CHAPTER 1

Dynamic programming

Dynamic programming is an algorithm which enables to solve a certain class of problems, by an induction argument which reduces them to simpler sub-problems. Or, to put it in the reverse direction, this approach allows to tackle difficult problems by solving simpler ones first and relating these solutions to the harder context by intrinsic recurrence relations. It plays an important role in computer science, as it can be used to construct effective algorithms of polynomial complexity. Furthermore, the method is utilized in optimal planning problems (e.g. in problems of the optimal distribution of resources available, the theory of inventory management, replacement of equipment, and so on). From the viewpoint of our further purposes, we would like to emphasize that the dynamic programming lies at the heart of the Bellman function technique which will be developed later.

The basic idea behind the dynamic programming can be formulated as follows. Suppose that a given system $\mathcal{S}$ is controlled with a procedure which consists of N steps, where N is a fixed positive integer. At the beginning, the system is in some state x_0. At the m-th step of the procedure, there is a possibility of applying a class of controls, each of which transforms the state x_{m-1} obtained through the previous operations into some new state x_m. Formally, if y_m denotes the control applied at m-th step, then we have the identity

$$x_m = f_m(x_{m-1}, y_m)$$

for some f_m (the "transition function associated with m-th step"). An important feature of the method is the absence of after-effects, i.e. the controls selected for a given step may only affect the state of the system at that moment.

As the result of the whole procedure y_1, y_2, ..., y_N, the system is converted from the state x_0 into the final state x_N. Now the problem can be stated as follows: given the initial state x_0 and a fixed objective function

$$F(x_0, x_1, \dots, x_N, y_1, y_2, \dots, y_N) = \sum_{m=1}^{N} g_m(x_{m-1}, y_m),$$

identify the controls y_1^*, y_2^*, ..., y_N^* which yield the optimal performance, i.e., such that we have

$$F(x_0, x_1, \dots, x_N, y_1^*, y_2^*, \dots, y_N^*) = \sup_{y_1, y_2, \dots, y_N} F(x_0, x_1, \dots, x_N, y_1, y_2, \dots, y_N).$$

Conventional methods of tackling such problems are often either inapplicable, or involve lengthy and elaborate calculations. Dynamic programming allows to solve the above problem by a recursive formula, which rests on the so-called optimality principle of Richard Bellman. For simplicity, let us assume that the controls y_1^*, y_2^*, ..., y_N^* exist (i.e., the optimal performance is not attained in the limit). Suppose that for some $1 \leq n \leq N$, we have chosen certain controls y_1, y_2, ..., y_n up to step

n, transforming the system from the state x_0 to the state x_n. So, to conclude the procedure, we need to select the remaining controls y_{n+1}, y_{n+2}, ..., y_N. Then, if the part of F corresponding to these remaining controls is not optimized, that is, if the maximum of the function

$$F_n(x_n, x_{n+1}, \dots, x_N, y_n, y_{n+1}, \dots, y_N) = \sum_{m=n+1}^{N} g_m(x_{m-1}, y_m)$$

is not attained, then the procedure as a whole will not be optimal. Now, for any $n = 0, 1, 2, \dots, N$ and any state x, define $B_n(x)$ to be the maximal possible value of the value function F_n assuming that after n-th step the system is in the state x. Then $B_N(x) = 0$, and the optimality principle explained above yields the identity

$$B_{n-1}(x) = \sup_y \{ B_n(f_n(x, y)) + g_n(x, y) \}, \qquad n = 1, 2, \dots, N, \tag{1.1}$$

called the *Bellman equation.* Solving this system of functional equations, we obtain $B_0(x_0)$, the value of the optimal performance under the assumption that the system is initially in the state x_0. We also identify the optimal controls y_1^*, y_2^*, ..., y_N^*: these are the parameters for which the suprema in (1.1) are attained. Specifically, having determined B_0, B_1, ..., B_N (called the Bellman sequence in the sequel), we find y_1^* by the requirement

$$B_0(x_0) = B_1(f_1(x_0, y_1^*)) + g_1(x_0, y_1^*).$$

Let $x_1 = f_1(x_0, y_1^*)$ be the position of the system after the optimal first move. Then we get the control y_2^* from the equation

$$B_1(x_1) = B_2(f_2(x_1, y_2^*)) + g_2(x_1, y_2^*),$$

and so on.

Summarizing, the approach rests on solving the initial problem by imbedding it into a class of similar sub-problems, the collection of which can be treated, as a whole, by means of the recursive formulas. We should point out, however, that still, in many cases, the analysis of the obtained setting can be technically involved and require plenty of laborious computations.

In what follows, we will see the above reasoning in various disguises. We have purposefully restrained ourselves from the discussion concerning the state space or the class of feasible controls; this would probably complicate the above presentation, as these objects can be multidimensional or change from step to step. Sometimes it will be convenient to enumerate the steps by numbers 1, 2, ..., N instead of 0, 1, ..., N. Moreover, in many cases it will be more natural to work with the reversed sequence B_N, B_{N-1}, ..., B_1 instead of B_1, B_2, ..., B_N (then the lower index indicates the number of steps up to the termination of the process). However, the main idea remains essentially unchanged.

Instead of exploring further the abstract description, we continue with the analysis of several examples which will serve as an illustration of the above concepts.

1.1. A toy example: AM-GM inequality

We will show how dynamic programming can yield one of the most fundamental inequalities in mathematics.

THEOREM 1.1. *For any positive integer N and any nonnegative numbers a_1, a_2, …, a_N we have*

$$\frac{a_1 + a_2 + \ldots + a_N}{N} \geq (a_1 a_2 \ldots a_N)^{1/N}.$$

Equality holds if and only if $a_1 = a_2 = \ldots = a_N$.

PROOF. The first step in the analysis is to rephrase the desired claim into the problem of the optimization of some value function. This can be done as follows. Fix a nonnegative number x and consider the quantity

$$\sup\Big\{a_1 a_2 \ldots a_N \,:\, a_1,\, a_2,\, \ldots,\, a_N \geq 0,\; a_1 + a_2 + \ldots + a_N = x\Big\}.$$

We need to show that for any x, the above supremum does not exceed $(x/N)^N$. This problem can be studied with the use of dynamic programming approach, where the sub-problems correspond to AM-GM inequalities applied to subsequences of a_1, a_2, …, a_N. More precisely, fix $n \in \{1, 2, \ldots, N\}$, a nonnegative number x and consider the quantity

$$B_n(x) = \sup\Big\{a_n a_{n+1} \ldots a_N\Big\},$$

where the supremum is taken over all sequences a_n, a_{n+1}, …, a_N of nonnegative numbers satisfying $a_n + a_{n+1} + \ldots + a_N = x$. To put this into the framework presented in the previous section, the state of the system at step k is described by x, the sum of the numbers $a_k + a_{k+1} + \ldots + a_N$, while the control in step k is the value of the variable $a_k \in [0, x]$, which transforms the state x into $x - a_k$.

The dynamic approach rests on writing the system of equations which govern the evolution of the sequence $(B_n)_{n=1}^N$. Bellman's optimality principle implies that for any $n = 0, 1, \ldots, N-1$ and any $x \geq 0$ we have

$$B_n(x) = \sup_{t\in[0,x]} \big\{B_{n+1}(x-t)\cdot t\big\}. \tag{1.2}$$

Indeed, if a_n, a_{n+1}, …, a_N are arbitrary nonnegative numbers summing up to x and we denote $t = a_n \in [0, x]$, then

$$a_n a_{n+1} \ldots a_N = t \cdot a_{n+1} a_{n+2} \ldots a_N \leq B_{n+1}(x-t)\cdot t,$$

by the definition of B_{n+1} and the fact that $a_{n+1} + a_{n+2} + \ldots + a_N = x - t$. Taking the supremum over all a_n, a_{n+1}, …, a_N as above, gives the inequality "$\leq$" in (1.2). To get the reverse, we fix $t \in [0, x]$ and nonnegative numbers a_{n+1}, a_{n+2}, …, a_N summing up to $x - t$. Then $t + a_{n+1} + a_{n+2} + \ldots + a_N = x$, so the definition of B_n yields

$$t \cdot a_{n+1} a_{n+2} \ldots a_n \leq B_n(x).$$

Now, taking the supremum over a_{n+1}, a_{n+2}, …, a_N as above gives $tB_{n+1}(x-t) \leq B_n(x)$, and since $t \in [0, x]$ was arbitrary, the identity (1.2) follows. The description of the sequence $(B_n)_{n=1}^N$ is completed by observing that $B_N(x) = x$, directly from the definition.

It remains to solve the recurrence. It is easy to compute that $B_{N-1}(x) = (x/2)^2$ and $B_{N-2}(x) = (x/3)^3$, which leads to the conjecture

$$B_n(x) = (x/(N-n+1))^{N-n+1}. \tag{1.3}$$

This can be verified by a straightforward induction. Let us briefly present the calculations, as they will be useful in the identification of the optimal controls. For

$n = N$ the hypothesis is true. Assuming the validity for a fixed $n+1 \in \{2, 3, \ldots, N\}$, we derive that the expression

$$B_{n+1}(x-t) \cdot t = \left(\frac{x-t}{N-n}\right)^{N-n} \cdot t,$$

considered as a function of t, attains its maximum for $t = x/(N-n+1)$ (only). Furthermore, this maximal value is equal to $(x/(N-n+1))^{N-n+1}$. This yields (1.3) and the claim follows.

The above calculations encode, for any fixed $x \geq 0$, the optimal sequences a_1^*, a_2^*, …, a_N^* for $B_N(x)$. To see this, assume that $x > 0$ (for $x = 0$ there is nothing to prove). We go back to the above proof of (1.2). We have

$$B_1(x) = \sup_{t \in [0,x]} \big\{ B_2(x-t) \cdot t \big\},$$

and the supremum is attained for the unique choice $t = x/N$. This necessarily implies that a_1^* must be equal to x/N. Otherwise, we would have

$$B_1(x) = \sup_{t \in [0,x]} \big\{ B_2(x-t) \cdot t \big\} > B_2(x - a_1^*) \cdot a_1^* \geq a_1^* \cdot a_2^* a_3^* \ldots a_N^*,$$

where in the last passage we have exploited the definition of B_2 and the fact that a_2^*, a_3^*, …, a_N^* sum up to $x - a_1^*$. This would contradict the optimality of $(a_k^*)_{k=1}^N$. Therefore we have $a_1^* = x/N$ and $a_2^* + a_3^* + \ldots + a_N^* = (N-1)x/N$. How to get any information on a_2^*, a_3^*, …, a_N^*? Write

$$a_1^* a_2^* \ldots a_N^* = B_1(x) = B_2\left(x - \frac{x}{N}\right) \cdot \frac{x}{N} = B_2\left(x - \frac{x}{N}\right) \cdot a_1^*,$$

or, equivalently (recall that we have assumed $x > 0$)

$$a_2^* a_3^* \ldots a_N^* = B_2\left(x - \frac{x}{N}\right).$$

This brings us to the same position as above, with the length of the unknown extremal sequence decreased by 1 (and the required sum x replaced by $x - x/N$). Repeating the arguments, we show that $a_2^* = (x - x/N)/(N-1) = x/N$, $a_3^* + a_4^* + \ldots + a_N = x - 2x/N$, and the numbers a_3^*, a_4^*, …, a_N^* satisfy

$$a_3^* a_4^* \ldots a_N^* = B_3\left(x - \frac{2x}{N}\right),$$

and so on. The procedure can be carried out until we get all the values of a_1^*, a_2^*, …, a_N^*: one easily checks by induction that $a_1^* = a_2^* = \ldots = a_N^* = x/N$ is the extremal sequence we have searched for. □

Remark 1.1. We could have shown the estimate by considering the “dual” problem

$$\inf \left\{ a_1 + a_2 + \ldots + a_N \,:\, a_1, a_2, \ldots, a_N \geq 0,\, a_1 a_2 \ldots a_N = x \right\}$$

for $x \geq 0$, imbedded into the simpler sub-problems

$$B_n(x) = \inf \left\{ a_n + a_{n+1} + \ldots + a_N \,:\, a_n, a_{n+1}, \ldots, a_N \geq 0,\, a_n a_{n+1} a_{n+2} \ldots a_N = x \right\}$$

for $n = 1, 2, \ldots, N$, $x \geq 0$.

1.2. Two one-dimensional DP problems

We continue the presentation of the examples, starting with dimension 1, i.e., with the case in which the Bellman sequence $(B_n)_n$ consists of functions of one variable.

THEOREM 1.2. *For any nonnegative numbers $a_1, a_2, \ldots, a_N$ we have*

$$(1 + a_1)(1 + a_2)\ldots(1 + a_N) \geq (1 + (a_1a_2\ldots a_N)^{1/N})^N. \tag{1.4}$$

The equality holds if and only if $a_1 = a_2 = \ldots = a_N$.

PROOF. We proceed as previously, but we will use a slightly different enumeration for the Bellman sequence. For any integer $n = 1, 2, \ldots, N$ and $x \geq 0$, let

$$B_n(x) = \inf\left\{(1 + a_1)(1 + a_2)\ldots(1 + a_n)\right\},$$

where the infimum is taken over all sequences $a_1, a_2, \ldots, a_n$ of nonnegative numbers satisfying $a_1a_2\ldots a_n = x$. We need to show that $B_N(x) \geq (1 + x^{1/N})^N$. Directly from the definition, we see that $B_1(x) = 1 + x$ and Bellman optimality principle yields the recurrence

$$B_{n+1}(x) = \inf_{t>0}\left\{(1 + t)B_n\left(\frac{x}{t}\right)\right\}. \tag{1.5}$$

We easily show inductively that $B_n(x) = (1 + x^{1/n})^n$ satisfies this recurrence (and the initial requirement $B_1(x) = 1 + x$). Furthermore, the infimum in (1.5) is attained for $t = x^{1/(n+1)}$ only. Consequently, the desired estimate holds, and the optimal controls $a_1^*, a_2^*, \ldots, a_N^*$ must satisfy $a_{n+1}^* = (a_1^*a_2^*\ldots a_{n+1}^*)^{1/(n+1)}$ for all $n = 1, 2, \ldots, N-1$. The latter condition is equivalent to saying $a_1^* = a_2^* = \ldots = a_N^* = x^{1/N}$ and the proof is complete. □

Now we will turn our attention to a slightly more involved result. The following Hardy-type estimate can be easily established with the use of Schwarz' inequality. However, it is instructive to present an alternative approach based on dynamic programming.

THEOREM 1.3. *For any positive integer n and any real numbers $a_1, a_2, \ldots, a_n$ we have the inequality*

$$\sum_{k=1}^n\left(\frac{a_1 + a_2 + \ldots + a_k}{k}\right) \leq \left(2n - \sum_{k=1}^n\frac{1}{k}\right)^{1/2}\left(\sum_{k=1}^n a_k^2\right)^{1/2}. \tag{1.6}$$

For any n, the constant $\left(2n - \sum_{k=1}^n \frac{1}{k}\right)^{1/2}$ cannot be improved.

PROOF OF (1.6). In the light of the previous estimates, it is natural to study the function

$$B_n(x) = \sup\left\{\sum_{k=1}^n\left(\frac{a_1 + a_2 + \ldots + a_k}{k}\right) : \sum_{k=1}^n a_k^2 = x\right\}$$

for a fixed $x \geq 0$. However, it seems that the successful treatment of such a problem with the use of dynamic programming requires the introduction of another parameter. Indeed, if we try to express $B_{n+1}(x)$ in terms of B_n at some point, we see that we must introduce the "free" variable t (i.e., the variable over which the supremum in the recurrence is taken) equal to a_{n+1}: this enables the handling of the sum $\sum_{k=1}^{n+1} a_k^2 = \sum_{k=1}^{n} a_k^2 + t^2$, as previously. However, we have

$$\sum_{k=1}^{n+1} \left(\frac{a_1 + a_2 + \ldots + a_k}{k} \right) = \sum_{k=1}^{n} \left(\frac{a_1 + a_2 + \ldots + a_k}{k} \right) + \frac{a_1 + a_2 + \ldots + a_n + t}{n+1},$$

and we do not have any control over the second term on the right. To remedy this, let us introduce the additional parameter $y = a_1 + a_2 + \ldots + a_n$. Then, if we set

$$B_n(x, y) = \sup \left\{ \sum_{k=1}^{n} \left(\frac{a_1 + a_2 + \ldots + a_k}{k} \right) \; : \; \sum_{k=1}^{n} a_k^2 = x, \sum_{k=1}^{n} a_k = y \right\},$$

then the optimality principle yields

$$B_{n+1}(x, y) = \sup \left\{ B_n(x - t^2, y - t) + \frac{y}{n+1} \right\},$$

where the supremum is taken over all $t \geq 0$ satisfying $t \leq y$ and $t^2 \leq x$. Thus, we have to perform the analysis of the two-dimensional problem, which in general is much more difficult than the one-dimensional ones.

This raises the question whether the estimate (1.6) can be reformulated so that it can be treated with Bellman sequence which consists of functions of one variable. The answer turns out to be positive and involves the linearization of the inequality first. Namely, the desired estimate is equivalent to

$$\sum_{k=1}^{n} \left(\frac{a_1 + a_2 + \ldots + a_k}{k} \right) - \sum_{k=1}^{n} a_k^2 \leq \frac{n}{2} - \frac{1}{4} \sum_{k=1}^{n} \frac{1}{k}. \tag{1.7}$$

Indeed, the above bound follows from (1.6) and Schwarz' inequality. To prove the reverse implication, note that given a_1, a_2, $\ldots$, a_n, the application of (1.7) to the sequence λa_1, λa_2, $\ldots$, λa_n with $\lambda = \left(\frac{n}{2} - \frac{1}{4} \sum_{k=1}^{n} \frac{1}{k} \right)^{1/2} / \left(\sum_{k=1}^{n} a_k^2 \right)^{1/2}$, gives an inequality equivalent to (1.6).

Now, for any positive integer n and any $x \in \mathbb{R}$, consider the quantity

$$B_n(x) = \sup \left\{ \sum_{k=1}^{n} \left(\frac{a_1 + a_2 + \ldots + a_k}{k} \right) - \sum_{k=1}^{n} a_k^2 \right\},$$

where the supremum is taken over all sequences a_1, a_2, $\ldots$, a_n of real numbers satisfying $a_1 + a_2 + \ldots + a_n = x$. Directly from this definition, we check that

$$B_1(x) = x - x^2. \tag{1.8}$$

Furthermore, Bellman optimality principle becomes, as the reader verifies readily,

$$B_{n+1}(x) = \sup_{t \in \mathbb{R}} \left\{ B_n(x - t) + \frac{x}{n+1} - t^2 \right\}. \tag{1.9}$$

To solve this recurrence, one can compute that

$$B_2(x) = x - \frac{x^2}{2} + \frac{1}{8}, \qquad B_3(x) = x - \frac{x^2}{3} + \frac{1}{8} + \frac{1}{6},$$

which suggests that for general n we have the formula

$$B_n(x) = x - \frac{x^2}{n} + b_n, \tag{1.10}$$

for some $(b_n)_{n\geq 1}$ to be found. We proceed by induction. We have already checked (1.10) for $n = 1, 2, 3$ with $b_1 = 0$, $b_2 = 1/8$ and $b_3 = 1/8 + 1/6$. Straightforward calculations show that the expression

$$B_n(x-t) + \frac{x}{n+1} - t^2 = x - t - \frac{(x-t)^2}{n} + b_n + \frac{x}{n+1} - t^2,$$

considered as a function of t, attains its maximal value at

$$t = (x - n/2)/(n+1), \tag{1.11}$$

equal to $x - x^2 + b_n + \frac{n}{4(n+1)}$. This proves the validity of (1.10), with $b_{n+1} = b_n + \frac{n}{4(n+1)}$. Consequently, we have

$$b_n = \sum_{k=1}^{n-1} \frac{k}{4(k+1)}$$

and by the very definition of $B_n(x)$, we have established the estimate

$$\sum_{k=1}^{n}\left(\frac{a_1 + a_2 + \ldots + a_k}{k}\right) - \sum_{k=1}^{n} a_k^2 \leq x - \frac{x^2}{n} + \sum_{k=1}^{n-1}\frac{k}{4(k+1)}.$$

It remains to observe that $x - x^2/n \leq n/4$ and do some simple manipulations with the right-hand side to obtain (1.7).

To identify the optimal controls $a_1^*, a_2^*, \ldots, a_n^*$, we must ensure that all the inequalities we obtained on our way become equalities. First, note that we must have $x = a_1^* + a_2^* + \ldots + a_n^* = n/2$: then $x - x^2/n = n/4$, so we do not lose anything in the last step of the proof of (1.7). Next, it follows from the proof of (1.9) and the formula (1.11) that for each $k \in \{0, 1, 2, \ldots, n-1\}$, the optimal number a_k^* must satisfy

$$a_{k+1}^* = \frac{a_1^* + a_2^* + \ldots + a_{k+1}^* - k/2}{k+1},$$

or $ka_{k+1}^* = a_1^* + a_2^* + \ldots + a_k^* - k/2$. This recurrence can be solved by induction, and the outcome is $a_k^* = \sum_{j=k}^{n}(2j)^{-1}$. □

1.3. A few higher-dimensional DP problems

The first example is the inequality equivalent to (1.4), but it is a nice starter for two-dimensional problems.

THEOREM 1.4. *For any nonnegative numbers* $a_1, a_2, \ldots, a_N, b_1, b_2, \ldots, b_N$ *we have*

$$(a_1 + b_1)(a_2 + b_2)\ldots(a_N + b_N) \geq ((a_1a_2\ldots a_N)^{1/N} + (b_1b_2\ldots b_N)^{1/N})^N. \tag{1.12}$$

PROOF. We may assume that all numbers a_i and b_j are non-zero (otherwise, the claim is obvious). For any $x, y > 0$, consider the function

$$B_n(x,y) = \sup\left\{(a_1 + b_1)(a_2 + b_2)\ldots(a_n + b_n)\right\},$$

where the supremum is taken over all sequences $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ of positive numbers such that $a_1a_2\ldots a_n = x$ and $b_1b_2\ldots b_n = y$. Clearly, we have $B_1(x,y) = x + y$, and Bellman optimality principle yields

$$B_{n+1}(x,y) = \sup\left\{(s+t)B_n(x/s, y/t)\right\}, \tag{1.13}$$

where the supremum is taken over all s, $t > 0$. One easily shows by induction that the solution to this recurrence is given by $B_n(x,y) = (x^{1/n} + y^{1/n})^n$, and the supremum in (1.13) is attained for s, t such that $y/x = (t/s)^{n+1}$. This establishes the desired inequality; furthermore, we obtain that the optimal controls a_1^*, a_2^*, ..., a_N^*, b_1^*, b_2^*, ..., b_N^* satisfy

$$\frac{a_1^* a_2^* \dots a_{n+1}^*}{b_1^* b_2^* \dots b_{n+1}^*} = \left(\frac{a_{n+1}^*}{b_{n+1}^*}\right)^{1/(n+1)}, \qquad n = 1, 2, \dots, N-1,$$

i.e., the equality in (1.12) is attained if and only if $a_1/b_1 = a_2/b_2 = \dots = a_N/b_N$. Of course, this is in perfect consistency with the assertion of Theorem 1.2. □

The next inequality is related to the inclusion-exclusion principle in probability theory.

