Stochastic Control
Lecture: Stochastic Optimal Control*

Álvaro Cartea
University of Oxford

January 20, 2017


1 Introduction

Stochastic control problems arise in many facets of financial modelling. The classical example is the optimal investment problem introduced and solved in continuous-time by Merton (1971). Of course there is a multitude of other applications, such as optimal dividend setting, optimal entry and exit problems, utility indifference valuation and so on. In general, the all-encompassing goal of stochastic control problems is to maximise (or minimise) some expected profit (cost) function by tuning a strategy which itself affects the dynamics of the underlying stochastic system, and to find the strategy which attains the maximum (minimum). For example, in the simplest form of the Merton problem, the agent is trying to maximise expected utility of future wealth by trading a risky asset and a risk-free bank account. The agent’s actions

*These notes are for University of Oxford students and should not be posted on the web or distributed. Email: alvaro.cartea@maths.ox.ac.uk
affect her wealth, but at the same time the uncertain dynamics in the traded asset modulate the agent’s wealth in a stochastic manner. The resulting optimal strategies are tied to the dynamics of the asset and perhaps also to the agent’s wealth. It is a surprising fact that, in many cases, the optimal strategies turn out to be Markov in the underlying state variables, even if the agent is considering non-Markovian controls (which may depend on the entire history of the system).

One tool keeps coming to the forefront when solving stochastic control problems: the dynamic programming principle (DPP) and the related non-linear partial differential equation (PDE) known as the Hamilton–Jacobi–Bellman (HJB) equation – also called the dynamic programming equation (DPE). The DPP allows a stochastic control problem to be solved from the terminal date backwards and the HJB equation / DPE can be viewed as its infinitesimal version.

Here, the subtle mathematical issues are not addressed and focus is instead placed on the mechanics which allow for immediate application to financial problems. The interested reader is referred to the many excellent texts which focus on the theoretical aspects of stochastic control for a thorough treatment of the subject: Yong & Zhou (1999), Fleming & Soner (2006), Øksendal & Sulem (2007), Pham (2010), and Touzi (2013).

2 Control for Diffusion Processes

In this section, the control problems are of the general form

\[ H(x) = \sup_{u \in \mathcal{A}_{0,T}} \mathbb{E} \left[ G(X^u_T) + \int_0^T F(s, X^u_s, u_s) \, ds \right], \]

where \( u = (u_t)_{0 \leq t \leq T} \) is the vector (dim \( p \)) valued control process, \( X^u = (X^u_t)_{0 \leq t \leq T} \) is the vector (dim \( n \)) valued controlled process assumed (in this section) to be an Itô diffusion satisfying

\[ dX^u_t = \mu(t, X^u_t, u_t) \, dt + \sigma(t, X^u_t, u_t) \, dW_t, \quad X^u_0 = x, \]

where \((W_t)_{0 \leq t \leq T}\) is a vector of independent Brownian motions, \( \mathcal{A} \) is a set (called the admissible set) of \( \mathcal{F} \)-predictable processes such that (2) admits a strong solution
(and may contain other constraints such as the process being bounded), \(G : \mathbb{R}^n \mapsto \mathbb{R}\) is a terminal reward and \(F : \mathbb{R}_+ \times \mathbb{R}^{n+p} \mapsto \mathbb{R}\) is a running penalty/reward. The running penalty/reward may, in general, be dependent on time \(t\), the current position of the controlled process \(X_t^u\), and the control itself \(u_t\), while the terminal reward depends solely on the terminal value of the controlled process. For simplicity, the functions \(G\) and \(F\) are assumed uniformly bounded and the vector of drifts \(\mu_t\) and volatilities \(\sigma_t\) are, as usual, Lipschitz continuous. The integrability assumption on the controls, drift and volatility are necessary to ensure that the steps outlined below can be made rigorous. The predictability assumption on the controls is necessary since otherwise the agent may be able to peek into the future to optimise her strategy, and strategies which do peek into the future cannot be implemented in the real world.

The value function \([1]\) has the interpretation that the agent wishes to maximise the total of terminal reward function \(G\) and running reward/penalty by acting in an optimal manner. Her actions \(u\) affect the dynamics of the underlying system in some generic way given by \([2]\). Thus, her past actions affect the future dynamics and she must therefore adapt and tune her actions to account for this feedback effect.

For an arbitrary admissible control \(u\), define the so-called performance criteria \(H^u(x)\) by

\[
H^u(x) = E \left[ G(X_T^u) + \int_0^T F(s, X_s^u, u_s) \, ds \right].
\]

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

\[
H(x) = \sup_{u \in A_{0, T}} H^u(x).
\]  

As mentioned in the introduction, rather than optimising \(H^u(x)\) directly, it is more convenient (and powerful) to introduce a time-indexed collection of optimisation problems on which a dynamic programming principle (DPP) can be derived. The DPP in infinitesimal form leads to a DPE – the Hamilton–Jacobi–Bellman (HJB) equation – which is a non-linear PDE whose solution is a tentative solution to the original problem. If a classical solution\(^1\) to the DPE exists, it is possible to prove, through a

\(^1\)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.
verification argument, that it is in fact the solution to the original control. The next three subsections take on this programme.

2.1 The Dynamic Programming Principle

The usual trick\textsuperscript{2} 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 \( t = 0 \). To this end, first define (with a slight abuse of notation)

\[
H(t, x) := \sup_{u \in A_{t,T}} H^u(t, x), \quad \text{and} \quad H^u(t, x) := \mathbb{E}_{t,x} \left[ G(X^u_T) + \int_t^T F(s, X^u_s, u_s) \, ds \right],
\]

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

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

More precisely, by iterated expectations the time-indexed performance criteria

\textsuperscript{2}There is another approach to controlling both deterministic and stochastic systems which makes use of Pontryagin’s maximum principle. See Yong & Zhou (1999) for a wonderful exposition of its use for stochastic systems and the connection between it and the DPP formulation.
Figure 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 $\tau$ and then rewrite the remaining expectation as the future performance criteria.

becomes

$$H^u(t, x) = \mathbb{E}_{t, x} \left[ G(X^u_T) + \int_t^\tau F(s, X^u_s, u_s) \, ds \right]$$

$$= \mathbb{E}_{t, x} \left[ \mathbb{E}_{\tau, X^u_{\tau}} \left[ G(X^u_{\tau}) + \int_\tau^T F(s, X^u_s, u_s) \, ds \right] + \int_t^\tau F(s, X^u_s, u_s) \, ds \right]$$

$$= \mathbb{E}_{t, x} \left[ H^u(\tau, X^u_{\tau}) + \int_t^\tau F(s, X^u_s, u_s) \, ds \right]. \quad (6)$$

Now, $H(t, x) \geq H^u(t, x)$ for an arbitrary admissible control $u$ (with equality holding if $u$ is the optimal control $u^*$ — assuming that $u^* \in A_{t,T}$, i.e. the supremum is attained by an admissible strategy[^1] and an arbitrary $x$). Hence, on the right-hand side of (6), the performance criteria $H^u(\tau, X^u_{\tau})$ at the stopping time $\tau$ is bounded above by the value function $H(\tau, X^u_{\tau})$. The equality can then be replaced by an inequality with the value function (and not the performance criteria) showing up under the expectation:

$$H^u(t, x) \leq \mathbb{E}_{t, x} \left[ H(\tau, X^u_{\tau}) + \int_t^\tau F(s, X^u_s, u_s) \, ds \right]$$

$$\leq \sup_{u \in A} \mathbb{E}_{t, x} \left[ H(\tau, X^u_{\tau}) + \int_t^\tau F(s, X^u_s, u_s) \, ds \right].$$

Note that on the right-hand side of the above, the arbitrary control $u$ only acts over the interval $[t, \tau]$ and the optimal one is implicitly incorporated in the value function $H(\tau, X^u_{\tau})$ but starting at the point to which the arbitrary control $u$ caused the process $X$ to flow, namely $X^u_{\tau}$.

