Selected Topics of the Optimal Stopping Theory

CHAPTER 1

Discrete time

1.1. Some motivating examples

We start with a few simple problems which illustrate well basic aspects of the optimal stopping theory. In all the examples, $S = (S_n)_{n \geq 0}$ is a symmetric random walk over integers, started at zero.

1. Assume further that we are interested in the quantity

$$V = \sup_{\tau} \mathbb{P}(S_\tau \geq 1), \tag{1.1}$$

where the supremum is taken over the class of all finite stopping time adapted to the natural filtration of S. More precisely, we want to compute the number V and exhibit the *optimal* stopping time τ_*, i.e., that for which the supremum is attained. The answer is trivial: we have $V \leq 1$, for obvious reasons, and the equality is attained for the stopping time $\tau_* = \inf\{n : S_n = 1\}$ (the finiteness of τ_* is a well-known consequence of the symmetry of S). Observe that the above choice is not unique: for example, the stopping time $\tilde{\tau}_* = \inf\{n : S_n = 2\}$ is also finite and satisfies $\mathbb{P}(S_{\tilde{\tau}_*} \geq 1) = 1$. However, it is easy to see that τ_* is the *smallest* optimal stopping time.

2. Next, suppose that we are interested in the quantity (1.1), but under the assumption that the stopping times τ are bounded. Then we still have $V = 1$. Indeed, the estimate $V \leq 1$ is obvious, and the reverse estimate is obtained by considering the stopping times $\tau_n = \tau \wedge n$, where $\tau = \inf\{n : S_n \geq 1\}$. Of course, each τ_n is bounded, and the finiteness of τ implies $\tau_n \to \tau$ and $S_{\tau_n} \to S_\tau$, so that

$$\lim_{n\to\infty} \mathbb{P}(S_{\tau_n} \geq 1) = \mathbb{P}(S_\tau \geq 1) = 1.$$

The difference in comparison to the previous example is that now there is no bounded optimal stopping time τ_*. Indeed, for any τ satisfying $\tau \leq N$ for some deterministic N, we have

$$\mathbb{P}(S_0 = 0, S_1 = -1, S_2 = -2, \ldots, S_N = -N) = 2^{-N}$$

and hence $\mathbb{P}(S_\tau \geq 1) \leq 1 - 2^{-N} < 1$.

3. Next, consider the following modification of (1.1):

$$V = \sup_{\tau} \Big(\mathbb{P}(S_\tau \geq 1) - \tfrac{1}{16}\mathbb{E}\tau \Big), \tag{1.2}$$

where the supremum is taken over the class of all finite stopping times. In other words, we want, as previously, to maximize the probability of getting to 1, but now we have the additional „cost of observation": with each time unit, our gain decreases by 1/16. At the first glance, the problem might seem a little more challenging, as there are two processes involved (the random walk and time): however, it is easy to reformulate the problem so that it depends on the process S only. To see this, note that in (1.2) we may restrict ourselves to *integrable* stopping times: indeed, if $\mathbb{E}\tau = \infty$, then the difference $\mathbb{P}(S_\tau \geq 1) - \frac{1}{16}\mathbb{E}\tau$ is equal to $-\infty$, so it does not contribute to the supremum in (1.2). Next, for $\tau \in L^1$, we have the Wald identity $\mathbb{E}S_\tau^2 = \mathbb{E}\tau$ and hence we may rewrite the problem in the form

$$V = \sup_{\tau \in L^1} \left(\mathbb{P}(S_\tau \geq 1) - \tfrac{1}{16}\mathbb{E}S_\tau^2 \right) = \sup_{\tau \in L^1} \mathbb{E}G(S_\tau),$$

where $G(x) = 1_{\{x \geq 1\}} - \frac{x^2}{16}$.

How can we compute the supremum above? One of the possible approaches is to try to *guess* the optimal stopping time, or at least to *restrict* considerations to some smaller class of stopping times. Let us go back to (1.2). It is evident that if the random walk reaches 1, then we should stop immediately: we cannot make the probability bigger, and the cost of observation will only increase. Next, a little thought suggests that if S drops too much (that is, we have $S_n = -b$ for some nonnegative threshold b), then we also should stop immediately: indeed, in such a case we would have to wait too long for S to climb to 1, and the cost of that would be too big. On the other hand, for the „intermediate" values (that is, for $S \in \{-b+1, -b+2, \ldots, 0\}$), it seems plausible to wait until we get to $-b$ or to 1.

Motivated by the above analysis, we introduce the stopping times

$$\sigma_b = \inf\{n : S_n \in \{-b, 1\}\}, \qquad b \in \{0, 1, 2, \ldots\}.$$

By the well-known properties of the symmetric random walk, we have

$$\mathbb{P}(S_{\sigma_b} = -b) = \frac{1}{b+1}, \qquad \mathbb{P}(S_{\sigma_b} = 1) = \frac{b}{b+1}$$

and hence

$$\mathbb{P}(S_{\sigma_b} \geq 1) - \tfrac{1}{16}\mathbb{E}S_{\sigma_b}^2 = \frac{b}{b+1} - \tfrac{1}{16}\left(\frac{b^2}{b+1} + \frac{b}{b+1}\right) = \frac{b}{b+1} - \frac{b}{16}.$$

By the very definition of V, this gives the estimate $V \geq \dfrac{b}{b+1} - \dfrac{b}{16}$. Optimizing over $b \in \{0, 1, 2, \ldots\}$, we see that the right hand side attains its maximum $\frac{9}{16}$ for $b = 3$. Consequently, we have proved that

$$V \geq \frac{9}{16}, \tag{1.3}$$

and it seems quite reasonable to *conjecture* that equality holds. To show this rigorously, we extend the problem to the case in which the random walk is allowed to start from an arbitrary point of the state space. That is, for any $x \in \mathbb{Z}$, we define

$$V(x) = \sup_{\tau \in L^1} \left\{ \mathbb{P}(S_\tau \geq 1) - \tfrac{1}{16}\mathbb{E}S_\tau^2 \,:\, S_0 = x \right\}.$$

Repeating the above argumentation, we get $V(x) = 1 - \frac{1}{16}x^2$ for $x \geq 1$ or $x \leq -b$ and

$$V(x) \geq \frac{b+x}{b+1} + \tfrac{1}{16}\big(bx - x - b\big)$$

otherwise. Optimizing over b, we get that the maximum is attained for $b = 3$ again (no matter what x is) and hence we may write

$$V(x) \geq \begin{cases} 1 - \frac{x^2}{16} & \text{if } x \leq -3 \text{ or } x \geq 1, \\ \frac{9}{16} + \frac{3}{8}x & \text{otherwise.} \end{cases}$$

Denote the right-hand side above by U. It is easy to see that U is a concave function on $\mathbb{Z}$ and we have $U \geq G$ (recall that $G(x) = 1_{\{x \geq 1\}} - \frac{x^2}{16}$.)

Now we return to the context in which S is assumed to start from zero. Suppose that τ is an arbitrary, finite stopping time. The concavity of U implies that the process $(U(S_n))_{n \geq 0}$ is a supermartingale and hence, by Doob's optional sampling theorem, the sequence $(U(S_{\tau \wedge n}))_{n \geq 0}$ also has this property. Thus, using the majorization $U \geq G$ we obtain

$$\mathbb{E}G(S_{\tau \wedge n}) \leq \mathbb{E}U(S_{\tau \wedge n}) \leq \mathbb{E}U(S_0) = U(0) = \frac{9}{16},$$

that is,

$$\mathbb{P}(S_{\tau \wedge n} \geq 1) - \tfrac{1}{16}\mathbb{E}(\tau \wedge n) = \mathbb{P}(S_{\tau \wedge n} \geq 1) - \tfrac{1}{16}\mathbb{E}S_{\tau \wedge n}^2 \leq \frac{9}{16}.$$

Letting $n \to \infty$ and using Lebesgue's monotone convergence theorem, we get

$$\mathbb{P}(S_\tau \geq 1) - \tfrac{1}{16}\mathbb{E}\tau \leq \frac{9}{16}.$$

But τ was arbitrary: hence, we obtain $V \leq \frac{9}{16}$, which combined with (1.3) gives

$$V = \frac{9}{16}.$$

The above analysis immediately yields the optimal stopping time: it is given by $\tau = \inf\{n : S_n \in \{-3, 1\}\}$. It is not difficult to see that this is a unique choice.

We conclude with several observations which will be important to us later.

- The above reasoning leads to the more general equality $V = U$ on $\mathbb{Z}$.
- As we have seen, the analysis has been split into the following steps:
  - (i) restriction to a smaller class of stopping times;
  - (ii) derivation of the corresponding expectations and optimization over parameters: this leads to the candidate U for the value function V, in particular this object satisfies $U \leq V$;
  - (iii) formal verification of the estimate $U \geq V$, with the use of martingale methods; more specifically, this was accomplished by noting that $(U(S_n))_{n \geq 0}$ is a supermartingale majorant of the sequence $(G(S_n))_{n \geq 0}$.

1.2. Martingale approach: description

Now we turn our attention to the more systematic presentation of the topic. In the optimal stopping theory, one distinguishes two techniques: the martingale approach and the Markovian approach, the purpose of this section is to describe the first of them.

Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space, equipped with a filtration $(\mathcal{F}_n)_{n \geq 0}$, i.e., a nondecreasing family of sub-σ-algebras of $\mathcal{F}$. Let $G = (G_n)_{n \geq 0}$ be an adapted sequence of random variables, i.e., we assume that for each $n \geq 0$ the random variable G_n is measurable with respect to $\mathcal{F}_n$. Our objective will be to stop this sequence so that the expected return is maximized. The stopping procedure is described by a random variable $\tau : \Omega \to \{0, 1, 2, \ldots\}$, which returns the value of the time when the sequence $(G_n)_{n\geq 0}$ should be stopped. Clearly, a reasonable procedure decides whether to stop the sequence at time n or not based on the observations up to time n; this means that for each n we have

$$\{\tau = n\} \in \mathcal{F}_n \qquad \text{for each } n \geq 0.$$

The random variable τ satisfying the above condition will be called a stopping time.

Let us put the discussion into a more precise framework. The optimal stopping problem concerns the study of

$$V_0 = \sup_{\tau} \mathbb{E} G_\tau, \tag{1.4}$$

where the supremum is taken over a certain family of adapted stopping times τ (which depends on the problem). We should point out that the study consists of two parts: (i) to compute the value V_0 as explicitly as possible; (ii) to identify the optimal stopping time τ_* (or the family of almost-optimal stopping times) for which the supremum V_0 is attained.

The first problem we encounter concerns the existence of $\mathbb{E}G_\tau$, to overcome which we need to impose some additional assumptions on G and τ. For example, if

$$\mathbb{E} \sup_{n\geq 0} |G_n| < \infty, \tag{1.5}$$

then the expectation $\mathbb{E}G_\tau$ is well defined for all stopping times τ. Another possibility is to restrict in (1.4) to those τ, for which the expectation exists. One way or another, we should emphasize that in general, this obstacle is just a technicality which is easily removed by some straightforward arguments (which might depend on the problem under the study). For the sake of simplicity and the clarity of the statements, we will assume that the condition (1.5) is satisfied, but it will be evident how to relax this requirement in other contexts.

So, let us assume that the supremum in (1.4) is taken over the family $\mathcal{M}$ of all stopping times τ. A successful treatment of this problem requires the introduction, for each $n \leq N$, the smaller family

$$\mathcal{M}_n^N = \{\tau \in \mathcal{M} : n \leq \tau \leq N\}.$$

We will also write $\mathcal{M}^N = \mathcal{M}_0^N$ and $\mathcal{M}_n = \mathcal{M}_n^\infty$. These families give rise to the related value functions

$$V_n^N = \sup_{\tau \in \mathcal{M}_n^N} \mathbb{E} G_\tau, \tag{1.6}$$

and we will use the notation $V^N = V_0^N$, $V_n = V_n^\infty$ and $V = V_0^\infty$. The primary goal of this section is to present the solution to (1.6) with the use of martingale approach.

1.2.1. Martingale approach: finite horizon. If $N < \infty$ (the case of "finite horizon"), then the problem (1.6) can be easily solved by means of the backward induction. Indeed, let us fix a nonnegative integer N and try to inspect the value functions as n decreases from N to 0. If $n = N$, then the class $\mathcal{M}_n^N$ consists of one stopping time $\tau \equiv N$ only and hence the optimal gain is equal to G_N (and $V_N^N = \mathbb{E}G_N$). If $n = N - 1$, then we have two choices for the stopping time: we can either stop at time $N - 1$ or continue and stop at time N. In the first case our gain is G_{N-1}; in the second case we do not know what the random variable G_N will be, so we can only say that on average, we will obtain $\mathbb{E}(G_N|\mathcal{F}_{N-1})$. Therefore, if

$G_{N-1} \geq \mathbb{E}(G_N|\mathcal{F}_{N-1})$, we should stop immediately; otherwise, we should continue. For smaller values of n we proceed similarly. More precisely, define recursively the sequence $(B_n^N)_{0\leq n\leq N}$, representing the optimal gains at times 0, 1, 2, ..., N, as follows:

$$
\begin{aligned}
B_N^N &= G_N, \\
B_n^N &= \max\left\{G_n, \mathbb{E}(B_{n+1}^N|\mathcal{F}_n)\right\}, \qquad n = N-1,\, N-2,\, \dots .
\end{aligned} \tag{1.7}
$$

The above discussion also suggests to consider the family of stopping times

$$
\tau_n^N = \inf\left\{k \in \{n,\, n+1,\, \dots,\, N\} \,:\, B_k^N = G_k\right\}, \tag{1.8}
$$

for $n = 0,\, 1,\, 2,\, \dots,\, N$.

THEOREM 1.1. *Suppose that N is a fixed integer and the sequence $G = (G_k)_{k=n}^N$ satisfies $\mathbb{E}\max_{n\leq k\leq N}|G_k| < \infty$. Consider the optimal stopping problem* (1.6) *and the sequence $(B_k^N)_{k=n}^N$, defined by* (1.7).

(i) The sequence $(B_k^N)_{k=n}^N$ is the smallest supermartingale majorizing $(G_k)_{k=n}^N$. In addition, the stopped sequence $\left(B_{\tau_n^N\wedge k}^N\right)_{k=n}^N$ is a martingale.

(ii) For any $0 \leq n \leq N$ we have, with probability 1,

$$
B_n^N \geq \mathbb{E}(G_\tau|\mathcal{F}_n) \qquad \text{for any } \tau \in \mathcal{M}_n^N, \tag{1.9}
$$

$$
B_n^N = \mathbb{E}(G_{\tau_n^N}|\mathcal{F}_n). \tag{1.10}
$$

(iii) The stopping time τ_n^N is optimal in (1.6) *and any other optimal stopping time τ_* satisfies $\tau_* \geq \tau_n^N$ almost surely.*

PROOF. (i) The supermartingale property and the majorization follow directly from the definition of the sequence $(B_n^N)_{n=0}^N$. If $(\bar{B}_n^N)_{n=0}^N$ is another supermartingale majorizing $(G_k)_{k=n}^N$, then the desired inequality $B_k^N \leq \bar{B}_k^N$ almost surely, $k = n,\, n+1,\, N+2,\, \dots,\, N$, can be proved by backward induction. Indeed, the estimate is trivial for $k = N$ (we have $B_N^N = G_N \leq \bar{B}_N^N$, by the majorization property of $\bar{B}$), and assuming its validity for k, we see that

$$
\bar{B}_{k-1}^N \geq \max\{G_{k-1}, \mathbb{E}(\bar{B}_k^N|\mathcal{F}_{k-1})\} \geq \max\{G_{k-1}, \mathbb{E}(B_k^N|\mathcal{F}_{k-1})\} = B_{k-1}^N.
$$

So, it remains to prove the martingale property of the stopped process $\left(B_{\tau_n^N\wedge k}^N\right)_{k=n}^N$. We compute directly that

$$
\begin{aligned}
\mathbb{E}\left[B_{\tau_n^N\wedge(k+1)}^N \,|\, \mathcal{F}_k\right] &= \mathbb{E}\left[B_{\tau_n^N\wedge(k+1)}^N 1_{\{\tau_n^N\leq k\}} \,|\, \mathcal{F}_k\right] + \mathbb{E}\left[B_{\tau_n^N\wedge(k+1)}^N 1_{\{\tau_n^N> k\}} \,|\, \mathcal{F}_k\right] \\
&= \mathbb{E}\left[B_{\tau_n^N\wedge k}^N 1_{\{\tau_n^N\leq k\}} \,|\, \mathcal{F}_k\right] + \mathbb{E}\left[B_{k+1}^N 1_{\{\tau_n^N> k\}} \,|\, \mathcal{F}_k\right] \\
&= B_{\tau_n^N\wedge k}^N 1_{\{\tau_n^N\leq k\}} + 1_{\{\tau_n^N> k\}}\mathbb{E}\left[B_{k+1}^N \,|\, \mathcal{F}_k\right].
\end{aligned}
$$

However, on the set $\{\tau_n^N > k\}$ we have $B_k^N > G_k$ and hence $B_k^N = \mathbb{E}(B_{k+1}^N|\mathcal{F}_k)$. This shows the identity $\mathbb{E}\left[B_{\tau_n^N\wedge(k+1)}^N \,|\, \mathcal{F}_k\right] = B_{\tau_n^N\wedge k}^N 1_{\{\tau_n^N\leq k\}} + 1_{\{\tau_n^N>k\}} B_k^N = B_{\tau_n^N\wedge k}^N$ and part (i) is established.

(ii) This follows at once from (i) and Doob's optional sampling theorem.

(iii) Taking the expectations in (1.9) and (1.10) gives $\mathbb{E}G_\tau \leq \mathbb{E}B_n^N = \mathbb{E}G_{\tau_n^N}$ for all $\tau \in \mathcal{M}_n^N$, showing that τ_n^N is indeed the optimal stopping time. Suppose that τ_*

is another optimal stopping time. Then $B^N_{\tau_*} = G_{\tau_*}$ almost surely, since otherwise we would have

$$\mathbb{E}G_{\tau_*} < \mathbb{E}B^N_{\tau_*} \le \mathbb{E}B^N_n = \mathbb{E}G_{\tau^N_n},$$

where the second inequality follows from Doob's optional sampling theorem and the supermartingale property of the sequence $(B^N_k)^N_{k=n}$. The contradiction shows that B_{τ_*} and G_{τ_*} must coincide, and clearly τ^N_n is the smallest stopping time which has this property. □

1.2.2. Martingale approach: infinite horizon. The above method required N to be a finite integer, since we have needed the variable G_N to start the backward recurrence. In the case $N = \infty$ one could try to use approximation-type arguments (of the form $V^\infty_n = \lim_{N\to\infty} V^N_n$), but these do not necessarily work in general, so we will proceed in a different manner. By (1.9) and (1.10) it seems tempting to write

$$B^N_n = \sup_{\tau \in \mathcal{M}^N_n} \mathbb{E}(G_\tau | \mathcal{F}_n).$$

However, two problems arise. The first is that (1.9) and (1.10) hold true on a set of full measure only which might depend on the stopping time, so the above identity might fail to hold. A second obstacle is that the supremum on the right need not be even measurable. To overcome these difficulties, a typical argument in the theory of optimal stopping is to introduce the concept of essential supremum.

Definition 1.1. Let $(Z_\alpha)_{\alpha\in I}$ be a family of random variables. Then there is a countable subset J of I such that the random variable $\overline{Z} = \sup_{\alpha\in J} Z_\alpha$ satisfies the following two properties:

(i) $\mathbb{P}(Z_\alpha \le \overline{Z}) = 1$ for each $\alpha \in I$,

(ii) if $\tilde{Z}$ is another random variable satisfying (i) in the place of $\overline{Z}$, then $\mathbb{P}(\overline{Z} \le \tilde{Z}) = 1$.

The random variable $\overline{Z}$ is called the *essential supremum* of $(Z_\alpha)_{\alpha\in I}$ and is denoted by $\operatorname{ess\,sup}_{\alpha\in I} Z_\alpha$. In addition, if $\{Z_\alpha : \alpha \in I\}$ is upwards directed in the sense that for any α, $\beta \in I$ there is $\gamma \in I$ such that $\max\{Z_\alpha, Z_\beta\} \le Z_\gamma$, then the countable set $J = \{\alpha_1, \alpha_2, \ldots\}$ can be chosen so that $Z_{\alpha_1} \le Z_{\alpha_2} \le \ldots$ and $\operatorname{ess\,sup}_{\alpha\in I} Z_\alpha = \lim_{n\to\infty} Z_{\alpha_n}$.

Now we see that (1.9) and (1.10) give the identity

$$B^N_n = \operatorname*{ess\,sup}_{\tau \in \mathcal{M}^N_n} \mathbb{E}(G_\tau | \mathcal{F}_n) \tag{1.11}$$

with probability 1. A nice feature of this alternative characterization of the sequence $(B^N_n)^N_{n=0}$ is that it extends naturally to the setting of infinite horizon (i.e., for $N = \infty$) and, as we shall prove now, provides the desired solution.

So, consider the optimal stopping problem (1.6) for $N = \infty$:

$$V_n = \sup_{\tau \ge n} \mathbb{E}G_\tau. \tag{1.12}$$

For $n = 0, 1, 2, \ldots$, introduce the random variable

$$B_n = \operatorname*{ess\,sup}_{\tau \ge n} \mathbb{E}(G_\tau | \mathcal{F}_n) \tag{1.13}$$

and the stopping time

$$\tau_n = \inf\{k \ge n : B_k = G_k\}, \tag{1.14}$$

with the usual convention $\inf \emptyset = \infty$. In the literature, the sequence $(B_n)_{n\geq 0}$ is often referred to as the Snell envelope of G.

We will establish the following analogue of Theorem 1.1.

THEOREM 1.2. *Suppose that the sequence $(G_n)_{n\geq 0}$ satisfies $\mathbb{E}\sup_{n\geq 0}|G_n| < \infty$ and consider the optimal stopping problem* (1.12). *Then the following statements hold true.*

(i) For any $n \geq 0$ we have the recurrence relation

$$B_n = \max(G_n, \mathbb{E}(B_{n+1}|\mathcal{F}_n)).$$

(ii) We have $\mathbb{P}(B_n \geq \mathbb{E}(G_\tau|\mathcal{F}_n)) = 1$ for all $\tau \in \mathcal{M}_n$ and, if the stopping time τ_n is finite almost surely, then $\mathbb{P}(B_n = \mathbb{E}(G_{\tau_n}|\mathcal{F}_n)) = 1$.

(iii) If $\mathbb{P}(\tau_n < \infty) = 1$, then τ_n is optimal in (1.12). *Furthermore, if τ_* is another optimal stopping time for* (1.12), *then $\tau_n \leq \tau_*$ almost surely.*

(iv) The sequence $(B_k)_{k\geq n}$ is the smallest supermartingale which majorizes $(G_k)_{k\geq n}$. Moreover, the stopped process $(B_{\tau_n\wedge k})_{k\geq n}$ is a martingale.

PROOF. We will only establish (i), the other parts can be shown with the argumentation similar to that used in the proof of Theorem 1.1. We need to show two inequalities to prove the identity. Take $\tau \in \mathcal{M}_n$ and let $\tau' = \tau \vee (n+1)$. Then $\tau' \in \mathcal{M}_{n+1}$ and since $\{\tau \geq n+1\} \in \mathcal{F}_n$, we may write

$$\begin{aligned}
\mathbb{E}(G_\tau|\mathcal{F}_n) &= \mathbb{E}(G_\tau 1_{\{\tau=n\}}|\mathcal{F}_n) + \mathbb{E}(G_\tau 1_{\{\tau\geq n+1\}}|\mathcal{F}_n) \\
&= 1_{\{\tau=n\}}G_n + 1_{\{\tau\geq n+1\}}\mathbb{E}(G_{\tau'}|\mathcal{F}_n) \\
&= 1_{\{\tau=n\}}G_n + 1_{\{\tau\geq n+1\}}\mathbb{E}\big[\mathbb{E}(G_{\tau'}|\mathcal{F}_{n+1})|\mathcal{F}_n\big] \\
&\leq 1_{\{\tau=n\}}G_n + 1_{\{\tau\geq n+1\}}\mathbb{E}(B_{n+1}|\mathcal{F}_n) \\
&\leq \max\{G_n, \mathbb{E}(B_{n+1}|\mathcal{F}_n)\}.
\end{aligned}$$

This proves the inequality "$\leq$". To show the reverse, observe that the family $\{\mathbb{E}(B_\tau|\mathcal{F}_{n+1}) \ : \ \tau \in \mathcal{M}_{n+1}\}$ is upwards directed. Indeed, if α, $\beta \in \mathcal{M}_{n+1}$ and we set $\gamma = \alpha 1_A + \beta 1_{\Omega\setminus A}$, where $A = \{\mathbb{E}(G_\alpha|\mathcal{F}_{n+1}) \geq \mathbb{E}(G_\beta|\mathcal{F}_{n+1})\}$, then γ is a stopping time belonging to $\mathcal{M}_{n+1}$ and

$$\begin{aligned}
\mathbb{E}(G_\gamma|\mathcal{F}_{n+1}) &= \mathbb{E}(G_\alpha 1_A + G_\beta 1_{\Omega\setminus A}|\mathcal{F}_{n+1}) \\
&= 1_A\mathbb{E}(G_\alpha|\mathcal{F}_{n+1}) + 1_{\Omega\setminus A}\mathbb{E}(G_\beta|\mathcal{F}_{n+1}) \\
&= \max\big\{\mathbb{E}(G_\alpha|\mathcal{F}_{n+1}), \mathbb{E}(G_\beta|\mathcal{F}_{n+1})\big\}.
\end{aligned}$$

Therefore, there is a sequence $\{\sigma_k : k \geq 1\}$ in $\mathcal{M}_{n+1}$ such that

$$\operatorname*{ess\,sup}_{\tau\in\mathcal{M}_{n+1}} \mathbb{E}(G_\tau|\mathcal{F}_{n+1}) = \lim_{k\to\infty} \mathbb{E}(G_{\sigma_k}|\mathcal{F}_{n+1})$$

and $\mathbb{E}(G_{\sigma_1}|\mathcal{F}_{n+1}) \leq \mathbb{E}(G_{\sigma_2}|\mathcal{F}_{n+1}) \leq \ldots$ with probability 1. Now we can write, by Lebesgue's monotone convergence theorem,

$$\begin{aligned}
\mathbb{E}(B_{n+1}|\mathcal{F}_n) &= \mathbb{E}\left(\operatorname*{ess\,sup}_{\tau\in\mathcal{M}_{n+1}} \mathbb{E}(G_\tau|\mathcal{F}_{n+1})\middle|\mathcal{F}_n\right) \\
&= \mathbb{E}\left(\lim_{k\to\infty} \mathbb{E}(G_{\sigma_k}|\mathcal{F}_{n+1})\middle|\mathcal{F}_n\right) \\
&= \lim_{k\to\infty} \mathbb{E}\big(\mathbb{E}(G_{\sigma_k}|\mathcal{F}_{n+1})|\mathcal{F}_n\big) = \lim_{k\to\infty} \mathbb{E}(G_{\sigma_k}|\mathcal{F}_n) \leq B_n.
\end{aligned}$$

Since $B_n \geq G_n$ (which can be trivially obtained by considering $\tau \equiv n$ in the definition of B_n), we get the desired identity. $\square$

In the remaining part of this subsection, let us inspect the connection between the contexts of finite and infinite horizons. One easily checks that the random variables B_n^N and τ_n^N do not decrease as we increase N. Consequently, the limits

$$B_n^\infty := \lim_{N\to\infty} B_n^N \qquad \text{and} \qquad \tau_n^\infty := \lim_{N\to\infty} \tau_n^N$$

exist on a set of full measure. Furthermore, we also see that the sequence $(V_n^N)_{N=n}^\infty$ is nondecreasing, so the quantity $V_n^\infty = \lim_{N\to\infty} V_n^\infty$ is well-defined. Now it follows directly from (1.8), (1.11), (1.13) and (1.14) that

$$B_n^\infty \leq B_n \qquad \text{and} \qquad \tau_n^\infty \leq \tau_n \tag{1.15}$$

almost surely. Moreover, we also have

$$V_n^\infty \leq V_n. \tag{1.16}$$

THEOREM 1.3. *Suppose that* $\mathbb{E}\sup_{n\geq 0}|G_n| < \infty$ *and consider the optimal stopping problems* (1.6) *and* (1.12). *Then equalities hold in* (1.15) *and* (1.16).

