0.1 Introduction

As previously mentioned, in many instances the value function arising from an optimal control problem may fail to be continuously differentiable. If that happens the derivation of the Hamilton-Jacobi equation is no longer valid, but more importantly the notion of classical solution to it does hold anymore. Therefore, we have to weaken the notion of solution in order to get a consistent and unique solution to the dynamic programming equation for non-differentiable value functions. The *viscosity solution* is exactly what we are searching for. It arises from a standard procedure called vanishing viscosity, which allows us to compute the solution of a fully non-linear first order PDE as the limiting solution of quasilinear parabolic PDEs, obtained via infinitesimal perturbations of second order derivatives.

0.1.1 Non-differentiable value functions

Let us consider the calculus of variation problem:

$$\inf_{x\in Lip([0,1];[-1,1])} \int_t^{t_1} 1 + \frac{1}{4}(\dot{x}(s))^2 \, ds, \tag{1}$$

where $Lip(I;U)$ is the collection of Lipschitz continuous functions from I to U. The Hamiltonian related to this problem is:

$$H(t,x,p) = \max_{v\in[-1,1]} \left\{ -v \cdot p - 1 - \frac{1}{4}v^2 \right\}.$$

We can explicitly compute the Hamiltonian and get:

$$H(t,x,p) = p^2 - 1.$$

Then the Hamilton-Jacobi equations read:

$$\begin{cases} \dot{x}^*(s) = -H_p(s, x^*(s), p^*(s)) = 2p^*(s) \\ \dot{p}^*(s) = H_x(s, x^*(s), p^*(s)) = 0, \end{cases}$$

therefore, we get:

$$\dot{x}(s)^* = 2p^*, \; s \in [0,1],$$

for some $p^* \in \mathbb{R}$. We now compute the exit time of $(s, x(s)) = (s, 2(s-t)p^* + x)$ with initial data (t,x). If $p = 0$ then:

$$\tau = 1, \; |x| < 1.$$

If $p > 0$ then $x(s) = 2(s-t)p + x$ is increasing, which implies that the system is going to exit from the right boundary, that is from $x(s) = 1$, and if that happens before time $s = 1$ the exit time will be determined by:

$$2(s-t)p + x = 1 \Rightarrow s = t + \frac{1-x}{2p}.$$

$x(s) = 1$ for $s < 1$ if:

$$2(1-t)p + x \geqslant 1 \Rightarrow p \geqslant t + \frac{1-x}{2p},$$

therefore:

$$\tau = \begin{cases} 1 & p \geqslant t + \frac{1-x}{2p} \\ t + \frac{1-x}{2p} & p > t + \frac{1-x}{2p}. \end{cases}$$

Analogously, if $p < 0$:

$$\tau = \begin{cases} 1 & p \leqslant t - \frac{1+x}{2p} \\ t - \frac{1+x}{2p} & p < t - \frac{1+x}{2p}. \end{cases} .$$

We now solve:

$$\inf_{p \in \mathbb{R}} \int_t^\tau 1+p^2\, ds = \inf_{p \in \mathbb{R}} (1+p^2)(\tau - t) = \begin{cases} (1+p^2)(1-t), & p = 0 \text{ or } p > 0 \wedge p \geqslant \frac{1-x}{2(1-t)} \text{ or } p < 0 \wedge p \leqslant \frac{-1-x}{2(1-t)} \\ (1+p^2)\frac{1-x}{2p}, & p > 0 \wedge p \geqslant \frac{1-x}{2(1-t)} \\ -(1+p^2)\frac{1+x}{2p}, & p \leqslant \frac{-1-x}{2(1-t)}, \end{cases}$$

which is solved as follows:

$$V(t,x) = \begin{cases} 1-t & |x| \leqslant t \\ 1-t & |x| \geqslant t, \end{cases}$$

which is continuous on the whole space, but clearly no differentiable in $|x| = t$.