[^1]: 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.
Taking supremum over admissible strategies on the left-hand side, so that the left-hand side also reduces to the value function, we have that

\[
H(t, x) \leq \sup_{u \in A} E_{t,x} \left[ H(\tau, X_\tau^u) + \int_t^\tau F(s, X_s^u, u_s) \, ds \right].
\]

This provides us with a first inequality.

Next, we aim to show that the inequality above can be reversed. Take an arbitrary admissible control \( u \in A \) and consider what is known as an \( \varepsilon \)-optimal control denoted by \( \nu^\varepsilon \in A \) and defined as a control which is better than \( H(t, x) - \varepsilon \), but of course not as good as \( H(t, x) \), i.e. a control such that

\[
H(t, x) \geq H(\nu^\varepsilon)(t, x) \geq H(t, x) - \varepsilon.
\]

Such a control exists, assuming that the value function is continuous in the space of controls. Consider next the modification of the \( \varepsilon \)-optimal control

\[
\tilde{\nu}^\varepsilon = u I_{t \leq \tau} + \nu^\varepsilon I_{t > \tau},
\]

i.e. the modification is \( \varepsilon \)-optimal after the stopping time \( \tau \), but potentially suboptimal on the interval \([t, \tau]\). Then we have that

\[
H(t, x) \geq H(\tilde{\nu}^\varepsilon)(t, x)
\]

\[
= E_{t,x} \left[ H(\tilde{\nu}^\varepsilon)(\tau, X_\tau^\tilde{\nu}^\varepsilon) + \int_t^\tau F(s, X_s^\tilde{\nu}^\varepsilon, \tilde{\nu}^\varepsilon_s) \, ds \right], \quad \text{(from (6)),}
\]

\[
= E_{t,x} \left[ H(\nu^\varepsilon)(\tau, X_\tau^\nu^\varepsilon) + \int_t^\tau F(s, X_s^\nu^\varepsilon, u_s) \, ds \right], \quad \text{(using (9)),}
\]

\[
\geq E_{t,x} \left[ H(\nu^\varepsilon)(\tau, X_\tau^\nu^\varepsilon) + \int_t^\tau F(s, X_s^\nu^\varepsilon, u_s) \, ds \right] - \varepsilon, \quad \text{(by (8)).}
\]

Taking the limit as \( \varepsilon \searrow 0 \), we have

\[
H(t, x) \geq E_{t,x} \left[ H(\tau, X_\tau^u) + \int_t^\tau F(s, X_s^u, u_s) \, ds \right].
\]

Moreover, since the above holds true for every \( u \in A \) we have that

\[
H(t, x) \geq \sup_{u \in A} E_{t,x} \left[ H(\tau, X_\tau^u) + \int_t^\tau F(s, X_s^u, u_s) \, ds \right].
\]

The upper bound (7) and lower bound (10) form the dynamic programming inequalities. Putting them together, we obtain the theorem below.

6
Figure 2: The DPE is an infinitesimal version of the DPP where the stopping time $\tau$ is the first exit time from a ball of size $\epsilon$ or a small time $h$, whichever occurs first. The figure shows three paths of $X^{u}_{\tau}$, where the stopping time $\tau$ is indicated by the black circles for $\epsilon = 0.05$ and $h = 0.01$.

**Theorem 1. Dynamic Programming Principle for Diffusions.** The value function $H(t, x)$ satisfies the DPP

\[
H(t, x) = \sup_{u \in A} \mathbb{E}_{t, x} \left[ H(\tau, X^{u}_{\tau}) + \int_{t}^{\tau} F(s, X^{u}_{s}, u_{s}) \, ds \right],
\]

for all $(t, x) \in [0, T] \times \mathbb{R}^{n}$ and all stopping times $\tau \leq T$.

This equation is really a sequence of equations that tie the value function to its future expected value, plus the running reward/penalty. Since it is a sequence of equations, an even more powerful equation can be found by looking at its infinitesimal version – the so-called DPE.

### 2.2 Dynamic Programming Equation / Hamilton–Jacobi–Bellman Equation

The DPE is an infinitesimal version of the dynamic programming principle (DPP) (11). There are two key ideas involved:

(i) Setting the stopping time $\tau$ in the DPP to be the minimum between (a) the time it takes the process $X^{u}_{\tau}$ to exit a ball of size $\epsilon$ around its starting point, and (b) a fixed (small) time $h$ – all while keeping it bounded by $T$. This can be
viewed as in Figure 2 and can be stated precisely as

$$\tau = T \land \inf \{s > t : (s - t, |X^u_s - x|) \notin [0, h) \times [0, \epsilon)\}.$$  

Notice that as $h \downarrow 0$, $\tau \downarrow t$, a.s. and that $\tau = t + h$ whenever $h$ is sufficiently small – since as the time span $h$ shrinks, it is less and less likely that $X$ will exit the ball first.

(ii) Writing the value function (for an arbitrary admissible control $u$) at the stopping time $\tau$ in terms of the value function at $t$ using Itô’s lemma. Specifically, assuming enough regularity of the value function, we can write

$$H(\tau, X^u_\tau) = H(t, x) + \int_t^\tau (\partial_t + L^u_s) H(s, X^u_s) ds + \int_t^\tau D_x H(s, X^u_s) \sigma^u_s dW_s,$$

where $\sigma^u_s := \sigma(t, X^u_t, u_t)$ for compactness, $L^u_t$ represents the infinitesimal generator of $X^u_t$, and $D_x H(\cdot)$ denotes the vector of partial derivatives with components $[D_x H(\cdot)]_i = \partial_{x_i} H(\cdot)$. For example, in the one-dimensional case,

$$L^u_t = \mu^u_t \partial_x + \frac{1}{2} (\sigma^u_t)^2 \partial_{xx}$$

$$= \mu(t, x, u) \partial_x + \frac{1}{2} \sigma^2(t, x, u) \partial_{xx}.$$  

Now we proceed as before and derive the DPE in two stages by obtaining two inequalities.

First, taking $u \in A$ to be constant over the interval $[t, \tau]$, applying the lower bound \(10\), and substituting \(12\) into the right-hand side implies that

$$H(t, x) \geq \sup_{u \in A} \mathbb{E}_{t,x} \left[ H(\tau, X^u_\tau) + \int_t^\tau F(s, X^u_s, u_s) ds \right]$$

$$\geq \mathbb{E}_{t,x} \left[ H(\tau, X^\nu_\tau) + \int_t^\tau F(s, X^\nu_s, \nu) ds \right]$$

$$= \mathbb{E}_{t,x} \left[ H(t, x) + \int_t^\tau (\partial_t + L^\nu_s) H(s, X^\nu_s) ds ight. $$

$$\left. + \int_t^\tau D_x H(s, X^\nu_s) \sigma^\nu_s dW_s + \int_t^\tau F(s, X^\nu_s, \nu) ds \right].$$

8
The integrand in the stochastic integral above, i.e. $D_x H(s, X^\nu_s) \sigma^\nu_s$, is bounded on the interval $[t, \tau]$ because we have ensured that $|X^\nu_t - x| \leq \epsilon$ on the interval. Hence, this stochastic integral is the increment of a martingale and we can be assured that its expectation is zero. Therefore,

$$H(t, x) \geq \mathbb{E}_{t, x} \left[ H(t, x) + \int_t^\tau \left\{ (\partial_t + \mathcal{L}^\nu_s) H(s, X^\nu_s) + F(s, X^\nu_s, \nu) \right\} ds \right],$$

and recall that $\tau = t + h$.

Moving the $H(t, x)$ on the left-hand side over to the right-hand side, dividing by $h$ and taking the limit as $h \searrow 0$ yields

$$0 \geq \lim_{h \downarrow 0} \mathbb{E}_{t, x} \left[ \frac{1}{h} \int_t^\tau \left\{ (\partial_t + \mathcal{L}^\nu_s) H(s, X^\nu_s) + F(s, X^\nu_s, \nu) \right\} ds \right] = (\partial_t + \mathcal{L}^\nu_t) H(t, x) + F(t, x, \nu).$$

The second line follows from:

(i) as $h \searrow 0$, $\tau = t + h$ a.s., because the process will not hit the barrier at $\epsilon$ in extremely short periods of time,

(ii) the condition that $|X^u_\tau - x| \leq \epsilon$, which implies that if the process does hit the barrier it is bounded,

(iii) the Mean-Value Theorem allows us to write $\lim_{h \downarrow 0} \frac{1}{h} \int_t^{t+h} \omega_s ds = \omega_t$, and

(iv) the process starts at $X^u_t = x$.

