Notes on Stochastic Control

Haosheng Zhou

Dec, 2022

The following contents about HJE & HJB refers to the Evans book.

Hamilton-Jacobi Equation (HJE)

The Hamilton-Jacobi equation is a non-linear first-order PDE with the form

$$\begin{cases} u_t + H(Du) = 0 & in\ \mathbb{R}^n \times (0,\infty) \\ u = g & on\ \mathbb{R}^n \times \{t=0\} \end{cases} \tag{1}$$

where $u = u(x,t) : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}$ is the function to solve out and $Du = (u_{x_1}, ..., u_{x_n})$ is the gradient of u w.r.t. space variable x. Here the **Hamiltonian** $H : \mathbb{R}^n \to \mathbb{R}$ is given and the initial condition $g : \mathbb{R}^n \to \mathbb{R}$ is given.

Connection with Hamilton's Equations

Let's first apply the method of characteristics to get some intuition by noticing that this equation is a first-order equation. We know that the method of characteristics does not necessarily hold in general (since it requires the existence of C^2 solution), but this may tell us how to proceed. In this section, we assume that HJE looks like

$$\begin{cases} u_t + H(Du, x) = 0 & in\ \mathbb{R}^n \times (0,\infty) \\ u = g & on\ \mathbb{R}^n \times \{t=0\} \end{cases} \tag{2}$$

where the Hamiltonian also depends on x.

Notice that here we **merge the time variable t with the space variable x and denote it as $x \in \mathbb{R}^{n+1}$, where $x^1, ..., x^n$ are components of x and x^{n+1} denotes the time**. Define

$$z(s) = u(x(s)) \tag{3}$$

as the version of u along the characteristic curve and

$$p(s) = Du(x(s)) \in \mathbb{R}^{n+1} \tag{4}$$

as the version of Du along the characteristic curve, note that here $p^1, ..., p^n$ are partial derivatives w.r.t. x-components and p_{n+1} is the partial derivative w.r.t. t. One would always set the characteristic direction to be

$$x'(s) = D_p F \tag{5}$$

where the original PDE can be written as $F(Du, u, x) = 0$ and in this case

$$F(p, z, y) = p^{n+1} + H(p^1, ..., p^n, x^1, ..., x^n) \tag{6}$$

As a result, one get the ODE system from the method of characteristics

$$\begin{cases} [x^i(s)]' = H_{p_i}(p^1, ..., p^n, x^1, ..., x^n) \ (i = 1, 2, ..., n) \\ [x^{n+1}(s)]' = 1 \end{cases} \tag{7}$$

so one can identify $x^{n+1}(s)$ as s, meaning that the parameter s is the same as the time variable $t = x^{n+1}$. The equation for $z(s)$ is $z'(s) = D_p F \cdot p(s)$, so

$$z'(s) = \sum_{i=1}^{n} H_{p_i}(p^1, ..., p^n, x^1, ..., x^n) \cdot p^i(s) + p^{n+1}(s) \tag{8}$$

$$= \sum_{i=1}^{n} H_{p_i}(p^1, ..., p^n, x^1, ..., x^n) \cdot p^i(s) - H(p^1, ..., p^n, x^1, ..., x^n) \tag{9}$$

The equation for $p(s)$ is $p'(s) = -D_x F - D_z F \cdot p(s)$, so

$$\begin{cases} [p^i(s)]' = -H_{x_i}(p^1, ..., p^n, x^1, ..., x^n) \ (i = 1, 2, ..., n) \\ [p^{n+1}(s)]' = 0 \end{cases} \tag{10}$$

with the last equation $[p^{n+1}(s)]' = 0$ as the redundant one since x^{n+1} has already been parameterized as s.

By cancelling all redundant equations and reorganizing the variables, we get the **characteristic ODE system** for HJE

$$\begin{cases} x'(s) = D_p H(p(s), x(s)) \\ z'(s) = D_p H(p(s), x(s)) \cdot p(s) - H(p(s), x(s)) \\ p'(s) = -D_x H(p(s), x(s)) \end{cases} \tag{11}$$

where $p(s) = (p^1(s), ..., p^n(s))$ and $x(s) = (x^1(s), ..., x^n(s))$ (the last component in $x(s), p(s)$ is ignored). **The Hamilton's equation** is defined as the system consisting of the first and third equation, i.e.

$$\begin{cases} x'(s) = D_p H(p(s), x(s)) \\ p'(s) = -D_x H(p(s), x(s)) \end{cases} \tag{12}$$

Remark. *The reason that we only take the equations w.r.t $x(s)$ and $p(s)$ in the Hamilton's equations is that those two equations have nothing to do with z, they already have $2n$ unknowns and $2n$ equations. In other words, the equation w.r.t. $z(s)$ does not provide any effective information for the derivation of $x(s), p(s)$, and after solving out $x(s), p(s)$, one can immediately know $z(s)$.*

A Problem in the Calculus of Variation

The connection between HJE and Hamilton's equations can also be shown in another perspective by considering a problem in the calculus of variation. The problem is formed as finding a best curve in an admissible class. The **admissible class** is defined as

$$\mathscr{A} = \left\{ w \in C^2, w : [0,t] \to \mathbb{R}^n : w(0) = y, w(t) = x \right\} \tag{13}$$

so any admissible curve is a C^2 path in $\mathbb{R}^n$ such that it starts from point y and ends at point x with $x, y \in \mathbb{R}^n, t > 0$ given. Imagine $w(s) \in \mathscr{A}$ as the moving trajectory of a particle, then $w'(s)$ is actually the speed of the particle at each time. The **action functional** is then defined as

$$I[w] = \int_0^t L(w'(s), w(s))\, ds \tag{14}$$

where $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a given smooth function called **Lagrangian** and we hope to find a curve $x(s) \in \mathscr{A}$ such that the action functional is minimized

$$I[x] = \inf_{w(s) \in \mathscr{A}} I[w] \tag{15}$$

Remark. *The Lagrangian has the meaning as the kinetic energy minus the potential energy in physics which has the meaning of "increments of distance". Here among all possible and smooth enough curves between two fixed points, we want to find $x(s)$ such that it minimizes the integral of the Lagrangian along the path, equivalent to saying that the optimal path is the one that* ***takes the "shortest" path****. If one still finds it hard to understand, think about how light travels, it always travels in the path such that the distance it goes through is the shortest, a natural minimization of a "trivial" action functional.*

Let's **assume** that the Lagrangian is given by $L = L(v, x)$ $(v, x \in \mathbb{R}^n)$ for the convenience of notations and that **the inf of $I[w]$ can be achieved by some $x(s) \in \mathscr{A}$ as the optimal path**. To build up a PDE for $x(s)$, choose smooth $y : [0,t] \to \mathbb{R}^n$ with $y(s) = (y^1(s), ..., y^n(s))$ such that $y(0) = y(t) = 0$ and consider perturbing the optimal path $x(s)$ by a small multiple of $y(s)$ to get

$$w(s) = x(s) + \tau y(s) \ (\tau \in \mathbb{R}) \tag{16}$$

since $w(s) \in \mathscr{A}$, one immediately sees that

$$I[w] \geq I[x] \tag{17}$$

Consider the action functional of the perturbed path

$$i : \mathbb{R} \to \mathbb{R}, i(\tau) = I[x + \tau y] \tag{18}$$

it's easy to see that it has minimum at $\tau = 0$ (assume it's differentiable) with

$$i'(\tau) = \frac{d}{d\tau}\int_0^t L(x'(s) + \tau y'(s), x(s) + \tau y(s))\, ds \tag{19}$$

$$= \int_0^t y'(s) \cdot L_v(x'(s) + \tau y'(s), x(s) + \tau y(s)) + y(s) \cdot L_x(x'(s) + \tau y'(s), x(s) + \tau y(s))\, ds \tag{20}$$

so

$$i'(0) = \int_0^t y'(s) \cdot L_v(x'(s), x(s)) + y(s) \cdot L_x(x'(s), x(s))\, ds \tag{21}$$

$$= \int_0^t \sum_{i=1}^n \left([y^i(s)]' \cdot L_{v_i}(x'(s), x(s)) + y^i(s) \cdot L_{x_i}(x'(s), x(s))\right)\, ds \tag{22}$$

$$= 0 \tag{23}$$

Do transformations to this integral to find

$$\sum_{i=1}^n \int_0^t \left([y^i(s)]' \cdot L_{v_i}(x'(s), x(s)) + y^i(s) \cdot L_{x_i}(x'(s), x(s))\right)\, ds \tag{24}$$

$$= \sum_{i=1}^n \int_0^t L_{v_i}(x'(s), x(s))\, dy^i(s) + \int_0^t L_{x_i}(x'(s), x(s)) \cdot y^i(s)\, ds \tag{25}$$

$$= \sum_{i=1}^n -\int_0^t y^i(s)\, dL_{v_i}(x'(s), x(s)) + \int_0^t L_{x_i}(x'(s), x(s)) \cdot y^i(s)\, ds \tag{26}$$

$$= \sum_{i=1}^n \int_0^t \left[-\frac{d}{ds}L_{v_i}(x'(s), x(s)) + L_{x_i}(x'(s), x(s))\right] y^i(s)\, ds \tag{27}$$

$$= 0 \tag{28}$$

which is valid for any smooth y such that $y(0) = y(t) = 0$. By density argument,

$$\forall i = 1, 2, ..., n, \forall s \in [0, t], -\frac{d}{ds}L_{v_i}(x'(s), x(s)) + L_{x_i}(x'(s), x(s)) = 0 \tag{29}$$

Theorem 1. ***(Euler-Lagrange Equation)*** *If path $x(s)$ is the optimal path and solves the variational problem mentioned above, then it must satisfy Euler-Lagrange equation that*

$$\forall s \in [0, t], -\frac{d}{ds}D_v L(x'(s), x(s)) + D_x L(x'(s), x(s)) = 0 \tag{30}$$

Remark. *The Euler-Lagrange equation consists of n* ***second-order ODEs****. Note that when $x(s)$ is the solution to the Euler-Lagrange equation, it does not necessarily achieve the inf of the action functional in the variational problem so* ***the converse of this theorem is not true****.*

In order to link Euler-Lagrange equation back to Hamilton's equations, let's first define

$$p(s) = D_v L(x'(s), x(s)) \tag{31}$$

as the **generalized momentum for position** $x(s)$ **and velocity** $x'(s)$ (we will see why this has something to do with momentum later). We have to **assume that given** $x, p \in \mathbb{R}^n$ **the equation** $p = D_v L(v, x)$ **can be uniquely solved for** v **as a smooth function of** p **and** x **as** $v(p, x)$. The **Hamiltonian** H **associated with Lagrangian** L is defined as

$$H(p, x) = p \cdot v(p, x) - L(v(p, x), x) \ (p, x \in \mathbb{R}^n) \tag{32}$$

for $v(p, x)$ satisfying $p = D_v L(v, x)$ for given x, p defined implicitly.

Remark. *To understand the motivation of those definitions, let's consider the classical setting in physics that*

$$L(v, x) = \frac{1}{2} m ||v||_2^2 - \phi(x) \tag{33}$$

where $\frac{1}{2} m ||v||_2^2$ *is the kinetic energy and* ϕ *is the potential energy with the mass* $m > 0$*. The Lagrangian immediately tells us that the Euler-Lagrange equation is*

$$-\frac{d}{ds} m x'(s) - D\phi(x(s)) = 0 \tag{34}$$

$$m \cdot x''(s) = -D\phi(x(s)) \tag{35}$$

where $D\phi$ *is the force field generated by* ϕ *and this is* ***Newton's second law*** *for the acceleration of a particle with mass* m *in such force field.*

Let's then try to calculate the generated momentum

$$p(s) = m \cdot x'(s) \tag{36}$$

which is consistent with the true momentum in this case. The implicit definition of v *is then*

$$p(s) = D_v L(v(s), x(s)) \tag{37}$$

$$m \cdot x'(s) = m \cdot v(s) \tag{38}$$

since it can be uniquely solved for v *as a smooth function, it must be true that* $v(p, x) = x'(s)$*, just the velocity of the particle. As a result, the Hamiltonian for such Lagrangian is*

$$H(p, x) = m \cdot v \cdot v - L(v, x) \tag{39}$$

$$= \frac{1}{2} m ||v||_2^2 + \phi(x) \tag{40}$$

so the Hamiltonian is the ***total energy*** *as the sum of kinetic and potential energy.*

Theorem 2. ***(Connection with Hamilton's Equation)*** *Let $x(s)$ be the optimal solution to the variational problem and $p(s)$ be its generalized momentum defined as $p(s) = D_v L(x'(s), x(s))$ above, then those two quantities satisfy Hamilton's equations*

$$\begin{cases} x'(s) = D_p H(p(s), x(s)) \\ p'(s) = -D_x H(p(s), x(s)) \end{cases} \tag{41}$$

for $s \in [0, t]$ and the mapping

$$s \to H(p(s), x(s)) \tag{42}$$

is constant.

Proof. Here is where the assumption that $p = D_v L(v, x)$ has unique smooth solution $v = v(p, x)$ comes in. By such assumption, we conclude that $v(p(s), x(s)) = x'(s)$.

After noticing this fact, we are left with pure calculations for $D_p H, D_x H$. By definition, $H(p, x) = p \cdot v(p, x) - L(v(p, x), x)$, so

$$H_{p_i}(p, x) = \sum_{j=1, j\neq i}^{n} p_j \cdot v^j_{p_i}(p, x) + v^i(p, x) + p_i \cdot v^i_{p_i}(p, x) - D_v L(v(p, x), x) \cdot D_{p_i} v(p, x) \tag{43}$$

$$= \sum_{j=1}^{n} [p_j \cdot v^j_{p_i}(p, x) - L_{v_j}(v(p, x), x) \cdot v^j_{p_i}(p, x)] + v^i(p, x) \tag{44}$$

$$= \sum_{j=1}^{n} [p_j - L_{v_j}(v(p, x), x)] \cdot v^j_{p_i}(p, x) + v^i(p, x) \tag{45}$$

$$= v^i(p, x) \tag{46}$$

since $p = D_v L(v, x)$ by the definition of v. As a result,

$$H_{p_i}(p(s), x(s)) = v^i(p(s), x(s)) = [x_i(s)]' \tag{47}$$

proved how the first n equations come.

For the next n equations, the calculation is similar

$$H_{x_i}(p,x) = \sum_{j=1}^{n} p_j v^j_{x_i}(p,x) - D_v L(v(p,x),x) \cdot D_{x_i} v(p,x) - D_x L(v(p,x),x) \tag{48}$$

$$= \sum_{j=1}^{n} [p_j v^j_{x_i}(p,x) - p_j \cdot v^j_{x_i}(p,x)] - D_x L(v(p,x),x) \tag{49}$$

$$= -D_x L(v(p,x),x) \tag{50}$$

by applying the definition of v once more. As a result,

$$H_{x_i}(p(s),x(s)) = -D_x L(v(p(s),x(s)),x(s)) = -D_x L(x'(s),x(s)) \tag{51}$$

proves the Hamilton's equations.

