PDE for Lookback Option

Math 622

February 25, 2015

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 5 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

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

where

$$v(t, x, y) = \mathbb{E}\Big[e^{-r(T-t)} \max\{Y_t, \max_{u \in [t,T]} S_u\} \Big| S_t = x, Y_t = y\Big] - x.$$

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

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

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 A summary of the main results

In what follows, we will discuss many technical details about the behavior of Y_t, the running max of S_t, and the extension of Ito's formula to functions depending on Y_t. Thus, it is easy to lose track of the main points of the discussion. To help the readers to follow, we list these points here.

a. Y_t is an increasing (non-decreasing) process in t.

b. Y_t has no dynamics. That is we cannot represent Y_t in the form

$$Y_t = Y_0 + \int_0^t \alpha_s ds + \int_0^t \sigma_s dW_s.$$

c. $\langle Y, S\rangle_t = \langle Y\rangle_t = 0$.

d. If $f(y)$ is a differentiable function then

$$f(Y_t) - f(Y_0) = \int_0^t f'(Y_s)dY_s.$$

e. Y_t can only increase on the set $\{t : Y_t = S_t\}$, which is closed and contains no interval.

f. If the term $\sigma(t)$ in the dynamics of S_t is positive for all t 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 $t > 0$.

(This allows us to set the terms other than the dW_t term equal to 0 when we derive the PDE).

3 The quadratic variation and covariation

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

Definition 3.1. *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 $TV_Y(T) < \infty$.

Definition 3.2. *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,
$0 = t_0^n < t_1^n < t_2^n < \ldots < t_n^n = t$ *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 3.3. *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,
$0 = t_0^n < t_1^n < t_2^n < \ldots < t_n^n = t$ *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.

4 An extension of Ito's formula

We now give answer to question (i) of Section 1. Let $W(t)$ be a Brownian Motion 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

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

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

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

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

Remark 4.1. *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 4.2. *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 4.3. *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 $\sum_{0<s\le t} f(X^i(s)) - f(X^i(s-))$. 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.*

4.1 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

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

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.

5 The integral dY_t

5.1 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 observations (recall that by definition, $S_t \leq Y_t$):

(i) Suppose that $S_t < Y_t$ for some t then there must exist an interval $[a, b]$ around t (that is, $t \in (a, b)$)so that Y is constant on $[a, b]$.

Reason: Since S_t is continuous, if $S_t < Y_t$ there must exist an interval $[a, b]$ around t so that $S_u < Y_t$ for all $u \in [a, b]$. Since $Y_t = Y_a \vee \max_{u\in[a,t]} S_u$, it is clear that $Y_a = Y_t$. Similar reasoning gives $Y_t = Y_b$. Because Y is increasing, it follows that Y is constant on $[a, b]$.

<u>Remark:</u> Observation (i) tells us that Y can only increase when $Y_t = S_t$. The next observation tells us how often this happens.

(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 $t \in [a, b]$. 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 by oservation (i) we can find an interval around u on which Y is constant, contradicting the assumption that Y is strictly increasing. Thus $S_t = Y_t$ for all t in $[a, b]$. The second conclusion is obvious.

<u>Remark:</u> Observation (ii) tells us that on any interval where Y_t is strictly increasing, S_t also has to be strictly increasing. Intuitively you can see that this will not happen on any interval $[a, b]$ if the volatility term σ_t of S_t is positive. The reason is if S_t is increasing, then its quadratic variation on $[a, b]$ must be 0 as we mentioned before. On the other hand, this should be $\int_a^b \sigma_t^2 dt$, which is a contradiction if $\sigma_t > 0$.

5.2 The set $C := \{t : Y_t = S_t\}$

From the above discussion, we see that Y_t can only increase on the set $C := \{t : Y_t = S_t\}$. We emphasize that C is a *random set*, that is we should write

$$C(\omega) := \{t : Y_t(\omega) = S_t(\omega)\}.$$

But our convention is that we just remember the fact that C is random and omit the writing of ω.

We also observe that C is a closed set (in the sense that its complement is an open set) from an elementary topological result that says the inverse image of a closed set is closed:

$$C = (Y_t - S_t)^{-1}(\{0\}).$$

From observation (ii) in the above section, C does not contain any interval. In this sense it is rather "small". Trivially then C cannot be the whole interval $[0, T]$. Equivalently, its nonempty complement C^c should be rather "large." But observe also that trivially, C^c cannot be too large (that is equal to $[0, T]$) because it would imply that

$$Y_T = Y_0,$$

which would force $S_t \leq S_0$ on $[0, T]$ and that is impossible.

Because C^c is open, for any point in C^c, we can find an open 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 because Y is constant on the interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ by the fact that $(t_0 - \varepsilon, t_0 + \varepsilon) \subseteq C^c$.

From this observation, coupled with the fact that any open set in $\mathbb{R}$ can be written as a countable union of open intervals, it follows that for any measurable function g

$$\int_0^T \mathbf{1}_{C^c}(s) g_s dY_s = 0.$$

And

$$\int_0^T g_s dY_s = \int_0^T \mathbf{1}_C(s) g_s dY_s.$$

Finally, we discuss the fact that C is the set of increase of Y_t in the following sense: for any $s < t$, let E be the event that there is a point $u \in (s, t)$ such that $S_u = Y_u$. Then

$$P(Y_s < Y_t | E) = 1.$$

Indeed, let $\tau := \inf\{u \geq s : S_u = Y_u\}$. Then τ is a stopping time. Conditioned on E, $\tau < t$ with probability 1. Then $S_u - S_\tau, u \geq \tau$ is an Ito process starting at τ. By exactly the same argument as we discussed in lecture 4, with probability 1, there are infinitely many points u_n close to τ so that $S_{u_n} > S_\tau$. That implies $Y_{u_n} > Y_\tau$. Thus $P(Y_t > Y_s | E) = 1$.

