PDE for Lookback Option

Math 622

February 27, 2014

Reading material: Ocone's Lecture note 6

1 Preliminary discussion

Let $S(t)$ satisfies

$$\begin{aligned} dS_t &= rS_t dt + \sigma(t, S_t) S_t dW_t \\ S(0) &= x > 0. \end{aligned}$$

Note that σ is a function of t, St here, in stead of being a constant. We call this the *local volatility* model and make the assumption that $\sigma(t, x) > 0$ for all t, x.
Consider the Lookback Option:

$$V_T = \max_{[0,T]} S_t - S_T.$$

Then by risk neutral pricing

$$V_t = E\Big(e^{-r(T-t)} \max_{u \in [0,T]} S_u \Big| \mathcal{F}(t)\Big) - S_t.$$

Similar to what we did in Lecture 5a notes, define

$$Y(t) = \max_{[0,t]} S(u),$$

then for $s > t$

$$Y(s) = \max\{Y(t), \max_{u \in [t,s]} S(u)\}$$

In Homework 5, we have discussed that when σ is constant, then $\{Y(t), S(t)\}$ is a Markov process. The argument is by Independence Lemma. For the current local volatility model, the Independence Lemma no longer applies, since we cannot conclude that $\int_t^T \sigma(u, S(u))dWu$ is independent of $\mathcal{F}(t)$. However, it is still true that $\{Y(t), S(t)\}$ is Markov. Indeed, we have the following principle:

Principle: If $S(t), t \geq 0$ is Markov with respect to $\mathcal{F}(t)$ and $Y(t) = \max_{u\in[0,t]} S(u)$ then $\{Y(t), S(t)\}$ is also Markov with respect to $\mathcal{F}(t)$.

We call it a principle instead of a theorem because we will not give it a proof due to technical details. Thus we also have

$$\begin{aligned}
V(t) &= \mathbb{E}\left[e^{-r(T-t)}V_T \middle| \mathcal{F}(t)\right] \\
&= \mathbb{E}\left[e^{-r(T-t)} \max_{[0,T]}\{S_t\} \middle| \mathcal{F}(t)\right] - S(t) \\
&= \mathbb{E}\left[e^{-r(T-t)} \max\{Y(t), \max_{u\in[t,T]} S_u\} \middle| \mathcal{F}(t)\right] - S(t) \\
&= v(t, S(t), Y(t)),
\end{aligned}$$

where

$$v(t, x, y) = \mathbb{E}\left[e^{-r(T-t)} \max\{Y(t), \max_{u\in[t,T]} S_u\} \middle| S(t) = x, Y(t) = y\right].$$

Remark 1.1. *Note that here $V_T = G(Y_T, S_T)$ where $G(x, y) = y - x$. For this case, we call the option* **floating strike lookback option**. *Clearly one can consider other types of function G as well. The only difference this would affect on the PDE is the boundary conditions. See Section* (6) *for more details.*

Now assuming that v is $C^{1,2,2}$, that is once continuously differetiable in t and twice continuously differentiable in x, y, we would like to derive a PDE that v satisfies. But note the following difference in our current case: $Y(t)$ *is not a $C^{1,2}$ function of* $S(t)$ so we cannot write down its dynamics using Ito's formula. In other words, we do not know what $dY(t)$ is explicitly.
However, observe that for $s < t$

$$Y(s) = \max_{u\in[0,s]} S(u) \leq \max_{u\in[0,t]} S(u) = Y(t),$$

simply because the max over a bigger set is not smaller than the max over a (smaller) set contained in it. Therefore $Y(t)$ is *an increasing (meaning it is non-decreasing)* function.

From the discussion of the Lebesgue-Stieltjes integral of Chapter 11, we have learned how to integrate with respect to functions of bounded variation. Recall that increasing function is of bounded variation. Therefore, it makes sense to talk about $dY(t)$ (in the Lebesgue-Stieltjes integral sense, that is).

