{t_{n+1}} + H(t_n, \hat{X}_{t_n}, \hat{Z}_{t_n})\Delta t - \hat{Z}_{t_n} \cdot \sigma(t_n, \hat{X}_{t_n})\Delta W_{t_{n+1}} \\ \hat{Y}_T = g(\hat{X}_T) \end{cases} \tag{3.5.3}$$

In (3.5.3), if we take conditional expectation given $\hat{X}_{t_n}$, we obtain

$$\mathbb{E}[\hat{Z}_{t_n} \cdot \sigma(t_n, \hat{X}_{t_n})\Delta W_{t_{n+1}}|\hat{X}_{t_n}] = \hat{Z}_{t_n} \cdot \sigma(t_n, \hat{X}_{t_n})\mathbb{E}[\Delta W_{t_{n+1}}|\hat{X}_{t_n}] = 0$$

and

$$\hat{Y}_{t_n} = \mathbb{E}[\hat{Y}_{t_{n+1}}|\hat{X}_{t_n}] + H(t_n, \hat{X}_{t_n}, \hat{Z}_{t_n})\Delta t \tag{3.5.4}$$

Furthermore, if we multiply both sides on (3.5.3) by $\Delta W_{t_{n+1}}$ and take conditional expectation given $\hat{X}_{t_n}$, we have

$$\begin{aligned} \hat{Y}_{t_n}\mathbb{E}[\Delta W_{t_{n+1}}|\hat{X}_{t_n}] =& \mathbb{E}[\hat{Y}_{t_{n+1}}\Delta W_{t_{n+1}}|\hat{X}_{t_n}] + H(t_n, \hat{X}_{t_n}, \hat{Z}_{t_n})\mathbb{E}[\Delta W_{t_{n+1}}|\hat{X}_{t_n}]\Delta t \\ &- \hat{Z}_{t_n}\sigma^{-1}(t, \hat{X}_t)\mathbb{E}[\Delta W_{t_{n+1}}^{\intercal}\Delta W_{t_{n+1}}|\hat{X}_{t_n}] \end{aligned} \tag{3.5.5}$$

In the above, $\Delta W_{t_{n+1}}^{\intercal}\Delta W_{t_{n+1}}$ is a $d \times d$ matrix obtained from multiplying a column vector to a row vector.

$$\Delta W_{t_{n+1}}^{\intercal}\Delta W_{t_{n+1}} = \begin{bmatrix} \Delta W_{1,t_{n+1}}^2 & \Delta W_{1,t_{n+1}}\Delta W_{2,t_{n+1}} & \cdots & \Delta W_{1,t_{n+1}}\Delta W_{d,t_{n+1}} \\ & & \vdots & \\ \Delta W_{d,t_{n+1}}\Delta W_{1,t_{n+1}} & \Delta W_{d,t_{n+1}}\Delta W_{2,t_{n+1}} & \cdots & \Delta W_{d,t_{n+1}}^2 \end{bmatrix}$$

Since $\mathbb{E}[\Delta W_{t_{n+1}}|\hat{X}_{t_n}] = 0$, we have

$$0 = \mathbb{E}[\hat{Y}_{t_{n+1}}\Delta W_{t_{n+1}}|\hat{X}_{t_n}] - \hat{Z}_{t_n}\sigma(t_n, \hat{X}_{t_n})\mathbb{E}[\Delta W_{t_{n+1}}^{\intercal}\Delta W_{t_{n+1}}|\hat{X}_{t_n}] \tag{3.5.6}$$

Note that

$$\mathbb{E}[\Delta W_{i,t_{n+1}}\Delta W_{j,t_{n+1}}|\hat{X}_{t_n}] = \begin{cases} 0 & i \neq j \\ \Delta t & i = j \end{cases}$$

Then,

$$\hat{Z}_{t_n} = \frac{1}{\Delta t}\sigma^{-1}(t_n, \hat{X}_{t_n})\mathbb{E}[\hat{Y}_{t_{n+1}}\Delta W_{t_{n+1}}|\hat{X}_{t_n}] \tag{3.5.7}$$