Since the above inequality holds for arbitrary $\nu \in \mathcal{A}$, it follows that

$$\partial_t H(t, x) + \sup_{\nu \in \mathcal{A}} (\mathcal{L}^\nu_t H(t, x) + F(t, x, \nu)) \leq 0. \quad (13)$$

Next, we show that the inequality is indeed an equality. To show this, suppose that $u^*$ is an optimal control, then from (11), we have

$$H(t, x) = \mathbb{E}_{t, x} \left[ H(\tau, X^{u^*_\tau}) + \int_\tau^\tau F(s, X^{u^*_s}, u^*) ds \right].$$
As above, by applying Itô’s lemma to write \( H(\tau, X^u_\tau) \) in terms of \( H(t, x) \) plus the integral of its increments, taking expectations, and then the limit as \( h \downarrow 0 \), we find that

\[
\partial_t H(t, x) + \mathcal{L}_t^{u^*} H(t, x) + F(t, x, u^*) = 0.
\]

(14)

Thus, combining (14) with (13), we arrive at the DPE (also known in this context as the Hamilton–Jacobi–Bellman equation)

\[
\partial_t H(t, x) + \sup_{u \in A} (\mathcal{L}_t^{u} H(t, x) + F(t, x, u)) = 0,
\]

\[
H(T, x) = G(x).
\]

(15)

The terminal condition above follows from the definition of the value function in (5) from which we see that the running reward/penalty drops out and \( G(X^u_T) \) is \( \mathcal{F}_T \) measurable.

Notice that the optimisation of the control in (15) is only over its value at time \( t \), rather than over the whole path of the control. Hence, it appears that the optimal control can be obtained pointwise. Treating the value function as known, the optimal control can often be found in feedback control form in terms of the value function itself. Substituting the feedback control back into (15) results in non-linear PDEs. In fact the function

\[
\mathcal{H}(t, x, D_x H, D_x^2 H) = \sup_{u \in A} (\mathcal{L}_t^u H(t, x) + F(t, x, u))
\]

is called the Hamiltonian of the associated stochastic control problem. Here \( D_x H \) and \( D_x^2 H \) represent the collection of first and second order derivatives of the value function \( H \) respectively (recall that in general \( X \) is vector-valued) and in particular

\[
[D H]_k = \partial_{x^k} H, \quad \text{and} \quad [D_x^2 H]_{jk} = \partial_{x^j x^k} H,
\]

which appear in the infinitesimal generator.

2.3 Verification

The derivation of the DPE (15) in the previous section provides a necessary condition for the value function. A pertinent question, however, is whether or not a solution
of the DPE does indeed provide the solution to the original control problem. The main workhorse for showing this is indeed the case, when classical solutions to the associated DPE exist, is the so-called verification theorem. We state the result below and refer the reader to many of the excellent texts on optimal control for its proof, see e.g., Yong & Zhou (1999), Fleming & Soner (2006), Øksendal & Sulem (2007), and Pham (2010).

**Theorem 2. Verification Theorem.** Let $\psi \in C^{1,2}([0,T] \times \mathbb{R}^n)$ and satisfies, for all $u \in A$,

$$
\begin{align*}
\partial_t \psi(t, x) + (L_u^u \psi(t, x) + F(t, x, u)) &\leq 0, \quad \forall (t, x) \in [0, T] \times \mathbb{R}^n, \\
G(x) - \psi(T, x) &\leq 0.
\end{align*}
$$

Then

$$
\psi(t, x) \geq H^u(t, x), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^n,
$$

for all Markov controls $u \in A$.

Moreover, if for every $(t, x) \in [0, T] \times \mathbb{R}^n$, there exists measurable $u^*(t, x)$ such that

$$
0 = \partial_t \psi(t, x) + \left( L_{u^*(t,x)} \psi(t, x) + F(t, x, u^*(t, x)) \right)
$$

$$
= \partial_t \psi(t, x) + \sup_{u \in A} \left( L_u^u \psi(t, x) + F(t, x, u) \right), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^n,
$$

with $\psi(T, x) = G(x)$, and the SDE

$$
dX_s^* = \mu(t, X_s^*, u^*(t, X_s^*)) dt + \sigma(t, X_s^*, u^*(t, X_s^*)) dW_s, \quad X_t^* = x,
$$

admits a unique solution and $\{u^*(s, X_s^*)\}_{t \leq s \leq T} \in A$, then

$$
H(t, x) = \psi(t, x), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^n,
$$

and $u^*$ is an optimal Markov control.

The theorem states that if we can find a solution to the DPE and demonstrate that it is a classical solution, i.e. once differentiable in time and twice differentiable in the state variables, and the resulting control is admissible, then the solution is
indeed the value function we seek and the resulting control is the optimal one – at least an optimal Markov control. Another key result is that, under some more technical assumptions, the optimal control is Markov, even if we search over general \( \mathcal{F} \)-predictable controls. See, e.g., Theorem 11.2.3 in Øksendal (2010).

3 Control for Counting Processes

In the previous section, diffusion processes were the driving sources of uncertainty in the control problem. In many circumstances, and in particular problems related to algorithmic and high-frequency trading, counting processes will be used to drive uncertainty. There are many features that can be incorporated into the analysis, but the general approach remains the same, and as such only the case of a single counting process with controlled intensity will be investigated. This amounts to treating doubly stochastic Poisson processes, or Cox processes, which are counting processes with intensity that itself is a stochastic process and in this case at least partially controlled.

Consider the situation in which the agent can control the frequency of the jumps in a counting process \( N \) and does so to maximise some target. In this case, the control problem is of the general form

\[
H(n) = \sup_{u \in A_{0,T}} \mathbb{E} \left[ G(N^n_T) + \int_0^T F(s, N^n_s, u_s) \, ds \right],
\]

where \( u = (u_t)_{0 \leq t \leq T} \) is the control process, \( (N^n_t)_{0 \leq t \leq T} \) is a controlled doubly stochastic Poisson process (starting at \( N_{0^-} = n \)) with intensity \( \lambda^u_t = \lambda(t, N^n_t, u_t) \) so that \( \hat{N}^u_t \) is a martingale, \( \hat{N}^u_t = N_t - \int_0^t \lambda^u_s \, ds \)

is a martingale, \( A \) is a set of \( \mathcal{F} \)-predictable processes such that \( \hat{N} \) is a true martingale, \( G : \mathbb{R} \mapsto \mathbb{R} \) is a terminal reward and \( F : \mathbb{R}_+ \times \mathbb{R}^2 \mapsto \mathbb{R} \) is a running reward/penalty.

As before, the functions \( G \) and \( F \) are assumed uniformly bounded.
For an arbitrary admissible control $u$, the performance criteria $H^u(n)$ is given by

$$H^u(n) = \mathbb{E} \left[ G(N_T^n) + \int_0^T F(s, N_s^n, u_s) \, ds \right],$$

and the agent seeks to maximise this performance criteria, i.e.

$$H(n) = \sup_{u \in A_{0,T}} H^u(n).$$

In the next two subsections we provide a concise derivation of the DPP and the associated DPE.

### 3.1 The Dynamic Programming Principle

As before, the original problem is embedded into a larger class of problems indexed by time $t \in [0, T]$ by first defining

$$H(t, n) := \sup_{u \in A_{t,T}} H^u(t, n), \quad \text{and} \quad (19a)$$

$$H^u(t, n) := \mathbb{E}_{t,n} \left[ G(N_T^n) + \int_t^T F(s, N_s^n, u_s) \, ds \right], \quad (19b)$$

where the notation $\mathbb{E}_{t,n}[\cdot]$ represents expectation conditional on $N_{t-} = n$. Next, take an arbitrary admissible strategy $u$ and flow the $N$ process forward in time from $t$ to an arbitrary stopping time $\tau \leq T$. Then, conditional on $N_{\tau}^u$, the contribution of the running reward/penalty from $\tau$ to $T$ and the terminal reward can be viewed as the performance criteria starting from the new value of $N_{\tau}^u$. This allows the value function to be written in terms of the expectation of its future value at $\tau$ plus the reward between now and $\tau$.