However, we did not discuss the Ito's formula for $v(t, S(t), Y(t))$ where $S(t)$ is an Ito process and $Y(t)$ is an increasing process. But suppose we just formally carry out the usual Ito's rule to $e^{-rt}v(t, S(t), Y(t))$, what we should get is

$$\begin{aligned} de^{-rt}v(t,S(t),Y(t)) &= e^{-rt}\Big\{[-rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2]\Big|_{(x,y)=(S_t,Y_t)}\Big\}dt \\ &\quad + e^{-rt}\frac{\partial}{\partial x}v(t,S_t,Y_t)\sigma(t,S_t)S_t dW_t + e^{-rt}\frac{\partial}{\partial y}v(t,S_t,Y_t)dY_t \\ &\quad + e^{-rt}\frac{\partial^2}{\partial xy}v(t,S_t,Y_t)d\langle S,Y\rangle(t) + e^{-rt}\frac{\partial^2}{\partial y^2}v(t,S_t,Y_t)d\langle Y\rangle(t). \end{aligned}$$

Remark 1.2. *We will discuss what $d\langle S,Y\rangle(t)$ and $d\langle Y\rangle(t)$ means in the following section. For now, you can formally replace $d\langle S,Y\rangle(t)$ with $dS(t)dY(t)$ and $d\langle Y\rangle(t)$ with $[dY(t)]^2$ to get an intuition.*

Since

$$e^{-rt}V_t = e^{-rt}v(t, S(t), Y(t)),$$

$e^{-rt}v(t, S(t), Y(t))$ is a martingale. On the RHS of the above equation, the only martingale term we have is

$$\frac{\partial}{\partial x}v(t, S_t)\sigma(t, S_t)S_t dW_t.$$

The principle of deriving our PDE is that any other terms that do not contribute to the martingale property of the RHS should be set to 0. But before we can do that, we need to understand the following:

(i) Is the Ito's rule that we just formally applied correct? (If it is not correct there is no point in discussing the items below).

(ii) What are $d\langle S,Y\rangle(t)$ and $d\langle Y\rangle(t)$?

(iii) How to understand $dY(t)$?

We will address these questions in the folowing order (ii), (i) and (iii) and then derive the PDE for $v(t, x, y)$ after that.

2 The quadratic variation and covariation

Fix $T > 0$. Let $X(t), Y(t)$ be functions defined on $[0, T]$. Recall the following definitions:

Definition 2.1. *The total variation of Y on $[0, T]$, denoted as $TV_Y(T)$ is defined as the smallest (finite) number such that for all partitions*
$0 = t_0 < t_1 < t_2 < ... < t_n = T$

$$\sum_{i=0}^{n-1} |Y(t_{i+1}) - Y(t_i)| \leq TV_Y(T).$$

If there is no such number, we define $TV_Y(T) = \infty$.
We also say Y is a function of bounded variation (on $[0, T]$) if $TV_f(T) < \infty$.

Definition 2.2. *The quadratic variation of Y on $[0, t]$, if it exists is defined as*

$$\langle Y \rangle(t) = \lim_{n\to\infty} \sum_{i=0}^{n-1} |Y(t_{i+1}^n) - Y(t_i^n)|^2,$$

where the limit is taken in probability, and for each fixed n,
$0 = t_0^n < t_1^n < t_2^n < ... < t_n^n = t$ *is a partition of $[0, t]$ such that its mesh size:*
$\max_i |t_{i+1}^n - t_i^n|$ *goes to 0 as $n \to \infty$.*

Definition 2.3. *The covariation between X and Y on $[0, t]$, if it exists is defined as*

$$\langle X, Y \rangle(t) = \lim_{n\to\infty} \sum_{i=0}^{n-1} \big(X(t_{i+1}^n) - X(t_i^n)\big)\big(Y(t_{i+1}^n) - Y(t_i^n)\big),$$

where the limit is taken in probability, and for each fixed n,
$0 = t_0^n < t_1^n < t_2^n < ... < t_n^n = t$ *is a partition of $[0, t]$ such that its mesh size:*
$\max_i |t_{i+1}^n - t_i^n|$ *goes to 0 as $n \to \infty$.*

Note: Some authors (including Shreve in our textbook, see Exercise 7.4) called the covariation the cross quadratic variation.
We will also state the following facts about quadratic variation and covariation. The proof is more or less contained in the extra credit problem in Homework 1.