THEOREM 1.5. *For any $N \geq 2$ and any numbers a_1, a_2, ..., a_N belonging to $[0,1]$ we have*

$$\prod_{n=1}^{N}(1-a_n) \leq 1 - \sum_{n=1}^{N} a_n + \sum_{1 \leq n < m \leq N} a_n a_m.$$

PROOF. We rewrite the inequality in the equivalent form

$$\prod_{n=1}^{N}(1-a_n) \leq 1 - \sum_{n=1}^{N} a_n + \frac{1}{2}\left(\left(\sum_{n=1}^{N} a_n\right)^2 - \sum_{n=1}^{N} a_n^2\right),$$

which suggests to consider the functions

$$B_n(x,y) = \sup \prod_{n=1}^{N}(1-a_n), \qquad x, y \geq 0,$$

the supremum taken over all sequences a_1, a_2, ..., a_n of numbers in $[0,1]$ such that $a_1 + a_2 + \dots + a_n = x$ and $a_1^2 + a_2^2 + \dots + a_n^2 = y$. The assumption a_1, a_2, ..., $a_n \in [0,1]$ implies that x and y must satisfy certain additional relations: for instance, we must have $y \leq \min\{x, x^2\}$. In other words, the domain of B_n has a little more complicated structure. In general, this is an important issue: the analysis of the Bellman sequence (or Bellman function) involves the careful handling of the corresponding domains (which is already visible from the examples we have studied so far: the domain imposes some restrictions on the controls allowed in the recurrence).

However, in the above example we need not worry about this. Directly from the definition, we have $B_2(x,y) = 1 - x + (x^2 - y)/2$. Furthermore, Bellman optimality principle reads

$$B_{n+1}(x,y) = \sup\left\{(1-t)B_n(x-t, y-t^2)\right\}, \tag{1.14}$$

where the supremum is taken over all $t \in [0,1]$ such that $(x-t, y-t^2)$ belongs to the domain of B_n. Solving this recurrence seems to be a little complicated task (there are several cases to consider). This is why we will develop a slightly different argument, which will be of fundamental importance in our future considerations. Namely, the assertion of the theorem is equivalent to checking that $B_n(x,y) \leq 1 - x + (x^2 - y)/2$. Consider the function $U(x,y) = 1 - x + (x^2 - y)$ given for all x, $y \geq 0$ satisfying $y \leq x^2$. We will show that this auxiliary function satisfies the equation (1.14); however, the crucial difference is that now in the supremum we allow all t such that $(x-t, y-t^2)$ lies in the domain of U. This is a strictly larger set of controls, which will actually make U bigger than B_n (this is why we will sometimes call U a *supersolution*). It is enough to prove that

$$U(x,y) \geq \sup\Big\{(1-t)U(x-t, y-t^2)\Big\}. \tag{1.15}$$

To this end, take t such that $x-t \geq 0$, $y-t^2 \geq 0$ and $y-t^2 \leq (x-t)^2$. Note that

$$\begin{aligned}(1-t)U(x-t, y-t^2) &= (1-t)\left(1-(x-t)+\frac{(x-t)^2-(y-t^2)}{2}\right)\\ &= \left(1-x+\frac{x^2-y}{2}\right)+t\cdot\frac{y-t^2-(x-t)^2}{2}\\ &\leq U(x,y),\end{aligned}$$

since $y-t^2 \leq (x-t)^2$ (the point $(x-t, y-t^2)$ lies in the domain of U). This yields the supersolution property (1.15). The point is that the function U carries all the information necessary to prove the desired bound. Indeed, by (1.15), for any a_1, a_2, …, $a_N \in [0,1]$ and any $n = 2, 3, \ldots, N$,

$$U(a_1+\ldots+a_n, a_1^2+\ldots+a_n^2) \geq (1-a_n)U(a_1+\ldots+a_{n-1}, a_1^2+\ldots+a_{n-1}^2),$$

which by induction implies

$$(1-a_2)\ldots(1-a_N)U(a_1, a_1^2) \leq U(a_1+a_2+\ldots+a_N, a_1^2+a_2^2+\ldots+a_N^2).$$

This is precisely the desired inequality. □

Now we turn our attention to the more involved estimate, the classical L^p estimate of Hardy, Littlewood and Pólya.

THEOREM 1.6. *Let $1 < p < \infty$. Then for any positive numbers a_1, a_2, … and λ_1, λ_2, … we have the inequality*

$$\sum_{n=1}^{\infty}\lambda_n\left(\frac{\lambda_1 a_1+\lambda_2 a_2+\ldots+\lambda_n a_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^p \leq \left(\frac{p}{p-1}\right)^p\sum_{n=1}^{\infty}\lambda_n a_n^p. \tag{1.16}$$

The constant $(p/(p-1))^p$ is the best possible.

REMARK 1.2. Setting $\lambda_1 = \lambda_2 = \ldots$ transforms the inequality into the more familiar form

$$\sum_{n=1}^{\infty}\left(\frac{a_1+a_2+\ldots+a_n}{n}\right)^p \leq \left(\frac{p}{p-1}\right)^p\sum_{n=1}^{\infty}a_n^p,$$

in which the constant $(p/(p-1))^p$ is also optimal.

PROOF OF (1.16). A natural idea is to try to find, for any $N \geq 1$, the best constant $C_{p,N}$ in the truncated inequality

$$\sum_{n=1}^{N}\lambda_n\left(\frac{\lambda_1 a_1+\lambda_2 a_2+\ldots+\lambda_n a_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^p \leq C_{p,N}\sum_{n=1}^{N}\lambda_n a_n^p$$

(to which the dynamic programming can obviously be applied) and then show that $\lim_{N\to\infty} C_{p,N} = (p/(p-1))^p$. However, the resulting system of functional equations seems to be very difficult to solve. We will proceed in a different manner. It is convenient to split the reasoning into a few intermediate steps.

Step 1. Bellman function and optimality principle. Clearly, it suffices to prove the inequality for finite sequences a_1, a_2, ... only (i.e., we may assume that there is N such that $a_N = a_{N+1} = a_{N+2} = \ldots = 0$). Furthermore, we may assume that the sequence a_1, a_2, ... is nonincreasing: indeed, if we replace it by its nonincreasing rearrangement, then the right-hand side does not change, while the left can only increase.

Let C_p be the best constant in the inequality (1.16). The first step is to identify the state space: it will be governed by the values of the initial parameters a_1, λ_1. Set

$$B(a_1, \lambda_1) = \sup \left\{ \sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^p - C_p \sum_{n=1}^{\infty} \lambda_n a_n^p \right\},$$

where the supremum is taken over all positive numbers a_2, a_3, ..., λ_2, λ_3, ... such that a_n's are zero for sufficiently large n. Actually, we will change slightly the definition of B later on - we shall see that in the analysis of (1.16) there is a bit more convenient expression to maximize. To get the functional equation governing B, note that

(1.17)

$$\begin{aligned} &\sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^p - C_p \sum_{n=1}^{\infty} \lambda_n a_n^p \\ &= \lambda_1 a_1^p - C_p \lambda_1 a_1^p + \left[\sum_{n=2}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^p - C_p \sum_{n=2}^{\infty} \lambda_n a_n^p \right]. \end{aligned}$$

Now the idea is to transform the expression in the square brackets into the shape similar to that appearing in the definition of B, by changing the parameters a_1, a_2, ..., λ_1, λ_2, ... appropriately. This is simple: the first summand in the first sum is

$$\lambda_2 \left(\frac{\lambda_1 a_1 + \lambda_2 a_2}{\lambda_1 + \lambda_2} \right)^p,$$

while the n-th summand is

$$\left(\frac{\frac{\lambda_1 a_1 + \lambda_2 a_2}{\lambda_1 + \lambda_2} \cdot (\lambda_1 + \lambda_2) + \lambda_3 a_3 + \ldots + \lambda_n a_n}{(\lambda_1 + \lambda_2) + \lambda_3 + \ldots + \lambda_n} \right)^p.$$

Therefore, we have

$$\begin{aligned} &\sum_{n=2}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^p - C_p \sum_{n=2}^{\infty} \lambda_n a_n^p \\ &\quad = \sum_{n=1}^{\infty} \lambda'_n \left(\frac{\lambda'_1 a'_1 + \lambda'_2 a'_2 + \ldots + \lambda'_n a'_n}{\lambda'_1 + \lambda'_2 + \ldots + \lambda'_n} \right)^p - C_p \sum_{n=1}^{\infty} \lambda'_n (a'_n)^p \\ &\quad\quad - \lambda_1 (a'_1)^p - C_p \lambda_2 a_2^p + C_p \lambda'_1 (a'_1)^p, \end{aligned}$$

where $a'_1 = (\lambda_1 a_1 + \lambda_2 a_2)/(\lambda_1 + \lambda_2)$, $\lambda'_1 = \lambda_1 + \lambda_2$, and $a'_n = a_{n+1}$, $\lambda'_n = \lambda_{n+1}$ for $n = 2, 3, \ldots$. Note that $a'_1 \leq a_1$ (since the sequence $(a_n)_{n \geq 1}$ is nonincreasing) and

$\lambda_1' > \lambda_1$. Plugging this into (1.17), we get

$$\begin{aligned}&\sum_{n=1}^{\infty}\lambda_n\left(\frac{\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^p - C_p\sum_{n=1}^{\infty}\lambda_na_n^p - \lambda_1a_1^p + C_p\lambda_1a_1^p\\ &= \sum_{n=1}^{\infty}\lambda_n'\left(\frac{\lambda_1'a_1'+\lambda_2'a_2'+\ldots+\lambda_n'a_n'}{\lambda_1'+\lambda_2'+\ldots+\lambda_n'}\right)^p - C_p\sum_{n=1}^{\infty}\lambda_n'(a_n')^p - \lambda_1'(a_1')^p + C_p\lambda_1'(a_1')^p\\ &\quad + (\lambda_1'-\lambda_1)(a_1')^p - C_p(\lambda_1'-\lambda_1)^{1-p}(\lambda_1'a_1'-\lambda_1a_1)^p,\end{aligned}$$

since $\lambda_2 = \lambda_1' - \lambda_1$ and $\lambda_2a_2 = \lambda_1'a_1' - \lambda_1a_1$. This identity indicates that it is more convenient to study the function

$$\begin{aligned}&B(a_1,\lambda_1)\\ &= \sup\left\{\sum_{n=1}^{\infty}\lambda_n\left(\frac{\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^p - C_p\sum_{n=1}^{\infty}\lambda_na_n^p + (C_p-1)\lambda_1a_1^p\right\},\end{aligned}$$

where the supremum is taken over the same parameters as previously. The preceding identity yields

$$\begin{aligned}&\sum_{n=1}^{\infty}\lambda_n\left(\frac{\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^p - C_p\sum_{n=1}^{\infty}\lambda_na_n^p - \lambda_1a_1^p + C_p\lambda_1a_1^p\\ &\le B(a_1',\lambda_1') + (\lambda_1'-\lambda_1)(a_1')^p - C_p(\lambda_1'-\lambda_1)^{1-p}(\lambda_1'a_1'-\lambda_1a_1)^p,\end{aligned}$$

and hence

$$B(a_1,\lambda_1) \le \sup\left\{B(a_1',\lambda_1') + (\lambda_1'-\lambda_1)(a_1')^p - C_p(\lambda_1'-\lambda_1)^{1-p}(\lambda_1'a_1'-\lambda_1a_1)^p\right\},$$

where the supremum is taken over all $a_1' \le a_1$ and $\lambda_1' \ge \lambda_1$ such that $\lambda_1'a_1' > \lambda_1a_1$ (the latter estimate being equivalent to $a_2 > 0$). A similar argument, involving taking the supremum over a_1, a_2, $\ldots$ first, gives the reverse estimate. Thus we have established the functional equation

$$\begin{aligned}&B(a_1,\lambda_1)\\ &= \sup\left\{B(a_1',\lambda_1') + (\lambda_1'-\lambda_1)(a_1')^p - C_p(\lambda_1'-\lambda_1)^{1-p}(\lambda_1'a_1'-\lambda_1a_1)^p\right\},\end{aligned} \tag{1.18}$$

the supremum taken over all a_1', λ_1' satisfying the same requirements as previously.

Step 2. On the guess *of C_p and B.* An important homogeneity-type property of B is that

$$B(a,\lambda) = \lambda a^pB(1,1) \qquad \text{for any } a,\ \lambda > 0. \tag{1.19}$$

Indeed, given any sequences $a_1, a_2, \ldots, \lambda_1, \lambda_2, \ldots$ as in the definition of $B(1,1)$, we see that

$$\begin{aligned}
&\lambda a^p \left[\sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}\right)^p - C_p \sum_{n=1}^{\infty} \lambda_n a_n^p + (C_p - 1)\lambda_1 a_1^p\right] \\
&= \sum_{n=1}^{\infty} \lambda\lambda_n \left(\frac{\lambda\lambda_1 aa_1 + \lambda\lambda_2 aa_2 + \ldots + \lambda\lambda_n aa_n}{\lambda\lambda_1 + \lambda\lambda_2 + \ldots + \lambda\lambda_n}\right)^p - C_p \sum_{n=1}^{\infty} \lambda\lambda_n (aa_n)^p \\
&\quad + (C_p - 1)\lambda\lambda_1 (aa_1)^p \\
&\leq B(a, \lambda),
\end{aligned}$$

since $aa_1 = a$ and $\lambda\lambda_1 = \lambda$. Taking the supremum over all $a_1, a_2, \ldots, \lambda_1, \lambda_2, \ldots$ as above gives $B(a,\lambda) \geq \lambda a^p B(1,1)$. The reverse estimate is shown similarly. Plugging (1.19) into the functional equation for B and dividing both sides by λa^p gives

$$B(1,1) = \sup_{\alpha \leq 1,\, r > 1,\, r\alpha > 1} \left\{ r\alpha^p B(1,1) + (r-1)\alpha^p - C_p (r-1)^{1-p} (r\alpha - 1)^p \right\},$$

where we have set $\alpha = a_1'/a_1$ and $r = \lambda_1'/\lambda_1$. Observe that the supremum is attained in the limit if we take $r \to 1$, $\alpha \to 1$. Let us look what happens with the identity when r and α are close to 1, say, $r = 1 + \delta$ and $\alpha = 1 - u\delta$, for some $u \in (0,1)$ and some small $\delta > 0$. We have

$$B(1,1) \geq r\alpha^p B(1,1) + (r-1)\alpha^p - C_p (r-1)^{1-p} (r\alpha - 1)^p.$$

If $u \in (1/p, 1)$ and δ is sufficiently small, then $r\alpha^p < 1$ and the inequality can be rewritten in the form

$$B(1,1) \geq \frac{\delta(1-u\delta)^p - C_p \delta^{1-p} \big((1+\delta)(1-u\delta) - 1\big)^p}{1 - (1+\delta)(1-u\delta)^p}.$$

Letting $\delta \to 0$ gives

$$B(1,1) \geq \frac{1 - C_p(1-u)^p}{pu - 1}.$$

Now we let $u \downarrow 1/p$. If $C_p < (p/(p-1))^p$, then the right-hand side converges to infinity: a contradiction, since $B(1,1) \leq C_p - 1$, by the very definition of B (and the fact that Hardy-Littlewood-Pólya inequality holds with the constant C_p). This shows how the constant $(p/(p-1))^p$ appears in the estimate. Suppose for a moment that $C_p = (p/(p-1))^p$; then letting $u \to 1/p$ returns $B(1,1) \geq \frac{p}{p-1}$. This suggests to assume that we actually have equality: $B(1,1) = \frac{p}{p-1}$ and by (1.19),

$$B(a,\lambda) = \frac{p}{p-1} \lambda a^p.$$

Step 3. A formal proof. Now we will show that the function B and the constant C_p we have guessed in the previous step actually work. The key step is the verification of the identity (1.18). It suffices to check the inequality "$\geq$", since the reverse bound is obtained by letting $a_1' \to a_1$ and $\lambda_1' \to \lambda$. By (1.19), the inequality is equivalent to

$$\frac{p}{p-1} \geq \frac{p}{p-1} r\alpha^p + (r-1)\alpha^p - \left(\frac{p}{p-1}\right)^p (r-1)^{1-p} (r\alpha - 1)^p$$

for $\alpha \in [0,1]$ and $r \geq \alpha^{-1}$. Substitute $x = \alpha^p$ and rewrite the bound in the form

$$F(r,x) = \left(\frac{p}{p-1}\right)^p (r-1)^{1-p}(rx^{1/p}-1)^p + \frac{p}{p-1}(1-rx) - (r-1)x \geq 0.$$

One easily checks that F does not have extremal points inside the domain which, by standard arguments, yields the inequality. So, we have

$$\begin{aligned} B(a_1,\lambda_1) - B(a_1',\lambda_1') &\geq (\lambda_1'-\lambda_1)(a_1')^p - \left(\frac{p}{p-1}\right)^p (\lambda_1'-\lambda_1)^{1-p}(\lambda_1'a_1'-\lambda_1 a_1)^p \\ &= \lambda_2\left(\frac{\lambda_1a_1+\lambda_2a_2}{\lambda_1+\lambda_2}\right)^p - \left(\frac{p}{p-1}\right)^p \lambda_2 a_2^p. \end{aligned}$$

Apply this inequality to the sequence a_1', a_2', …, λ_1', λ_2', …. As the result, we get

$$\begin{aligned} &B\left(\frac{\lambda_1a_1+\lambda_2a_2}{\lambda_1+\lambda_2}, \lambda_1+\lambda_2\right) + B\left(\frac{\lambda_1a_1+\lambda_2a_2+\lambda_3a_3}{\lambda_1+\lambda_2+\lambda_3}, \lambda_1+\lambda_2+\lambda_3\right) \\ &= B(a_1',\lambda_1') - B(a_1'',\lambda_1'') \\ &\geq \lambda_2'\left(\frac{\lambda_1'a_1'+\lambda_2'a_2'}{\lambda_1'+\lambda_2'}\right)^p - \left(\frac{p}{p-1}\right)^p \lambda_2'(a_2')^p \\ &= \lambda_3\left(\frac{\lambda_1a_1+\lambda_2a_2+\lambda_3a_3}{\lambda_1+\lambda_2+\lambda_3}\right)^p - \left(\frac{p}{p-1}\right)^p \lambda_3 a_3^p, \end{aligned}$$

and so on, by induction, if $\Lambda_n = \lambda_1 + \lambda_2 + \ldots + \lambda_n$, $A_n = (\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n)/\Lambda_n$, then

$$B(A_n,\Lambda_n) - B(A_{n+1},\Lambda_{n+1}) \geq \lambda_{n+1}A_{n+1}^p - \left(\frac{p}{p-1}\right)^p \lambda_{n+1}a_{n+1}^p.$$

Write down these inequalities for $n = 1, 2, \ldots, N-1$ and sum them up to obtain

$$B(a_1,\lambda_1) - B(A_N,\lambda_N) \geq \sum_{n=2}^N \lambda_n A_n^p - \left(\frac{p}{p-1}\right)^p \sum_{n=2}^N \lambda_n a_n^p.$$

However, $B(A_N,\lambda_N) \geq 0$ and

$$B(a_1,\lambda_1) = \frac{p}{p-1}\lambda_1 a_1^p \leq \left(\left(\frac{p}{p-1}\right)^p - 1\right)\lambda_1 a_1^p.$$

Combining this with the previous estimate yields

$$\sum_{n=1}^N \lambda_n A_n^p \leq \left(\frac{p}{p-1}\right)^p \sum_{n=1}^N \lambda_n a_n^p,$$

which is the desired assertion, since N was arbitrary. □

REMARK 1.3. The above proof of (1.16) could be shortened significantly. Indeed, as the reader immediately verifies, the essential argumentation for the proof is contained in Step 3 only, where we have derived the crucial inductive argument. The preceding two steps have preparatory character and their purpose is to indicate how the inductive step should look like.

1.4. Integral inequalities

1.4.1. Description of the method. Fix $d \geq 1$. Introduce a function $\Phi : (0,1] \times [0,\infty) \to \mathbb{R}^d$ and, for each $t \in (0,1]$, the auxiliary functional X_t acting on nonnegative functions on $(0,t]$ by the formula

$$X_t(f) = \frac{1}{t}\int_0^t \Phi(u, f(u))\mathrm{d}u, \qquad t \in (0,1]. \tag{1.20}$$

Suppose that all functionals X_t have equal range, denoted by $\mathfrak{D} \subseteq \mathbb{R}^d$. Next, let $F : (0,1] \times [0,\infty) \times \mathfrak{D} \to [0,\infty)$, $G : \mathfrak{D} \to [0,\infty)$ be two given Borel functions. Suppose that we are interested in showing the estimate

$$\int_0^1 F(t, f(t), X_t(f))\,\mathrm{d}t \leq G(X_1(f)) \tag{1.21}$$

for all f. For instance, fix $p \in (1,\infty)$ and let $\Phi(u,v) = (v, v^p)$; then

$$X_t(f) = \left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u,\ \frac{1}{t}\int_0^t f(u)^p\,\mathrm{d}u\right)$$

takes values in the set $\mathfrak{D} = \{(x,y) \in [0,\infty)^2 : x^p \leq y\}$. Now, if we additionally take $F(t,u,x) = F(t,u,x_1,x_2) = x_1^p$ and $G(x) = G(x_1,x_2) = c_p x_2$, where c_p is a given constant, then (1.21) becomes the "localized" Hardy inequality

$$\int_0^1 \left(\frac{1}{t}\int_0^t f(u)\,\mathrm{d}u\right)^p \mathrm{d}t \leq \int_0^1 f^p(t)\,\mathrm{d}t$$

(where by "localized" we mean that the integrals are taken over the interval $[0,1]$; the "non-localized" version would involve integration over the whole half-line $[0,\infty)$). To give another important example, take X_t as above, but set $F(t,u,x_1,x_2) = t^\alpha x_1^q$ and $G(x,y) = c_{p,q} x_2^{q/p}$, where α, q and $c_{p,q}$ are some positive constants. Then (1.21) becomes the "localized" Bliss' inequality

$$\int_0^1 t^\alpha \left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u\right)^q \mathrm{d}t \leq c_{p,q}\left(\int_0^1 f^p(t)\mathrm{d}t\right)^{q/p}.$$

The inequality (1.21) can be studied with the use of a version of dynamic programming. Suppose there exists a family $(U_t)_{t\in(0,1]}$ of functions on $\mathfrak{D}$, satisfying

1° (Nonnegativity) For any $t \in (0,1]$ and $x \in \mathfrak{D}$, we have $U_t(x) \geq 0$.
2° (Majorization) For any $x \in \mathfrak{D}$ we have $U_1(x) \leq G(x)$.
3° (Monotonicity) For any nonnegative f on $(0,1]$, the function

$$t \mapsto U_t(X_t(f)) + \int_t^1 F(u, f(u), X_u(f))\mathrm{d}u$$

is nondecreasing on $(0,1]$.

Lemma 1.1. *If there exists a family satisfying 1°, 2° and 3°, then* (1.21) *holds.*

Proof. Fix a nonnegative function f on $(0,1]$. Applying the majorization condition 2° and then the monotonicity property 3°, we get

$$\begin{aligned} G(X_1(f)) \geq U_1(X_1(f)) &= U_1(X_1(f)) + \int_1^1 F(u, f(u), X_u(f))\mathrm{d}u \\ &\geq U_t(X_t(f)) + \int_t^1 F(u, f(u), X_u(f))\mathrm{d}u \end{aligned}$$

for any $t \in (0,1)$. Combining this with the nonnegativity condition 1°, we obtain

$$\int_t^1 F(u, f(u), X_u(f))\mathrm{d}u \le G(X_1(f))$$

for any t. Letting $t \to 0$ yields the desired bound, by virtue of Fatou's lemma. □

A beautiful fact is that the implication of Lemma 1.1 can be reversed. For any $T \in (0,1]$, consider the function $B_T : \mathfrak{D} \to \mathbb{R}$ given by

$$B_T(x) = \sup \left\{ \int_0^T F(t, f(t), X_t(f))\,\mathrm{d}t \right\},$$

where the supremum is taken over all $f : (0,T] \to [0,\infty)$ satisfying $X_T(f) = x$. The family $(B_T)_{T\in(0,1]}$ is the continuous version of the Bellman sequence $(B_n)_{n=1}^N$ appearing in the preceding considerations.