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

$$H^u(t, n) = \mathbb{E}_{t,n} \left[ G(N_T^n) + \int_t^T F(s, N_s^n, u_s) \, ds + \int_t^\tau F(s, N_s^n, u_s) \, ds \right]$$

$$= \mathbb{E}_{t,n} \left[ \mathbb{E}_{\tau,N_\tau} \left[ G(N_T^n) + \int_\tau^T F(s, N_s^n, u_s) \, ds \right] + \int_t^\tau F(s, N_s^n, u_s) \, ds \right]$$

$$= \mathbb{E}_{t,n} \left[ H^u(\tau, N_{\tau}^u) + \int_t^\tau F(s, N_s^n, u_s) \, ds \right].$$

(20)
Now, \( H(t, n) \geq H^u(t, n) \) for an arbitrary admissible control \( u \) (with equality holding if \( u \) is the optimal control \( u^* \)) and an arbitrary \( n \). Hence, on the right-hand side of (20) the performance criteria \( H^u(\tau, N^u_\tau) \) is bounded above by the value function \( H(\tau, N^u_\tau) \) and the equality can be replaced by an inequality with the value function (and not the performance criteria) showing up under the expectation:

\[
H^u(t, n) \leq \mathbb{E}_{t,n} \left[ H(\tau, N^u_\tau) + \int_t^\tau F(s, N^u_s, u_s) \, ds \right] \\
\leq \sup_{u \in \mathcal{A}} \mathbb{E}_{t,n} \left[ H(\tau, N^u_\tau) + \int_t^\tau F(s, N^u_s, u_s) \, ds \right].
\] (21)

Note that on the right-hand side of the first line of (21), the arbitrary control \( u \) only acts over the interval \([t, \tau]\) and the optimal control is implicitly incorporated in the value function \( H(\tau, N^u_\tau) \), but with the state variable \( N \) beginning at the point where the arbitrary control \( u \) caused \( N \) to flow, namely \( N^u_\tau \). Next, taking the best control on the right-hand side must also dominate the left-hand side, since the arbitrary one does. Then, the right-hand side is no longer dependent on the arbitrary control. Finally, taking a supremum over all admissible controls on the left-hand side allows us to replace the left-hand side with the value function and provides us with the first inequality

\[
H(t, n) \leq \sup_{u \in \mathcal{A}} \mathbb{E}_{t,n} \left[ H(\tau, N^u_\tau) + \int_t^\tau F(s, N^u_s, u_s) \, ds \right].
\] (22)

To obtain the reverse inequality, take an \( \epsilon \)-optimal control denoted by \( \nu^\epsilon \in \mathcal{A} \) such that

\[
H(t, x) \geq H^{\nu^\epsilon}(t, x) \geq H(t, x) - \epsilon.
\] (23)

Such a control exists if the value function is continuous in the space of controls. Consider its modification up to time \( \tau \):

\[
\tilde{\nu}^\epsilon = u_t \mathbbm{1}_{t \leq \tau} + \nu^\epsilon \mathbbm{1}_{t > \tau},
\] (24)
where \( u \in \mathcal{A} \) is an arbitrary admissible control. Then we have that

\[
H(t, n) \geq H^{\bar{\nu}}(t, n)
\]

\[
= \mathbb{E}_{t, n}\left[ H^{\bar{\nu}}(\tau, N^{\bar{\nu}}_{\tau}) + \int_{t}^{\tau} F(s, N_{s}^{\bar{\nu}}, \bar{\nu}_{s}) \, ds \right], \quad \text{(from (20))},
\]

\[
= \mathbb{E}_{t, n}\left[ H^{\bar{\nu}}(\tau, N_{\tau}^{u}) + \int_{t}^{\tau} F(s, N_{s}^{u}, u_{s}) \, ds \right], \quad \text{(using (24))},
\]

\[
\geq \mathbb{E}_{t, n}\left[ H(\tau, N_{\tau}^{u}) + \int_{t}^{\tau} F(s, N_{s}^{u}, u_{s}) \, ds \right] - \epsilon, \quad \text{(by (23))}.
\]

Since the above holds for every \( u \) and every \( \epsilon \), it holds for the sup, and we take the limit as \( \epsilon \downarrow 0 \) to find the inequality

\[
H(t, n) \geq \sup_{u \in \mathcal{A}} \mathbb{E}_{t, n}\left[ H(\tau, N_{\tau}^{u}) + \int_{t}^{\tau} F(s, N_{s}^{u}, u_{s}) \, ds \right]. \tag{25}
\]

Putting (25) together with (22), we obtain the theorem below.

**Theorem 3. Dynamic Programming Principle for Counting Processes.**

The value function (18) satisfies the DPP

\[
H(t, n) = \sup_{u \in \mathcal{A}} \mathbb{E}_{t, n}\left[ H(\tau, N_{\tau}^{u}) + \int_{t}^{\tau} F(s, N_{s}^{u}, u_{s}) \, ds \right], \tag{26}
\]

for all \((t, n) \in [0, T] \times \mathbb{Z}_{+}\) and all stopping times \( \tau \leq T \).

### 3.2 Dynamic Programming Equation / Hamilton–Jacobi–Bellman Equation

This section develops the DPE satisfied by the value function, which we obtain by looking at the DPP over infinitesimal amounts of time. Analogous to the diffusion case, set the stopping time \( \tau \) in the DPP to be the minimum of (i) the time it takes the process \( N_{s}^{u} \) to exit a ball of size \( \epsilon \) around its starting point, and (ii) a fixed (small) time \( h \), all while keeping it bounded by \( T \).

\[
\tau = T \wedge \inf\{ s > t : (s - t, |N_{s}^{u} - n|) \notin [0, h] \times [0, \epsilon] \}.
\]

15
Next, write the value function (for an arbitrary admissible control $u$) at the stopping time $\tau$ in terms of the value function at $t$ using Itô’s lemma. Specifically,

$$H(\tau, N^u_\tau) = H(t, n) + \int_t^{\tau} (\partial_t + L^u_s) H(s, N^u_s) \, ds + \int_t^{\tau} [H(s, N^u_{s^-}) + 1] - H(s, N^u_{s^-}) \, d\hat{N}^u_s ,$$

where $L^u_t$ represents the infinitesimal generator of $N^u_t$ and acts on functions $h : \mathbb{R}_+ \times \mathbb{Z}_+ \mapsto \mathbb{R}$ as follows:

$$L^u_t h(t, n) = \lambda(t, n, u) [h(t, n + 1) - h(t, n)].$$

(27)

Taking $\nu \in \mathcal{A}$ to be constant over the interval $[t, \tau]$, applying the lower bound (25), and substituting (27) into the right-hand side implies that

$$H(t, n) \geq \sup_{u \in \mathcal{A}} \mathbb{E}_{t,n} \left[ H(\tau, N^u_\tau) + \int_t^{\tau} F(s, N^u_s, u_s) \, ds \right] = \mathbb{E}_{t,n} \left[ H(\tau, N^\nu_\tau) + \int_t^{\tau} F(s, N^\nu_s, \nu) \, ds \right] = \mathbb{E}_{t,n} \left[ H(t, n) + \int_t^{\tau} ((\partial_t + L^\nu_s) H(s, N^\nu_s) + F(s, N^\nu_s, \nu)) \, ds \right].$$

The third equality follows because the stochastic integral with respect to $\hat{N}_t$ has zero mean since $\hat{N}_t$ is a martingale. The martingale property is guaranteed because we have $|N^\nu_t - n| \leq \epsilon$, and so $L^\nu_t H(t, N^\nu_t)$ is bounded. Subtracting $H(t, n)$, dividing by $h$, and taking the limit as $h \downarrow 0$ implies that

$$0 \geq \lim_{h \downarrow 0} \mathbb{E}_{t,n} \left[ \frac{1}{h} \int_t^{\tau} ((\partial_t + L^\nu_s) H(s, N^\nu_s) + F(s, N^\nu_s, \nu)) \, ds \right] = (\partial_t + L^\nu_t) H(t, n) + F(t, n, \nu).$$

(28)

In the above, the control is constant, and equals $\nu$, on the interval $[t, \tau]$. The second line follows, as before, from:

(i) as $h \downarrow 0, \tau = t + h$ a.s. since the process will not hit the barrier of $\epsilon$ in extremely short periods of time,
(ii) the condition that $|N^u_t - n| \leq \epsilon$, so that if the process does hit the barrier, it is bounded,

(iii) the Fundamental Theorem of Calculus: $\lim_{h \downarrow 0} \frac{1}{h} \int_{t}^{t+h} \omega_s \, ds = \omega_t$, and

(iv) $N^u_{t^-} = n$.