(i) If Y is increasing then Y is of bounded variation.

(ii) If Y is continuous and of bounded variation, then $\langle Y \rangle(t) = 0$.

(iii) If Y is of bounded variation and X is continuous, then $\langle X, Y\rangle(t) = 0$.

(iv) The quadratic variation $\langle X\rangle(t)$ and covariation $\langle X, Y\rangle(t)$ of any two processes X, Y, if exist, are of bounded variation on $[0, T]$. Therefore, it makes sense to talk about $d\langle X\rangle(t)$ and $d\langle X, Y\rangle(t)$.

Applying these facts to our situation, we see that indeed

$$\langle Y\rangle(t) = 0$$
$$\langle S, Y\rangle(t) = 0$$

So question (ii) of Section 1 is answered.

3 An extension of Ito's formula

We now give answer to question (i) of Section 1. Let $W(t)$ be a BM and $\mathcal{F}(t)$ a filtration for $W(t)$.

$$X^i(t) = X^i(0) + \int_0^t \alpha^i(s)ds + \int_0^t \sigma^i(s)dW(s) + A^i(t), i = 1, 2$$

where α^i, σ^i, A^i are *stochastic processes* adapted to $\mathcal{F}(t)$, σ^i are chosen so that the stochastic integral is well-defined, and $A^i(t)$ are *continuous functions of bounded variations.* Let $f(t, x_1)$ be a $C^{1,2}$ function. Then

$$\begin{aligned} df(t, X^1(t)) &= \Big[\frac{\partial}{\partial t}f + \left(\frac{\partial}{\partial x_1}f\right)\alpha^1_t + \frac{1}{2}\left(\frac{\partial^2}{\partial (x_1)^2}f\right)(\sigma^1_t)^2\Big]dt \\ &\quad + \left(\frac{\partial}{\partial x_1}f\right)\sigma^1(t)dW(t) + \left(\frac{\partial}{\partial x_1}f\right)dA^1(t). \end{aligned}$$

Let $f(t, x_1, x_2)$ be a $C^{1,2,2}$ function. Then

$$\begin{aligned} df(t, X^1(t), X^2(t)) &= \Big[\frac{\partial}{\partial t}f + \sum_{i=1}^{2}\Big\{\left(\frac{\partial}{\partial x_i}f\right)\alpha^i_t + \frac{1}{2}\left(\frac{\partial^2}{\partial (x_i)^2}f\right)(\sigma^i_t)^2 + \left(\frac{\partial^2}{\partial x_1 x_2}f\right)\sigma^1_t\sigma^2_t\Big\}\Big]dt \\ &\quad + \Big\{\sum_{i=1}^{2}\left(\frac{\partial}{\partial x_i}f\right)\sigma^i(t)\Big\}dW(t) + \sum_{i=1}^{2}\left(\frac{\partial}{\partial x_i}f\right)dA^i(t). \end{aligned}$$

where by f we understand as $f(t, X^1(t), X^2(t))$.

Remark 3.1. *We do not give a proof of this extension. But you can see the formula is just the application of Ito's formula as we used to do, combined with the facts about quadratic variation and covariation of bounded variation process that we discussed in Section 2.*

Remark 3.2. *It is true that* $\int_0^t \alpha^i(s)ds$ *is also a continuous function of bounded variation. So what is the difference between* $\int_0^t \alpha^i(s)ds$ *and* $A^i(t)$*? Can we combine them into just 1 term? The answer is no, because these two terms have very different property. We say the term* $\int_0^t \alpha^i(s)ds$ *is* **absolutely continuous** *(with respect to the Lebesgue measure dt). Basically this means it can be represented as an integral with respect to dt (which it is already in that form). The term* $A^i(t)$ *in this formula is meant to be* **singularly continuous** *(with respect to the Lebesgue measure dt). For our purpose, what it means is that even though* $A^i(t)$ *is continuous we cannot represent* $A^i(t)$ *as an integral with respect to dt. Therefore, the two terms should be kept separate.*