Moreover,

$$\frac{d}{ds} H(p(s),x(s)) = D_p H(p(s),x(s)) \cdot p'(s) + D_x H(p(s),x(s)) \cdot x'(s) \tag{52}$$

$$= x'(s) \cdot p'(s) - p'(s) \cdot x'(s) \tag{53}$$

$$= 0 \tag{54}$$

and this is telling us that the Hamiltonian is invariant w.r.t. time. □

Remark. *To briefly conclude what we have talked about in this section, we start from introducing Lagrangian as the "running loss function" of the variational problem and hope to find the optimal path $x(s)$ minimizing the loss. Such optimal path shall then satisfy the Euler-Lagrange equation consisting of n second-order ODEs.*

From the Euler-Lagrange equations, one can further introduce the generalized momentum $p(s)$ and the velocity $v(p,x)$ as the unique smooth solution to $p = D_v L(v,x)$ (such $v(s) = x'(s)$ is the unique velocity such that the generalized momentum is the given p). The Hamiltonian is defined and the optimal path $x(s)$ and the generalized momentum $p(s)$ must satisfy the Hamilton's equation. Moreover, ***the Hamiltonian won't change as time goes by****.*

As a result, we have interpreted the meaning of the Hamilton equations derived from the method of characteristics.

Legendre Transform & Frenchel Conjugate

Now let's turn back to HJE

$$\begin{cases} u_t + H(Du) = 0 & in\ \mathbb{R}^n \times (0,\infty) \\ u = g & on\ \mathbb{R}^n \times \{t=0\} \end{cases} \tag{55}$$

with **the dependence on x of Hamiltonian H cancelled**. Now the Lagrangian $L(v)$ only depends on v. Let's **assume that the Lagrangian is a convex function with** $\lim_{||v||\to\infty} \frac{L(v)}{||v||} = +\infty$ so of course it's continuous.

The **Legendre transform** provides the **Frenchel conjugate** of the Lagrangian as

$$L^*(p) = \sup_{v\in\mathbb{R}^n} \{p\cdot v - L(v)\} \tag{56}$$

The motivation of considering Frenchel conjugate comes from the fact that in previous discussions the Hamiltonian is defined as $H(p,x) = p\cdot v(p,x) - L(v(p,x),x)$, a form very similar to the conjugate of Lagrangian. To figure out the relationship between Hamiltonian and Lagrangian, notice that under the assumptions for Lagrangian, $p\cdot v - L(v)$ is concave and continuous in v. For each fixed $p\in\mathbb{R}^n$,

$$\frac{p\cdot v - L(v)}{||v||} = p\cdot\frac{v}{||v||} - \frac{L(v)}{||v||} \to -\infty\ (||v||\to\infty) \tag{57}$$

so there must **exist $v^*\in\mathbb{R}^n$ such that the sup can be attained**, i.e. $L^*(p) = p\cdot v^* - L(v^*)$. Note that **if L is differentiable at** v^*, then

$$p - DL(v^*) = 0 \tag{58}$$

since v^* achieves the sup. This gives us the equation $p = DL(v^*)$ which is just the definition equation for velocity $v(p)$ in the context above. As a result, $v(p) = v^*$ is the solution (although no uniqueness ensured). Replacing v^* with the velocity $v(p)$ one can see

$$p\cdot v(p) - L(v(p)) = L^*(p) \tag{59}$$

and the LHS is an analogue to the definition of the Hamiltonian at p! Heuristically, this gives rise to the convex duality construction of Lagrangian and Hamiltonian.

Theorem 3. ***(Convex Duality of Lagrangian and Hamiltonian)*** *Assume that Lagrangian $L = L(v)$ is convex and $\lim_{||v||\to\infty} \frac{L(v)}{||v||} = +\infty$ and **define** Hamiltonian $H = L^*$, then H is still convex, $\lim_{||p||\to\infty} \frac{H(p)}{||p||} = +\infty$ and $L = H^*$.*

In particular, when H is differentiable at p and L is differentiable at v, then the followings are equivalent:

$$\begin{cases} p \cdot v = L(v) + H(p) \\ p = DL(v) \\ v = DH(p) \end{cases} \tag{60}$$

Proof. Note that $H = L^*$ so $H^* = L^{**}$. Note that since L is a convex and closed function, its Frenchel conjugate must be itself (since double Frenchel conjugate gives the convex envelope), so $H^* = L$ is still convex and closed.

Notice that $H(p) = \sup_{v \in \mathbb{R}^n} \{p \cdot v - L(v)\}$, so

$$\forall \lambda > 0, H(p) \geq p \cdot \lambda \frac{p}{||p||} - L\left(\lambda \frac{p}{||p||}\right) \tag{61}$$

$$\geq \lambda ||p|| - \sup_{B(0,\lambda)} L \tag{62}$$

it's then obvious that $\lim_{||p|| \to \infty} \frac{H(p)}{||p||} \geq \lambda$, so $\lim_{||p|| \to \infty} \frac{H(p)}{||p||} = +\infty$.

When H is differentiable at p and L is differentiable at v, note that if $p \cdot v = L(v) + H(p)$ then v is achieving the sup in $H(p) = \sup_{v \in \mathbb{R}^n} \{p \cdot v - L(v)\}$ so

$$p - DL(v) = 0 \tag{63}$$

and p is achieving the sup in $L(v) = \sup_{p \in \mathbb{R}^n} \{p \cdot v - H(p)\}$ so

$$v - DH(p) = 0 \tag{64}$$

Conversely, if $p = DL(v)$, then it's true that $H(p) = p \cdot v - L(v)$ so it's proved. □

Remark. *Consider the previous example that*

$$L(v) = \frac{1}{2} m ||v||^2 \tag{65}$$

then $H(p) = \sup_{v \in \mathbb{R}^n} \left\{p \cdot v - \frac{1}{2} m ||v||^2\right\}$ and the sup is achieved at $v^* = \frac{1}{m} p$

$$H(p) = \frac{1}{2m} ||p||^2 \tag{66}$$

if $p = DL(v) = mv$, then the Hamiltonian is actually

$$H(p) = \frac{1}{2} m ||v||^2 \tag{67}$$

which is equal to the Lagrangian when there's no potential and $H(p) + L(v) = p \cdot v$.

Remark. *Let's compute some more examples to illustrate the connection between Lagrangian and Hamiltonian.*

Consider $H(p) = \frac{1}{r}||p||^r$ $(1 < r < \infty)$*, so*

$$L(v) = \sup_{p \in \mathbb{R}^n} \{p \cdot v - H(p)\} \tag{68}$$

and the sup is achieved when $v = p \cdot ||p||^{r-2}$*, so* v *is parallel to* p *with* $p = kv$ $(k > 0)$*. Plug in to find*

$$L(v) = \sup_{k>0} \left\{ k||v||^2 - \frac{k^r}{r}||v||^r \right\} \tag{69}$$

and take another derivative w.r.t. k *to find that the sup is achieved when* $k = ||v||^{\frac{2-r}{r-1}}$*, so*

$$L(v) = \frac{r-1}{r}||v||^{\frac{r}{r-1}} \tag{70}$$

$$= \frac{1}{s}||v||^s \tag{71}$$

where $\frac{1}{s} + \frac{1}{r} = 1$*, so* s *is the Holder conjugate of* r*.*

Consider $H(p) = \frac{1}{2}p^T Ap + b \cdot p$*, where* A *is symmetric, positive definite and* $b \in \mathbb{R}^n$*, then*

$$L(v) = \sup_{p \in \mathbb{R}^n} \{p \cdot v - H(p)\} \tag{72}$$

and the sup is achieved when $p = A^{-1}(v - b)$*, so*

$$L(v) = \frac{1}{2}(v-b)^T A^{-1}(v-b) \tag{73}$$

Remark. *For convex function, one can define the subdifferential of* H *at* p *so that the Frenchel inequality holds*

$$H(p) + L(v) \geq p \cdot v \tag{74}$$

and the equality holds if and only if $v \in \partial H(p)$ *if and only if* $p \in \partial L(v)$*, a generalization of the conclusion in the theorem above.*

Hopf-Lax Formula

We still consider the HJE with Hamiltonian H not depend on x but only depends on Du. The characteristic ODEs then become

$$\begin{cases} p'(s) = 0 \\ z'(s) = DH(p(s)) \cdot p(s) - H(p(s)) \\ x'(s) = DH(p(s)) \end{cases} \tag{75}$$

with the equation for $p'(s), x'(s)$ being Hamilton's equations. Note that since $x'(s) = DH(p(s))$, by the theorem we have proved above, $L(x'(s)) + H(p(s)) = p \cdot x'(s)$. So **the equation of $z'(s)$ is describing the fact that** $z'(s) = L(x'(s))$. From the method of characteristics,

$$z(t) = u(x(t), t) = \int_0^t L(x'(s))\, ds + g(x(0)) \tag{76}$$

since $z(0) = u(x(0), 0) = g(x(0))$ by the initial value condition, providing us an ansatz of the solution. However, this construction of the solution $u(x, t)$ assumes the smoothness of the solution, which is often not the case for HJE. To think about modifying the construction of the solution such that it also works for non-smooth solution $u(x, t)$, we notice that

$$\int_0^t L(x'(s))\, ds \tag{77}$$

is the "running loss function" of the variational problem we have mentioned above and $x(s)$ is the optimal path found in that problem. As a result, we can think about **defining**

$$u(x, t) \overset{def}{=} \inf_w \left\{ \int_0^t L(w'(s))\, ds + g(w(0)) : w : [0, t] \to \mathbb{R}^n, w \in C^1, w(t) = x \right\} \tag{78}$$

as the optimal "loss" determined by the Lagrangian among all paths that hits x at time t. To see how this works as the solution to the HJE, refer to the following theorem. We **assume that H is smooth and convex with $\lim_{||p|| \to \infty} \frac{H(p)}{||p||} = +\infty$ and $g : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz in the following context**.

Theorem 4. ***(Hopf-Lax Formula)*** *For fixed $x \in \mathbb{R}^n, t > 0$,*

$$u(x, t) = \inf_w \left\{ \int_0^t L(w'(s))\, ds + g(w(0)) : w : [0, t] \to \mathbb{R}^n, w \in C^1, w(t) = x \right\} \tag{79}$$

$$= \inf_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x - y}{t} \right) + g(y) \right\} \tag{80}$$

Proof. Consider $\forall y \in \mathbb{R}^n$ and the path $w(s) = y + \frac{s}{t}(x - y)$ so $w(t) = x$ (constructed based on $\frac{x-y}{t}$ inside the Lagrangian), it's obvious that

$$u(x, t) \leq \int_0^t L\left(\frac{x - y}{t} \right) ds + g(y) \tag{81}$$

so by taking inf w.r.t. y on both sides

$$u(x, t) \leq \inf_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x - y}{t} \right) + g(y) \right\} \tag{82}$$

Conversely, for any C^1 path w such that $w(t) = x$, take $y = w(0)$

$$tL\left(\frac{x-y}{t}\right) + g(y) = tL\left(\frac{x-w(0)}{t}\right) + g(w(0)) \tag{83}$$

$$= tL\left(\frac{1}{t}\int_0^t w(s)\,ds\right) + g(w(0)) \tag{84}$$

$$\leq \int_0^t L(w'(s))\,ds + g(w(0)) \tag{85}$$

because of Jensen's inequality applied for $\frac{1}{t}\int_0^t f(s)\,ds$, the integral average of f on $[0,t]$

$$\frac{1}{t}\int_0^t L\left(w'(s)\right)\,ds \geq L\left(\frac{1}{t}\int_0^t w'(s)\,ds\right) \tag{86}$$

by taking inf w.r.t. all paths w on both sides, one can conclude that

$$u(x,t) \geq \inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\} \tag{87}$$

so the theorem is proved. □

Remark. *The shifting from the ansatz $u(x(t),t) = \int_0^t L(x'(s))\,ds + g(x(0))$ to Hopf-Lax formula is critical! The main thought comes from the variational problem viewing $x(s)$ as the optimal path and the integral of Lagrangian as running loss.*

Actually, from another perspective, one may be able to see the spirit of stochastic control out of the Hopf-Lax formula. Notice that w can be view as a stochastic process instead of a deterministic function, and the $\int_0^t L(w'(s))\,ds$ can be viewed as a running loss with $g(w(0))$ as terminal loss (conditional on the filtration $\mathscr{F}_t$, i.e. all information available until time t, that's why the domain of w is $[0,t]$. Then $u(x,t)$ is essentially a value function conditioning on $w(t) = x$, i.e. the process passes through x at time t. In such sense, ***HJE is actually characterizing the value function of a stochastic control problem in a deterministic way!***

Remark. ***In Hopf-Lax formula, the inf can always be attained.*** *Note that $f(y) = tL\left(\frac{x-y}{t}\right) + g(y)$ is continuous in y and*

$$\frac{f(y)}{||y||} = \frac{L\left(\frac{x-y}{t}\right)}{\frac{||y||}{t}} + \frac{g(y)}{||y||} \tag{88}$$

with $L = H^*$ so $\lim_{||v||\to\infty}\frac{L(v)}{||v||} = +\infty$ and since g is Lipschitz, $\frac{g(y)}{||y||} \leq \frac{g(0)+Lips(g)||y||}{||y||} \leq Lips(g) + \varepsilon$ for large enough $||y||$. As a result,

$$\frac{f(y)}{||y||} \to +\infty\ (||y|| \to \infty) \tag{89}$$

combining with continuity, we see that the minimum of $f(y)$ must be attained by some $y \in \mathbb{R}^n$.

Hopf-Lax Formula as Solution to HJE

Now let's argue that the heuristic definition of such $u(x,t)$ by the Hopf-Lax formula is actually a solution to HJE. In order to prove this, let's first consider some useful propositions.