PROOF. Letting $N \to \infty$ in the recurrence relation (1.7) yields

$$B_n^\infty = \max\{G_n, \mathbb{E}(B_{n+1}^\infty|\mathcal{F}_n)\}, \qquad n = 0, 1, 2, \ldots,$$

by Lebesgue's monotone convergence theorem. Consequently, $(B_n^\infty)_{n\geq 0}$ is an adapted supermartingale dominating $(G_n)_{n\geq 0}$. Thus $B_n^\infty \geq B_n$ for each n, by the fourth part of the preceding theorem. This shows the identity $B^\infty = B$ almost surely, and the remaining equalities follow immediately. $\square$

EXAMPLE 1.1. Strict inequalities may hold in (1.15) and (1.16) if the integrability condition on $\sup_{n\geq 0}|G_n|$ is not imposed. To see this, let ε_0, ε_1, ε_2, $\ldots$ be a sequence of independent Rademacher variables and set $G_n = \varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_n$. Then the process $(G_n)_{n\geq 0}$ is a martingale with respect to the natural filtration, so $V_n^N = 0$, $B_n^N = G_n$ and $\tau_n^N = n$ for all $0 \leq n \leq N < \infty$. Consequently, these identities are preserved in the limit: $V_n^\infty = 0$, $B_n^\infty = G_n$ and $\tau_n^\infty = n$ for all n. On the other hand, it is well-known that for any positive integer a, the stopping time $\sigma_n = \inf\{k \geq n : G_k = a\}$ is finite almost surely and hence $V_n \geq \mathbb{E}G_{\sigma_n} = a$. Since a was arbitrary, we get $V_n = \infty$, $B_n = \infty$ and $\tau_n = \infty$ with probability 1.

1.2.3. An application: a difference prophet inequality for bounded random variables. We take the opportunity to present here some information about the so-called "prophet inequalities", a distinguished class of estimates arising in the theory of optimal stopping. As in the preceding section, we assume that we are given a sequence G (finite or infinite) of random variables adapted to some filtration. The idea is as follows: under some boundedness condition on G, find universal inequalities which compare $M = \mathbb{E}\sup_n G_n$, the expected supremum of the sequence, with $V = \sup_\tau \mathbb{E}G_\tau$, the optimal stopping value of the sequence; here τ runs over the class of all finite stopping times adapted to the underlying filtration. The term "prophet inequality" arises from the optimal-stopping interpretation of M, which is the optimal expected return of a player endowed with complete foresight; this player observes the sequence G and may stop whenever he wants, incurring a reward equal to the variable at the time of stopping. With complete foresight, such a player obviously stops always when the largest value is observed, and on the average, his reward is equal to M. On the other hand, the quantity V corresponds to the optimal return of the non-prophet player.

The literature on the subject is quite large (see the bibliographic details at the end of this chapter). We will mention only two results here. The first result in this direction is the estimate of Krengel, Sucheston and Garling, which asserts that if $G_1, G_2, \ldots$ are independent and nonnegative, then we have

$$M \leq 2V$$

and the constant 2 is the best possible. Sometimes such estimates are called ratio prophet inequalities, as they provide an efficient upper bound for the ratio M/V. The above estimate is left as an exercise: see Problem 1.4 below. Instead, we will study the following related *difference* prophet inequality.

THEOREM 1.4. *If $G_1, G_2, \ldots$ are independent random variables taking values in an interval $[a, b]$, then*

$$M \leq V + \frac{b-a}{4},$$

and the constant $(b-a)/4$ cannot be decreased.

Note that it suffices to show the claim for $a = 0$ and $b = 1$ only, by simple translation and dilation arguments. The proof of the above statement will involve a certain transformation of random variables, which we describe now. Namely, if Y is an arbitrary integrable random variable (taking values in $\mathbb{R}$) and $a < b$, then we define

$$Y_a^b(\omega) = \begin{cases} Y(\omega) & \text{if } Y(w) \notin [a,b], \\ a \text{ or } b & \text{if } Y(w) \in [a,b], \end{cases}$$

with

$$\mathbb{P}(Y_a^b = a) = \frac{1}{b-a}\mathbb{E}(b-Y)1_{\{Y \in [a,b]\}}, \quad \mathbb{P}(Y_a^b = b) = \frac{1}{b-a}\mathbb{E}(Y-a)1_{\{Y \in [a,b]\}}.$$

The probabilities $\mathbb{P}(Y_a^b = a)$, $\mathbb{P}(Y_a^b = b)$ are uniquely determined by the requirement $\mathbb{E}Y_a^b = \mathbb{E}Y$. We will need the following two simple lemmas.

LEMMA 1.1. *If X is a random variable independent of Y and Y_a^b, then*

$$\mathbb{E}\max\{X, Y\} \leq \mathbb{E}\max\{X, Y_a^b\}.$$

PROOF. By the convexity of the function $y \mapsto \max\{x, y\}$ (where $x \in \mathbb{R}$ is fixed), we get

$$\begin{aligned} &\mathbb{E}\max\{x, Y\}1_{\{Y \in [a,b]\}} \\ &\leq \frac{1}{b-a}\mathbb{E}(b-Y)1_{\{Y \in [a,b]\}}\max\{x, a\} + \frac{1}{b-a}\mathbb{E}(Y-a)1_{\{Y \in [a,b]\}}\max\{x, b\}. \end{aligned}$$

Therefore, since X and Y are independent, we get

$$\begin{aligned} &\mathbb{E}\max\{X, Y\}1_{\{Y \in [a,b]\}} \\ &\leq \frac{1}{b-a}\mathbb{E}(b-Y)1_{\{Y \in [a,b]\}}\max\{X, a\} + \frac{1}{b-a}\mathbb{E}(Y-a)1_{\{Y \in [a,b]\}}\max\{X, b\}. \end{aligned}$$

Thus we obtain

$$\begin{aligned}\mathbb{E}\max\{X,Y\} &= \mathbb{E}\max\{X,Y\}1_{\{Y\notin[a,b]\}} + \mathbb{E}\max\{X,Y\}1_{\{Y\in[a,b]\}}\\ &\le \mathbb{E}\max\{X,Y_a^b\}1_{\{Y\notin[a,b]\}} + \frac{1}{b-a}\mathbb{E}(b-Y)1_{\{Y\in[a,b]\}}\max\{X,a\}\\ &\quad + \frac{1}{b-a}\mathbb{E}(Y-a)1_{\{Y\in[a,b]\}}\max\{X,b\}\\ &= \mathbb{E}\max\{X,Y_a^b\}. \qquad\square\end{aligned}$$

In the second lemma, we will use the notation

$$D(G_1,G_2,\ldots,G_n) = \mathbb{E}\max_{1\le k\le n} G_k - \sup_{1\le\tau\le n}\mathbb{E}G_\tau.$$

The purpose of the statement below is to show that for any sequence of $n>2$ random variables there is a sequence of $n-1$ random variables offering at least as large an additive advantage to the prophet.

LEMMA 1.2. *For any $n\ge 2$ and any independent random variables G_1, G_2, …, G_n with values in $[0,1]$, there is a random variable W with values in $\{0,1\}$ independent of G_2, …, G_{n-2}, satisfying*

$$D(G_1,G_2,\ldots,G_n) \le D(\mu,G_2,\ldots,G_{n-2},W),$$

where $\mu = \sup_{2\le\tau\le n}\mathbb{E}G_\tau$.

PROOF. By the theory of optimal stopping, we know that

$$\begin{aligned}V_1^n = V_1^n(G_1,G_2,\ldots,G_n) &= \mathbb{E}\max\{G_1,\mathbb{E}\max\{G_2,\ldots\}\}\\ &= \mathbb{E}\max\{G_1,\mu\} = \mu + \mathbb{E}(G_1-\mu)^+.\end{aligned}$$

Furthermore,

$$\mathbb{E}\max\{G_1,G_2,\ldots,G_n\} \le \mathbb{E}\max\{\mu,G_2,\ldots,G_n\} + \mathbb{E}(G_1-\mu)^+,$$

and hence $D(G_1,G_2,\ldots,G_n)\le D(\mu,G_2,\ldots,G_n)$. Now consider the random variables $Y=(G_{n-1})^1_{\mathbb{E}G_n}$ and $Z=(G_n)^1_0$, independent of each other and of the sequence G_2, G_3, …, G_{n-2}. We have

$$\mathbb{E}\max\{Y,\mathbb{E}Z\} = \mathbb{E}\max\{Y,\mathbb{E}G_n\} = \mathbb{E}\max\{G_{n-1},G_n\},$$

which gives $V^n(\mu,G_2,\ldots,G_n) = V^n(\mu,G_2,\ldots,G_{n-2},Y,Z)$. By the previous lemma,

$$\mathbb{E}\max\{\mu,G_2,\ldots,G_n\} \le \mathbb{E}\max\{\mu,G_2,\ldots,G_{n-2},Y,Z\},$$

so $D(\mu,G_2,\ldots,G_n)\le D(\mu,G_2,\ldots,G_{n-2},Y,Z)$. Take any random variable W independent of G_2, …, G_{n-2} and satisfying $\mathbb{P}(W=1)=\mathbb{E}\max\{Y,\mathbb{E}Z\}=1-\mathbb{P}(W=0)$. We have $\mathbb{E}W=\mathbb{E}\max\{Y,\mathbb{E}Z\}$, so

$$V_1^n(\mu,G_2,\ldots,G_n) = V_1^{n-1}(\mu,G_2,\ldots,G_{n-2},W).$$

Next, we have

$$\begin{aligned}\mathbb{P}(W=1) = \mathbb{E}\max\{Y,\mathbb{E}Z\} &= \mathbb{P}(Y=1)+\mathbb{E}G_n\mathbb{P}(Y<1)\\ &= \mathbb{P}(Y=1)+\mathbb{P}(Z=1)\mathbb{P}(Y<1) = \mathbb{P}(Y=1 \text{ or } Z=1).\end{aligned}$$

Finally, we have $\mathbb{E}G_n\le\mu$, which combined with the above identity gives

$$\mathbb{E}\max\{\mu,G_2,\ldots,G_n\} = \mathbb{E}\max\{\mu,G_2,\ldots,G_{n-2},W\}.$$

Consequently, we get

$$D(\mu, G_2, \dots, G_{n-2}, Y, Z) = D(\mu, G_2, \dots, G_{n-2}, W)$$

and the proof is complete. □

PROOF OF THEOREM 1.4. Clearly, it suffices to show the claim for finite number of variables. By the second lemma above, we may reduce this number to 2: $D(G_1, G_2) \leq \frac{1}{4}$. However, the above proof shows that if we set $\mu = \mathbb{E}G_2$, then $D(G_1, G_2) \leq D(\mu, Z)$, where $\mathbb{P}(Z = 1) = \mu = 1 - \mathbb{P}(Z = 0)$. It suffices to note that

$$D(\mu, Z) = \mathbb{E}\max\{\mu, Z\} - \mathbb{E}\max\{\mu, \mathbb{E}Z\} = \mu \cdot 1 + (1 - \mu) \cdot \mu - \mu = \mu - \mu^2 \leq 1/4.$$

The equality holds for $\mu = 1/2$, and the above considerations immediately yield the example for which the difference $1/4$ is attained. Indeed, take the pair (G_1, G_2), where $G_1 \equiv 1/2$ and the distribution of G_2 is given by $\mathbb{P}(G_2 = 0) = \mathbb{P}(G_2 = 1) = 1/2$. □

1.2.4. An example. Let X_0, X_1, X_2, … be i.i.d. random variables following the $Exp(1)$ law, and let $c > 0$ be a fixed constant. We will solve the optimal stopping problem

$$V = \sup_{\tau \in \mathcal{M}} \mathbb{E}\Big[\max\{X_0,\, X_1,\, X_2,\, \dots,\, X_\tau\} - c\tau\Big].$$

For the sake of clarity, we will split the reasoning into three separate steps. To put this problem into the general framework of the optimal stopping theory, we set

$$G_n = \max\{X_0,\, X_1,\, X_2,\, \dots,\, X_n\} - cn.$$

Then $V = \sup_{\tau \in \mathcal{M}} G_\tau$ and we may proceed.

STEP 1. GUESSING THE OPTIMAL STOPPING RULE AND THE ASSOCIATED EXPECTATION. This is an informal step and it requires some thought and experimentation. It seems reasonable to conjecture that the optimal stopping rule should be of the following threshold type:

$$\tau_a = \inf\{n : X_n \geq a\}$$

for some unknown constant a. To find a, let us first compute the corresponding expectation

$$\mathbb{E}\Big[\max\{X_0,\, X_1,\, X_2,\, \dots,\, X_{\tau_a}\} - c\tau_a\Big]. \tag{1.17}$$

Note that τ_a has the geometric distribution with parameter $\mathbb{P}(X_0 \geq a)$ and hence

$$\mathbb{E}\tau_a = \frac{\mathbb{P}(X_0 < a)}{\mathbb{P}(X_0 \geq a)} = \frac{1 - e^{-a}}{e^{-a}} = e^a - 1.$$

Furthermore, we have

$$\begin{aligned}\mathbb{E}\max\{X_0,\, X_1,\, X_2,\, \dots,\, X_{\tau_a}\} &= \mathbb{E}X_{\tau_a} = \mathbb{E}(X_0 | X_0 \geq a)\\ &= \frac{1}{\mathbb{P}(X_0 \geq a)} \int_{\{X_0 \geq a\}} X_0 \mathrm{d}\mathbb{P}\\ &= e^a \left(ae^{-a} + 1 \cdot e^{-a}\right) = a + 1\end{aligned}$$

and hence the expectation (1.17) equals $a+1-c(e^a-1)$. We want to maximize this expectation (over all possible stopping times, and so, in particular, over τ_a): we easily check that

$$\max_a \Big\{a+1-c(e^a-1)\Big\} = \begin{cases} 1 & \text{if } c \geq 1 \text{ (maximum attained at } a_* = 0), \\ c - \ln c & \text{if } c < 1 \text{ (maximum attained at } a_* = -\ln c). \end{cases}$$

Let us denote the right-hand side by $\tilde{V}$. This is the candidate for the value of our optimal stopping problem.

STEP 2. GUESSING THE SNELL ENVELOPE. Actually, the computation from the previous step easily yields the corresponding candidate for the Snell envelope. From the general theory, we know that

$$B_n = \operatorname*{esssup}_{\tau \geq n} \mathbb{E}(G_\tau|\mathcal{F}_n).$$

Take $\tau = \tau_{a_*} \vee n$ (the additional maximum with n is to enforce the estimate $\tau \geq n$): by the above computations, for this special stopping time, we have

$$\mathbb{E}(G_\tau|\mathcal{F}_n) = \begin{cases} \max\{X_0, X_1, \ldots, X_n\} - cn & \text{if } \max\{X_0, X_1, \ldots, X_n\} \geq a_*, \\ \tilde{V} - c(n+1) & \text{if } \max\{X_0, X_1, \ldots, X_n\} < a_*. \end{cases}$$

Let us denote the right-hand side by $\tilde{B}_n$: this is our candidate for the Snell envelope of $(G_n)_{n\geq 0}$. By the very definition, we have $\tilde{B}_n \leq B_n$.

STEP 3. VERIFICATION OF THE PROPERTIES OF $\tilde{B}$. Now we will check that $(\tilde{B}_n)_{n\geq 0}$ is a supermartingale majorizing $(G_n)_{n\geq 0}$. Then by the general theory we will obtain the reverse estimate $\tilde{B}_n \geq B_n$, which will show that $\tilde{B}$ coincides with the Snell envelope and the stopping time τ_{a_*} is optimal.

We start with the majorization. On the set $\{\max\{X_0, X_1, \ldots, X_n\} \geq a_*\}$ we have $\tilde{B}_n = G_n$. On the other hand, on $\{\max\{X_0, X_1, \ldots, X_n\} < a_*\}$ (which is nonempty iff $a_* > 0$, i.e., $c < 1$), the majorization is equivalent to

$$\tilde{V} - c(n+1) \geq \max\{X_0, X_1, \ldots, X_n\} - cn$$

or $-\ln c \geq \max\{X_0, X_1, \ldots, X_n\}$: but this is trivial, since $-\ln c = a_*$.

It remains to check the supermartingale property of $\tilde{B}_n$:

$$\mathbb{E}(\tilde{B}_{n+1}|\mathcal{F}_n) \leq \tilde{B}_n, \qquad n = 0, 1, 2, \ldots. \tag{1.18}$$

On the set $\{\max\{X_0, X_1, \ldots, X_n\} \geq a_*\} \in \mathcal{F}_n$ we automatically have the bound $\max\{X_0, X_1, \ldots, X_{n+1}\} \geq a_*$ and hence

$$\mathbb{E}(\tilde{B}_{n+1}|\mathcal{F}_n) = \mathbb{E}\Big(\max\{X_0, X_1, \ldots, X_{n+1}\} - c(n+1)|\mathcal{F}_n\Big).$$

Now, consider the random variable $\xi = \max\{X_0, X_1, \ldots, X_n\}$. But for any $a > a_*$,

$$\mathbb{E}\Big(\max\{a, X_{n+1}\} - c(n+1)\Big) \leq a - cn$$

(this is equivalent to $e^{-a} \leq c$ and holds true, since $e^{-a_*} \leq c$). Hence

$$\begin{aligned}\mathbb{E}\Big(\max\{X_0, X_1, \ldots, X_{n+1}\} - c(n+1)|\mathcal{F}_n\Big) &= \mathbb{E}\Big(\max\{a, X_{n+1}\} - c(n+1)\Big)\Big|_{a=\xi} \\ &\leq \xi - c_n = \tilde{B}_n.\end{aligned}$$

Next, we analyze (1.18) on the set $\{\max\{X_0, X_1, \ldots, X_n\} < a_*\}$ (which is nonempty iff $a_* > 0$, i.e., $c < 1$). On this set, we have the identity

$$\tilde{B}_{n+1} = \begin{cases} X_{n+1} - c(n+1) & \text{if } X_{n+1} \geq a_*, \\ \tilde{V} - c(n+2) & \text{if } X_{n+1} < a_*, \end{cases}$$

which is independent of $\mathcal{F}_n$. Consequently, the conditional expectation $\mathbb{E}(\tilde{B}_{n+1}|\mathcal{F}_n)$ is equal to the average of the right-hand side above, i.e., to

$$\begin{aligned} &\mathbb{E}\Big[(X_{n+1} - c(n+1))1_{\{X_{n+1}\geq a_*\}} + (\tilde{V} - c(n+2))1_{\{X_{n+1}<a_*\}}\Big] \\ &= \mathbb{E}X_{n+1}1_{\{X_{n+1}\geq a_*\}} - c(n+1) + (\tilde{V} - c)\mathbb{P}(X_{n+1} < a_*) \\ &= e^{-a_*}(a_* + 1) - c(n+1) + (\tilde{V} - c)(1 - e^{-a_*}) = V - c(n+1) = \tilde{B}_n, \end{aligned}$$

where we have used the identity $a_* = -\ln c$.

This completes the proof of (1.18) and finishes the analysis of the optimal stopping problem.

1.3. Markovian approach

Throughout this section, we assume that $X = (X_0, X_1, X_2, \ldots)$ is a Markov family defined on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, (P_x)_{x\in E})$, taking values in some topological space $(E, \mathcal{B}(E))$. For the sake of simplicity, we will assume that $E = \mathbb{R}^d$ for some $d \geq 1$, though the reasoning remains essentially the same for other topological spaces. As usual, we assume that for each $x \in E$, we have $X_0 \equiv x$ $\mathbb{P}_x$-almost surely. Let us also introduce the transition operator T of X, which acts by the formula

$$Tf(x) = \mathbb{E}_x f(X_1) \qquad \text{for } x \in E,$$

on the class I of all measurable functions $f : E \to \mathbb{R}$ such that $f(X_1)$ is $\mathbb{P}_x$-integrable for all $x \in E$.

Suppose that N is a nonnegative integer and let $G : E \to \mathbb{R}$ be a measurable function satisfying

$$\mathbb{E}_x\left(\sup_{0\leq n\leq N} |G(X_n)|\right) < \infty \qquad \text{for all } x \in E. \tag{1.19}$$

Consider the associated finite-horizon optimal stopping problem

$$V^N(x) = \sup \mathbb{E}_x G(X_\tau), \tag{1.20}$$

where $x \in E$ and the supremum is taken over all $\tau \in \mathcal{M}^N$. Obviously, if we define $G_n = G(X_n)$ for $n = 0, 1, 2, \ldots$, then for each separate x this problem is of the form considered in the preceding sections (with $\mathbb{P}$ and $\mathbb{E}$ replaced by $\mathbb{P}_x$ and $\mathbb{E}_x$). However, the joint study of the whole family of optimal stopping problems depending on the initial value x enables the exploitation of the additional Markovian structure of the sequence $(X_n)_{n\geq 0}$.

For a given x, let us consider the random variables B_n^N and the stopping times τ_n^N, $n = 0, 1, 2, \ldots, N$, defined by (1.7) and (1.8). We also introduce the sets

$$\begin{aligned} C_n &= \{x \in E : V^{N-n}(x) > G(x)\}, \\ D_n &= \{x \in E : V^{N-n}(x) = G(x)\}, \end{aligned}$$

for $n = 0, 1, 2, \dots, N$; we will call these the continuation and stopping regions, respectively. Finally, define the stopping time

$$\tau_D = \inf\{0 \le n \le N \,:\, X_n \in D_n\}.$$

Since $V^0 = G$, by the very definition (1.20), we see that $X_N \in D_N$ and hence the stopping time τ_D is finite (it does not exceed N).

THEOREM 1.5. *Assume that the function G satisfies the integrability condition* (1.19) *and consider the optimal stopping problem* (1.20).

(i) For any $n = 0, 1, 2, \dots, N$ we have $B_n^N = V^{N-n}(X_n)$.

(ii) The function $x \mapsto V^n(x)$ satisfies the Wald-Bellman equation

$$V^n(x) = \max\{G(x), TV^{n-1}(x)\}, \qquad x \in E, \tag{1.21}$$

for $n = 1, 2, \dots, N$.

(iii) The stopping time τ_D is optimal in (1.20). *If τ_* is another optimal stopping time, then $\tau_D \le \tau_*$ $\mathbb{P}_x$-almost surely for all $x \in E$.*

(iv) For each $x \in E$, the sequence $(V^{N-n}(X_n))_{n=0}^N$ is the smallest $\mathbb{P}_x$-supermartingale majorizing $(G(X_n))_{n=0}^N$, and the stopped sequence $\big(V^{N-n\wedge\tau_D}(X_{n\wedge\tau_D})\big)_{n=0}^N$ is a $\mathbb{P}_x$-martingale.

PROOF. We only need to establish (i) and (ii); the remaining parts follow at once from Theorem 1.1. To verify (i), recall that

$$B_n^N = \mathbb{E}_x\left[G(X_{\tau_n^N})|\mathcal{F}_n\right]$$

for all $n = 0, 1, 2, \dots, N$. This shows the claim for $n = 0$, by the very definition of $V^N(x)$. On the other hand, for $n \ge 1$ we apply the Markov property to get

$$B_n^N = \mathbb{E}_y\left[G(X_{\tau_0^{N-n}})\right]\Big|_{y=X_n} = V^{N-n}(y)\big|_{y=X_n} = V^{N-n}(X_n).$$

(ii) We apply the definition of the sequence $(B_n^N)_{n=0}^N$ and part (i) to obtain that $\mathbb{P}_x$-almost surely,

$$\begin{aligned} V^N(x) = V^N(X_0) = B_0^N &= \max\{G(X_0), \mathbb{E}_x(B_1^N|\mathcal{F}_0)\} \\ &= \max\{G(x), \mathbb{E}_x(V^{N-1}(X_1)|\mathcal{F}_0)\} \\ &= \max\{G(x), TV^{N-1}(x)\}. \end{aligned}$$

□

Part (ii) above gives the following iterative method of solving (1.20). Define the operator Q acting on $f \in I$ by the formula

$$Qf(x) = \max\{G(x), Tf(x)\}, \qquad x \in E.$$

COROLLARY 1.1. *We have $V^N(x) = Q^N G(x)$ for all $x \in E$ and all integers N.*

Let us illustrate the above considerations by analyzing the following simple example.

EXAMPLE 1.2. Let $(S_n)_{n\ge0}$ be a symmetric random walk over the space $E = \{-2, -1, 0, 1, 2\}$ stopped at $\{-2, 2\}$. Clearly, $(S_n)_{n\ge0}$ is a Markov family on E. Set $G(x) = x^2(x+2)$ and consider the optimal stopping problem

$$V^N(x) = \sup_{\tau\le N} \mathbb{E}_x G(S_\tau), \qquad x \in E.$$

To treat the problem successfully, we compute the sequence V^0, V^1, V^2, V^3, ... Directly from (1.21), we have

$$V^n(x) = \max\{G(x), TV^{n-1}(x)\}$$
$$= \begin{cases} \max\{G(x), V^{n-1}(x)\} & \text{if } x \in \{-2, 2\}, \\ \max\left\{G(x), \frac{1}{2}\big(V^{n-1}(x-1) + V^{n-1}(x+1)\big)\right\} & \text{if } x \in \{-1, 0, 1\}. \end{cases}$$

For notational simplicity, let us identify a function $f : E \to \mathbb{R}$ with the sequence of its values $f(-2)$, $f(-1)$, $f(0)$, $f(1)$, $f(2)$. Using the above recurrence, we compute that

$$\begin{array}{rccccc} V^0 = G: & 0, & 1, & 0, & 3, & 16, \\ V^1: & 0, & 1, & 2, & 8, & 16, \\ V^2: & 0, & 1, & 4\frac{1}{2}, & 9, & 16, \\ V^3: & 0, & 2\frac{1}{4}, & 5 & 10\frac{1}{4}, & 16, \\ V^4: & 0, & 2\frac{1}{2}, & 6\frac{1}{4}, & 10\frac{1}{2}, & 16, \\ \dots & & & & & \end{array}$$

and so on. Suppose that we want to solve the problem

$$V^4(x) = \sup_{\tau \le 4} \mathbb{E}_x G(S_\tau), \qquad x \in E.$$

The value function V^4 has been derived above; to describe the optimal stopping strategy, let us write down the continuation and stopping regions C_i and D_i, $i = 0, 1, 2, 3, 4$. Directly from the above formulas for V^i, we see that

$$\begin{aligned} C_0 &= \{-1, 0, 1\}, & D_0 &= \{-2, 2\}, \\ C_1 &= \{-1, 0, 1\}, & D_1 &= \{-2, 2\}, \\ C_2 &= \{0, 1\}, & D_2 &= \{-2, -1, 2\}, \\ C_3 &= \{0, 1\}, & D_3 &= \{-2, -1, 2\}, \\ C_4 &= \emptyset & D_4 &= \{-2, -1, 0, 1, 2\}. \end{aligned}$$

The optimal strategy is to wait for the first step n at which we visit the corresponding stopping set D_n; then we stop the process ultimately.

We turn our attention to the case of infinite horizon, i.e., we consider the optimal stopping problem (or rather a family of optimal stopping problems)

$$V(x) = \sup \mathbb{E}_x G(X_\tau), \qquad x \in E, \tag{1.22}$$

where the supremum is taken over the class $\mathcal{M}$ of all adapted stopping times. Recall that the class I consists of all measurable functions $f : E \to \mathbb{R}$ such that $f(X_1)$ is $\mathbb{P}_x$-integrable for all $x \in E$. The following notion will be crucial in our further considerations.