LEMMA 1.2. *If the inequality* (1.21) *holds true, then the family* $(B_t)_{t\in(0,1]}$ *satisfies the conditions 1°, 2° and 3°.*

PROOF. The property 1° follows from the fact that F is nonnegative, while the majorization 2° is a direct consequence of (1.21). Let us express 3° in a slightly more general form, exhibiting the connection between this condition and the optimality principle. Namely, we will show that for any $0 < S < T \le 1$ we have

$$B_T(x) = \sup \left\{ B_S(X_S(f)) + \int_S^T F(t, f(t), X_t(f))\,\mathrm{d}t \right\}, \tag{1.22}$$

where the supremum is taken over all $f : (0,T] \to [0,\infty)$ satisfying $X_T(f) = x$. To this end, take an arbitrary f on $(0,T]$ satisfying $X_T(f) = x$. Then, directly from the definition of B_S applied to $f|_{(0,S]}$,

$$\begin{aligned}\int_0^T F(t, f(t), X_t(f))\,\mathrm{d}t &= \int_0^S F(t, f(t), X_t(f))\,\mathrm{d}t + \int_S^T F(t, f(t), X_t(f))\,\mathrm{d}t \\ &\le B_S(X_S(f)) + \int_S^T F(t, f(t), X_t(f))\,\mathrm{d}t.\end{aligned}$$

Taking the supremum over all f as above yields the inequality "$\le$" in (1.22). To prove the reverse bound, take an arbitrary function $f : (0,T] \to [0,\infty)$ satisfying $X_T(f) = x$. Next, pick any $\bar{f} : (0,S] \to [0,\infty)$ such that $X_S(\bar{f}) = X_S(f)$, and let us splice f with $\bar{f}$ by the formula $g = \bar{f}\chi_{(0,S]} + f\chi_{(S,T]}$. Observe that if $t \in (S,T]$, then

$$\begin{aligned}X_t(g) = \frac{1}{t}\int_0^t \Phi(u, g(u))\,\mathrm{d}u &= \frac{1}{t}\left(\int_0^S \Phi(u, \bar{f}(u))\,\mathrm{d}u + \int_S^t \Phi(u, f(u))\,\mathrm{d}u \right), \\ &= \frac{1}{t}\left(\int_0^S \Phi(u, f(u))\,\mathrm{d}u + \int_S^t \Phi(u, f(u))\,\mathrm{d}u \right) = X_t(f).\end{aligned}$$

Consequently,

$$\begin{aligned}B_T(x) &\ge \int_0^T F(t, g(t), X_t(g))\,\mathrm{d}t \\ &= \int_0^S F(t, \bar{f}(t), X_t(\bar{f}))\mathrm{d}t + \int_S^T F(t, f(t), X_t(f))\mathrm{d}t,\end{aligned}$$

so taking the supremum over all $\bar{f}$ yields the inequality "$\geq$" in (1.22) and completes the proof. $\square$

Let us also record here the following "minimality principle" for the Bellman family $(B_t)_{t\in(0,1]}$ constructed above.

LEMMA 1.3. *The Bellman family is the pointwise least family satisfying 1°, 2° and 3°.*

PROOF. We will prove that if a family $(U_t)_{t\in(0,1]}$ satisfies 1° and 3°, then we have the pointwise bound $B_t \leq U_t$ for all $t \in (0,1]$. This will clearly yield the claim. For given $T > 0$ and $x \in \mathfrak{D}$, we take a nonnegative function f on $(0,T]$ satisfying $X_T(f) = x$ and observe that

$$\begin{aligned} U_T(x) &= U_T(X_T(f)) + \int_T^T F(u, f(u), X_u(f))\mathrm{d}u \\ &\geq U_t(X_t(f)) + \int_t^T F(u, f(u), X_u(f))\mathrm{d}u \end{aligned}$$

for any $t > 0$. Since $U_t(X_t(f)) \geq 0$, letting $t \to 0$ above yields

$$U_T(x) \geq \int_0^T F(u, f(u), X_u(f))\mathrm{d}u$$

and taking the supremum over all f as above gives $B_T(x) \leq U_T(x)$. $\square$

Summarizing the above discussion, the problem of showing (1.21) is reduced to that of finding a family satisfying 1°, 2° and 3°. While the first two conditions are simple pointwise estimates, the third property might look a little more difficult. However, under appropriate regularity requirements, one may differentiate (with respect to t) the expression on the right in 3°, making it much much easier to handle. To state this observation precisely, let us introduce the function $u(t,x) := U_t(x)$, defined for $(t,x) \in (0,1] \times \mathfrak{D}$.

LEMMA 1.4. *Assume that Φ is a continuous function. Furthermore, suppose that u is continuous, of class C^1 in the interior of its domain and satisfies*

$$u_t(t,x) + \left\langle \nabla_x u(t,x), \frac{\Phi(t,d) - x}{t} \right\rangle - F(t,d,x) \geq 0 \tag{1.23}$$

for any $(t,x) \in (0,1] \times \mathfrak{D}$ and $d \geq 0$. Then 3° holds true for any continuous f.

PROOF. This follows from a direct differentiation of the function

$$t \mapsto U_t(X_t(f)) + \int_t^1 F(u, f(u), X_u(f))\mathrm{d}u. \quad \square$$

Therefore, if the assumptions of the above lemma are satisfied, then the estimate (1.21) holds true for all continuous functions f. In most cases, the passage to general f is carried out by straightforward approximation arguments and the analysis is complete.

Let us present several examples illustrating the above approach.

1.4.2. A weak-type estimate. We will prove the following statement.

THEOREM 1.7. *Let $p \in [0,1)$. Then for any $f \in L^1(0,\infty)$ and any $\lambda > 0$ we have*

$$\lambda \left| \left\{ t > 0 \,:\, t^{p-1} \int_0^t |f(u)| du \geq \lambda \right\} \right|^{1-p} \leq \int_0^\infty |f(u)| du. \tag{1.24}$$

PROOF. Clearly, we may restrict ourselves to nonnegative functions and, by homogeneity, we may assume that $\lambda = 1$. It is enough to show that for any $T > 0$,

$$\left| \left\{ t \in (0,T] \,:\, t^{p-1} \int_0^t f(u)\,\mathrm{d}u \geq 1 \right\} \right|^{1-p} \leq \int_0^T f(u)\,\mathrm{d}s. \tag{1.25}$$

However, if we fix T and substitute $t = Ts$, then the above inequality becomes

$$\left| \left\{ s \in (0,1] \,:\, s^{p-1} \int_0^s f(u)\mathrm{d}u > 1 \right\} \right|^{1-p} \leq \int_0^1 f(u)\mathrm{d}u,$$

with the new function $f(u) := T^p f(Tu)$, $u \in (0,1]$. Hence, the validity of (1.25) for all $T > 0$ is actually equivalent to the validity of the single estimate with $T = 1$. We can rewrite this bound in the form (1.21) as follows. First, we set $\Phi(u,v) = v \in [0,\infty)$ so that $X_t(f) = \frac{1}{t}\int_0^t f(u)\mathrm{d}u \in \mathfrak{D} := [0,\infty)$. Next, take

$$F(t,u,x) = \chi_{\{t^p x > 1\}} \qquad \text{and} \qquad G(x) = x^{1/(1-p)}.$$

The associated Bellman family is given by

$$B_t(x) = \sup \left| \left\{ s \in (0,t] \,:\, s^{p-1} \int_0^s f(u)\mathrm{d}u > 1 \right\} \right|, \qquad t \in (0,1],$$

where the supremum runs over all $f \geq 0$ satisfying $X_t(f) = x$. It is easy to find the above supremum: it is clear that we should consider only those functions, which have very small support concentrated near 0 (for instance, of the form $f^{(a)} = \frac{x}{a}\chi_{[0,a]}$, for some $a \in (0,t]$). Computing the integral over these functions and letting $a \to 0$ leads to the following candidate for B:

$$U_t(x) = \min\left\{ (tx)^{1/(1-p)}, t \right\}.$$

Of course, for any $t \in (0,1]$ we have $U_t \leq B_t$ directly from the definition of the Bellman sequence. To show the reverse bound, it is enough to verify the condition 3° (see Corollary 1.3). To this end, observe first that for any $x \geq 0$, $t \mapsto U_t(x)$ is nondecreasing. Next, pick $0 \leq t < u \leq 1$; we must show that for any f on $(0,u]$,

$$\min\left\{ (uX_u(f))^{1/(1-p)}, u \right\} - \min\left\{ (tX_t(f))^{1/(1-p)}, t \right\} \geq \int_t^u \chi_{\{s^p X_s(f) > 1\}}\mathrm{d}s.$$

Fix t and look at the behavior of both sides as u increases. The right-hand side increases on the open set $A = \{s \in (t,u) : s^p X_s(f) > 1\}$ only and its derivative is equal to 1 there. On the other hand, the left-hand side is a nondecreasing function, and we have $\min\left\{ (sX_s(f))^{1/(1-p)}, s \right\} = s$ on A, implying that the derivative of the left-hand side on A is also equal to 1. This proves 3° and establishes the identity $U_t = B_t$ for all t. Since $B_1(x) \leq G(x)$, the desired inequality follows. □

1.4.3. Hardy inequality in L^p. Our next example is the following.

THEOREM 1.8. *For any $1 < p < \infty$ and any $f \in L^p(0,\infty)$ we have*

$$\int_0^\infty \left|\frac{1}{t}\int_0^t f(u)\,du\right|^p dt \leq \left(\frac{p}{p-1}\right)^p \int_0^\infty |f(t)|^p\,dt, \tag{1.26}$$

and the constant is the best possible.

In the proof of this result, we will need the following auxiliary fact.

LEMMA 1.5. *There is an increasing continuous function $\varphi : [1,\infty) \to \mathbb{R}$, satisfying the differential equation*

$$(p-1)\left[1-\left(\frac{s\varphi'(s)-\varphi(s)}{\varphi'(s)}\right)^{1/(p-1)}\right](s\varphi'(s)-\varphi(s)) = 1 \tag{1.27}$$

for $s \in (1,\infty)$ and the initial condition $\varphi(1) = 1$. Furthermore,

$$\varphi(s) \leq \left(\frac{p}{p-1}\right)^p s \qquad \text{for } s \geq 1. \tag{1.28}$$

PROOF. Consider the function $\psi : [1,\infty) \to \mathbb{R}$ given by

$$\psi(s) = s\left(1-\frac{1}{p}+\frac{1}{ps}\right)^p. \tag{1.29}$$

One easily verifies that this function is smooth, strictly increasing and maps $[1,\infty)$ onto itself. Let φ be the inverse to ψ; then $\varphi(1) = 1$ and we have

$$s = \left(1-\frac{1}{p}+\frac{1}{p\varphi(s)}\right)^p \varphi(s). \tag{1.30}$$

This, by a direct differentiation, yields

$$1 = \varphi'(s)\left(1-\frac{1}{p}+\frac{1}{p\varphi(s)}\right)^p - \frac{\varphi'(s)}{\varphi(s)}\left(1-\frac{1}{p}+\frac{1}{p\varphi(s)}\right)^{p-1},$$

or, equivalently,

$$\frac{1}{\varphi'(s)} = \frac{p-1}{p}\left(1-\frac{1}{p}+\frac{1}{p\varphi(s)}\right)^{p-1}\left(1-\frac{1}{\varphi(s)}\right). \tag{1.31}$$

Multiply both sides by $\varphi(s)$ and subtract the obtained equality from (1.30). We get

$$s-\frac{\varphi(s)}{\varphi'(s)} = \left(1-\frac{1}{p}+\frac{1}{p\varphi(s)}\right)^{p-1} \tag{1.32}$$

and hence (1.31) can be rewritten in the form

$$\frac{1}{\varphi'(s)} = (p-1)\left(s-\frac{\varphi(s)}{\varphi'(s)}\right)\left[1-\left(s-\frac{\varphi(s)}{\varphi'(s)}\right)^{1/(p-1)}\right],$$

which is the desired differential equation (1.27). To show (1.28), note that

$$\left(\frac{\varphi(s)}{s}\right)' = \frac{\varphi'(s)s-\varphi(s)}{s^2} \geq 0,$$

where the latter bound comes from (1.32) (and the estimate $\varphi'(s) > 0$, which is a consequence of the strict monotonicity of ψ). Thus, (1.28) follows at once from

$$\lim_{s\to\infty} \frac{\varphi(s)}{s} = \lim_{s\to\infty} \frac{s}{\psi(s)} = \left(\frac{p}{p-1}\right)^p.$$

This finishes the proof. □

A DIRECT PROOF OF (1.26). Suppose that c_p is the smallest constant such that

$$\int_0^\infty \left|\frac{1}{t}\int_0^t f(u)\,\mathrm{d}u\right|^p \mathrm{d}t \leq c_p \int_0^\infty |f(t)|^p\,\mathrm{d}t$$

for all $f \in L^p(0,\infty)$. Clearly, we may restrict ourselves to nonnegative and continuous functions (having established the above bound for such functions, we easily get the general case by standard density arguments). Using a similar truncation as in the proof of the previous estimate, we see that it is enough to show that

$$\int_0^1 \left(\frac{1}{t}\int_0^t f(u)\,\mathrm{d}u\right)^p \mathrm{d}t \leq c_p \int_0^1 f^p(t)\,\mathrm{d}t, \tag{1.33}$$

i.e., we may deal with functions on $(0,1]$ only. As we have explained above, if we take $\Phi(u,v) = (v, v^p)$, then

$$X_t(f) = \left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u, \frac{1}{t}\int_0^t f^p(u)\mathrm{d}u\right) \in \mathfrak{D} = \{(x,y) \in [0,\infty)^2 : x^p \leq y\}$$

and the choice $F(t,x_1,x_2) = x_1^p$, $G(x_1,x_2) = c_p x_2$ corresponds to (1.33). Write down the formula for the corresponding Bellman family:

$$B_t(x,y) = \sup\left\{\int_0^t \left(\frac{1}{s}\int_0^s f(r)\mathrm{d}r\right)^p \mathrm{d}s\right\},$$

where the supremum is taken over all nonnegative f satisfying $X_t(f) = (x,y)$. Directly from this definition we extract several structural properties of the family. First, observe that $X_t(f) = (x,y)$ if and only if the dilated function $f^{(T)} : u \mapsto f(Tu)$ satisfies $X_{Tt}(f^{(T)}) = (x,y)$. Thus, substituting in the integrals defining the Bellman family, we see that

$$B_t(x,y) = t\sup\left\{\int_0^1 \left(\frac{1}{s}\int_0^s f^{(t)}(r)\mathrm{d}r\right)^k \mathrm{d}s\right\} = tB_1(x,y).$$

Another important property comes from the observation that for any $\lambda > 0$, we have $X_t(f) = (x,y)$ if and only if $X_t(\lambda f) = (\lambda x, \lambda^p y)$. This implies $B_t(\lambda x, \lambda^p y) = \lambda^p B_t(x,y)$, and hence, taking $\lambda = 1/x$, we obtain

$$B_t(x,y) = tB_1(x,y) = tx^p B_1(1, y/x^p)$$

for all (x,y,t). Denote $\varphi(s) = B_1(s)$: all we need is to find the explicit formula for φ (there is no abuse of notation: as we will see shortly, φ is precisely the function studied in Lemma 1.5 above). Here the reasoning will be a little informal: we will obtain the desired function after imposing some (reasonable) assumptions. The key indication for the search is contained in the inequality (1.23). Plugging $u(t,x,y) = B_t(x,y)$ there, we get

$$x^p\varphi(y/x^p) + p\left(x^p\varphi(y/x^p) - y\varphi'(y/x^p)\right)\left(\frac{d}{x} - 1\right) + \varphi'(y/x^p)(d^p - y) - x^p \geq 0$$

for all $d \geq 0$. Divide both sides by x^p and substitute $s = y/x^p$, $r = d/x$, obtaining

$$\varphi(s) + p(\varphi(s) - s\varphi'(s))(r-1) + \varphi'(s)(r^p - s) - 1 \geq 0. \tag{1.34}$$

This inequality should hold for all $r \geq 0$, if we want our method to work. Suppose for a while that $\varphi'(s) > 0$ (when we have found the candidate for φ, we will verify that this condition is satisfied). Then, optimizing over r, we see that the above inequality yields

$$(p-1)\left[1 - \left(s - \frac{\varphi(s)}{\varphi'(s)}\right)^{1/(p-1)}\right](s\varphi'(s) - \varphi(s)) - 1 \geq 0.$$

Since we are studying the sharp estimate, it is natural to expect that we actually should have equality here. This leads us to the differential equation (1.27). To get the initial condition for $\varphi(1) = B_1(1,1)$, note that there is only one continuous function f for which $X_1(f) = (1,1)$, the constant function $f \equiv 1$. By the definition of B_1, we get $\varphi(1) = 1$ and hence φ is the function constructed in Lemma 1.5 (in particular, we have $\varphi' > 0$ which we needed above).

The above reasoning has brought us the candidate U_t given by $U_t(x,y) = tx^p\varphi(y/x^p)$, $t \in (0,1]$. By the very construction, we see that the family $(U_t)_{t\in(0,1]}$ satisfies 1° and 3°. Furthermore, (1.28) shows that 2° is also true, if we only assume that $c_p \geq (p/(p-1))^p$. Hence, we set $c_p = (p/(p-1))^p$: the method implies that the inequality (1.33), and hence also the inequality of Theorem 1.8, are valid. □

Sharpness. A beautiful feature of the approach is that it also gives an indication about the extremal functions in (1.26). The idea is very simple: we want to find an f such that all the intermediate inequalities appearing during the application of the method become equalities. Let us be more specific. As we already know, for a given smooth $f : (0,1] \to [0,\infty)$, the function

$$t \mapsto B(X_t(f), Y_t(f), t) + \int_t^1 F(X_s(f), Y_s(f), s)\mathrm{d}s$$

is nondecreasing. This, by the direct differentiation, gives (1.34) with the parameters $s = Y_t(f)/X_t^k(f)$ and $r = f(t)/X_t(f)$. Then the optimization over r leads to (1.27): the extremal choice for r is $(s - \varphi(s)/\varphi'(s))^{1/(p-1)}$, as one easily verifies. This, in the light of (1.32), is equal to $1 - p^{-1} + (p\varphi(s))^{-1}$, or just $(s/\varphi(s))^{1/p}$, by the definition of the function ψ. Since the extremal function f should correspond to this extremal choice of r, we should have

$$\frac{f(t)}{X_t(f)} = \frac{(Y_t(f)/X_t^p(f))^{1/p}}{\varphi(Y_t(f)/X_t^k(f))^{1/p}}, \tag{1.35}$$

or $\varphi(Y_t(f)/X_t^p(f)) = Y_t(f)/f^p(t)$. Plugging this into the definition (1.29) of ψ and working a little bit, we get

$$\frac{f(t)}{X_t(f)} = 1 - \frac{1}{p} + \frac{f^p(t)}{pY_t(f)},$$

or

$$\frac{f(t)}{\int_0^t f(s)\mathrm{d}s} = \left(1 - \frac{1}{p}\right)t^{-1} + \frac{f^p(t)}{p\int_0^t f^p(s)\mathrm{d}s}.$$

The latter inequality means that the function

$$t \mapsto \log\left[\int_0^t f(s)\mathrm{d}s\right] - \left(1 - \frac{1}{p}\right)\log t - \frac{1}{p}\log\left[\int_0^t f^p(s)\mathrm{d}s\right]$$

is constant. Equivalently, we have $Y_t(f)/X_t^p(f) = c$ for some constant c. Coming back to (1.35), we see that f and $X_\cdot(f)$ should be proportional: $f(t) = CX_t(f)$, where $C = c^{1/p}/\varphi(c)^{1/p}$. This, in turn, implies $tf(t) = C\int_0^t f(s)\mathrm{d}s$ and hence $f(t) = at^{C-1}$ for some constant a. One easily verifies that this candidate indeed yields the sharpness of (1.33) and hence also (1.26): it suffices to take $a = 1$ and C close to (but a little larger than) $1 - 1/p$. □

A SIMPLER PROOF OF THEOREM 1.8. The above direct application of the method may look a little complicated, but fortunately one can simplify the argumentation. It turns out that a proper reformulation of the inequality leads us to Bellman functions of one variable only. To see this, rewrite (1.33) in the form

$$\int_0^1 \left\{ \left(\frac{1}{t} \int_0^t f(u)\,\mathrm{d}u \right)^p - c_p f^p(t) \right\} \mathrm{d}t \leq 0.$$

Take $\Phi(u, v) = v$: then $X_t(f) = \frac{1}{t}\int_0^t f(u)\mathrm{d}u \in \mathfrak{D} := [0, \infty)$, and the choice $F(t, u, x) = x^p - c_p u^p$, $G(x) \equiv 0$ transforms (1.21) into the above bound. The evident obstacle we face here is that the function F may take negative values; on the other hand, in the above abstract discussion on the method we have used the inequality $F \geq 0$ several times. However, as we will see, we can modify the arguments at appropriate places and make the technique applicable also in this more general setting.

