INTRODUCTION TO THE HAMILTON-JACOBI-BELLMAN EQUATION

ADAM ANDERSSON

This text is a summary of important parts of chapter 3 and 4 in the book (*Controlled Markov Processes and Viscosity Solutions*, Fleming and Soner) [1]. It first states the optimal control problem over a finite time interval, or horizon. It then contains a formal derivation of the Hamilton-Jacobi-Bellman partial differential equation. In the third section some existence results are stated without proofs for the Hamilton-Jacobi-Bellman equation under a non-degeneracy condition. In the fourth and final section a verification theorem is stated. It gives the rigorous connection between the solution of the HJB-equation and the original stochastic control problem.

1. The control problem

The following diffusion type SDE will be considered:

$$\mathrm{d}X(t) = f(t, X(t), u(t))\,\mathrm{d}t + \sigma(t, X(t), u(t))\,\mathrm{d}W(t), \quad t \geq 0, \quad X(0) = x_0.$$

The state $\{X(t)\}_{t\geq 0}$ here depends on the process $\{u(t)\}_{t\geq 0}$ that we refer to as a control process. For the ease of notation this dependence is not made explicit.

To settle the framework let $\{W(t)\}_{t\geq 0}$ be a d-dimensional Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. The control process $\{u(t)\}_{t\geq 0}$ takes its values in a closed subset $U \subset \mathbb{R}^m$. The coefficients $f : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n\times d}$ are deterministic functions. They are assumed to have continuous and bounded first derivatives in t and x and moreover satisfy a linear growth condition in the control variable u.

The purpose of stochastic control is to control the diffusion to behave in a certain way. This is done by stating and solving a minimization or maximization problem. Define, for $(t, x, v) \in [0,T] \times \mathbb{R}^n \times U$, the **cost functional**

$$J(t, x, u) = \mathbb{E}\Big[\int_t^\tau L(s, X(s), u(s))ds + \Psi(\tau, X(\tau))\Big| X(t) = x\Big].$$

Here $\tau = \min(\tilde{\tau}, T)$, where $\tilde{\tau} \geq t$ is the stoping time when X leaves the open set $O \subset \mathbb{R}^n$, that may be $\mathbb{R}^n$ itself. This means that the X lives in O and is stopped when hitting the boundary ∂O. The function $L : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}$ is called the running cost function and $\Psi : ([0,T) \times \partial O) \cup (\{T\} \times \overline{O}) \to \mathbb{R}$ the terminal cost function. The running cost is assumed to be bounded in t and of polynomial growth in x and u. The terminal cost is assumed to have polynomial growth. If $L \equiv 0$ the control problem is said to be on Meyer form and when $\Psi \equiv 0$ the problem is said to be on Lagrange form.

It is common practice to state the problem as a minimization problem, choosing u to minimize J. This we will do in those notes. In finance on the other hand one often needs to maximize the utility of an investment, where, with our notation, Ψ is the utility function and $L \equiv 0$. The utility function describes the investors risk aversion.

A process $u : [t,T] \times \Omega \to U$ is called progressively measurable if its restriction to $[t, s] \times \Omega$ is $\mathcal{B}_{[t,s]} \times \mathcal{F}_s$-measurable. A control $u : [t,T] \times \Omega \to U$ is said to be admissible if it is progressively measurable and

$$\mathbb{E}\int_t^T |u(s)|^m\,\mathrm{d}s < \infty, \quad \forall m \in \mathbb{N}.$$

An easy way to make this assumption hold is to assume U to be compact. The class of admissible controls will be denoted $\mathcal{A}_t$.

