MA668: Algorithmic and High Frequency Trading
Lecture 22

Prof. Siddhartha Pratim Chakrabarty
Department of Mathematics
Indian Institute of Technology Guwahati

Control for Diffusion Processes (Contd ...)

1. The running penalty/reward may, in general, be dependent on time t, the current position of the controlled process $\mathbf{X}_t^{\mathbf{u}}$ and the control itself $\mathbf{u}_t$.
2. On the other hand, the terminal reward depends solely on the terminal value of the controlled process.
3. For simplicity, the functions G and F are assumed to be uniformly bounded and the vector of drifts μ_t and volatilities σ_t are, as usual, Lipschitz continuous.
4. The integrability assumption on the controls, drift and volatility are necessary to ensure that the steps outlined below can be made rigorous.
5. The predictability assumption on the controls is necessary, since otherwise the agent may be able to peek into the future to optimise her/his strategy, and strategies which do peek into the future cannot be implemented in the real world.

Control for Diffusion Processes (Contd ...)

1. The value function (equation (10) of previous lecture) has the interpretation that the agent wishes to maximize the total of terminal reward function G and running reward/penalty F, by acting in an optimal manner.
2. Her/his actions $\mathbf{u}$ affect the dynamics of the underlying system in some generic way given by (equation (11) of previous lecture).
3. Thus, her/his past actions affect the future dynamics and she/he must therefore adapt and tune her/his actions to account for this feedback effect.

Control for Diffusion Processes (Contd ...)

1. For an arbitrary admissible control $\mathbf{u}$, we define the so-called performance criteria $H^{\mathbf{u}}(\mathbf{x})$ by:

$$H^{\mathbf{u}}(\mathbf{x}) = \mathbb{E}\left[G\left(\mathbf{X}_T^{\mathbf{u}}\right) + \int\limits_0^T F\left(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s\right) ds \right]. \tag{1}$$

2. The agent therefore seeks to maximise this performance criteria, and naturally:

$$H(\mathbf{x}) = \sup_{\mathbf{u} \in \mathcal{A}_{0,T}} H^{\mathbf{u}}(\mathbf{x}). \tag{2}$$

Control for Diffusion Processes (Contd ...)

1. As mentioned earlier, rather than optimizing $H^{\mathbf{u}}(\mathbf{x})$ directly, it is more convenient (and powerful) to introduce a time-indexed collection of optimization problems on which a dynamic programming principle (DPP) can be derived.
2. The DPP in infinitesimal form leads to a Dynamic Programming Equation (DPE) or the Hamilton-Jacobi-Bellman (HJB) equation: This is a non-linear PDE whose solution is a tentative solution to the original problem.
3. If a classical solution [a] to the DPE exists, then it is possible to prove, through a verification argument, that it is in fact the solution to the original control.
4. We discuss the three preceding points in the next three topics of discussion.

---

[a] Here, a classical solution means that the solution is once differentiable in time and twice in all (diffusive) state variables, so that the infinitesimal generator can be applied to it

The Dynamic Programming Principle

1. The usual trick to solving stochastic (and deterministic!) control problems is to embed the original problem into a larger class of problems indexed by time $t \in [0, T]$ but equal to the original problem at $t = 0$.
2. To this end, first define (with a slight abuse of notation):

$$H(t, \mathbf{x}) := \sup_{\mathbf{u} \in \mathcal{A}_{t,T}} H^{\mathbf{u}}(t, \mathbf{x}), \tag{3}$$

and

$$H^{\mathbf{u}}(t, \mathbf{x}) := \mathbb{E}_{t,\mathbf{x}} \left[G\left(\mathbf{X}_T^{\mathbf{u}}\right) + \int_0^T F\left(s, \mathbf{X}_s^{\mathbf{u}}, \mathbf{u}_s\right) ds \right], \tag{4}$$

where the notation $\mathbb{E}_{t,\mathbf{x}}[\cdot]$ represents the expectation conditional on $\mathbf{X}_t^{\mathbf{u}} = \mathbf{x}$.
3. These two objects are the time indexed analog of the original control problem and the performance criteria.
4. In particular, $H(0, \mathbf{x})$ coincides with the original control (equation (10) of previous lecture) and $H^{\mathbf{u}}(0, \mathbf{x})$ with the performance criteria (1).