So, let us proceed as above and introduce the associated Bellman family by

$$B_T(x) = \sup \left\{ \int_0^T \left(\frac{1}{t} \int_0^t f(u)\,\mathrm{d}u \right)^p - c_p f^p(t)\,\mathrm{d}t \right\}, \qquad x \geq 0,\ T \in (0, 1],$$

where the supremum is taken over all nonnegative and continuous functions f on $(0, T]$ satisfying $\frac{1}{T}\int_0^T f(u)\mathrm{d}u = x$. Arguing as above, we show the following structural properties of $(B_t)_{t\in(0,1]}$:

$$B_T(x) = TB_1(x) \qquad \text{for all } x \geq 0 \text{ and } T \in (0, 1],$$

and

$$B_T(x) = x^p B_T(1) \qquad \text{for all } x \geq 0 \text{ and } T \in (0, 1].$$

Putting these two observations together, we see that

$$B_T(x) = -\alpha T x^p, \tag{1.36}$$

for some constant α depending only on p; we have put the minus sign above, since B should be nonpositive, if the inequality (1.33) is to hold. Let us try to verify the validity of (1.23). One easily computes that it reads

$$-\alpha p x^{p-1}(d - x) + (-\alpha - 1)x^p + c_p d^p \geq 0.$$

For a fixed x, the left-hand side, considered as a function of d, attains its minimal value for d satisfying $d = (\alpha/c_p)^{1/(p-1)}x$. Plugging this above, we see that α must satisfy the equation

$$\frac{(1-p)\alpha^{p/(p-1)}}{c_p^{1/(p-1)}} + \alpha(p - 1) - 1 \geq 0.$$

The left-hand side, considered as a function of α, attains its maximum for $\alpha = \left(\frac{p}{p-1}\right)^{1-p} c_p$, and this maximal value is equal to $\left(\frac{p}{p-1}\right)^{-p} c_p - 1$. Therefore, summarizing the above calculations, if we take $c_p = \left(\frac{p}{p-1}\right)^p$, then there is α (equal to $p/(p-1)$) such that the family $(B_T)_{T\in(0,1]}$, given by (1.36), satisfies (1.23). Consequently, arguing as in the proof of Lemma 1.1, we get

$$\begin{aligned} 0 = G(X_1(f)) &\geq B_t(X_t(f)) + \int_t^1 F(u, f(u), X_u(f))\mathrm{d}u \\ &= -\alpha t \left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u\right)^p + \int_t^1 \left\{\left(\frac{1}{s}\int_0^s f(u)\,\mathrm{d}u\right)^p - c_p f^p(s)\right\}\mathrm{d}s. \end{aligned}$$

The point is that we cannot remove the term $-\alpha t\left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u\right)^p$, since it is negative; furthermore, letting $t \to 0$ in the second term is also problematic, since the integrand takes negative values. However, these obstacles are easily removed. We rewrite the inequality in the form

$$c_p^p \int_t^1 f^p(u)\mathrm{d}u \geq -\alpha t \left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u\right)^p + \int_t^1 \left(\frac{1}{s}\int_0^s f(u)\,\mathrm{d}u\right)^p \mathrm{d}s \tag{1.37}$$

and let $t \to 0$. Then $\int_t^1 f^p(u)\mathrm{d}u \to \int_0^1 f^p(u)\mathrm{d}u$ and, by Hölder's inequality,

$$t\left(\frac{1}{t}\int_0^t f(u)\mathrm{d}u\right)^p \leq \int_0^t f^p(u)\mathrm{d}u \to 0.$$

Finally, the integrand in the last integral in (1.37) is nonnegative. Therefore, Fatou's lemma yields (1.33). □

1.5. Problems

1. Prove that for any $n \geq 1$ and any numbers $a_1, a_2, \ldots, a_n \in [0,1]$ we have

$$(1-a_1)(1-a_2)\ldots(1-a_n) \geq 1 - a_1 - a_2 - \ldots - a_n.$$

2. Prove that for any $n \geq 1$ and any real numbers $a_1, a_2, \ldots, a_n$ we have

$$a_1 a_2 \ldots a_n \leq \frac{a_1^2}{2} + \frac{a_2^4}{4} + \frac{a_3^8}{8} + \ldots + \frac{a_n^{2^n}}{2^n} + \frac{1}{2^n}.$$

3. Show that for any $n \geq 2$ and any positive numbers $a_1, a_2, \ldots, a_n$ we have

$$\sum_{k=1}^n \frac{1}{1+a_k} \geq \min\left\{1, \frac{n}{1+(a_1 a_2 \ldots a_n)^{1/n}}\right\}.$$

4. (J. V. Whittaker) Nonnegative numbers $a_1, a_2, \ldots, a_N$ satisfy the condition $\sum_{n=1}^N (1+a_n)^{-1} \leq a$. Find the maximal value of the expression $\sum_{n=1}^N 2^{-a_n}$.

5. Prove that for any positive numbers $a_1, a_2, \ldots, a_n$ satisfying $a_1 + a_2 + \ldots + a_n < 1$ we have

$$\frac{a_1 a_2 \ldots a_n\big(1 - (a_1 + a_2 + \ldots + a_n)\big)}{(a_1 + a_2 + \ldots + a_n)(1-a_1)(1-a_2)\ldots(1-a_n)} \leq \frac{1}{n^{n+1}}.$$

6. Prove that for any positive numbers $a_1, a_2, \ldots, a_n$ satisfying $a_1 a_2 \ldots a_n = 1$ we have

$$\frac{1}{n-1+a_1} + \frac{1}{n-1+a_2} + \ldots + \frac{1}{n-1+a_n} \leq 1.$$

7. Prove that for any positive numbers $a_1, a_2, \ldots, a_n$ we have

$$\left(1+\frac{1}{a_1}\right)\left(1+\frac{1}{a_2}\right)\ldots\left(1+\frac{1}{a_n}\right) \geq \left(1+\frac{n}{a_1+a_2+\ldots+a_n}\right)^n.$$

8. Prove that for any positive numbers $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ we have

$$\left(\frac{a_1+a_2+\ldots+a_n}{b_1+b_2+\ldots+b_n}\right)^2 \leq \left(\frac{a_1}{b_1}\right)^2+\left(\frac{a_2}{b_2}\right)^2+\ldots+\left(\frac{a_n}{b_n}\right)^2.$$

9. (R. Bellman [**1**]) At any time, a particle can be in one of two states, called S and T. At the beginning, the particle is in the state T. We perform a sequence of operations of type A and L. If the operation A is used and the probability that the particle is in the state T is equal to x, then this probability decreases to ax ($0 < a < 1$ is a given constant). The operation L consists of observing the particle and tells us definitely which state it is in. Given the strategy, let τ be the number of operations after which the particle is observed in S (with certainty). Find the strategy giving the minimal expectation of τ.

10. (E. Cerri, M. Silvestri and S. Villan) Let $g : [1, \infty) \to \mathbb{R}$ be a continuous function and let

$$V_n(x) = \inf\left\{g(a_1)+\frac{g(a_2)}{a_1}+\frac{g(a_3)}{a_1a_2}+\ldots+\frac{g(a_n)}{a_1a_2\ldots a_{n-1}}\right\},$$

where the infimum is taken over all $a_1, a_2, \ldots, a_n \geq 1$ satisfying $a_1a_2\ldots a_n = x$. Show that for any $n \geq 1$ we have

$$V_{n+1}(x) = \inf_{t\geq 1}\left\{g(t)+\frac{1}{t}V_n\left(\frac{x}{t}\right)\right\}.$$

Solve the problem explicitly for $g(y) = y^p$, $p > 0$.

11. (T. Carleman [**4**]) Using dynamic programming, prove that for any nonnegative numbers $a_1, a_2, \ldots$ we have the sharp estimate

$$\sum_{n=1}^{\infty}(a_1a_2\ldots a_n)^{1/n} \leq e\sum_{n=1}^{\infty}a_n.$$

12. (G. H. Hardy, J. E. Littlewood, G. Pólya [**10**]) Using dynamic programming, prove that if $0 < p < 1$ and $a_1, a_2, \ldots$ is a sequence of nonnegative numbers such that $\sum_{n=1}^{\infty} a_n^p < \infty$, then we have the sharp estimate

$$\left(1+\frac{1}{1-p}\right)\left(\frac{a_1+a_2+\ldots}{1}\right)^p+\sum_{n=2}^{\infty}\left(\frac{a_n+a_{n+1}+\ldots}{n}\right)^p > \left(\frac{p}{1-p}\right)^p\sum_{n=1}^{\infty}a_n^p.$$

13. (F. Carlson [**5**]) Prove that for any $f : [0, \infty) \to \mathbb{R}$ we have

$$\int_0^{\infty}|f(t)|\mathrm{d}t \leq \sqrt{\pi}\left(\int_0^{\infty}f^2(u)\mathrm{d}u\right)^{1/4}\left(\int_0^{\infty}f^2(u)u^2\mathrm{d}u\right)^{1/4}.$$

CHAPTER 2

Optimal stopping

2.1. Discrete time

2.1.1. Martingale approach: description. 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

$$V_0 = \sup_\tau \mathbb{E} G_\tau, \tag{2.1}$$

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

$$\mathbb{E} \sup_{n\geq 0} |G_n| < \infty, \tag{2.2}$$

then the expectation $\mathbb{E}G_\tau$ is well defined for all stopping times τ. Another possibility is to restrict in (2.1) 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 (2.2) is satisfied, but it will be evident how to relax this requirement in other contexts.

So, let us assume that the supremum in (2.1) 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

$$V_n^N = \sup_{\tau \in \mathcal{M}_n^N} \mathbb{E} G_\tau, \tag{2.3}$$

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 (2.3) with the use of martingale approach.

2.1.2. Martingale approach: finite horizon. If $N < \infty$ (the case of "finite horizon"), then the problem (2.3) 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:

$$\begin{aligned} B_N^N &= G_N, \\ B_n^N &= \max\left\{G_n, \mathbb{E}(B_{n+1}^N|\mathcal{F}_n)\right\}, \qquad n = N-1,\, N-2,\, \ldots. \end{aligned} \tag{2.4}$$

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

$$\tau_n^N = \inf\left\{k \in \{n,\, n+1,\, \ldots,\, N\} \;:\; B_k^N = G_k\right\}, \tag{2.5}$$

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

THEOREM 2.1. *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* (2.3) *and the sequence $(B_k^N)_{k=n}^N$, defined by* (2.4).

(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,

$$B_n^N \geq \mathbb{E}(G_\tau|\mathcal{F}_n) \qquad \textit{for any } \tau \in \mathcal{M}_n^N, \tag{2.6}$$

$$B_n^N = \mathbb{E}(G_{\tau_n^N}|\mathcal{F}_n). \tag{2.7}$$

(iii) The stopping time τ_n^N is optimal in (2.3) *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, $k = n,\, n+1,\, N+2,\, \ldots,\, N$, 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 $\left(B^N_{\tau^N_n\wedge k}\right)_{k=n}^N$. We compute directly that

$$\begin{aligned}\mathbb{E}\left[B^N_{\tau^N_n\wedge(k+1)}\,|\,\mathcal{F}_k\right] &= \mathbb{E}\left[B^N_{\tau^N_n\wedge(k+1)}1_{\{\tau^N_n\le k\}}\,|\,\mathcal{F}_k\right]+\mathbb{E}\left[B^N_{\tau^N_n\wedge(k+1)}1_{\{\tau^N_n> k\}}\,|\,\mathcal{F}_k\right]\\ &= \mathbb{E}\left[B^N_{\tau^N_n\wedge k}1_{\{\tau^N_n\le k\}}\,|\,\mathcal{F}_k\right]+\mathbb{E}\left[B^N_{k+1}1_{\{\tau^N_n> k\}}\,|\,\mathcal{F}_k\right]\\ &= B^N_{\tau^N_n\wedge k}1_{\{\tau^N_n\le k\}}+1_{\{\tau^N_n> k\}}\mathbb{E}\left[B^N_{k+1}\,|\,\mathcal{F}_k\right].\end{aligned}$$

However, on the set $\{\tau^N_n > k\}$ we have $B^N_k > G_k$ and hence $B^N_k = \mathbb{E}(B^N_{k+1}|\mathcal{F}_k)$. This shows the identity $\mathbb{E}\left[B^N_{\tau^N_n\wedge(k+1)}\,|\,\mathcal{F}_k\right] = B^N_{\tau^N_n\wedge k}1_{\{\tau^N_n\le k\}}+1_{\{\tau^N_n> k\}}B^N_k = B^N_{\tau^N_n\wedge k}$ and part (i) is established.

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

(iii) Taking the expectations in (2.6) and (2.7) gives $\mathbb{E}G_\tau \le \mathbb{E}B^N_n = \mathbb{E}G_{\tau^N_n}$ 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. □

Remark 2.1. The following observation is an immediate consequence of the above considerations. Namely, as we have just proved, the stopping time τ^N_0 is optimal for V^N_0. If it was not optimal to stop at the instances 0, 1, …, $n-1$, then the optimal policy coincides with that corresponding to the problem V^N_n. This naturally brings us into the context of dynamic programming studied in the previous chapter.

2.1.3. 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 (2.6) and (2.7) 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 (2.6) and (2.7) 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 2.1. 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 $\mathbb{P}(\overline{Z} \leq \tilde{Z}) = 1$.

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 $\max\{Z_\alpha, Z_\beta\} \leq Z_\gamma$, then the countable set $J = \{\alpha_1, \alpha_2, \ldots\}$ can be chosen so that $Z_{\alpha_1} \leq Z_{\alpha_2} \leq \ldots$ and $\operatorname{ess\,sup}_{\alpha \in I} Z_\alpha = \lim_{n \to \infty} Z_{\alpha_n}$.

Now we see that (2.6) and (2.7) give the identity

$$B_n^N = \operatorname*{ess\,sup}_{\tau \in \mathcal{M}_n^N} \mathbb{E}(G_\tau | \mathcal{F}_n) \tag{2.8}$$

with probability 1. A nice feature of this alternative characterization of the sequence $(B_n^N)_{n=0}^N$ 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 (2.3) for $N = \infty$:

$$V_n = \sup_{\tau \geq n} \mathbb{E} G_\tau. \tag{2.9}$$

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

$$B_n = \operatorname*{ess\,sup}_{\tau \geq n} \mathbb{E}(G_\tau | \mathcal{F}_n) \tag{2.10}$$

and the stopping time

$$\tau_n = \inf\{k \geq n : B_k = G_k\}, \tag{2.11}$$

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 2.1.

THEOREM 2.2. *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 (2.9). 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 (2.9). Furthermore, if τ_* is another optimal stopping time for (2.9), 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 2.1. 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. □

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 (2.5), (2.8), (2.10) and (2.11) that

$$B_n^\infty \leq B_n \qquad \text{and} \qquad \tau_n^\infty \leq \tau_n \tag{2.12}$$

almost surely. Therefore, we also have

$$V_n^\infty \leq V_n. \tag{2.13}$$

THEOREM 2.3. *Suppose that* $\mathbb{E}\sup_{n\geq 0}|G_n| < \infty$ *and consider the optimal stopping problems* (2.3) *and* (2.9). *Then equalities hold in* (2.12) *and* (2.13).

PROOF. Letting $N \to \infty$ in the recurrence relation (2.4) 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. □

EXAMPLE 2.1. Strict inequalities may hold in (2.12) and (2.13) if the integrability condition on $\sup_{n\geq 0}|G_n|$ is not imposed. To see this, let ε_0, ε_1, ε_2, … 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.

2.1.4. An application: a difference prophet inequality for bounded random variables. We take the opportunity to present here some information about the so-called "prohet 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, … 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 2.3 below. Instead, we will study the following related *difference* prophet inequality.

THEOREM 2.4. *If G_1, G_2, … 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 2.1. *If X is a random variable independent of Y and Y_a^b, then*

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

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

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

Therefore, since X and Y are independent, we get

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

Thus we obtain

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

□

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 2.2. *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

$$V_1^n = V_1^n(G_1, G_2, \dots, G_n) = \mathbb{E}\max\{G_1, \mathbb{E}\max\{G_2, \dots\}\}$$
$$= \mathbb{E}\max\{\mu, \mathbb{E}\max\{G_2, \dots\}\} + \mathbb{E}(G_1 - \mu)^+$$

(note that the first summand in the last expression is just $\mathbb{E}\max\{\mu,\mu\} = \mu$, but we write it in the above more complicated form). Furthermore,

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

and hence $D(G_1, G_2, \dots, G_n) \le D(\mu, G_2, \dots, G_n)$. 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

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

and hence $V^n(\mu, G_2, \dots, G_n) = V_n(\mu, G_2, \dots, G_{n-2}, Y, Z)$. By the previous lemma,

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

so $D(\mu, G_2, \dots, G_n) \le D(\mu, G_2, \dots, G_{n-2}, Y, Z)$. 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 $\mathbb{E}W = \mathbb{E}\max\{Y, \mathbb{E}Z\}$ and $\mathbb{E}Z = \mathbb{E}X_n \le \mu$, so

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

and the proof is complete. □

Proof of Theorem 2.4. Clearly, it suffices to show the claim for finite number of variables. By the second lemma above, we may reduce this number to 2: $D(G_1, G_2) \le \frac{1}{4}$. However, the above proof shows that if we set $\mu = \mathbb{E}G_2$, then $D(G_1, G_2) \le D(\mu, Z)$, where $\mathbb{P}(Z=1) = \mu = 1 - \mathbb{P}(Z=0)$. It suffices to note that

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

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 $\mathbb{P}(G_2 = 0) = \mathbb{P}(G_2 = 1) = 1/2$. □

2.2. Markovian approach

Throughout this section, we assume that $X = (X_0, X_1, X_2, \dots)$ is a Markov family defined on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\ge 0}, (P_x)_{x\in E})$, 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 $d \ge 1$, 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

$$\mathbb{E}_x\left(\sup_{0\le n\le N} |G(X_n)|\right) < \infty \qquad \text{for all } x \in E. \tag{2.14}$$

Consider the associated finite-horizon optimal stopping problem

$$V^N(x) = \sup \mathbb{E}_x G(X_\tau), \tag{2.15}$$

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 (2.4) and (2.5). We also introduce the sets

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

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

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

Since $V^0 = G$, by the very definition (2.15), we see that $X_N \in D_N$ and hence the stopping time τ_D is finite (it does not exceed N).

THEOREM 2.5. *Assume that the function G satisfies the integrability condition* (2.14) *and consider the optimal stopping problem* (2.15).

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

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

$$V^n(x) = \max\{G(x), TV^{n-1}(x)\}, \qquad x \in E, \tag{2.16}$$

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

(iii) The stopping time τ_D is optimal in (2.15). *If τ_* is another optimal stopping time, then $\tau_D \leq \tau_*$ $\mathbb{P}_x$-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 $\left(V^{N-n\wedge\tau_D}(X_{n\wedge\tau_D})\right)_{n=0}^N$ is a $\mathbb{P}_x$-martingale.

PROOF. We only need to establish (i) and (ii); the remaining parts follow at once from Theorem 2.1. To verify (i), recall that

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

for all $n = 0, 1, 2, \ldots, N$. This shows the claim for $n = 0$, by the very definition of $V^N(x)$. On the other hand, for $n \geq 1$ 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 (2.15). Define the operator Q acting on $f \in I$ by the formula

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

COROLLARY 2.1. *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 2.2. Let $(S_n)_{n\geq 0}$ be a symmetric random walk over the space $E = \{-2, -1, 0, 1, 2\}$ stopped at $\{-2, 2\}$. Clearly, $(S_n)_{n\geq 0}$ is a Markov family on E. Set $G(x) = x^2(x+2)$ and consider the optimal stopping problem

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

To treat the problem successfully, we compute the sequence V^0, V^1, V^2, V^3, ... Directly from (2.16), we have

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

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

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

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

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

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)

$$V(x) = \sup \mathbb{E}_x G(X_\tau), \qquad x \in E, \tag{2.17}$$

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 2.2. 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 2.3. *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,

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

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

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

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 2.6. *Consider the optimal stopping problem* (2.17) *and assume that*

$$\mathbb{E}_x \sup_{n\ge 0} |G(X_n)| < \infty, \qquad x \in E. \tag{2.18}$$

Then the following holds.

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

$$V(x) = \max\{G(x), TV(x)\}. \tag{2.19}$$

(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 $\tau_* \ge \tau_D$ $\mathbb{P}_x$-almost surely.

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

(iv) The stopped sequence $(V(X_{\tau_D \wedge n}))_{n\ge 0}$ 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 2.3. $\square$

Let us make here an important comment on the uniqueness of the solutions to the Wald-Bellman equations (2.16) and (2.19). Clearly, in the case of finite horizon there is only one solution: indeed, the starting function V^0 coincides with G and the formula (2.16) 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 $c' \ge c$ satisfies the Wald-Bellman equation. However, any solution to (2.19) is a superharmonic function majorizing G, so part (iii) of Theorem 2.6 immediately yields the following "minimality principle".

COROLLARY 2.2. *The value function V is the minimal solution to* (2.19).

EXAMPLE 2.3. Let us provide solution to the infinite-horizon version of Example 2.2. 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 2.6 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 2.4. 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 (2.18) (with $G(x) = x^+$) is satisfied. This follows from the well-known fact that

$$\mathbb{P}\big(\sup_{n\geq 0}(\xi_1 + \xi_2 + \ldots + \xi_n) \geq k\big) = \left(\frac{p}{q}\right)^k, \qquad k = 0, 1, 2, \ldots. \tag{2.20}$$

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

$$0 < V(x) - G(x) \le V(x-1) - G(x-1) \tag{2.21}$$

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,

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

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 (2.20) that $V(x) \to 0$ as $x \to -\infty$ (simply use the estimate $\mathbb{E}_x S_\tau^+ \le \mathbb{E}_x \sup_{n\ge 0} S_n^+$): this implies $\alpha = 0$ and $\beta \ge 0$. This also shows that $b < \infty$. Indeed, otherwise $V(x)$ would explode exponentially as $x \to \infty$, but on the other hand, by (2.21), for $x > 0$ we would have

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

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

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

Secondly, again by Wald-Bellman equation, we see that $V(b) \ge pV(b+1) + qV(b-1)$, which is equivalent to $b \ge p/(q-p)$. 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 $b \ge p/(q-p)$, then the function

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

is excessive. Let us now check for which b the inequality $\mathcal{V} \ge G$ 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

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

Observe that by (2.20), the stopping time

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

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 $M \to -\infty$.

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

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

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

$$\mathbb{E} \sup_{0 \leq n \leq N} G_n \leq \sup_{\tau \in \mathcal{M}_0^N} \mathbb{E} G_\tau + \left(\frac{n}{n+1} \right)^{n+1}$$

and show that the constant $(n/(n+1))^{n+1}$ cannot be decreased.

4. 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

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

5. 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

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

6. (Chow and Robbins [**6**]) Let G_0, G_1, G_2, ... be i.i.d. nonnegative random variables and let $c > 0$ be a fixed constant. Solve the optimal stopping problem

$$V = \sup_{\tau \in \mathcal{M}} \mathbb{E} \Big[\max\{G_0, G_1, G_2, \ldots, G_\tau\} - c\tau \Big].$$

7. (Robbins [**18**]) 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

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

8. 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) integrable stopping times τ, (iii) bounded stopping times τ.

9. 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 got at least three heads. What is the strategy which yields the largest expected gain?

10. 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

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

11. (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.

CHAPTER 3

Inequalities for martingale transforms and differentially subordinated processes

3.1. Description of the method

We turn our attention to another class of estimates arising in the probability theory and harmonic analysis. Though we will mainly focus on the martingale setting, we should emphasize that the results we will obtain have their analytic counterparts which can be expressed in terms of unconditional-type properties of the Haar system. Furthermore, the martingale inequalities which will be studied can be applied to obtain corresponding tight estimates for wide classes of Fourier multipliers. In other words, despite the probabilistic language, the contents of this chapter is meaningful from the point of view of harmonic analysis.

As we will see below, our approach to the estimates for martingale transforms/under the differential subordination is quite similar to that used in the previous chapter in the context of Markovian optimal stopping. We will use an analogous notation to make this similarity even more visible. We assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, filtered by $(\mathcal{F}_n)_{n\geq 0}$, a non-decreasing family of sub-σ-algebras of $\mathcal{F}$. Let $f = (f_n)_{n\geq 0}$, $g = (g_n)_{n\geq 0}$ be real-valued martingales, with the difference sequences $df = (df_n)_{n\geq 0}$, $dg = (dg_n)_{n\geq 0}$ given by

$$df_0 = f_0, \qquad df_n = f_n - f_{n-1}, \quad n = 1, 2, \ldots,$$

and similarly for dg. The martingale f is *simple*, if for every n the random variable f_n takes only a finite number of values. We say that g is a transform of f, if there is a predictable sequence $v = (v_n)_{n\geq 0}$ such that $dg_n = v_n df_n$ for each $n \geq 1$ (here by predictability of v we mean that for any $n \geq 0$, the variable v_n is $\mathcal{F}_{(n-1)\vee 0}$-measurable). This definition is slightly different from that used in the literature, where it is assumed that the equality $dg_n = v_n df_n$ holds also for $n = 0$ (in such a case, we will say that g is a *full* transform of f by v). We will say that g is a (full) ± 1-transform of f, if the transforming sequence v takes values in the set $\{-1, 1\}$.