0.1.2 Vanishing viscosity

We now euristichally expose a technique called vanishing viscosity, which is widely used in calculus of variations problems and will show us one of the origins of the viscosity solution notion. Let us consider the initial value problem:

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

The method of characteristics shows that there cannot be a smooth solution of the above problem over the whole positive real line. Indeed, a weaker notion of solution is needed. One approach is to use Hopf-Lax solution concept. We are not interested in it, instead we start by perturbing the system as:

$$\begin{cases} u^\epsilon_t + H(u^\epsilon, Du^\epsilon) - \epsilon \Delta u^\epsilon = 0, & \mathbb{R}^n \times (0, +\infty) \\ u^\epsilon = g, & \mathbb{R}^n \times \{t = 0\}, \end{cases} \tag{3}$$

so that the fully non-linear system in 2 becomes a semilinear one, which turn out to have a smooth solution. Then, we take $\epsilon \to 0$. We expect the solution u^ϵ to lose the bounds on the derivatives, as they strongly depend on the regularization effect of $\epsilon\Lambda$. Turns out that many times Ascoli-Arzela theorem's hypotheses are satisfied, that is $(u^\epsilon)_\epsilon$ is uniformly bounded and equicontinuous, then we have local uniform convergence along a subsequence u^{ϵ_j}. We now use the limit $u \xleftarrow{j \to +\infty} u^{\epsilon_j}$ as a solution. We now it to be continuous but we lack information about its derivatives. We will then verify these information using test functions. Unlike the classical variational weak solution concept, where integration by part play the central role, we will use the maximum principle to translate the derivates of u onto the test functions.
Let us take $v \in C^\infty(\mathbb{R}^n \times (0, +\infty))$ and suppose that $u - v$ has a strict local maximum at (x_0, t_0), then:

$$(u - v)(x_0, t_0) > (u - v)(x, t),$$

for all (t, x) sufficiently close to (x_0, t_0). It can be shown that it implies that there exists $J > 0$ such that for all $j > J$ there exists $(x_{\epsilon_j}, t_{\epsilon_j})$ such that:

$$(u^\epsilon - v)(x_{\epsilon_j}, t_{\epsilon_j}) \geqslant (u^\epsilon - v)(x, t),$$

for (x, t) sufficiently close to $(x_{\epsilon_j}, t_{\epsilon_j})$ and such that:

$$(x_{\epsilon_j}, t_{\epsilon_j}) \xrightarrow{j \to +\infty} (x_0, t_0).$$

Because $(u^\epsilon - v)$ has a local maximum at $(x_{\epsilon_j}, t_{\epsilon_j})$:

$$Du^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) = Dv(x_{\epsilon_j}, t_{\epsilon_j}),\ u_t^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) = v(x_{\epsilon_j}, t_{\epsilon_j}),$$

and:

$$-\Lambda u^{\epsilon_j}(x_{\epsilon_j}, t_{\epsilon_j}) \geqslant -\Lambda v(x_{\epsilon_j}, t_{\epsilon_j})$$

.

Therefore, we get:

$$v_t(x_{\epsilon_j}, t_{\epsilon_j}) + H(Dv(x_{\epsilon_j}, t_{\epsilon_j}), x_{\epsilon_j}) \leqslant \Lambda v(x_{\epsilon_j}, t_{\epsilon_j}) \xrightarrow{j \to +\infty} 0$$

Analogous computations can be done for local minimum of $u - v$, obtaining the opposite inequality above.
We can now grasp the intuition behind the following definition.

Definition 0.1.1. *A viscosity solution of 2 is a function u bounded and uniformly continuous on $\mathbb{R}^n \times [0, T]$ for all $T > 0$ such that for all $v \in C^{+\infty}(\mathbb{R}^n \times (0, +\infty))$:*

$$v_t(x, t) + H(Dv(x, t), x) \leqslant 0$$

for all $(x, t) \in \arg\max\{u - v\}$ and:

$$v_t(x, t) + H(Dv(x, t), x) \geqslant 0$$

for all $(x, t) \in \arg\min\{u - v\}$. Furthermore, $u \equiv g$ for $t = 0$.