Remark 3.3. *You should compare and contrast these Ito formulas with the ones we obtained in Chapter 11. There, the* $A^i(t)$ *are the pure jump processes. So while here we have the term* $\left(\frac{\partial}{\partial x_i} f\right) dA^i(t)$*; in Chapter 11 the corresponding term is* $\sum_{0<s\leq t} f(X^i(s)) - f(X^i(s-))$*. We mentioned that it's not always possible to get the differential form in the Ito's formula in Chapter 11. Here note that it is always in differential form.*

3.1 An intuition on the difference between the 2 Ito's formulae

For simplicity, let's just consider 2 cases:
(i) $X^1(t) = A(t)$
A is continuous and of bounded variation.
(ii) $X^2(t) = J(t)$,
J is pure jump.
Let f be C^1 function. Then (from Ito's formula) we have

$$f(X_t^1) = f(X_s^1) + \int_s^t f'(X_u^1)dX^1(u);$$
$$f(X_t^2) = f(X_s^2) + \sum_{s<u\leq t} f(X^2(u)) - f(X^2(u-)).$$

Note the derivative in the 1st case and the original function in the 2nd case. Why is this? We always have the following identity (assuming $X_t^1 - X_s^1 \neq 0$)

$$f(X_t^1) - f(X_s^1) = \frac{f(X_t^1) - f(X_s^1)}{X_t^1 - X_s^1}(X_t^1 - X_s^1).$$

And so intuitively, we have for s very close to t,

$$\begin{aligned} f(X_t^1) - f(X_s^1) &\approx f'(X^1(s))(X_t^1 - X_s^1) \\ &\approx f'(X^1(s))dX^1(s). \end{aligned}$$

This is correct, **if** $X_t^1 \to X_s^1$ **as** $t \to s$, which requires the continuity of $X^1(t)$ (so that the difference quotient approximates the derivative of f). But in the case of X^2, if X^2 is not continuous at s (it has a jump at s), then we cannot say the difference quotient approximates the derivative of f at $X^2(s)$ in any sense. Therefore, we cannot write it in the differential form, and can only write it as the form we always used in Chapter 11.

4 The integral dYt

4.1 Some preliminary discussion

Recall that we define $Y_t := \max_{u \in [0,t]} S_u$ to be the running max of S_t. We have observed that Y_t is non-decreasing. But can we say more? For example, is there any interval where Y is strictly increasing, not just non-decreasing? To anwer that, we make the following observation (recall that by definition, $S(t) \leq Y(t)$):

(i) Suppose that $S(t) < Y(t)$ for all $t \in [a, b]$. Then $Y(t)$ is constant on $[a, b]$ (Note the strict inequality).

Reason:

$$Y(b) = \max(Y(a), \max_{[a,b]} S(t)).$$

Suppose by contradiction that $Y(b) > Y(a)$. Then it must follow that

$$Y(b) = \max_{[a,b]} S(t).$$

However, this contradicts with our assumption that $S(t) < Y(t)$ for all $t \in [a, b]$ since it would imply $S(t) < Y(t) \le Y(b)$ for all $t \in [a, b]$ which lead to $\max_{[a,b]} S(t) < Y(b)$, not equality.

(ii) Suppose $Y(t)$ is strictly increasing on $[a, b]$ (that is for all $u < v$ in $[a, b], Y(u) < Y(v)$) then $Y(t) = S(t)$ for all $t \in [a, b]$ and thus it also follows that $S(t)$ is also strictly increasing on $[a, b]$.

Reason:
Suppose there is u in $[a, b]$ such that $S(u) < Y(u)$. Then because $S(t)$ is continuous, we can find $\varepsilon > 0$ small enough such that $S(t) \le Y(u)$ on $[u, u + \varepsilon]$. But then it follows that for $u < v$ in $[a, b]$ we have