Theorem 5. ***(Flow Property)*** *For each $x \in \mathbb{R}^n$ and $s \in [0,t]$,*

$$u(x,t) = \inf_{y\in\mathbb{R}^n}\left\{(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s)\right\} \tag{90}$$

Proof. Let's start by noticing that for $\forall y \in \mathbb{R}^n, s \in [0,t]$, there always exists $z \in \mathbb{R}^n$ such that the inf in Hopf-Lax formula is attained, i.e.

$$u(y,s) = sL\left(\frac{y-z}{s}\right) + g(z) \tag{91}$$

in order to connect it with $\frac{x-y}{t-s}$, consider the convex representation and apply the convexity of L that

$$\frac{t-s}{t}\frac{x-y}{t-s} + \frac{s}{t}\frac{y-z}{s} = \frac{x-z}{t} \tag{92}$$

$$\frac{t-s}{t}L\left(\frac{x-y}{t-s}\right) + \frac{s}{t}L\left(\frac{y-z}{s}\right) \geq L\left(\frac{x-z}{t}\right) \tag{93}$$

so that

$$(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) = (t-s)L\left(\frac{x-y}{t-s}\right) + sL\left(\frac{y-z}{s}\right) + g(z) \geq tL\left(\frac{x-z}{t}\right) + g(z) \tag{94}$$

take inf w.r.t. y on both sides, one would see that

$$\inf_{y\in\mathbb{R}^n}\left\{(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s)\right\} \geq tL\left(\frac{x-z}{t}\right) + g(z) \geq u(x,t) \tag{95}$$

On the other hand, let's try to find $y \in \mathbb{R}^n$ such that $(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \leq u(x,t)$. Apply the Hopf-Lax formula again to find $w \in \mathbb{R}^n$ such that $u(x,t) = tL\left(\frac{x-w}{t}\right) + g(w)$. Consider applying the convexity of L again, to set

$$y = \frac{s}{t}x + \frac{t-s}{t}w \tag{96}$$

$$\frac{x-y}{t-s} = \frac{x-w}{t} \tag{97}$$

and apply Hopf-Lax formula for $u(y,s)$ once more to find

$$(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \leq (t-s)L\left(\frac{x-w}{t}\right) + u(y,s) \leq (t-s)L\left(\frac{x-w}{t}\right) + sL\left(\frac{y-w}{s}\right) + g(w) \tag{98}$$

note that $\frac{y-w}{s} = \frac{x-w}{t}$, so

$$(t-s)L\left(\frac{x-y}{t-s}\right) + u(y,s) \leq tL\left(\frac{x-w}{t}\right) + g(w) = u(x,t) \tag{99}$$

by taking inf w.r.t. y on both sides, we proved the conclusion. □

Remark. *Note that* ***the inf in this theorem can always be attained****. This requires proving the fact that $y \to u(y,s)$ is continuous, which will be proved in a later context.*

Remark. *The reason why we are calling this property the flow property is that this is telling us that we can act as if we are starting at time $s < t$ with initial value $u(y,s)$. Then the Hopf-Lax formula still holds for such problem and will generate the same u as what we would derive with an initial value condition at time 0. This is actually very similar to the flow property of diffusion process.*

Under the assumption that g is Lipschitz, one would see that such u is also Lipschitz in $\mathbb{R}^n \times [0,\infty)$ and it agrees with the initial value condition g, i.e. $\forall x \in \mathbb{R}^n, u(x,0) = g(x)$.

Theorem 6. ***(Lipschitz Continuity)*** *Such u is Lipschitz in $\mathbb{R}^n \times [0,\infty)$, and $\forall x \in \mathbb{R}^n, u(x,0) = g(x)$.*

Proof. First prove that $u(x,t)$ is Lipschitz in x. By Hopf-Lax formula, there exists $y \in \mathbb{R}^n$ such that $u(x,t) = tL\left(\frac{x-y}{t}\right) + g(y)$. As a result, for $\forall x, x' \in \mathbb{R}^n$,

$$\begin{aligned} u(x',t) - u(x,t) &= \inf_z \left\{ tL\left(\frac{x'-z}{t}\right) + g(z) \right\} - tL\left(\frac{x-y}{t}\right) - g(y) && (100)\\ &\leq tL\left(\frac{x'-(x'-x+y)}{t}\right) + g(x'-x+y) - tL\left(\frac{x-y}{t}\right) - g(y) && (101)\\ &= g(x'-x+y) - g(y) \leq Lips(g) \cdot ||x'-x|| && (102) \end{aligned}$$

so

$$|u(x',t) - u(x,t)| \leq Lips(g) \cdot ||x'-x|| \tag{103}$$

by interchanging x and x'.

Now let's prove that u and g agree when $t = 0$. Note that by Hopf-Lax formula, $u(x,t) \leq tL(0) + g(x)$. Set $t = 0$ to find $u(x,0) \leq g(x)$. For the other direction, we would need to use the conjugacy of Lagrangian and Hamiltonian.

$$\begin{aligned} u(x,t) &= \inf_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\} && (104)\\ &= g(x) + \inf_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) - g(x) \right\} && (105)\\ &\geq g(x) - t \sup_{y \in \mathbb{R}^n} \left\{ -L\left(\frac{x-y}{t}\right) + Lips(g) \cdot \frac{||y-x||}{t} \right\} && (106) \end{aligned}$$

by setting $z = \frac{x-y}{t}$ as a new variable, one can see the structure of this sup

$$u(x,t) \geq g(x) - t \sup_{z \in \mathbb{R}^n} \{-L(z) + Lips(g) \cdot ||z||\} \tag{107}$$

in order to connect this sup with the Frenchel conjugate of Lagrangian which is the Hamiltonian, we would like to see the forms like $w \cdot z - L(z)$. That's why we view $Lips(g) \cdot ||z||$ as $Lips(g)\frac{z}{||z||} \cdot z$ with $w = Lips(g)\frac{z}{||z||}$

$$u(x,t) \geq g(x) - t \sup_{w \in B(0,Lips(g))} \sup_{z \in \mathbb{R}^n} \{-L(z) + w \cdot z\} \tag{108}$$

$$= g(x) - t \sup_{w \in B(0,Lips(g))} H(w) \tag{109}$$

and since H is continuous and convex, $\sup_{w \in B(0,Lips(g))} H(w) < \infty$, setting $t = 0$ to see

$$u(x,0) \geq g(x) \tag{110}$$

and we conclude that such u is equal to g when $t = 0$.

At last, prove that $u(x,t)$ is Lipschitz in t. For $\forall 0 < t < t'$, by the flow property,

$$u(x,t') - u(x,t) = \inf_{y \in \mathbb{R}^n} \left\{ (t'-t) L\left(\frac{x-y}{t'-t}\right) + u(y,t) \right\} - u(x,t) \tag{111}$$

$$\leq (t'-t) L(0) + u(x,t) - u(x,t) \tag{112}$$

$$= (t'-t) \cdot L(0) \tag{113}$$

on the other hand, let's apply the trick above one more time

$$u(x,t') = u(x,t) + \inf_{y \in \mathbb{R}^n} \left\{ (t'-t) L\left(\frac{x-y}{t'-t}\right) + u(y,t) - u(x,t) \right\} \tag{114}$$

$$\geq u(x,t) + (t'-t) \inf_{y \in \mathbb{R}^n} \left\{ L\left(\frac{x-y}{t'-t}\right) - Lips(u) \cdot \frac{||y-x||}{t'-t} \right\} \tag{115}$$

consider $z = \frac{y-x}{t'-t}$ and transform $Lips(u) \cdot \frac{||y-x||}{t'-t}$ into the inner product form to see

$$u(x,t') - u(x,t) \geq -(t'-t) \sup_{z \in \mathbb{R}^n} \{-L(z) + Lips(u) \cdot ||z||\} \tag{116}$$

$$= -(t'-t) \sup_{w \in B(0,Lips(u))} \sup_{z \in \mathbb{R}^n} \{-L(z) + w \cdot z\} \tag{117}$$

$$= -(t'-t) \sup_{w \in B(0,Lips(u))} H(w) \tag{118}$$

in all, we see that

$$|u(x,t')-u(x,t)| \le C\cdot|t'-t|, C=\max\left\{|L(0)|, \sup_{w\in B(0,Lips(u))}|H(w)|\right\} \tag{119}$$

and such constant C has no dependence on x and t, that's why u is also Lipschitz w.r.t. time t.

□

Theorem 7. ***(Hopf-Lax Formula as Solution to HJE)*** *For u defined by the Hopf-Lax formula, if it's differentiable at a point (x,t), then $u_t(x,t)+H(Du(x,t))=0$. In particular, such u is differentiable almost everywhere and it's the solution to HJE in the almost everywhere sense.*

Proof. By Rademacher's theorem, Lipschitz function on an open subset of $\mathbb{R}^n$ is almost everywhere differentiable. So we only have to prove that HJE holds whenever u is differentiable at (x,t).

let's first calculate the directional derivative of u along any vector v. By flow property,

$$u(x+hv,t+h)=\inf_{y\in\mathbb{R}^n}\left\{hL\left(\frac{x+hv-y}{h}\right)+u(y,t)\right\} \tag{120}$$

$$\le hL(v)+u(x,t) \tag{121}$$

as a result,

$$\forall v\in\mathbb{R}^n, v\cdot Du(x,t)+u_t(x,t)=\lim_{h\to0^+}\frac{u(x+hv,t+h)-u(x,t)}{h}\le L(v) \tag{122}$$

note that the Hamiltonian is the Frenchel conjugate of Lagrangian, so

$$u_t(x,t)+H(Du(x,t))=u_t(x,t)+\sup_{v\in\mathbb{R}^n}\{v\cdot Du(x,t)-L(v)\}\le 0 \tag{123}$$

To prove the other side, we have to choose v in the sup carefully. By Hopf-Lax formula, there exists $z\in\mathbb{R}^n$ such that $u(x,t)=tL\left(\frac{x-z}{t}\right)+g(z)$. Take $v=\frac{x-z}{t}$ in the sup to find

$$u_t(x,t)+H(Du(x,t))\ge u_t(x,t)+\frac{x-z}{t}\cdot Du(x,t)-L\left(\frac{x-z}{t}\right) \tag{124}$$

again we have to use finite difference to approximate the partial derivatives

$$u(x,t)-u\left(\frac{t-h}{t}x+\frac{h}{t}z,t-h\right)=tL\left(\frac{x-z}{t}\right)+g(z)-u\left(\frac{t-h}{t}x+\frac{h}{t}z,t-h\right) \tag{125}$$

$$\ge tL\left(\frac{x-z}{t}\right)+g(z)-(t-h)L\left(\frac{x-z}{t}\right)-g(z) \tag{126}$$

$$=hL\left(\frac{x-z}{t}\right) \tag{127}$$

setting $h \to 0^+$ to know

$$u_t(x,t) + \frac{x-z}{t} \cdot Du(x,t) \geq L\left(\frac{x-z}{t}\right) \tag{128}$$

Finally, we have proved that

$$u_t(x,t) + H(Du(x,t)) = 0 \tag{129}$$

□

The theorem above ends our discussion on the solution to **a particular kind of HJE (Hamiltonian only depends on Du and is convex with Lipschitz initial value condition)**. To see a direct example of the application of Hopf-Lax formula, consider the following PDE

$$\begin{cases} u_t + ||Du||^2 = 0 & in\ \mathbb{R}^n \times (0,\infty) \\ u = +\infty \cdot \mathbb{I}_{E^c} & on\ \mathbb{R}^n \times \{0\} \end{cases} \tag{130}$$

with E as a closed subset in $\mathbb{R}^n$. Now the Hamiltonian is $H(p) = ||p||^2$ so Lagrangian is its Frenchel conjugate

$$L(v) = \sup_{p \in \mathbb{R}^n} \{p \cdot v - H(p)\} = \frac{1}{4}||v||^2 \tag{131}$$

Apply the Hopf-Lax formula to find

$$u(x,t) = \inf_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\} \tag{132}$$

$$= \inf_{y \in E} \left\{ \frac{1}{4t}||x-y||^2 \right\} \tag{133}$$

$$= \frac{1}{4t} dist^2(x,E) \tag{134}$$

the solution has something to do with the distance between x and E.

Optimal Control Problem and Hamilton-Jacobi-Bellman Equation

In this section we state the deterministic optimal control problem and find the connection between optimal control problem, HJE, Hamilton-Jacobi-Bellman equation (HJBE) and -Lax formula.

Problem Formulation

All control problems have a certain dynamics telling us how the system evolves. In optimal control problem, the dynamics is given by an ODE

$$\begin{cases} x'(s) = f(x(s), \alpha(s)) \ (s \in [t, T]) \\ x(t) = x \end{cases} \tag{135}$$

where the dynamics works in time interval $[t, T]$ with T fixed and an initial value condition given at time t. We will be varying the time t and the initial value x shortly afterwards to get a PDE describing such an optimal control problem. **Note that $x(s)$ denotes the state of the problem at time s while x denotes the initial value condition.** Viewing $x'(s)$ as $\frac{x(s+h)-x(s)}{h}$ for $h \to 0^+$, the ODE is describing how the change of state from time s to time $s+h$ happens given the current state $x(s)$ and the current **control** $\alpha(s)$. (so it's actually a **Markovian** setting since $x'(s)$ has nothing to do with $\{x(t)\}|_{t<s}$ given $x(s)$.) The control can be understood as the "action" in discrete-time Markov decision process that changes the state evolution and has something to do with the rewards.

The control is nothing complicated but a set of parameters given at each time that will change the dynamics of the system, eventually changing the state evolution of the system. Let's denote $A \subset \mathbb{R}^m$ as some given compact set consisting of all possible values the control **at a given time** $\alpha(s)$ can take. The **admissible set**

$$\mathscr{A} = \{\alpha : [0, T] \to A : \alpha(\cdot) \ measurable\} \tag{136}$$

then denotes all possible controls across the whole time interval $[0, T]$ (since control may change over time, it maps each time point to the value of control at that time point). It's then clear that the function

$$f : \mathbb{R}^n \times A \to \mathbb{R}^n \tag{137}$$

is mapping a $m+n$-dimensional vector to a n-dimensional vector. Let's **assume that f is a given bounded Lipschitz function**. This assumption is made to ensure that the ODE always has unique solution for every given control $\alpha(\cdot) \in \mathscr{A}$ denoted $x(\cdot) = x^{\alpha(\cdot)}(\cdot)$. **Our goal in optimal control problem is to find the optimal control $\alpha^*(\cdot)$ under some criteria**.

In order to define the optimality, we introduce the **cost functional** that represents the cost one has to pay selecting control α with initial value condition $x(t) = x$

$$C_{x,t}[\alpha] = \int_t^T r(x(s), \alpha(s))\, ds + g(x(T)) \tag{138}$$

here $\int_t^T r(x(s),\alpha(s))\,ds$ is the running cost and $g(x(T))$ is the terminal cost where $r:\mathbb{R}^n\times A\to\mathbb{R}, g:\mathbb{R}^n\to\mathbb{R}$ are **assumed to be bounded and Lipschitz in variable** x.

To sum up, given time $t\in[0,T]$ and the initial value condition $x(t)=x$, we want to find **the optimal control** α^* such that

$$C_{x,t}[\alpha^*]=\inf_{\alpha\in\mathscr{A}} C_{x,t}[\alpha] \tag{139}$$

Value Function

Let's consider the **value function** $u(x,t)$ as the least possible cost among all admissible control with initial value condition $x(t)=x$ (with dynamic programming approach), i.e.

$$u(x,t)=\inf_{\alpha\in\mathscr{A}} C_{x,t}[\alpha] \tag{140}$$

then we hope to find a PDE that characterizes such value function u.

Theorem 8. ***(Optimality Condition)*** *For fixed $x\in\mathbb{R}^n, 0\le t<T$ and $h>0$ such that $t+h\le T$,*

$$u(x,t)=\inf_{\alpha\in\mathscr{A}}\left\{\int_t^{t+h} r(x(s),\alpha(s))\,ds+u(x(t+h),t+h)\right\} \tag{141}$$

where $x(\cdot)=x^{\alpha(\cdot)}(\cdot)$ is the solution to the ODE for fixed control $\alpha(\cdot)$.

