# NOTE ON HAMILTON-JACOBI-BELLMAN EQUATION

TAI-HO WANG

February 21, 2013

## 1. Heuristic derivation of the HJB equation

Consider the controlled stochastic differential equation

$$dX_t^{(v)} = \mu(t, X_t^{(v)}, v_t)dt + \sigma(t, X_t^{(v)}, v_t)dW_t.$$

Note that we include the super index $v$ to indicate that the solution $X^{(v)}$ depends on the choice of the control $v$. Let $\mathcal{G}[t,T]$ denote the collection of all admissible controls within the time interval $[t,T]$, for $0 \le t \le T$. The objective functional, for a given (admissible) control $v \in \mathcal{G}[0,T]$, consists of two parts:

$$g(X_T^{(v)}) + \int_0^T h(s, X_s^{(v)}, v_s)ds,$$

where $h$ is called the running reward/cost, and $g$ is the final reward/cost. The goal is optimize the expected objective functional over the collection $\mathcal{G}[0,T]$ of all admissible controls. Precisely, we seek to maximize

$$\max_{v \in \mathcal{G}[0,T]} \mathbb{E}\left[ g(X_T^{(v)}) + \int_0^T h(s, X_s^{(v)}, v_s)ds \right].$$

To this end, let's introduce the concept of the so-called *value function*. Define, for a given admissible control $v \in \mathcal{G}[t,T]$, the expected objective functional $J^{(v)}$ as

$$J^{(v)}(t,x) = \mathbb{E}\left[ g(X_T^{(v)}) + \int_0^T h(s, X_s^{(v)}, v_s)ds \middle| \mathcal{F}_t \right].$$

The value function $J(t,x)$ for a stochastic control problem is defined as

$$J(t,x) = \max_{v \in \mathcal{G}[t,T]} J^{(v)}(t,x).$$

In other words, the value function at $(t,x)$ is the optimal value of the control problem conditioned on the process starting at $(t,x)$ and applying the optimal control thereafter. To simplify the notation, we shall denote the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t]$ by $\mathbb{E}_t[\cdot]$ hereafter.

Obviously, for any admissible control $v \in \mathcal{G}[t,T]$, $J^{(v)}(t,x) \le J(t,x)$. Moreover, consider the following chain of inequality/equalities: for any $v \in \mathcal{G}[t,T]$,

$$
\begin{aligned}
& J^{(v)}(t,x) \\
= \; & \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds\right] + \mathbb{E}_t\left[\mathbb{E}_{t+\epsilon}\left\{g(X_T^{(v)}) + \int_{t+\epsilon}^T h(s,X_s^{(v)},v_s)ds\right\}\right] \\
\le \; & \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds\right] + \mathbb{E}_t\left[\max_{u\in\mathcal{G}[t+\epsilon,T]} \mathbb{E}_{t+\epsilon}\left\{g(X_T^{(u)}) + \int_{t+\epsilon}^T h(s,X_s^{(u)},u_s)ds\right\}\right] \\
= \; & \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds\right] + \mathbb{E}_t\left[J(t,X_{t+\epsilon}^{(v)})\right] \\
= \; & \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds + J(t,X_{t+\epsilon}^{(v)})\right]
\end{aligned}
$$

Therefore, if we maximize the two ends of the inequality over all admissible controls in $\mathcal{G}[t,T]$, we end up with

$$J(t,x) = \max_{v\in\mathcal{G}[t,T]} J^{(v)}(t,x) \le \max_{v\in\mathcal{G}[t,T]} \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds + J(t+\epsilon,X_{t+\epsilon}^{(v)})\right].$$

Notice that, since the control $v \in \mathcal{G}[t,T]$ on the right hand side of the inequality has no effect after time $t+\epsilon$, the maximization can be restricted to the admissible controls $v \in \mathcal{G}[t,t+\epsilon]$ in the smaller time interval from $t$ to $t+\epsilon$:

$$J(t,x) \le \max_{v\in\mathcal{G}[t,t+\epsilon]} \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds + J(t+\epsilon,X_{t+\epsilon}^{(v)})\right].$$

In fact, the Bellman's principle suggests that the inequality is indeed an equality. (Why?) In other words, the value function $J$ satisfies, for any $\epsilon > 0$,

$$J(t,x) = \max_{v\in\mathcal{G}[t,t+\epsilon]} \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v)},v_s)ds + J(t+\epsilon,X_{t+\epsilon}^{(v)})\right]. \tag{1.1}$$

**Remark 1.** Let's parse the Bellman's principle (1.1) a bit as follows. Pretend there is an optimal control $v^* \in \mathcal{G}[t,T]$ and the value function is attained at the optimal control $v^*$, i.e.