The central problem of this chapter can be formulated as follows. Let $G : \mathbb{R}^2 \to \mathbb{R}$ be a given function and suppose that we are interested in the inequality

$$\mathbb{E}G(f_n, g_n) \leq 0, \tag{3.1}$$

for all n and all pairs (f, g) of simple martingales such that g is a full transform of f by a predictable sequence bounded in absolute value by 1 (i.e., we assume that for any n, the random variable v_n takes values in $[-1, 1]$). Note that we do not need to assume anything about the regularity of integrability of G: the fact that f and g are simple guarantees the existence of the expectation. Motivated by the results obtained in the preceding chapter, we may search, for each pair (f, g) as above, for a supermartingale $(U_n)_{n\geq 0}$ majorizing $(G(f_n, g_n))_{n\geq 0}$ and satisfying

$U_0 \le 0$. Clearly, if such a supermartingale exists, then (3.1) holds true. The Bellman function approach rests on reducing the search to the supermartingales of the form $(U(f_n, g_n))_{n\ge 0}$ for some special function U to be found; now the condition $U_0 \le 0$, the majorization $U_n \ge G(f_n, g_n)$ and the supermartingale property translate into appropriate pointwise properties of U. Specifically, consider the following conditions a function $U : \mathbb{R}^2 \to \mathbb{R}$ might satisfy.

1° (initial condition) We have $U(x, y) \le 0$ for all $|y| \le |x|$.

2° (majorization) We have $U \ge G$ on $\mathbb{R}^2$.

3° (concavity) The function U is concave along any line of slope in $[-1, 1]$.

The statement below describes the relation between the above set of conditions and the desired estimate (3.1).

LEMMA 3.1. *If there is a function U satisfying 1°, 2° and 3°, then* (3.1) *holds.*

PROOF. Fix any pair (f, g) of simple martingales such that g is the full transform of f by a certain predictable sequence v bounded in absolute value by 1. The concavity condition implies that $(U(f_n, g_n))_{n\ge 0}$ is a supermartingale: for any $n \ge 0$ we have

$$\begin{aligned}\mathbb{E}\big[U(f_{n+1}, g_{n+1})|\mathcal{F}_n\big] &= \mathbb{E}\big[U(f_n + df_{n+1}, g_n + dg_{n+1})|\mathcal{F}_n\big] \\ &= \mathbb{E}\big[U(f_n + df_{n+1}, g_n + v_{n+1}df_{n+1})|\mathcal{F}_n\big] \\ &\le U(f_n, g_n),\end{aligned}$$

by 3° and the conditional version of Jensen's inequality. Therefore, applying 2° and then 1°, we obtain that for each n,

$$\mathbb{E}G(f_n, g_n) \le \mathbb{E}U(f_n, g_n) \le \mathbb{E}U(f_0, g_0) \le 0. \quad \square$$

There is a natural question about the existence of a special function U and how to identify this object. To understand this issue, we will adopt the reasoning developed in the preceding chapter. Our starting observation is that the inequality (3.1) is closely related to the following optimal stopping problem

$$V^N = \sup \mathbb{E}G(f_N, g_N), \tag{3.2}$$

where the supremum is taken over all pairs (f, g) of simple martingales such that g is a full transform of f by a predictable sequence bounded in absolute value by 1. Here the filtration is also allowed to vary as well as the probability space. At the first glance, the phrase "optimal stopping problem" seems a bit misleading here, since no stopping times enter the above problem. However, a little thought reveals that the stopping times actually do appear: one easily checks that if g is a full transform of f and $\tau \le N$ is a stopping time, then the stopped process g^τ is a full transform of f^τ (by the same sequence). Thus (3.2) can indeed be regarded as an optimal stopping problem, and the significant complication (in comparison to the preceding chapter) lies in the fact that we do not stop a given Markov process, but a much larger class of martingale pairs.

The problem (3.2) is of finite-horizon-type and admits a natural version for $N = \infty$:

$$V = V^\infty = \sup \mathbb{E}G(f_n, g_n), \tag{3.3}$$

where the supremum is taken over all n and all pairs (f, g) as above. Motivated by the theory of optimal stopping, our first step is to enlarge the class of problems so that the initial values of the martingales f, g can be taken into account. Namely, for each $(x,y)\in\mathbb{R}$ and any N, set

$$V^N(x,y)=\sup\mathbb{E}G(f_N,g_N), \tag{3.4}$$

where the supremum is taken over all pairs (f,g) of simple martingales starting from (x,y) such that g is a transform of f by some predictable sequence bounded in absolute value by 1. An infinite-horizon version of (3.4) is introduced analogously:

$$V(x,y)=\sup\mathbb{E}G(f_n,g_n), \tag{3.5}$$

where the supremum is taken over all n and all pairs (f,g) as above.

THEOREM 3.1. *Consider the finite-horizon problem* (3.4). *The sequence* $(V^n)_{n\geq 0}$ *can be computed inductively from the relations* $V^0=G$ *and, for* $n\geq 1$,

$$V^n(x,y)=\sup\Big\{\alpha_1V^{n-1}(x+t_1,y+at_1)+\alpha_2V^{n-1}(x+t_2,y+at_2)\Big\}, \tag{3.6}$$

where the supremum is taken over all numbers α_1, $\alpha_2\geq 0$, t_1, $t_2\in\mathbb{R}$ *and* $a\in[-1,1]$ *such that* $\alpha_1+\alpha_2=1$ *and* $\alpha_1t_1+\alpha_2t_2=0$.

The equality (3.6) can be regarded as a version of Wald-Bellman equation. It has a very nice geometric meaning: to find $V^n(x,y)$, we take an arbitrary $a\in[-1,1]$ and consider the restriction of V^{n-1} to the line of slope a passing through the point (x,y), i.e., the function $t\mapsto V^{n-1}(x+t,y+at)$. If ζ_a is the smallest concave function which majorizes this restriction, then $V^n(x,y)$ is the supremum of $\zeta_a(0)$ over all a. Directly from this interpretation, we see that (3.6) can be rewritten in the form

$$V^n(x,y)=\sup\mathbb{E}V^{n-1}(x+\xi,y+a\xi), \tag{3.7}$$

where the supremum is taken over all random variables a with values in $[-1,1]$ and all simple centered random variables ξ satisfying $\mathbb{E}(\xi|a)=0$.

PROOF. The equality $V^0=G$ follows from the very definition of the sequence $(V^n)_{n\geq0}$. To show the recurrence (3.6), fix α_1, α_2, t_1, t_2 and a as in its statement and take any martingales (f^1,g^1) and (f^2,g^2) as in the definition of $V^{n-1}(x+t_1,y+at_1)$ and $V^{n-1}(x+t_2,y+at_2)$, respectively. Suppose that these pairs are given on two probability spaces $(\Omega,\mathcal{F}^1,\mathbb{P}^1)$ and $(\Omega^2,\mathcal{F}^2,\mathbb{P}^2)$. Let us glue these pairs into one pair (f,g). To this end, let $\Omega=\Omega^1\cup\Omega^2$, $\mathcal{F}=\sigma(\mathcal{F}^1\cup\mathcal{F}^2)$ and define the probability measure $\mathbb{P}$ on $\mathcal{F}$ by requiring that $\mathbb{P}(A_1\cup A_2)=\alpha_1\mathbb{P}^1(A_1)+\alpha_2\mathbb{P}^2(A_2)$ for any $A_1\in\mathcal{F}^1$ and $A_2\in\mathcal{F}^2$. The pair (f,g) is given by $(f_0,g_0)=(x,y)$ and

$$(f_k(\omega),g_k(\omega))=\begin{cases}(f^1_{k-1}(\omega),g^1_{k-1}(\omega)) & \text{if } \omega\in\Omega^1,\\ (f^2_{k-1}(\omega),g^2_{k-1}(\omega)) & \text{if } \omega\in\Omega^2.\end{cases}$$

Notice that the sequence (f,g) is a martingale with respect to its natural filtration $(\mathcal{F}_n)_{n\geq0}$: this follows at once from the fact that (f^1,g^1), (f^2,g^2) are martingales and the identity

$$\mathbb{E}((f_1,g_1)|\mathcal{F}_0)=\mathbb{E}(f_1,g_1)=\alpha_1(x+t_1,y+at_1)+\alpha_2(x+t_2,y+at_2)=(x,y).$$

Consequently, by the very definition of the functions V^n and V^{n-1},

$$V^n(x,y)\geq\mathbb{E}G(f_n,g_n)=\alpha_1\mathbb{E}^1G(f^1_{n-1},g^1_{n-1})+\alpha_2\mathbb{E}^2G(f^2_{n-1},g^2_{n-1})$$

(where $\mathbb{E}^i$ is the expectation with respect to $\mathbb{P}^i$), so taking the supremum over all (f^1, g^1), (f^2, g^2) as above, we get

$$V^n(x,y) \geq \alpha_1 V^{n-1}(x+t_1, y+at_1) + \alpha_2 V^{n-1}(x+t_2, y+at_2).$$

Taking the supremum over all α_i, t_i and a, we obtain the inequality "$\geq$" in (3.6) and hence also in (3.7). To show the reverse estimate, take an arbitrary pair (f,g) as in the definition of $V^n(x,y)$ and note that

$$V^{n-1}(f_1, g_1) \geq \mathbb{E}\big[G(f_n, g_n)|(f_1, g_1)\big],$$

by the very definition of V^{n-1} applied conditionally on the σ-algebra generated by (f_1, g_1). Therefore

$$\mathbb{E}G(f_n, g_n) = \mathbb{E}\left[\mathbb{E}(G(f_n, g_n)|(f_1, g_1)\right] \leq \mathbb{E}V^{n-1}(x + df_1, y + v_0 df_1)$$

does not exceed the right-hand side of (3.7). Taking the supremum over all (f,g) we get the claim. $\square$

The passage to the case of infinite horizon requires a simple limiting argument. Indeed, it follows directly from (3.2) and (3.3) that the functional sequence $(V^n)_{n\geq 0}$ is pointwise increasing (any martingale f_0, f_1, f_2, $\ldots$, f_{n-1} of length n can be treated as a martingale f_0, f_1, f_2, $\ldots$, f_{n-1}, f_{n-1} of length n) and $V(x,y) = V^\infty(x,y) = \lim_{n\to\infty} V^n(x,y)$. Therefore, we obtain the following

Corollary 3.1. *Consider the problem* (3.5). *Then V is the smallest function which satisfies the conditions 2° and 3°. Furthermore, if the inequality* (3.1) *is valid, then V satisfies 1° as well.*

Proof. We have $V^n \geq V^0 = G$, so indeed V majorizes G. Letting $n \to \infty$ in (3.6), we see that

$$V(x,y) = \sup\left\{\alpha_1 V(x+t_1, y+at_1) + \alpha_2 V(x+t_2, y+at_2\right\},$$

where the supremum is taken over all t_i, α_i and a as above. This implies that V has the required concavity property. To see that V is the smallest, fix an arbitrary function $\mathcal{V}$ satisfying 2° and 3°, a point (x,y) and a pair (f,g) as in the definition of $V(x,y)$. As we have seen in the proof of Lemma 3.1, then the sequence $(\mathcal{V}(f_n, g_n))_{n\geq 0}$ is a supermartingale majorizing $(G(f_n, g_n))_{n\geq 0}$, so

$$\mathbb{E}G(f_n, g_n) \leq \mathbb{E}\mathcal{V}(f_n, g_n) \leq \mathbb{E}\mathcal{V}(f_0, g_0) = \mathcal{V}(x,y).$$

Taking the supremum over all (f,g) yields the desired bound $V(x,y) \leq \mathcal{V}(x,y)$. It remains to show that if (3.1) holds true, then 1° holds. This follows immediately from the fact that if $|y| \leq |x|$, the martingale pair (f,g) starts from (x,y) and g is a transform of f by a predictable sequence v bounded in absolute value by 1, then actually g is a full transform of f (with $v_0 \in [-1,1]$ determined by the condition $y = v_0 x$). $\square$

The above discussion concerns the case when the transforming sequence takes values in $[-1,1]$. However, the approach can be easily modified to apply to other possibilities, e.g. to the case of ± 1-transforms (when $v_n \in \{-1,1\}$ for all n), or to the case when $v_n \in [0,1]$ for all n, etc. Let us be more precise. Suppose we are interested in proving (3.1) for all (f,g) such that g is a full transform of f by a predictable sequence v taking values in some fixed set $A \subset \mathbb{R}$. A straightforward modification of the above arguments shows that the validity of this estimate is equivalent to the existence of a function U satisfying

1° (initial condition) We have $U(x,y) \le 0$ for all (x,y) such that $y = vx$ for some $v \in A$.

2° (majorization) We have $U \ge G$ on $\mathbb{R}^2$.

3° (concavity) The function U is concave along any line of slope in A.

An important observation is that the passage from the set $[-1,1]$ to $\{-1,1\}$ simplifies the analysis and in many cases does not reduce the validity of the results. More precisely, we will frequently perform the following procedure. Suppose that we want to establish (3.1) for all f, g such that g is a full transform of f by a predictable sequence with values in $[-1,1]$. The first step is to consider the more restrictive (and a little simpler) case of ± 1-transforms first and identify the corresponding special function U. The second step is to verify that this function U actually satisfies all the properties needed for the more general case of $[-1,1]$-valued transforming sequences. This phenomenon occurs in most interesting estimates and is what one might expect: when studying the $[-1,1]$-case, it seems reasonable that the extremal martingales (for which the equality or almost equality occurs) should be constructed with the use of extremal transforming sequences, with values in $\{-1,1\}$. Analogous reasoning can be applied when reducing the case of $[0,1]$-valued to $\{0,1\}$-valued transforming sequences, and so on.

Actually, the special functions U obtained via the analysis of ± 1-transforms often lead to much wider class of inequalities for martingales satisfying the so-called differential subordination, a condition which is of significant importance from the point of view of applications. Here is the precise definition.

DEFINITION 3.1. A martingale g is differentially subordinate to f if for any $n \ge 0$ we have $|dg_n| \le |df_n|$ almost surely.

Of course, if g is a transform of f by a predictable sequence with values in $[-1,1]$, then g is differentially subordinate to f. However, the differential subordination allows a much wider class of martingale pairs (f,g). Nevertheless, a similar approach to that used above can be used to the study of estimates in this new setting. Suppose that $G : \mathbb{R}^2 \to \mathbb{R}$ is a given function and assume we are interested in showing the inequality (3.1) for any pair (f,g) of simple martingales such that g is differentially subordinate to f. We should stress here that in a typical situation, the assumption on the simplicity of the sequences can be skipped, mostly often by imposing some boundedness-type conditions on f; for the sake of clarity, we will work with simple martingales only. Suppose that $U : \mathbb{R}^2 \to \mathbb{R}$ satisfies the following properties:

1° (initial condition) We have $U(x,y) \le 0$ for all (x,y) such that $|y| \le |x|$.

2° (majorization) We have $U \ge G$ on $\mathbb{R}^2$.

3° (concavity) For any $(x,y) \in \mathbb{R}^2$ there are two numbers $A = A(x,y)$ and $B = B(x,y)$ such that if h, $k \in \mathbb{R}$ satisfy $|k| \le |h|$, then

$$U(x+h, y+k) \le U(x,y) + A(x,y)h + B(x,y)k. \tag{3.8}$$

As we have mentioned above, in many cases the special function obtained from the analysis of the corresponding estimate for ± 1-transforms satisfies these conditions; in 3°, typically one takes the partial derivatives $A(x,y) = U_x(x,y)$ and $B(x,y) = U_y(x,y)$, or their one-sided versions.

LEMMA 3.2. *If there is a function* U *satisfying* $1°$, $2°$ *and* $3°$, *then* (3.1) *is valid for any pair* (f, g) *such that* g *is differentially subordinate to* f.

PROOF. As in the case of transforms, the main ingredient of the proof is the supermartingale property of the sequence $(U(f_n, g_n))_{n\geq 0}$. Fix $n \geq 0$ and apply $3°$ to $x = f_n$, $y = g_n$, $h = df_n$ and $k = dg_n$ (the condition $|k| \leq |h|$ follows from the differential subordination) to obtain

$$U(f_{n+1}, g_{n+1}) \leq U(f_n, g_n) + A(f_n, g_n)dg_{n+1} + B(f_n, g_n)dg_{n+1}.$$

Taking the conditional expectation with respect to $\mathcal{F}_n$ yields the desired supermartingale property. The remainder of the proof is as previously:

$$\mathbb{E}G(f_n, g_n) \leq \mathbb{E}U(f_n, g_n) \leq \mathbb{E}U(f_0, g_0) \leq 0. \quad \Box$$

The remaining part of this section is a list of useful comments and observations which will be used frequently in our later considerations.

(a) In general, the special function U satisfying $1°$, $2°$ and $3°$, if it exists, is not unique. In some situations the choice of the appropriate function does simplify the calculations involved.

(b) In many cases, the special function inherits certain structural properties from G. For example, if G is symmetric with respect to x-variable (i.e., we have $G(x, y) = G(-x, y)$ for all x, y), then we may search for U in the class of functions enjoying this property. That is, if there is U satisfying $1°$, $2°$ and $3°$, then there exists a function $\mathcal{U}$ which enjoys these conditions as well as the additional property $\mathcal{U}(x, y) = \mathcal{U}(-x, y)$ for all x, y. To see this, one easily verifies that

$$\mathcal{U}(x, y) = \min\{U(x, y), U(-x, y)\},$$

or

$$\mathcal{U}(x, y) = \frac{U(x, y) + U(-x, y)}{2},$$

is such a function. Alternatively, one can check that the function V given by (3.5) meets this requirement (directly from the definition). Similarly, if G is homogeneous of order p and there exists a function U satisfying $1°$, $2°$ and $3°$, then there is also a function $\mathcal{U}$ enjoying these properties which is homogeneous of order p. Indeed, it suffices to take

$$\mathcal{U}(x, y) = \inf_{\lambda > 0} \left\{\lambda^p U(x/\lambda, y/\lambda)\right\},$$

or check that the solution to (3.5) has the required property.

(c) In the above discussion we have considered the real-valued martingales f and g only. This can be easily modified to the case when the sequences take values in some other domains. For instance, suppose we are interested in showing (3.1) under the assumption that f takes values in $[0, 1]$ and g is its transform by a predictable sequence with values in $[-1, 1]$ (in particular, g *may* take negative values). Then only some minor straightforward modifications of the approach are required. Namely, one needs to construct a function on $[0, 1] \times \mathbb{R}$ which satisfies

$1°$ (initial condition) We have $U(x, y) \leq 0$ for all $|y| \leq x \leq 1$.

$2°$ (majorization) We have $U \geq G$ on $[0, 1] \times \mathbb{R}$.

$3°$ (concavity) The function U is concave along any line segment of slope in $[-1, 1]$, contained in $[0, 1] \times \mathbb{R}$.

Analogously, one can extend the method so that it works for Hilbert or Banach-space valued sequences. We will not pursue the discussion in this direction. See the bibliographical details at the end of the chapter.

(d) It should also be mentioned here that in the above martingale context, there are very natural analogues of the continuation and the stopping regions C and D which were so meaningful in the previous chapter. These sets can be defined with the same formula as in the theory of optimal stopping:

$$C = \{(x,y) \in R^2 : V(x,y) > G(x,y)\}, \qquad D = \{(x,y) \in \mathbb{R}^2 : V(x,y) = G(x,y)\}.$$

If one looks at the formula (3.5), these sets have very natural interpretation. Namely, if (x,y) belongs to the set D, then the optimal choice for the pair (f,g) (for which the supremum defining $V(x,y)$ is attained) is just the constant pair $(f,g) \equiv (x,y)$. In other words, when starting from the set D, it is optimal to "stop" the martingale pair instantly. In contrast, when $(x,y) \in C$, the optimal pair (f,g) for $V(x,y)$ must make some nontrivial moves, that is, it must "continue" its evolution.

3.2. Examples

In this section we will illustrate the above method on several important examples. We assume throughout that all the processes we work with are simple.

Example 3.1. Suppose that g is differentially subordinate to f and we are interested in the best constant C in the inequality

$$\lambda \mathbb{P}(|g_n| \geq \lambda) \leq C\mathbb{E}|f_n|^2, \qquad n = 0,\, 1,\, 2,\, \ldots, \tag{3.9}$$

where $\lambda > 0$ is a fixed parameter. Note that we may assume that $\lambda = 1$, replacing f, g with f/λ and g/λ, respectively (this replacement does not affect the differential subordination). There are two objects to be determined: a priori unknown optimal value of the constant C and an appropriate special Bellman function. As we have explained above, we first restrict ourselves to the case of ± 1-transforms and hope that the function U we will obtain will also work in the general setting. We rewrite the estimate in the form

$$\mathbb{E}G(f_n, g_n) \leq 0,$$

with $G(x,y) = 1_{\{|y| \geq 1\}} - C|x|^2$. We will identify the sequence V^0, V^1, V^2, ... By the very definition, $V^0 = G$. Note that the function $\mathcal{V}(x,y) = 1 - C|x|^2$ is concave; this, by a straightforward induction, implies $V^n \leq \mathcal{V}$ on $\mathbb{R}^2$. Since $G = \mathcal{V}$ outside the strip $\mathbb{R} \times (-1,1)$, we see that

$$V^n(x,y) = \mathcal{V}(x,y) = 1 - C|x|^2 \qquad \text{if } |y| \geq 1.$$

To find V^n on the remaining part of the domain, fix $(x,y) \in \mathbb{R} \times (-1,1)$ and write down the "Wald-Bellman" equation for V^1:

$$V^1(x,y) = \sup\Big\{\alpha_1 G(x+t_1, y+at_1) + \alpha_2 G(x+t_2, y+at_2)\Big\}.$$

Here the supremum is taken over all $\alpha_i \geq 0$, $t_i \in \mathbb{R}$ and $a = \pm 1$ such that $\alpha_1 + \alpha_2 = 1$ and $\alpha_1 t_1 + \alpha_2 t_2 = 0$. So, we need to find the least concave majorant

ζ_1 of $t \mapsto G(x+t, y+t)$, the least majorant ζ_{-1} of $t \mapsto G(x+t, y-t)$ and set $V^1(x,y) = \max\{\zeta_1(0), \zeta_{-1}(0)\}$. The function $\xi(t) = G(x+t, y+t)$ has the formula

$$\xi(t) = \begin{cases} -C(x+t)^2 & \text{if } t \in (-1-y, 1-y), \\ 1 - C(x+t)^2 & \text{otherwise} \end{cases}$$

and hence

$$\zeta_1(0) \geq \frac{1+y}{2}\xi(1-y) + \frac{1-y}{2}\xi(-1-y) = 1 - C + C(y^2 - x^2). \tag{3.10}$$

This inequality yields a useful lower bound for C. Namely, if we take $x = y \in (0,1)$, then, by 1°,

$$0 \geq V(x,y) \geq V^1(x,y) \geq 1 - C + C(y^2 - x^2) = 1 - C,$$

or $C \geq 1$. under this additional assumption, it is not difficult to check that

$$\zeta_1(t) = \begin{cases} 1 - C + C((y+t)^2 - (x+t)^2) & \text{if } t \in (-1-y, 1-y), \\ 1 - C(x+t)^2 & \text{otherwise,} \end{cases}$$

and

$$\zeta_{-1}(t) = \begin{cases} 1 - C + C((y-t)^2 - (x+t)^2) & \text{if } t \in (-1+y, 1+y), \\ 1 - C(x+t)^2 & \text{otherwise,} \end{cases}$$

so $V^1(x,y) = 1 - C + C(y^2 - x^2)$ for $(x,y) \in \mathbb{R} \times (-1,1)$. Now, one easily checks that V^1 is concave along the lines of slope ± 1. This implies that the sequence $(V^n)_{n \geq 0}$ stabilizes: we have $V^1 = V^2 = \ldots = V$. Since $V^1(x,y) \leq 0$ for $y = \pm x$, we have proved the validity of (3.9) for any $C \geq 1$ (under the assumption that g is a ± 1-transform of g). Therefore, $C = 1$ is the best constant there.