Proof. For any control $\alpha_1\in\mathscr{A}$, the ODE has solution $x_1(\cdot)$. Now we want to prove that LHS is less than RHS, so we have to argue that $\forall\varepsilon>0$,

$$u(x,t)\le\int_t^{t+h} r(x_1(s),\alpha_1(s))\,ds+u(x_1(t+h),t+h)+\varepsilon \tag{142}$$

In order to achieve this goal, expand the inf in the definition of $u(x,t)$ for time $t+h$ and initial value $x_1(t+h)$ to find that $\forall\varepsilon>0$, there exists $\alpha_2\in\mathscr{A}$ and the solution to the ODE for fixed control α_2 which is $x_2(\cdot)$ such that

$$u(x_1(t+h),t+h)+\varepsilon\ge C_{x_1(t+h),t+h}[\alpha_2]=\int_{t+h}^T r(x_2(s),\alpha_2(s))\,ds+g(x_2(T)) \tag{143}$$

so far, we have successfully figured out a lower bound for $u(x_1(t+h),t+h)$. To connect it with $u(x,t)$ and any control α_1, we can construct a new control α_3 that sticks to α_1 before time $t+h$ but shifts to α_2 after time $t+h$.

$$\alpha_3(s)=\alpha_1(s)\cdot\mathbb{I}_{t\le s\le t+h}+\alpha_2(s)\cdot\mathbb{I}_{t+h\le s\le T} \tag{144}$$

under our assumption, the original ODE has unique solution, and it's easy to see that

$$x_3(s)=x_1(s)\cdot\mathbb{I}_{t\le s\le t+h}+x_2(s)\cdot\mathbb{I}_{t+h\le s\le T} \tag{145}$$

is the solution to the ODE for fixed control α_3 since

$$\forall t \le s \le t+h, x_3'(s) = x_1'(s) = f(x_1(s), \alpha_1(s)) = f(x_3(s), \alpha_3(s)) \tag{146}$$
$$\forall t+h \le s \le T, x_3'(s) = x_2'(s) = f(x_2(s), \alpha_2(s)) = f(x_3(s), \alpha_3(s)) \tag{147}$$
$$x_3(t) = x_1(t) = x \tag{148}$$

now we can see that

$$u(x,t) \le C_{x,t}[\alpha_3] \tag{149}$$
$$= \int_t^T r(x_3(s), \alpha_3(s))\, ds + g(x_3(T)) \tag{150}$$
$$= \int_t^{t+h} r(x_1(s), \alpha_1(s))\, ds + \int_{t+h}^T r(x_2(s), \alpha_2(s))\, ds + g(x_2(T)) \tag{151}$$
$$\le \int_t^{t+h} r(x_1(s), \alpha_1(s))\, ds + u(x_1(t+h), t+h) + \varepsilon \tag{152}$$

so we have proved that

$$u(x,t) \le \inf_{\alpha \in \mathscr{A}} \left\{ \int_t^{t+h} r(x(s), \alpha(s))\, ds + u(x(t+h), t+h) \right\} \tag{153}$$

On the other hand, $\forall \varepsilon > 0$, there exists control $\alpha_4 \in \mathscr{A}$ such that

$$u(x,t) + \varepsilon \ge C_{x,t}[\alpha_4] \tag{154}$$
$$= \int_t^T r(x_4(s), \alpha_4(s))\, ds + g(x_4(T)) \tag{155}$$
$$= \int_t^{t+h} r(x_4(s), \alpha_4(s))\, ds + \int_{t+h}^T r(x_4(s), \alpha_4(s))\, ds + g(x_4(T)) \tag{156}$$

by the inf in the definition of value function. However, by applying again the inf for $u(x_4(t+h), t+h)$

$$u(x_4(t+h), t+h) \le C_{x_4(t+h), t+h}[\alpha_4] \tag{157}$$
$$= \int_{t+h}^T r(x_4(s), \alpha_4(s))\, ds + g(x_4(T)) \tag{158}$$

so we have proved that

$$u(x,t) + \varepsilon \ge \int_t^{t+h} r(x_4(s), \alpha_4(s))\, ds + u(x_4(t+h), t+h) \tag{159}$$

and we will find

$$u(x,t) \geq \inf_{\alpha \in \mathscr{A}} \left\{ \int_t^{t+h} r(x(s),\alpha(s))\, ds + u(x(t+h), t+h) \right\} \tag{160}$$

so the theorem is proved. □

Remark. *The optimality condition is telling us a very intuitive fact: the optimal control for the process starting from x at time t, has already taken the optimal control for the process starting from $x(t+h)$ at time $t+h$ into consideration. As a result, we can view $u(x(t+h), t+h)$ as the terminal cost (only depends on the endpoint $x(t+h)$) and $\int_t^{t+h} r(x(s),\alpha(s))\, ds$ as the running cost (depends on how $x(t)$ behaves in time $[t, t+h]$). One might be able to see the Markovian structure again from this expression.*

To set up a PDE for value function $u(x,t)$, it's natural for us to prove that u is Lipschitz (so it's almost everywhere differentiable and the PDE can hold in the almost everywhere sense).

Theorem 9. ***(Boundedness and Lipschitz Continuity of Value Function)*** *The value function $u(x,t)$ under the assumptions above is bounded and Lipschitz on $\mathbb{R}^n \times [0,T]$.*

Proof. Since $u(x,t) = \inf_{\alpha \in \mathscr{A}} C_{x,t}[\alpha]$ and r, g are assumed to be bounded, it's obvious that u is also bounded

$$u(x,t) \leq \sup |r| \cdot T + \sup |g| \tag{161}$$

Now fix $t \in [0,T]$ and consider $x, \hat{x} \in \mathbb{R}$, apply the inf in the definition of value function, so $\forall \varepsilon > 0$, there exists control $\hat{\alpha}$ and $\hat{x}(s)$ as the solution to the ODE with fixed control $\hat{\alpha}$ and initial value condition $\hat{x}(t) = \hat{x}$ such that

$$u(\hat{x},t) + \varepsilon \geq \int_t^T r(\hat{x}(s),\hat{\alpha}(s))\, ds + g(\hat{x}(T)) \tag{162}$$

so let's estimate the difference

$$u(x,t) - u(\hat{x},t) \leq u(x,t) - \int_t^T r(\hat{x}(s),\hat{\alpha}(s))\, ds - g(\hat{x}(T)) + \varepsilon \tag{163}$$

$$\leq \int_t^T r(x(s),\hat{\alpha}(s))\, ds + g(x(T)) - \int_t^T r(\hat{x}(s),\hat{\alpha}(s))\, ds - g(\hat{x}(T)) + \varepsilon \tag{164}$$

note that here we are taking $x(s)$ as the solution to the ODE with initial value condition $x(t) = x$ that

$$x'(s) = f(x(s),\hat{\alpha}(s)) \tag{165}$$

since r, g are Lipschitz with Lipschitz constant C_r, C_g,

$$u(x,t) - u(\hat{x},t) \leq C_r \int_t^T ||x(s) - \hat{x}(s)||\, ds + C_g ||x(T) - \hat{x}(T)|| + \varepsilon \tag{166}$$

in order to estimate $\int_t^T ||x(s) - \hat{x}(s)|| \, ds$, note that since f is also Lipschitz with constant C_f,

$$||x'(s) - \hat{x}'(s)|| = ||f(x(s), \hat{\alpha}(s)) - f(\hat{x}(s), \hat{\alpha}(s))|| \tag{167}$$

$$\leq C_f ||x(s) - \hat{x}(s)|| \tag{168}$$

by Grownwall's inequality,

$$||x(s) - \hat{x}(s)|| \leq C||x(t) - \hat{x}(t)|| = C||x - \hat{x}|| \tag{169}$$

that's why

$$u(x,t) - u(\hat{x},t) \leq CT||x - \hat{x}|| + \varepsilon \tag{170}$$

for some constant C and thus u is Lipschitz in variable x (the other side is similar).

To prove that it's also Lipschitz in variable t, let's fix $x \in \mathbb{R}^n$ and consider $t, \hat{t} \in [0,T]$. For $\forall \varepsilon > 0$, there exists control α and the solution $x(\cdot)$ to the ODE with fixed control α such that

$$u(x,t) + \varepsilon \geq C_{x,t}[\alpha] = \int_t^T r(x(s), \alpha(s)) \, ds + g(x(T)) \tag{171}$$

consider the time-shifted control $\hat{\alpha}(s) = \alpha(s + t - \hat{t})$ and $\hat{x}$ as the solution to the ODE with fixed control $\hat{\alpha}$, one may find $\hat{x}'(s) = f(\hat{x}(s), \hat{\alpha}(s))$ and $\frac{d}{ds} x(s + t - \hat{t}) = x'(s + t - \hat{t}) = f(x(s + t - \hat{t}), \alpha(s + t - \hat{t})) = f(x(s + t - \hat{t}), \hat{\alpha}(s))$. By the uniqueness of the solution, we know that $\hat{x}(s) = x(s + t - \hat{t}), \hat{x}(\hat{t}) = x(t) = x$, so

$$u(x,\hat{t}) - u(x,t) \leq u(x,\hat{t}) - \int_t^T r(x(s), \alpha(s)) \, ds - g(x(T)) + \varepsilon \tag{172}$$

$$\leq \int_{\hat{t}}^T r(\hat{x}(s), \hat{\alpha}(s)) \, ds + g(\hat{x}(T)) - \int_t^T r(x(s), \alpha(s)) \, ds - g(x(T)) + \varepsilon \tag{173}$$

$$\leq \int_T^{T - \hat{t} + t} r(x(s), \alpha(s)) \, ds + g(\hat{x}(T)) - g(x(T)) + \varepsilon \tag{174}$$

$$\leq \sup |r| \cdot |t - \hat{t}| + C_g \cdot ||\hat{x}(T) - x(T)|| + \varepsilon \tag{175}$$

$$\leq C \cdot |t - \hat{t}| + \varepsilon \tag{176}$$

since $||\hat{x}(T) - x(T)|| \leq \sup |f| \cdot |T + t - \hat{t} - T| = \sup |f| \cdot |t - \hat{t}|$ so we have proved that $u(x,t)$ is also Lipschitz in t (the other side is similar). □

Hamilton-Jacobi-Bellman Equation (HJBE)

Now from the optimality condition and the Lipschitz continuity of the value function derived above, we can set up a PDE describing the evolution of value function $u(x,t)$.

Theorem 10. ***(HJBE for Value Function)*** *The value function under assumptions above satisfies the HJBE*

$$\begin{cases} u_t + \inf_{\alpha\in\mathscr{A}} \{f(x,\alpha)\cdot Du + r(x,\alpha)\} = 0 & in\ \mathbb{R}^n \times (0,T) \\ u = g & on\ \mathbb{R}^n \times \{t\} \end{cases} \tag{177}$$

Proof. When $t = T$, $u = \inf_{\alpha\in\mathscr{A}} C_{x,T}[\alpha] = \int_T^T r(x(s),\alpha(s))\,ds + g(x(T)) = g(x)$ gives the terminal condition.

When $0 < t < T$, recall the optimality condition that for $h > 0$ such that $t + h \le T$,

$$u(x,t) = \inf_{\alpha\in\mathscr{A}} \left\{ \int_t^{t+h} r(x(s),\alpha(s))\,ds + u(x(t+h), t+h) \right\} \tag{178}$$

where $x(\cdot)$ is the solution to the ODE for fixed control α. Let's modify both sides of this property to get HJBE, be careful with the difference between x and $x(\cdot)$ since the previous one denotes the initial value while the latter one denotes the solution to the PDE

$$\frac{u(x,t) - u(x,t+h)}{h} = \inf_{\alpha\in\mathscr{A}} \left\{ \frac{1}{h}\int_t^{t+h} r(x(s),\alpha(s))\,ds + \frac{u(x(t+h),t+h) - u(x,t+h)}{h} \right\} \tag{179}$$

$$= \inf_{\alpha\in\mathscr{A}} \left\{ \frac{1}{h}\int_t^{t+h} r(x(s),\alpha(s))\,ds + \frac{u(x(t+h),t+h) - u(x(t),t+h)}{h} \right\} \tag{180}$$

setting $h \to 0^+$ on both sides to find

$$-u_t(x,t) = \inf_{\alpha\in\mathscr{A}} \{r(x(t),\alpha(t)) + Du(x(t),t)\cdot x'(t)\} \tag{181}$$

$$= \inf_{\alpha\in\mathscr{A}} \{r(x(t),\alpha(t)) + Du(x,t)\cdot f(x(t),\alpha(t))\} \tag{182}$$

now let's neglect the initial time t and initial value x to denote the PDE as

$$u_t + \inf_{\alpha\in\mathscr{A}} \{r(x,\alpha) + Du\cdot f(x,\alpha)\} = 0 \tag{183}$$

note that u being Lipschitz guarantees that the partial derivatives w.r.t. each variable exists almost everywhere. □

Remark. *We can find the connection between HJE and HJBE that if we set the Hamiltonian as*

$$H(p,x) = \inf_{\alpha\in\mathscr{A}} \{f(x,\alpha)\cdot p + r(x,\alpha)\} \tag{184}$$

then HJBE is just HJE $u_t + H(Du,x) = 0$ but with a ***terminal value condition*** *instead of an initial value condition.*

Remark. *One may still recall the Hopf-Lax formula mentioned above to solve HJE $u_t + H(Du) = 0$ with initial value condition $u(x,0) = g(x)$ that*

$$u(x,t) = \inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right) + g(y)\right\} \tag{185}$$

with the Lagrangian L as the Frenchel conjugate of the Hamiltonian H. We can verify that such $u(x,t)$ also provides us with the solution to a special kind of HJBE.

Now that HJE has initial value condition but HJBE has terminal value condition, the most natural way is to do the time reflection $v(x,t) = u(x,T-t)$ such that the terminal value condition of u actually gives the initial value condition of v. It's easy to see that

$$v(x,0) = u(x,T) = g(x) \tag{186}$$

then notice that $v_t = -u_t, Dv = Du$, so the HJBE for u can be reformulated as the following HJE for v that

$$\begin{cases} v_t + H(Dv,x) = 0 & in\ \mathbb{R}^n\times(0,T) \\ v = g & on\ \mathbb{R}^n\times\{0\} \end{cases} \tag{187}$$

with Hamiltonian

$$H(p,x) = -\inf_{\alpha\in\mathscr{A}}\{f(x,\alpha)\cdot p + r(x,\alpha)\} \tag{188}$$

However, in order to let the Hopf-Lax formula work, we have to ***assume that*** $r(x,\alpha) = r(\alpha), f(x,\alpha) = f(\alpha)$***, i.e. both running reward and the dynamics does not depend on the state*** x*. So the HJE and the Hamiltonian becomes*

$$\begin{cases} v_t + H(Dv) = 0 & in\ \mathbb{R}^n\times(0,T) \\ v = g & on\ \mathbb{R}^n\times\{0\} \end{cases} \tag{189}$$

and

$$H(p) = -\inf_{\alpha\in\mathscr{A}}\{f(\alpha)\cdot p + r(\alpha)\} \tag{190}$$

So the Frenchel conjugate is

$$L(v) = \sup_{p\in\mathbb{R}^n}\left\{p\cdot v + \inf_{\alpha\in\mathscr{A}}\{f(\alpha)\cdot p + r(\alpha)\}\right\} \tag{191}$$

and the solution to HJE is given by

$$v(x,t) = \inf_{y\in\mathbb{R}^n}\left\{tL\left(\frac{x-y}{t}\right)+g(y)\right\} \tag{192}$$

$$= \inf_{y\in\mathbb{R}^n}\left\{\sup_{p\in\mathbb{R}^n}\left\{p\cdot(x-y)+t\inf_{\alpha\in\mathscr{A}}\{f(\alpha)\cdot p+r(\alpha)\}\right\}+g(y)\right\} \tag{193}$$

as a result, ***the solution to HJBE*** *is*

$$u(x,t) = v(x,T-t) = \inf_{y\in\mathbb{R}^n}\left\{\sup_{p\in\mathbb{R}^n}\left\{p\cdot(x-y)+(T-t)\inf_{\alpha\in\mathscr{A}}\{f(\alpha)\cdot p+r(\alpha)\}\right\}+g(y)\right\} \tag{194}$$

under the assumption that g ***is Lipschitz,*** H ***is convex and*** $\lim_{||p||\to\infty}\frac{H(p)}{||p||}=+\infty$.