Since the above inequality holds for arbitrary $u \in A$, it follows that

$$\partial_t H(t, n) + \sup_{u \in A} (\mathcal{L}_t^u H(t, n) + F(t, n, u)) \leq 0.$$  \hfill (29)

Next, we show the inequality is indeed an equality. For this purpose, suppose that $u^* \in A$ is an optimal control, then from the DPP (26) we have

$$H(t, n) = \mathbb{E}_{t, n} \left[ H(\tau, N^u_{\tau}) + \int_{t}^{\tau} F(s, N^u_s, u^*_s) \, ds \right].$$

Repeating the steps from above (i.e. applying Itô’s lemma, taking expectations, and the limit as $h \downarrow 0$) we find that

$$\partial_t H(t, n) + \mathcal{L}_t^{u^*} H(t, n) + F(t, n, u^*) = 0.$$ 

Combining this result with (29), we arrive at the DPE (also known in this context as the Hamilton–Jacobi–Bellman equation)

$$\begin{cases} 
\partial_t H(t, n) + \sup_{u \in A} (\mathcal{L}_t^u H(t, n) + F(t, n, u)) = 0, \\
H(T, n) = G(n). 
\end{cases} \hfill (30)$$

The terminal condition follows from the observation that the integral in the optimisation problem (19) vanishes as $t \nearrow T$. Recall that the generator $\mathcal{L}_t^u$ for the controlled process acts as follows:

$$\mathcal{L}_t^u H(t, n) = \lambda(t, n, u) [H(t, n + 1) - H(t, n)].$$

As in the diffusion case, the optimisation of the control in (30) is only over its value at time $t$, rather than over the whole path of the control. Hence, it appears
that the optimal control can be obtained pointwise. Treating the value function as known, the optimal control can often be found in feedback control form in terms of the value function itself. Substituting the feedback control back into (30) typically results in non-linear PDEs. In this Poisson case, the supremum in (30) has a very simple form since

\[
\sup_{u \in A_t} \left( \mathcal{L}_t H(t, n) + F(t, n, u) \right) = \sup_{u \in A_t} \{ \lambda(s, n, u) \left[ H(t, n + 1) - H(t, n) \right] + F(t, n, u) \},
\]

and hence, if \( F = 0 \), the optimal choice of the control is to make \( \lambda(s, n, u) \) as large as possible if \( H(t, n + 1) - H(t, n) > 0 \) and as small as possible if \( H(t, n + 1) - H(t, n) < 0 \). Such controls are called bang-bang controls because the quantity being controlled (in this case the intensity) reaches its extremal points. To make the problem more interesting we either need to (i) include a running reward/penalty (i.e. \( F \neq 0 \)) or, more interestingly (and relevant), (ii) introduce another stochastic process which is driven by the counting process, and therefore controlled indirectly, and have this stochastic process affect the agent’s terminal and running rewards.

### 3.3 Combined Diffusion and Jumps

As already hinted above, there are many situations in which the agent is exposed to more than one source of uncertainty. Typically, an agent is faced with control problems where both diffusive and jump uncertainty appear, and she may be able to control all or only parts of the system. Such scenarios will appear in several of the algorithmic trading problems that arise in the sections ahead and here we simply state the main results for a fairly general class of models.

First, let \( \mathbf{N}^u = (\mathbf{N}^u_t)_{0 \leq t \leq T} \) denote a collection of counting processes (of dim \( p \)) with controlled intensities \( \lambda^u = (\lambda^u_t)_{0 \leq t \leq T} \), and let \( \mathbf{u} = (\mathbf{u}_t)_{0 \leq t \leq T} \) denote the control processes (of dim \( m \)). Furthermore, let \( \mathbf{W} = (\mathbf{W}_t)_{0 \leq t \leq T} \) denote a collection of independent Brownian motions (of dim \( m \)). Next, let \( \mathbf{X}^u = (\mathbf{X}^u_t)_{0 \leq t \leq T} \) denote the controlled processes (of dim \( m \)) which will appear directly in the agent’s performance criteria. If one (or more) of the counting processes \( \mathbf{N}^i_t \) should appear in the perfor-
mance criteria or in the dependence in $\mu^u_t$, $\sigma^u_t$ and/or $\gamma^u_t$, shown below, then take $X^j_t = N^j_t$ for some $j$, i.e. include it through one of the components in $X$. We next assume that the controlled $X^u$ processes satisfy the SDEs

$$dX^u_t = \mu^u_t dt + \sigma^u_t dW^u_t + \gamma^u_t dN^u_t. \quad (31a)$$

With a slight abuse of notation we write the controlled drift, volatility and jump size as

$$\mu^u_t := \mu(t, X^u_t, u_t), \quad (m \times 1 \text{ vector}), \quad (31b)$$

$$\sigma^u_t := \sigma(t, X^u_t, u_t), \quad (m \times m \text{ matrix}), \quad (31c)$$

$$\gamma^u_t := \gamma(t, X^u_t, u_t), \quad (m \times p \text{ matrix}). \quad (31d)$$

Furthermore, we assume that the controlled intensity takes the form

$$\lambda^u_t := \lambda(t, X^u_t, u_t). \quad (31e)$$

The modelling approach above implies that the agent can control, in general, the drift, volatility, jump size and jump arrivals. Precisely how much control she has depends on the specific form of the various functions appearing in (31b), (31c), (31d) and (31e).

The agent’s performance criteria is given by

$$H^u(t, x) = \mathbb{E}_{t, x} \left[ G(X^u_T) + \int_t^T F(s, X^u_s, u_s) \, ds \right], \quad (32a)$$

and the value function is then given by the usual expression

$$H(t, x) = \sup_{u \in A_{t, T}} H^u(t, x). \quad (32b)$$

Repeating the analysis along similar lines to the previous sections, we arrive at the DPP for the combined problem.

**Theorem 4. Dynamic Programming Principle for Jump-Diffusions.** The value function (32) satisfies the DPP

$$H(t, x) = \sup_{u \in A_{t, \tau}} \mathbb{E}_{t, x} \left[ H(\tau, X^u_\tau) + \int_t^\tau F(s, X^u_s, u_s) \, ds \right], \quad (33)$$

for all $(t, x) \in [0, T] \times \mathbb{R}^m$ and all stopping times $\tau \leq T$. 19
Moreover, following similar arguments as before, one can develop a DPE

\[
\begin{aligned}
\partial_t H(t, x) + \sup_{u \in A_t} (L_t^u H(t, x) + F(t, x, u)) &= 0, \\
H(T, x) &= G(x),
\end{aligned}
\]

(34)

where the infinitesimal generator \( L_t^u \) acts as follows:

\[
L_t^u H(t, x) = \mu(t, x, u) \cdot D_x H(t, x) + \frac{1}{2} \sigma(t, x, u) \sigma(t, x, u)' D_x^2 H(t, x) \\
+ \sum_{j=1}^p \lambda_j(t, x, u) \left[ H(t, x + \gamma_j(t, x, u)) - H(t, x) \right],
\]

where \( \gamma_j \) denotes the vector corresponding to the \( j \)th column of \( \gamma \), \( D_x H \) represents the vector of partial derivatives with respect to \( x \), and \( D_x^2 H \) represents the matrix of (mixed) second order partial derivatives with respect to \( x \).