Now we can pass to the more general classes of martingales. Setting $C = 1$, one can check that the function

$$V(x,y) = \begin{cases} y^2 - x^2 & \text{if } |y| < 1, \\ 1 - x^2 & \text{if } |y| \geq 1 \end{cases}$$

is concave along lines of slope belonging to $[-1,1]$ and satisfies $V(x,y) \leq 0$ for $|y| \leq |x|$. Therefore, we obtain that the inequality (3.9) holds for transforming sequences with values in $[-1,1]$. Actually, this inequality holds even in the less restrictive setting of differential subordination. Indeed, the appropriate versions of the initial and majorization conditions are satisfied, so it remains to verify the concavity inequality (3.8). We set $A(x,y) = -2x$, $B(x,y) = 2y1_{\{|y|<1\}}$ (note that these are essentially the partial derivatives of V) and consider two cases. If $|y| \geq 1$, we use the fact that the function $\mathcal{V}(x,y) = 1 - x^2$ is concave and majorizes V. Therefore, for *any* h, $k \in \mathbb{R}$,

$$\begin{aligned} V(x+h, y+k) &\leq \mathcal{V}(x+h, y+k) \\ &\leq \mathcal{V}(x,y) + \mathcal{V}_x(x,y)h + \mathcal{V}_y(x,y)k \\ &= V(x,y) + A(x,y)h + B(x,y)k. \end{aligned}$$

On the other hand, if $|y| < 1$ and $|k| \leq |h|$, then

$$\begin{aligned} V(x,y) + A(x,y)h + B(x,y)k &= (y+k)^2 - (x+h)^2 + h^2 - k^2 \\ &\geq (y+k)^2 - (x+h)^2 \\ &\geq V(x+h, y+k). \end{aligned}$$

This proves the validity of (3.9) under the assumption of the differential subordination of g to f.

We conclude this lengthy analysis of the weak-type inequality by the observation that the function V we used above is the smallest possible, but not the "simplest possible". There is a different, nicer choice for the function which yields the validity of (3.9): it is easy to see that $U(x,y)=y^2-x^2$ has all the required properties. See the Remark (a) above.

However, in general the computation of the whole sequence V^0, V^1, V^2, ... is a formidable task and the calculations are hard to push through. Therefore, a typical approach rests on the direct search for the function V. Let us move to the next example.

EXAMPLE 3.2. We will identify the best constant C_1 in the estimate

$$\lambda\mathbb{P}(|g_n|\geq\lambda)\leq C_1\mathbb{E}|f_n|, \qquad n=0,\,1,\,2,\,\ldots,$$

where $\lambda>0$ and g is assumed to be differentially subordinate to f. As previously, we start the analysis in the more restrictive setting of ±1-transforms and assume that $\lambda=1$. Then the problem can be rewritten in the form (3.1) with $G(x,y)=1_{\{|y|\geq1\}}-C_1|x|$. To gain some intuition about the special function to be found, let us write down the definition of V:

$$V(x,y)=\sup\left\{\mathbb{P}(|g_n|\geq1)-C_1\mathbb{E}|f_n|\right\}, \tag{3.11}$$

the supremum taken over the usual parameters. We divide the analysis into a few intermediate steps.

Step 1. The case $|y|\geq1$. We know that V is the smallest majorant of G which is concave along the lines of slope ±1. On the other hand, observe that the function $\mathcal{V}(x,y)=1-C_1|x|$ is concave and majorizes G: this implies $V\leq\mathcal{V}$ on the whole $\mathbb{R}^2$. Since $\mathcal{V}$ and G coincide on $\{(x,y):|y|\geq1\}$, we must have

$$V(x,y)=G(x,y)=1-C_1|x| \qquad \text{provided } |y|\geq1.$$

Step 2. The case $|x|+|y|\geq1$. Actually, the above equality holds for all x, y satisfying $|x|+|y|\geq1$. This can be easily seen geometrically, by looking at the graph of the function G; the formal proof goes as follows. Suppose that $x,\,y\geq0$ (for the remaining cases the reasoning is analogous) and $y<1$. We have

$$V(x,y)\geq\frac{1-y}{2}V\left(x+y+1,-1\right)+\frac{1+y}{2}V\left(x+y-1,1\right)=1-C_1x=\mathcal{V}(x,y).$$

Since $\mathcal{V}\geq V$ on the whole $\mathbb{R}^2$, we must actually have equality above.

Step 3. The case $|x|+|y|<1$. Here we make a guess. Take a line of slope 1 passing through (x,y): this line intersects the set $\{(x,y):|x|+|y|=1\}$ at two points

$$P_1=\left(\frac{1+x-y}{2},\frac{1-x+y}{2}\right), \qquad P_2=\left(\frac{-1+x-y}{2},\frac{-1-x+y}{2}\right).$$

We have, by the concavity of V,

$$V(x,y)\geq\frac{1+x+y}{2}V(P_1)+\frac{1-x-y}{2}V(P_2)=1-C_1(1+x^2-y^2)/2.$$

We assume that we actually have equality here. Since $V(0,0)$ must be nonpositive, this implies $C_1 \geq 2$ (note that this is a formal proof that the weak-type constant cannot be smaller than 2). This leads us to the candidate

$$U(x,y) = \begin{cases} 1 - C_1(1+x^2-y^2) & \text{if } |x|+|y|<1, \\ 1 - C_1|x| & \text{if } |x|+|y| \geq 1. \end{cases}$$

It is easy to check that this function has all the required properties and the inequality $\mathbb{P}(|g_n| \geq 1) \leq C_1 \mathbb{E}|f_n|$ is established (in the setting of ± 1-transforms). Since $C_1 \geq 2$ was arbitrary, we see that the weak-type constant is equal to 2.

The passage to more general contexts of $[-1,1]$-valued transforming sequences and the setting of differential subordination goes along the same lines as previously. Clearly, it is enough to study the less restrictive context of differential subordination. Fix $C_1 = 2$. Then the function

$$U(x,y) = \begin{cases} y^2 - x^2 & \text{if } |x|+|y|<1, \\ 1-2|x| & \text{if } |x|+|y| \geq 1 \end{cases}$$

satisfies the appropriate initial condition and majorization, so it suffices to check (3.8). We leave the straightforward verification to the reader and just mention that for the function A and B, one can take

$$A(x,y) = \begin{cases} -2x & \text{if } |x|+|y|<1, \\ -2\operatorname{sgn} x & \text{if } |x|+|y| \geq 1, \end{cases} \qquad B(x,y) = \begin{cases} 2y & \text{if } |x|+|y|<1, \\ 0 & \text{if } |x|+|y| \geq 1 \end{cases}$$

(as in the previous example, A and B are essentially the partial derivatives of U). This establishes the sharp inequality

$$\lambda \mathbb{P}(|g_n| \geq \lambda) \leq 2\mathbb{E}|f_n|, \qquad n = 0, 1, 2, \ldots,$$

for differentially subordinate martingales.

It is instructive to see that the Steps 1, 2 and 3 have a very transparent probabilistic interpretation. The idea can be described as follows. Let us first look at (3.11). A little informally, we want to make the probability $\mathbb{P}(|g_n| \geq 1)$ as large as possible and the expectation $\mathbb{E}|f_n|$ as little as possible. Furthermore, the process $(|f_n|)_{n \geq 0}$ is a submartingale, so its expectation does not decrease as we increase n; however, it stays on the same level if the process does not change its sign.

Step 1. The case $|y| \geq 1$. Here the reasoning is simple: the supremum defining $V(x,y)$ must be attained for the constant pair $(f,g) \equiv (x,y)$. Indeed, for such a pair the probability $\mathbb{P}(|g_n| \geq 1)$ is one (so we cannot make it larger); on the other hand, any nontrivial movement of f can only increase $\mathbb{E}|f_n|$. Thus $V(x,y) = 1 - C_1|x|$.

Step 2. The case $|x|+|y| \leq 1$. Here it is again possible to send g outside the strip $\mathbb{R} \times (-1,1)$ almost surely and simultaneously keep $\mathbb{E}|f_n|$ at the level $|x|$. Indeed, suppose that $x \geq 0$ and $y \geq 0$ (and $y < 1$), for the remaining cases the construction is similar. We consider the martingale pair (f,g) starting from (x,y) and moving along the line of slope -1 at the first step: formally, we require that $(f_1, g_1) \in \{(x+y-1, 1), (x+y+1, -1)\}$ (the corresponding probabilities are uniquely determined by the fact that $\mathbb{E}((f_1,g_1)|\mathcal{F}_0) = (x,y)$). The construction is completed by the condition $f_1 = f_2 = f_3 = \ldots$ and $g_1 = g_2 = g_3 = \ldots$ almost surely. Then $\mathbb{P}(|g_1| \geq 1) = 1$ and $\mathbb{E}|f_1| = |x|$, so $V(x,y) = 1 - C_1|x|$.

Step 3. The case $|x|+|y| < 1$. Here we need to experiment a bit. A little thought leads to the following idea: start the pair (f,g) at (x,y), then, at the first step, send it to the set $\{(x,y) : x+y \in \{-1,1\}\}$, and then move according to the pattern described in Step 2. Precisely, consider the following Markov martingale (f,g):

(i) It starts from (x,y): $(f_0,g_0) \equiv (x,y)$.
(ii) The random variable $df_1 = dg_1$ is centered and takes values in $\{(1-x-y)/2, (-1-x-y)/2\}$.
(iii) Conditionally on $\{df_1 > 0\}$ and conditionally on $\{df_1 < 0\}$, the random variable $df_2 = -dg_2$ is centered and takes values in $\{-f_1, g_1+1\}$.
(iv) Put $df_n = dg_n \equiv 0$ for $n \geq 3$.

Then we have $\mathbb{P}(|g_2| \geq 1) = 1$. Furthermore, we easily derive that df_1 takes values $(1-x-y)/2$ and $(-1-x-y)/2$ with probabilities $p_- = (1+x+y)/2$ and $p_+ = (1-x-y)/2$, respectively. In consequence, since f_2 has the same sign as f_1, we may write

$$\begin{aligned}\mathbb{E}|f_2| = \mathbb{E}|f_1| &= \left|x + \frac{1-x-y}{2}\right| \cdot \frac{1+x+y}{2} + \left|x + \frac{-1-x-y}{2}\right| \cdot \frac{1-x-y}{2}\\ &= \frac{1+|x|^2-|y|^2}{2}.\end{aligned}$$

and hence we get

$$V(x,y) \geq 1 - C_1(1+|x|^2-|y|^2)/2.$$

The remaining analysis is the same as previously. The above probabilistic analysis very clearly illustrates the notions of the continuation and stopping regions described in Remark (d) above. Indeed, from the formulas for G and V we infer that

$$\begin{aligned}C &= \{(x,y) : V(x,y) > G(x,y)\} = \mathbb{R} \times (-1,1),\\ D &= \{(x,y) : V(x,y) = G(x,y)\} = \mathbb{R} \times ((-\infty,-1] \cup [1,\infty)),\end{aligned}$$

and it is evident from Steps 1, 2 and 3 above that the optimal pairs (f,g) corresponding to $V(x,y)$ have some nontrivial evolution if and only if $(x,y) \in C$.

EXAMPLE 3.3. Fix $1 < p < 2$. The purpose of the example is to identify the best constant C_p in the L^p estimate

$$\mathbb{E}|g_n|^p \leq C_p^p \mathbb{E}|f_n|^p, \qquad n = 0,\,1,\,2,\,\ldots,$$

under the assumption of the differential subordination of g to f. As previously, our first step is to consider the case of ± 1-transforms. The inequality is of the form (3.5) with $G(x,y) = |y|^p - C_p^p|x|^p$ and the formula for the smallest Bellman function becomes

$$V(x,y) = \sup\left\{\mathbb{E}|g_n|^p - C_p^p\mathbb{E}|f_n|^p\right\}, \tag{3.12}$$

where the supremum is taken over all n and all pairs (f,g) starting from (x,y) such that g is a ± 1-transform of f. As we have explained in Remark (b), we may search for the Bellman function in the class of homogeneous functions:

$$U(\lambda x, \pm\lambda y) = \lambda^p U(x,y), \qquad x,\, y \in \mathbb{R},\, \lambda > 0.$$

Thus it is enough to find the formula for the restriction

$$u(x) = U(x, 1-x). \tag{3.13}$$

This function must be concave and majorize the function $w(x) = G(x,1-x)$ on $[0,1]$. Clearly, this is not a full set of requirements: we must somehow guarantee

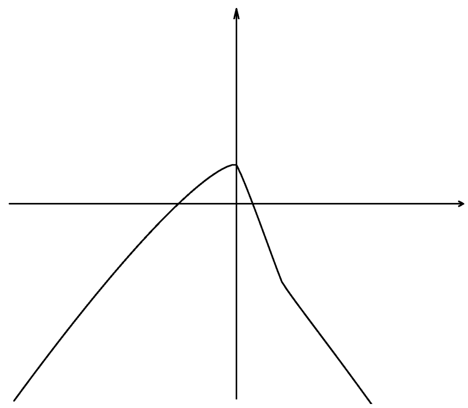

FIGURE 3.1. The graph of the function $w(x) = G(x, 1-x)$ for $p = 6/5$ and $C_p = 2$.

that the function $x \mapsto U(x, 1-x)$ is concave on the full real line. To take this into account, let us first inspect the behavior of this restriction on $(1, \infty)$. To this end, fix $x > 1$ and note that

$$U(x, 1-x) = U(x, x-1) = (2x-1)^p u\left(\frac{x}{2x-1}\right)$$

and hence, the analysis the one-sided derivatives yields

$$2pu(1) - u'(1-) = \frac{d}{dx}U(x, 1-x)\bigg|_{x=1+} \le \frac{d}{dx}U(x, 1-x)\bigg|_{x=1-} = u'(1-),$$

or

$$u'(1-) \ge pu(1). \tag{3.14}$$

A natural guess for u is to take a linear function, whose graph is tangent to that of w at some point $x_0 \in (0, 1)$. In other words, as a candidate for u, let us take

$$u(x) = w(x_0) + w'(x_0)(x - x_0),$$

for some x_0 to be found. An application of (3.14) yields $p\big[w(x_0) + w'(x_0)(1-x_0)\big] \le w'(x_0)$ or, equivalently,

$$C_p^p \ge \frac{(1-x_0)^{p-2} - (p-1)(1-x_0)^{p-1}}{(p-1)x_0^{p-1}}. \tag{3.15}$$

Now we specify x_0 by requiring that the right-hand side above is the least possible, and assume that C_p^p is equal to that minimal value. Simple calculations show that we must take $x_0 = 1 - 1/p$ and $C_p = (p-1)^{-1}$; this leads us to the candidate

$$u(x) = -\frac{p^{3-p}}{p-1}\left(x - 1 + \frac{1}{p}\right)$$

and the function

$$U(x, y) = (|x| + |y|)^p u\left(\frac{|x|}{|x| + |y|}, \frac{|y|}{|x| + |y|}\right) = p^{2-p}\left(|y| - \frac{|x|}{p-1}\right)(|x| + |y|)^{p-1}.$$

Now one has to verify rigorously that the properties 1°, 2° and 3° are satisfied. One can perform this analysis right away in the context of the differential subordination (we omit the details) and thus obtain that

$$\mathbb{E}|g_n|^p \leq (p-1)^{-p}\mathbb{E}|f_n|^p, \qquad n = 0, 1, 2, \ldots.$$

In contrast to the preceding examples, the above analysis does not imply that the constant $(p-1)^{-1}$ is the best possible. To prove this, we will exploit the abstract properties of the function V. Suppose that the L^p inequality holds with some constant C_p and introduce the function $v(x) = V(x, 1-x)$, $x \in [0,1]$. We have

$$v((C_p+1)^{-1}) \geq w((C_p+1)^{-1}) = \big(1-(C_p+1)^{-1}\big)^p - C_p^p(C_p+1)^{-p} = 0. \tag{3.16}$$

Therefore, exploiting the concavity of v, we get

$$\frac{C_p+1}{C_p}v(1) \geq \frac{v(1)-v((C_p+1)^{-1})}{1-(C_p+1)^{-1}} \geq v'(1-) \geq pv(1).$$

In the last passage we have used that fact that v, being appropriately symmetric and concave, must satisfy (3.14). The above estimate is equivalent to $C_p \geq (p-1)^{-1}$, since $v(1)$ is strictly negative. The latter follows from concavity of v, (3.16) and the estimate $v(0) \geq w(0) = 1$.

We should point out here that the function U discovered above does not coincide with the smallest Bellman function V. One can show that the two functions coincide only on a part of $\mathbb{R}^2$, more precisely, we have

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

It is instructive to see that the above result can be obtained with the use of a different argumentation which is of probabilistic nature. Let us write down the formula for the function V:

$$V(x,y) = \sup\Big\{\mathbb{E}|g_n|^p - C_p^p\mathbb{E}|f_n|^p\Big\},$$

where the supremum is taken over all n and all martingale pairs (f,g) starting from (x,y) such that g is a ± 1 transform of f. As in the example concerning the weak-type estimate, we begin by a little informal observation. Namely, when searching for the supremum, we need to make $\mathbb{E}|g_n|^p$ relatively big in comparison to $\mathbb{E}|f_n|^p$. Both processes $(|g_n|^p)_{n\geq 0}$ and $(|f_n|^p)_{n\geq 0}$ are submartingales, so their moments grow as we increase n. However, we can steer the pair (f,g) so that the p-th moment of g grows appropriately faster than that of f. To do this, note that the second derivative of the function $t \mapsto |t|^p$ is decreasing: this implies that it is "profitable" to evolve the pair (f,g) when g_n "small" and f_n "big"; on the other hand, if g is big when compared to f, the best strategy seems to be to stop the processes at once. This observation can be very nicely expressed in terms of the continuation and the stopping regions C and D: we should have $C = \{(x,y) : |y| < \gamma_p|x|\}$ and $D = \{(x,y) : |y| \geq \gamma_p|x|\}$, for some parameter γ_p to be found. In other words,

$$V(x,y) = |y|^p - C_p^p|x|^p \qquad \text{if } |y| \geq \gamma_p|x|$$

and $V(x,y) > |y|^p - C_p^p|x|^p$ for $|y| < \gamma_p|x|$. To identify V on $C = \{(x,y) : |y| < \gamma_p|x|\}$, we make the second observation. Namely, a little thought suggests that the following principle should hold: if g approaches 0, then f should be go away from this value, and vice versa. This means that when $(x,y) \in C \cap \{xy > 0\}$, then V

should be linear along the lines of slope -1, and linear along the lines of slope 1 on the remaining part of C. This actually brings us back to the analytic reasoning similar to that presented above. Indeed, one first notes that V is homogeneous of order p and thus it is enough to identify $v(x) = V(x, 1-x)$ for $x \in [0,1]$. The above analysis suggests that if $1 - x \geq \gamma_p x$ (equivalently: $x \leq (1+\gamma_p)^{-1}$, then

$$v(x) = w(x) = (1-x)^p - C_p^p x^p.$$

On the other hand, on $[(1+\gamma_p)^{-1}, 1]$ the function v should be linear and majorize w: this uniquely determines v:

$$v(x) = w'((1+\gamma_p)^{-1})(x - (1+\gamma_p)^{-1}) + w((1+\gamma_p)^{-1}),$$

and the condition $v'(1-) \geq pv(1)$ implies (3.15), with $x_0 = (1+\gamma_p)^{-1}$.

There is a natural question about the explicit examples showing that the constant $(p-1)^{-1}$ cannot be improved. The idea behind the construction is straightforward: we want to start the pair (f,g) at some point (x,x) for which $U(x,x) = 0$ and then require that the martingales evolve on the line segments along which U is linear. If, in addition, we ensure that the final pair (f_N, g_N) terminates at the set $D = \{(x,y) : |y| = (p-1)^{-1}|x|\}$, then all the inequalities

$$\mathbb{E}|g_N|^p - C_p^p \mathbb{E}|f_N|^p \leq \mathbb{E}U(f_N, g_N) \leq \mathbb{E}U(f_{N-1}, g_{N-1}) \leq \ldots \mathbb{E}U(f_0, g_0) \leq 0$$

would become equalities, so the constant C_p^p would be attained. Unfortunately, this cannot be done in such a simple manner. The only pair (f,g) which satisfies the above set of conditions is the pair $(f,g) \equiv (0,0)$, and clearly, the equality $\mathbb{E}|g_N|^p = C_p^p \mathbb{E}|f_N|^p$ is not meaningful. We will show that for any $\varepsilon > 0$, there is a martingale pair (f,g) such that if n is sufficiently large, then $\mathbb{E}|g_n|^p > \big((p-1)^{-1} - \varepsilon\big)^p \mathbb{E}|f_n|^p$. To guarantee this inequality, we may relax a little the requirements formulated above. Fix $\beta \in (1, (p-1)^{-1})$. First, we will allow the martingale (f,g) to start from the point $(1,1)$: we will see that the "loss" $\mathbb{E}U(f_0, g_0) < 0$ we experience here is insignificant in comparison to the overall size of the martingales f and g. Next, consider the following Markov transitions:

(i) The states lying in the set $\{(x,y) : |y| \geq \beta|x|\}$ are absorbing.
(ii) The state (x,y) with $0 < y < \beta x$, leads to $(x+y, 0)$ or to $\left(\frac{x+y}{\beta+1}, \frac{\beta(x+y)}{\beta+1}\right)$ (the move along the line of slope -1).
(iii) The state $(x,0)$ with $x > 0$ leads to $(x + \delta x, \delta x)$ or to $\left(\frac{x}{\beta+1}, -\frac{\beta x}{\beta+1}\right)$ (the move along the line of slope 1).
(iv) The remaining states (x,y) behave in a symmetrical way when compared to (ii) and (iii).

It is easy to check that g is a ± 1-transform of f. Furthermore, (f,g) converges to a nontrivial random variable (f_∞, g_∞) with values $\{(x,y) : |y| = \beta|x|\}$. Therefore we will be done if we check that f is L^p bounded. This can be verified readily from the above construction, we leave the details to the reader.

Example 3.4. The purpose of our final example is to illustrate the modification of the method when the values of f are restricted to the interval $[-1,1]$ and the transforming sequence takes values in $[0,1]$. Namely, in such a setting, we will identify the best constant C_λ in the estimate

$$\mathbb{P}(|g_n| \geq \lambda) \leq C_\lambda, \qquad n = 0,\, 1,\, 2,\, \ldots, \tag{3.17}$$

where $\lambda \geq 3/2$. As in the previous examples, we start with the extremal case in which the transforming sequence takes values in $\{0, 1\}$. The inequality (3.17) can be rewritten in the form

$$\mathbb{E}G(f_n, g_n) \leq C_\lambda, \qquad n = 0, 1, 2, \ldots,$$

with $G(x, y) = 1_{\{|y| \geq \lambda\}}$, $(x, y) \in [-1, 1] \times \mathbb{R}$. This inequality is of slightly different form than (3.1), due to the appearance of the term C_λ on the right. Of course, we could have put it on the left and use the function $G(x, y) = 1_{\{|y| \geq \lambda\}} - C_\lambda$, but instead we prefer to say that the method described above works perfectly fine, the only change we need is the requirement $U(x, y) \leq C_\lambda$ in the initial condition 1°.