0.1.3 Abstract dynamic programming and viscosity solutions

We now present an abstraction of the dynamic programming principle, which will allow us to define the viscosity solutions of the dynamic programming equation. Let Σ be a closed subset of a Banach space and $\mathcal{C}$ a collection of functions on Σ, closed under addition:

$$\phi, \psi \in \mathcal{C} \Rightarrow \phi + \psi \in \mathcal{C}.$$

We consider the family of operators $\{\mathcal{T}_{tr}\}_{t_0 \leqslant t \leqslant r \leqslant t_1}$ such that:

$$\mathcal{T}_{tt}\phi = \phi,\ \ \forall\ \phi \in \mathcal{C}, \tag{4}$$

$$\mathcal{T}_{tr}\phi \leqslant \mathcal{T}_{ts}\psi \text{ if } \phi \leqslant \mathcal{T}_{rs}\psi, \tag{5}$$

and:

$$\mathcal{T}_{tr}\phi \geqslant \mathcal{T}_{ts}\psi \text{ if } \phi \geqslant \mathcal{T}_{rs}\psi. \tag{6}$$

Conditions 5 and 6 are a weaker version of monotonicity; they imply it together with 4. Moreover, they also imply the semigroup property, provided that $\mathcal{T}_{rt} : \mathcal{C} \to \mathcal{C}$. Under this assumption, the two conditions are equivalent to monotonicity. The semigroup property:

$$\mathcal{T}_{tr}\left(\mathcal{T}_{rs}\psi\right) = \mathcal{T}_{ts}\psi,\ \mathcal{T}_{rs}\psi \in \mathcal{C}, \tag{7}$$

is going to be the dynamic programming principle. Indeed, let us consider the classical optimal control problem defined on a bounded set $O \subset \mathbb{R}^n$, which we set to be $\Sigma = \overline{O}$ and $\mathcal{C} = \mathcal{M}(\Sigma)$, the collection of measurable functions bounded by below. Then as in chapter 1 we aim at minimize a functional, we set this functional to be the operator $\mathcal{T}$. Let us define:

$$\mathcal{T}_{t,r;u}\psi(x) = \int_t^{\tau\wedge r} L(s, x(s), u(s)),\, ds + g(\tau, x(\tau))\chi_{\tau<r} + \psi(x(r))\chi_{\tau\geqslant r}, \tag{8}$$

which gives:

$$\mathcal{T}_{tr}\psi = \inf_{u\in\mathcal{U}(t,x)} \mathcal{T}_{t,r;u}\psi. \tag{9}$$

Under the usual assumption on the running and terminal costs, as well as on the control space U we now that the value function defined in ref4-1-valuefunctreform is measurable and bounded by below. Therefore, $\mathcal{T}_{rt} : \mathcal{C} \to \mathcal{C}$ and we can formulate the semigroup property just by asking $\psi \in \mathcal{C}$. It is clear by its definition that the dynamic programming principle is translated as:

$$\mathcal{T}_{tt_1}\psi(x) = \mathcal{T}_{tr}\left(\mathcal{T}_{rt_1}\psi\right)(x),$$

for $(t, x) \in \overline{Q}$ and $\psi \in \mathcal{C}$. We now derive the abstract dynamic programming equation. The same procedure as in chapter 1 gives:

$$-\frac{1}{h}\left[\mathcal{T}_{tt+h}V(t+h,\cdot)(x) - V(t,x)\right] = 0.$$

What happens if we let $h \to 0$? We ask for the existence of a family of non-linear operators which will play the role of the Hamiltonian. Let $\Sigma' \subset \Sigma$ and $\mathcal{D} \subset C([t_0, t_1) \times \Sigma')$ and $\{\mathcal{G}_t\}_{t\in[t_0,t_1]}$ functions on Σ such that:

$$\lim_{h\to 0}\frac{1}{h}\left[\mathcal{T}_{tt+h}V(t+h,\cdot)(x) - V(t,x)\right] = \frac{\partial}{\partial t}w(t,x) - (\mathcal{G}_t w(t,\cdot))(x) \tag{10}$$

for all $w \in \mathcal{D}$, $(t, x) \in Q = [t_0, t_1) \times \Sigma'$. The space $\mathcal{D}$ is such that:

$$\text{For every } w \in \mathcal{D} \text{ the functions } \frac{\partial w}{\partial t} \text{ and } \mathcal{G}_t w(t,\cdot) \text{ are continuous on } Q \text{ and } w(t,\cdot) \in \mathcal{C} \text{ for all } t \in [t_0, t_1], \tag{11}$$

and $\mathcal{D}$ is a vector space:

$$w, \tilde{w} \in \mathcal{D} \Rightarrow w + \tilde{w} \in \mathcal{D},\ \lambda w \in \mathcal{D}. \tag{12}$$