2. A formal derivation of the HJB-equation

For simplicity we here take $O = \mathbb{R}^n$. It then make sense to consider a terminal cost function $\Psi : \mathbb{R}^n \to \mathbb{R}$, only depending on the state since $\tilde{\tau} = \infty$ in this case. Define the **value function**

$$V(t,x) = \inf_{u \in \mathcal{A}_t} J(t,x,u) \tag{2.1}$$

It plays an important role in control theory. We will derive formally an equation for V called the dynamic programming equation. This can be done for any Markov process, e.g., Levy processes or finite state Markov Chains. For diffusions the equation becomes a non-linear second order PDE called the Hamilton-Jacobi-Bellman (HJB) equation. Often it can be deduced from the equation what the optimal control is. It is then often of the form $u^*(s) = \pi^*(s, X(s))$, where $\pi : [0,T] \times \mathbb{R}^n \to U$ is a deterministic function, see (2.7) below. Such a control is called a **Markov control policy**.

Bellman's dynamic programming principle reads

$$V(t,x) = \inf_{u \in \mathcal{A}_t} \mathbb{E}\Big[\int_t^{t+h} L(s, X(s), u(s)) ds + V(t+h, X(t+h)) \Big| X(t) = x \Big].$$

The intuition is that the minimal cost on $[t,T]$ is achieved when running optimally in $[t, t+h]$ and then continue optimally in [t+h,T] with $X(t+h)$ as initial value. We will accept this heuristic argument in order to give a formal derivation of the HJB-equation. The important implications goes the other way. Once we have a smooth enough solution to the HJB-equation, we can prove the dynamic programming principle and other important results. So called **verification theorems** is used for this purpose.

We now start deriving the dynamic programming equation. Let the control be constant $u(s) = v$ for $s \in [t, t+h]$. Then the dynamic programming principle yields

$$V(t,x) \leq \mathbb{E}\Big[\int_t^{t+h} L(s, X(s), v) ds + V(t+h, X(t+h)) \Big| X(t) = x \Big].$$

Subtracting $V(t,x)$ from both sides and dividing by h gives

$$\begin{aligned} 0 &\leq \frac{1}{h} \mathbb{E}\Big[\int_t^{t+h} L(s, X(s), v) ds \Big| X(t) = x \Big] + \frac{1}{h} \mathbb{E}\Big[(V(t+h, X(t+h)) - V(t,x)) \Big| X(t) = x \Big] \\ &= I_1^h + I_2^h \end{aligned}$$

Using Fubinis theorem for conditional expectation and letting $h \to 0$ we have that

$$\begin{aligned} I_1^h &= \frac{1}{h} \int_t^{t+h} \mathbb{E}[L(s, X(s), v) | X(t) = x] \, \mathrm{d}s \\ &\to L(t,x,v). \end{aligned}$$

The second term I_2^h needs a little more work. Itô's formula yields

$$\begin{aligned} &V(t+h, X(t+h)) - V(t,x) \\ &\quad = \int_t^{t+h} A^v V(s, X(s)) \, \mathrm{d}s + \int_t^{t+h} V_x(s, X(s)) \cdot \sigma(s, X(s), v) \, \mathrm{d}W(s), \end{aligned} \tag{2.2}$$

where the **backward operator**[1] is given by

$$A^v\Phi(t,x) = \Phi_t(t,x) + \Phi_x(t,x)\cdot f(t,x,v) + \frac{1}{2}\operatorname{Tr}\{\Phi_{xx}(t,x)\sigma(t,x,v)\sigma^*(t,x,v)\}.$$

The trace of a square matrix is the sum of the diagonal elements. The above trace becomes for $a = \sigma\sigma^*$

$$\operatorname{Tr}\{\Phi_{xx}(t,x)\sigma(t,x,v)\sigma^*(t,x,v)\} = \sum_{i,j=1}^{n} a_{ij}(t,x,v)\Phi_{x_i x_j}(t,x).$$