The numerical scheme for BSDE is given by

$$\begin{cases} \hat{X}_{t_{n+1}} = \hat{X}_{t_n} + \sigma(t_n, \hat{X}_{t_n})\Delta W_{t_{n+1}} \\ \hat{Z}_{t_n} = \frac{1}{\Delta t}\sigma^{-1}(t_n, \hat{X}_{t_n})\mathbb{E}[\hat{Y}_{t_{n+1}}\Delta W_{t_{n+1}}|\hat{X}_{t_n}] \\ \hat{Y}_{t_n} = \mathbb{E}[\hat{Y}_{t_{n+1}}|\hat{X}_{t_n}] + H(t_n, \hat{X}_{t_n}, \hat{Z}_{t_n})\Delta t \\ \hat{Y}_T = g(\hat{X}_T) \end{cases} \tag{3.5.8}$$

3.6 Deep BSDE

One of the most recent advancement in the numerical methods based on BSDE was achieved by Han et al. [2018], where they introduced a method based on deep neural networks to solve the BSDEs. Consider the

BSDE

$$\begin{cases} dY_t = -H(t, X_t, Z_t)dt + Z_t \cdot \sigma(t, X_t)dW_t \\ Y_T = g(X_T) \end{cases} \tag{3.6.1}$$

Given two functions $\psi(t, x)$ and $phi(x)$, we can write the BSDE as a forward process

$$\begin{cases} d\tilde{Y}_t^{\phi,\psi} = -H(t, X_t, \psi(t, X_t))dt + \psi(t, X_t) \cdot \sigma(t, X_t)dW_t \\ \tilde{Y}_0^{\phi,\psi} = \phi(X_0) \end{cases} \tag{3.6.2}$$

If $\psi(t, x) = \nabla V(t, x)$ and $Y_0 = V(0, x)$, then we must have $Y = \tilde{Y}^{\phi,\psi}$ and $Z_t = \nabla V(t, X_t) = \psi(t, X_t)$ and the terminal condition is satisfied $\tilde{Y}_T^{\phi,\psi} = V(T, X_T) = g(X_T)$. Therefore, solving the BSDE is equivalent to finding the minimizer of

$$\inf_{\phi(x),\psi(t,x)} \mathbb{E}\left[\left(\tilde{Y}_T^{\phi,\psi} - g(X_T)\right)^2\right]$$

To find the minimizer approximately, we discretize

$$\hat{X}_{t_{n+1}} = \hat{X}_{t_n} + \sigma(t_n, \hat{X}_{t_n})\Delta W_{t_{n+1}}$$

and

$$\begin{cases} \hat{Y}_{t_{n+1}}^{\phi,\psi} = \hat{Y}_{t_n}^{\phi,\psi} - H(t_n, X_{t_n}, \psi(t_n, X_{t_n}))\Delta t + \psi(t_n, X_{t_n}) \cdot \sigma(t_n, X_{t_n})(W_{t_{n+1}} - W_{t_n}) \\ \hat{Y}_0^{\phi,\psi} = \phi(X_0) \end{cases} \tag{3.6.3}$$

and minimize

$$\inf_{\phi(x),\psi(t,x)} \mathbb{E}\left[\left(\hat{Y}_T^{\phi,\psi} - g(\hat{X}_T)\right)^2\right]$$

The above minimization problem is over an infinite dimensional space of functions $(\phi(x), \psi(t, x))$. To make it practical, we reduce it to an optimization problem over a parametrized set of functions:

$$\inf_{\theta} \mathbb{E}\left[\left(\hat{Y}_T^{\theta} - g(\hat{X}_T)\right)^2\right]$$

where

$$\begin{cases} \hat{Y}_{t_{n+1}}^{\theta} = \hat{Y}_{t_n}^{\theta} - H(t_n, X_{t_{n+1}}, \psi(t_n, X_{t_{n+1}}; \theta))\Delta t + \psi(t_n, X_{t_{n+1}}\theta) \cdot \sigma(t_n, X_{t_{n+1}})(W_{t_{n+1}} - W_{t_n}) \\ \hat{Y}_0^{\theta} = \phi(X_0; \theta) \end{cases}$$

$$\tag{3.6.4}$$

Furthermore, we evaluate it empirically by generating J sample paths of $\hat{X}$:

$$\inf_{\theta} \sum_{i=1}^{J}\left[\left(\hat{Y}_T^{j,\theta} - g(\hat{X}_T^j)\right)^2\right]$$

where

$$\begin{cases} \hat{Y}_{t_{n+1}}^{j,\theta} = \hat{Y}_{t_n}^{j,\theta} - H(t_n, X_{t_{n+1}}, \psi(t_n, X_{t_{n+1}}^j; \theta))\Delta t + \psi(t_n, X_{t_{n+1}}^j\theta) \cdot \sigma(t_n, X_{t_{n+1}}^j)(W_{t_{n+1}}^j - W_{t_n}^j) \\ \hat{Y}_0^{j,\theta} = \phi(X_0^j; \theta) \end{cases}$$

$$\tag{3.6.5}$$

Exercise 36. *Solve modified version of Exercise 10 with modification*

$$dX_t = (cX_t + du_t)dt + (eX_t + fu_t)dW_t \tag{3.6.6}$$

Exercise 37. *Extend the stochastic linear quadratic problem in Exercise 36 to higher dimension and write the HJB.*