DEFINITION 1.2. The function $f \in I$ is called *superharmonic* (or *excessive*) if we have

$$Tf(x) \le f(x) \qquad \text{for all } x \in E.$$

We have the following simple observation.

LEMMA 1.3. *The function $f \in I$ is superharmonic if and only if $(f(X_n))_{n \ge 0}$ is a supermartingale under each $\mathbb{P}_x$, $x \in E$.*

PROOF. If f is superharmonic, then by Markov property,

$$\mathbb{E}_x(f(X_{n+1})|\mathcal{F}_n) = \mathbb{E}_y f(X_1)\big|_{y=X_n} = Tf(X_n) \leq f(X_n),$$

for each n. To show the reverse implication, observe that if $(f(X_n))_{n\geq 0}$ is a supermartingale under each $\mathbb{P}_x$, then in particular

$$Tf(x) = \mathbb{E}_x(f(X_1)|\mathcal{F}_0) \leq f(x). \quad \square$$

To formulate the main theorem, we introduce the corresponding continuation set C and stopping set D by

$$C = \{x \in E \,:\, V(x) > G(x)\},$$
$$D = \{x \in E \,:\, V(x) = G(x)\}.$$

Moreover, we define the stopping time $\tau_D = \inf\{n \,:\, X_n \in D\}$. In contrast to the case of finite horizon, this stopping time need not be finite (which will force us to impose some additional assumptions: see the statement below).

THEOREM 1.6. *Consider the optimal stopping problem* (1.22) *and assume that*

$$\mathbb{E}_x \sup_{n\geq 0} |G(X_n)| < \infty, \qquad x \in E. \tag{1.23}$$

Then the following holds.

(i) The function V satisfies the Wald-Bellman equation

$$V(x) = \max\{G(x), TV(x)\}. \tag{1.24}$$

(ii) If τ_D is finite $\mathbb{P}_x$-almost surely for all $x \in E$, then τ_D is the optimal stopping time. If τ_* is another optimal stopping time, then $\tau_* \geq \tau_D$ $\mathbb{P}_x$-almost surely.

(iii) The value function V is the smallest superharmonic function which majorizes the gain function G on E.

(iv) The stopped sequence $(V(X_{\tau_D \wedge n}))_{n\geq 0}$ is a $\mathbb{P}_x$-martingale for every $x \in E$.

PROOF. This follows immediately from the case of finite horizon and the limit Theorem 1.3. $\square$

Let us make here an important comment on the uniqueness of the solutions to the Wald-Bellman equations (1.21) and (1.24). Clearly, in the case of finite horizon there is only one solution: indeed, the starting function V^0 coincides with G and the formula (1.21) produces a unique sequence $V^1, V^2, \ldots, V^N$. In the case of infinite horizon, the situation is less transparent. For instance, if G is a constant function, say, $G \equiv c$, then any constant function $V \equiv c'$ for some $c' \geq c$ satisfies the Wald-Bellman equation. However, any solution to (1.24) is a superharmonic function majorizing G, so part (iii) of Theorem 1.6 immediately yields the following "minimality principle".

COROLLARY 1.2. *The value function V is the minimal solution to* (1.24).

EXAMPLE 1.3. Let us provide solution to the infinite-horizon version of Example 1.2. Under the notation used there, we study the optimal stopping problem

$$V(x) = \sup_{\tau \in \mathcal{M}} \mathbb{E}G(S_\tau), \qquad x \in E.$$

The function G is bounded, so the integrability assumption of Theorem 1.6 is satisfied. Thus, we know that V is the least superharmonic function which majorizes G: here the superharmonicity means that

$$V(x) \geq \frac{1}{2}\big(V(x-1)+V(x+1)\big), \qquad \text{for } x \in \{-1,0,1\}.$$

In other words, we search for the smallest concave function on $\{-2,-1,0,1,2\}$ majorizing the function G. One easily checks that the function $x \mapsto 4(x+2)$ is concave (since it is linear), majorizes G and coincides with G at the endpoints ± 2. Thus it is the smallest majorant of G and hence it must be equal to the value function V.

EXAMPLE 1.4. Let ξ_1, ξ_2, ... be a sequence of i.i.d. random variables with the distribution given by $\mathbb{P}(\xi_i = 1) = p$, $\mathbb{P}(\xi_i = -1) = q$, where $p + q = 1$ and $p < q$. For a given integer x, define $S_n = x + \xi_1 + \xi_2 + \ldots + \xi_n$, $n = 0, 1, 2, \ldots$. Then the sequences $(S_n)_{n\geq 0}$ (with varying x) form a Markov family. Consider the optimal stopping problem

$$V(x) = \sup_{\tau\in\mathcal{M}} \mathbb{E}_x S_\tau^+, \qquad x \in E.$$

One easily checks the integrability assumption (1.23) (with $G(x) = x^+$) is satisfied. This follows from the well-known fact that

$$\mathbb{P}\Big(\sup_{n\geq 0}(\xi_1+\xi_2+\ldots+\xi_n) \geq k\Big) = \left(\frac{p}{q}\right)^k, \qquad k = 0,\, 1,\, 2,\, \ldots. \tag{1.25}$$

Thus, we need to find the least superharmonic majorant of G: V is the least function on $\mathbb{Z}$ satisfying

$$V(x) = \max\left\{x^+, pV(x+1) + qV(x-1)\right\}, \qquad x \in \mathbb{Z}.$$

To identify this object, let us try to inspect the properties of the continuation set C and the stopping region D. A little thought suggests that these sets should be of the form $C = \{\ldots, b-2, b-1\}$, $D = \{b, b+1, \ldots\}$ for some positive integer b (possibly infinite). While this is more or less clear by some intuitive argumentation, we should point out here that this can also be shown rigorously. Indeed, pick $x \in \mathbb{Z}_-$ and take the stopping time $\tau \equiv -x+1$. Then $V(x) \geq \mathbb{E}_x S_\tau^+ = p^{-x+1} > 0 = G(x)$, so in particular C contains all nonpositive integers. Furthermore, if $x > 0$ lies in C, then so does $x - 1$. To see this, note that for any $a \in \mathbb{Z}$ we have

$$(x+a)^+ - x^+ \leq (x-1+a)^+ - (x-1)^+$$

(which is equivalent to the trivial bound $(x+a)^+ \leq (x-1+a)^+ + 1$) and hence for any stopping time τ, if we plug $a = \xi_1 + \xi_2 + \ldots + \xi_\tau$,

$$\mathbb{E}_x S_\tau^+ - G(x) \leq \mathbb{E}_{x-1} S_\tau^+ - G(x-1).$$

This yields

$$0 < V(x) - G(x) \leq V(x-1) - G(x-1) \tag{1.26}$$

and thus $x - 1 \in C$, as we have claimed. This shows that C and D are of the form postulated above and hence, by the general theory,

$$V(x) = \begin{cases} x & \text{if } x \geq b, \\ pV(x+1) + qV(x-1) & \text{if } x < b. \end{cases}$$

Let us first identify V on C. Solving the linear recurrence, we check that

$$V(x) = \alpha + \beta \left(\frac{q}{p}\right)^x, \qquad x < b,$$

for some constants α, $\beta \in \mathbb{R}$. It follows from (1.25) that $V(x) \to 0$ as $x \to -\infty$ (simply use the estimate $\mathbb{E}_x S_\tau^+ \leq \mathbb{E}_x \sup_{n\geq 0} S_n^+$): this implies $\alpha = 0$ and $\beta \geq 0$. This also shows that $b < \infty$. Indeed, otherwise $V(x)$ would explode exponentially as $x \to \infty$, but on the other hand, by (1.26), for $x > 0$ we would have

$$V(x) \leq G(x) + V(0) - G(0) = x + V(0) - G(0).$$

It remains to find β and the boundary b. First, exploiting the Wald-Bellman equation, we see that $V(b-1) = pV(b) + qV(b-2)$. This implies $V(b) = \beta(q/p)^b$ and hence

$$V(x) = b\left(\frac{q}{p}\right)^{x-b} \qquad \text{for } x \leq b.$$

Secondly, again by Wald-Bellman equation, we see that $V(b) \geq pV(b+1)+qV(b-1)$, which is equivalent to $b \geq p/(q-p)$. Finally, observe that if $x > b$, then

$$V(x) = x = px + qx > p(x+1) + q(x-1) = pV(x+1) + qV(x-1).$$

Therefore, if b satisfies the inequality $b \geq p/(q-p)$, then the function

$$\mathcal{V}(x) = \begin{cases} x & \text{if } x \geq b, \\ b\,(q/p)^{x-b} & \text{if } x < b. \end{cases}$$

is excessive. Let us now check for which b the inequality $\mathcal{V} \geq G$ holds. This majorization is clear on $\{b, b+1, b+2, \ldots\}$. Since the function $x \mapsto (q/p)^{x-b}$ is nonnegative, convex and coincides with G at $x = b$, it suffices to check whether it is bigger than G at $x = b-1$. The latter bound is equivalent to $b < q/(q-p) = p/(q-p) + 1$. This actually *forces* us to take $b = \lceil p/(q-p)\rceil$: this is the only choice for the parameter such that the resulting function $\mathcal{V}$ is superharmonic and majorizes G. Summarizing, we have shown that

$$V(x) = \begin{cases} x & \text{if } x \geq \lceil p/(q-p)\rceil, \\ \lceil p/(q-p)\rceil\,(q/p)^{x-\lceil p/(q-p)\rceil} & \text{if } x < \lceil p/(q-p)\rceil. \end{cases}$$

Observe that by (1.25), the stopping time

$$\tau = \inf\Big\{n\,:\, S_n \geq \lceil p/(q-p)\rceil\Big\}$$

is infinite with positive probability. Therefore, there is no optimal stopping time τ^* which would be finite $\mathbb{P}_x$-almost surely for all x. Hence, the value function is attained asymptotically at the stopping times

$$\tau^{(M)} = \inf\Big\{n : S_n \notin \big[M, \lceil p/(q-p)\rceil\big]\Big\},$$

as $M \to -\infty$.

We proceed to the analysis of further examples, which will be useful in our later considerations.

EXAMPLE 1.5. Let X_1, X_2, ... be a sequence of independent and identically distributed random variables satisfying $\mathbb{P}(X_i = -1) = \frac{2}{3} = 1 - \mathbb{P}(X_i = 2)$, and set $S_0 = 0$, $S_n = X_1 + X_2 + \ldots + X_n$ for $n = 1, 2, \ldots$. We will solve the optimal stopping problem

$$V = \sup_{\tau \in L^1} \mathbb{E}\left(|S_\tau| - \frac{1}{2}\tau\right). \tag{1.27}$$

It is convenient to split the analysis into several separate steps.

Step 1. Dimension reduction. At the first glance, the problem is two-dimensional, i.e., it involves the stopping of the two-dimensional process (S_n, n). It is possible to reduce the dimension to one, by the following simple observation. Namely, note that the processes $(S_n)_{n\geq 0}$ and $(S_n^2 - 2n)_{n\geq 0}$ are martingales. This is evident for the first process, to check the second we compute that

$$\begin{aligned}\mathbb{E}\left[S_{n+1}^2 - 2(n+1)\Big|\mathcal{F}_n\right] &= S_n^2 - 2n + \mathbb{E}(2S_nX_{n+1} + X_{n+1}^2 - 2|\mathcal{F}_n)\\ &= S_n^2 - 2n + \mathbb{E}(X_{n+1}^2 - 2) = S_n^2 - 2n,\end{aligned}$$

since $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}X_{n+1} = 0$ and $\mathbb{E}X_{n+1}^2 = 2$. Consequently, by Doob's optional sampling theorem, for any $\tau \in L^1$ we have

$$\mathbb{E}S_{\tau\wedge n}^2 = 2\mathbb{E}(\tau \wedge n). \tag{1.28}$$

The right-hand side is uniformly bounded in n and converges to $\mathbb{E}\tau$, by Lebesgue's monotone convergence theorem. Therefore, the martingale $(S_{\tau\wedge n})_{n\geq 0}$ is bounded in L^2, and hence also L^2-convergent. Thus, letting $n \to \infty$ in (1.28) gives $\mathbb{E}S_\tau^2 = 2\mathbb{E}\tau$ and hence we may rewrite (1.27) as

$$V = \sup_{\tau \in L^1} \mathbb{E}\left(|S_\tau| - \frac{S_\tau^2}{4}\right).$$

Now the right-hand side depends on the process S only.

Step 2. General theory. To put the above problem into the general framework, we set $G(x) = |x| - x^2/4$ and note that the problem reads

$$V = \sup_{\tau \in L^1} \mathbb{E}G(S_\tau).$$

The process S extends to a Markov family on $E = \mathbb{Z}$, with the transity function determined by $p_{n,n-1} = 2/3$ and $p_{n,n+2} = 1/3$ for all $n \in \mathbb{Z}$. Let $V : \mathbb{Z} \to \mathbb{R}$ be given by

$$V(x) = \sup_{\tau \in L^1} \mathbb{E}_x G(S_\tau).$$

In order to apply the general theory, we need to check the condition $\mathbb{E}\sup_{n\geq 0} G(S_n) < \infty$. This is a little technical, so we take the opportunity to present a different approach. Namely, we will construct the least excessive majorant of the function G, and then exploit its properties to solve rigorously the problem under consideration.

Step 3. On the search of the excessive majorants. Let U be the least excessive majorant of G. In our setting, excessiveness amounts to

$$U(n) \geq \frac{2}{3}U(n-1) + \frac{1}{3}U(n+2) \qquad \text{for all } n \in \mathbb{Z}.$$

So, if a function is concave on $\mathbb{Z}$, then it is automatically excessive; as we shall see during the analysis, the reverse implication does not hold in general (which might be a little surprising at the first glance). A quick look at the graph of G shows that the function

$$H(x) = \begin{cases} G(x) & \text{if } |x| \geq 2, \\ 1 & \text{if } |x| \leq 2 \end{cases}$$

is concave and majorizes G. Thus, we have $H \geq U$ on $\mathbb{Z}$. On the other hand, we have $U(x) \geq G(x) = H(x)$ for $|x| \geq 2$: this shows that $U(x) = G(x)$ for $|x| \geq 2$. Next, observe that

$$\begin{aligned} U(1) &\geq \frac{2}{3}U(0) + \frac{1}{3}U(3) = \frac{2}{3}U(0) + \frac{1}{3} \cdot \frac{3}{4}, \\ U(0) &\geq \frac{2}{3}U(-1) + \frac{1}{3}U(2) = \frac{2}{3}U(-1) + \frac{1}{3}, \\ U(-1) &\geq \frac{2}{3}U(-2) + \frac{1}{3}U(1) = \frac{2}{3} + \frac{1}{3}U(1). \end{aligned}$$

Combining these estimates, we get

$$\begin{aligned} U(1) &\geq \frac{2}{3}U(0) + \frac{1}{4} \geq \frac{2}{3}\left(\frac{2}{3}U(-1) + \frac{1}{3}\right) + \frac{1}{4} \\ &\geq \frac{4}{9}\left(\frac{2}{3} + \frac{1}{3}U(1)\right) + \frac{17}{36} = \frac{4}{27}U(1) + \frac{83}{108}, \end{aligned}$$

or $U(1) \geq 83/92$. Plugging this above, we get $U(-1) \geq 267/276$ and $U(0) \geq 45/46$. Assuming equalities, we obtain the excessive function U; furthermore, one verifies directly that such U majorizes G at -1, 0, 1. Hence U is an excessive majorant of G.

Step 4. Coming back to the optimal stopping problem. Suppose that τ is an arbitrary and integrable stopping time. The function U constructed above is excessive, so the process $(U(S_n))_{n\geq 0}$ is a $\mathbb{P}_0$-supermartingale. Furthermore, since $U \geq G$, Doob's optional sampling theorem gives

$$\mathbb{E}G(S_{\tau\wedge n}) \leq \mathbb{E}U(S_{\tau\wedge n}) = \mathbb{E}U(0) = \frac{45}{46},$$

that is,

$$\mathbb{E}\left(|S_{\tau\wedge n}| - \frac{S_{\tau\wedge n}^2}{4}\right) \leq \frac{45}{46}.$$

However, as we have proved in Step 1 above, the process $(S_{\tau\wedge n})_{n\geq 0}$ is an L^2-bounded supermartingale. Thus, letting $n \to \infty$ yields

$$\mathbb{E}\left(|S_\tau| - \frac{1}{2}\tau\right) \leq \frac{45}{46}.$$

Directly from the analysis in Step 3, the equality is attained for the stopping time $\tau = \inf\{n : |S_n| \geq 2\}$.

The next example is more "continuous" in nature.

EXAMPLE 1.6. Suppose that $\xi_1, \xi_2, \ldots$ and $\varepsilon_1, \varepsilon_2, \ldots$ are independent random variables, such that ξ_n has exponential law with parameter $\lambda > 0$ and $\mathbb{P}(\varepsilon_n = 1) = p = 1 - \mathbb{P}(\varepsilon_n = 0)$, $n = 1, 2, \ldots$. For a given parameter $x > 0$, we define the sequence $(X_n)_{n\geq 0}$ by $X_0 \equiv x$ and $X_{n+1} = \varepsilon_{n+1}(X_n + \xi_{n+1})$, $n = 0, 1, 2, \ldots$. Finally, we fix $c > 0$ and consider the optimal stopping problem

$$V = \sup_\tau \mathbb{E}(X_\tau - c\tau), \tag{1.29}$$

where the supremum is taken over all integrable stopping times τ.

Step 1. We start by putting the problem into the general framework developed above. It is easy to see that $(X_n)_{n\geq 0}$ is a time-homogeneous Markov process with the transition function given by

$$P(x,0)=1-p \qquad \text{and} \qquad P(x,x+A)=p\int_A \lambda e^{-\lambda a}\mathrm{d}a \quad \text{for } A\subseteq[0,\infty).$$

Actually, for the problem (1.29) we need to consider the space-time version (n,X_n), which is also a Markov process (this time on the state space $\mathbb{N}\times[0,\infty)$). We extend it to the Markov family and, as usual, denote the corresponding initial probabilities by $\mathbb{P}_{n,x}$. For any (n,x), we consider the auxiliary optimal stopping problem

$$V(n,x)=\sup_\tau \mathbb{E}_{n,x}G(\tau,X_\tau),$$

where the supremum is taken over all finite stopping times τ and $G(n,x)=x-cn$.

Step 2. Now we will reduce the dimension of the problem, by observing a certain homogeneity-type condition on V. Namely, the identity $G(n,x)=G(0,x)-cn$ immediately gives

$$V(n,x)=\sup_\tau \mathbb{E}_{n,x}G(\tau,X_\tau)=\sup_\tau \mathbb{E}_{0,x}G(\tau,X_\tau)-cn=V(0,x)-cn$$

and hence it is enough to identify the function $f(x):=V(0,x)$.

Step 3. We introduce the continuation and the stopping sets C and D by

$$C=\{(n,x):V(n,x)>G(n,x)\},\qquad D=\{(n,x):V(n,x)=G(n,x)\}.$$

Here is some initial analysis of the stopping domain D. Namely, we will show that if $(0,x)\in D$ and $x'>x$, then necessarily $(0,x')\in D$. To this end, note that for any stopping time τ, we have

$$\mathbb{E}_{0,x'}(X_\tau-c\tau)-\mathbb{E}_{0,x}(X_\tau-c\tau)\leq x'-x.$$

Indeed, one can couple the trajectories of X under $\mathbb{P}_{(0,x)}$ and $\mathbb{P}_{(0,x')}$ in such a way that their difference is equal to $x-x'$, until they both drop to zero and coincide from that time. Since $x'-x=G(0,x')-G(0,x)$, we obtain

$$V(0,x')-G(0,x')\leq V(0,x)-G(0,x)=0,$$

which gives $(0,x')\in D$. By the homogeneity established in Step 2, this gives the aforementioned property of the set D. Actually, we see that D must be of the form

$$D=\{(n,x):x\geq b\},$$

for some unknown parameter b (to be found).

Step 4. It follows from the general theory that V is the least excessive majorant of the function G. Here the excessiveness means

$$V(n,x)\geq(1-p)V(n+1,0)+p\int_0^\infty V(n+1,x+a)\cdot\lambda e^{-\lambda a}\mathrm{d}a,$$

or, by the homogeneity proved in Step 2 above,

$$f(x)-cn\geq(1-p)(f(0)-c(n+1))+p\int_0^\infty (f(x+a)-c(n+1))\cdot\lambda e^{-\lambda a}\mathrm{d}a.$$

This is equivalent to the inequality

$$f(x)+c\geq(1-p)f(0)+p\int_0^\infty f(x+a)\cdot\lambda e^{-\lambda a}\mathrm{d}a. \tag{1.30}$$

This allows us to write the corresponding system of requirements: the so-called boundary value problem for V (or rather, for f). Namely,

$$f(x) = x \qquad \text{for } x \geq b, \tag{1.31}$$

$$f(x) + c = (1-p)f(0) + p\int_0^\infty f(x+a)\cdot \lambda e^{-\lambda a}\mathrm{d}a \qquad \text{for } x < b. \tag{1.32}$$

Step 5. Let us try to find the candidate for the solution to the above system. A direct differentiation of (1.32) yields

$$f'(x) = p\int_0^\infty f'(x+a)\cdot \lambda e^{-\lambda a}\mathrm{d}a = -p\lambda f(x) + p\lambda \int_0^\infty f(x+a)\cdot\lambda e^{-\lambda a}\mathrm{d}a,$$

where the last passage follows from the integration by parts. We apply (1.32) again, obtaining

$$f'(x) = -p\lambda f(x) + \lambda(f(x) + c - (1-p)f(0)) = \lambda(1-p)\Big[f(x) + \frac{c}{1-p} - f(0)\Big].$$

This can be solved directly: we get

$$f(x) = f(0) - \frac{c}{1-p} + \alpha e^{\lambda(1-p)x} \qquad x < b,$$

for some parameter α. Plugging $x = 0$, we get $\alpha = c/(1-p)$ and hence

$$f(x) = \begin{cases} f(0) + \frac{c}{1-p}\left(e^{\lambda(1-p)x} - 1\right) & \text{if } x < b,\\ x & \text{if } x \geq b.\end{cases}$$

It is plausible to conjecture that f is continuous at b: this implies

$$f(0) = b - \frac{c}{1-p}\left(e^{\lambda(1-p)b} - 1\right)$$

and we finally obtain

$$f(x) = \begin{cases} b + \frac{c}{1-p}\left(e^{\lambda(1-p)x} - e^{\lambda(1-p)b}\right) & \text{if } x < b,\\ x & \text{if } x \geq b.\end{cases}$$

Step 6. It remains to find the boundary point b. To accomplish this, we verify the excessiveness inequality (1.30) on $[b,\infty)$. Since $f(x) = x$ on this interval, the inequality reads

$$x + c \geq (1-p)\left(b - \frac{c}{1-p}\left(e^{\lambda(1-p)b} - 1\right)\right) + p\left(x + \frac{1}{\lambda}\right).$$

This requirement is most restrictive for smallest x, i.e., for $x = b$. For this particular choice, the estimate becomes

$$\exp(\lambda(1-p)b) \geq \frac{p}{\lambda c}. \tag{1.33}$$

If $p/(\lambda c) \leq 1$, then this estimate holds for all $b \geq 0$. In other words, the identity function $f(x) = x$ leads to the excessive function $V(n,x) = x - cn$. In this case, we have $V = G$ and the optimal stopping rule is to stop instantaneously.

If $p/(\lambda c) > 1$, then assuming equality in (1.33), we obtain

$$b = \frac{1}{\lambda(1-p)}\ln\frac{p}{\lambda c}$$

and

$$f(x) = \begin{cases} \frac{1}{\lambda(1-p)} \ln \frac{p}{\lambda c} + \frac{c}{1-p} e^{\lambda(1-p)x} - \frac{p}{1-p} & \text{if } x < b, \\ x & \text{if } x \geq b. \end{cases}$$

Step 7. We need to emphasize that the analysis in Step 6 was informal: we obtained the *candidate* for V (or rather, for f) under a number of additional assumptions. It remains to check that V satisfies all the necessary requirements. Namely, it actually follows from the above analysis that V is excessive. The majorization $V \geq G$ is equivalent to the estimate $f(x) \geq x$ on $[0, b]$; since both sides are equal for $x = b$ and f is strictly convex on $[0, b]$, it is enough to verify that $f'(b-) \leq 1$. This is equivalent to the estimate

$$c\lambda \cdot \frac{p}{\lambda c} \leq 1,$$

which holds trivially. Finally, let us show that V is the smallest excessive majorant. Assume conversely that this is not the case: the optimal function $\tilde{V}$ satisfies $\tilde{V}(n, x) < V(n, x)$ for some (n, x). Then $\tilde{V}$ satisfies the identity $\tilde{V}(n, y) = \tilde{f}(y) - cn$ for some function $\tilde{f}$ on $[0, \infty)$, and we have $\tilde{f}(x) < f(x)$. We have $\tilde{f}(y) \geq y$ for all y, and hence we must necessarily have $x < b$ By the excessiveness condition (1.30), we get

$$\tilde{f}(x) - f(x) \geq p \int_0^\infty (\tilde{f}(x+a) - f(x+a)) \cdot \lambda e^{-\lambda a} \mathrm{d}a.$$

Now suppose that $x_0 = \sup\{t : \tilde{f}(t) - f(t) \leq \tilde{f}(x) - f(x)\}$: clearly x_0 is finite, since $\tilde{f}$ and f coincide on $[b, \infty)$. Then $\tilde{f}(u) - f(u) > \tilde{f}(x) - f(x)$ on (x_0, ∞) and hence we obtain

$$\tilde{f}(x) - f(x) \geq p \int_0^\infty (\tilde{f}(x_0+a) - f(x_0+a)) \cdot \lambda e^{-\lambda a} \mathrm{d}a \geq p(\tilde{f}(x) - f(x)),$$

which implies $\tilde{f}(x) > f(x)$, a contradiction.

We conclude the analysis with the comment that the optimal stopping rule is given by $\tau_* = \inf\{n : X_n \geq b\}$.

1.4. Problems

1. Let G_1, G_2, ... be a sequence of independent random variables, each of which has the uniform distribution on $[0, 1]$. Solve the optimal stopping problems

$$V^N = \sup_{\tau \in \mathcal{M}^N} \mathbb{E} G_\tau \qquad \text{and} \qquad V_0 = \sup_{\tau \in \mathcal{M}} \mathbb{E} G_\tau,$$

where N is an arbitrary integer.

2. We flip a coin at most five times, at each point we may decide whether to stop or not (in particular, we are allowed to stop at the very beginning, without flipping the coin even once). Having stopped, we look at the outcomes we have obtained. We get 1 if there are no heads and get 2 if we obtained at least three heads. What is the strategy which yields the largest expected gain?

3. The Secretary Problem.

4. (Robbins [**19**]) We flip a coin infinitely many times. For $n \geq 1$, let $G_n = n2^n/(n+1)$ of there were no tails in the first n flips, and $G_n = 0$ otherwise. Solve the optimal stopping problem

$$V = \sup_{\tau \in \mathcal{M}} \mathbb{E} G_\tau.$$

5. Let $(S_n)_{n\geq 0}$ be a symmetric random walk over the integers and let $G(x) = \arctan x - x^-$, $x \in \mathbb{Z}$. Solve the optimal stopping problem

$$V(x) = \sup \mathbb{E}_x G(S_\tau), \qquad x \in \mathbb{Z},$$

where the supremum is taken over (i) all stopping times τ, (ii) bounded stopping times τ.