The various terms in the generator can be easily interpreted: in the first line, the first term represents the (controlled) drift of \( X \), and the second term represents the (controlled) volatility of \( X \); in the second line, the intensity for each counting process is shown separately, and each term in the sum has the (controlled) rate of arrival of that jump component through \( \lambda_j \), the difference terms that appear are due to the jump arriving and causing that component of \( N \) to increment, and simultaneously cause (potentially all components of) \( X \) to jump by \( \gamma_j \).

4 Optimal Stopping

In many circumstances the agent wishes to find the best time at which to enter or exit a given strategy. The classical finance example of this is the ‘American put option’ in which an agent who owns the option has the right, at any point in time up to, and including, the maturity date \( T \), to exercise the option by receiving the amount of cash \( K \) in exchange for the underlying asset. The net cash value of this transaction at the exercise date \( \tau \) is \((K - S_\tau)\). Naturally, the agent will not exercise when \( S_\tau > K \), hence the effective payoff is \((K - S_\tau)_+\) where \((\cdot)_+ = \max(\cdot, 0)\). This simple observation provides only a trivial part of the strategy: to determine the full
strategy, the agent seeks the stopping time $\tau$ which maximises the discounted value of the payoff, i.e. she searches for the stopping time which attains the supremum (if possible)

$$\sup_{\tau \in \mathcal{T}} \mathbb{E} \left[ e^{-r\tau} (K - S_\tau)_+ \right],$$

where $\mathcal{T}$ are the $\mathcal{F}$-stopping times bounded by $T$. This is just one of many such problems, and in general, problems which seek optimal stopping times are called \textbf{optimal stopping problems}\textsuperscript{1}. Similar to optimal control problems, optimal stopping problems admit a DPP and have an infinitesimal version, i.e. a DPE. In this section we provide a concise outline of the DPP and DPE for optimal stopping problems.

Rather than first developing the diffusion case, then the jump, and then the jump-diffusion, here we begin immediately with a fairly general jump-diffusion model. For this purpose, let $X = (X_t)_{0 \leq t \leq T}$ denote a vector-valued process of dim $m$ satisfying the SDE

$$dX_t = \mu(t, X_t) \, dt + \sigma(t, X_t) \, dW_t + \gamma(t, X_t) \, dN_t,$$

where $\mu : [0, T] \times \mathbb{R}^m \mapsto \mathbb{R}^m$, $\sigma : [0, T] \times \mathbb{R}^m \mapsto \mathbb{R}^m \times \mathbb{R}^m$, $\gamma : [0, T] \times \mathbb{R}^m \mapsto \mathbb{R}^m \times \mathbb{R}^m$, $W$ is an $m$-dimensional Brownian motion with independent components, and $N$ is an $m$-dimensional counting process with intensities $\lambda(t, X_t)$. The filtration $\mathcal{F}$ is the natural one generated by $X$, and the generator of the process $\mathcal{L}_t$ acts on twice differentiable functions as follows:

$$\mathcal{L}_t h(t, x) = \mu(t, x) \cdot D_x h(t, x) + \frac{1}{2} \text{tr} \sigma(t, x) \sigma(t, x)' D_x^2 h(t, x) + \sum_{j=1}^p \lambda_j(t, x) \left[ h(t, x + \gamma_j(t, x)) - h(t, x) \right].$$

Recall that $D_x h$ represents the vector of partial derivatives with respect to $x$, and $D_x^2 h$ represents the matrix of (mixed) second order partial derivatives with respect to $x$.

The agent then has a performance criterion, for each $\mathcal{F}$-stopping time $\tau \in \mathcal{T}_{[t, T]}$, given by

$$H^\tau(t, x) = \mathbb{E}_{t,x} \left[ G(X_\tau) \right], \quad (35a)$$
where $G(\mathbf{X}_\tau)$ is the reward upon exercise, and she seeks to find the value function

$$H(t, \mathbf{x}) = \sup_{\tau \in \mathcal{T}_t} H^\tau(t, \mathbf{x}),$$

(35b)

and the stopping time $\tau$ which attains the supremum if it exists. At first glance, it appears that we have omitted the running reward/penalty $\int_t^\tau F(s, \mathbf{X}_s) \, ds$ that we included when studying optimal control. Such terms can, however, be cast into the above form by choosing one of the components of $\mathbf{X}$, say $X^1$, to satisfy

$$dX_t^1 = F(t, X_t^2, \ldots, X_t^m) \, dt,$$

and writing $G(\mathbf{x}) = x^1 + \tilde{G}(x^2, \ldots, x^m)$. It is therefore no loss of generality to consider only terminal rewards. In the stochastic control case, we kept the explicit running reward/penalty since it traditionally appears there. Now it will prove more convenient to absorb the running reward/penalty in $G$. Similarly, we can incorporate discounting by introducing a second state process which equals the discount factor up to that point in time, and modify $G$ accordingly.

From the posing of the stopping problem in (35), intuitively, we see that the agent is attempting to decide between stopping ‘now’ and receiving the reward $G$, or continuing to hold off in hopes of receiving a larger reward in the future. However, on closer examination it seems clear that she should continue to wait at the point $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^m$ as long as the value function has not attained a value of $G(\mathbf{x})$ there. This motivates the definition of the stopping region $S$, which we define as

$$S = \left\{ (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^m : H(t, \mathbf{x}) = G(\mathbf{x}) \right\}.$$

It should be evident that whenever $(t, \mathbf{x}) \in S$, i.e. if the state of the system lies in $S$, it is optimal to stop immediately – since the agent cannot improve beyond $G$ by the definition of the value function in (35). The complement of this region, $S^c$, is known as the continuation region $C$. In this region, the agent can still improve her value and therefore continues to wait. Both regions can be generically visualised as in Figure 3.

The difficulties in optimal stopping problems arise because the region $S$ (or equivalently its boundary $\partial S$) must be solved simultaneously with the value function itself.
Hence, the DPEs which arise in this context are free-boundary problems, also called obstacle problems, or variational inequalities. Such PDEs are more difficult to solve, and even very simple examples typically do not admit explicit solutions – take for example the American put option when the underlying asset is a geometric Brownian motion. Nonetheless, a non-linear PDE is often enough to characterise the solution and many numerical schemes exist for solving them.

4.1 The Dynamic Programming Principle

Reasoning along similar lines as before, we can show that the following DPP applies.

**Theorem 5. Dynamic Programming Principle for Stopping Problems.** Dynamic Programming Principle!stoppingThe value function (35) satisfies the DPP

\[
H(t, x) = \sup_{\tau \in [t, T]} \mathbb{E}_{t, x} \left[ G(X_\tau) 1_{\tau < \theta} + H(\theta, X_\theta) 1_{\tau \geq \theta} \right],
\]

for all \((t, x) \in [0, T] \times \mathbb{R}^m\) and all stopping times \(\theta \leq T\).

The intuition of this DPP is that if the optimal stopping time \(\tau^*\) occurs prior to \(\theta\), then the agent’s value function equals the reward at \(\tau^*\). If, however, the agent has not stopped by \(\theta\), then at \(\theta\) she receives the value function evaluated at the current state of the system.
4.2 Dynamic Programming Equation

As in the optimal control problems studied earlier, the DPP for optimal stopping can be recast in its infinitesimal version – a much more useful form for computation – i.e. it can be used to derive a DPE. This time, we state the result first and then provide the proof. We follow closely the proof in Touzi (2013) as it provides a nice alternative to the usual construction. It also opens the door for the interested reader to move on to studying the viscosity solution approach to stochastic optimal control and optimal stopping, which relies heavily on the proof by contradiction approach taken here.

Theorem 6. Dynamic Programming Equation for Stopping Problems. Dynamic Programming Equation for Stopping Problems. Assume that the value function $H(t, x)$ is once differentiable in $t$ and all second order derivatives in $x$ exist, i.e. $H \in C^{1,2}([0, T], \mathbb{R}^m)$, and that $G : \mathbb{R}^m \mapsto \mathbb{R}$ is continuous. Then $H$ solves the variational inequality Variational Inequality, also known as an obstacle problem, or free-boundary problem:

$$\max \left\{ \frac{\partial}{\partial t} H + \mathcal{L}_t H , \ G - H \right\} = 0 , \text{ on } \mathcal{D} ,$$

(37)

where $\mathcal{D} = [0, T] \times \mathbb{R}^m$.