Example 18 (Merton optimal investment problem). *Remember a self-financing portfolio with a Black-Scholes risky asset,* $dS_t = S_t(\mu dt + \sigma dW_t)$, *under zero interest rate satisfies*

$$dX_t = \theta_t(\mu dt + \sigma dW_t) \tag{3.6.7}$$

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:

$$\sup_\theta \mathbb{E}[U(X_T^\theta)] \tag{3.6.8}$$

The HJB for this problem is given by

$$\begin{cases} 0 = V_t + \sup_\theta \left\{ \frac{\sigma^2}{2} V_{xx} + \mu V_x \right\} \\ V(T, x) = U(x) \end{cases} \tag{3.6.9}$$

After simplification, we have

$$\begin{cases} 0 = V_t - \frac{\mu^2 (V_x)^2}{2\sigma^2 V_{xx}} \\ V(T, x) = U(x) \end{cases} \tag{3.6.10}$$

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 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.*

$$dX_t = \theta_t(\mu dt + \sigma dW_t) - c_t dt \tag{3.6.11}$$

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

$$\sup_{c \geq 0, \, \theta} \mathbb{E}\left[\int_0^\infty e^{-\gamma t} U(c_t) dt \right], \quad \gamma > 0 \tag{3.6.12}$$

Write the HJB for this problem.

Remark 3.6.1. $\gamma > 0$ *in the above example represents preference of current consumption over future consumption.*

Remark 3.6.2. *If the control problem is with infimum instead of supremum,*

$$V(t, x) = \inf_{u \in \mathcal{A}_{t,T}} \mathbb{E}\left[\int_t^T C(s, X_s^{t,x,u}, u_s) dt + g(X_T^{t,x,u}) \right], \tag{3.6.13}$$

Theorem 2.5.2 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.

.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:

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

Define the Lagrangian:

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

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

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

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

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

is called the <u>dual problem</u> for the <u>primal problem</u> (.1.1).

The key to the success of the Lagrange multiplier method is the <u>strong duality</u>.

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

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

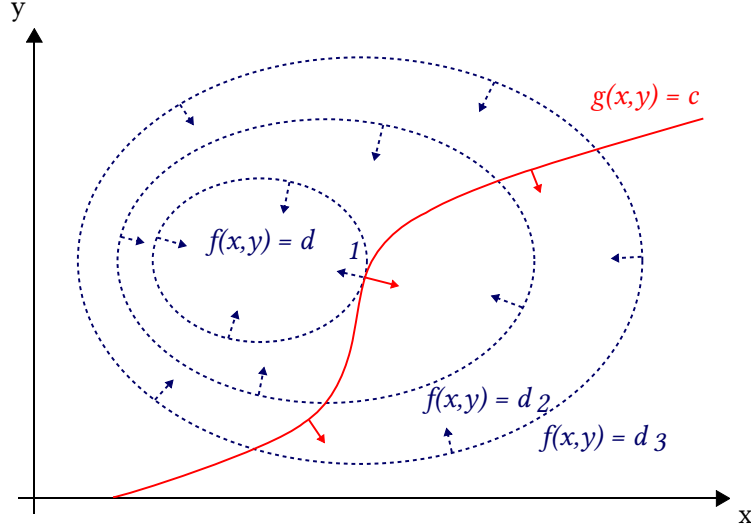$$\sup_\lambda L(x, \lambda) = \sup_\lambda \{ f(x) - \lambda \cdot g(x) \} = \infty \tag{.1.6}$$

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

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

It is not always easy to check if strong duality holds. However, <u>Karush-Kuhn-Tucker</u> 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*

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

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

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 qualities, i.e.,

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

the range of dual variable in Hamiltonian changes:

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

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

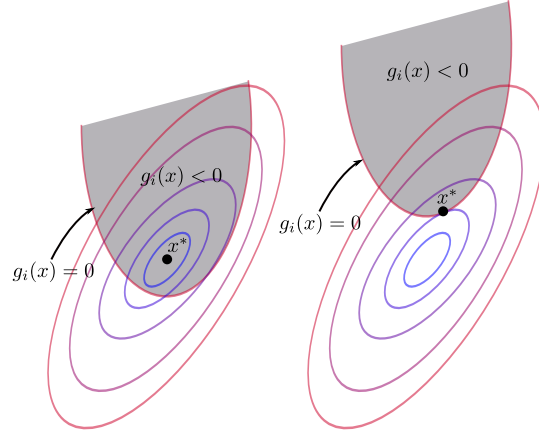The reason for such modification can formally be seen after switching $\inf_x$ and $\sup_{\mu \geq 0}$ in the above

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

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:

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

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

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:

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

subject to $x_T = 0$. The Lagrangian is

$$\int_0^T \left(x_t^2 - \alpha_t x_t \right) dt + \lambda x_T \tag{.1.15}$$

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.

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

KKT condition becomes $x_t^* = (\alpha_t + \lambda^*)/2$ and $\int_0^T x_t^* = 0$. 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 $f : \mathbb{R}^d \to \mathbb{R}$ 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 <u>gradient descent</u>, 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 $\alpha > 0$ $x - \alpha\nabla(x)$ is closer to a local minimum than x was. The most important parameter in GD is α, <u>learning rate</u>. Ideally, when we are further from a local

---

Algorithm 8: 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 $\quad$ $x \leftarrow x - \alpha_n\nabla f(x)$;
3 $\quad$ $n \leftarrow n + 1$

4 **return** $\underline{x}$

---

minimum, 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 $\alpha_n \to 0$ 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 jutification 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 $\int_0^\infty \alpha_t \mathrm{d}t = \infty$ 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 $\mathrm{d}x_t = -\alpha_t \nabla f(x_t)\mathrm{d}t$ converges to x_*.

Proof. We use chain rule to evaluate

$$\mathrm{d}|x_t - x_*|^2 = 2(x_t - x_*) \cdot \mathrm{d}x_t = -2\alpha(x_t - x_*) \cdot \nabla f(x_t)\mathrm{d}t$$

By the assumption of the theorem, we have

$$\mathrm{d}|x_t - x_*|^2 = -2\alpha_t(x_t - x_*) \cdot \nabla f(x_t)\mathrm{d}t \leq -2\lambda\alpha_t|x_t - x_*|^2\mathrm{d}t$$

or

$$\mathrm{d}|x_t - x_*|^2 + 2\lambda\alpha_t|x_t - x_*|^2\mathrm{d}t \leq 0$$

By multiplying the above by $\exp(2\lambda\int_0^t \alpha_s\mathrm{d}s)$ and integrating, we obtain

$$\exp(2\lambda \int_0^t \alpha_s\mathrm{d}s)|x_t - x_*|^2 \leq |x_0 - x_*|^2$$

or

$$|x_t - x_*|^2 \leq |x_0 - x_*|^2 \exp(-2\lambda \int_0^t \alpha_s \mathrm{d}s)$$

The right-hand above goes to zero as $t \to \infty$. $\hfill \square$

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, in used. SGD randomly chooses small number of directions from $\{1, ..., d\}$ e.g. 10 out of $1e^6$ 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.

Patrick Cheridito, H. Mete Soner, Nizar Touzi, and Nicolas Victoir. Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs. Comm. Pure Appl. Math., 60(7):1081–1110, 2007. ISSN 0010-3640. doi: 10.1002/cpa.20168. URL `https://doi.org/10.1002/cpa.20168`.

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.

Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. Proceedings of the National Academy of Sciences, 115(34):8505–8510, 2018. doi: 10.1073/pnas.1718942115. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1718942115`.

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`.

Étienne Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic pdes of second order. In Stochastic Analysis and Related Topics VI: Proceedings of the Sixth Oslo—Silivri Workshop Geilo 1996, pages 79–127. Springer.

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`.

H. Mete Soner, Nizar Touzi, and Jianfeng Zhang. Wellposedness of second order backward SDEs. Probab. Theory Related Fields, 153(1-2):149–190, 2012. ISSN 0178-8051,1432-2064. doi: 10.1007/s00440-011-0342-y. URL `https://doi.org/10.1007/s00440-011-0342-y`.

Nizar Touzi. Optimal stochastic control, stochastic target problems, and backward SDE, volume 29. Springer, 2012.