6. We toss a fair coin and for each $n \geq 0$, we denote by X_n the length of the current sequence of consecutive tails after n flips:

$$\overbrace{\ldots H \underbrace{TT\ldots T}_{X_n}}^{n \text{ flips}}.$$

Solve the optimal stopping problem

$$V = \sup_{\tau \in \mathcal{M}} \mathbb{E}\left(X_\tau - \frac{1}{16}\tau \right).$$

7. (Krengel, Sucheston and Garling [**12, 13**]) Let G_1, G_2, ..., be a family of independent, nonnegative random variables. Show that

$$\mathbb{E} \sup_{n\geq 1} G_n \leq 2 \sup_{\tau \in \mathcal{M}} \mathbb{E} G_\tau.$$

8. Suppose that $(G_n)_{n\geq 0}$ is an adapted sequence of random variables satisfying $\mathbb{E}\sup_{n\geq 0}|G_n| < \infty$ and let $(B_n)_{n\geq 0}$ be the associated Snell envelope. Prove the identity

$$\sup_{\tau \in \mathcal{M}} \mathbb{E} G_\tau = \sup_{\tau \in \mathcal{M}} \mathbb{E} B_\tau.$$

9. (Hill and Kertz [**11**]) Let G_0, G_1, G_2, ... be a family of arbitrarily dependent random variables taking values in $[0, 1]$. Prove the inequality

$$\mathbb{E} \sup_{n\geq 0} G_n \leq \sup_{\tau \in \mathcal{M}} \mathbb{E} G_\tau + \frac{1}{e}$$

and show that the constant $1/e$ cannot be decreased.

10. Let ε_1, ε_2, ... be the sequence of independent Rademacher variables and set $X_0 \equiv 0$ and $X_n = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$ for $n = 1, 2, \ldots$. Find the smallest constant C such that for any stopping time τ adapted to the natural filtration of X we have

$$\mathbb{E} X_\tau^4 \leq C \mathbb{E} \tau^2.$$

11. Let ε_1, ε_2, ... be the sequence of independent Rademacher variables and set $X_0 \equiv 0$ and $X_n = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$ for $n = 1, 2, \ldots$. For any $1 < p < \infty$, find the smallest constant C_p such that for any stopping time $\tau \in L^{p/2}$ adapted to the natural filtration of X we have

$$\mathbb{E} \sup_{n\leq \tau} |X_n|^p \leq C_p \mathbb{E}|X_\tau|^p.$$

12. Let $(S_n)_{n\geq 0}$ be a symmetric random walk over the integers and let $\beta \in (0, 1)$ be a fixed parameter. Solve the optimal stopping problem

$$V(x) = \sup_{\tau \in \mathcal{M}} \mathbb{E}_x \left[\beta^\tau (1 - \exp(S_\tau))^+ \right], \qquad x \in \mathbb{Z}.$$

13. (Dubins, Shepp and Shiryaev [**9**]) Let $(S_n)_{n\geq 0}$ be a symmetric random walk over the integers, started at some point x. Prove that for any stopping time τ we have

$$\mathbb{E}\sup_{n\leq\tau} S_n \leq \sqrt{\mathbb{E}\tau}$$

and that the inequality is sharp.

14. A pawn moves over the set $\{1,2,\ldots,n\}$, according to the following rules. If at some time it is located at the point k, then at the next step it jumps, independently from its evolution in the past, to one of the points k, $k+1$, …, n (each choice has the same probability). Let X_j be the location of the pawn at the time j. Assuming that $X_0 = 1$, describe the stopping time τ which maximizes the expectation $\mathbb{E}V(X_\tau)$, where $V(x) = x1_{\{x<n\}}$.

15. Let $p, \beta \in (0,1)$ be fixed parameters and set $q = 1-p$. Let X_1, X_2, … be a sequence of independent random variables with the same distribution given by $\mathbb{P}(X_j = 1) = p = 1 - \mathbb{P}(X_j = 0)$. Let $S_0 = 0$ and $S_n = X_1 + X_2 + \ldots + X_n$ for $n \geq 1$. Solve the optimal stopping problem $\sup_\tau \mathbb{E}\beta^\tau S_\tau$.

CHAPTER 2

Continuous time

2.1. Preliminaries

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space: here by completeness we mean that for any $A \in \mathcal{F}$ of probability zero and any $B \subset A$ we have $B \in \mathcal{F}$. We equip the space with a filtration, i.e., a nondecreasing family $(\mathcal{F}_t)_{t\geq 0}$ of sub-σ-fields of $\mathcal{F}$. For technical reasons, we assume that the filtration satisfies the usual conditions, that is, it is right-continuous (we have $\bigcap_{s\geq t} \mathcal{F}_s = \mathcal{F}_t$ for all $t \geq 0$) and $\mathcal{F}_0$ contains all the events of probability zero.

A family $X = (X_t)_{t\geq 0}$ taking values in some measure space $(\mathcal{E}, \mathcal{B})$ will be called a stochastic process. The process is adapted to $(\mathcal{F}_t)_{t\geq 0}$, if for any t the random variable X_t is $\mathcal{F}_t$-measurable. Quite often it is convenient to treat X as a random variable taking values in the set of trajectories (equipped with the appropriate cylindric σ-algebra). Throughout, we will assume that X is cádlág, i.e., the trajectories of X are right-continuous and have limits from the left.

A random variable τ with values in $[0, \infty]$ is called a stopping time (relative to $(\mathcal{F}_t)_{t\geq 0}$), if for any $t \geq 0$ we have $\{\tau \leq t\} \in \mathcal{F}_t$. Any such variable gives rise to the σ-algebra

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : \{\tau \leq t\} \cap A \in \mathcal{F} \text{ for all } t \geq 0\}.$$

Furthermore, given a stochastic process $X = (X_t)_{t\geq 0}$ and a stopping time τ, the associated stopped version of X is given by $X^\tau = (X_{\tau\wedge t})_{t\geq 0}$.

Definition 2.1. An adapted process $X = (X_t)_{t\geq 0}$ is a martingale (respectively, submartingale or supermartingale), if $X_t \in L^1$ for all $t \geq 0$ and

$$\mathbb{E}(X_t|\mathcal{F}_s) = X_s \qquad \text{almost surely}$$

for all $s \leq t$ (respectively, $\geq$ or $\leq$).

Theorem 2.1 (Doob's optional sampling theorem). *Suppose that X is a cádlág martingale and τ_1, τ_2 are bounded stopping times such that $\tau_1 \leq \tau_2$. Then*

$$\mathbb{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) = X_{\tau_1}.$$

A similar fact holds for sub- and supermartingales.

Definition 2.2. An adapted process $X = (X_t)_{t\geq 0}$ is a local martingale, if there exists a nondecreasing sequence $(\tau_n)_{n\geq 0}$ of stopping times such that $\lim_{n\to\infty} \tau_n = \infty$ almost surely and for each n, the stopped process X^{τ_n} is a martingale. Analogously one introduces the concept of a local submartingale/supermartingale.

Definition 2.3. An adapted process $X = (X_t)_{t\geq 0}$ is a semimartingale, if it admits the representation

$$X = X_0 + M + A,$$

where $M = (M_t)_{t\geq 0}$ is a local martingale started at zero and $A = (A_t)_{t\geq 0}$ is a process of bounded variation (i.e., we have $\int_0^t |\mathrm{d}A_s| < \infty$ for all $t > 0$; the integral means the total variation of A on $[0,t]$) started at zero.

It should be pointed out that in general, the above decomposition of a semimartingale is not unique. However, we have the uniqueness if the processes X, M and A are assumed to have continuous trajectories.

We will also need the notion of predictability.

DEFINITION 2.4. Consider the σ-algebra $\mathcal{P}$ of subsets of $\Omega \times [0, \infty)$, generated by the sets of the form

· $A \times \{0\}$, $A \in \mathcal{F}_0$,
· $A \times (s, t]$, $A \in \mathcal{F}_s$ and $0 \leq s \leq t$.

This algebra is called predictable. A process $X = (X_t)_{t\geq 0}$ is called predictable, if it is $\mathcal{P}$-measurable (treated as a function on $\Omega \times [0, \infty)$).

One can show that X is predictable, if it is an almost sure limit of a sequence of continuous-path processes.

THEOREM 2.2 (Doob-Meyer decomposition). *Every submartingale $X = (X_t)_{t\geq 0}$ admits the unique decomposition (up to indistinguishability)*

$$X = X_0 + M + A,$$

where M is a local martingale started zt zero and A is a nondecreasing, predictable and integrable process started at zero.

DEFINITION 2.5. Let us assume that X is an L^2-bounded martingale: we have $\sup_{t\geq 0} \mathbb{E}X_t^2 < \infty$. We define $\langle X\rangle$, the skew bracket (quadratic variation) of X, as the predictable component in the Doob-Meyer decomposition $X^2 = X_0^2 + M + \langle X\rangle$. In other words, $\langle X\rangle$ is the unique nondecreasing, predictable process starting from zero such that $X^2 - \langle X\rangle$ is a local martingale.

By a straightforward localization, the above definition enables the introduction of $\langle X\rangle$ for an arbitrary local martingale X. We extend it further to the context of semimartingales: if $X = X_0 + M + A$ is a semimartingale, we let $\langle X\rangle := \langle M\rangle$. One can show that this object is well-defined (recall that the semimartingale decomposition is not unique in general). Before we proceed, let us mention that if a semimartingale X has continuous trajectories, then there is an alternative definition of $\langle X\rangle_t$, as the limit in probability of the sums

$$\sum_{k=1}^{k_n} \left(X_{t_k^{(n)}} - X_{t_{k-1}^{(n)}} \right)^2.$$

Here $(t_k^{(n)})_{k=0}^{k_n}$ is a sequence of partitions of $[0,t]$, with mesh converging to zero.

By a standard polarization, the process $\langle X\rangle$ leads to the corresponding bracket. Namely, given two semimartingales X, Y, we define

$$\langle X, Y\rangle = \frac{\langle X+Y\rangle - \langle X-Y\rangle}{4}.$$

Directly by the definition of the skew bracket, if X, Y are local martingales, then the difference $XY - \langle X, Y\rangle$ is a local martingale. This is not true if X, Y are only assumed to be semimartingales (e.g., take $X_t = Y_t = t$).

Now we introduce the process which will be fundamental in our further considerations.

DEFINITION 2.6. A real-valued process $W = (W_t)_{t\geq 0}$ is a Wiener process, if it satisfies the following conditions.

(i) We have $W_0 = 0$ almost surely.

(ii) The process W has independent increments: for any $0 \leq t_0 < t_1 < t_2 < \ldots < t_n$, the variables W_{t_0}, $W_{t_1} - W_{t_0}$, $W_{t_2} - W_{t_1}$, ..., $W_{t_n} - W_{t_{n-1}}$ are independent.

(iii) For any $0 \leq s < t$, the random variable $W_t - W_s$ has the normal distribution of mean zero and variance $t - s$.

(iv) The process W has continuous trajectories.

A Wiener process $\mathbb{W}$ in $\mathbb{R}^n$ is the collection $(W^{(1)}, W^{(2)}, \ldots, W^{(n)})$ of independent one-dimensional Wiener processes $W^{(1)}$, $W^{(2)}$, ..., $W^{(n)}$. One easily checks that the processes W and $\mathbb{W}$ are martingales (with respect to the natural filtrations). Furthermore, one computes that the skew bracket of W is equal to t: we have $\langle W\rangle_t = t$ for all $t \geq 0$.

We proceed to the presentation of some basic facts about the stochastic integration. For the sake of simplicity, we will focus mainly on the context in which the integrator has continuous-paths. So, suppose that X is a continuous-path semimartingale with continuous trajectories and let H be some process: our aim is to define the integral $(H \cdot X)_t = \int_0^t H_s \cdot \mathrm{d}X_s$. Assume first that H is "very simple", that is, we have

$$H = \xi 1_{\{0\}} \quad \text{for some } \mathcal{F}_0\text{-measurable random variable } \xi$$

or, for some $0 \leq s \leq t$,

$$H = \xi 1_{(s,t]} \quad \text{for some } \mathcal{F}_s\text{-measurable random variable } \xi.$$

Then for any r, we define

$$\int_0^r H_u \mathrm{d}X_u = 0$$

in the first case, and

$$\int_0^r H_u \mathrm{d}X_u = \xi(X_{r\wedge t} - X_{s\wedge t})$$

in the second case. By linearity, this definition extends to integrals of simple integrands H (linear combinations of very simple integrands). By a straightforward limiting argument, one can define the stochastic integrals, with respect to X, of *locally bounded predictable processes* H. The key property which enables the extension is the identity

$$\mathbb{E}\left(\int_0^t H_s \cdot \mathrm{d}X_s\right)^2 = \mathbb{E}\int_0^t H_s^2 \mathrm{d}\langle X\rangle_s,$$

provided X is a martingale bounded in L^2. Actually, pushing the localization a little further, we obtain the following generalization. Suppose that $X = X_0 + M + A$ is a semimartingale and let H be a predictable process such that

$$\int_0^t |H_s||\mathrm{d}A|_s < \infty \qquad \text{almost surely for all } t > 0$$

and

$$\int_0^t H_s^2 \mathrm{d}\langle M\rangle_s < \infty \qquad \text{almost surely for all } t > 0.$$

Then we define the stochastic integral via

$$H \cdot X = H \cdot A + H \cdot M,$$

where $H \cdot A$ is the Lebesgue-Stieltjes integral and $H \cdot M$ is the integral with respect to the local martingale M. The latter stochastic integral is defined as the limit in probability of $H^{(n)} \cdot X$, where $H^n = H1_{\{|H|\leq n\}}$ is a predictable and bounded process.

Let us list some basic properties of stochastic integrals and skew brackets introduced above.

THEOREM 2.3. *(i) The mapping $H \mapsto H \cdot X$ is linear.*

(ii) The process $H \cdot X$ is a semimartingale.

(iii) If X is a local martingale, then $H \cdot X$ is also a local martingale.

(iv) If X has a locally bounded variation, then $H \cdot X$ has locally bounded variation.

(v) If H is a bounded integrable process, then $\langle H \cdot X, Y\rangle = H \cdot \langle X, Y\rangle$. In particular, $\langle H \cdot X\rangle = H^2 \cdot \langle X\rangle$.

(vi) If X has a locally bounded variation, then $\langle X, Y\rangle = 0$.

One of the fundamental statements in the theory of stochastic integration is the following Itô's formula. We state it in the context of continuous-path semimartingales.

THEOREM 2.4. *Suppose that $X = (X^1, X^2, \ldots, X^d)$ is a d-dimensional semimartingale with continuouos trajectories. Assume further that $F : \mathbb{R}^d \to \mathbb{R}$ is a C^2 function. Then the process $F(X)$ is again a semimartingale and we have*

$$F(X_t) = I_0 + I_1 + I_2 + I_3,$$

where

$$\begin{aligned} I_0 &= F(X_0), \\ I_1 &= \sum_{j=1}^{d} \int_0^t F_{x_j}(X_s) \cdot dX_s^j, \\ I_2 &= \frac{1}{2} \sum_{j,k \leq d} \int_0^t F_{x_j x_k}(X_s) d\langle X^j, X^k\rangle_s. \end{aligned}$$

We will often work in the very special case in which the semimartingale X is the space-time Brownian motion: $X_t = (t, W_t)$ for $t \geq 0$. In such a case, Itô's formula simplifies significantly and reads

$$\begin{aligned} F(t, W_t) &= F(0,0) + \int_0^t F_t(s, W_s)\mathrm{d}s + \int_0^t F_x(s, W_s) \cdot \mathrm{d}W_s + \frac{1}{2}\int_0^t F_{xx}(s, W_s)\mathrm{d}s \\ &= F(0,0) + \int_0^t \left(F_t(s, W_s) + \frac{1}{2}F_{xx}(s, W_s)\right)\mathrm{d}s + \int_0^t F_x(s, W_s) \cdot \mathrm{d}W_s \end{aligned}$$

for any C^2 function F on $[0, \infty) \times \mathbb{R}$. This immediately gives the following fact.

COROLLARY 2.1. *If a C^2 function F satisfies the heat equation*

$$F_t + \frac{1}{2}F_{xx} = 0,$$

then the process $(F(t, W_t))_{t\geq 0}$ is a local martingale.

We turn our attention to some basic facts about Markov processes. Suppose that $(E, \mathcal{B})$ is a state space. Assume in addition that $(\Omega, \mathcal{F})$ is a measure space, equipped with the filtration $(\mathcal{F}_t)_{t\geq 0}$ and a family $(\mathbb{P}_x)_{x\in E}$ of probability measures. Finally, suppose that $X = (X_t)_{t\geq 0}$ is an adapted process.

DEFINITION 2.7. Suppose that $P : [0,\infty)\times E\times\mathcal{B} \to [0,1]$ is a function satisfying the following requirements:
(i) For any $t \geq 0$ and $x \in E$, $P_t(x;\cdot)$ is a probability measure.
(ii) For any $t \geq 0$ and $A \in \mathcal{B}$, the map $x \mapsto P_t(x;A)$ is measurable.
(iii) For any $s, t \geq 0$ and $A \in \mathcal{B}$ we have the Chapman-Kolmogorov identity

$$P_{s+t}(x;A) = \int_E P_t(y;A)P_s(x;dy).$$

Then P is called a (time-homogeneous) transition function.

DEFINITION 2.8. Let P be a (time-homogeneous) transition function. A process X is said to be a (time-homogeneous) Markov process with the transition function P, if for any $s, t \geq 0$, any $x \in E$ and any $A \in \mathcal{B}(E)$ we have

$$\mathbb{P}_x(X_{s+t} \in A|\mathcal{F}_s) = P_t(X_s;A).$$

The process X is said to have a strong Markov property, if for any stopping time τ, any $s > 0$, any $x \in E$ and any $A \in \mathcal{B}(E)$ we have

$$\mathbb{P}_x(X_{s+\tau} \in A|\mathcal{F}_s) = P_s(X_\tau;A).$$

For a given Markov process X, suppose that T_t is a shift operator, acting on measurable functions f by

$$T_t f(x) = \mathbb{E}_x f(X_t) = \int_E f(y)P_t(x;dy)$$

(we assume that f has the property that the integrals above exist).

DEFINITION 2.9. A Markov process $(X_t)_{t\geq 0}$ is called a Feller process if the associated semigroup $(T_t)_{t\geq 0}$ is a Feller semigroup, i.e.
(i) $T_t f \in C_\infty(E)$ if $f \in C_\infty(E)$ (Feller property);
(ii) $\|T_t f - f\|_\infty \to 0$ as $t \to 0$ for any $f \in C_\infty(E)$ (strong continuity at $t = 0$).
(Here $C_\infty(E)$ is the class of all continuous functions which vanish at infinity).

Now we can define the associated infinitesimal operator $\mathbb{L} = \mathbb{L}_X$. First we introduce the domain of the operator, by

$$\mathcal{D}(\mathbb{L}) = \left\{ f \in C_\infty(E) \,:\, \exists_{g\in C_\infty(E)} \left\| \frac{T_t f - f}{t} - g \right\|_\infty \xrightarrow{t\to 0} 0 \right\}.$$

The generator $\mathbb{L}$ acts on $f \in \mathcal{D}(\mathbb{L})$ by

$$\mathbb{L}f(x) = \lim_{t\downarrow 0} \frac{T_t f(x) - f(x)}{t}, \qquad f \in \mathcal{D}(\mathbb{L}).$$

THEOREM 2.5. *Suppose that* $(X_t)_{t\geq 0}$ *is a Feller process with generator* $\mathbb{L}_X$*. Then for any* x *and any* $f \in \mathcal{D}(\mathbb{L})$ *the process*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \mathbb{L} f(X_s) ds$$

is a $\mathbb{P}^x$*-martingale.*

PROOF. We need to show that

$$\mathbb{E}^x \left(f(X_t) - \int_s^t \mathbb{L} f(X_r) \mathrm{d}r \,\middle|\, \mathcal{F}_s \right) = f(X_s).$$

By Markov property, we may write

$$\begin{aligned}\mathbb{E}^x \left(f(X_t) - \int_s^t \mathbb{L} f(X_r) \mathrm{d}r \,\middle|\, \mathcal{F}_s \right) &= \mathbb{E}^x (f(X_t)|\mathcal{F}_s) - \mathbb{E}^x \left(\int_0^{t-s} \mathbb{L} f(X_{s+r}) \mathrm{d}r \,\middle|\, \mathcal{F}_s \right) \\ &= \mathbb{E}^{X_s} f(X_{t-s}) - \int_0^{t-s} \mathbb{E}^{X_s} \mathbb{L} f(X_r) \mathrm{d}r \\ &= T_{t-s} f(X_s) - \int_0^{t-s} T_r \mathbb{L} f(X_s) \mathrm{d}r \\ &= f(X_s),\end{aligned}$$

where in the last line we have exploited the identity $\frac{d}{dr} T_r f = T_r \mathbb{L} f$. Indeed, using the semigroup structure, we get

$$\frac{d}{dr} T_r f = \lim_{t\to 0} \frac{T_t T_r f - T_r f}{t} = \lim_{t \to 0} T_r \left(\frac{T_t f - f}{t} \right) = T_r \mathbb{L} f,$$

where the last passage is due to the strong Feller property. □

2.2. Martingale approach

We proceed to the theory of optimal stopping of continuous-time processes. Suppose that $T \in [0, \infty]$ is a given and fixed time horizon. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. Let $G = (G_t)_{t\geq 0}$ be a fixed gain process, which will be assumed to be cádlág. For technical reasons, we impose the integrability condition

$$\mathbb{E} \sup_{0\leq t\leq T} |G_t| < \infty \tag{2.1}$$

(we use the convention $G_\infty = 0$). Our primary goal concerns the optimal stopping problem

$$V_t^T = \sup_{t \leq \tau \leq T} \mathbb{E} G_\tau.$$

There are two natural approaches. The first is to use approximation, in order to reduce the problem to the discrete-time context. More precisely, we consider an increasing sequence $(t_k^{(n)})$ of partitions of $[0, T]$, with the mesh $\sup_k (t_{k+1}^{(n)} - t_k^{(n)})$ converging to zero as $n \to \infty$. One introduces the associated discrete-time gain processes $G^{(n)} = (G_{t_k^{(n)}})$, applies the results from the discrete setting and passes to the limit $n \to \infty$. This approach is quite useful and meaningful from the viewpoint of numerical approximation. However, we will use a different way, namely, we will extend the method of essential supremum.

Introduce the process $B = (B_t)_{t\geq 0}$ given by

$$B_t^T = \operatorname*{esssup}_{\tau \geq t} \mathbb{E}(G_\tau | \mathcal{F}_t), \qquad t \geq 0.$$

This process will be referred to as the Snell envelope of G. Next, put

$$\tau_t = \inf\{s \geq t : B_s^T = G_s\},$$

with the standard convention $\inf \emptyset = \infty$.

THEOREM 2.6. *Suppose that the gain process G satisfies the integrability condition* (2.1)*. Consider the family of optimal stopping problems*

$$V_t^T = \sup_{\tau \geq t} \mathbb{E} G_\tau, \qquad t \geq 0. \tag{2.2}$$

Assume further that the stopping time τ_t is finite almost surely for all $t \geq 0$.

(i) For any $t \geq 0$ we have

$$B_t^T \geq \mathbb{E}(G_\tau | \mathcal{F}_t) \qquad \text{for all } \tau \text{ satisfying } \tau \geq t$$

and

$$B_t^T = \mathbb{E}(G_{\tau_t} | \mathcal{F}_t).$$

(ii) For any $t \geq 0$, the stopping time τ_t is optimal for the problem $V_t^T = \sup_{\tau \geq t} \mathbb{E} G_\tau$.

(iii) If τ_t^* is another optimal stopping time for $V_t^T = \sup_{\tau\geq t} \mathbb{E} G_\tau$, then $\tau_t \leq \tau_t^*$ almost surely.

(iv) The process $(B_s^T)_{s\geq t}$ is the smallest right-continuous supermartingale majorizing $(G_s)_{s\geq t}$. Furthermore, the stopped process $(B_{\tau_t \wedge s}^T)_{s \geq t}$ is a right-continuous martingale.

As we see, the above statement is quite parallel to Theorem 1.2 in discrete time. We will not present the proof, referring the interested reader to Chapter I, Section 2 in [**18**] for details.

Now we will analyze several exemplary applications of the above statement.

EXAMPLE 2.1. Suppose that X is a Bessel process of dimension 2, started at zero. We will find the best constant C in the estimate

$$\mathbb{P}\Big(\sup_{0 \leq t \leq \tau} X_t \geq 1 \Big) \leq C \mathbb{E} \tau, \tag{2.3}$$

to be valid for all integrable stopping times τ. To this end, note that X is the norm of the two-dimensional Brownian motion $\mathbb{W} = (W^{(1)}, W^{(2)})$: $X_t^2 = (W_t^{(1)})^2 + (W_t^{(2)})^2$ for all $t \geq 0$. We fix $C > 0$ and consider the associated optimal stopping problem

$$V = \sup_\tau \mathbb{E} \left[1_{\{\sup_{0\leq t\leq\tau}((W_t^{(1)})^2 + (W_t^{(2)})^2) \geq 1\}} - C\tau \right].$$

Note that the problem of finding the optimal constant C in (2.3) is equivalent to that of finding the least C for which the value V is nonpositive. Let B be the corresponding Snell envelope: for any $t \geq 0$ we put

$$B_t = \sup_{\tau \geq t} \mathbb{E} \left[1_{\{\sup_{0\leq s\leq\tau}((W_s^{(1)})^2 + (W_s^{(2)})^2) \geq 1\}} - C\tau \,\middle|\, \mathcal{F}_t \right].$$

We will try to guess the optimal stopping time for B_t. Roughly speaking, we want to make the indicator as large as possible, keeping $\mathbb{E}\tau$ as small as possible at the same time. Obviously, if $\sup_{0\leq s\leq t}\Big[(W_s^{(1)})^2+(W_s^{(2)})^2\Big]\geq 1$, then it is optimal to stop immediately: the indicator function already returns one, while the expectation $\mathbb{E}\tau$ can only increase. Thus, all we need is to identify the rule when $\sup_{0\leq s\leq t}\Big[(W_s^{(1)})^2+(W_s^{(2)})^2\Big]$ is less than one. A little thought leads to the conjecture that in such a case it is optimal to wait until the supremum reaches one. Summarizing, we conjecture that the optimal stopping time for B_t is given by

$$\tau_*=\inf\Big\{u\geq t:\ \sup_{0\leq s\leq u}\Big[(W_s^{(1)})^2+(W_s^{(2)})^2\Big]\geq 1\Big\}.$$

For this stopping time, the conditional expectation

$$\mathcal{B}_t=\mathbb{E}\bigg[1_{\{\sup_{0\leq s\leq\tau_*}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1\}}-C\tau_*\Big|\mathcal{F}_t\bigg]$$

is given as follows. If $\sup_{0\leq s\leq t}\Big[(W_s^{(1)})^2+(W_s^{(2)})^2\Big]\geq 1$, then $\mathcal{B}_t=1-Ct$. If $\sup_{0\leq s\leq t}\Big[(W_s^{(1)})^2+(W_s^{(2)})^2\Big]<1$, then we use the fact that $\Big((W_s^{(1)})^2+(W_s^{(2)})^2-2s\Big)_{s\geq 0}$ is a martingale to obtain

$$\begin{aligned}\mathcal{B}_t&=1-C\mathbb{E}\big(\tau_*|\mathcal{F}_t\big)\\&=1-\frac{C}{2}\mathbb{E}\Big((W_{\tau_*}^{(1)})^2+(W_{\tau_*}^{(2)})^2|\mathcal{F}_t\Big)+\frac{C}{2}\mathbb{E}\Big((W_{\tau_*}^{(1)})^2+(W_{\tau_*}^{(2)})^2-2\tau_*|\mathcal{F}_t\Big)\\&=1-\frac{C}{2}+\frac{C}{2}\Big((W_t^{(1)})^2+(W_t^{(2)})^2-2t\Big).\end{aligned}$$

The process $\mathcal{B}$ computed above is the *candidate* for the (explicit form of) Snell envelope B. By the very definition of B, we have $B\geq\mathcal{B}$; to check the reverse inequality, we need to verify that $\mathcal{B}$ is a supermartingale dominating the gain process

$$G=\Big(1_{\{\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1\}}-Ct\Big)_{t\geq 0}.$$

The supermartingale property is straightforward, so we turn our attention to the majorization. If $\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1$, then $\mathcal{B}_t=G_t$. If the reverse estimate holds, then we have

$$1-\frac{C}{2}+\frac{C}{2}\Big((W_t^{(1)})^2+(W_t^{(2)})^2-2t\Big)\geq 1-\frac{C}{2}-Ct\geq -Ct=G_t$$

provided $C\leq 2$. So, for such C, we have $B=\mathcal{B}$ and hence

$$V=\mathbb{E}B_0=1-\frac{C}{2}.$$

Now, if $C=2$, then $V=0$ and hence the desired estimate (2.3) holds; if $C<2$, then $V>0$ and therefore the inequality under the investigation fails to hold for the stopping time τ_* defined above. This shows that the best constant in (2.3) is equal to 2.