Let us start with the smallest Bellman function

$$V(x, y) = \sup \mathbb{E}G(f_n, g_n) = \sup \Big\{ \mathbb{P}(|g_n| \geq \lambda) - C_\lambda \Big\},$$

where the supremum is taken over all associated parameters. The function G is symmetric with respect to x and y; this implies that we have $V(x, y) = V(-x, -y)$ for all $(x, y) \in [-1, 1] \times \mathbb{R}$ (on contrary, we do *not* have $V(x, -y) = V(x, y)$: this is due to the fact that the transforming sequence is non-symmetric, i.e., takes values in a non-symmetric set).

Step 1. As in the preceding examples, it is easy to identify V on some part of its domain, directly from the definition. It will be convenient to express the observations in terms of the continuation and stopping sets C and D. Directly from the above definition of V, we see that our objective is to send g outside $(-1, 1)$ with as large probability as possible, keeping f inside the interval $[-1, 1]$. This observation immediately gives $C = (-1, 1) \times (-\lambda, \lambda)$ and $D = ([-1, 1] \times \mathbb{R}) \setminus C$. Indeed, we have $G = 0$ on $(-1, 1) \times (-\lambda, \lambda)$ and for any $(x, y) \in (-1, 1) \times (-\lambda, \lambda)$ one easily constructs a martingale pair (f, g) starting from (x, y) for which the probability $\mathbb{P}(|g_n| \geq \lambda)$ is strictly positive for some n. On the other hand, for any $(x, y) \notin (-1, 1) \times (-\lambda, \lambda)$, the best choice is to take the constant martingale $(f, g) \equiv (x, y)$. Indeed, if $|x| = 1$ this is due to the fact that f must be stopped immediately (otherwise it leaves $[-1, 1]$), while for remaining points the process $|g|$ already reaches the desired set $[\lambda, \infty)$ at its initial position. This reasoning shows that $V(x, y) = 1_{\{|y| \geq \lambda\}}$ on D and it remains to identify the formula for V on C. We will actually guess this formula basing on a number of (reasonable) assumptions: this is done in the next four steps and our reasoning will be a little informal. The rigorous verification that the obtained candidate enjoys all the required properties is a separate issue (see Step 6). To stress that we are working with a candidate, we will use a different letter and, from now on, denote the investigated function by U.

Step 2. In all the considerations below, we will treat U as a C^1 function (though the candidate we will end up with will not even be continuous). Let us consider the case when $x - 1 + \lambda \leq y < \lambda$. Consider the line of slope 1 passing through (x, y). The restriction of G to this line (more formally, to an appropriate line segment) is given by

$$\xi(t) = \begin{cases} 0 & \text{if } t \in [-1 - x, \lambda - y), \\ 1 & \text{if } t \in [\lambda - y, 1 - x], \end{cases}$$

so the least concave majorant is $\zeta(t) = \min\{(t+x+1)/(x-y+\lambda+1), 1\}$. Consequently, we have

$$V(x,y) \geq \frac{x+1}{x-y+\lambda+1}.$$

The right hand side is linear along lines of slope 1 and concave along lines of slope 0 when $y < \lambda$. These desired properties suggest to assume that

$$U(x,y) = \frac{x+1}{x-y+\lambda+1} \qquad \text{if } x-1+\lambda \leq y < \lambda$$

(with a symmetric formula for $-x-1+\lambda \leq -y < \lambda$).

Step 3. Let us turn our attention to the case $y < x - 1 + \lambda$. Suppose, for a while, that y is positive and not too close to 0. A little thought and experimentation suggest the following behavior of U: there should be some curve γ splitting the set $\{(x,y) \in [-1,1] \times \mathbb{R} : y < x - 1 + \lambda\}$ such that U is linear along the lines of slope 1 on the left of γ, and linear along the horizontal lines on the right of γ: see Figure 3.2 below. Let us parametrize this curve as $\{(\gamma(y), y) : y \in I\}$ for some interval I. So, if A denotes the restriction of U to the curve $\{(\gamma(y), y) : y \in I\}$ (meaning that $A(y) = U(\gamma(y), y))$ and $x > \gamma(y)$, then

$$U(x,y) = \frac{x-\gamma(y)}{1-\gamma(y)} V(1,y) + \frac{1-x}{1-\gamma(y)} A(y) = \frac{1-x}{1-\gamma(y)} A(y). \tag{3.18}$$

Let us assume that the function U is of class C^1 on the set $y < x - 1 + \lambda$ (ac-

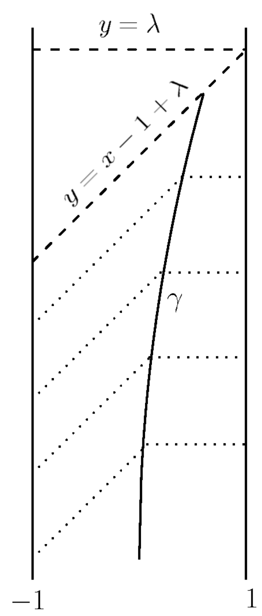

FIGURE 3.2. The curve γ. The dotted lines indicate the directions along which U should be linear.

tually, this condition will not hold, but it leads to the right candidate). Then the aforementioned linearity of U along line segments of slope -1 lying on the left of γ implies that

$$U_x(\gamma(y), y) + U_y(\gamma(y), y) = \frac{U(\gamma(y), y) - U(-1, y - \gamma(y) - 1)}{\gamma(y) + 1} = \frac{A(y)}{\gamma(y) + 1}.$$

The partial derivatives can be identified with the use of (3.18): as the result, we obtain the differential equation

$$\frac{A'(y)}{A(y)} + \frac{\gamma'(y)}{1 - \gamma(y)} = \frac{2}{1 - \gamma^2(y)}.$$

Solving this equation, we get

$$\frac{A(y)}{1 - \gamma(y)} = c \exp\left(\int_0^y \frac{2\mathrm{d}u}{1 - \gamma^2(u)}\right),$$

for some unknown constant c. To find this parameter, let $(\gamma(y_0), y_0)$ be the intersection of the curve γ with the line $y = x - 1 + \lambda$ (the "upper end" of γ: see Figure 3.2). We have assumed that U is continuous, so by the formula guessed at the previous step we may write

$$c \exp\left(\int_0^{y_0} \frac{2\mathrm{d}u}{1 - \gamma^2(u)}\right) = \frac{A(y_0)}{1 - \gamma(y_0)} = \frac{1 + \gamma(y_0)}{2(1 - \gamma(y_0))}.$$

Consequently, we have obtained that

$$\frac{A(y)}{1 - \gamma(y)} = \frac{1 + \gamma(y_0)}{2(1 - \gamma(y_0))} \exp\left(\int_{y_0}^y \frac{2\mathrm{d}u}{1 - \gamma^2(u)}\right).$$

Now suppose that y is a positive number and suppose that $x > \gamma(y)$. Exploiting the linearity of U along the horizontal segments as in the Figure 3.2, we may write

$$\begin{aligned} U(x, y) &= \frac{1 - x}{1 - \gamma(y)} A(y) + \frac{x - \gamma(y)}{1 - \gamma(y)} U(1, y) \\ &= \frac{1 + \gamma(y_0)}{2(1 - \gamma(y_0))} (1 - x) \exp\left(\int_{y_0}^y \frac{2\mathrm{d}u}{1 - \gamma^2(u)}\right). \end{aligned} \tag{3.19}$$

At this point it is not difficult to see what the optimal choice for γ should be. The function γ should be continuous, take values in $(-1, 1)$ and satisfy $\gamma(y) \geq y + 1 - \lambda$ (the latter condition means that the curve γ lies below or on the line $y = x - 1 + \lambda$). Now if we keep y_0 and $\gamma(y_0)$ fixed, it is clear that in order to maximize $U(x, y)$, we need to make γ as close to 0 as possible: this will guarantee that the integral $\int_{y_0}^y 2(1 - \gamma^2(u))^{-1}\mathrm{d}u$ will be maximal (recall that $y \leq y_0$). This leads to the assumption $\gamma \equiv 0$ and $y_0 = \lambda - 1$. Having assumed this, we can find U on a large part of the domain. Namely, if $\lambda - 1 \leq y < x + \lambda - 1$, then

$$U(x, y) = \frac{1 - x}{\lambda - y} U(y + 1 - \lambda, y) + \frac{x - y - 1 + \lambda}{\lambda - y} U(1, y) = \frac{(1 - x)(y - \lambda + 2)}{2(\lambda - y)}.$$

If $x \in [0, 1]$ and $y < \lambda - 1$ is sufficiently big (this will be made more precise later), then the above considerations (see (3.19)) give

$$U(x, y) = \frac{1 - x}{2} \exp(2(y - \lambda + 1)). \tag{3.20}$$

Finally, if $x \in [-1,0]$ and $y < x+\lambda-1$ is large enough, we have

$$\begin{aligned} U(x,y) &= -xU(-1,-x+y-1)+(1+x)U(0,y-x) \\ &= (1+x)A(-x+y) \\ &= \frac{1+x}{2}\exp(2(-x+y-\lambda+1)). \end{aligned} \tag{3.21}$$

Step 4. Now it is time to specify what we have meant by saying that the formulas (3.20) and (3.21) hold for sufficiently large y. Clearly, they cannot hold for all $y \leq \lambda-1$, since then the symmetry condition $U(x,y)=U(-x,-y)$ would be violated. A closer look at this symmetry condition suggests that (3.20) should hold true for $D_1=\{(x,y): x\in[0,1],\, 1/2\leq y\leq \lambda-1\}$ and (3.21) should be valid for $D_2=\{(x,y): x\in[-1,0], x-1/2\leq y\leq x+1-\lambda\}$: indeed, the lower boundaries $y=1/2$ in D_1 and $y=x-1/2$ are the smallest numbers with the property that the interiors of the sets D_1, D_2 and their reflections $-D_1=\{(x,y):(-x,-y)\in D_1\}$, $-D_2$ are disjoint. See Figure 3.3.

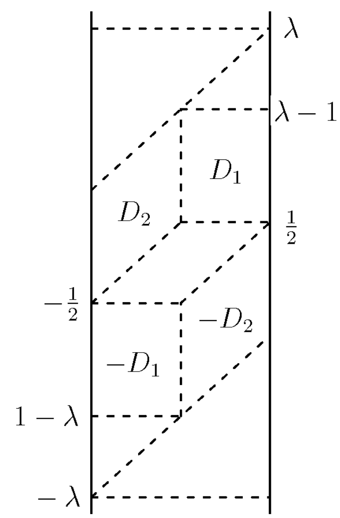

FIGURE 3.3. The sets D_1, D_2, $-D_1$ and $-D_2$ have pairwise disjoint interiors.

Step 5. It remains to find U on the set $\{(x,y): y-1/2\leq x\leq y+1/2,\, y\in[-1,1]\}$. Now we make the following guess. Namely, take the line segment of slope 1 passing through (x,y), with endpoints lying on the lines $y=\pm 1/2$. We assume that U is linear on this line segment, which leads us to

$$\begin{aligned} U(x,y) &= \left(\frac{1}{2}-y\right)U\left(x-y-\frac{1}{2},-\frac{1}{2}\right)+\left(y+\frac{1}{2}\right)U\left(x-y+\frac{1}{2},\frac{1}{2}\right) \\ &= \exp(3-2\lambda)\left(y^2-xy+\frac{1}{4}\right). \end{aligned}$$

Summarizing, we have obtained the function uniquely determined by the condition $U(x,y) = U(-x,-y)$ and the formula

$$U(x,y) = \begin{cases} 1 & \text{if } y \geq \lambda, \\ \dfrac{x+1}{x-y+\lambda+1} & \text{if } x-1+\lambda \leq y < \lambda, \\ \dfrac{(1-x)(y-\lambda+2)}{2(\lambda-y)} & \text{if } \lambda - 1 \leq y < x+\lambda-1, \\ \dfrac{1-x}{2}\exp(2(y-\lambda+1)) & \text{if } x \in [0,1],\ 1/2 \leq y < \lambda-1, \\ \dfrac{1+x}{2}\exp(2(-x+y-\lambda+1)) & \text{if } x \in [-1,0],\ -1/2 \leq y-x < \lambda-1, \\ \exp(3-2\lambda)\left(y^2 - xy + \dfrac{1}{4}\right) & \text{if } -1/2 \leq x-y \leq 1/2,\ y \in [-1,1]. \end{cases}$$

The initial condition 1° reads

$$U(x,y) \leq C_\lambda \qquad \text{if } x \in [-1,1] \text{ and } y \in \{0,x\},$$

so in particular, taking $x = y = 0$, we get $C_\lambda \geq \exp(3-2\lambda)/4$. We *assume* that we have equality here.

Step 6. The above U is just a candidate for the Bellman function. Furthermore, it was constructed under the assumption that the transforming sequence takes values in $\{0,1\}$. The remaining part of the analysis is to verify the conditions

1° We have $U(x,y) \leq \exp(3-2\lambda)/4$ for all $x \in [-1,1]$ and $y \in \{0,x\}$.

2° We have $U(x,y) \geq 1_{\{|y|\geq 1\}}$ for all $(x,y) \in [-1,1]\times\mathbb{R}$.

3° The function U is concave along any line segment of slope belonging to $[0,1]$, contained in $[-1,1]\times\mathbb{R}$.

We leave the lengthy, but rather straightforward verification to the reader. Having done this, we will have shown the estimate

$$\mathbb{P}(|g_n| \geq \lambda) \leq \exp(3-2\lambda)/4$$

under the assumption that f is bounded by 1 and g is its full transform by a predictable sequence with values in $[0,1]$.

Step 7. Finally, let us address the issue of the sharpness of the constant C_λ. It is not difficult to see the structure of the (almost) extremal examples. Informally speaking, the corresponding pair (f,g) should start from some point (x,x) for which we have $U(x,x) = C_\lambda$, and then it should move along the segments of linearity of U. This will guarantee that in the chain of inequalities

$$\mathbb{E}U(f_n,g_n) \leq \mathbb{E}U(f_{n-1},g_{n-1}) \leq \ldots \mathbb{E}U(f_0,g_0) \leq C_\lambda,$$

we have actually a chain of equalities (or almost equalities). If in addition we ensure that (f,g) terminates at the stopping set D, we will be able to write that for a large n we have $\mathbb{P}(|g_n| \geq 1) = \mathbb{E}G(f_n,g_n) = C_\lambda$ (or rather $\approx C_\lambda$) and we will be done.

Let us put the above idea into a rigorous framework: for simplicity, we will consider the strict inequality $\lambda > 3/2$ only. We assume that (f,g) starts from the point $(1/2,1/2)$ and, at the first step, jumps either to $(0,1/2)$ or to $(1,1/2)$ (with probabilities 1/2). The latter point belongs to the stopping set $\{(x,y) : U(x,y) = G(x,y)\}$, so we finish the evolution there (we can also argue that the martingale f must stop, since otherwise it would leave the interval $[-1,1]$). On contrary, if $(f_1,g_1) = (0,1/2)$, then the movement continues. Unfortunately, there is no line segment of slope 0 or 1, containing $(0,1/2)$ inside, along which U is linear. To overcome this difficulty, we fix a small positive δ and move (f, g) along the line segment with endpoints $(-1, -1/2)$ and $(\delta, 1/2 + \delta)$: we expect that the "loss" obtained due to this non-optimal move (i.e., the difference $\mathbb{E}(U(f_2, g_2)|(f_1, g_1) = (0, 1/2)) - U(0, 1/2))$ will be of order $o(\delta)$ and hence insignificant if we let $\delta \to 0$ at the very end. Formally, on the set $\{(f_1, g_1) = (0, 1/2)\}$, we assume that the random variable (f_2, g_2) takes values in the set $\{(-1, -1/2), (\delta, 1/2+\delta)\}$ (the corresponding probabilities $\delta/(1+\delta)$ and $1/(1+\delta)$ are determined by the requirement that (f, g) is a martingale). If $(f_2, g_2) = (-1, -1/2)$, the movement stops; if $(f_2, g_2) = (\delta, 1 + \delta)$, we let (f_3, g_3) jump to $(0, 1+\delta)$ or to $(1, 1+\delta)$. If the latter occurs, the evolution is over; if $(f_3, g_3) = (0, 1+\delta)$, we move (f_4, g_4) along the line segment with endpoints $(-1, -1/2+\delta)$ and $(\delta, 1+2\delta)$, and so on.

Now suppose that δ is of special form: $\delta = (\lambda - 3/2)/N$ for some large integer N. It is easy to see that after $2N+1$ steps we have two possibilities: either $f_{2N+1} = \pm 1$ (and we have stopped the evolution of the martingale pair), or $(f_{2N+1}, g_{2N+1}) = (0, \lambda - 1)$. If the latter occurs, we change the above scheme and let (f_{2N}, g_{2N}) move along the line segment of slope 1, with endpoints $(-1, \lambda - 2)$ and $(1, \lambda)$. Then, after this $2N + 2$-nd step, we ultimately stop the process.

It is clear that the martingale f just constructed takes its values in $[-1, 1]$ and g is its full transform by the sequence with values in $\{0, 1\}$. Furthermore, we have

$$\begin{aligned}\mathbb{P}(g_{2N+2} \geq \lambda) &= \mathbb{P}(f_0 = 1/2, f_1 = 0, f_2 = \delta, f_3 = 0, \dots, f_{2N+1} = 0, f_{2N+2} = 1) \\ &= \frac{1}{2} \cdot \frac{1}{1+\delta} \cdot (1-\delta) \cdot \frac{1}{1+\delta} \cdot (1-\delta) \cdot \ldots \cdot (1-\delta) \cdot \frac{1}{2} \\ &= \frac{1}{4} \left(\frac{1-\delta}{1+\delta} \right)^N .\end{aligned}$$

If we recall that $\delta = (\lambda - 3/2)/N$ and let $N \to \infty$, we see that the above quantity converges to C_λ. This proves the sharpness of the estimate and completes the analysis.

3.3. Continuous-time analogues

Now we will discuss the possibility of extending the above approach and results to continuous-time setting. First we need to introduce the necessary background. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, which is filtered by a nondecreasing right-continuous family $(\mathcal{F}_t)_{t \geq 0}$ of sub-σ-fields of $\mathcal{F}$. In addition, for technical reasons, we will assume that $\mathcal{F}_0$ contains all the events of probability 0. Suppose that $X = (X_t)_{t \geq 0}$, $Y = (Y_t)_{t \geq 0}$ are adapted real-valued martingales. We assume that these processes have right-continuous paths with left limits. The symbol $[X, X] = ([X, X]_t)_{t \geq 0}$ will stand for the quadratic variation (square bracket) of the process X.

It is straightforward to see that the continuous-time counterpart of a martingale transform is the notion of a stochastic integral:

$$Y_t = (H \star X)_t := H_0 X_0 + \int_{0+}^t H_s \mathrm{d}X_s, \qquad t \geq 0,$$

where $H = (H_t)_{t \geq 0}$ is a predictable process. If we assume in addition that H takes values in the set $\{-1, 1\}$, we get precisely the continuous-time version of full ± 1-transform. To see that the above concept does generalize martingale transforms, suppose that $f = (f_n)_{n \geq 0}$ and $g = (g_n)_{n \geq 0}$ are discrete-time martingales such that

g is the full transform of f by a certain predictable sequence v. Let us embed these sequences into continuous time via the formulas

$$X_t = f_{\lfloor t \rfloor}, \qquad Y_t = g_{\lfloor t \rfloor} \qquad \text{and} \qquad H_t = v_{\lfloor t \rfloor} \tag{3.22}$$

for $t \geq 0$. Then we have the obvious identity $Y = H \star X$, proving the desired consistency.

To introduce the continuous-time differential subordination, we rewrite its discrete version in the equivalent form:

$$\text{the sequence } \left(\sum_{k=0}^{n} |df_k|^2 - \sum_{k=0}^{n} |dg_k|^2 \right)_{n \geq 0} \text{ is nonnegative and nondecreasing.}$$

This immediately suggests the following definition.

Definition 3.2. Suppose that X, Y are two martingales. We say that Y is differentially subordinate to X, if the process $([X, X]_t - [Y, Y]_t)_{t \geq 0}$ is nonnegative and nondecreasing.

Treating discrete-time martingales as continuous-time processes as in (3.22), we check, as previously, that this definition is consistent with the previous one. Furthermore, in analogy to the discrete-time setting, if Y is a stochastic integral, with respect to X, of some predictable process H taking values in $[-1, 1]$, then Y is differentially subordinate to X. This follows at once from the identity

$$[X, X]_t - [Y, Y]_t = X_0^2(1 - H_0^2) + \int_{0+}^{t} (1 - H_s^2) \mathrm{d}[X, X]_s.$$

Finally, we will also need the following fact from stochastic analysis, the proof of which is left as an easy exercise for the reader.

Lemma 3.3. *(i) For any martingale X there exists a unique continuous martingale part X^c of X satisfying*

$$[X, X]_t = |X_0|^2 + [X^c, X^c]_t + \sum_{0 < s \leq t} |\Delta X_s|^2$$

for all $t \geq 0$ (here $\Delta X_s = X_s - X_{s-}$ is the jump of X at time s). Furthermore, $[X^c, X^c]$ is equal to $[X, X]^c$, the pathwise continuous part of $[X, X]$.

(ii) If X and Y are martingales, then Y is differentially subordinate to X if and only if Y^c is differentially subordinate to X^c, the inequality $|\Delta Y_t| \leq |\Delta X_t|$ holds for all $t > 0$ and $|Y_0| \leq |X_0|$.

Now we can pose a similar problem to that introduced at the beginning of this chapter. Namely, suppose that $G : \mathbb{R}^2 \to \mathbb{R}$ is a given function (for technical reasons, let us assume that it is Borel measurable) and assume we are interested in proving the inequality

$$\mathbb{E} G(X_t, Y_t) \leq 0, \qquad t \geq 0, \tag{3.23}$$

under one of the above dominations. Here, from the formal reason, we need to assume that we work with processes for which the above expectation makes sense. For the sake of simplicity, we will restrict ourselves to bounded martingales X, Y and locally bounded function G, but we would like to stress that in general these assumptions can be relaxed significantly (depending on the problem).

Roughly speaking, any "reasonable" ineqiuality which holds for full martingale transforms extends, with unchanged constants, to the continuous-time setting, even in the context of the differential subrodination. More precisely, we will prove the following statement.

THEOREM 3.2. *Suppose that the inequality* (3.1) *holds in the context of full martingale transforms by sequences taking values in* $[-1,1]$ *and assume further that there is a* continuous *Bellman function* U *leading to this estimate.*

(i) Then (3.23) *is valid for stochastic integrals of predictable processes with values in* $[-1,1]$.