$$J(t,x) = \mathbb{E}_t\left[g(X_T^{(v^*)}) + \int_0^T h(s,X_s^{(v^*)},v_s^*)ds\right].$$

The Bellman's principle (1.1) is then rephrased as

$$J(t,x) = \mathbb{E}_t\left[\int_t^{t+\epsilon} h(s,X_s^{(v^*)},v_s^*)ds + J(t+\epsilon,X_{t+\epsilon}^{(v^*)})\right]. \tag{1.2}$$

The right hand side of the equality basically says that we start the process at $(t,x)$ and apply the control $v^*$ until $t+\epsilon$, for $\epsilon > 0$, of course at this time the process is suspended at $X_{t+\epsilon}^{(v^*)}$, then we restart the process at $(t+\epsilon, X_{t+\epsilon}^{(v^*)})$ and apply an optimal control in $\mathcal{G}[t+\epsilon,T]$ for the rest of the life of the process. Lastly, average out the possibilities at $t+\epsilon$, i.e., take the conditional expectation conditioned on $\mathcal{F}_t$. Therefore, the equality (1.2) simply says, if $v^*$ is indeed an optimal control in $\mathcal{G}[t,T]$, then $v^*$ shouldn't be affected if we suspend the process at a later time and restart it afresh, which, in principle, is the Bellman's principle. Note that the Bellman's principle (1.1) holds for any $\epsilon > 0$.

Now let's derive an infinitesimal version of the Bellman's principle (1.1). Rearrange terms in (1.1) we have

$$\max_{v\in\mathcal{G}[t,t+\epsilon]} \mathbb{E}_t\left[J(t+\epsilon, X^{(v)}_{t+\epsilon}) - J(t,x) + \int_t^{t+\epsilon} h(s, X^{(v)}_s, v_s)ds\right] = 0 \tag{1.3}$$

$$\Longrightarrow \quad \max_{v\in\mathcal{G}[t,t+\epsilon]} \left\{\frac{1}{\epsilon}\left(\mathbb{E}_t\left[J(t+\epsilon, X^{(v)}_{t+\epsilon})\right] - J(t,x)\right) + \mathbb{E}_t\left[\frac{1}{\epsilon}\int_t^{t+\epsilon} h(s, X^{(v)}_s, v_s)ds\right]\right\} = 0$$

Let's simplify the expressions inside the brackets. For the first part, apply Ito's formula (assume the value function $J$ is smooth enough), we have

$$\mathbb{E}_t\left[J(t+\epsilon, X^{(v)}_{t+\epsilon})\right] - J(x,t) = \int_t^{t+\epsilon} \mathbb{E}_t\left[\partial_t J(s, X^{(v)}_s) + \mathcal{L}^{(v)} J(s, X^{(v)}_s)\right] ds$$

$$\Longrightarrow \quad \frac{1}{\epsilon}\left\{\mathbb{E}_t\left[J(t+\epsilon, X^{(v)}_{t+\epsilon})\right] - J(x,t)\right\} = \frac{1}{\epsilon}\int_t^{t+\epsilon} \mathbb{E}_t\left[\partial_t J(s, X^{(v)}_s) + \mathcal{L}^{(v)} J(s, X^{(v)}_s)\right] ds$$

where $\mathcal{L}^{(v)}$ is the infinitesimal generator associated with the controlled process $X^{(v)}$. Precisely,

$$\mathcal{L}^{(v)} = \frac{1}{2}\sigma^2(t,x,v)\partial_x^2 + \mu(t,x,v)\partial_x.$$

Hence, (1.3) now reads

$$\max_{v\in\mathcal{G}[t,t+\epsilon]} \left\{\mathbb{E}_t\left[\frac{1}{\epsilon}\int_t^{t+\epsilon} \partial_t J(s, X^{(v)}_s) + \mathcal{L}^{(v)} J(s, X^{(v)}_s) + h(s, X^{(v)}_s, v_s)ds\right]\right\} = 0.$$

Taking the limit as $\epsilon \to 0^+$, we end up with

$$\partial_t J(t,x) + \max_{v\in\mathcal{G}[t]}\{\mathcal{L}^{(v)} J(t,x) + h(t,x,v)\} = 0,$$

which is the celebrated HJB equation for the value function $J$. Finally, taking into account the final reward/cost, the value function $J$ satisfies the terminal value problem

$$\begin{aligned} &\partial_t J(t,x) + \max_{v\in\mathcal{G}[t]}\{\mathcal{L}^{(v)} J(t,x) + h(t,x,v)\} = 0, \text{ for } t < T, \\ &J(T,x) = g(x). \end{aligned}$$