The elements of $\mathcal{D}$ are called test functions and $\mathcal{G}_t$ the infinitesimal generator of $\mathcal{T}_{tr}$. Explicit choices of $\mathcal{C}$ and $\mathcal{D}$ will vary from case to case, usually they are chosen to satisfy certain integrability conditions on the functions.

If we require the existence of a test function space and an infinitesimal generator of the semigroup given by the value function, the dynamic programming equation becomes:

$$-\frac{\partial}{\partial t}V(t,x) + (\mathcal{G}_t V(t,\cdot))(x) = 0,\ (t,x) \in Q. \tag{13}$$

Then if a function $V \in \mathcal{D}$ satisfies 13 point wise it is called a classical solution of it. Thanks to this reformulation we will be able to weaken this notion of solution and finally get the viscosity solution of the dynamic programming equation.

We point out that in the canonical deterministic optimal control studied in chapter 1, the infinitesimal generator has the form:

$$(\mathcal{G}_t\phi)(x) = H(t, x, D\phi(x)) = \sup_{v\in U} \{-f(t, x, v) \cdot D\phi(x) - L(t, x, v)\},$$

with test functions space $\mathcal{D} = C^1(Q) \cap \mathcal{M}(\overline{Q})$ and $\Sigma' = O$.
In view of what we saw in previous sections, we give the definition of viscosity solution.

Definition 0.1.2. *Let $W \in C([t_0, t_1] \times \Sigma)$. Then:*

1. *W is a viscosity subsolution of 13 in Q if for every $w \in \mathcal{D}$:*

$$-\frac{\partial}{\partial t} w\left(\bar{t}, \overline{x}\right) + \left(\mathcal{G}_{\bar{t}} w\left(\bar{t}, \cdot\right)\right)(\overline{x}) \leqslant 0, \tag{14}$$

at every:

$$\left(\bar{t}, \overline{x}\right) \in \arg\max_{(t,x)\in Q} \{(W - w)(t, x)\},$$

and $W\left(\bar{t}, \overline{x}\right) = w\left(\bar{t}, \overline{x}\right)$.

2. *W is a viscosity supersolution of 13 in Q if for every $w \in \mathcal{D}$:*

$$-\frac{\partial}{\partial t} w\left(\bar{t}, \overline{x}\right) + \left(\mathcal{G}_{\bar{t}} w\left(\bar{t}, \cdot\right)\right)(\overline{x}) \geqslant 0, \tag{15}$$

at every:

$$\left(\bar{t}, \overline{x}\right) \in \arg\min_{(t,x)\in Q} \{(W - w)(t, x)\},$$

and $W\left(\bar{t}, \overline{x}\right) = w\left(\bar{t}, \overline{x}\right)$.

3. *W is a viscosity solution if it is both a subsolution and a supersolution.*

As a first step after the definition of viscosity solutions, we prove its consistency with the classical notion. To do so we require the operator to satisfy a maximum principle, that is if $\mathcal{G}_t$ is a general operator and:

$$\mathcal{D} = \{W \in C([t_0, t_1] \times \Sigma) \,|\, W_t(t, x), (\mathcal{G}_t W(t, \cdot))(x) \in C(Q)\},$$

then $\mathcal{G}_t$ satisfies the *maximum principle* if:

$$\mathcal{G}_t\phi(\overline{x}) \geqslant \mathcal{G}_t\psi(\overline{x})$$

for every $\overline{x} \in \arg\max\{(\phi - \psi)(x) \,|\, x \in \Sigma\} \cap \Sigma'$ with $\phi(\overline{x}) = \psi(\overline{x})$. Now, if W is a classical solution of 13, then for a $w \in \mathcal{D}$ and $\left(\bar{t}, \overline{x}\right) \in \arg\max_{(t,x)\in Q} \{(W - w)(t, x)\}$ and $W\left(\bar{t}, \overline{x}\right) = w\left(\bar{t}, \overline{x}\right)$ then:

$$-\frac{\partial}{\partial t} w\left(\bar{t}, \overline{x}\right) + \left(\mathcal{G}_{\bar{t}} w\left(\bar{t}, \cdot\right)\right)(\overline{x}) \leqslant -\frac{\partial}{\partial t} W\left(\bar{t}, \overline{x}\right) + \left(\mathcal{G}_{\bar{t}} W\left(\bar{t}, \cdot\right)\right)(\overline{x}) = 0, \tag{16}$$

since we asked continuity of time derivatives. For the supersolution recall that:

$$\max\{(\phi - \psi)(x) \,|\, x \in \Sigma\} = \min\{-(\phi - \psi)(x) \,|\, x \in \Sigma\}.$$

If $mathcalG_t$ is the infinitesimal generator of a two-parameter semigroup the connection between classical and viscosity solutions is even stronger.

Proposition 0.1.1. *Let $W \in \mathcal{D}$. Then W is a classical solution of 13 if and only if it is a viscosity solution of 13 in Q.*

Proof. If W is a viscosity solution, since it is also a test function then **??** and **??** hold for every point $(t,x) \in Q$, which implies that:

$$-\frac{\partial}{\partial t}w(t,x) + \left(\mathcal{G}_t w(t,\cdot)\right)(x) = 0, \ \ \forall\ (t,x) \in Q.$$

If W is a classical solution and prove the subsolution property. Let $w \in \mathcal{D}$ and (t,x) as usual. Since $w \geqslant W$:

$$\begin{aligned}
-\frac{\partial}{\partial t}w(t,x) + (\mathcal{G}_t w(t,\cdot))(x) &= -\lim_{h\to 0}\frac{1}{h}\left[(\mathcal{T}_{tt+h}w(t+h,\cdot))(x) - w(t,x)\right] \\
&\leqslant -\lim_{h\to 0}\frac{1}{h}\left[(\mathcal{T}_{tt+h}W(t+h,\cdot))(x) - W(t,x)\right] \\
&= -\frac{\partial}{\partial t}W(t,x) + (\mathcal{G}_t W(t,\cdot))(x) = 0.
\end{aligned}$$

The supersolution property is proven similarly. □

We now prove that the family of linear operators $\mathcal{T}_{tt_1}$ defined in 9 defines viscosity solution of the dynamic programming equation 13.

Theorem 0.1.2. *Let $\{\mathcal{T}_{tr}\}_{t_0 \leqslant t \leqslant r \leqslant t_1}$ such that 4,6,5 and also there exists a vector space $\mathcal{D}$ and another family of operator $\{\mathcal{G}_t\}_{t\in[t_0,t_1]}$ such that 11 and 10 hold. Let:*

$$V(t,x) = \left(\mathcal{T}_{tt_1}\psi\right)(x).$$

If $V \in C(Q)$ then it is a viscosity solution of 13.

Proof. Let us prove the subsolution condition. Let $w \in \mathcal{D}$ and (t,x) maximizer of $V - w$ in $\overline{Q}$ and $V(t,x) = w(t,x)$. Then $V \leqslant w$ on Q, which implies by 6 that we have:

$$(\mathcal{T}_{tr}w(r,\cdot))(x) \geqslant (\mathcal{T}_{tt_1}\psi)(x) = V(t,x) = w(t,x). \tag{17}$$

Now, because of 11 and 10 we can compute:

$$-\frac{\partial}{\partial t}w(t,x) + (\mathcal{G}_t w(t,\cdot))(x) = -\lim_{h\to 0}\frac{1}{h}\left[(\mathcal{T}_{tt+h}w(t+h,\cdot))(x) - w(t,x)\right] \leqslant 0$$

because of 17. The supersolution condition is proven analogously. □