REMARK 2.1. Since $\mathbb{E}\tau=\mathbb{E}(W_\tau^{(1)})^2$ for any $\tau\in L^1$, the above analysis establishes the sharp estimate

$$\mathbb{P}\Big(\sup_{0\leq t\leq\tau}((W_t^{(1)})^2+(W_t^{(2)})^2)\geq 1\Big)\leq 2\mathbb{E}(W_\tau^{(1)})^2.$$

Now we will consider a slight modification of the above inequality.

EXAMPLE 2.2. We will identify the optimal constant C in the estimate

$$\mathbb{P}\Big(\sup_{0\leq t\leq\tau}((W_t^{(1)})^2+(W_t^{(2)})^2)\geq 1\Big)\leq C\mathbb{E}|W_\tau^{(1)}|, \tag{2.4}$$

to be valid for all $\tau\in L^{1/2}$. We proceed as previously, introducing the associated problem

$$V=\sup_\tau\mathbb{E}\Big[1_{\{\sup_{0\leq t\leq\tau}((W_t^{(1)})^2+(W_t^{(2)})^2)\geq 1\}}-C|W_\tau^{(1)}|\Big]$$

and the Snell envelope

$$B_t=\sup_{\tau\geq t}\mathbb{E}\Big[1_{\{\sup_{0\leq s\leq\tau}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1\}}-C|W_\tau^{(1)}|\,\Big|\mathcal{F}_t\Big].$$

Again, we conjecture that the optimal stopping time is given by

$$\tau_*=\inf\Big\{u\geq t\,:\,\sup_{0\leq s\leq u}\Big[(W_s^{(1)})^2+(W_s^{(2)})^2\Big]\geq 1\Big\}$$

and denote

$$\mathcal{B}_t=\mathbb{E}\Big[1_{\{\sup_{0\leq s\leq\tau_*}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1\}}-C|W_{\tau_*}^{(1)}|\,\Big|\mathcal{F}_t\Big].$$

If $\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1$, then $\mathcal{B}_t=1-C|W_t^{(1)}|$. The analysis becomes a little more involved in the case $\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)<1$: then we have $\mathcal{B}_t=1-C\mathbb{E}[|W_{\tau_*}^{(1)}||\mathcal{F}_t]$ and there is a question how to compute the latter expectation. This leads us to the following simple Dirichlet problem. Suppose that u is a harmonic function on the unit disc $\mathbb{D}$, continuous to the boundary, such that $u(x,y)=|x|$ for $(x,y)\in\partial\mathbb{D}$. By the classical theory, this function is given explicitly, in polar coordinates, by the convolution

$$u(re^{i\theta})=\frac{1}{2\pi}\int_{-\pi}^{\pi}P_r(\theta-t)|e^{it}|\mathrm{d}t,$$

where P is the Poisson kernel for the disc, given by

$$P_r(\theta)=\sum_{n=-\infty}^{\infty}r^{|n|}e^{in\theta}=\frac{1-r^2}{1-2r\cos\theta+r^2}.$$

By Itô's formula, we have

$$|W_{\tau_*}^{(1)}|=u(W_{\tau_*}^{(1)},W_{\tau_*}^{(2)})=u(W_t^{(1)},W_t^{(2)})+\int_t^{\tau_*}\nabla u(W_s^{(1)},W_s^{(2)})\cdot\mathrm{d}\mathbb{W}_s$$

and hence $\mathbb{E}[|W_{\tau_*}^{(1)}||\mathcal{F}_t]=u(W_t^{(1)},W_t^{(2)})$. Thus we have obtained the candidate

$$\mathcal{B}_t=1-Cu(W_t^{(1)},W_t^{(2)})\qquad\text{for }\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)<1.$$

As before, we need to check that $\mathcal{B}$ is a supermartingale majorizing

$$G=\Big(1_{\{\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1\}}-C|W_t^{(1)}|\Big)_{t\geq 0}.$$

We will focus on the majorization. If $\sup_{0\leq s\leq t}((W_s^{(1)})^2+(W_s^{(2)})^2)\geq 1$, then $\mathcal{B}_t=G_t$. If the reverse estimate holds, then the majorization becomes

$$1-Cu(W_t^{(1)},W_t^{(2)})\geq -C|W_t^{(1)}|.$$

So, we will be done if we prove that $1 - Cu(x,y) \geq -C|x|$ for $(x,y) \in \mathbb{D}$. By symmetry, it is enough to show this estimate for $x \geq 0$: then the bound becomes $1 - Cu(x,y) \geq -Cx$. Since both sides are harmonic and $u(x,y) = x$ for $(x,y) \in \mathbb{D} \cap \{x \geq 0\}$, it suffices to prove the estimate for $x = 0$. In this special case, the majorization reduces to $u(0,y) \leq 1/C$, and again by the symmetry we may restrict ourselves to $y \geq 0$. However, by the definition of u, we have

$$u(0,y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_y(\pi/2 - t)|\cos t|\mathrm{d}t$$

and one can show that the right-hand side is decreasing with y. Hence the majorization holds when $C \leq 1/u(0,0) = \pi/2$. Consequently, for such C we have $V = \mathbb{E}B_0 = 1 - Cu(0,0)$. Hence we need to take $C = \pi/2$: then $V = 0$ and the estimate holds; the choice of smaller C violates (2.4) for the stopping time τ_*.

The purpose of the next example is to develop a certain argument which will be needed in our further considerations.

Example 2.3. Suppose that X is a martingale starting from 1, satisfying the stochastic differential equation $dX_t = X_t dW_t$ (that is, such that $X_t = X_0 + \int_0^t X_s \mathrm{d}W_s$ for all $t \geq 0$) and stopped upon reaching 0 or 2. We will solve the optimal stopping problem

$$V = \sup_\tau \mathbb{E}\big(X_\tau^2 - \tau\big).$$

It is perhaps more natural to solve this problem using the Markovian approach, but it is also instructive to apply the martingale theory. In the previous examples, our first step was to guess the optimal stopping time, compute the candidate for the Snell envelope and verify that this candidate has all the necessary properties (which guarantee that it does coincide with the envelope). Here we will proceed a little differently: using some algebraic manipulations, we will construct the candidate for the Snell envelope directly and then come up with the optimal stopping time.

To this end, let us try to analyze supermartingale majorants of the process $G = (X_t^2 - t)_{t\geq 0}$, of the form $\mathcal{B}_t = f(X_t) - t$ for some (sufficiently regular) f. We will be very imprecise in the considerations below: the purpose is to reveal some very general facts about the envelope, which are then established rigorously. The first (trivial) step is to try the process G itself, i.e., to take $f(x) = x^2$. Itô's formula implies that the supermartingale property will hold if f satisfies

$$\frac{1}{2}f''(x)x^2 - 1 \leq 0. \tag{2.5}$$

Hence $(X_t^2 - t)_{t\geq 0}$ would be a supermartingale, if it took values in the interval $[0,1]$: for $f(x) = x^2$, we have $\frac{1}{2}f''(x)x^2 - 1 = x^2 - 1$. Since the range of the process is larger, the formula for f should be modified, at least on some interval of the form $[x_0, 2]$. Assuming equality in (2.5), we see that f must be of the form $f(x) = -2\log x + ax + b$ for some unknown constants a and b. Now, from the general theory, we search for the *least* supermartingale majorant of G; by the above observations, it seems natural to consider functions of the form

$$f(x) = \begin{cases} x^2 & \text{if } x \leq x_0, \\ -2\log x + ax + b & \text{if } x \geq x_0. \end{cases} \tag{2.6}$$

To guarantee the minimality of the majorant, we will impose the following *principle of smooth-fit*: $f'(x_0-) = f'(x_0+)$. Furthermore, the optimality condition implies that f should satisfy the requirement $f(2) = 2^2$. Putting the above facts together, we obtain the following system of equations:

$$\begin{aligned} x_0^2 &= -2\log x_0 + ax_0 + b && \text{(continuouos fit)},\\ 2x_0 &= -\frac{2}{x_0} + a && \text{(smooth fit)},\\ 4 &= -2\log 2 + 2a + b && (f(2) = 2^2). \end{aligned}$$

By a direct computation, we check that x_0 is the unique root of the equation $x_0^2 - 4x_0 - 4x_0^{-1} - 2\log x_0 + 6 + 2\log 2 = 0$ lying in $(0, 1)$, and $a = 2(x_0 + x_0^{-1})$, $b = x_0^2 - ax_0 + 2\log x_0$.

The above procedure was informal and exploited a number of guesses and additional assumptions. Now we check rigorously that the function f constructed above does lead to the Snell envelope corresponding to our problem. First, note that the function (2.6) is of class C^1 and $(f(X_t) - t)_{t\geq 0}$ is a supermartingale. Note that $f(x) \geq x^2$. Indeed, for $x \leq x_0$ we have equality, so we may focus on the interval $(x_0, 2]$. Note that $f(x_0) - x_0^2 = f'(x_0) - 2x_0 = 0$ and $(f(x) - x^2)'' = 2x^{-2} - 2$, so $f(x) - x^2$ is convex on some subinterval of the form $(x_0, x_1]$ and concave on $[x_1, 2]$. Since $f(2) - 2^2 = 0$, this establishes the desired majorization. Hence, the Snell envelope B of G does not exceed $(f(X_t) - t)_{t\geq 0}$. To see the reverse, we need to find a stopping time for which the corresponding conditional expectation (defining B_t) yields $f(X_t) - t$. We fix t; by the general theory, the optimal stopping time is given as the first time when the Snell envelope meets the gain process. This leads us to the choice $\tau_* = \inf\{s \geq t : X_s \leq x_0 \text{ or } X_s = 2\}$. Now, by the construction, the process $(f(X_{\tau_*\wedge s}) - \tau_* \wedge s)$ is a martingale (since we have equality in (2.5) for $x \in (x_0, 2)$) and hence

$$B_t \geq \mathbb{E}(X_{\tau_*}^2 - \tau_*|\mathcal{F}_t) = \mathbb{E}(f(X_{\tau_*}) - \tau_*|\mathcal{F}_t) = f(X_t) - t.$$

(We have carried out some standard limiting arguments in Doob's optional sampling theorem, which available due to the boundedness of f). This solves the above optimal stopping problem: we have $V = \mathbb{E}B_0 = a + b$ and the stopping time τ_* is optimal.

REMARK 2.2. The above reasoning can be easily generalized. Suppose that X is a process satisfying the differential equation $dX_t = \sigma(X_t)dW_t$ for $t \geq 0$. Let $g : \mathbb{R} \to \mathbb{R}$ be a given function. Assume further that we are interested in the optimal stopping problem

$$V = \sup_\tau \mathbb{E}[g(X_\tau) - \tau].$$

Then one needs to search for the Snell envelope of the form $(f(X_t)-t)_{t\geq 0}$, where f is a (smallest) majorant of g, satisfying $\frac{1}{2}f''(x)\sigma^2(x) - 1 \leq 0$ on a large part of its domain. This observation will be pushed much further in the Markovian approach to optimal stopping problems.

EXAMPLE 2.4 (Sequential testing of a Wiener process). The last example in this section will be more elaborate, but it is very natural and interesting from the viewpoint of applications. The idea of the problem is the following: suppose that we observe a trajectory of a Wiener process with unknown drift θ which is 0 or 1 with probability 1/2. The goal is to determine the drift after a (relatively) short time.

Formally, suppose that W is a standard one-dimensional Brownian motion (with drift 0) and let θ be a random variable independent of W, following the law $\mathbb{P}(\theta = 0) = \mathbb{P}(\theta = 1) = 1/2$. Assume further that we observe the trajectory of the process $X_t = \theta t + W_t$, $t \geq 0$. We search for the decision rule (τ, d), where τ is the stopping time of X (that is, a stopping time relative to the filtration $\mathcal{F}^X = (\sigma(X_s : 0 \leq s \leq t))_{t\geq 0}$) and an $\mathcal{F}^X_\tau$-measurable random variable d taking values in $\{0, 1\}$; after the terminal observation at time τ, the decision d indicates the drift. The associated risk function is given by

$$V = \inf_{(\tau,d)} \mathbb{E}\big(\tau + a1_{\{d=0,\theta=1\}} + b1_{\{d=1,\theta=0\}}\big),$$

and the goal is to find the decision rule (τ, d) for which the infimum is attained. Let us briefly comment on the components of the number V. The term τ corresponds to the cost of observation; the remaining two summands are related to the loss incurred by the wrong decision.

It is convenient to split the analysis into several steps.

Step 1. Introduce the *a posteriori process* $\pi_t = \mathbb{P}(\theta = 1|\mathcal{F}^X_t)$, $t \geq 0$. One can show that if τ is the terminal stopping time, then the decision d must be of the form

$$d = \begin{cases} 1 & \text{if } \pi_\tau \geq c, \\ 0 & \text{if } \pi_\tau < c. \end{cases}$$

To determine c, we compute that

$$\begin{aligned} \mathbb{E}(a1_{\{d=0,\theta=1\}} + b1_{\{d=1,\theta=0\}}) &= \mathbb{E}\big[\mathbb{E}(a1_{\{d=0,\theta=1\}} + b1_{\{d=1,\theta=0\}}|\mathcal{F}^X_\tau)\big] \\ &= \mathbb{E}\big[a1_{\{d=0\}}\pi_\tau + b1_{\{d=1\}}(1-\pi_\tau)\big] \\ &= \mathbb{E}\big[a1_{\{\pi_\tau<c\}}\pi_\tau + b1_{\{\pi_\tau\geq c\}}(1-\pi_\tau)\big] \\ &\geq \mathbb{E}\min\big\{a\pi_\tau, b(1-\pi_\tau)\big\} \end{aligned}$$

and we have equality in the last passage if and only if $c = b/(a+b)$. Thus, we come up with the following optimal stopping problem:

$$V = \inf_\tau \mathbb{E}\big[\tau + \min\big\{a\pi_\tau, b(1-\pi_\tau)\big\}\big].$$

Step 2. Now we will express the process π in terms of X. Fix an arbitrary integer n, short intervals $A_1, A_2, \ldots, A_n \subset \mathbb{R}$ and times $0 = t_0 < t_1 < t_2 < \ldots < t_n = t$. Finally, take the event $\mathcal{A} = \{X_{t_1} \in A_1, X_{t_2} - X_{t_1} \in A_2, \ldots, X_{t_n} - X_{t_{n-1}} \in A_n\}$. By Bayes' theorem,

$$\begin{aligned} \mathbb{P}(\theta = 0|\mathcal{A}) &= \frac{\mathbb{P}(\mathcal{A}|\theta = 0)\mathbb{P}(\theta = 0)}{\mathbb{P}(\mathcal{A}|\theta = 0)\mathbb{P}(\theta = 0) + \mathbb{P}(\mathcal{A}|\theta = 1)\mathbb{P}(\theta = 1)} \\ &= \frac{\mathbb{P}(\mathcal{A}|\theta = 0)}{\mathbb{P}(\mathcal{A}|\theta = 0) + \mathbb{P}(\mathcal{A}|\theta = 1)}. \end{aligned}$$

The conditional probabilities on the right can be easily computed, using the independence of increments of Brownian motion: we have

$$\mathbb{P}(\mathcal{A}|\theta = 0) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi(t_k - t_{k-1})}} \int_{A_k} \exp\left(-\frac{x^2}{2(t_k - t_{k-1})}\right) \mathrm{d}x$$

and

$$\mathbb{P}(\mathcal{A}|\theta=1)=\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi(t_k-t_{k-1})}}\int_{A_k}\exp\left(-\frac{(x-(t_k-t_{k-1}))^2}{2(t_k-t_{k-1})}\right)\mathrm{d}x$$
$$\approx\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi(t_k-t_{k-1})}}\int_{A_k}\exp\left(-\frac{x^2}{2(t_k-t_{k-1})}\right)\mathrm{d}x\cdot e^{\eta-t/2},$$

where $\eta=a_1+a_2+\ldots+a_n$ and $a_j\in A_j$. Hence, letting $|A_i|\to 0$ and $n\to\infty$, we obtain the identity

$$\mathbb{P}(\theta=0|\mathcal{F}_t^X)=\frac{1}{1+e^{X_t-t/2}}$$

and therefore

$$\pi_t=\mathbb{P}(\theta=1|\mathcal{F}_t^X)=\frac{e^{X_t-t/2}}{1+e^{X_t-t/2}}. \tag{2.7}$$

Step 3. Now we will analyze the behavior of π. Obviously, it is a martingale with respect to $(\mathcal{F}_t^X)_{t\geq 0}$. Furthermore, by Itô's formula, we get

$$d\pi_t=\pi_t(1-\pi_t)d(X_t-t/2)+\frac{1}{2}\pi_t(1-\pi_t)(1-2\pi_t)\mathrm{d}\left\langle X-t/2\right\rangle_t. \tag{2.8}$$

Now, the process $(X_t-\theta t)_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$ martingale, so applying the conditional expectation with respect to $\mathcal{F}^X$ we obtain

$$\mathbb{E}(X_t-\theta t|\mathcal{F}_s^X)=\mathbb{E}(X_s-\theta s|\mathcal{F}_s^X).$$

But $\mathbb{E}(\theta|\mathcal{F}_s^X)=\mathbb{E}(1_{\{\theta=1\}}|\mathcal{F}_s^X)=\mathbb{P}(\theta=1|\mathcal{F}_s^X)=\pi_s$ for all s. This observation combined with the previous identity implies $\mathbb{E}(X_t-t\pi_t|\mathcal{F}_s^X)=X_s-s\pi_s$, so $(X_t-t\pi_t)_{t\geq 0}$ is a martingale. This can be further simplified. Note that integration by parts (or Itô's formula) gives

$$t\pi_t=\int_0^t\pi_s ds+\int_0^t t d\pi_s$$

and hence $\left(X_t-\int_0^t\pi_s\mathrm{d}s\right)_{t\geq 0}$ is a local martingale. We go back to (2.8) and rewrite it in the form

$$d\pi_t=\pi_t(1-\pi_t)d\left(X_t-\int_0^t\pi_s\mathrm{d}s\right)+I,$$

where

$$I=\pi_t(1-\pi_t)d\left(\int_0^t\pi_s\mathrm{d}s-\frac{t}{2}\right)+\frac{1}{2}\pi_t(1-\pi_t)(1-2\pi_t)\mathrm{d}\left\langle X\right\rangle_t$$
$$=\pi_t(1-\pi_t)\left(\pi_t-\frac{1}{2}\right)\mathrm{d}(t-\langle X\rangle_t)=0$$

(recall that π is a martingale). This implies that $\langle X-\int_0^\cdot\pi_s\mathrm{d}s\rangle_t=t$, so the process $(X_t-\int_0^t\pi_s)_{t\geq 0}$ is an $\mathcal{F}^X$-Brownian motion. This brings us to the context studied in the previous example.

Step 4. Now we will use the observation from Remark 2.2. Our optimal stopping problem can be rewritten in the form

$$-V=\sup_\tau\mathbb{E}\big[g(\pi_\tau)-\tau\big],$$

where $g(x) = -\min\{ax, b(1-x)\}$. To solve it, we search for the least majorant f of g, satisfying the condition $\frac{1}{2}f''(x)x^2(1-x)^2 - 1 \leq 0$. Some lengthy (but rather straightforward) calculations yield the existence of unique numbers $0 < \alpha_* < c < \beta_* < 1$ for which there is a C^1 function f satisfying $f = g$ for $x \leq \alpha_*$ or $x \geq \beta_*$, and

$$\frac{1}{2}f''(x)x^2(1-x)^2 - 1 = 0$$

for $x \in (\alpha_*, \beta_*)$. We omit the technical details - they are quite similar (yet a little more involved) to those presented in the analysis of the previous example.

Now, a direct application of Itô's formula implies that $(f(\pi_t) - t)_{t\geq 0}$ is a supermartingale majorizing $(g(\pi_t) - t)_{t\geq 0}$ and hence the Snell envelope B is not smaller than $(f(\pi_t) - t)_{t\geq 0}$. To obtain the reverse estimate, consider the stopping time $\tau = \inf\{t : \pi_t \leq \alpha_* \text{ or } \pi_t \geq \beta_*\}$.

It remains to note that the outcome is very intuitive: we wait until the process π_t given in (2.7) is either 'big' or 'small', and depending on which scenario happens, we decide whether $d = 1$ or $d = 0$.

2.3. Markovian approach

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. Assume further that $X = (X_t)_{t\geq 0}$ is a strong Markov family on the state space $E = \mathbb{R}^d$, with the corresponding family $(\mathbb{P}_x)_{x\in E}$ of probabilities such that $\mathbb{P}_x(X_0 = x) = 1$ and such that $x \mapsto \mathbb{P}_x(A)$ is measurable for each $A \in \mathcal{F}$. As before, for a given Borel function $G : E \to \mathbb{R}$ (called the gain function) and an arbitrary state $x \in E$, we consider the optimal stopping problem

$$V(x) = \sup_\tau \mathbb{E}_x G(X_\tau), \tag{2.9}$$

where the supremum is taken over all finite stopping times relative to $(\mathcal{F}_t)_{t\geq 0}$. For technical reasons, we assume that G satisfies

$$\mathbb{E}_x \sup_{t\geq 0} |G(X_t)| < \infty \qquad \text{for all } x \in E. \tag{2.10}$$

The function V defined in (2.9) is called the value function.

We introduce the associated continuation and stopping regions, given by

$$C = \{x \in E \,:\, V(x) > G(x)\}, \qquad D = \{x \in E \,:\, V(x) = G(x)\}.$$

Definition 2.10. Let X be a topological space and let $x_0 \in X$. A function $f : X \to \mathbb{R}$ is lower semicontinuous at x_0, if for any $y < f(x_0)$ there is a neighborhood U of x_0 such that $f(x) > y$ for all $x \in U$. Equivalently, we have $\liminf_{x\to x_0} f(x) \geq f(x_0)$. A function f is lower semicontinuous on X, if it is lower semicontinuous at each point $x_0 \in X$. A function $f : X \to \mathbb{R}$ is upper semicontinuous, if $-f$ is lower semicontinuous.

One can show that if V is lower semicontinuous and G is upper semicontinuous, then C is open and D is closed. We consider the associated first-entry time

$$\tau_D = \inf\{t \geq 0 \,:\, X_t \in D\}.$$

Note that if D is closed, then τ_D is a stopping time since both X and $(\mathcal{F}_t)_{t\geq 0}$ are right-continuous.

DEFINITION 2.11. Suppose that $F : E \to \mathbb{R}$ is a Borel function satisfying $\mathbb{E}_x|F(X_\tau)| < \infty$ for all x and all stopping times τ. Then F is called excessive (superharmonic), if for any x and any stopping time τ we have

$$\mathbb{E}_x F(X_\tau) \leq F(x).$$

As in the discrete-time setting, one easily checks that F is excessive if and only if $(F(X_t))_{t\geq 0}$ is a supermartingale under each $\mathbb{P}_x$, $x \in E$.

THEOREM 2.7. *Suppose that for any $x \in E$ there exists an optimal stopping time τ_* in the problem*

$$V(x) = \sup_\tau \mathbb{E}_x G(X_\tau),$$

that is, $V(x) = \mathbb{E}_x G(X_{\tau_*})$.

(i) The value function V is the smallest excessive function majorizing the gain function G on E.

Assume further that V is lower semicontinuous and G is upper semicontinuous.

(ii) The stopping time τ_D is optimal and satisfies $\tau_D \leq \tau_*$.

(iii) The stopped process $(V(X_{\tau_D \wedge t}))_{t\geq 0}$ is a right-continuous $\mathbb{P}_x$-martingale for all $x \in E$.

THEOREM 2.8. *Suppose that the integrability condition* (2.10) *holds and consider the optimal stopping problem*

$$V(x) = \sup_\tau \mathbb{E}_x G(X_\tau).$$

Assume that there exists the smallest excessive function $\hat{V}$ majorizing the gain function G on E. In addition, suppose that $\hat{V}$ is lower semicontinuous and G is upper semicontinuous. Consider the associated stopping set

$$\hat{D} = \{x \in E : \hat{V}(x) = G(x)\}$$

and let $\hat{\tau_D} = \inf\{t : X_t \in \hat{D}\}$.

(i) If $\mathbb{P}_x(\hat{\tau_D} < \infty) = 1$ for all $x \in E$, then $\hat{V} = V$ and $\hat{\tau_D}$ is the optimal stopping time.

(ii) If $\mathbb{P}_x(\hat{\tau_D} < \infty) < 1$ for some $x \in E$, then there is no optimal stopping time.

COROLLARY 2.2. *Suppose that V is lower semicontinuous and G is upper semicontinuous. If $\mathbb{P}_x(\tau_D < \infty) = 1$ for all $x \in E$, then τ_D is an optimal stopping time. If $\mathbb{P}_x(\tau_D < \infty) < 1$ for some $x \in E$, then there is no optimal stopping time.*

REMARK 2.3. If the function $x \mapsto \mathbb{E}_x G(X_\tau)$ is continuous (or lower semicontinuous) for every stopping time τ, then $x \mapsto V(x)$ is lower semicontinuous and the above corollary can be applied.

The above statements imply that the analysis of the optimal stopping problem

$$V = \sup_\tau \mathbb{E}_x G(X_\tau)$$

is equivalent to the search for the smallest excessive majorant $\hat{V}$ of G on E. Having found such an object, we obtain $V = \hat{V}$ and the optimal stopping time is given by $\tau_D = \inf\{t : \hat{V}(X_t) = G(X_t)\}$.

In our considerations below, we will search for the smallest excessive majorant by solving the corresponding free-boundary problem. The idea is the following: the function V should satisfy the conditions

$$\begin{aligned} \mathbb{L}_X V = 0 \quad & \text{on } C, \\ V = G \quad & \text{on } D. \end{aligned}$$

Assuming that G is smooth in some neighborhood of ∂C, we have two possibilities.

(i) If X after starting at ∂C enters immediately into $\text{int}D$ (for instance, when X is a diffusion and ∂C is sufficiently regular) then we should additionally have $\partial_x V = \partial_x G$ on ∂C (the principle of smooth fit).

(ii) If X after starting at ∂C does not enter immediately into $\text{int}D$ (for instance, when X has jumps) then we should additionally have $V = G$ on ∂C (the principle of continuous fit).

Finally, let us present some basic information on diffusion processes. Suppose that b, $\sigma : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}$ are fixed continuous functions, let $B = (B_t)_{t\geq 0}$ be a standard one-dimensional Brownian motion and let ξ be an $\mathcal{F}_0$-measurable random variable. We say that X is a (strong) solution to the equation

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t \tag{2.11}$$

with $X_0 = \xi$, if it satisfies the identity

$$X_t = \xi + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s) \cdot \mathrm{d}B_s.$$

Suppose in addition that b, σ satisfy the following Lipschitz conditions:

$$\begin{aligned} |b(t,x) - b(t,y)| \leq C|x-y|, \qquad & |b(t,x)| \leq L\sqrt{1+|x|^2}, \\ |\sigma(t,x) - \sigma(t,y)| \leq C|x-y|, \qquad & |\sigma(t,x)| \leq L\sqrt{1+|x|^2}. \end{aligned}$$

We have the following fact.