However, one might realize that although we have got an analytic solution for HJBE, the assumption that the running reward and the dymanics both do not depend on state is too strong that most of the interesting examples would not satisfy such assumption. This assumption only works well for a problem setting with a single state and many actions to be chosen, i.e. the continuous-time bandit problem but fails for most reinforcement learning problems.

Although one would not be able to solve the HJBE analytically in all cases, our previous discussion about general HJE $u_t+H(Du,x)=0$ *still provides some insights. One can consider the Hamilton's equation and the Euler-Lagrange equations associated with the HJBE.*

Infinite-Horizon Problem

Among our discussion, we are assuming that there exists some upper time limit $T<\infty$ and the dynamics works in time interval $[0,T]$. However, one can also consider the infinite-horizon problem by taking $T=\infty$. Let's adopt all same assumptions for A,f,r,g above, and consider the admissible set

$$\mathscr{A} = \{\alpha:[0,\infty)\to A:\alpha(\cdot)\ measurable\} \tag{195}$$

with $x(\cdot)$ as the unique solution to ODE

$$\begin{cases} x'(s) = f(x(s),\alpha(s)) \\ x(0)=x \end{cases} \tag{196}$$

for fixed control α. In order to ensure that the cost is well-defined on infinite time horizon, let's introduce $\lambda>0$ as continuous-time discount rate and define the cost as

$$C_x[\alpha] = \int_0^\infty e^{-\lambda s}r(x(s),\alpha(s))\,ds \tag{197}$$

and the value function as

$$u(x) = \inf_{\alpha \in \mathscr{A}} C_x[\alpha] \tag{198}$$

note that **the biggest difference is that infinite time horizon problem under the Markovian setting has time-homogeneous value function**.

Remark. *To see this, let's assume that former definition still applies*

$$C_{x,t}[\alpha] = \int_t^\infty e^{-\lambda s} r(x(s), \alpha(s))\, ds \tag{199}$$

and the value function is

$$u(x,t) = \inf_{\alpha \in \mathscr{A}} C_{x,t}[\alpha] \tag{200}$$

with the ODE having initial value condition $x(t) = x$. Now consider $\forall t > 0$,

$$C_{x,t}[\alpha] = \int_t^\infty e^{-\lambda s} r(x(s), \alpha(s))\, ds \tag{201}$$

$$= \int_0^\infty e^{-\lambda(s+t)} r(x(s+t), \alpha(s+t))\, ds \tag{202}$$

where $x'(s) = f(x(s), \alpha(s)), x(t) = x$. However, let's consider another solution $\hat{x}(s)$ to the ODE with fixed control $\hat{\alpha}(s) = \alpha(s+t)$ such that $\hat{x}'(s) = f(\hat{x}(s), \hat{\alpha}(s)), \hat{x}(0) = x$, according to the uniqueness of the solution to the ODE, we immediately know that $\hat{x}(s) = x(s+t)$. So now

$$C_{x,t}[\alpha] = \int_0^\infty e^{-\lambda(s+t)} r(x(s+t), \alpha(s+t))\, ds \tag{203}$$

$$= e^{-\lambda t} \cdot \int_0^\infty e^{-\lambda s} r(\hat{x}(s), \hat{\alpha}(s))\, ds \tag{204}$$

$$= e^{-\lambda t} \cdot C_{x,0}[\hat{\alpha}] \tag{205}$$

and by taking inf on both sides, one would see that

$$u(x,t) = e^{-\lambda t} \cdot u(x,0) \tag{206}$$

so the time t only appears in the discount factor $e^{-\lambda t}$. That's why we only need to consider $u(x,0)$ and denote it as $u(x)$ by taking the time t as 0 by default.

Under all assumptions made above, one can see that u **is bounded and if** $\lambda > Lips(f)$ **then** u **is Lipschitz**. To argue this, one do the similar thing as done in the previous proofs. $\forall x, \hat{x} \in \mathbb{R}^n, \forall \varepsilon > 0$, there exists control

$\hat{\alpha} \in \mathscr{A}$ and the solution $\hat{x}(s)$ to the ODE with fixed control $\hat{\alpha}$ and initial value condition $\hat{x}(0) = \hat{x}$ such that

$$u(\hat{x}) + \varepsilon \geq \int_0^\infty e^{-\lambda s} r(\hat{x}(s), \hat{\alpha}(s))\, ds \tag{207}$$

now by definition,

$$u(x) - u(\hat{x}) \leq u(x) - \int_0^\infty e^{-\lambda s} r(\hat{x}(s), \hat{\alpha}(s))\, ds + \varepsilon \tag{208}$$

$$\leq \int_0^\infty e^{-\lambda s} r(x(s), \hat{\alpha}(s))\, ds - \int_0^\infty e^{-\lambda s} r(\hat{x}(s), \hat{\alpha}(s))\, ds + \varepsilon \tag{209}$$

where $x(s)$ is the solution to the ODE with fixed control $\hat{\alpha}$ and initial value condition $x(0) = x$. So we know that

$$u(x) - u(\hat{x}) \leq \int_0^\infty e^{-\lambda s} [r(x(s), \hat{\alpha}(s)) - r(\hat{x}(s), \hat{\alpha}(s))]\, ds + \varepsilon \tag{210}$$

$$\leq C_r \cdot \int_0^\infty e^{-\lambda s} \cdot ||x(s) - \hat{x}(s)||\, ds + \varepsilon \tag{211}$$

and $||x'(s) - \hat{x}'(s)|| = ||f(x(s), \hat{\alpha}(s)) - f(\hat{x}(s), \hat{\alpha}(s))|| \leq C_f \cdot ||x(s) - \hat{x}(s)||$ so by Grownwall's inequality, we conclude that

$$||x(s) - \hat{x}(s)|| \leq e^{C_f s} \cdot ||x(0) - \hat{x}(0)|| = e^{C_f s} \cdot ||x - \hat{x}|| \tag{212}$$

so the estimates look like

$$u(x) - u(\hat{x}) \leq C_r \cdot ||x - \hat{x}|| \cdot \int_0^\infty e^{(C_f - \lambda)s}\, ds + \varepsilon \tag{213}$$

so when $C_f = Lips(f) < \lambda$, the integral converges and is a constant, that's why u is Lipschitz and is differentiable almost everywhere.

To get the HJBE for such value function $u(x)$, let's plug in

$$u(x, t) = e^{-\lambda t} \cdot u(x, 0) \tag{214}$$

into the HJBE we derived for general optimal control problem to see that

$$u_t(x, t)|_{t=0} = -\lambda \cdot u(x, 0) \tag{215}$$

so

$$-\lambda \cdot u(x, 0) + \inf_{\alpha \in \mathscr{A}} \{f(x, \alpha) \cdot Du + r(x, \alpha)\} = 0 \tag{216}$$

and we get **the HJBE for infinite-horizon optimal control problem**

$$\lambda u - \inf_{\alpha \in \mathscr{A}} \{f(x, \alpha) \cdot Du + r(x, \alpha)\} = 0 \tag{217}$$

for value function $u = u(x)$.

Till now, we have finished the discussion on optimal control problems. In the following context, we will talk about stochastic control problem where the dynamics is not an ODE but an SDE. One would see that the PDE approach is still similar to what we have done here but the probabilistic approach would be very different.

For the following contents, we refer to the book *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications* written by Rene Carmona.

Stochastic Control Problem, PDE Approach

Problem Setting

In the setting of stochastic control, the state process is denoted as $\{X_t\}$, a stochastic process in $\mathbb{R}^d$, generated as the solution to a SDE for given control (action) $\{\alpha_t\}$, which is also a stochastic process. Similar to the deterministic case, let's first specify the set of all admissible controls one can choose from. Note that different from the deterministic case, here we also have to specify the **measurability** of those controls, i.e. one cannot make use of the information that can only be known in the future to determine the best control for the time being.

Let's assume that the control α_t at each fixed time t can take value in A, a subset of a Polish space. Most often, we assume that $A \subset \mathbb{R}^k$ is a compact subset and $\mathscr{A}$ denotes the set of all **admissible controls**, i.e.

$$\mathscr{A} = \{\alpha = \{\alpha_t\} : \forall t \geq 0, \alpha_t \in A\} \tag{218}$$

let's denote α as the whole stochastic process $\{\alpha_t\}$ in the following context. Sometimes, there will be uniform bounded condition added for $\alpha \in \mathscr{A}$ and sometimes we would assume that

$$\mathbb{E}\int_0^T ||\alpha_t||^2 \, dt < \infty \tag{219}$$

, i.e. $\alpha \in L^2([0,T] \times \Omega)$ is in the L^2 Hilbert space of stochastic processes on time interval $[0,T]$. However, those conditions are added as required and there's no standard formulation of the admissible set.

Let's then consider the measurability of admissible controls. Assume that we are **in the finite time horizon case and the time has upper limit** T. Then for each $t \in [0,T]$, when one wants to choose the control, one obviously cannot use all the information of $\{X_t\}_{t\in[0,T]}$ since one cannot make any current decision based on future information. Let's denote $\{\mathcal{I}_t\}$ as a filtration standing for **the information available to the controller at time** t, i.e. $\alpha_t \in \mathcal{I}_t$. There are mainly four different kinds of settings for the measurability conditions of the admissible set.

- **OL (Open Loop)** The setting where $\mathcal{I}_t = \sigma\{X_0\}$ and $\alpha_t = \alpha_t(t, X_0)$.
- **CLPS (Closed Loop Perfect State)** The setting where $\mathcal{I}_t = \sigma\{X_s : s \in [0,t]\}$ and $\alpha_t = \alpha_t(t, \{X_s\}|_{s\in[0,t]})$.
- **MPS (Memoryless Perfect State)** The setting where $\mathcal{I}_t = \sigma\{X_0, X_t\}$ and $\alpha_t = \alpha_t(t, X_0, X_t)$.
- **FPS (Feedback Perfect State/Markovian)** The setting where $\mathcal{I}_t = \sigma\{X_t\}$ and $\alpha_t = \alpha_t(t, X_t)$.

In OL, the information available to the controller at time t is always trivial (only initial state). In MPS, the information available to the controller at time t is the initial state and the current state. In FPS, the information

available to the controller at time t is only the current state but the initial state can not be observed and in CLPS, all history states are known to the controller. One might be able to find that OL is the most specific setting while CLPS is the most general setting. By mentioning Markov games, we take the FPS setting by default.

Now the dynamics of the state process is given by the SDE

$$dX_t = b(t, X_t, \alpha_t)\, dt + \sigma(t, X_t, \alpha_t)\, dB_t \tag{220}$$

where the drift and diffusion coefficient $b : [0,T] \times \mathbb{R}^d \times A \to \mathbb{R}^d, \sigma : [0,T] \times \mathbb{R}^d \times A \to \mathbb{R}^{d\times m}$. So X_t **is a process in** $\mathbb{R}^d$**, the BM** B_t **here is of** m**-dimension** and for each given control α_t one can solve the SDE to know X_t (the choice of action changes the state evolution). For the purpose of simplicity, we want to ensure **the existence and uniqueness of the strong solution to such SDE**. Note that both coefficients depend on the control α_t, so it's natural to make some **additional assumptions to the admissible set** that

$$\mathscr{A} = \left\{ \alpha : \mathbb{E} \int_0^T ||b(t,0,\alpha_t)||^2 + ||\sigma(t,0,\alpha_t)||^2\, dt < \infty \right\} \tag{221}$$

now if we also **assume that** $b(t,x,\alpha), \sigma(t,x,\alpha)$ **are both Lipschitz in** x, then the existence and uniqueness of the strong solution can be guaranteed. Let's denote $X^{t,x,\alpha} = \{X_s^{t,x,\alpha}\}\,|_{s\in[t,T]}$ as the unique solution to the following SDE with initial value condition and given control $\alpha \in \mathscr{A}$

$$\begin{cases} dX_t = b(t, X_t, \alpha_t)\, dt + \sigma(t, X_t, \alpha_t)\, dB_t \\ X_t = x \end{cases} \tag{222}$$

The **cost functional** is defined as

$$J(\alpha) = \mathbb{E}\left[\int_0^T f(s, X_s, \alpha_s)\, ds + g(X_T) \right] \tag{223}$$

consisting of two parts, the running cost and the terminal cost, and we **assume that** $f(t,x,\alpha)$ **is Lipschitz in** x. Now the objective of stochastic control problem is to **find the optimal control** $\alpha = \alpha^*$ **such that it minimizes the cost functional** $J(\alpha)$.

Remark. *Of course, such optimal control α^* does not necessarily exist. To prove the existence, one needs to show that $\mathscr{A}$ is a convex subset and J is convex, l.s.c. with compact level sets so that the existence of the minimum is ensured. However, this does not seem to be very interesting in the scope of our discussion.*

Remark. *To mention the technique of absorbing the running cost and maintaining only the terminal cost, let's consider a new process*

$$Y_t = \int_0^t f(s, X_s, \alpha_s)\, ds \tag{224}$$

so $J(\alpha) = \mathbb{E}\left[Y_T + g(X_T)\right] = \mathbb{E}\tilde{g}(X_T, Y_T)$ *if the function* $\tilde{g}$ *is defined as*

$$\tilde{g}(x, y) = y + g(x) \tag{225}$$

As a result, under the new setting, our state process becomes $\tilde{X}_t = (X_t, Y_t)$*, and the cost functional is* $J(\alpha) = \mathbb{E}\tilde{g}(\tilde{X}_T)$ *with the same set of admissible controls. However, the cost of doing this is that: (i): the increase in the dimension of state process (ii):* Y_t *has dynamics* $dY_t = f(t, X_t, \alpha_t)\,dt$ *that has no diffusion terms.*

At last, we introduce the Hamiltonian of the system. We define the **Hamiltonian** H as

$$H(t, x, y, z, \alpha) = b(t, x, \alpha) \cdot y + \sigma(t, x, \alpha) \cdot z + f(t, x, \alpha) \tag{226}$$

where $\cdot$ means the standard inner product and y, z are **dual variables**. To be specific, for $y \in \mathbb{R}^d, z \in \mathbb{R}^{d\times m}$, $b(t, x, \alpha) \cdot y$ is the inner product between two vectors and $\sigma(t, x, \alpha) \cdot z$ is the inner product between two matrices defined as $\langle A, B\rangle = tr(A^T B)$. In particular, when the control variable α does not appear in the diffusion coefficient σ, we use the **reduced Hamiltonian** defined as

$$\tilde{H}(t, x, y, \alpha) = b(t, x, \alpha) \cdot y + f(t, x, \alpha) \tag{227}$$

which does not depend on the dual variable z. The minimization of Hamiltonian w.r.t. the control α will be done in the later context.