Proof. The proof is broken up into two steps by showing (on $\mathcal{D}$) that (i) the left-hand side is first smaller than or equal to zero, and that (ii) by contradiction the left-hand side is also greater than or equal to zero. Therefore, equality must hold.

1. First we establish that

$$\max \left\{ \frac{\partial}{\partial t} H + \mathcal{L}_t H , \ G - H \right\} \leq 0 , \text{ on } \mathcal{D} .$$

To show this, first note that the constant stopping rule $\tau = t$ is admissible. Therefore, the performance criteria on this constant rule equals $G$ and we must have $H \geq G$ so that $G - H \leq 0$.

Next, take any point $(t_0, x_0) \in \mathcal{D}$. We show that the desired inequality holds at this point. Consider the sequence of stopping times indexed by $h > 0$,

$$\theta_h = \inf \left\{ t > t_0 : (t, ||X_t - x_0||) \notin [t_0, t_0 + h] \times 1 \right\} ,$$

24
where $||\cdot||$ denotes the Euclidean norm. That is, we take a stopping time equal to the minimum of $t_0 + h$ and the time it takes $X$ to exit a ball of size 1 from its current position. As long as $h < T - t_0$, this is an admissible stopping time. Therefore, by taking $\tau = t_0$, the DPP (36) implies that

$$H(t_0, x_0) \geq E_{t_0, x_0} [H(\theta h, X_{\theta h})].$$

We then expand the term under the expectation using Itô’s lemma for jump-diffusions to write

$$H(\theta h, X_{\theta h}) = H(t_0, x_0) + \int_{t_0}^{\theta h} \left\{ \partial_t H(t, X_t) + \mathcal{L}_t H(t, X_t) \right\} dt + \int_{t_0}^{\theta h} (\sigma(t, X_t) D_x H(t, X_t))' dW_t + \sum_{j=1}^p \int_{t_0}^{\theta h} \left[ H(t, X_t + \gamma_j(t, X_t)) - H(t, X_t) \right] d\widehat{N}_t^j,$$

where $\widehat{N}_t = N_t - \int_0^t \lambda(s, X_s) ds$ are the compensated versions of the counting process, and $\gamma_j$ denotes the $j^{th}$ column of $\gamma$.

Since the stopping time $\theta h$ is chosen so that the process $X$ remains bounded by the ball of size 1 around $x_0$ plus the potential of a jump (which we assume bounded), it follows that the stochastic integrals with respect to both the Brownian motions and the compensated counting processes vanish under the expectation. Hence, we have

$$0 \geq E_{t_0, x_0} \left[ \int_{t_0}^{\theta h} \left\{ \partial_t H(t, X_t) + \mathcal{L}_t H(t, X_t) \right\} dt \right].$$

Dividing by $h$ and taking $h \searrow 0$, in which case $\theta h \searrow h$ a.s. (since $X_t$ will a.s. not hit the edge of the bounding ball), the Mean-Value Theorem implies that

$$\partial_t H(t, x_0) + \mathcal{L}_t H(t, x_0) \leq 0.$$

This completes the first part of the proof.

2. We next show by contradiction that

$$\max \left\{ \partial_t H + \mathcal{L}_t H , G - H \right\} \geq 0, \quad \text{on } \mathcal{D}.$$
If the above inequality does not hold on $\mathcal{D}$, then there exists a point $(t_0, x_0) \in \mathcal{D}$ such that

$$G(t_0, x_0) - H(t_0, x_0) < 0 \quad \text{and} \quad (\partial_t + L_t)H(t_0, x_0) < 0. \quad (38)$$

We show that (38) contradicts the DPP in (36). To this end, introduce a new function $\varphi_\varepsilon$ which approximates the value function near $(t_0, x_0)$ but locally dominates it:

$$\varphi_\varepsilon(t, x) := H(t, x) + \varepsilon \left( ||x - x_0||^4 + |t - t_0|^2 \right), \quad \forall (t, x) \in \mathcal{D}, \quad \varepsilon > 0.$$ 

Under the assumption (38), for $\varepsilon > 0$ but sufficiently small, there is a small neighbourhood around $(t_0, x_0)$ on which the value function is at least $\delta$ larger than the reward $G$ and for which the operator $(\partial_t + L_t)$ renders the approximation $\varphi_\varepsilon$ negative. More precisely, there exists $h > 0$ and $\delta > 0$ such that

$$H \geq G + \delta \quad \text{and} \quad (\partial_t + L_t)\varphi_\varepsilon \leq 0 \quad \text{on} \quad \mathcal{D}_h := [t_0, t_0 + h] \times B_h, \quad (39)$$

where $B_h$ is a ball of size $h$ around $x_0$, i.e. $B_h = \{ x \in \mathbb{R}^m : ||x - x_0|| \leq h \}$. Also, near $(t_0, x_0)$, $\varphi_\varepsilon$ is locally larger than $H$, hence,

$$-\zeta := \max_{\partial \mathcal{D}_h} (H - \varphi_\varepsilon) < 0, \quad (40)$$

where $\partial \mathcal{D}_h$ represents the boundary of the set $\mathcal{D}_h$.

We now take a stopping time equal to the first time the process exits this ball:

$$\theta := \inf \left\{ t > t_0 : (t, X_t) \notin \mathcal{D}_h \right\}.$$ 

Take a second stopping rule, this time arbitrary, $\tau \in \mathcal{T}_{[t, T]}$, and let $\psi = \tau \wedge \theta$. Then we have

$$H(\psi, X_\psi) - H(t_0, x_0) = (H - \varphi_\varepsilon)(\psi, X_\psi) + (\varphi_\varepsilon(\psi, X_\psi) - \varphi_\varepsilon(t_0, x_0)), \quad (41)$$

since $\varphi_\varepsilon$ and $H$ coincide at $(t_0, x_0)$. From Itô’s lemma, and the fact that $X_\psi$ is bounded due to stopping the first time $X$ exits the ball $B_h$, we have

$$\mathbb{E}_{t_0, x_0} [\varphi_\varepsilon(\psi, X_\psi) - \varphi_\varepsilon(t_0, x_0)] = \mathbb{E}_{t_0, x_0} \left[ \int_{t_0}^{\psi} (\partial_t + L_t) \varphi_\varepsilon(t, X_t) dt \right] \leq 0.$$
The diffusive and jump terms vanish because they are martingales, and the inequality follows from the second inequality in (39).

Hence, putting this together with (41), we have

\[ \mathbb{E}_{t_0, x_0} [H(\psi, X_\tau) - H(t_0, x_0)] \leq \mathbb{E}_{t_0, x_0} [(H - \varphi_\epsilon)(\psi, X_\psi)] \leq -\zeta \mathbb{P}(\tau \geq \theta), \]

where the second inequality follows from (40). By rearranging to isolate \( H(t_0, x_0), \)

we have

\[ H(t_0, x_0) \geq \zeta \mathbb{P}(\tau \geq \theta) + \mathbb{E}_{t_0, x_0} [H(\psi, X_\psi)] \]

By the first inequality in (39), \( H \geq G + \delta \) on \( D_h, \) so therefore

\[ H(t_0, x_0) \geq \zeta \mathbb{P}(\tau \geq \theta) + \mathbb{E}_{t_0, x_0} [(G(X_\tau) + \delta) \mathbf{1}_{\tau < \theta} + H(\theta, X_\theta) \mathbf{1}_{\tau \geq \theta}], \]

where we have replaced \( H(\tau, X_\tau) \) by its lower bound \( (G(X_\tau) + \delta) \) on \( \{\tau < \theta\}, \)

since in that event we are still in \( D_h. \) Finally, since \( \mathbb{E}_{t_0, x_0} [\mathbf{1}_{\tau < \theta}] = \mathbb{P}(\tau < \theta), \)

we have

\[ H(t_0, x_0) \geq \zeta \mathbb{P}(\tau \geq \theta) + \delta \mathbb{P}(\tau < \theta) + \mathbb{E}_{t_0, x_0} [(G(X_\tau) \mathbf{1}_{\tau < \theta} + H(\theta, X_\theta) \mathbf{1}_{\tau \geq \theta}] \]

By the arbitrariness of \( \tau \in T_{[t,T]}, \)

and the fact that the constants added to the expectation above are positive, we arrive at a contradiction to the DPP (36) and the proof is complete.