THEOREM 2.9. *Suppose that b, σ satisfy the above conditions and $\xi \in L^2$. Then there exists a unique strong solution to the equation* (2.11).

Suppose that b and σ do not depend on time. Then the solution to (2.11), guaranteed by the above theorem, is called the diffusion process. It is a homogeneous Markov process with the generator $\mathbb{L}_X$ acting via

$$\mathbb{L}_X f = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)$$

for all $f \in C^2(\mathbb{R})$.

We turn to examples. The first of them is left as an exercise. In what follows, we will use the notation

$$B_t^* = \sup_{0\leq s\leq t} B_s, \qquad |B|_t^* = \sup_{0\leq s\leq t} |B_s|$$

for the one- and two-sided maximal function of the Brownian motion B.

EXAMPLE 2.5 (Burkholder-Davis-Gundy inequality). For any $0 < p < \infty$ there is a finite constant c_p such that

$$c_p^{-1}\mathbb{E}\tau^{p/2} \leq (\mathbb{E}|B|_\tau^*)^p \leq c_p\mathbb{E}\tau^{p/2} \tag{2.12}$$

for all stopping times τ.

EXAMPLE 2.6. Now we will study Doob's maximal estimate for the stopped Brownian motion. We will show that if $1 < p < \infty$ and τ is an arbitrary stopping time satisfying $\tau \in L^{p/2}$, then

$$\mathbb{E}(B_\tau^*)^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|B_\tau|^p. \tag{2.13}$$

For the sake of clarity, we split the analysis into several steps.

Step 1. Formulation of the optimal stopping problem. It is easy to check that (B, B^*) is a Markov process on the state space $E = \{(x,y) \in \mathbb{R}^2 : y \geq x, y \geq 0\}$. As usual, we extend it to a Markov family and denote the collection of the corresponding probability measures by $(\mathbb{P}_{x,y})_{(x,y)\in E}$. Observe that the distribution of (B, B^*) under $\mathbb{P}_{x,y}$ is the same as that of $(x+B, (x+B)^* \vee y)$ under $\mathbb{P} = \mathbb{P}_{(0,0)}$.

Let β be a fixed positive constant and consider the stopping time

$$V(x,y) = \sup_\tau \mathbb{E}_{x,y} G(B_\tau, B_\tau^*),$$

where $G(x,y) = y^p - \beta^p|x|^p$ and the supremum is taken over all stopping times $\tau \in L^{p/2}$. We will prove that $V(0,0) = \infty$ if $\beta < (p/(p-1))^p$ and $V(0,0) \leq 0$ if $\beta = p/(p-1)$. This will clearly yield the estimate (2.13) and show that the constant $p/(p-1)$ is optimal. From now on, let us assume that we pick β such that $V(0,0)$ is finite. This implies that $V(x,y)$ is finite for all $(x,y) \in E$, $y > 0$ (we leave this as an easy exercise).

Step 2. Homogeneity. Stopping and continuation regions. For an arbitrary positive parameter λ, the process $(\tilde{B}_t)_{t\geq 0} = (\lambda B_{t/\lambda^2})_{t\geq 0}$ is a Brownian motion (under $\mathbb{P}$). Furthermore, if τ is a stopping time of B, then $\tilde{\tau} = \tau\lambda^2$ is a stopping time of $\tilde{B}$, and vice versa. Therefore, for any $(x,y) \in E$ we may write

$$\begin{aligned}
V(x,y) &= \sup_\tau \mathbb{E}_{x,y} G(B_\tau, B_\tau^*) \\
&= \sup_\tau \mathbb{E} G(x + B_\tau, (x+B)_\tau^* \vee y) \\
&= \sup_{\tilde{\tau}} \mathbb{E} G(x + \tilde{B}_{\tilde{\tau}}, (x+\tilde{B})_{\tilde{\tau}}^* \vee y) \\
&= \sup_\tau \mathbb{E} G(x + \lambda B_\tau, (x+\lambda B)_\tau^* \vee y) \\
&= \lambda^p \sup_\tau \mathbb{E} G\left(\frac{x}{\lambda} + B_\tau, \left(\frac{x}{\lambda} + B\right)_\tau^* \vee \frac{y}{\lambda}\right) = \lambda^p V\left(\frac{x}{\lambda}, \frac{y}{\lambda}\right).
\end{aligned}$$

Note that G enjoys the same homogeneity. Therefore, the continuation and the stopping regions

$$C = \{(x,y) \in E : V(x,y) = G(x,y)\}, \qquad D = \{(x,y) \in E : V(x,y) = G(x,y)\}$$

satisfy the following scaling condition: if $(x,y) \in C$, then $(\lambda x, \lambda y) \in C$ for all λ, and similarly for D.

Step 3. Free boundary problem. Based on the general theory of optimal stopping of Markov processes, we write down the system of requirements V should satisfy:

$$\begin{aligned}
\mathbb{L}_{(B,B^*)} V = 0 \qquad &\text{on } C, \\
V = G \qquad &\text{on } D, \\
V \text{ is of class } C^1 \qquad &\text{on } E.
\end{aligned} \tag{2.14}$$

This system can be simplified when one applies the homogeneity conditions established in the previous step; however, let us postpone this argument until we understand the action of the infinitesimal operator. Instead, let s briefly comment on the structure of the stopping set. We have

$$V(x,y)=\sup_\tau \mathbb{E}_{x,y}\Big\{\big[(x+B)^*_\tau\vee y\big]^p-\beta^p|x+B_\tau|^p\Big\}$$

and hence we want to maximize $\mathbb{E}\big[(x+B)^*_\tau\vee y\big]^p$, keeping $\mathbb{E}|x+B_\tau|^p$ relatively small. Note that the first expectation increases only when the process $(x+B)^*\vee y$ increases; on the other hand, the second expectation is constantly growing. Therefore, if the starting point (x,y) is very far from diagonal, then it is optimal to stop immediately: when the process $x+B$ gets to the point y (before this happens, $(x+B)^*\vee y$ is constant), the loss incurred by the term $\mathbb{E}|x+B_\tau|^p$ might be already too big. Taking into account the shape of the continuation and the stopping regions, it seems plausible to expect that they are of the form

$$C=\{(x,y)\in E\,:\,x>\gamma y\}\quad\text{ and }\quad D=\{(x,y)\in E\,:\,x\leq\gamma y\}$$

for some constant $\gamma<1$ to be found. We guess that γ needs to be positive.

Step 4. The generator. Suppose that f is a compactly supported C^2 function, defined on some neighborhood of E. The application of Itô's formula to $f(B,B^*)$ yields

$$\begin{aligned}&f(B_t,B^*_t)\\&=f(B_0,B^*_0)+\int_0^t f_x(B_s,B^*_s)\mathrm{d}B_s+\int_0^t f_y(B_s,B^*_s)\mathrm{d}B^*_s+\frac12\int_0^t f_{xx}(B_s,B^*_s)\mathrm{d}s.\end{aligned}$$

Taking the expectation with respect to $\mathbb{P}_{x,y}$ and dividing both sides by t gives

$$\begin{aligned}&\frac{\mathbb{E}_{x,y}f(B_t,B^*_t)-f(x,y)}{t}\\&\quad=\mathbb{E}_{x,y}\left(\frac1t\int_0^t f_y(B_s,B^*_s)\mathrm{d}B^*_s\right)+\mathbb{E}_{x,y}\left(\frac{1}{2t}\int_0^t f_{xx}(B_s,B^*_s)\mathrm{d}s\right).\end{aligned}$$

Observe that the left-hand side is precisely $(T_tf(x,y)-f(x,y))/t$. Letting $t\to0$, we see that

$$\mathbb{E}_{(x,y)}\left(\frac{1}{2t}\int_0^t f_{xx}(B_s,B^*_s)\mathrm{d}s\right)\to\frac12 f_{xx}(x,y)$$

(actually, the convergence can be shown to be uniform). Furthermore, observe that if $x\neq y$, then

$$\mathbb{E}_{x,y}\left(\frac1t\int_0^t f_y(B_s,B^*_s)\mathrm{d}B^*_s\right)\xrightarrow{t\to0}0.$$

On the other hand, if $x=y$, then

$$\begin{aligned}\lim_{t\to0}\mathbb{E}_{x,x}\left(\frac1t\int_0^t f_y(B_s,B^*_s)\mathrm{d}B^*_s\right)&=\lim_{t\to0}\mathbb{E}_{x,x}\left(\frac1t\int_0^t f_y(x,x)\mathrm{d}B^*_s\right)\\&=f_y(x,x)\lim_{t\to0}\frac{\mathbb{E}_{x,x}B^*_t-x}{t}\\&=f_y(x,x)\lim_{t\to0}\frac{\mathbb{E}_{x,x}|B_t|}{t}=\infty\end{aligned}$$

(here we have used the reflection principle for the Brownian motion), unless $f_y(x,x)=0$. Therefore, the domain of $\mathbb{L}_{B,B^*}$ must be contained in the class of functions f

whose partial derivative with respect to y vanish on the diagonal $x = y$. For such f, we have $L_{(B,B^*)}f = \frac{1}{2}f_{xx}$.

Step 5. Finding the candidate for the value function. We come back to the boundary value problem (2.14) and rewrite it as

$$\begin{aligned} V_{xx}(x,y) = 0 \text{ and } V_y(x,x) = 0 \qquad & \text{for all } (x,y) \in C, \\ V(x,y) = G(x,y) \qquad & \text{for all } (x,y) \in D, \\ V \text{ is of class } C^1 \qquad & \text{on } E. \end{aligned}$$

Now, let us look at the function $v(x) = V(x,1)$. It is of class C^1 and satisfies $v(x) = 1 - \beta^p x^p$ for $x \le \gamma$, $v''(x) = 0$ for $x \in (\gamma, 1)$. This allows us to compute the explicit formula of v on $(\gamma, 1)$:

$$\begin{aligned} v(x) = v(\gamma) + v'(\gamma)(x-\gamma) &= 1 - \beta^p\gamma^p - p\beta^p\gamma^{p-1}(x-\gamma) \\ &= 1 + (p-1)\beta^p\gamma^p - p\beta^p\gamma^{p-1}x. \end{aligned}$$

Consequently,

$$V(x,y) = \begin{cases} y^p - \beta^p x^p & \text{if } x \le \gamma y, \\ (1+(p-1)\beta^p\gamma^p)y^p - p\beta^p\gamma^{p-1}xy^{p-1} & \text{if } x > \gamma y. \end{cases} \tag{2.15}$$

The requirement $V_y(x,x) = 0$ is equivalent to the equation

$$p(1+(p-1)\beta^p\gamma^p) - p(p-1)\beta^p\gamma^{p-1} = 0,$$

or

$$\beta^p = \frac{1}{(p-1)\gamma^{p-1}(1-\gamma)}.$$

Now, a straightforward analysis shows that the right-hand side attains its maximal value $\left(\frac{p}{p-1}\right)^p$ for $\gamma = (p-1)/p$. Hence, if $\beta < p/(p-1)$, then there is no γ which would satisfy the above equality. This contradicts our assumption on β made at the end of Step 1. On the other hand, if we take $\beta = p/(p-1)$, then the above analysis forces us to take $\gamma = (p-1)/p$ and we arrive at the function

$$V(x,y) = \begin{cases} y^p - \left(\frac{p}{p-1}\right)^p x^p & \text{if } x \le (p-1)y/p, \\ py^{p-1}\left(y - \frac{p}{p-1}x\right) & \text{if } x > (p-1)y/p. \end{cases} \tag{2.16}$$

From the formal point of view, the above function has been found after a number of guesses and informal arguments; this is actually a *candidate* for the value function and, as such, will be denoted by $\tilde{V}$.

Step 6. Formal verification. The function $\tilde{V}$ is of class C^1 and satisfies $\tilde{V}_{xx} \le 0$, $\tilde{V}_y(x,x) = 0$. In addition, one easily checks that $V \ge G$. Consequently, for any stopping time $\tau \in L^{p/2}$ we have, by Itô's formula,

$$\tilde{V}(B_{\tau\wedge t}, B^*_{\tau\wedge t}) \le \tilde{V}(B_0, B^*_0) + \int_0^{\tau\wedge t} \tilde{V}_x(B_s, B^*_s)\mathrm{d}B_s.$$

The latter integral is an L^2 bounded martingale, so its expectation vanishes and we get $\mathbb{E}_{x,y}G(B_{\tau\wedge t}, B^*_{\tau\wedge t}) \le \mathbb{E}_{x,y}\tilde{V}(B_{\tau\wedge t}, B^*_{\tau\wedge t}) \le \tilde{V}(x,y)$. That is,

$$\mathbb{E}_{x,y}(B^*_{\tau\wedge t})^p - \left(\frac{p}{p-1}\right)^p \mathbb{E}_{x,y}|B_{\tau\wedge t}|^p \le \tilde{V}(x,y).$$

Now we let $t \to \infty$. By (2.12), we have $\mathbb{E}(|B|_\tau^*)^p < \infty$, so Lebesgue's dominated convergence theorem gives $\mathbb{E}_{x,y}G(B_\tau, B_\tau^*) \leq \tilde{V}(x,y)$. Taking the supremum over all τ as above, we obtain $V \leq \tilde{V}$. To prove the reverse estimate, consider the stopping time $\sigma = \inf\{t : B_t \leq (p-1)B_t^*/p\}$. If $x \leq (p-1)y/p$, then $\sigma = 0$ is bounded and hence $V(x,y) \geq \mathbb{E}G(B_\sigma, B_\sigma^*) = G(x,y) = \tilde{V}(x,y)$. Suppose then, that $x > (p-1)y/p$ and let γ' be a number slightly bigger than $(p-1)/p$, but so that $x > \gamma' y$. Consider the function $\hat{V}$ as (2.15), corresponding to the parameters γ' and $\beta = p/(p-1)$, and modify the stopping time σ slightly: set $\sigma = \inf\{t : B_t \leq \gamma' B_t^*\}$. We easily check that $\tilde{V} \geq G$ and $\tilde{V}_y(x,x) \geq 0$, so the application of Itô's formula, with $\tau = \sigma \wedge t$, gives

$$\hat{V}(B_\tau, B_\tau^*) \geq \hat{V}(B_0, B_0^*) + \int_0^\tau \hat{V}_x(B_s, B_s^*)\mathrm{d}B_s,$$

since on the time interval $[0,\tau]$, the process (B, B^*) evolves along the set on which $\hat{V}_{xx} = 0$. Consequently, $\mathbb{E}_{x,y}\hat{V}(B_\tau, B_\tau^*) \geq \hat{V}(x,y)$. Now, observe that $\hat{V}(B_\tau, B_\tau^*)$ is negative; thus, the application of Fatou's lemma (and letting $t \to \infty$) yields

$$\mathbb{E}_{x,y}\hat{V}(B_\sigma, B_\sigma^*) \geq \hat{V}(x,y).$$

By the very definition of σ and $\hat{V}$, we have $\hat{V}(B_\sigma, B_\sigma^*) = G(B_\sigma, B_\sigma^*)$ and hence

$$\mathbb{E}_{x,y}(B_\sigma^*)^p - \left(\frac{p}{p-1}\right)^p \mathbb{E}_{x,y}|B_\sigma|^p \geq \hat{V}(x,y).$$

But $B_\sigma = \gamma' B_\sigma^*$; since $\gamma' > (p-1)/p$, the above identity shows that $|B|_\sigma^* = B_\sigma^* \in L^p$, and hence (2.12) proves that $\sigma \in L^{p/2}$ (with respect to $\mathbb{P}_{x,y}$). Now, if γ' is chosen sufficiently close to $(p-1)/p$, then $\hat{V}(x,y)$ can be made arbitrarily close to $\tilde{V}(x,y)$. Consequently, we see that for any $\varepsilon > 0$ there is a stopping time $\tau \in L^{p/2}$ for which

$$\mathbb{E}_{x,y}(B_\tau^*)^p - \left(\frac{p}{p-1}\right)^p \mathbb{E}_{x,y}|B_\tau|^p \geq \tilde{V}(x,y) - \varepsilon.$$

This gives $V(x,y) \geq \tilde{V}(x,y)$ and completes the proof.

Remark 2.4. As a by-product of the above reasoning, we have obtained that the stopping time

$$\tau = \inf\{t > 0 : B_t < \gamma B_t^*\}$$

is $L^{p/2}$-integrable (with respect to $\mathbb{P}_{1,1}$, or any other measure $\mathbb{P}_{x,x}$, $x > 0$) if and only if $\gamma > (p-1)/p$.

It is also worth mentioning that the above analysis immediately leads to the solution of the corresponding bound for the *two-sided* maximal function. Suppose that we are interested in the best constant C_p in the estimate

$$\mathbb{E}(|B|_\tau^*)^p \leq C_p \mathbb{E}|B_\tau|^p,$$

where the supremum is taken over all $\tau \in L^{p/2}$. Since $|B|_\tau^* \geq |B_\tau^*|$, we must have $C_p \geq \left(\frac{p}{p-1}\right)^p$, by the above considerations. However, we have equality. To see this, introduce the associated optimal stopping problem

$$\mathcal{V}(x,y) = \mathbb{E}_{x,y}\left[(|B|_\tau^*)^p - \left(\frac{p}{p-1}\right)^p |B_\tau|^p\right],$$

on the underlying state space $E = \{(x,y) : |x| \leq y\}$. Let V be the value function of the previous problem, defined via (2.16). The function $\mathcal{V}$ is symmetric with respect to x: $\mathcal{V}(x,y)=\mathcal{V}(-x,y)$, which can be easily checked by the symmetry of the Brownian motion. Furthermore, we have $\mathcal{V}(x,y)\geq V(x,y)$ on E, as we have noted above. Consequently, we also have

$$\mathcal{V}(x,y)\geq \max\{V(x,y),V(-x,y)\}=V(|x|,y).$$

We check directly that

$$V(|x|,y))=\begin{cases} y^p-\left(\frac{p}{p-1}\right)^p|x|^p & \text{if } |x|\leq (p-1)y/p,\\ py^{p-1}\left(y-\frac{p}{p-1}|x|\right) & \text{if } |x|>(p-1)y/p \end{cases}$$

is excessive. Thus, by the previous estimate, we must have $\mathcal{V}=V$ and the result follows.

EXAMPLE 2.7. Now we will discuss a procedure which is called *the method of measure change.* Roughly speaking, the purpose of this technique is to replace the underlying probability measure $\mathbb{P}$ by a different measure $\tilde{\mathbb{P}}$, under which the problem is less-dimensional. Suppose that X is a time-homogeneous diffusion process which satisfies the SPDE

$$\mathrm{d}X_t=\rho(X_t)\mathrm{d}t+\sigma(X_t)\mathrm{d}B_t$$

with the initial condition $X_0=x$. Consider the optimal stopping problem

$$V=\sup_{0\leq\tau\leq T}\mathbb{E}G(I_\tau,X_\tau,X^*_\tau), \tag{2.17}$$

where $I_t=\int_0^t X_s\mathrm{d}s$ and X^* is the (one-sided) maximal function of X; here G is a measurable function satisfying appropriate boundedness conditions (which guarantee the existence of the expectations above). The first step is to introduce the auxiliary exponential process

$$\mathcal{E}_t=\exp\left(\int_0^t H_s\mathrm{d}B_s-\frac{1}{2}\int_0^t H_s^2\mathrm{d}s\right),\qquad t\geq 0,$$

where H is an appropriate predictable process, so that $\mathcal{E}$ is a martingale: for example, this holds if H satsifies the Novikov condition $\mathbb{E}\exp\left(\frac{1}{2}\int_0^T H_s^2\mathrm{d}s\right)<\infty$. We consider the new probability measure

$$\tilde{\mathbb{P}}(A)=\int_A\mathcal{E}_T\mathrm{d}\mathbb{P}$$

and rewrite the problem (2.17) as follows. Set $Z=(I,X,X^*)$ and use the identities

$$\mathbb{E}G(Z_\tau)=\mathbb{E}\mathcal{E}_\tau\frac{G(Z_\tau)}{\mathcal{E}_\tau}=\mathbb{E}\mathcal{E}_T\frac{G(Z_\tau)}{\mathcal{E}_\tau}=\tilde{\mathbb{E}}\left(\frac{G(Z_\tau)}{\mathcal{E}_\tau}\right).$$

If the ratio $G(Z_t)/\mathcal{E}_t$, $t\geq 0$, can be rewritten as $\tilde{G}(Y_t)$ for some (strong) Markov process Y, then (2.17) becomes

$$V=\sup_{0\leq\tau\leq T}\tilde{\mathbb{E}}\tilde{G}(Y_\tau)$$

(the change of measure does not affect stopping times). Note that the latter problem involves stopping of a one-dimensional process: this is the aforementioned reduction of the difficulty.

It should be emphasized that the above approach requires quite restrictive environment: the key Markov property seems to hold only for special choices of the diffusion processes X. The reason for the introduction of the exponential process

$\mathcal{E}$ and the associated measure $\tilde{\mathbb{P}}$ stems from Girsanov theorem: under this new measure, the process $\left(B_t - \int_0^t H_s \mathrm{d}s\right)_{t\geq 0}$ is a Brownian motion. This in particular implies the independence of the appropriate increments and might lead to the Markov property of the process Y mentioned above. We should also stress here that the method works in the case of infinite horizon (i.e., for $T = \infty$), with no essential changes.

We will illustrate the above method on the following specific example. Consider the optimal stopping problem

$$V = \sup_\tau \mathbb{E}\big(e^{-2\tau}(I_\tau - X_\tau)\big),$$

where the supremum is taken over all stopping times τ. Here X is the geometric Brownian motion satisfying

$$\mathrm{d}X_t = X_t \mathrm{d}t + \sqrt{3} X_t \mathrm{d}B_t, \qquad X_0 = 1.$$

The unique strong solution to this equation is given by $X_t = \exp\left(\sqrt{3}B_t - t/2\right) = e^t \mathcal{E}_t$, where $\mathcal{E} = \left(\exp(\sqrt{3}B_t - 3t/2)\right)_{t\geq 0}$ is an exponential martingale. Observe that X takes only positive values (in particular, $X_t \neq 0$ almost surely for all t). In addition, note that the above stopping problem is three-dimensional: there are three processes involved, X, I and time.

Step 1. Change of measure. Now we will apply the above procedure. We write

$$\mathbb{E}\big(e^{-2\tau}(I_\tau - X_\tau)\big) = \mathbb{E}\left[X_\tau e^{-2\tau}\left(\frac{I_\tau}{X_\tau} - 1\right)\right] = \tilde{\mathbb{E}}\left[e^{-\tau}\left(\frac{I_\tau}{X_\tau} - 1\right)\right],$$

where the latter expectation is taken with respect to the new measure $\mathrm{d}\tilde{\mathbb{P}} = \mathcal{E}_T \mathrm{d}\mathbb{P}$. Now we will show that $Y = I/X$ is a strong Markov process with respect to $\tilde{\mathbb{P}}$: we have

$$\begin{aligned}
&Y_{t+h} \\
&= \frac{I_{t+h}}{X_{t+h}} \\
&= \frac{X_t}{X_{t+h}} \cdot \frac{\int_0^t X_s \mathrm{d}s + \int_t^{t+h} X_s \mathrm{d}s}{X_t} \\
&= \frac{1}{\exp(\sqrt{3}(B_{t+h} - B_t) - h/2)} \left(\frac{I_t}{X_t} + \int_t^{t+h} \frac{X_s}{X_t} \mathrm{d}s\right) \\
&= \frac{1}{\exp(\sqrt{3}(B_{t+h} - B_t) - h/2)} \left(\frac{I_t}{X_t} + \int_t^{t+h} \exp(\sqrt{3}(B_s - B_t) - (s-t)/2) \mathrm{d}s\right).
\end{aligned}$$

Now, by Girsanov theorem, the process $(B_t - \sqrt{3}t)_{t\geq 0}$ is a Brownian motion under $\tilde{\mathbb{P}}$. In particular, this shows that the above expressions involving the differences of B are independent of $Y_t = I_t/X_t$. Consequently, we have $\tilde{\mathbb{E}}(Y_{t+h}|\mathcal{F}_t) = \tilde{\mathbb{E}}(Y_{t+h}|Y_t)$, which gives the desired Markov property. In addition, observe that the Markov process is time-homogeneous. Actually, the same argument, with t replaced by an arbitrary stopping time, gives the strong Markov property of Y. Note that the resulting problem

$$V = \sup_\tau \tilde{\mathbb{E}}\big(e^{-\tau}(Y_\tau - 1)\big)$$

is two-dimensional.

Step 2. Now we will compute the infinitesimal operator of Y. By Itô's formula, we have

$$\mathrm{d}Y_t = \mathrm{d}t - \frac{Y_t}{X_t}\mathrm{d}X_t + \frac{Y_t}{X_t^2}\mathrm{d}\langle X\rangle_t = (1-Y_t)\mathrm{d}t - \sqrt{3}Y_t\mathrm{d}(B_t - \sqrt{3}t).$$

As a by-product, we obtain an alternative proof of the fact that Y is a strong time-homogeneous Markov process: by the above equation, we see that Y is a diffusion with coefficients not depending on t.

Now, since $(B_t - \sqrt{3}t)_{t\geq 0}$ is a $\tilde{\mathbb{P}}$-Brownian motion (by Girsanov theorem), the infinitesimal generator of Y acts via

$$\mathbb{L}_Y u(y) = (1-y)u'(y) + \frac{3}{2}y^2 u''(y).$$

Consequently, the generator of the space-time process (t, Y) is

$$\mathbb{L}_{t,Y} v(t,y) = v_t(t,y) + (1-y)v_y(t,y) + \frac{3}{2}y^2 v_{yy}(t,y).$$

Step 3. We are ready for the study of the optimal stopping problem

$$V = \sup_\tau \tilde{\mathbb{E}}\big(e^{-\tau}(Y_\tau - I)\big) = \sup_\tau \tilde{\mathbb{E}}G(\tau, Y_\tau),$$

where $G(t,y) = e^{-t}(y-1)$. As usual, we extend it to the arbitrary starting point of the state space $E = [0,\infty)\times(0,\infty)$. Namely, we consider the appropriate family $(\tilde{\mathbb{P}}_{t,y})_{(t,y)\in E}$ of probability measures and, for $(t,y)\in E$, we set

$$V(t,y) = \sup_\tau \tilde{\mathbb{E}}_{t,y}\big(e^{-\tau}(Y_\tau - 1)\big) = \sup_\tau \tilde{\mathbb{E}}\big(e^{-t-\tau}(Y_\tau - 1)\big).$$

Directly from the latter formula, we see that V satisfies the homogeneity condition $V(t,y) = e^{-t}V(0,y)$. So, if we introduce the continuation and stopping regions

$$C = \{(t,y) : V(t,y) \geq G(t,y)\}, \quad D = \{(t,y) : V(t,y) = G(t,y)\},$$

then the above homogeneity property of V gives that $(t,y)\in C$ if and only if $(0,y)\in C$.

Step 4. We write down the corresponding free-boundary problem:

$$\mathbb{L}_{t,Y}V = 0 \text{ on } C,$$
$$V = G \text{ on } D,$$
$$V \text{ is of class } C^1 \text{ on } E.$$

The first condition becomes, under the substitution $v(y) = V(0,y)$,

$$-v(y) + (1-y)v'(y) + \frac{3y^2}{2}v''(y) = 0. \tag{2.18}$$

Let us search for the solution in the class of power series $v(y) = \sum_{n=0}^\infty a_n y^n$. Plugging this into (2.18), we see that the coefficient standing in front of y^n is equal to $-a_n - na_n + (n+1)a_{n+1} + \frac{n(n-1)}{2}a_n$. This must be equal to zero, which is equivalent to saying that

$$(n+1)a_{n+1} = -\left(\frac{3}{2}n(n-1) - (n+1)\right)a_n \qquad \text{for all } n \geq 0.$$

We easily see that this holds if and only if $a_0 = a_1 = a_2$ and $a_3 = a_4 = \ldots = 0$. In other words, we obtain the one-parameter family of solutions of the form $a(1+x+x^2)$, $a\in\mathbb{R}$.