Example

Consider the problem where a single firm facing regulations for pollution permits in time $[0, T]$. The firm has cumulative emissions E_t up to time t generated by the SDE

$$\begin{cases} dE_t = (b_t - \xi_t)\,dt + \sigma_t\,dB_t \\ E_0 = 0 \end{cases} \tag{228}$$

where b_t is the expected rate of emission change if there's no regulation and ξ_t is the rate of abatement chosen by the firm (so it's an action/control). However, the larger rate of abatement the form chooses, the less it can produce, so there is a cost function $c : \mathbb{R} \to \mathbb{R}$ characterizing the cost of lowering the emission. On the other hand, the form can also choose to hold θ_t quantity of pollution permits at time t, with Y_t characterizing the price of each pollution permit (there is an allowance market where firms can trade permits). At last, our goal is to figure out the best control ξ^*, θ^* such that the utility of the firm is maximized for a given utility function U.

Now we make assumptions that b_t, σ_t are adapted and bounded, c is C^1, nondecreasing, strictly convex and $c'(-\infty) = -\infty, c'(+\infty) = +\infty, c(0) = 0$, U is C^1, increasing, strictly concave and $U'(-\infty) = -\infty, U'(+\infty) = +\infty$ (the Inada condition). Note that here BM B_t and E_t are both 1-dimensional.

Let's denote X_T as the total wealth of the company at terminal time T with initial wealth $X_0 = x$, then

$$X_T = x + \int_0^T \theta_t \, dY_t - \int_0^T c(\xi_t) \, dt - E_T Y_T \tag{229}$$

here the second term on RHS stands for the wealth the firm gets in the allowance market by trading permits through time $[0, T]$, the third term on RHS is the cost in production caused by the abatement in the emission, and the last term on RHS is the final cost to eliminate all emissions with permits (there's E_T emissions altogether and each permit costs Y_T). Our goal is to find the optimal control ξ^*, θ^* such that

$$\mathbb{E}U\left(X_T^{\xi^*,\theta^*}\right) = \sup_{(\xi,\theta)\in\mathscr{A}} \mathbb{E}U\left(X_T^{\xi,\theta}\right) \tag{230}$$

the admissible set here only requires the integrability condition

$$\mathbb{E}\int_0^T ||b_t - \xi_t||^2 + ||\sigma_t||^2 \, dt < \infty \tag{231}$$

Let's prove that **the optimal abatement strategy is**

$$\xi_t^* = (c')^{-1}(Y_t) \tag{232}$$

let's rewrite the terminal wealth by replacing E_t

$$E_T Y_T = Y_T \left(\int_0^T (b_t - \xi_t) \, dt + \int_0^T \sigma_t \, dB_t \right) \tag{233}$$

$$= Y_T \left(\int_0^T b_t \, dt + \int_0^T \sigma_t \, dB_t \right) - Y_T \int_0^T \xi_t \, dt \tag{234}$$

note that

$$\int_0^T (Y_T - Y_t)\xi_t \, dt = \int_0^T \left(\int_t^T dY_s \right) \xi_t \, dt \tag{235}$$

$$= \int_0^T \left(\int_0^s \xi_t \, dt \right) dY_s \tag{236}$$

plug in to find

$$X_T = x - \int_0^T c(\xi_t)\,dt + \int_0^T \theta_t\,dY_t - Y_T\left(\int_0^T b_t\,dt + \int_0^T \sigma_t\,dB_t\right) + \int_0^T Y_t\xi_t\,dt + \int_0^T \left(\int_0^s \xi_t\,dt\right)\,dY_s \tag{237}$$

$$= x - \int_0^T [c(\xi_t) - Y_t\xi_t]\,dt + \int_0^T \theta_t\,dY_t - Y_T\int_0^T b_t\,dt - Y_T\int_0^T \sigma_t\,dB_t + \int_0^T \left(\int_0^t \xi_s\,ds\right)\,dY_t \tag{238}$$

$$= x - \int_0^T [c(\xi_t) - Y_t\xi_t]\,dt + \int_0^T \left[\theta_t + \int_0^t \xi_s\,ds\right]\,dY_t - Y_T\int_0^T b_t\,dt - Y_T\int_0^T \sigma_t\,dB_t \tag{239}$$

call the first two terms on RHS as B_T^ξ and the remaining terms on RHS as $A_T^{\tilde\theta}$ with the new control defined as $\tilde\theta_t = \theta_t + \int_0^t \xi_s\,ds$, so now

$$\begin{cases} B_T^\xi = x - \int_0^T [c(\xi_t) - Y_t\xi_t]\,dt \\ A_T^{\tilde\theta} = \int_0^T \tilde\theta_t\,dY_t - Y_T\int_0^T b_t\,dt - Y_T\int_0^T \sigma_t\,dB_t \end{cases} \tag{240}$$

those two parts are separated such that B_T^ξ has nothing to do with $\tilde\theta$ and $A_T^{\tilde\theta}$ has nothing to do with ξ if we see $\xi, \tilde\theta$ as two independent controls (although they are actually not since the definition of $\tilde\theta$ contains ξ). However, we can notice that when (θ, ξ) traverses through the admissible set $\mathscr{A}$, $(\tilde\theta, \xi)$ also traverses through the admissible set $\mathscr{A}$ and vice versa. As a result,

$$\sup_{(\xi,\theta)\in\mathscr{A}} \mathbb{E}U\left(X_T^{\xi,\theta}\right) = \sup_{(\xi,\tilde\theta)\in\mathscr{A}} \mathbb{E}U\left(X_T^{\xi,\theta}\right) \tag{241}$$

$$= \sup_{(\xi,\tilde\theta)\in\mathscr{A}} \mathbb{E}U\left(A_T^{\tilde\theta} + B_T^{\xi}\right) \tag{242}$$

$$= \sup_{\tilde\theta\in\mathscr{A}}\sup_{\xi\in\mathscr{A}} \mathbb{E}U\left(A_T^{\tilde\theta} + B_T^{\xi}\right) \tag{243}$$

so the optimal abatement rate ξ_t^* is the ξ_t that maximizes B_T^ξ (under the maximization of xi, $A_T^{\tilde\theta}$ is a constant and note that the utility is increasing, under the assumptions, the maximum exists and $(c')^{-1}$ exists).

$$\xi_t^* = \arg\max_{\xi_t}\left\{x - \int_0^T [c(\xi_t) - Y_t\xi_t]\,dt\right\} \tag{244}$$

$$\xi_t^* = (c')^{-1}(Y_t) \in \mathcal{I}_t \tag{245}$$

Remark. *The trick applied in this example is to set up a new control such that the wealth is a* ***separable*** *and argue that the old set of controls traverse through the admissible set if and only if the new set of controls traverse through the admissible set. As a result, the new controls can be seen as independent controls and two maximization can be dealt with separately.*

Note that one has to verify that the optimal control one get satisfies measurability requirements. For example, in this example, we are taking the Markovian setting so ξ_t^* can only depend on the value of all observable processes

at time t.

Remark. *To get the intuition of such optimal abatement rate, it's telling us that on observing the price of the pollution permit Y_t at time t, the firm shall always make sure that* ***the marginal production cost*** $c'(\xi_t)$ ***is equal to the marginal emission cost*** Y_t*. In economics, it's rational to only compare the marginal so we would get the same conclusion from intuition.*

Example

Let's use a slightly different example to illustrate the same trick once again. Now still consider a single firm with regulation for emission allowances. Now the firm produce a source with price P_t following BS model such that

$$\frac{dP_t}{P_t} = \mu(P_t)\,dt + \sigma(P_t)\,dB_t \tag{246}$$

at each time the form can choose its rate of production q_t with production costs $c(q_t)$. Similar to the example above, the form has to buy permits for all the emission it produces. The price of the permit is denoted as Y_t and now the cumulative emission until time t, denoted E_t, is proportional to the production amount until time t for fixed $\varepsilon > 0$

$$E_t = \varepsilon Q_t, E_0 = 0 \tag{247}$$

$$Q_t = \int_0^t q_s\,ds \tag{248}$$

The firm has to decide θ_t, the quantity of permit to hold and q_t, the rate of production at time t, so the control is made up of the pair (θ_t, q_t). Let's still use X_T for the total wealth of the firm at time T with initial wealth $X_0 = x$, then

$$X_T = x + \int_0^T P_t q_t\,dt - \int_0^T c(q_t)\,dt + \int_0^T \theta_t\,dY_t - E_T Y_T \tag{249}$$

$$= x + \int_0^T P_t q_t\,dt - \int_0^T c(q_t)\,dt + \int_0^T \theta_t\,dY_t - \varepsilon Q_T Y_T \tag{250}$$

the utility function U is provided and we wish to find optimal control θ^*, q^* to maximize the expected terminal utility

$$\mathbb{E}U\left(X_T^{\theta^*,q^*}\right) = \sup_{\theta,q\in\mathscr{A}} \mathbb{E}U\left(X_T^{\theta,q}\right) \tag{251}$$

Likewise, we make the following assumptions that μ, σ are C^1 with bounded derivatives, cost function c is C^1 and strictly convex, satisfies the Inada condition, i.e. $c'(-\infty) = -\infty, c'(+\infty) = +\infty$, and the utility function U is C^1, increasing, strictly concave and satisfy the Inada condition, i.e. $U'(-\infty) = -\infty, U'(+\infty) = +\infty$. The admissible set of controls only have adaptability and integrability conditions as stated in the previous context.

Now **the optimal production strategy** should be

$$q_t^* = (c')^{-1}(P_t - \varepsilon Y_t) \tag{252}$$

since the marginal cost of producing is $c'(q_t) + \varepsilon Y_t$ (the rising in cost and the need to buy permit for increased emission) and the marginal profit of producing is P_t. By previous explanations on the intuitions, it's easy to see that optimal control is achieved when **the marginals are equal**.

Let's apply the same trick of separating two control variables here by transforming the term $Q_T Y_T$ using Ito formula (note that Q_t has finite variation)

$$d(Q_t Y_t) = Q_t \, dY_t + Y_t \, dQ_t \tag{253}$$

$$Q_T Y_T = \int_0^T Q_t \, dY_t + \int_0^T Y_t \, dQ_t \tag{254}$$

$$= \int_0^T \left(\int_0^t q_s \, ds \right) dY_t + \int_0^T Y_t q_t \, dt \tag{255}$$

plug into the expression for X_T to see that

$$X_T = x + \int_0^T \left[(P_t - \varepsilon Y_t) q_t - c(q_t) \right] dt + \int_0^T \theta_t \, dY_t - \varepsilon \int_0^T \left(\int_0^t q_s \, ds \right) dY_t \tag{256}$$

$$= x + \int_0^T \left[(P_t - \varepsilon Y_t) q_t - c(q_t) \right] dt + \int_0^T \left[\theta_t - \varepsilon \int_0^t q_s \, ds \right] dY_t \tag{257}$$

denote the new control $\tilde{\theta}_t = \theta_t - \varepsilon \int_0^t q_s \, ds$ to find that

$$X_T = x + \int_0^T \left[(P_t - \varepsilon Y_t) q_t - c(q_t) \right] dt + \int_0^T \tilde{\theta}_t \, dY_t \tag{258}$$

and separate it into two parts

$$\begin{cases} A_T^{\tilde{\theta}} = \int_0^T \tilde{\theta}_t \, dY_t \\ B_T^q = x + \int_0^T \left[(P_t - \varepsilon Y_t) q_t - c(q_t) \right] dt \end{cases} \tag{259}$$

note that when (θ, q) traverse through the admissible set, so does $(\tilde{\theta}, q)$, so

$$\sup_{\theta, q \in \mathscr{A}} \mathbb{E} U \left(X_T^{\theta, q} \right) = \sup_{\tilde{\theta}, q \in \mathscr{A}} \mathbb{E} U \left(X_T^{\theta, q} \right) \tag{260}$$

$$= \sup_{\tilde{\theta}, q \in \mathscr{A}} \mathbb{E} U \left(A_T^{\tilde{\theta}} + B_T^q \right) \tag{261}$$

$$= \sup_{\tilde{\theta} \in \mathscr{A}} \sup_{q \in \mathscr{A}} \mathbb{E} U \left(A_T^{\tilde{\theta}} + B_T^q \right) \tag{262}$$

and the optimal production strategy will be attained when B_T^q attains its maximum (since now $\tilde{\theta}, q$ are considered as independent controls and utility function is increasing, with B_T^q only depending on q and $A_T^{\tilde{\theta}}$ only depending on $\tilde{\theta}$)

$$q_t^* = \arg\max_{q_t} \left\{ x + \int_0^T \left[(P_t - \varepsilon Y_t) q_t - c(q_t) \right] dt \right\} \tag{263}$$

$$q_t^* = (c')^{-1}(P_t - \varepsilon Y_t) \in \mathcal{I}_t \tag{264}$$

we can check that at time t, under the Markovian setting, P_t, Y_t are observable to the controller so $q_t^* \in \mathcal{I}_t$ satisfies the measurability condition.

Remark. *One might hope to use the similar technique to figure out the optimal quantity of permit to hold since*

$$\tilde{\theta}_t^* = \arg\max_{\tilde{\theta}_t} \left\{ \int_0^T \tilde{\theta}_t \, dY_t \right\} \tag{265}$$

however, one may find that by setting the derivative w.r.t. $\tilde{\theta}_t$ as 0, one cannot find the optimal $\tilde{\theta}_t$ because of the measurability issue (Y_T is not known at time t). So one cannot find an admissible control from this problem as we have done for q_t^* and one has to consider instead

$$\tilde{\theta}_t^* = \arg\max_{\tilde{\theta}_t} \left\{ \mathbb{E} U \left(\int_0^T \tilde{\theta}_t \, dY_t + B_T^{q^*} \right) \right\} \tag{266}$$

so by taking the derivative, one would get

$$\mathbb{E}\left[U' \left(\int_0^T \tilde{\theta}_t^* \, dY_t + B_T^{q^*} \right) \cdot (Y_T - Y_0) \right] = 0 \tag{267}$$

and we will see that $\int_0^T \tilde{\theta}_t^* \, dY_t$ has something to do with the process Y after time t, so the optimal control $\tilde{\theta}_t^*$ is hard to figure out (especially to ensure the measurability). This is telling us that the two examples shown above have easy and intuitive optimal control solution because of the simplicity of the example and generally it's hard to find the ***admissible*** *optimal control.*

Value Function, HJBE and the PDE Approach

Now the PDE approach to stochastic control focuses on applying the **dynamic programming principle**, setting up **value functions** and deriving **HJBE of the value functions** to solve the problem.