It is here important to choose a suitable domain $\mathcal{D}$, for A^v, common for all $v \in U$, since we later want to vary v. Moreover the functions of this domain must be such that $A^v V$ is continuous and the Itô term in (2.2), for $V \in \mathcal{D}$, is a martingale. Assume here that this is the case. Then, when taking expectation in (2.2), we get that

$$\begin{aligned} I_2^h &= \frac{1}{h}\mathbb{E}\Big[\int_t^{t+h} A^v V(s,X(s))\,\mathrm{d}s\Big|X(t)=x\Big] \\ &= \frac{1}{h}\int_t^{t+h} \mathbb{E}[A^v V(s,X(s))|X(t)=x]\,\mathrm{d}s \\ &\to A^v V(t,x) \end{aligned}$$

as $h \to 0$. To conclude, for all $v \in U$,

$$0 \le A^v V(t,x) + L(t,x,v). \tag{2.3}$$

Assume now that the optimal control is given by an optimal Markov control policy, i.e, $u^*(s) = \pi^*(s, X^*(s))$. Here X^* is the optimal state process, controlled under u^*. The dynamic programming principle then takes the form

$$V(t,x) = \mathbb{E}\Big[\int_t^{t+h} L(s,X^*(s),\pi^*(s,X^*(s)))ds + V(t+h,X^*(t+h))\Big|X^*(t)=x\Big].$$

Using this, noticing that the backward operator of X^* is $A^{\pi^*} := A^{\pi^*(t,x)}$ and making similar calculation as those above one shows that

$$0 = A^{\pi^*}V(t,x) + L(t,x,\pi^*(t,x)). \tag{2.4}$$

For the limit argument to hold in this case continuity of π^* is needed, something we boldly assume. Combining (2.3) and (2.4) yields the **dynamic programming equation**

$$0 = \inf_{v\in U}[A^v V(t,x) + L(t,x,v)], \tag{2.5}$$

for $(t,x) \in [0,T]\times\mathbb{R}^n$, with terminal data

$$V(T,x) = \Psi(x). \tag{2.6}$$

If we accept this then a reasonable candidate for an optimal control policy is

$$\pi^*(t,x) = \operatorname{argmin}[A^v V(t,x) + L(t,x,v)]. \tag{2.7}$$

[1] The operator A^v is called backward since it is the operator appearing in the backward Kolmogorov equation $A^v\Phi = 0$, for $\Phi(T,x) = \phi(x)$. Its solution is given by $\Phi(t,x) = \mathbb{E}[\phi(X(T))|X(t)=x]$. In Markov theory standard notation reeds $A^v = \partial_t + G^v$, where G^v is the infinitesimal generator of the Markov semigroup.

If it exist almost everywhere and the solution V to (2.5) and (2.6) is sufficiently smooth, then a verification theorem guarantees that so is the case. The same theorem states that V really is the value function we defined in (2.1).

We rewrite the dynamic programming equation in terms of the Hamiltonian

$$\mathcal{H}(t,x,p,A) = \sup_{v\in U}\Big[-f(t,x,v)\cdot p - \frac{1}{2}\operatorname{Tr}\{A\sigma(t,x,v)\sigma^*(t,x,v)\} - L(t,x,v)\Big].$$

The equation then takes the form of the Hamilton-Jacobi-Bellman equation

$$-\frac{\partial V}{\partial t} + \mathcal{H}(t,x,D_xV,D_x^2V) = 0, \quad (t,x)\in(0,T)\times\mathbb{R}^n$$

satisfying

$$V(T,x) = \Psi(x), \quad x\in\mathbb{R}^n.$$

3. HJB in the case of a non-degenerate diffusion

Taking into account the more general case of an arbitrary open $O\subset\mathbb{R}^n$ the Hamilton-Jacobi-Bellman equation becomes

$$-\frac{\partial V}{\partial t} + \mathcal{H}(t,x,D_xV,D_x^2V) = 0, \quad (t,x)\in(0,T)\times O \tag{3.1}$$

with the boundary condition

$$V(t,x) = \Psi(t,x), \quad (t,x)\in([0,t)\times\partial O)\cup(T\times\overline{O}). \tag{3.2}$$

So, by leaving O at $x\in\partial O$ a time $t<T$ costs $\Psi(t,x)$, while as before, we pay $\Psi(T,x)$ if X remains inside O for all $t<T$ and has value x at the final time.