The approximating function \( \varphi_\epsilon \) in the second part of the above proof can be seen as an upper-semicontinuous approximation to the value function, and the viscosity solution approach makes heavy use of both upper and lower semicontinuous envelopes of the value function when it is not smooth enough to differentiate. We do not divert into these discussions here, and instead refer the interested reader to the excellent monographs Touzi (2013) as well as Pham (2010) and Fleming & Soner (2006).

It is worth briefly exploring the interpretation of the variational inequality appearing above, repeated here for convenience:

\[ \max \left\{ \partial_t H + \mathcal{L}_t H , \ G - H \right\} = 0 , \quad \text{on} \quad D. \]
This really represents two possibilities: either we have
\[(i) \quad \partial_t H + \mathcal{L}_t H = 0 \quad \text{and} \quad H < G,\]
or we have
\[(ii) \quad H = G \quad \text{and} \quad \partial_t H + \mathcal{L}_t H < 0.\]
The first of these possibilities corresponds to the value function \(H\) being lower than the reward \(G\), and in this region the value function satisfies a linear PDE. If we introduce the stochastic process \(h_t = H(t, X_t)\), then due to the linear PDE having zero right-hand side, \(h_t\) is a martingale. The region in which (i) holds is what we identified earlier as the continuation region \(S^c\), since it is suboptimal to stop there.
The second of these possibilities corresponds to the value function \(H\) equalling the reward \(G\), and hence occurs in the stopping region \(S\). Moreover, the linear operator \(\partial_t + \mathcal{L}_t\) renders the value function negative. In this region, if we did not pin the value function to the reward, we see that the process \(h_t\) would be a supermartingale; however, pinning it to the reward constrains the process to become a martingale. Hence, we see that the stochastic process \(h_t\) corresponding to the flow of the value function is in fact a martingale on the entire \(\mathcal{D}\).

5 Combined Stopping and Control

There are many instances in which an agent wishes to solve a combined optimal stopping and control problem – i.e. she wishes to solve simultaneously for the optimal timing and optimal strategy in order to maximise a reward. Such problems inherit features from both problem types and we will see that the resulting DPEs are in effect a combination of the HJB equation and variational inequality.

In this case, we adopt the modelling assumptions and notation from subsection 3.3 but repeat it here. First, let \(\mathbf{N}^u = (\mathbf{N}^u_t)_{0 \leq t \leq T}\) denote a collection of counting processes (of dim \(p\)) with controlled intensities \(\mathbf{\lambda}^u = (\mathbf{\lambda}^u_t)_{0 \leq t \leq T}\), and let \(\mathbf{u} = (u_t)_{0 \leq t \leq T}\) denote the control processes (of dimension \(m\)). Furthermore, let \(\mathbf{W} = (W_t)_{0 \leq t \leq T}\) denote a collection of independent Brownian motions (of dim \(m\)). Next, let \(\mathbf{X}^u =\)
\((X_t^u)_{0 \leq t \leq T}\) denote the controlled processes (of dim \(m\)) which will appear directly in the agent’s performance criteria. If one (or more) of the counting processes \(N_i^t\) should appear in the performance criteria or in the dependence in \(\mu_i^u\), \(\sigma_i^u\) and/or \(\gamma_i^u\), shown below, then take \(X_j^t = N_i^t\) for some \(j\). i.e. include it through one of the components in \(X\). We next assume that the controlled \(X^u\) processes satisfy the SDEs
\[
dX^u_t = \mu^u_t dt + \sigma^u_t dW_t + \gamma^u_t dN^u_t. \tag{42a}
\]
With a slight abuse of notation we write the controlled drift, volatility and jump size as
\[
\mu^u_t := \mu(t, X^u_t, u_t), \quad (m \times 1 \text{ vector}), \tag{42b}
\]
\[
\sigma^u_t := \sigma(t, X^u_t, u_t), \quad (m \times m \text{ matrix}), \tag{42c}
\]
\[
\gamma^u_t := \gamma(t, X^u_t, u_t), \quad (m \times p \text{ matrix}). \tag{42d}
\]
Furthermore, we assume that the controlled intensity takes the form
\[
\lambda^u_t := \lambda(t, X^u_t, u_t). \tag{42e}
\]

The modelling approach above implies that the agent can control, in general, the drift, volatility, jump size and jump arrivals. Precisely how much control she has depends on the specific form of the various functions appearing in (42b), (42c), (42d) and (42e).

Next, the agent’s performance criteria for a given admissible control \(u\) and admissible stopping time \(\tau\) is given by
\[
H^{u,\tau}(t, x) = \mathbb{E}_{t,x}[G(X^u_\tau)], \tag{43a}
\]
and her value function is
\[
H(t, x) = \sup_{\tau \in \mathcal{T}_t} \sup_{u \in \mathcal{A}_{[t,\tau]}} H^{u,\tau}(t, x). \tag{43b}
\]
In this manner, the agent selects a stopping rule, seeks the best strategy, and then selects the stopping rule which provides the best overall performance. As such, she is aiming to decide whether to stop ‘now’ and receive the reward at the point in state
space to which her strategy took her, or wait and continue to control the process in hopes of receiving a better reward later on.

A DPP still applies in the current setting; however, now we must take care of both optimal stopping and control. Following similar arguments as in the previous sections we arrive at the following DPP.

**Theorem 7. Dynamic Programming Principle for Optimal Stopping and Control.**

The value function \( H(t, x) \) satisfies the DPP

\[
H(t, x) = \sup_{\tau \in T_{[t,T]}} \sup_{u \in A_{[t,T]}} \mathbb{E}_{t,x} \left[ G(X^u_\tau) 1_{\tau < \theta} + H(\theta, X^u_\theta) 1_{\tau \geq \theta} \right],
\]

for all \((t, x) \in [0,T] \times \mathbb{R}^m\) and all stopping times \(\theta \leq T\).

The intuition here is similar to the previous DPPs. If stopping has not yet occurred \((\tau \geq \theta)\), then the agent receives the value function at that point in time, at that point in state space, where the optimal control has driven her to. If stopping has already occurred \((\tau < \theta)\), then the agent has already received the reward, and her value equals the reward at the time it was paid, at the point in state space where the optimal control drove her to.

Next, we state the DPE arising in this class of problems.

**Theorem 8. Dynamic Programming Equation for Stopping and Control Problems.** Assume that the value function \( H(t, x) \) is once differentiable in \( t \) and all second order derivatives in \( x \) exist, i.e. \( H \in C^{1,2}([0,T], \mathbb{R}^m) \), and that \( G : \mathbb{R}^m \rightarrow \mathbb{R} \) is continuous. Then \( H \) solves the quasi-variational inequality (QVI),

\[
\max \left\{ \partial_t H + \sup_{u \in A_t} L^u_t H , \ G - H \right\} = 0, \quad \text{on} \quad \mathcal{D},
\]

where \( \mathcal{D} = [0,T] \times \mathbb{R}^m \).

This is similar in spirit to the variational inequality (37) for the optimal stopping problem, but now contains an optimisation over the control process as well. The
optimisation results in the equation becoming non-linear, as in the HJB equation \([34]\), and hence the above equation is referred to as a quasi-variational inequality rather than a variational inequality.

We once again have the notion of a stopping region \(S\) and continuation region \(S^c\), where

\[
S = \{(t, x) \in [0, T] \times \mathbb{R}^m : H(t, x) = G(x)\}.
\]

The only difference is that in the continuation region \(S^c\), the value function should satisfy an HJB equation

\[
\partial_t H + \sup_{u \in A_t} \mathcal{L}_t^u H = 0, \quad \text{on} \quad S^c,
\]

rather than a linear PDE as in the stopping problem.

References


Pham, Huyên (2010), *Continuous-time Stochastic Control and Optimization with Financial Applications*, Springer.