The Dynamic Programming Principle (Contd ...)

1. Next, we take an arbitrary admissible strategy $\mathbf{u}$ and imagine flowing the $\mathbf{X}$ process forward in time from t to an arbitrary stopping time $\tau \leq T$.
2. Then, conditional on $\mathbf{X}_\tau^{\mathbf{u}}$, the contribution of the running reward/penalty from τ to T and the terminal reward can be viewed as the performance criteria starting from the new value of $\mathbf{X}_\tau^{\mathbf{u}}$ (see Figure 5.1).
3. This allows the value function to be written in terms of the expectation of its future value at τ plus the reward between now and τ.

Figure 5.1

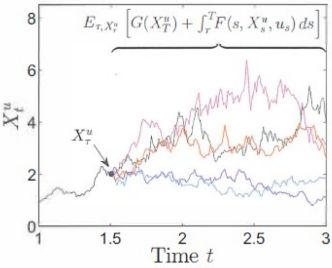

Figure 5.1 The DPP allows the value function to be written as an expectation of the future value function. The key idea is to flow the dynamics of the controlled process from t to τ and then rewrite the remaining expectation as the future performance criteria.

Figure: Figure 5.1

The Dynamic Programming Principle (Contd ...)

1. More precisely, by iterated expectations, the time-indexed performance criteria becomes:

$$
\begin{aligned}
H^{\mathbf{u}}(t,\mathbf{x}) &= \mathbb{E}_{t,\mathbf{x}}\left[G\left(\mathbf{X}^{\mathbf{u}}_T\right) + \int_\tau^T F\left(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s\right)ds + \int_t^\tau F\left(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s\right)ds\right], \\
&= \mathbb{E}_{t,\mathbf{x}}\left[\mathbb{E}_{\tau,\mathbf{x}^{\mathbf{u}}_\tau}\left[G\left(\mathbf{X}^{\mathbf{u}}_T\right) + \int_\tau^T F\left(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s\right)ds\right]\right. \\
&\qquad \left. + \int_t^\tau F\left(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s\right)ds\right], \\
&= \mathbb{E}_{t,\mathbf{x}}\left[H^{\mathbf{u}}(\tau,\mathbf{X}^{\mathbf{u}}_\tau) + \int_t^\tau F\left(s,\mathbf{X}^{\mathbf{u}}_s,\mathbf{u}_s\right)ds\right]. \qquad (5)
\end{aligned}
$$

The Dynamic Programming Principle (Contd ...)

1. Now, $H(t, \mathbf{x}) \geq H^{\mathbf{u}}(t, \mathbf{x})$ for an arbitrary admissible control $\mathbf{u}$ (with equality holding if $\mathbf{u}$ is the optimal control $\mathbf{u}^*$ assuming that $\mathbf{u}^* \in \mathcal{A}_{t,T}$, *i.e.,* the supremum is attained by an admissible strategy [a] and an arbitrary $\mathbf{x}$.
2. Hence, on the right-hand side of (5) the performance criteria $H^{\mathbf{u}}(\tau, \mathbf{X}^{\mathbf{u}}_\tau)$ at the stopping time T is bounded above by the value function $H(\tau, \mathbf{X}^{\mathbf{u}}_\tau)$.
3. The equality can then be replaced by an inequality with the value function (and not the performance criteria) showing up under the expectation:

$$
\begin{aligned}
H^{\mathbf{u}}(t, \mathbf{x}) \quad &\leq \quad \mathbb{E}_{t,\mathbf{x}} \left[H(\tau, \mathbf{X}^{\mathbf{u}}_\tau) + \int_t^\tau F\left(s, \mathbf{X}^{\mathbf{u}}_s, \mathbf{u}_s\right) ds \right], \\
&\leq \quad \sup_{\mathbf{u} \in \mathcal{A}} \mathbb{E}_{t,\mathbf{x}} \left[H(\tau, \mathbf{X}^{\mathbf{u}}_\tau) + \int_t^\tau F\left(s, \mathbf{X}^{\mathbf{u}}_s, \mathbf{u}_s\right) ds \right].
\end{aligned}
$$

4. This provides us with a first inequality.

---

[a] It may be the case that the supremum is obtained by a limiting sequence of admissible strategies for which the limiting strategy is in fact not admissible