$$Y(u + \varepsilon) = \max(Y(u), \max_{t\in[u,u+\varepsilon]} S(t)) = Y(u),$$

contradicting our assumption that $Y(t)$ is strictly increasing. So it must be that $S(t) = Y(t)$ for all $t \in [a, b]$. Now since they're equal, the 2nd conclusion obviously follows.
Thus from observation (ii), we see that if we can find any interval $[a, b]$ where $Y(t)$ is strictly increasing (so that $dY(t) > 0$ on $[a, b]$) then $S(t)$ also have to be strictly increasing there as well. There is a fact of real analysis which says a function that is strictly increasing must be differentiable almost every where and its derivative positive where it is differentiable. However, we also know that BM is nowhere differentiable (with probability 1). $S(t)$ being an exponential function of BM also must be nowhere differentiable. Thus it cannot be the case that we can find an interval where $Y(t)$ is strictly increasing.
There are examples in real analysis of functions on $[0, 1]$ which is continuous, increasing, whose derivative equals to 0 almost everywhere, yet $f(0) = 0$ and $f(1) = 1$. The Cantor function is such an example. In this case we can also not find any interval where the Cantor function is strictly increasing. For this reason, we say that the running max $Y(t)$ exhibits "Cantor-like" property.

4.2 Support of $Y(t)$

Reading material: Ocone's lecture note 6 Section 2

4.2.1 Preliminary discussion

From observation (ii) of the above subsection, we see that the set of points t where $dY(t) > 0$ must be small (in the sense that it cannot be an interval). So the set $\{dY(t) = 0\}$ should be large in some sense. Observe, however, that it cannot be the case that $dY(t) = 0$ for all $t \in [0, T]$ either. Because it would imply that

$$Y(T) = Y(0) + \int_0^T dY(s) = Y(0),$$

which would force $S(t)$ to be constant on $[0, T]$ which is not the case. Indeed, we can repeat the argument to get that on any subinterval $[u, v]$ of $[0, T]$ the probability that $Y(v) > Y(u)$ is positive (otherwise, with probability 1, $Y(u) = Y(v)$ which would force $S(u) = S(v)$ with probability 1, which is again not the case).
So there must be a set of points on $[0, T]$ where $dY(t) > 0$ with probability 1. From observation (i) and (ii) again, this set must be a subset of the set

$$C = \{t \in [0, T] : Y(t) = S(t)\}.$$

(It could be equal to C, whether or not this is the case is not important to us).
Note that C is a random set: for different event ω, $C(\omega)$ may be different. *We call C the support of Y.* See prof. Ocone's notes for more discussion on support of a continuous increasing function.

Remark 4.1. *Since $Y(t)$ is continuous, it follows that for any fixed $t_0 \in [0, T]$, $dY(t_0) = 0$. So in the above discussion, when we say $dY(t) > 0$ on some set E, what we really mean is*

$$\int_0^T \mathbf{1}_E(s) dY(s) > 0.$$

and when we say $dY(t) = 0$ on some set E, what we really mean is

$$\int_0^T \mathbf{1}_E(s) dY(s) = 0.$$

Remark 4.2. *The one important property you should remember about the support C is this: if E is a measurable set contained in C^c, then*

$$\int_0^T \mathbf{1}_E(s) dY(s) = 0.$$

The intuition if that outside the support C (on the set $\{t \in [0, T] : S(t) < Y(t)\}$), $Y(t)$ is constant so integrate against $dY(t)$ outside the support should give you 0.

Remark 4.3. *The one important topological property of C is that it is closed, yet it is not the whole interval $[0, T]$ (see the below section on Main results). This has the important consequence that its nonempty complement is open, in the sense that for any point in C^c, we can find an interval around it that is also in C^c. In other words, for any $t_0 \in C^c$, we can find $\varepsilon > 0$ so that*

$$\int_{t_0-\varepsilon}^{t_0+\varepsilon} g(s)dY(s) = 0,$$

for any measurable function g. This is important because we would like to differentiate the function

$$G(t) := \int_0^t g(s)dY(s),$$

at the point t_0. What the above equation says is $G(t)$ is constant on $[t_0 - \varepsilon, t_0 + \varepsilon]$, which allows us to take the derivative at t_0. If we cannot find any $\varepsilon > 0$ such that on $[t_0 - \varepsilon, t_0 + \varepsilon]$, the function $G(t)$ is well-behaved, (remember the integral is against $dY(t)$, so automatic regularity of $G(t)$ is not guaranteed) then taking derivative at t_0 is not justified.