Step 5. Guessing the candidate for V. Putting all the above facts together, it seems plausible to conjecture that C and D are quadrants and $\partial C \cap \partial D = [0,\infty) \times b$ for some $b > 0$. Since V is of class C^1, we must have $V(t,b) = G(t,b)$ and $V_y(t,b) = G_y(t,b)$ for all $t \geq 0$. This is equivalent to the system of equations

$$\begin{cases} a(1+b+b^2) = b-1, \\ a(1+b) = 1, \end{cases}$$

which is solved by $a = (3 \pm 2\sqrt{3})^{-1}$ and $b = 1 \pm \sqrt{3}$. We have to take the plus sign: otherwise a would be negative and the majorization $V \geq G$ would not hold. This gives us the "splitting level" b and it remains to specify where C and D lie. Note that on D, the function V must be excessive, i.e., we must have $\mathbb{L}_{t,Y} G \leq 0$. The latter is equivalent to $y \geq 2$; thus, finally we have

$$C = \{(t,y) : y < 1+\sqrt{3}\}, \qquad D = \{(t,y) : y \geq 1+\sqrt{3}\}$$

and the candidate for V is

$$\tilde{V}(t,y) = \begin{cases} (3+2\sqrt{3})^{-1} e^{-t}(1+y+y^2/2) & \text{on } C, \\ e^{-t}(y-1) & \text{on } D. \end{cases}$$

Step 6. Formal verification. The function $\tilde{V}$ is of class C^1, it is excessive, majorizes G and is bounded from below by 1. Thus we have $\tilde{V} \geq V$ (which can be proved by Itô's formula and Fatou's lemma). To show the reverse bound $\tilde{V}(t,y) \leq V(t,y)$, we may restrict ourselves to $y < 1+\sqrt{3}$. We will check that $\tilde{V}(t,y)$ can be attained by $\tilde{\mathbb{E}}_{t,y} G(\tau, Y_\tau)$ for an appropriate choice of τ. Motivated by the general theory and the above analysis, we consider the stopping time

$$\tau = \inf\{t : (t, Y_t) \in D\} = \inf\{t : Y_t \geq 1+\sqrt{3}\}.$$

Since $L_{t,Y}\tilde{V} = 0$ on C, we have, by Itô's formula,

$$\tilde{V}(\tau \wedge T, Y_{\tau\wedge T}) = \tilde{V}(t, Y_t) + \int_t^{\tau\wedge T} \tilde{V}_y(s, Y_s) \cdot (-\sqrt{3} Y_s) \mathrm{d}(B_s - \sqrt{3}s),$$

for any $t < T$. On the time interval $[t, \tau \wedge T]$, the process $\tilde{V}_y(s,Y_s)Y_s$ is bounded, so the expectation of the stochastic integral is zero and hence

$$\tilde{\mathbb{E}}_{t,y} \tilde{V}(\tau \wedge T, Y_{\tau\wedge T}) = \tilde{V}(t,y).$$

However, $\tilde{V}(\tau \wedge T, Y_{\tau\wedge T})$ is bounded; therefore, letting $T \to \infty$ and using Lebesgue's dominated convergence theorem, we get

$$\tilde{\mathbb{E}} G(\tau, Y_\tau) = \tilde{\mathbb{E}} \tilde{V}(\tau, Y_\tau) = \tilde{V}(t,y).$$

This gives $V(t,y) \geq \tilde{V}(t,y)$, by the very definition of V. The proof is complete.

As we see, the optimal strategy is to wait until the ratio I_t/X_t reaches (or exceeds) $1+\sqrt{3}$; in other words, we wait until $X_t \leq I_t/(1+\sqrt{3})$. This is not a very intuitive procedure, if we look at the initial formulation of the problem.

EXAMPLE 2.8. Suppose that $p > 2$ is a fixed constant. The purpose of the example is to identify the best constant c_p in the estimate

$$\mathbb{P}(\tau \geq 1) \leq c_p \mathbb{E}|B_\tau|^p,$$

to be valid for all bounded stopping times τ. To put the problem in the right framework, introduce the gain function $G : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by $G(x,t) = 1_{\{t \geq 1\}} - |x|^p$ and the value function $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by

$$V(x,t) = \sup \mathbb{E} G(x + B_\tau, t + \tau), \tag{2.19}$$

where the supremum is taken over all adapted bounded stopping times τ.

Clearly, we have $V(x,t) = 1 - |x|^p$ for $t \geq 1$, and if $t < 1$, we may restrict ourselves to stopping times τ which are bounded by $1 - t$. For the sake of clarity and convenience, we have decided to split the remaining reasoning into eleven intermediate steps.

Step 1. First we will show that the function V is continuous. Pick $t < 1$, x, $y \in \mathbb{R}$ and let τ be an arbitrary stopping time bounded by $1 - t$. The process $(|x + B_s|^{p-1})_{s \geq 0}$ is a submartingale, so by mean value property,

$$\mathbb{E}|y + B_\tau|^p - \mathbb{E}|x + B_\tau|^p \leq p\mathbb{E}|y + B_\tau|^{p-1}|y - x| \leq p|y - x| \cdot \mathbb{E}|y + B_{1-t}|^{p-1}$$

and thus

$$V(x,t) \leq V(y,t) + p|y - x| \cdot \mathbb{E}|y + B_{1-t}|^{p-1},$$

by the definition of V. Therefore, by symmetry,

$$|V(x,t) - V(y,t)| \leq p|y - x| \mathbb{E}(|x| + |y| + |B_{1-t}|)^{p-1}, \tag{2.20}$$

so for each fixed t, the function $x \mapsto V(x,t)$ is locally Lipschitz. Next, fix $s < t \leq 1$. For a given x, the function $u \mapsto G(x,u)$ is nondecreasing and hence V also has this property: thus, $V(x,s) \leq V(x,t)$. Furthermore, if τ is an arbitrary stopping time bounded by $1 - t$ and we put $\sigma = \tau + t - s$, then

$$\begin{aligned}
&\mathbb{P}(\tau \geq 1 - t) - \mathbb{E}|x + B_\tau|^p \\
&\quad = \mathbb{P}(\sigma \geq 1 - s) - \mathbb{E}|x + B_\sigma|^p + \mathbb{E}\Big[|x + B_{\tau+t-s}|^p - |x + B_\tau|^p\Big] \\
&\quad \leq V(x,s) + p\mathbb{E}|x + B_\tau|^{p-1}|B_{\tau+t-s} - B_\tau| \\
&\quad \leq V(x,s) + p\mathbb{E}|x + B_\tau|^{p-1}\mathbb{E}|B_{\tau+t-s} - B_\tau| \\
&\quad \leq V(x,s) + \sqrt{\frac{2(t-s)}{\pi}} p\mathbb{E}|x + B_{1-t}|^{p-1}.
\end{aligned}$$

Therefore, we have

$$0 \leq V(x,t) - V(x,s) \leq \sqrt{\frac{2(t-s)}{\pi}} p\mathbb{E}|x + B_{1-t}|^{p-1},$$

which combined with (2.20) yields the continuity of V.

Step 2. Let us provide an abstract formula for the optimal stopping time in (2.19). Introduce the continuation set C and the stopping region D by

$$C = \{(x,t) \in \mathbb{R} \times \mathbb{R} : V(x,t) > G(x,t)\}$$

and

$$D = \{(x,t) \in \mathbb{R} \times \mathbb{R} : V(x,t) = G(x,t)\}.$$

The gain function G is upper semicontinuous and the function V is lower semicontinuous (since it is continuous, in view of Step 1). Therefore, by the general theory of optimal stopping (cf. Corollary 2.9 in Peskir and Shiryaev [**18**]), for a given state $(x,t) \in \mathbb{R} \times \mathbb{R}$, the stopping time

$$\tau_D = \inf\{s \geq 0 : (t + s, x + B_s) \in D\}$$

is optimal in (2.19). Now, standard arguments based on the strong Markov property and classic results from PDEs (see e.g. Chapter III in [**18**]) show that V is of class $C^{2,1}$ on C and satisfies the heat equation

$$V_t + \frac{1}{2}V_{xx} = 0 \tag{2.21}$$

on this set. These facts will be freely used in the considerations below.

Step 3. Let us provide some insight into the shape of C and D. By the symmetry of Brownian motion, we immediately get that C and D are symmetric with respect to y-axis. By the upper semicontinuity of G and the continuity of V, we conclude that C is open and D is closed. As we have already observed, $\mathbb{R} \times [1, \infty) \subset D$. Now, for any $x \in \mathbb{R}$ we have $\lim_{t\uparrow 1}(1 - |x + B_{1-t}|^p) = 1 - |x|^p > -|x|^p$; thus, for a given x, we have $V(x, t) > G(x, t)$ for t sufficiently close to 1, and hence $(x, t) \in C$. Next, take $0 \leq x < y$ and a bounded stopping time τ. Put $2\delta = y - x$ and set $\sigma = \inf\{s : B_s = -x - \delta\}$. Consider the Brownian motion reflected at time σ, given by

$$W_s = \begin{cases} B_s & \text{if } s < \sigma, \\ 2B_\sigma - B_s & \text{if } s \geq \sigma. \end{cases} \tag{2.22}$$

It is easy to check the majorization

$$|x + B_s| \leq |y + W_s| \qquad \text{for all } s. \tag{2.23}$$

Indeed, if $s \leq \sigma$, then $B_s \geq -x - \delta$, so

$$|y + W_s| = |x + 2\delta + B_s| = x + 2\delta + B_s \geq |x + B_s|.$$

On the other hand, if $s > \sigma$, then $y + W_s = x + 2\delta - 2x - 2\delta - B_s = -x - B_s$ and $|y + W_s| = |x + B_s|$. Thus, (2.23) follows. Now, if τ is a stopping time bounded by $1 - t$, then Itô's formula gives

$$\begin{aligned} V(x,t) - G(x,t) &\geq \mathbb{P}(t + \tau \geq 1) - \mathbb{E}|x + B_\tau|^p - G(x,t) \\ &= \mathbb{P}(t + \tau \geq 1) - \mathbb{E}|x + B_\tau|^p + |x|^p \\ &= \mathbb{P}(t + \tau \geq 1) - \frac{p(p-1)}{2}\mathbb{E}\int_0^\tau |x + B_s|^{p-2}\mathrm{d}s \\ &\geq \mathbb{P}(t + \tau \geq 1) - \frac{p(p-1)}{2}\mathbb{E}\int_0^\tau |y + W_s|^{p-2}\mathrm{d}s \\ &= \mathbb{P}(t + \tau \geq 1) - \mathbb{E}|y + W_\tau|^p + |y|^p. \end{aligned}$$

Hence, taking supremum over τ yields

$$V(x,t) - G(x,t) \geq V(y,t) - G(y,t).$$

Therefore, C enjoys the following: if $(y, t) \in C$, then $[-y, y] \times \{t\} \subset C$.

To prove another geometrical property of the continuation set, note that for a fixed x, the function $t \mapsto V(x, t)$ is nondecreasing (we have already pointed this out above). Hence the function $t \mapsto V(x, t) - G(x, t)$, $t \in (-\infty, 1)$, also satisfies this condition; thus we may conclude that if $(x, t) \in C$, then the whole line segment $\{x\} \times [t, 1)$ is contained in C.

Now we will show that if $t < 1$ is fixed, then $(x, t) \in D$ for sufficiently large $|x|$. Otherwise, by the previous facts, we would have $\mathbb{R} \times [t, 1) \subset C$ and hence $\tau \equiv 1 - t$

would be optimal for (x,t). But, by Itô's formula,

$$\mathbb{E}|x+B_{1-t}|^p - |x|^p = \frac{p(p-1)}{2}\int_0^{1-t}\mathbb{E}|x+B_s|^{p-2}\mathrm{d}s \to \infty$$

as $|x|\to\infty$. So, if $|x|$ is sufficiently large, then $V(x,t) = 1-\mathbb{E}|x+B_{1-t}|^p < -|x|^p = G(x,t)$, a contradiction.

The final insight into the structure of C is gained with the use of Burkholder-Davis-Gundy inequality. Namely, for any bounded stopping time τ we have

$$\mathbb{P}(t+\tau\geq 1) \leq \mathbb{P}(\tau \geq 1-t) \leq \frac{||\tau^{1/2}||_p^p}{(1-t)^{p/2}} \leq \frac{C_p^p\mathbb{E}|B_\tau|^p}{(1-t)^{p/2}}.$$

Therefore, if $1-t > C_p^2$, then $\mathbb{P}(t+\tau\geq 1) - \mathbb{E}|B_\tau|^p \leq 0 = G(0,t)$ and hence $(0,t)\in D$ (which implies that the whole horizontal line $\mathbb{R}\times\{t\}$ is contained in D, in view of the above properties). The combination of the facts proved above leads to the following statement: there is a nondecreasing function $b:(-\infty,1)\to[0,\infty)$ which vanishes on some interval $(-\infty,t_0]$ and tends to ∞ as $t\uparrow 1$, such that

$$C = \{(x,t)\in\mathbb{R}\times(-\infty,1) : |x| < b(t)\}.$$

See Figure 2.1 below.

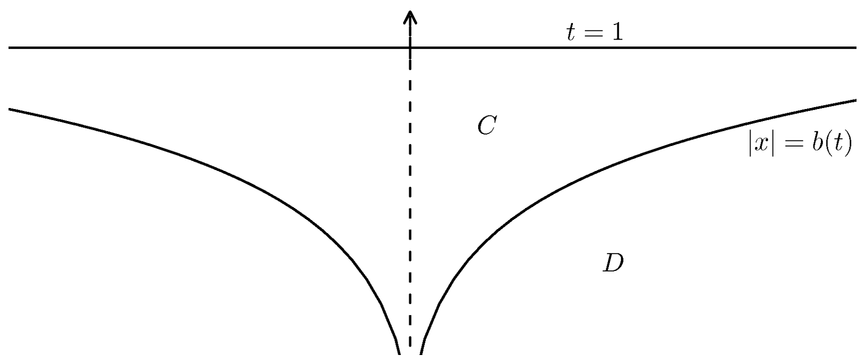

FIGURE 2.1. The continuation set lies between the curves $|x| = b(t)$ and $t=1$.

Step 4. Our next task is to prove that the function b is continuous. Let us first focus on the left-continuity of b. For this, fix $t<1$ and consider a sequence $(t_n)_{n\geq 0}$ which increases to t as $n\to\infty$. Since b is nondecreasing, the limit $b(t-) = \lim_{n\to\infty} b(t_n)$ exists. Because $(b(t_n),t_n)$ belongs to D for all n and D is closed, it follows that $(b(t-),t)$ belongs to D. This implies $b(t-)\geq b(t)$, and we get the reverse bound by the monotonicity of b. Consequently, b is left-continuous, as claimed.

We turn to the right-continuity of b. Assume on contrary that there is $t>0$ such that $b(t) < b(t+)$ and pick $x,\, y\in(b(t),b(t+))$ such that $x<y$. Define the stopping time

$$\tau' = \inf\{s>0 : (y+B_s, t+s)\in D\}$$

(the difference between τ' and τ_D lies in the fact that in the above infimum we consider *positive* s). The process $(y+B_s,t+s)_{0<s<\tau'}$ takes values in C, so by Itô's formula and (2.21), we have

$$\begin{aligned} 0 = V(y,t) + |y|^p &= \mathbb{E}V(y + B_{\tau'}, t + \tau') - G(y,t) \\ &= \mathbb{E}G(y + B_{\tau'}, t + \tau') - G(y,t) \\ &= \mathbb{P}(t + \tau' \geq 1) - \frac{p(p-1)}{2}\mathbb{E}\int_0^{\tau'} |y + B_s|^{p-2}\mathrm{d}s. \end{aligned} \tag{2.24}$$

Now we repeat the coupling argument from Step 3: let W be the reflected Brownian motion, corresponding to $x < y$, given by (2.22). Directly from its construction and the monotonicity of b, we infer that the process $(x + W_s, t + s)_{0<s<\tau'}$ takes values in C. Indeed, if $(x + W_s, t + s) \in D$ for some $s \in (0, \tau')$, then $\sigma > s$, since otherwise we would have $|x + W_s| = |y + B_s|$, a contradiction with the definition of τ'. But if $\sigma > s$, then $-b(t+s) < -\delta < x + W_s < |y + B_s| < b(t+s)$, which again makes the condition $(x + W_s, t + s) \in D$ impossible. Thus, $(x + W_s, t + s) \in C$ for $s \in (0, \tau)$; furthermore, we have $x + B_{\tau'} > 0$ with positive probability, which implies that $\mathbb{P}((x + W_{\tau'}, t + \tau') \in C) > 0$. Therefore, by Itô's formula,

$$\begin{aligned} 0 = V(x,t) + |x|^p &= \mathbb{E}V(x + W_{\tau'}, t + \tau') - G(y,t) \\ &> \mathbb{E}G(x + W_{\tau'}, t + \tau') - G(y,t) \\ &= \mathbb{P}(t + \tau' \geq 1) - \frac{p(p-1)}{2}\mathbb{E}\int_0^{\tau'} |x + W_s|^{p-2}\mathrm{d}s. \end{aligned}$$

Combining this with (2.24) yields

$$\mathbb{E}\int_0^{\tau'} |x + W_s|^{p-2}\mathrm{d}s > \mathbb{E}\int_0^{\tau'} |y + B_s|^{p-2}\mathrm{d}s,$$

which contradicts the inequality $|x + W_s| \leq |y + B_s|$ which can be proved as in Step 3. This gives the desired continuity of b.

Step 5. Now we will show the following smooth-fit property: for each $t < 1$ the function $x \mapsto V(x,t)$ is differentiable at the point $b(t)$ and satisfies $V_x(b(t),t) = G_x(b(t),t)$. Clearly, it suffices to compare the left derivatives of V and G. Since $V(b(t),t) = G(b(t),t)$, we may write

$$\frac{V(b(t),t) - V(b(t) - \varepsilon, t)}{\varepsilon} \leq \frac{G(b(t),t) - G(b(t) - \varepsilon, t)}{\varepsilon}$$

for all $\varepsilon > 0$ and hence

$$\limsup_{\varepsilon \downarrow 0} \frac{V(b(t),t) - V(b(t) - \varepsilon, t)}{\varepsilon} \leq G_x(b(t),t).$$

Let $\tau_\varepsilon = \tau_D(b(t) - \varepsilon, t)$ be optimal for $V(b(t) - \varepsilon, t)$. Then by the mean value theorem we have

$$\begin{aligned} &\frac{V(b(t),t) - V(b(t) - \varepsilon, t)}{\varepsilon} \\ &\quad \geq \frac{1}{\varepsilon}\Big(\mathbb{E}G(b(t) + B_{\tau_\varepsilon}, t + \tau_\varepsilon) - \mathbb{E}G(b(t) - \varepsilon + B_{\tau_\varepsilon}, t + \tau_\varepsilon)\Big) \\ &\quad = \mathbb{E}G_x(\xi_\varepsilon, t + \tau_\varepsilon), \end{aligned}$$

where ξ_ε lies between $b(t) - \varepsilon + B_{\tau_\varepsilon}$ and $b(t) + B_{\tau_\varepsilon}$. Since $t \mapsto b(t)$ is nondecreasing and $t \mapsto \lambda t$ is a lower function for B at $0+$ for every $\lambda \in \mathbb{R}$, one easily proves that

$\tau_\varepsilon \to 0$ as $\varepsilon \downarrow 0$. Consequently, $\xi_\varepsilon \to b(t)$ and $G_x(\xi_\varepsilon, t+\tau_\varepsilon) \to G_x(b(t),t)$ as $\varepsilon \downarrow 0$. In addition,

$$|G_x(\xi_\varepsilon, t+\tau_\varepsilon)| \leq p\big(|b(t)+B_{\tau_\varepsilon}|+\varepsilon\big)^{p-1} \leq p\big(\sup_{0\leq s\leq 1-t}|b(t)+B_s|+\varepsilon\big)^{p-1}$$

and the latter variable is integrable. Therefore, we may conclude that

$$\liminf_{\varepsilon\downarrow 0} \frac{V(b(t),t)-V(b(t)-\varepsilon,t)}{\varepsilon} \geq G_x(b(t),t)$$

by Lebesgue's dominated convergence theorem. This proves the desired smoothness of V.

Step 6. Now we will apply Itô's formula to V. Although V does not have to be of class C^2, we have the smooth-fit condition and hence the application is possible, by a local time-space formula of Peskir. As the result, for $t<1$ and $s \leq 1-t$,

$$V(x+B_s, t+s) = V(x,t) + I + II, \tag{2.25}$$

where

$$I = \int_0^s V_x(x+B_u, t+u)\mathrm{d}B_u,$$
$$II = \int_0^s \left(V_t + \frac{1}{2}V_{xx}\right)(x+B_u, t+u)1_{\{|x+B_u|\neq b(t+u)\}}\mathrm{d}u.$$

Let us analyze the terms I and II. The process in I is a local martingale; in fact, the partial derivative V_x is locally bounded, so the process is actually a true martingale and hence the stochastic integral has mean zero. The term II can be computed directly:

$$II = -\frac{p(p-1)}{2}\int_0^s |x+B_u|^{p-2}1_{\{|x+B_u|>b(t+u)\}}\mathrm{d}u.$$

Consequently, taking $s=1-t$ and integrating both sides of (2.25), we get

$$1-\mathbb{E}|x+B_{1-t}|^p = V(x,t) - \frac{p(p-1)}{2}\int_0^{1-t}\mathbb{E}|x+B_u|^{p-2}1_{\{|x+B_u|>b(t+u)\}}\mathrm{d}u.$$

Now take $x=b(t)$. Then $V(x,t)=-b(t)^p$ and by Itô's formula, we get

$$\int_0^{1-t}\mathbb{E}|x+B_u|^{p-2}1_{\{|x+B_u|\leq b(t+u)\}}\mathrm{d}u = \frac{2}{p(p-1)}.$$

Since $x+B_u \sim \mathcal{N}(x,u)$, we obtain the following functional equation for b:

$$\int_t^1 \int_{|r|\leq b(u)} \frac{|r|^{p-2}}{\sqrt{2\pi(u-t)}}\exp\left(-\frac{(r-b(t))^2}{2(u-t)}\right)\mathrm{d}r\mathrm{d}u = \frac{2}{p(p-1)}. \tag{2.26}$$

Step 7. The purpose of the four steps below is to show that the solution to (2.26) is unique. So, suppose that $c:(-\infty,1)\to[0,\infty)$ is a nondecreasing continuous function satisfying the above functional equation. Note that $\lim_{t\uparrow 1} c(t)=\infty$; otherwise, the left-hand side of (2.26) would be bounded from above by an expression of the form

$$\int_t^1 \frac{\alpha_1}{\sqrt{u-t}}\exp\left(-\frac{\alpha_2}{2(u-t)}\right)\mathrm{d}u$$

(for some positive constants α_1, α_2), which converges to 0 as $t\uparrow 1$.

Motivated by the above considerations leading to (2.26), we introduce the auxiliary function $U^c : \mathbb{R} \times (-\infty, 1] \to \mathbb{R}$ given by

$$U^c(x,t) = \mathbb{E}G(x + B_{1-t}, 1) - \frac{p(p-1)}{2} \int_0^{1-t} |x + B_u|^{p-2} 1_{\{|x+B_u|>c(t+u)\}} \mathrm{d}u.$$

The assumption that c solves (2.26) is equivalent to saying that

$$\int_0^{1-t} \mathbb{E}|x + B_u|^{p-2} 1_{\{|x+B_u|\le c(t+u)\}} \mathrm{d}u = \frac{2}{p(p-1)},$$

which, in turn, implies that $U^c(c(t), t) = G(c(t), t)$ for all t. In the remainder of this step we shall prove that $U^c(x, t) = G(x, t)$ provided $|x| \ge c(t)$; this will be accomplished with the use of the following martingale methods.

Observe that if X is a Markov process and we set $F(x, t) = \mathbb{E}_x G(X_{1-t})$ for an integrable function G (where $\mathbb{P}_x$ is a probability measure on the sample space such that $\mathbb{P}_x(X_0 = x) = 1$), the the Markov property of X implies that $\big(F(X_t, t)\big)_{t\in[0,1]}$ is a martingale under $\mathbb{P}_x$. Similarly, if we set $F(x,t) = \mathbb{E}_x \left(\int_0^{1-t} H(X_s)\mathrm{d}s\right)$ for a sufficiently regular function H, then $\left(F(X_t, t) + \int_0^t H(X_s)\mathrm{d}s\right)_{t\in[0,1]}$ is a martingale under $\mathbb{P}_x$. Applying these facts to the space-time Markov process $((x+B_s, t+s))_{s\ge 0}$, we get that for a fixed number t,

$$\left(U^c(x + B_s, t + s) + \frac{p(p-1)}{2} \int_0^s |x + B_u|^{p-2} 1_{\{|x+B_u|>c(t+u)\}} \mathrm{d}u\right)_{s\le 1-t} \tag{2.27}$$

is a martingale. On the other hand, we have

$$G(x + B_s, t + s) = G(x,t) - \frac{p(p-1)}{2} \int_0^s |x + B_u|^{p-2} \mathrm{d}u + M_s,$$

where

$$M_s = p \int_0^s |x + B_u|^{p-2}(x + B_u)\mathrm{d}B_u, \qquad 0 < s < 1 - t,$$

is a martingale. Suppose that $|x| > c(t)$ and consider the stopping time

$$\sigma_c = \inf\{0 < s < 1 - t : |x + B_s| = c(t + s)\}.$$

The stopping time σ_c is bounded by $1-t$, since c is continuous and explodes at the right endpoint of its domain. As we have already pointed out above, we have the equality $U^c(x + B_{\sigma_c}, t + \sigma_c) = G(x + B_{\sigma_c}, t + \sigma_c)$ and thus

$$\begin{aligned}
&U^c(x,t)\\
&= \mathbb{E}\left[U^c(x + B_{\sigma_c}, t + \sigma_c) + \frac{p(p-1)}{2} \int_0^{\sigma_c} |x + B_u|^{p-2} 1_{\{|x+B_u|>c(t+u)\}} \mathrm{d}u\right]\\
&= \mathbb{E}\left[G(x + B_{\sigma_c}, t + \sigma_c) + \frac{p(p-1)}{2} \int_0^{\sigma_c} |x + B_u|^{p-2} \mathrm{d}u\right]\\
&= G(x,t).
\end{aligned}$$

Step 8. Let us prove that $U^c \le V$. To do this, fix $x \in \mathbb{R}$, $t < 1$ and consider the stopping time

$$\tau_c = \inf\{0 \le s < 1 - t : |x + B_s| \ge c(t + s)\},$$

with the convention $\inf \emptyset = 1 - t$. By the previous step and the equality $U^c(x, 1) = G(x, 1)$ (which is obvious from the formula for U^c) we have $U^c(x + B_{\tau_c}, t + \tau_c) =$

$G(x+B_{\tau_c}, t+\tau_c)$. Using the martingale property of the process (2.27) and the fact that the integrand appearing in its definition vanishes for $u < \tau_c$, we obtain

$$U^c(x,t) = \mathbb{E}U^c(x+B_{\tau_c}, t+\tau_c) = \mathbb{E}G(x+B_{\tau_c}, t+\tau_c) \leq V(x,t),$$

as desired.