Let's assume that we are under **the Markovian setting** and the cost after time t sticking to control α with initial value condition $X_t = x$ is denoted as

$$J(t,x,\alpha) = \mathbb{E}\left[\int_t^T f(s,X_s,\alpha_s)\,ds + g(X_T)\Big|X_t = x\right] \tag{268}$$

(note that here X_s is the solution to the SDE for given control α) and let's denote $\mathscr{A}_t$ as the admissible set of controls α over time interval $[t,T]$ with measurability and integrability conditions

$$\mathbb{E}\int_t^T ||b(s,X_s,\alpha_s)||^2 + ||\sigma(s,X_s,\alpha_s)||^2\,ds < \infty \tag{269}$$

and the HJB value function is defined as

$$v(t,x) = \inf_{\alpha\in\mathscr{A}_t} J(t,x,\alpha) \tag{270}$$

the lowest possible cost with initial condition $X_t = x$ over all admissible controls. Since we are planning to find the HJBE that such value function is satisfying, it's natural to ask whether the value function is differentiable at all points and what conditions are needed such that a PDE for the value function can be constructed.

Remark. *We can also denote* $J(t,x,\alpha) = \mathbb{E}\left[\int_t^T f(s,X_s^{t,x,\alpha},\alpha_s)\,ds + g(X_T^{t,x,\alpha})\right]$ *and* $X_s^{t,x,\alpha}$ *denotes the value of the solution to the SDE at time* s *with fixed control* α *and initial value condition* $X_t = x$*. It's easy to see that those two definitions are equivalent.*

Example: Regularity Issues of Value Function

Let's look at an example where $d = 1$, i.e. X_t is 1-dimensional process with $A = [-1,1], \sigma = 0, f = 0, b(t,x,\alpha) = \alpha, g(x) = -x^2$. So now we know that

$$dX_t = \alpha_t\,dt \tag{271}$$

$$J(t,x,\alpha) = \mathbb{E}\left[-X_T^2\Big|X_t = x\right] \tag{272}$$

now conditioning on $X_t = x$, we know $X_s = x + \int_t^s \alpha_r \, dr$ so

$$J(t,x,\alpha) = -\left(x + \int_t^T \alpha_r \, dr\right)^2 \tag{273}$$

$$v(t,x) = \inf_{\alpha \in \mathscr{A}_t} \left\{ -\left(x + \int_t^T \alpha_r \, dr\right)^2 \right\} \tag{274}$$

it's clear that if $x \geq 0$ then since $\forall s \in [t,T], \alpha_s \in A = [-1,1]$, the inf is attained when $\forall s \in [t,T], \alpha_s = 1$ and if $x \leq 0$ then the inf is attained when $\forall s \in [t,T], \alpha_s = -1$. So the value function is

$$v(t,x) = \begin{cases} -(x+T-t)^2 & x \geq 0 \\ -(x-T+t)^2 & x < 0 \end{cases} \tag{275}$$

which is continuous but not differentiable at 0 even under this extremely simple setting.

Example: Value Function as Convex Envelope

Consider another example where $d = k = 1, b = 0$ so X_t and the control α_t are still 1-dimensional and $\sigma(t,x,\alpha) = \alpha$ with $f = 0$, g continuous and bounded from above and $A = \mathbb{R}$. So now

$$dX_t = \alpha_t \, dB_t \tag{276}$$

$$J(t,x,\alpha) = \mathbb{E}\left[g(X_T) | X_t = x\right] \tag{277}$$

since we will be varying the time variable from t to $t+h$ a little bit for $h \to 0^+$ to set up a PDE for the value function (as shown later), it's natural to see that we would want to apply Ito formula for the value function v, so whether $v \in C^{1,2}$ is then a problem of our concern. However, in this example, we can show that **if $v \in C^{1,2}$, then v is independent of time t and is equal to the convex envelope g^{**} of g** (g^{**} is the double Frenchel conjugate of g, it can be proved that it's the convex envelope).

Now that $X_s|_{X_t = x} = x + \int_t^s \alpha_r \, dB_r$ and note that $\{\alpha_r, r \in [t,s]\} \in \mathscr{A}_t$ satisfies the integrability condition that $\mathbb{E}\int_t^T \alpha_s^2 \, ds < \infty$, it's obvious that $X_s|_{X_t=x}$ is a MG in s for any given control $\alpha \in \mathscr{A}_t$. Then we find

$$v(t,x) = \inf_{\alpha \in \mathscr{A}_t} \mathbb{E}\left[g(X_T) | X_t = x\right] \tag{278}$$

$$\geq \inf_{\alpha \in \mathscr{A}_t} \mathbb{E}\left[g^{**}(X_T) | X_t = x\right] \tag{279}$$

$$\geq \inf_{\alpha \in \mathscr{A}_t} g^{**}(\mathbb{E}(X_T | X_t = x)) \tag{280}$$

$$= \inf_{\alpha \in \mathscr{A}_t} g^{**}(x) \tag{281}$$

$$= g^{**}(x) \tag{282}$$

by applying Jensen's inequality, so v has to be larger than the convex envelope for $\forall x \in \mathbb{R}$.

For the other side, if $v \in C^{1,2}$, Ito formula holds and

$$v(t+h, X_{t+h}) = v(t, X_t) + \int_t^{t+h} \partial_t v(s, X_s)\, ds + \int_t^{t+h} \partial_x v(s, X_s)\, dX_s + \frac{1}{2}\int_t^{t+h} \partial_{xx} v(s, X_s)\, d\langle X, X\rangle_s \tag{283}$$

$$= v(t, X_t) + \int_t^{t+h} \left(\partial_t + \frac{\alpha_s^2}{2}\partial_{xx}\right) v(s, X_s)\, ds + \int_t^{t+h} \partial_x v(s, X_s) \cdot \alpha_s\, dB_s \tag{284}$$

assume that the last stochastic integral is a MG, we can find that

$$\mathbb{E}\left[v(t+h, X_{t+h})|X_t = x\right] = v(t,x) + \mathbb{E}\left[\int_t^{t+h} \left(\partial_t + \frac{\alpha_s^2}{2}\partial_{xx}\right) v(s, X_s)\, ds \Big| X_t = x\right] \tag{285}$$

by applying the property of value function that $\mathbb{E}\left[v(t+h, X_{t+h})|X_t = x\right] - v(t,x) \geq 0$ (which will be proved below), we get that

$$\forall h > 0, \mathbb{E}\left[\int_t^{t+h} \left(\partial_t + \frac{\alpha_s^2}{2}\partial_{xx}\right) v(s, X_s)\, ds \Big| X_t = x\right] \geq 0 \tag{286}$$

dividing both sides by h and take $h \to 0^+$ to find

$$\forall (t, x, \alpha) \in [0,T] \times \mathbb{R} \times \mathbb{R}, \left(\partial_t + \frac{\alpha_t^2}{2}\partial_{xx}\right) v(t,x) \geq 0 \tag{287}$$

by taking $\alpha_t = 0$ as the constant control, we find that $\partial_t v \geq 0$, so for each fixed $x \in \mathbb{R}$, v is always increasing w.r.t. time t. Also note that $\partial_{xx} v \geq 0$ must hold since otherwise we can always take α_t to be large enough such that the inequality above fails, so v has to be convex in x. By Fatou's lemma and continuity of g,

$$\forall 0 \leq t < T, v(t,x) \leq \lim_{s \nearrow T} v(s,x) \tag{288}$$

$$= \lim_{s \nearrow T} \inf_{\alpha \in \mathscr{A}_s} \mathbb{E}\left[g(X_T)|X_s = x\right] \tag{289}$$

$$\leq \overline{\lim}_{s \nearrow T} \mathbb{E}\left[g(X_T)|X_s = x\right] \tag{290}$$

$$\leq \mathbb{E}\left[\overline{\lim}_{s \nearrow T} g(X_T^{s,x})\right] \tag{291}$$

$$= g(x) \tag{292}$$

where $X_T^{s,x}$ denotes the solution to the SDE with initial value condition $X_s = x$. So for any fixed time t, v is always a convex function dominated by g, $v(t,x) \leq g^{**}(x)$ by the maximality of convex envelope, and we conclude that

$$v(t,x) = g^{**}(x) \tag{293}$$

actually has nothing to do with t.

Remark. *The inequality* $\mathbb{E}\left[v(t+h, X_{t+h})|X_t = x\right] - v(t,x) \geq 0$ *we are using here is a natural property of the value function. The meaning is that since value function is already the optimal cost among all admissible controls based on the observation of the initial value condition* $X_t = x$, ***if one adopts the control that is the optimal control at time*** t ***in time interval*** $[t, t+h)$ ***but follows the optimal control at time*** $t+h$ ***in time interval*** $[t+h, T]$**,** ***then such strategy cannot be better than following the optimal control at time*** t ***in time interval*** $[t, T]$*. This would be explained in a later context.*

Remark. *This example shows that we can easily make up a "bad" value function. Let's pick* g *continuous and upper bounded on* $\mathbb{R}$ *with the convex envelope* g^{**} *being not* C^2*, then obviously* $v(t,x)$ *for this stochastic control problem would not be* $C^{1,2}$*. For example, consider* $g(x) = -|x|$ *then* $g^{**} = g = v$*.*

However, as proved in the deterministic case for HJE, when $f = 0$ and g is Lipschitz, we can make sure that v is Lipschitz in x for fixed time $t \in [0,T]$ and when A is bounded one can also get some estimates on $|v(t,x) - v(\hat{t}, x)|$. Since **Lipschitz functions are almost everywhere differentiable**, this makes it possible for us to set up a PDE for the value function. We only prove the Lipschitz property here for simplicity.

Theorem 11. ***(Lipschitz Value Function in*** x***)*** *When* $f = 0$ *and* g *is Lipschitz, the value function* v *is Lipschitz in* x *for fixed time* $t \in [0,T]$*.*

Proof. Fix time $t \in [0,T]$ and consider $\forall x, \hat{x} \in \mathbb{R}^d$, by the definition of value function, $\forall \varepsilon > 0, \exists \hat{\alpha}, v(t,\hat{x}) + \varepsilon \geq \mathbb{E}\left[g(X_T)|X_t = \hat{x}\right]$ with the X_t as the solution to the SDE with fixed control $\hat{\alpha}$

$$v(t,x) - v(t,\hat{x}) \leq \inf_{\alpha \in \mathscr{A}_t} \mathbb{E}[g(X_T)|X_t = x] - \mathbb{E}\left[g(X_T)|X_t = \hat{x}\right] + \varepsilon \tag{294}$$

$$\leq \mathbb{E}g(X_T^{t,x,\hat{\alpha}}) - \mathbb{E}g(X_T^{t,\hat{x},\hat{\alpha}}) + \varepsilon \tag{295}$$

$$\leq Lips(g) \cdot \mathbb{E}|X_T^{t,x,\hat{\alpha}} - X_T^{t,\hat{x},\hat{\alpha}}| + \varepsilon \tag{296}$$

let's consider $h(s) = \mathbb{E}|X_s^{t,x,\hat{\alpha}} - X_s^{t,\hat{x},\hat{\alpha}}|$, then

$$h(s) = \mathbb{E}\left| x - \hat{x} + \int_t^s \left[b(r, X_r^{t,x,\hat{\alpha}}, \hat{\alpha}_r) - b(r, X_r^{t,\hat{x},\hat{\alpha}}, \hat{\alpha}_r)\right] dr + \int_t^s \left[\sigma(r, X_r^{t,x,\hat{\alpha}}, \hat{\alpha}_r) - \sigma(r, X_r^{t,\hat{x},\hat{\alpha}}, \hat{\alpha}_r)\right] dr \right| \tag{297}$$

$$\leq |x - \hat{x}| + (Lips(b) + Lips(\sigma)) \cdot \mathbb{E}\int_t^s |X_r^{t,x,\hat{\alpha}} - X_r^{t,\hat{x},\hat{\alpha}}|\, dr \tag{298}$$

$$= |x - \hat{x}| + (Lips(b) + Lips(\sigma)) \cdot \int_t^s h(r)\, dr \tag{299}$$

since we have assumed that b, σ are both Lipschitz. By Grownwall's inequality,

$$\forall s \in [t,T], h(s) \leq |x - \hat{x}| \cdot e^{C(s-t)} \tag{300}$$

so we conclude that $v(t,x) - v(t,\hat{x}) \leq C \cdot h(T) + \varepsilon \leq C' \cdot |x - \hat{x}|$ which completes half of the proof. The other side can be done similarly. □

Dynamic Programming Principle (DPP)

Theorem 12. ***(Dynamic Programming Principle)*** *If the value function v is continuous, then for any initial value condition (t,x), and any stopping time τ that takes values in $[t,T]$,*

$$v(t,x)=\inf_{\alpha\in\mathscr{A}_t}\mathbb{E}\left[\int_t^\tau f(s,X_s^{t,x,\alpha},\alpha_s)\,ds+v(\tau,X_\tau^{t,x,\alpha})\right] \tag{301}$$

Remark. *One may find that this theorem is an analogue of the optimality condition we have mentioned above for deterministic optimal control problem. The meaning of the theorem is that the best control α at time t can be found by minimizing the sum of two parts: (i): the contribution of cost in time interval $[t,\tau]$ sticking to control α (ii): the contribution of cost in time interval $[\tau,T]$ sticking to the optimal control at time τ. So this theorem is showing us* ***the time consistency condition for the value function****.*

Note that the difference of this property in deterministic control and stochastic control lies in the fact that: (i): we are taking inf of an expectation since the cost is actually random (ii): the deterministic perturbed time $t+h$ can be replaced by any stopping time that takes values in $[t,T]$, allowing us to have more freedom.