4.2.2 Main results

We already discussed the fact that C cannot contain any interval. But even more is true by the following theorem:

Theorem 4.4. *Let $S(t)$ satisfies*

$$\begin{aligned} dS_t &= \alpha(t)S_t dt + \sigma(t)S_t dW_t \\ S(0) &= x > 0, \end{aligned}$$

where α, σ can be random processes. **We require** *$\sigma(t) > 0$. Define the set C as the support of Y as above. Then with probability 1,*

$$\int_0^\infty \mathbf{1}_{C(\omega)}(s)ds = 0.$$

In words, the Lesbegue measure of the set $C(\omega)$ is 0 with probability 1. The importance of this theorem to us (in deriving the PDE) is in the following corollary.

Corollary 4.5. *Suppose $S(t)$ satisfies the hypothesis of theorem* (4.4)*. Let $\alpha(s)$, $g(s)$ be continuous process. Then*

$$\int_0^t \alpha(s)ds + \int_0^t g(s)dY(s) = 0 \forall t > 0$$

if and only if $\alpha(t) = 0$ and $g(t)\mathbf{1}_{\{Y(t)=S(t)\}} = 0$ for all $t > 0$.

Proof. Suppose $\alpha(t) = 0$ and $g(t)\mathbf{1}_{\{Y(t)=S(t)\}} = 0$. Then since on C^c, $dY(t) = 0$, we have

$$\int_0^t \alpha(s)ds + \int_0^t g(s)dY(s) = \int_0^t g(s)\mathbf{1}_C dY(s) + \int_0^t g(s)\mathbf{1}_{C^c} dY(s) = 0.$$

We now show the forward direction. Let $t_0 \in C^c$. Then this implies that $S(t_0) < Y(t_0)$ and since $S(t)$ is continuous, there exists an interval $[a, b]$ around t_0 so that $S(t) \leq Y(t)$ on $[a, b]$. Thus $dY = 0$ on $[a, b]$ as well and it follows that $\int_0^t g(s)dY(s)$ is a constant on $[a, b]$ and hence differentiable at t_0 with derivative being 0. On the other hand $\frac{\partial}{\partial t} \int_0^t a(s)ds = a(t)$ for all t. Thus by taking derivative of the equation

$$\int_0^t \alpha(s)ds + \int_0^t g(s)dY(s) = 0$$

at t_0, we conclude that $a(t_0) = 0$ for all $t_0 \in C^c$.

Now let $t_0 \in C$. Since C contains no interval, for any n, there exists $t_n \in C^c$ in $[t_0 - 1/n, t_0 + 1/n]$. But then $t_n \to t$ and hence $a(t) = 0$ by continuity of a. Thus $a(t) = 0$ for all t. It follows that

$$\int_0^t g(u)\mathbf{1}_{\{Y(u)=S(u)\}} dY(u) = 0, \forall t,$$

or

$$\int_s^t g(u)\mathbf{1}_{\{Y(u)=S(u)\}} dY(u) = 0, \forall s \leq t.$$

By choosing s close to t, we can assume that g is either non-negative or non-positive on $[s, t]$ by its continuity. Then this equality says the integral of g, over the support of Y intersect with $[s, t]$ is 0. The support of Y is the closure of the set of points where $dY(t) > 0$. Thus it must follow that $g(u) = 0$ on the support of Y (by choosing s close to t we do not have to worry about g changing sign on $[s, t]$). Outside the support of Y, $\mathbf{1}_{\{Y(u)=S(u)\}} = 0$. Thus $g(u)\mathbf{1}_{\{Y(u)=S(u)\}} = 0$ for all u.

5 The dynamics of Y

From section 3, we gave the Ito's formula extension to stochastic process with singular continuous component $A^i(t)$. We want to apply the Ito's formula to $v(t, X_t, Y_t)$ where Y_t is *a singularly continuous process.* But so far we still have not argued about this fact. In other words, is there a possibility that Y can still be represented as an Ito's process? That is, does there exist $\alpha(t), \beta(t)$ such that

$$dY(t) = \alpha_t dt + \beta_t dW(t). \tag{1}$$

This indeed cannot happen, as the following Lemma shows

Lemma 5.1. *Let Y_t be the running maximum of S_t where S_t satisfies the hypothesis of Theorem* (4.4)*. Then Y_t* **cannot be representable** *in the form of equation* (1)*. In other words, Y_t is a singularly continuous process.*

Proof. The proof of the Lemma relies on Theorem (4.4) and the following result:

For any $t > 0, \mathbb{P}(Y_t > Y_0) = 1$.