There is one property of the diffusion that splits the problem into two categories. In the first category the HJB-equation has a unique classical solution. In the second the solution has a generalized solution in terms of viscosity solutions, possible to handle but more difficult. The property that makes this clear division is that of non-degeneracy.

The diffusion (1) is called non-degenerate if the diffusion matrix $a=\sigma\sigma^*$ satisfies the uniform elipticity condition

$$\sum_{i,j=1}^{n} a_{ij}(t,x,v)\xi_i\xi_j \geq C|\xi|^2. \tag{3.3}$$

The HJB-equation is then uniformly parabolic, allowing for classical solutions. Condition (3.3) implies that a is invertible. This can only happen if $\operatorname{rank}(\sigma)=n$ and hence $d\geq n$. We now interpreters this in probabilistic terms. That $d\geq n$ means that there are no less Brownian motions than space dimensions, i.e., there is enough noise to disturb the solution in any dimension. That $\operatorname{rank}(\sigma)=n$ means that σ distributes the noise in the n linearly independent directions of the row vectors of σ. Finally condition (3.3) guarantees that the noise is bounded away from zero, i.e., the behavior of the diffusion is never dominated by the drift. An equivalent definition is that X is non-degenerate iff it has a probability density for all $t>0$.

We here state known existence and uniqueness results from PDE-theory for the Hamilton-Jacobi-Bellman equation in the non-degenerate case.

Theorem 3.1. *Under the assumptions*

- *U is compact;*
- *O is bounded with ∂O being a manifold of class C^3;*
- *$a=\sigma\sigma^*$, f, L have one continuous t-derivative and two continuous x-derivatives;*

- *Ψ has three continuous derivatives in both t and x;*
- *σ satisfies (3.3)*

equation (3.1) and (3.2) has a unique solution $W \in C^{1,2}((0,T) \times O) \cap C([0,T] \times \overline{O})$

Theorem 3.2. *Under the assumptions*

- *U is compact;*
- *$O = \mathbb{R}^n$;*
- *$a = \sigma\sigma^*$, f, L are bounded and have one continuous t-derivative and two continuous x-derivatives;*
- *Ψ has three continuous and bounded derivatives in x (t-independent since $O = \mathbb{R}^n$);*
- *σ satisfies (3.3)*

equation (3.1) and (3.2) has a unique solution $W \in C_b^{1,2}([0,T] \times \mathbb{R}^n)$

4. A Verification theorem

We now have the Hamilton-Jacobi-Bellman equation and some existence results for it. Consider this as the starting point. The following verification theorem gives us the connection to the optimal control problem.

Theorem 4.1. *Let $W \in C^{1,2}((0,T) \times O) \cap C_p([0,T] \times \overline{O})$ be a solution to (3.1) and (3.2). Then*

- *$W(t,x) \leq J(t,x,u)$, for all $(t,x) \in (0,T) \times O$ and any admissible control u.*
- *If there exists an admissible control u^* such that*

$$u^*(s) \in \operatorname{argmin} \Big[f(s, x^*(s), v) \cdot W_x(s, x^*(s)) + \frac{1}{2} \operatorname{Tr}\{\Phi_{xx}(t,x)\sigma(t,x,v)\sigma^*(t,x,v)\} + L(s,x,v) \Big]$$

  for $ds \times \mathbb{P}$ almost every $(s,\omega) \in [t,\tau] \times \Omega$, then $W(t,x) = J(t,x,u^*)$.
- *The dynamic programming principle holds.*

The proof contains much of the spirit of the formal derivation of the HJB-equation, but is done in the right direction.

References

[1] W.H. Fleming and H.M. Soner *Controlled Markov Processes and Viscosity Solutions,* Applications of mathematics 25, Springer-Verlag **1993**