5.2.1 Main result

Theorem 5.1. *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 **with** *$\sigma(t) > 0$. Let $\beta(s)$, $g(s)$ be continuous process. Then*

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

if and only if $\beta(t) = 0$ and $g(t)\mathbf{1}_{\{Y_t = S_t\}} = 0$ for all $t > 0$.

Proof. Suppose $\beta(t) = 0$ and $g(t)\mathbf{1}_{\{Y_t = S_t\}} = 0$. We have

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

We now show the other direction. Let $t_0 \in C^c$. Since C^c is open, we can find an interval (a, b) around t_0 so that $(a, b) \subseteq C^c$. Thus

$$\int_a^t g(s) dY_s = 0, a \leq t \leq b.$$

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 $\frac{\partial}{\partial t} \int_0^t \beta(s) ds = \beta(t)$ for all t. Thus by taking derivative of both sides of the equation

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

at t_0, we conclude that $\beta(t_0) = 0$ for all $t_0 \in C^c$.
Now let $t_0 \in C$. Since $\sigma_t > 0$, C contains no interval from our above discussion. Thus for any n, the interval $[t_0 - 1/n, t_0 + 1/n]$ has non empty intersection with C^c. That is we can find $t_n \in C^c$ in any interval $[t_0 - 1/n, t_0 + 1/n]$, $t_n \neq t_0$.
Since $t_n \in C^c, \beta(t_n) = 0$ from the previous paragraph. Obseve also that $t_n \to t_0$ and hence $\beta(t_0) = 0$ by continuity of β. Thus $\beta(t) = 0$ for all t.
We then have

$$\int_0^t g(u)dY(u) = 0, \forall t,$$

which implies

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

Note that since $\int_0^t \mathbf{1}_{C^c}(s)g(s)dY_s = 0$, we cannot conclude anything about the values of g on C^c.
Now since C is the set of increase of Y_t as discussed above, it follows that $g(s)\mathbf{1}_C(s) = 0$. Otherwise, suppose that there is a point $t_0 \in C$ so that $g(t_0) > 0$. Since g is continuous, we can find ε and an interval (a, b) around t_0 so that $g \geq \varepsilon$ on (a, b). But then

$$\int_a^b g(s)\mathbf{1}_C(s)dYs \geq \int_a^b \varepsilon\mathbf{1}_C(s)dYs \geq \varepsilon(Y_b - Y_a) > 0,$$

which is a contradiction.

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 (5.1) we conclude that

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

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

$$\begin{aligned}
-rv(t,x,y) + \frac{\partial}{\partial t}v(t,x,y) + \frac{\partial}{\partial x}v(t,x,y)rx + \frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x,y)\sigma^2(t,x)x^2 &= 0; t < T, 0 < x \le y < \infty \\
\frac{\partial}{\partial y}v(t,y,y) &= 0; t \le T, y > 0 \qquad (4) \\
v(T,x,y) &= G(x,y); 0 \le x \le y \qquad (5) \\
v(t,0,y) &= e^{-r(T-t)}G(0,y). \qquad (6)
\end{aligned}$$

7 Appendix: The dynamics of Y

The results in this section is not used in deriving the PDE for the Lookback Option. However, it is interesting in shedding some more light into the behavior of Y_t. In

particular, we show here that Y is a singularly continuous process. That is it cannot be represented as a Ito process.
The question we'll address is: does there exist $\alpha(t), \beta(t)$ such that

$$dY_t = \alpha_t dt + \beta_t dW(t)? \tag{7}$$

This indeed cannot happen, as the following Lemma shows

Lemma 7.1. *Let Y_t be the running maximum of S_t where the volatility σ_t of S_t is positive. Then Y_t* **cannot be represented** *in the form of equation* (7)*. In other words, Y_t is a singularly continuous process.*

Proof.
Suppose there exist $\alpha(t), \beta(t)$ such that

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

Then since Y_t is increasing, $\langle Y \rangle_t = \int_0^t \beta_s^2 ds = 0$. 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 the above discussion, 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 (see the following Lemma), it follows that

$$Y_t = Y_0 + \int_0^t \alpha_s ds = Y_0,$$

with probability 1. We have observed that this forces $S_u \leq S_0$ on $[0, t]$ with probability 1 and this is not possible.

The above proof relies on the following lemma that tells us precisely how small the set $\{Y_t = S_t\}$ is:

Lemma 7.2. *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 **with** *$\sigma(t) > 0$. Define the set $C := \{t : Y_t = S_t\}$. Then with probability 1,*

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

In other words, the Lebesgue measure of the set $C(\omega)$, or the total length of $C(\omega)$, is 0 with probability 1.

Proof. By Fubini's theorem:

$$E\left(\int_0^T \mathbf{1}_{S_t=Y_t} dt\right) = \int_0^T E(\mathbf{1}_{S_t=Y_t})dt.$$

Since $\sigma_t > 0$, (S_t, Y_t) has a joint p.d.f. that can be explicitly computed. Thus $E(\mathbf{1}_{S_t=Y_t}) = 0$ for all t. It follows that

$$E\left(\int_0^T \mathbf{1}_{S_t=Y_t} dt\right) = 0.$$

Since the term inside the expectation is non-negative, it must be 0 with probability 1.