We have discussed before that for $t > 0, \mathbb{P}(Y_t > Y_0)$ must be > 0. Turns out it is true that this probability is actually 1. We now prove the Lemma by contradiction. Suppose there exist $\alpha(t), \beta(t)$ such that

$$dY(t) = \alpha_t dt + \beta_t dW(t).$$

Then since $Y(t)$ is increasing, by the result in Section 2, $\langle Y \rangle_t = \int_0^t \beta_s^2 ds = 0$. But this means $\beta_s = 0, \forall s$. Then $Y(t) = Y_0 + \int_0^t \alpha_s ds$. But that means Y is differentiable and $Y'(t) = \alpha_t$ for all t. From Theorem (4.4), we learned that $Y'(t) = 0$ on the set $\{S(t) < Y(t)\}$. But that means $\alpha_t = 0$ on $\{S(t) < Y(t)\}$. Since the set $\{S(t) = Y(t)\}$ has Lebesgue measure 0, it follows that

$$Y(t) = Y_0 + \int_0^t \alpha_s ds = Y_0,$$

which contradicts the fact we just state and the proof is complete.

6 Derivation of the PDE

Putting all the above information together, we have

$$\begin{aligned} e^{-rt}v(t,S(t),Y(t)) &= v(0,S(0),Y(0)) + \int_0^t e^{-ru}[\mathcal{L}v(u,S_u,Y_u) - rv(u,S_u,Y_u)]du \\ &\quad + \int_0^t e^{-ru}\frac{\partial}{\partial y}v(u,S_u,Y_u)dY_u + \int_0^t e^{-ru}\frac{\partial}{\partial x}v(u,S_u,Y_u)\sigma(u,S_u)S_u dW_u. \end{aligned}$$

where

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

Since the LHS is a martingale, we set

$$\int_0^t e^{-ru}[\mathcal{L}v(u,S_u,Y_u) - rv(u,S_u,Y_u)]du + \int_0^t e^{-ru}\frac{\partial}{\partial y}v(u,S_u,Y_u)dY_u = 0, \forall t.$$

Apply Corollary (4.5) we conclude that

$$\begin{aligned} -rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2 &= 0; t < T, 0 < x \le y < \infty \\ \frac{\partial}{\partial y}v(t,y,y) &= 0; t \le T, y > 0 \qquad (2) \\ v(T,x,y) &= y - x; \qquad (3) \\ v(t,0,y) &= e^{-r(T-t)}y. \qquad (4) \end{aligned}$$

Condition (2),(3),(4) are boundary conditions. Condition (2) is called a Neumann condition, since it imposes the value of a derivative of v on the boundary .
Condition (4) comes from the fact that once $S(t)$ hits 0 at time t it stays there so the running max is a constant on $[t,T]$: $Y(u) = Y(t), u \ge t$. Thus we get

$$\begin{aligned} v(t,0,Y(t)) &= E(e^{-r(T-t)}Y(T)|\mathcal{F}(t)) \\ &= E(e^{-r(T-t)}Y(t)|\mathcal{F}(t)) = e^{-r(T-t)}Y(t). \end{aligned}$$

More generally, suppose we consider the generalized lookback option:

$$V(T) = G(S(T),Y(T)),$$

then the same argument shows that $V(t) = v(t,S(t),Y(t))$ where $v(t,x,y)$ satisfies the PDE

$$
\begin{aligned}
-rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2 &= 0; t < T, 0 < x \leq y < \infty \\
\frac{\partial}{\partial y}v(t,y,y) &= 0; t \leq T, y > 0 \qquad (5) \\
v(T,x,y) &= G(x,y); 0 \leq x \leq y \qquad (6) \\
v(t,0,y) &= e^{-r(T-t)}G(0,y). \qquad (7)
\end{aligned}
$$