Step 9. We are ready to show that $c(t) \leq b(t)$ for all $t < 1$. Suppose that there is $t < 1$ for which the reverse inequality $c(t) > b(t)$ holds. Introduce the stopping time

$$\sigma_b = \inf\{0 \leq s < 1-t : c(t) + B_s = b(t+s)\}$$

(since $\lim_{u\uparrow 1} b(u) = \infty$, the above definition makes sense). By Itô formula, we have

$$\mathbb{E}V(c(t)+B_{\sigma_b}, t+\sigma_b) = V(c(t),t) - \frac{p(p-1)}{2}\mathbb{E}\int_0^{\sigma_b} |c(t)+B_u|^{p-2}\mathrm{d}u$$

and, by the martingale property of the process (2.27),

$$\begin{aligned}&\mathbb{E}U^c(c(t)+B_{\sigma_b}, t+\sigma_b)\\ &= U^c(c(t),t) - \frac{p(p-1)}{2}\mathbb{E}\int_0^{\sigma_b} |c(t)+B_u|^{p-2}1_{\{|c(t)+B_u|>c(t+u)\}}\mathrm{d}u.\end{aligned}$$

However, we have $V(c(t),t) = G(c(t),t) = U^c(c(t),t)$ and, by the previous step, $U^c(c(t)+B_{\sigma_b}, t+\sigma_b) \leq V(c(t)+B_{\sigma_b}, t+\sigma_b)$. Consequently, the two equalities above imply that

$$\mathbb{E}\int_0^{\sigma_b} |c(t)+B_u|^{p-2}1_{\{|c(t)+B_u|\leq c(t+u)\}}\mathrm{d}u \leq 0,$$

which is impossible in view of the continuity of b and c.

Step 10. Finally, we show that $b \leq c$, which will complete the proof of the uniqueness. Suppose on contrary that there is t for which $c(t) < b(t)$ and pick $x \in (c(t), b(t))$. Let

$$\tau_D = \inf\{s > 0 : (x+B_s, t+s) \in D\}.$$

We have

$$\mathbb{E}G(x+B_{\tau_D}, t+\tau_D) = V(x,t)$$

and, since $G(x+B_{\tau_D}, t+\tau_D) = U^c(x+B_{\tau_D}, t+\tau_D)$, we get

$$\begin{aligned}\mathbb{E}G(x+B_{\tau_D}, t+\tau_D) &= U^c(x,t)\\ &\quad - \frac{p(p-1)}{2}\mathbb{E}\left(\int_0^{\tau_D} |x+B_{t+u}|^{p-2}1_{\{|x+B_u|>c(t+u)\}}\mathrm{d}u\right).\end{aligned}$$

However, we have $V \geq U^c$ (see Step 8), so the two identities above imply

$$\mathbb{E}\left(\int_0^{\tau_D} |x+B_{t+u}|^{p-2}1_{\{|x+B_u|>c(t+u)\}}\mathrm{d}u\right) \leq 0,$$

which cannot hold, because of the continuity of b and c. This completes the proof of the uniqueness of the solution to (2.26).

Step 11. We are ready for the proof of the weak-type estimate. Directly from the above analysis, we have $\mathbb{P}(\tau \geq 1-t_0) \leq \mathbb{E}|B_\tau|^p$ and hence, since $(\lambda B_{t/\lambda^2})_{t\geq 0}$ is a Brownian motion, we get

$$\mathbb{P}(\tau \geq 1) \leq (1-t_0)^{p/2}\mathbb{E}|B_\tau|^p.$$

To see that the constant $(1-t_0)^{1/2}$ cannot be improved, pick an arbitrary number $t > t_0$ and let τ be the optimal stopping time for $(0,t)$. We have $V(0,t) > 0$ and hence

$$\mathbb{P}(\tau \geq 1-t) - \mathbb{E}|B_\tau|^p = V(0,t) > 0,$$

which, after the homogeneity-type property of Brownian motion, implies

$$\mathbb{P}(\tau \geq 1) - (1-t)^{p/2}\mathbb{E}|B_\tau|^p > 0.$$

Therefore, the best constant is not smaller than $(1-t)^{1/2}$ and letting $t \downarrow t_0$ yields the desired lower bound.

In the final part of these notes, we will discuss a few examples involving discontinuous processes. We start with the definition of the Poisson process.

Definition 2.12. A cádlag process $N = (N_t)_{t\geq 0}$ is a Poisson process with intensity $\lambda > 0$ if it satisfies the following conditions.
1. We have $N_0 = 0$ almost surely.
2. The increments of N are independent.
3. For any $0 \leq s < t$, the random variable $N_t - N_s$ has Poisson distribution with parameter $\lambda(t-s)$.

Example 2.9. Suppose that N is a Poisson process with intensity λ and $\gamma > 0$ is a fixed parameter. We will solve the optimal stopping problem

$$V = \sup_\tau \mathbb{E}e^{-\gamma\tau}N_\tau,$$

where the supremum is taken over all stopping times τ. We proceed in a standard manner. The process (t, N_t) is a time-homogeneous Markov process on $E = [0,\infty)\times\mathbb{N}$, which can be extended to a Markov family with the corresponding family $(P_{t,n})_{(t,n)\in E}$ of initial distributions. We generalize the above problem to

$$V(t,n) = \sup_\tau \mathbb{E}_{t,n}G(\tau, N_\tau) = \sup_\tau \mathbb{E}e^{-\gamma\tau}N_\tau,$$

where the supremum is taken over the same set as above. We easily check the homogeneity condition

$$V(t+s,n) = e^{-\gamma t}V(s,n) \qquad \text{for all } s,\ t > 0 \text{ and } n \in \mathbb{N},$$

which implies that $V(t,n) = e^{-\gamma t}v(n)$ for some unknown function $v : \mathbb{N} \to \mathbb{R}$. This also implies that the corresponding continuation/stopping sets C, D depend on the size of the variable n only.

It follows from the general theory of optimal stopping that V is an excessive majorant of the function G. To understand what the excessivity means, let us first analyze the action of the infinitesimal generator of the process N. For any $n \geq 0$ and any appropriately bounded function f on $\mathbb{N}$, we have

$$\begin{aligned}\mathbb{L}_N f(n) = \lim_{t\downarrow 0}\frac{\mathbb{E}f(n+N_t) - f(n)}{t} &= \lim_{t\downarrow 0} t^{-1}\left(\sum_{k=0}^{\infty} f(k+n)e^{-\lambda t}\frac{(\lambda t)^k}{k!} - f(0)\right)\\ &= \lim_{t\downarrow 0} t^{-1}\left(f(n)(e^{-\lambda t}-1) + f(n+1)\cdot\lambda t e^{-\lambda t}\right)\\ &= \lambda(f(n+1) - f(n)).\end{aligned}$$

This implies that for any (appropriately bounded) function F on E, differentiable with respect to t, we have

$$\mathbb{L}_{t,N}F(t,n) = \frac{\partial F}{\partial t}(t,n) + \lambda(F(t,n+1) - F(t,n)).$$

We are ready for the search for V. On the continuation set C, we must have $\mathbb{L}_{t,N}V = 0$, which combined with the homogeneity condition gives

$$e^{-\gamma t}\left(-\gamma v(n) + \lambda(v(n+1) - v(n))\right) = 0,$$

that is, $v(n+1) = \frac{\lambda+\gamma}{\lambda}v(n)$. To guess the shape of the stopping set, note that

$$\mathbb{L}_{t,N}G(t,n) = e^{-\gamma t}\left(-\gamma n + \lambda\right) \tag{2.28}$$

is positive for small n and negative for large n. This suggests that D should be of the form $D = \{(t,n) \in E : n \geq n_0\}$, which combined with the above discussion yields the candidate

$$\tilde{V}(t,n) = \begin{cases} Ae^{-\gamma t}\left(\dfrac{\gamma+\lambda}{\lambda}\right)^n & \text{if } n < n_0, \\ e^{-\gamma t}n & \text{if } n \geq n_0, \end{cases}$$

for some parameter $A > 0$ to be determined. To find A and n_0, we make use of the condition that V must be the *least* majorant of G. We define A to be the smallest positive number such that $A\left(\frac{\gamma+\lambda}{\lambda}\right)^n \geq n$ for all $n \in \mathbb{N}$ and let n_0 be the equality point (if there are two of them, we pick the smaller one). We easily check that then $\tilde{V}$ is an excessive majorant of G. Indeed, letting $H(t,n) = Ae^{-\gamma t}\left(\frac{\gamma+\lambda}{\lambda}\right)^n$ for all (t,n), we have $0 = \mathbb{L}_{t,N}H = 0$ (by the above construction), so the majorization gives $\mathbb{L}_{t,N}G(t,n_0) \leq 0$. A glimpse at (2.28) reveals that we must also have $\mathbb{P}_{t,N}\tilde{V}(t,n) \leq 0$ for all $n \geq n_0$. Consequently, we have $\tilde{V} \geq V$; to show the reverse bound, we fix (t,n) and consider the stopping time $\tau = \inf\{t : N_t = n_0\}$. The process $(\tilde{V}(\tau \wedge t, N_{\tau\wedge t}))_{t\geq 0}$ is a martingale (this can be verified directly), so

$$\tilde{V}(t,n) = \mathbb{E}_{t,n}\tilde{V}(\tau \wedge t, N_{\tau\wedge t}).$$

But by the very definition of $\tilde{V}$ and τ, the process $(\tilde{V}(\tau \wedge t, N_{\tau\wedge t}))_{t\geq 0}$ is bounded and τ is finite almost surely; thus, Lebesgue's dominated convergence theorem gives

$$\tilde{V}(t,n) = \mathbb{E}_{t,n}\tilde{V}(\tau, N_\tau) = \mathbb{E}_{t,n}G(\tau, N_\tau).$$

This gives $V(t,n) \geq \tilde{V}(t,n)$ and the analysis is complete.

EXAMPLE 2.10. Now we will analyze a different problem. Suppose that B is a standard Brownian motion and, for a given $t \geq 0$, define

$$Z_t = \sup\{s \leq t : B_s = 0\}$$

to be the last zero of B before time t. We will identify the best constant in the estimate

$$\mathbb{E}\sqrt{\tau Z_\tau} \leq C\mathbb{E}Z_\tau,$$

where τ is an arbitrary integrable stopping time. As usual, we transform the problem into the optimal stopping theory, rewriting it as

$$V = \sup_\tau \mathbb{E}\left[\sqrt{\tau Z_\tau} - CZ_\tau\right]. \tag{2.29}$$

The analysis is split into several parts.

Step 1. We start with the identification of the conditional distribution of B_t given $Z_t = s$. Suppose that $\varepsilon > 0$ is an auxiliary parameter and consider

$$\mathbb{P}(B_t \geq b \text{ and } B_u > -\varepsilon \text{ on } (s,t) | B_s = 0)$$

(of course, this conditional probability must be interpreted appropriately: the conditioning event has probability zero). It is equal to

$$\mathbb{P}(B_t \geq b | B_s = 0) - \mathbb{P}(B_t \geq b \text{ and } B_u = -\varepsilon \text{ for some } u \in (s,t) | B_s = 0).$$

The first probability is equal to $\mathbb{P}(B_{t-s} \geq b)$, by the reflection principle, the second probability is equal to $\mathbb{P}(B_{t-s} \geq b + 2\varepsilon)$. Finally, note that, again by the reflection principle, $\mathbb{P}(B_u > -\varepsilon \text{ on } (s,t) | B_s = 0) = 1 - 2\mathbb{P}(B_{t-s} \geq \varepsilon)$. Putting all the above facts together, we obtain that

$$\begin{aligned}\mathbb{P}(B_t \geq b | B_u > -\varepsilon \text{ on } (s,t) \text{ and } B_s = 0) &= \frac{\mathbb{P}(B_{t-s} \geq b) - \mathbb{P}(B_{t-s} \geq b + 2\varepsilon)}{1 - 2\mathbb{P}(B_{t-s} \geq \varepsilon)} \\ &= \frac{\Phi((b+2\varepsilon)/\sqrt{t-s}) - \Phi(b/\sqrt{t-s})}{2\Phi(\varepsilon/\sqrt{t-s}) - 1},\end{aligned}$$

where Φ is the distribution function of the standard normal variable. Letting $\varepsilon \to 0$ and applying de l'Hospital rule, we get

$$\mathbb{P}(|B_t| \geq b | Z_t = s) = \exp\left(-\frac{b^2}{2(t-s)}\right).$$

By the symmetry of the Brownian motion, the conditional density of B_t given $Z_t = s$ (for $s \neq t$) equals

$$\frac{|b|}{2(t-s)} \exp\left(-\frac{b^2}{2(t-s)}\right).$$

Clearly, if $s = t$, then B_t is concentrated at zero.

Step 2. Now we will identify the conditional distribution of Z_u given $Z_t = s$ (for a fixed $u > t$). Namely, given $r \in (t, u]$, we may write, by the reflection principle,

$$\mathbb{P}(B \text{ has zero in } (u,r) | B_t = b, Z_t = s) = 2\mathbb{P}(B_{r-t} \geq |b|)$$

and hence, integrating over b, we get

$$\begin{aligned}\mathbb{P}(B \text{ has zero in } (t,r) | Z_t = s) &= \int_{\mathbb{R}} 2\mathbb{P}(B_{r-t} \geq |b|) \cdot \frac{|b|}{2(t-s)} \exp\left(-\frac{b^2}{2(t-s)}\right) \mathrm{d}b \\ &= 2\int_0^\infty \left(1 - \Phi\left(\frac{b}{\sqrt{r-t}}\right)\right)\left(-\exp\left(-\frac{b^2}{2(t-s)}\right)\right)' \mathrm{d}b \\ &= 1 - \frac{2}{\sqrt{r-t}} \int_0^\infty \Phi'\left(\frac{b}{\sqrt{r-t}}\right) \exp\left(-\frac{b^2}{2(t-s)}\right) \mathrm{d}b \\ &= 1 - \frac{2}{\sqrt{r-t}} \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{b^2(r-s)}{2(r-t)(t-s)}\right) \mathrm{d}b \\ &= 1 - \frac{\sqrt{t-s}}{\sqrt{r-s}}.\end{aligned}$$

This has two consequences. First, we have $\mathbb{P}(Z_u = s | Z_t = s) = \sqrt{t-s}/\sqrt{u-s}$; second, the „partial" density of Z_u given $Z_t = s$, is equal to

$$\frac{\sqrt{t-s}}{2(r-s)^{3/2}}, \text{ for } r \in (t, u].$$

Step 3. It is easy to see that $Y = (t - Z_t)_{t \geq 0}$ is a time-homogeneous Markov process. We will compute its infinitesimal generator. Using the calculations from the previous step, we compute that for a C^1 bounded function f on $[0, \infty)$ and any $s < t$,

$$\frac{\mathbb{E}\big[f(u - Z_u) - f(t - Z_t) | t - Z_t = t - s\big]}{u - t}$$
$$= \frac{f(u-s) - f(t-s)}{u - t} \cdot \frac{\sqrt{t-s}}{\sqrt{u-s}} + \frac{1}{u-t} \int_t^u (f(u-r) - f(t-s)) \frac{\sqrt{t-s}}{2(r-s)^{3/2}} \mathrm{d}r$$
$$\xrightarrow{u \downarrow t} f'(t-s) + \frac{f(0) - f(t-s)}{2(t-s)}.$$

This is equivalent to saying that for $y > 0$,

$$\mathbb{L}_Y f(y) = f'(y) + \frac{f(0) - f(y)}{2y}.$$

For $y = 0$, we carry out the limiting procedure, obtaining $\mathbb{L}_Y f(0) = f'(0)/2$. Passing to the space-time version, we get

$$\mathbb{L}_{t,Y} F(t, y) = \frac{\partial F}{\partial t}(t, y) + \frac{\partial F}{\partial y}(t, y) + \frac{F(t, 0) - F(t, y)}{2y}$$

for $y > 0$, with the appropriate limit version for $y = 0$.

Step 4. To solve (2.29), we extend it into

$$V(t, y) = \sup_{\tau} \mathbb{E}_{t,y} G(\tau, \tau - Z_\tau),$$

where $G(t, y) = \sqrt{t(t-y)} - C(t - y)$. The Brownian motion is homogeneous: this gives also the appropriate homogeneity of the process $(t, t - Z_t)$ and we obtain $V(t, y) = tV(1, y/t) = tv(y/t)$ for all t, $y > 0$. As usual, we write down the continuation and the stopping sets C and D, and analyze the behavior of V on the first of these sets. Applying the generator $\mathbb{L}_{t,Y}$, we get the differential equation

$$v(s) - sv'(s) + v'(s) + \frac{v(0) - v(s)}{2s} = 0,$$

where $s = y/t \in [0, 1]$. It is easy to check that the solutions are the affine functions $v(s) = a(1 - 2s)$ for $a \in \mathbb{R}$. Thus, we expect to have

$$V(t, y) = \begin{cases} a(t - 2y) & \text{if } (t, y) \in C, \\ \sqrt{t(t-y)} - C(t-y) & \text{if } (t, y) \in D. \end{cases}$$

The question is: for which C there is an a such that the above function is excessive? Note that the function $a(t - 2y)$ is excessive itself, so in particular, we must have $G(t, y) \leq a(t - 2y)$ for all t, y. Dividing by t and substituting $s = y/t$ as above, the latter estimate becomes

$$\sqrt{1-s} - (C + 2a)(1 - s) \leq -a.$$

If we let $s \to 1$, we see that a must be nonpositive. Now, assuming that $C + 2a > 0$, we have

$$\sqrt{1-s} - (C + 2a)(1 - s) \leq \frac{1}{4(C + 2a)},$$

which forces us to take $\frac{1}{4(C+2a)} = -a$, i.e., $C = -(4a)^{-1} - 2a$. The latter condition implies $C \geq \sqrt{2}$, and the equality holds if and only if $a = -\sqrt{2}/4$. We easily see that for these choices of parameters, we have the majorization $a(t-2y) \geq G(t,y)$. Consequently, for any integrable stopping time τ and any $t \geq 0$ we have

$$\mathbb{E}\sqrt{(\tau \wedge t) Z_{\tau\wedge t}} \leq \sqrt{2}\mathbb{E}Z_{\tau\wedge t}.$$

Letting $t \to \infty$ and using Lebesgue's dominated convergence theorem (together with the obvious bound $Z_t \leq t$), we get

$$\mathbb{E}\sqrt{\tau Z_\tau} \leq \sqrt{2}\mathbb{E}Z_\tau.$$

To show that the constant $\sqrt{2}$ is optimal, pick an arbitrary $c > 1/2$, a small $\varepsilon > 0$ and consider the stopping time $\tau = \inf\{t : Z_t \leq ct - \varepsilon\}$. The process $(t - 2Z_t)_{t\geq 0}$ is a martingale, as shown above, so

$$\mathbb{E}(\tau \wedge t - 2Z_{\tau\wedge t}) = 0.$$

Furthermore, $Z_{\tau\wedge t} \geq c\tau \wedge t - \varepsilon$ for any t (since Z does not have negative jumps), which gives

$$\frac{1}{2}\mathbb{E}\tau \wedge t = \mathbb{E}Z_{\tau\wedge t} \geq c\mathbb{E}\tau \wedge t - \varepsilon.$$

This implies $\mathbb{E}\tau \wedge t \leq \varepsilon/(c - 1/2)$ and hence, letting $t \to \infty$, we get $\mathbb{E}\tau < \infty$, by Lebesgue's monotone convergence theorem; in particular, τ is finite almost surely. It remains to note that $Z_\tau = c\tau - \varepsilon$ and consequently

$$\mathbb{E}\sqrt{\tau Z_\tau} = \mathbb{E}\sqrt{\frac{Z_\tau + \varepsilon}{c} \cdot Z_\tau} \geq c^{-1/2}\mathbb{E}Z_\tau.$$

But $c > 1/2$ was chosen arbitrarily. Thus, the constant $\sqrt{2}$ is indeed the best possible.

2.4. Problems

In all exercises below, W denotes the standard one-dimensional Brownian motion.

1. Solve the optimal stopping problems

$$V = \sup_\tau \mathbb{E}\sqrt{|W_\tau|}, \quad V = \sup_\tau \mathbb{E}\frac{1}{1+W_\tau^2}, \quad V = \sup_{\tau\in L^1} \mathbb{E}(|W_\tau| - \tau), \quad V = \sup_\tau \mathbb{E}\frac{|W_\tau|}{1+\tau}.$$

2. Solve the optimal stopping problems

$$V = \sup_{0\leq\tau\leq 1} \mathbb{E}\tau W_\tau^2, \qquad V = \sup_{0\leq\tau\leq 1} \mathbb{P}(W_\tau \geq 1).$$

3. Put

$$D_t = |\{s \leq t : W_s \in [0,1]\}| - |\{s \leq t : W_s \in [-2,0)\}|, \qquad t \geq 0.$$

Solve the optimal stopping problem $V = \sup_\tau \mathbb{E}D_\tau$.

4. Solve the optimal prediction problem

$$V = \inf_{0\leq\tau\leq 1} \mathbb{E}\left(\int_0^1 W_s \mathrm{d}s - W_\tau\right)^2.$$

5. Solve the optimal prediction problem

$$V = \sup_\tau \mathbb{E}\int_\tau^\infty e^{-s/2} W_s \mathrm{d}s.$$

6. Solve the optimal prediction problem

$$V = \sup_{0 \le \tau \le 1} \mathbb{E}\left[\sup_{\tau \le t \le 1} W_t - \tau\right].$$

7. Solve the optimal prediction problem

$$V = \sup_{0 \le \tau \le 1} \mathbb{E}\int_{\tau}^{1} W_s \mathrm{d}s.$$

8. Solve the optimal stopping problems

$$V = \sup_{\tau} \mathbb{E}\big(W_\tau - \tau\big), \qquad V = \sup_{\tau} \mathbb{E}\big(|W_\tau| - \tau\big),$$

where the suprema are taken over a) all integrable stopping times τ; b) all integrable stopping times τ with respect to $\mathcal{F}^Z$, where $Z_t = \sup\{s \le t : W_s = 0\}$.

9. Find the best constants c, C in the estimates

$$\mathbb{E}B_\tau^* \le c\sqrt{\mathbb{E}\tau}, \qquad \mathbb{E}|B|_\tau^* \le C\sqrt{\mathbb{E}\tau},$$

to be valid for all $\tau \in L^1$.

10. For any $1 < p < \infty$, find the best constant C_p in the inequality

$$\mathbb{E}|B_\tau|^* \le C_p\big(\mathbb{E}|B_\tau|^p\big)^{1/p},$$

to be valid for all stopping times $\tau \in L^{p/2}$.

11. For any $1 < p < \infty$, find the best constant C_p in the inequality

$$\mathbb{E}|B_\tau|^* \le C_p\big(\mathbb{E}|B_\tau|^p\big)^{1/p},$$

to be valid for all stopping times $\tau \in L^{p/2}$.

12. Suppose that X is the solution to the differential equation

$$\mathrm{d}X_t = rX_t\mathrm{d}t + \sigma X_t \mathrm{d}B_t,$$

satisfying the initial condition $X_0 = x > 0$. For a given positive constant K, solve the optimal stopping problem

$$V(x) = \sup_{\tau} \mathbb{E}_x e^{-r\tau}(K - X_\tau)^+,$$

where the supremum is taken over all stopping times τ.

13. Suppose that X is the solution to the differential equation

$$\mathrm{d}X_t = rX_t\mathrm{d}t + \sigma X_t \mathrm{d}B_t,$$

satisfying the initial condition $X_0 = x > 0$. Solve the optimal stopping problem

$$V = \sup_{\tau} \mathbb{E}e^{-r\tau}X_\tau^*,$$

where the supremum is taken over all stopping times τ.

14. Suppose that V is the value function of the optimal stopping problem

$$V(t, x) = \sup_{\tau} \mathbb{E}G(\tau, B_\tau),$$

where the supremum is taken over all integrable stopping times τ. Suppose that for any (t, x) there is an optimal stopping time τ and V, G are continuous. Prove that V is of class $C^{1,2}$ on the continuation region and satisfies $V_t + \frac{1}{2}V_{xx} = 0$ there.

15. For any $1 < p < 2$, find the best constant C_p in the inequality

$$\|B_\tau^*\|_{L^p} \leq C_p \|\tau^{1/2}\|_{L^2},$$

to be valid for all integrable stopping times τ.

16. Let N be the Poisson process of intensity λ and let γ be a positive number. Solve the optimal stopping problem

$$V = \sup_\tau \mathbb{E} e^{-\gamma\tau}(N_\tau - \lambda\tau),$$

where the supremum is taken over all stopping times τ.

Bibliography

[1] Bellman, R. *Dynamic programming.* Reprint of the 1957 edition. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2010.

[2] Burkholder, D. L. *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab. **12** (1984), pp. 647–702.

[3] Burkholder, D. L. *An extension of a classical martingale inequality*, Prob. Theory and Harmonic Analysis (J.-A. Chao and A. W. Woyczyński, eds.), Marcel Dekker, New York, 1986, pp. 21–30.

[4] Carleman, T. *Sur les fonctions quasi-analytiques.* Conférences faites au cinquieme congres des mathématiciens Scandinaves, Helsinki (1923), 181–196.

[5] Carlson, F. *Une inegalité*, Ark. Math. Astron. Fys. **25** (1934), 1–5.

[6] Chow Y. S. and Robbins, H. *A martingale system theorem and applications*, Proc. Fourth Berkeley Symp. on Math., Stat. and Prob., **1** (1960) 93–104.

[7] Coifman R. R. and Rochberg R. *Another characterization of BMO*, Proc. Amer. Math. Soc. **79** (1980), pp. 249–254.

[8] Cox A. *Optimal stopping and applications*, available at www.maths.bath.ac.uk/mapamgc /optimalstopping__v5.pdf.

[9] Dubins, L. E., Shepp, L. A. and Shiryaev, A. N., *Optimal stopping rules and maximal inequalities for Bessel process*, Theory Probab. Appl. **38** (1993), 226–261.

[10] Hardy, G. H.; Littlewood, J. E. and Pólya, G. *Inequalities.* Reprint of the 1952 edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988.

[11] Hill, T.P. and Kertz R. P. *Stop rule inequalities for uniformly bounded sequences of random variables*, Trans. Amer. Math. Soc. **278** (1983), 197–207.

[12] Krengel, U. and Sucheston, L. *Semiamarts and finite values*, Bull. Amer. Math. Soc. **83** (1977), 745–747.

[13] Krengel, U. and Sucheston, L. *On semiamarts, amarts, and processes with finite value*, Probability on Banach Spaces, Ed. by J. Kuelbs, Marcel Dekker, New York, 1978.

[14] Nikolidakis, E. N. *Extremal problems related to maximal dyadic-like operators*, J. Math. Anal. Appl. 369 (2010), 377–385.

[15] Osękowski, A. *Sharp LlogL inequality for Differentially Subordinated Martingales*, Illinois J. Math. **52**, Vol. 3 (2008), pp. 745–756.

[16] Osękowski, A. *Sharp moment inequalities for differentially subordinated martingales*, Studia Math. **201** (2010), pp. 103–131.

[17] Osękowski, A. *A sharp one-sided bound for the Hilbert transform*, Proc. Amer. Math. Soc. **141** (2013), pp. 873–882.

[18] G. Peskir and A. Shiryaev, Optimal stopping and free boundary problems, Lect. Notes in Math. ETH Zurich, 2006.

[19] Robbins, H. *Optimal stopping*, Amer. Math. Mothly **77** (1970), pp. 333-343.

[20] Strzelecki, M. *A note on sharp one-sided bounds for the Hilbert transform*, Proc. Amer. Math. Soc. **144** (2016), pp. 1171–1181.

[21] Suh, Y. *A sharp weak type (p,p) inequality $(p > 2)$ for martingale transforms and other subordinate martingales*, Trans. Amer. Math. Soc. **357** (2005), pp. 1545–1564.

[22] Tomaszewski, B. *The best constant in a weak-type H^1-inequality, complex analysis and applications*, Complex Variables, Vol. 4, 1984, pp. 35–38.