Proof. Notice that $\forall\alpha\in\mathscr{A}_t, v(t,x)\le J(t,x,\alpha)$ so let us write $J(t,x,\alpha)$ in terms of the stopping time $\forall\tau\in\tau_{t,T}$ (where $\tau_{t,T}$ denotes the set of all stopping time that takes values in $[t,T]$)

$$\begin{aligned}
J(t,x,\alpha)&=\mathbb{E}\left[\int_t^T f(s,X_s,\alpha_s)\,ds+g(X_T)\Big|X_t=x\right] &(302)\\
&=\mathbb{E}\left[\mathbb{E}\left(\int_t^T f(s,X_s,\alpha_s)\,ds+g(X_T)\Big|\mathscr{F}_\tau\right)\Big|X_t=x\right] &(303)\\
&=\mathbb{E}\left[\int_t^\tau f(s,X_s,\alpha_s)\,ds+\mathbb{E}\left(\int_\tau^T f(s,X_s,\alpha_s)\,ds+g(X_T)\Big|\mathscr{F}_\tau\right)\Big|X_t=x\right] &(304)\\
&=\mathbb{E}\left[\int_t^\tau f(s,X_s,\alpha_s)\,ds+J(\tau,X_\tau,\alpha)\Big|X_t=x\right] &(305)
\end{aligned}$$

by tower property. Now replace the J inside the condition expectation with value function v to find

$$J(t,x,\alpha)\ge\mathbb{E}\left[\int_t^\tau f(s,X_s,\alpha_s)\,ds+v(\tau,X_\tau)\Big|X_t=x\right] \tag{306}$$

and take inf on both sides w.r.t. $\tau\in\tau_{t,T}$, take inf on both sides w.r.t. control $\alpha\in\mathscr{A}_t$ to find

$$v(t,x)\ge\inf_{\alpha\in\mathscr{A}_t}\inf_{\tau\in\tau_{t,T}}\mathbb{E}\left[\int_t^\tau f(s,X_s^{t,x,\alpha},\alpha_s)\,ds+v(\tau,X_\tau^{t,x,\alpha})\right] \tag{307}$$

On the other hand, since value function is the inf of cost, $\forall\varepsilon>0,\forall\tau\in\tau_{t,T}$, there exists **$\varepsilon$-optimal strategy** $\alpha^\varepsilon\in\mathscr{A}_\tau$ such that

$$v(\tau,X_\tau)+\varepsilon\ge J(\tau,X_\tau,\alpha^\varepsilon) \tag{308}$$

in order to argue that the value function has some upper bound, let's figure out the best control we are able to design so far. Now α^ε is nearly the best control at time τ, so we would expect to see that sticking to any current control $\alpha \in \mathscr{A}_t$ until stopping time τ and switch to the nearly best control α^ε after time τ would be a good strategy (Note that here we are **switching to the best strategy at time** τ since the admissible control $\alpha^\varepsilon \in \mathscr{A}_\tau$ should satisfy the measurability condition under the Markov setting that $\alpha^\varepsilon = \alpha^\varepsilon(\tau, X_\tau) \in \mathscr{F}_\tau$ so there's no way to know α^ε before time τ). So we construct

$$\hat{\alpha}_s = \begin{cases} \alpha_s & s \in [t, \tau] \\ \alpha_s^\varepsilon & s \in [\tau, T] \end{cases} \in \mathscr{A}_t \tag{309}$$

and

$$\begin{aligned} \forall \alpha \in \mathscr{A}_t, v(t,x) &\leq J(t,x,\hat{\alpha}) && (310)\\ &= \mathbb{E}\left[\int_t^T f(s, X_s^{\hat{\alpha}}, \hat{\alpha}_s)\, ds + g(X_T^{\hat{\alpha}}) \Big| X_t = x\right] && (311)\\ &= \mathbb{E}\left[\int_t^\tau f(s, X_s^{\alpha}, \alpha_s)\, ds + \int_\tau^T f(s, X_s^{\hat{\alpha}}, \hat{\alpha}_s)\, ds + g(X_T^{\hat{\alpha}}) \Big| X_t = x\right] && (312)\\ &= \mathbb{E}\left[\int_t^\tau f(s, X_s^{\alpha}, \alpha_s)\, ds + J(\tau, X_\tau, \hat{\alpha}) \Big| X_t = x\right] && (313)\\ &\leq \mathbb{E}\left[\int_t^\tau f(s, X_s^{\alpha}, \alpha_s)\, ds + v(\tau, X_\tau) \Big| X_t = x\right] + \varepsilon && (314) \end{aligned}$$

by first taking the sup w.r.t. $\tau \in \tau_{t,T}$ on both sides and then the inf on both sides w.r.t. control $\alpha \in \mathscr{A}_t$ to find

$$v(t,x) \leq \inf_{\alpha \in \mathscr{A}_t} \sup_{\tau \in \tau_{t,T}} \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\alpha}, \alpha_s)\, ds + v(\tau, X_\tau^{t,x,\alpha})\right] \tag{315}$$

Combining two inequalities to see the DPP

$$v(t,x) = \inf_{\alpha \in \mathscr{A}_t} \sup_{\tau \in \tau_{t,T}} \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\alpha}, \alpha_s)\, ds + v(\tau, X_\tau^{t,x,\alpha})\right] = \inf_{\alpha \in \mathscr{A}_t} \inf_{\tau \in \tau_{t,T}} \mathbb{E}\left[\int_t^\tau f(s, X_s^{t,x,\alpha}, \alpha_s)\, ds + v(\tau, X_\tau^{t,x,\alpha})\right] \tag{316}$$

□

Remark. *We have actually shown that* ***the selection of stopping time has no impact on the value function****, so any stopping time $\tau \in \tau_{t,T}$ works. This is because we are first fixing the stopping time and then select a good enough control,* ***the inf w.r.t. control α has already taken the stopping time into consideration!***

HJBE of Stochastic Control Problem

Taking the stopping time in DPP as the trivial one $\tau = t + h$ such that $h > 0, t + h \leq T$, one would get the following HJBE for the stochastic control problem. Let's denote L^α as the **infinitesimal generator** of the diffusion process X_t for fixed control α. Then from stochastic calculus, we know that

$$L^\alpha f(x) = b(t,x,\alpha) \cdot \nabla_x f(x) + \frac{1}{2} Tr\left(\sigma(t,x,\alpha) \cdot \sigma^T(t,x,\alpha) \cdot \nabla_x^2 f(x)\right) \tag{317}$$

where $b \in \mathbb{R}^d, \sigma \in \mathbb{R}^{d\times m}$ are drift and diffusion coefficients of the dynamics, $\nabla_x f$ is the gradient of f w.r.t. variable x and $\nabla_x^2 f$ is the Hessian of f w.r.t. variable x.

Theorem 13. ***(HJBE of Stochastic Control Problem)*** *Assume that $v \in C^{1,2}([0,T] \times \mathbb{R}^d)$ and $f \in C([0,T] \times \mathbb{R}^d \times A)$ for each fixed control $\alpha \in \mathscr{A}$ and assume the existence of the optimal control $\alpha^* \in \mathscr{A}$. Then*

$$\forall (t,x) \in [0,T] \times \mathbb{R}^d, \partial_t v(t,x) + \inf_{\alpha \in \mathscr{A}} \{L^\alpha v(t,x) + f(t,x,\alpha)\} = 0 \tag{318}$$

Proof. By taking the stopping time in DPP as the trivial one $\tau = t + h$ such that $h > 0, t + h \leq T$, we find that

$$v(t,x) = \inf_{\alpha \in \mathscr{A}_t} \mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x,\alpha}, \alpha_s)\, ds + v(t+h, X_{t+h}^{t,x,\alpha})\right] \tag{319}$$

apply Ito formula to see

$$v(t+h, X_{t+h}^{t,x,\alpha}) = v(t, X_t^{t,x,\alpha}) + \int_t^{t+h} \partial_t v(s, X_s^{t,x,\alpha})\, ds + \int_t^{t+h} \partial_x v(s, X_s^{t,x,\alpha})\, dX_s^{t,x,\alpha} \tag{320}$$

$$+ \frac{1}{2}\int_t^{t+h} \partial_{xx} v(s, X_s^{t,x,\alpha})\, d\langle X^{t,x,\alpha}, X^{t,x,\alpha}\rangle_s \tag{321}$$

$$= v(t,x) + \int_t^{t+h} (\partial_t + L^\alpha)\, v(s, X_s^{t,x,\alpha})\, ds + \int_t^{t+h} \partial_x v(s, X_s^{t,x,\alpha}) \cdot \sigma(s, X_s^{t,x,\alpha}, \alpha_s)\, dB_s \tag{322}$$

assume that the last stochastic integral is a MG, we can find that

$$\mathbb{E}\left[v(t+h, X_{t+h}) | X_t = x\right] = v(t,x) + \mathbb{E}\left[\int_t^{t+h} (\partial_t + L^\alpha)\, v(s, X_s^\alpha)\, ds \Big| X_t = x\right] \tag{323}$$

by noticing that from DPP we have

$$\forall \alpha \in \mathscr{A}_t, v(t,x) \leq \mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x,\alpha}, \alpha_s)\, ds + v(t+h, X_{t+h}^{t,x,\alpha})\right] \tag{324}$$

$$= v(t,x) + \mathbb{E}\left[\int_t^{t+h} (\partial_t + L^\alpha)\, v(s, X_s^\alpha) + f(s, X_s^\alpha, \alpha_s)\, ds \Big| X_t = x\right] \tag{325}$$

dividing both sides by h and apply the intermediate value theorem for integral, we get (note that the diffusion process

X_t is chosen as the version with continuous sample path and f is continuous)

$$\forall \alpha \in \mathscr{A}_t, 0 \leq (\partial_t + L^\alpha)v(t,x) + f(t,x,\alpha) \tag{326}$$

taking inf on both sides to see

$$\inf_{\alpha \in \mathscr{A}_t} \{(\partial_t + L^\alpha)v(t,x) + f(t,x,\alpha)\} \geq 0 \tag{327}$$

The equality directly comes from the assumption that the optimal control $\alpha^* \in \mathscr{A}_t$ exists and can attain the inf in the value function. As a result, the inequality in DPP becomes equality and

$$(\partial_t + L^{\alpha^*})v(t,x) + f(t,x,\alpha^*) = 0 \tag{328}$$

that's why we get the HJBE

$$\inf_{\alpha \in \mathscr{A}_t} \{(\partial_t + L^\alpha)v(t,x) + f(t,x,\alpha)\} = 0 \tag{329}$$

□

Remark. *Alike the forward/backward Kolmogorov's equations, HJBE is actually a* ***characterization*** *of the value function in that if the HJBE holds and the value function has specific regularity arguments, then the solution to the HJBE must be the value function. However, we do not provide the proof here and just mention this fact since it's not useful in the stochastic control setting.*

Remark. *Now one can see why the argument we have made in the example above that* $\mathbb{E}[v(t+h, X_{t+h})|X_t = x] \geq v(t,x)$ *is true. This is just a simple corollary of the DPP.*

How to Solve a Stochastic Control Problem with PDE Approach

Now that we have described the main tools for the PDE approach to solve stochastic control problems. However, our description is not very organized since we have made a lot of different assumptions within the discussions. Let's collect all non-trivial assumptions we have made so far and present a systematic way for the PDE approach.

- Assume the value function $v \in C^{1,2}$ (in order to apply Ito's formula)
- Assume that the stochastic integral in the calculation is a true MG (in order to ignore it after taking expectation)
- Assume that the optimal control $\alpha^* \in \mathscr{A}$ always exists (in order to turn the inequality into equality in HJBE)
- Assume that for each initial value pair (t, x), there always exists $\hat{\alpha} = \hat{\alpha}(t, x)$ that minimizes $L^{\alpha}v(t, x) + f(t, x, \alpha)$ (in order to make the HJBE easier to solve)
- After solving out the optimal control, check that it's admissible, the value function is $C^{1,2}$, and the local MG is actually a MG
- Note that $v \in C^{1,2}$ can also be checked by applying uniform ellipticity (lowest eigenvalue of $\sigma\sigma^T$ bounded away from 0)
- The existence of the minimizer $\hat{\alpha} = \hat{\alpha}(t, x)$ can be checked through convex analysis and implicit function theorem

Infinite Horizon Case

The infinite horizon case is actually the same as what we have done for HJE, the deterministic optimal control problem. For the purpose of completeness, we state it again here. The difference in formulation is the introduction of the **discount factor** $\beta > 0$ and the fact that no terminal reward exists. The expected cost after time t following control α is

$$J(t,x,\alpha) = \mathbb{E}\left[\int_t^\infty e^{-\beta s} f(s, X_s^{t,x,\alpha}, \alpha_s)\, ds\right] \tag{330}$$

the introduction of discount factor is to ensure that the integral will be finite for a general class of f. However, by changing variables $u = s - t$ and **assuming that f is time-homogeneous** ($f(s, X_s^{t,x,\alpha}, \alpha_s) = f(X_s^{t,x,\alpha}, \alpha_s)$) one might find that

$$J(t,x,\alpha) = e^{-\beta t} \cdot \mathbb{E}\left[\int_0^\infty e^{-\beta s} f(X_{s+t}^{t,x,\alpha}, \alpha_{s+t})\, ds\right] \tag{331}$$

$$= e^{-\beta t} \cdot \mathbb{E}\left[\int_0^\infty e^{-\beta s} f(X_s^{0,x,\alpha}, \alpha_s)\, ds\right] \tag{332}$$

$$= e^{-\beta t} \cdot J(0,x,\alpha) \tag{333}$$

under the Markovian setting. As a result, we remove the time variable for simplicity and consider a slightly different definition that

$$J(x,\alpha) \overset{def}{=} J(0,x,\alpha) = \mathbb{E}\left[\int_0^\infty e^{-\beta s} f(X_s^{0,x,\alpha}, \alpha_s)\, ds\right] \tag{334}$$

as a result, the value function is defined as

$$v(x) \overset{def}{=} \inf_{\alpha \in \mathscr{A}} J(x,\alpha) \tag{335}$$

independent of time and the connection of this value function with the previously defined value function is that

$$v(x) = v(0,x) = e^{\beta t} \cdot v(t,x) \tag{336}$$

Remark. *The β can be understood as the opposite of the continuous-time interest rate. Under the condition that the discount factor has the formulation as $e^{-\beta t}$ and the fact that f is time-homogeneous, we can eliminate the time variable in t (failure in either assumption would cause mistake in doing so).*

The reason we mention the connection between the value function that only works for the special infinite horizon case and that for the general case is that we would still be able to apply the general results to get the HJBE.

Theorem 14. ***(HJBE of Stochastic Control Problem in Infinite Horizon Case)*** *Assume that $v \in C^2(\mathbb{R}^d)$*

and $f \in C(\mathbb{R}^d \times A)$ for each fixed control $\alpha \in \mathscr{A}$ and assume the existence of the optimal control $\alpha^* \in \mathscr{A}$. Then

$$\forall x \in \mathbb{R}^d, -\beta v(x) + \inf_{\alpha \in \mathscr{A}} \{L^\alpha v(x) + f(x, \alpha)\} = 0 \tag{337}$$

Proof. Denote the general value function as $u(t, x)$ so now $v(x) = u(0, x) = e^{\beta t} \cdot u(t, x)$. Since $\partial_t u(t, x) = -\beta e^{-\beta t} \cdot v(x), L^\alpha u(t, x) = e^{-\beta t} \cdot L^\alpha v(x)$, apply the HJBE for $u(t, x)$ to know that

$$\forall t > 0, x \in \mathbb{R}^d, , -\beta e^{-\beta t} \cdot v(x) + \inf_{\alpha \in \mathscr{A}} \left\{e^{-\beta t} \cdot L^\alpha v(x) + f(x, \alpha)\right\} = 0 \tag{338}$$

set $t = 0$ to get

$$\forall x \in \mathbb{R}^d, -\beta v(x) + \inf_{\alpha \in \mathscr{A}} \{L^\alpha v(x) + f(x, \alpha)\} = 0 \tag{339}$$

□