(ii) If in addition U *is convex in its second variable, then* (3.23) *holds true under the differential subordination.*

PROOF. We will exploit the properties of the Bellman function U, combined with Itô's formula. The latter requires sufficient regularity of U, so we start with a standard mollification argument. Fix a C^∞ function $h:\mathbb{R}^2\to[0,\infty)$, supported on the unit ball of $\mathbb{R}^2$ and satisfying $\int_{\mathbb{R}^2} h=1$. Given a positive parameter δ, define the function $U^\delta:\mathbb{R}^2\to\mathbb{R}$ by the convolution

$$U^\delta(x,y)=\int_{\mathbb{R}^2} U(x+\delta u,y+\delta v)h(u,v)\mathrm{d}u\mathrm{d}v. \tag{3.24}$$

Clearly, this function is of class C^∞ and inherits the concavity along lines of slope belonging to $[-1,1]$. In particular, this implies that if $(x,y)\in\mathbb{R}^2$ and $|k|\leq|h|$, then

$$U^\delta_{xx}(x,y)h^2\pm 2U^\delta_{xy}(x,y)hk+U^\delta_{yy}(x,y)k^2\leq 0 \tag{3.25}$$

and

$$U^\delta(x+h,y+k)-U^\delta(x,y)-\langle\nabla U^\delta(x,y),(h,k)\rangle\leq 0. \tag{3.26}$$

Now, the application of Itô's formula yields

$$U^\delta(X_t,Y_t)=U^\delta(X_0,Y_0)+I_1+I_2+I_3/2, \tag{3.27}$$

where

$$\begin{aligned}
I_1&=\int_{0+}^t U^\delta_x(X_{s-},Y_{s-})\mathrm{d}X_s+\int_{0+}^t U^\delta_y(X_{s-},Y_{s-})\mathrm{d}Y_s,\\
I_2&=\sum_{0<s\leq t}\Big[U^\delta(X_s,Y_s)-U^\delta(X_{s-},Y_{s-})-\big\langle\nabla U^\delta(X_{s-},Y_{s-}),(\Delta X_s,\Delta Y_s)\big\rangle\Big],\\
I_3&=\int_{0+}^t U^\delta_{xx}(X_{s-},Y_{s-})\mathrm{d}[X,X]^c_s+2\int_{0+}^t U^\delta_{xy}(X_{s-},Y_{s-})\mathrm{d}[X,Y]^c_s\\
&\quad+\int_{0+}^t U^\delta_{yy}(X_{s-},Y_{s-})\mathrm{d}[Y,Y]^c_s.
\end{aligned}$$

Let us study the properties of the terms I_1, I_2 and I_3. The first term, treated as a function of t, defines a centered, L^2-bounded martingale: this is due to the fact that the processes X and Y are bounded and U^δ_x, U^δ_y are locally bounded. Consequently, $\mathbb{E}I_1=0$. The second term is nonpositive by (3.26), since $|\Delta Y_s|\leq|\Delta X_s|$. To deal with I_3, we need to apply different arguments depending on whether we prove (i) or (ii). First, if Y is the stochastic integral, with respect to X, of some predictable process H taking values in $[-1,1]$, then

$$I_3=\int_{0+}^t U^\delta_{xx}(X_{s-},Y_{s-})+2U^\delta_{xy}(X_{s-},Y_{s-})H_s+U^\delta_{yy}(X_{s-},Y_{s-})H_s^2\mathrm{d}[X,X]_s\leq 0,$$

by (3.25). Suppose, on the other hand, that Y is differentially subordinate to X and U is convex in y. The inequality (3.25) implies

$$U_{xx}^\delta + 2|U_{xy}^\delta| + U_{yy}^\delta \leq 0 \qquad \text{on } \mathbb{R}^2,$$

and hence, for any $(x,y) \in \mathbb{R}^2$ and any $h,\, k \in \mathbb{R}^2$ (we do not require $|k| \leq |h|$ here),

$$U_{xx}^\delta(x,y)(h^2+k^2) \pm 4U_{xy}^\delta(x,y)hk + U_{yy}^\delta(x,y)(h^2+k^2) \leq 0.$$

This is equivalent to saying that

$$U_{xx}^\delta(x,y)h^2 \pm 2U_{xy}^\delta(x,y)hk + U_{yy}^\delta(x,y)k^2 \leq \frac{U_{xx}^\delta(x,y) - U_{yy}^\delta(x,y)}{2}(h^2-k^2).$$

Approximating I_3 with discrete sums, we easily check that the above bound implies

$$I_3 \leq \int_{0+}^t \frac{U_{xx}^\delta(X_{s-},Y_{s-}) - U_{yy}^\delta(X_{s-},Y_{s-})}{2} \mathrm{d}([X,X]_s^c - [Y,Y]_s^c) \leq 0,$$

since $U_{xx}^\delta \leq 0$ (see (3.25)), $U_{yy}^\delta \geq 0$ (here we use the assumption of convexity of U in y) and $[X,X]^c - [Y,Y]^c$ is nondecreasing (since Y^c is differentially subordinate to X^c). Taking the expectation of both sides of (3.27) and putting all the above facts together, we obtain $\mathbb{E}U^\delta(X_t,Y_t) \leq \mathbb{E}U^\delta(X_0,Y_0)$. It remains to perform a limiting procedure. Since U is continuous, we have $U^\delta \to U$ pointwise. But X, Y are assumed to be bounded, so Lebesgue's dominated convergence theorem implies

$$\mathbb{E}U(X_t,Y_t) \leq \mathbb{E}U(X_0,Y_0).$$

Applying the properties 1° and 2°, we get the desired bound (3.23). The proof is complete. □

Though we have worked with Bellman functions defined on $\mathbb{R}^2$, a similar proof can be carried out if the domain of U is a strict subset of the plane. One can also try to relax the assumption on the convexity of U in y, appearing in the second part of the theorem above: for example, it is clear that it suffices to guarantee that the expression $U_{xx}^\delta - U_{yy}^\delta$ is nonpositive, or apply appropriate localizing techniques. We will not go into details here and refer the reader to the literature (see the bibliographical notes at the end of this chapter).

We can summarize the above discussion on the inequalities for continuous-time martingales as follows. If one wants to show (3.23) (say, under the assumption of the differential subordination), one should first take a look at the counterpart of this inequality in the discrete-time setting, in the context of ± 1-transforms. Having successfully identified the corresponding Bellman function, one often realizes that it enjoys all the structural properties required to yield the validity of the more general continuous-time version.

3.4. Orthogonal martingales and conjugate harmonic functions

In this section we will study pairs of processes enjoying the differential subordination and the following additional structural property.

DEFINITION 3.3. Two martingales $X = (X_t)_{t\geq 0}$, $Y = (Y_t)_{t\geq 0}$ are said to be orthogonal, if their square bracket $[X,Y]$ is constant as a function of time.

It follows from Lemma 3.3 and a standard polarization argument that for any martingales X, Y we have

$$[X,Y]_t = X_0Y_0 + [X^c,Y^c]_t + \sum_{0<s\leq t} \Delta X_s \cdot \Delta Y_s.$$

Therefore, in particular, the orthogonality condition implies $\Delta X_s \cdot \Delta Y_s = 0$ for all $t > 0$. This has a very important further consequence: if Y is differentially subordinate to X and the processes are orthogonal, then necessarily Y has continuous trajectories.

We will present two examples, which will be important to our subsequent considerations below.

EXAMPLE 3.5. Suppose that (X,Y) is a two-dimensional Brownian motion starting from some given point (x,y) with $|y| \leq |x|$. Then Y is differentially subordinate to X and both processes are orthogonal.

The second example is very important from the viewpoint of applications to analysis. It links the above probabilistic concepts with that of conjugate harmonic function (Hilbert transform). Let us discuss this issue for a while.

EXAMPLE 3.6. Suppose that D is the unit disc in $\mathbb{R}^2$ and let $u, v : D \to \mathbb{R}$ be two harmonic functions, normalized so that $|v(0)| \leq |u(0)|$. We say that v is conjugate to u, if the Cauchy-Riemann equations

$$u_x = v_y, \quad u_y = -v_x \tag{3.28}$$

hold on D. Equivalently, this amounts to saying that the function $u + iv$ is an analytic function on D. Let $D_0 \subset D$ be a disc of center 0 and radius $r < 1$ and let $W = (W^1, W^2)$ be the standard Brownian motion in $\mathbb{R}^2$, started at the origin, and stopped upon reaching the boundary of D_0. Define the processes X, Y by

$$X_t = u(W_{\tau\wedge t}), \quad Y_t = v(W_{\tau\wedge t}), \qquad t \geq 0.$$

Clearly, X and Y are bounded processes. Actually, it follows directly from Itô's formula that X and Y are martingales:

$$X_t = u(0) + \int_{0+}^{\tau\wedge t} \nabla u(W_s)\cdot \mathrm{d}W_s, \qquad Y_t = v(0) + \int_{0+}^{\tau\wedge t} \nabla v(W_s)\cdot \mathrm{d}W_s.$$

Since

$$[X,X]_t = u(0)^2 + \int_{0+}^{\tau\wedge t} |\nabla u(W_s)|^2 \mathrm{d}s, \qquad [Y,Y]_t = v(0)^2 + \int_{0+}^{\tau\wedge t} |\nabla v(W_s)|^2 \mathrm{d}s$$

and $|v(0)| \leq |u(0)|$, we see that Y is differentially subordinate to X (indeed: by (3.28), we have $|\nabla u| = |\nabla v|$). Furthermore, we have

$$[X,Y]_t = u(0)v(0) + \int_{0+}^{\tau\wedge t} \langle \nabla u(W_s), \nabla v(W_s)\rangle \mathrm{d}s = u(0)v(0),$$

so X and Y are orthogonal. Therefore, any estimate of the form (3.23) has an appropriate counterpart in the analytic setting: indeed, if we manage to show that

$$\mathbb{E}G(u(W_{\tau\wedge t}), v(W_{\tau\wedge t})) = \mathbb{E}G(X_t, Y_t) \leq 0,$$

then letting $t \to \infty$ we obtain, by Lebesgue's dominated convergence theorem (recall that u, v, G are bounded on D_0)

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} G(u(re^{i\theta}), v(re^{i\theta}))\mathrm{d}\theta = \mathbb{E}G(u(W_\tau), v(W_\tau)) \leq 0.$$

So, for instance, if $G(x,y) = |y|^p - C_p^p|x|^p$ is the function corresponding to sharp moment inequality for orthogonal, differentially subordinate martingales, we obtain

$$\begin{aligned} ||v||_{L^p(\mathbb{T})} &= \sup_{0<r<1}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}|v(re^{i\theta})|^p\mathrm{d}\theta\right]^{1/p} \\ &\leq C_p \sup_{0<r<1}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}|u(re^{i\theta})|^p\mathrm{d}\theta\right]^{1/p} = C_p||u||_{L^p(\mathbb{T})}. \end{aligned} \tag{3.29}$$

This inequality can be expressed in terms of the so-called periodic Hilbert transform $\mathcal{H}^{\mathbb{T}}$. This operator is a singular integral operator, acting on integrable functions $f : \mathbb{T} \to \mathbb{C}$ by the formula

$$\mathcal{H}^{\mathbb{T}} f(e^{i\theta}) = \frac{1}{2\pi}\text{p.v.}\int_{-\pi}^{\pi} f(e^{it})\cot\frac{\theta - t}{2}\mathrm{d}t.$$

More explicitly, if $f(\zeta) = \sum_{n\in\mathbb{Z}} a_n\zeta^n$, $\zeta \in \mathbb{C}$, then we have

$$\mathcal{H}^{\mathbb{T}} f(\zeta) = -i\sum_{n\in\mathbb{Z}}\text{sgn}(n)a_n\zeta^n, \qquad \zeta \in \mathbb{T}.$$

In other words, $\mathcal{H}^{\mathbb{T}}$ is the Fourier multiplier with the symbol $m(n) = -i\,\text{sgn}(n)$.

The connection between this setup and the context of the conjugate harmonic functions is the following. Suppose that f is an integrable function on the unit circle $\mathbb{T}$. Then we can extend f to a harmonic function u on the whole disc $\overline{D}$ with the use of the Poisson formula:

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta - t)f(e^{it})\,\mathrm{d}t, \qquad 0 \leq r < 1,\ \theta \in (-\pi, \pi],$$

where P_r denotes the Poisson kernel

$$P_r(\theta) = \Re\left(\frac{1 + re^{i\theta}}{1 - re^{i\theta}}\right) = \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

It turns out that u has a radial limit $\lim_{r\to 1} u(re^{i\theta})$ which is equal to $f(e^{i\theta})$ almost everywhere. Now, let v be a harmonic conjugate function to u, satisfying $v(0) = 0$. Then one can show that the radial limit $\lim_{r\to 1-} v(re^{i\theta})$ exists for almost all $\theta \in (-\pi, \pi]$, the limit is precisely $\mathcal{H}^{\mathbb{T}} f(e^{i\theta})$. Since $0 = |v(0)| \leq |u(0)|$, we can use (3.29), which combined with Fatou's lemma implies

$$||\mathcal{H}^{\mathbb{T}} f||_{L^p(\mathbb{T})} \leq C_p||f||_{L^p(\mathbb{T})}.$$

The above analysis can be carried out in any locally connected domain $D \subset \mathbb{R}^2$. Distinguish a point $\xi \in D$ and let u, $v : D \to \mathbb{R}$ be two harmonic functions satisfying (3.28) and $|v(\xi)| \leq |u(\xi)|$. Then if D_0 is any bounded subdomain of D satisfying $\overline{D_0} \subset D$ and $\mu_{D_0,\xi}$ is the harmonic measure on ∂D_0 corresponding to ξ, then the inequality (3.23) implies

$$\int_{\partial D_0} G(u, v)\mathrm{d}\mu_{D_0,\xi} \leq 0. \tag{3.30}$$

EXAMPLE 3.7. Let us continue the above considerations in the context of L^p estimates for non-periodic Hilbert transform $\mathcal{H}^{\mathbb{R}}$. This operator is a singular integral operator, acting on integrable functions $f : \mathbb{R} \to \mathbb{R}$ by the formula

$$\mathcal{H}^{\mathbb{R}} f(x) = \frac{1}{\pi}\text{p.v.} \int_{\mathbb{R}} \frac{f(y)}{x-y} \mathrm{d}y.$$

One can show that $\mathcal{H}^{\mathbb{R}}$ is the Fourier multiplier with the symbol $m(\xi) = -i\,\mathrm{sgn}(\xi)$, $\xi \in \mathbb{R}$, that is,

$$\widehat{\mathcal{H}^{\mathbb{R}} f}(\xi) = -i\,\mathrm{sgn}(\xi)\widehat{f}(\xi), \qquad \xi \in \mathbb{R}.$$

To link this operator with the context of orthogonal harmonic functions, we set $D = \mathbb{R} \times (0, \infty)$, the upper halfplane. Suppose that f is a function belonging to the Schwartz class $\mathcal{S}(\mathbb{R})$. We can extend it to a harmonic function u on $\overline{D}$ by the Poisson formula

$$u(x, y) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{y f(t)}{(x-t)^2 + y^2} \mathrm{d}t, \qquad (x, y) \in D.$$

This function is indeed an extension: one easily shows that $\lim_{y\downarrow 0} u(x, y) = f(x)$ for all $x \in \mathbb{R}$. Note that we can rewrite the formula for u in the form

$$u(x, y) = \Re\left(\frac{i}{\pi}\int_{\mathbb{R}} \frac{f(t)}{z-t}\mathrm{d}z\right),$$

where $z = x + iy$. Therefore, the function

$$v(x, y) = \Im\left(\frac{i}{\pi}\int_{\mathbb{R}} \frac{f(t)}{z-t}\mathrm{d}z\right) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{f(t)(x-t)}{(x-t)^2 + y^2}\mathrm{d}t$$

is conjugate to u, since the function $z \mapsto \frac{i}{\pi}\int_{\mathbb{R}} \frac{f(t)}{z-t}\mathrm{d}t$ is analytic on D. It can be shown that if $y \downarrow 0$, then $v(x, y)$ converges to $\mathcal{H}^{\mathbb{R}} f(x)$ for almost all x. To make use of (3.29), we need an appropriate behavior of u and v at some base point ξ. To this end, fix $y_0 > 0$, take $\xi = iy_0$ and consider the pair u, $v_{y_0} = v - v(\xi)$ of harmonic functions. Of course, their conjugacy is preserved and we have $|u(\xi)| \geq 0 = |v_{y_0}(\xi)|$. For a given $\varepsilon > 0$, consider the halfdisc

$$D_0 = \{(x, y) : y \geq \varepsilon,\, x^2 + (y-\varepsilon)^2 \leq \varepsilon^{-1}\}.$$

If ε is sufficiently small, then ξ belongs to D_0 and hence (3.30) yields

$$\int_{D_0} |v|^p \mathrm{d}\mu_{D_0,\xi} \leq C_p^p \int_{D_0} |u|^p \mathrm{d}\mu_{D_0,\xi}.$$

If we let $\varepsilon \to 0$, then $\mu_{D_0,\xi}$ converges weakly to

$$d\mu_\xi(x) = \frac{1}{\pi}\frac{y_0}{x^2 + y_0^2},$$

the harmonic measure on $\mathbb{R}$ with respect to the halfline $\mathbb{R} \times (0, \infty)$ and the point ξ. But u is bounded (which follows directly from the assumption $f \in \mathcal{S}(\mathbb{R})$), so Lebesgue's dominated convergence theorem and Fatou's lemma give

$$\frac{1}{\pi}\int_{\mathbb{R}} |\mathcal{H}^{\mathbb{R}} f(x) - v(iy_0)|^p \frac{y_0}{x^2 + y_0^2}\mathrm{d}x \leq \frac{C_p^p}{\pi}\int_{\mathbb{R}} |f(x)|^p \frac{y_0}{x^2 + y_0^2}\mathrm{d}x.$$

Multiply both sides by πy_0 and let $y_0 \to \infty$. Then $v(iy_0) \to 0$ and $y_0^2/(x^2 + y_0^2)$ increases to 1 and the above bound yields $||\mathcal{H}^{\mathbb{R}} f||_{L^p(\mathbb{R})} \leq C_p ||f||_{L^p(\mathbb{R})}$.

Let us now illustrate on some examples how the Bellman function method works in the above context. That is, suppose that G is a Borel, locally bounded function and assume that we are interested in showing the estimate

$$\mathbb{E}G(X_t, Y_t) \leq 0, \qquad t \geq 0, \tag{3.31}$$

for all bounded, orthogonal martingales X, Y such that Y is differentially subordinate to X. A closer look at the reasoning presented in the previous section reveals that it is enough to find a continuous function $U : \mathbb{R}^2 \to \mathbb{R}$ satisfying the following conditions:

1° $U(x, y) \leq 0$ if $|y| \leq |x|$.
2° $U \geq G$ on $\mathbb{R}^2$.
3° U is superharmonic on $\mathbb{R}^2$ and concave in its first variable.

THEOREM 3.3. *If U satisfies 1°, 2° and 3°, then* (3.31) *holds true.*

PROOF. We argue as in the preceding section; just a few modifications are necessary. Pick a mollifying function $h : \mathbb{R}^2 \to [0, \infty)$ and define U^δ by the formula (3.24). This new function is of class C^∞, so Itô's formula implies

$$U^\delta(X_t, Y_t) = U^\delta(X_0, Y_0) + I_1 + I_2 + I_3/2, \tag{3.32}$$

where (recall that Y has continuous paths)

$$\begin{aligned} I_1 &= \int_{0+}^t U_x^\delta(X_{s-}, Y_s)\mathrm{d}X_s + \int_{0+}^t U_y^\delta(X_{s-}, Y_s)\mathrm{d}Y_s, \\ I_2 &= \sum_{0<s\leq t} \Big[U^\delta(X_s, Y_s) - U^\delta(X_{s-}, Y_s) - U_x^\delta(X_{s-}, Y_s)\Delta X_s\Big], \\ I_3 &= \int_{0+}^t U_{xx}^\delta(X_{s-}, Y_s)\mathrm{d}[X, X]_s^c + 2\int_{0+}^t U_{xy}^\delta(X_{s-}, Y_s)\mathrm{d}[X, Y]_s^c \\ &\quad + \int_{0+}^t U_{yy}^\delta(X_{s-}, Y_s)\mathrm{d}[Y, Y]_s \\ &= \int_{0+}^t U_{xx}^\delta(X_{s-}, Y_s)\mathrm{d}[X, X]_s^c + \int_{0+}^t U_{yy}^\delta(X_{s-}, Y_s)\mathrm{d}[Y, Y]_s, \end{aligned}$$

the last equation following from the orthogonality of X and Y. Now, arguing as previously, we have $\mathbb{E}I_1 = 0$. Furthermore, the function U^δ inherits the property 3° and hence each summand in I_2 is nonpositive. Finally, since Y is differentially subordinate to X^c (see Lemma 3.3), we see that

$$I_3 \leq \int_{0+}^t U_{xx}^\delta(X_{s-}, Y_s)\mathrm{d}[Y, Y]_s + \int_{0+}^t U_{yy}^\delta(X_{s-}, Y_s)\mathrm{d}[Y, Y]_s \leq 0,$$

where in the last passage we have exploited the superharmonicity of U^δ. Putting all the above facts together, we see that

$$\mathbb{E}U^\delta(X_t, Y_t) \leq \mathbb{E}U^\delta(X_0, Y_0).$$

Letting $\delta \to 0$ we get $\mathbb{E}U(X_t, Y_t) \leq \mathbb{E}U(X_0, Y_0)$ (here we use the continuity of U and the boundedness properties of X and Y). It remains to apply 1° and 2° to obtain the desired claim. □

REMARK 3.1. Two important observations are in order.

(a) In some problems it is enough to consider only those pairs (X, Y) of orthogonal differentially subordinate martingales, for which the dominated process Y starts from 0 (see the above discussion concerning conjugate harmonic functions on the unit disc). Then the inequality of the condition 1° needs to be checked for $x \in \mathbb{R}$ and $y = 0$.

(b) Sometimes we are interested in the inequalities for orthogonal processes X, Y satisfying $|Y_0| \leq |X_0|$ and $[Y, Y] = [X, X]$. Then in 3° we only need to verify the superharmonicity property, the concavity along the first variable is not necessary (indeed: this is sufficient for the nonpositivity of the term I_3 in the above proof).

EXAMPLE 3.8. Let $1 < p < 2$. We will identify the best constant C_p in the inequality

$$\mathbb{E}|Y_t|^p \leq C_p^p \mathbb{E}|X_t|^p, \qquad t \geq 0, \tag{3.33}$$

where X, Y are bounded orthogonal martingales such that Y is differentially subordinate to X. We rewrite the above estimate in the form (3.31), with $G(x, y) = |y|^p - C_p^p|x|^p$. As we have discussed above, we need to find a function $U_p : \mathbb{R}^2 \to \mathbb{R}$ satisfying the conditions 1°, 2° and 3°. Clearly, we may assume that U_p is symmetric with respect to x and y, since G has this property. It is instructive to come back to L^p estimate for differentially subordinated martingales (without the orthogonality assumption). Recall the least special function corresponding to this inequality: when $1 < p \leq 2$, then

$$\overline{U}(x, y) = \begin{cases} \alpha_p(|y| - (p-1)^{-1}|x|)(|x| + |y|)^{p-1} & \text{if } |y| < (p-1)^{-1}|x|, \\ |y|^p - (p-1)^{-p}|x|^p & \text{if } |y| \geq (p-1)^{-1}|x|, \end{cases}$$

for an appropriate $\alpha_p > 0$. It is natural to conjecture that the special function U_p in the orthogonal case should also be of this form. So, we assume that there is a threshold $\gamma_p > 0$ such that

$$U_p(x, y) > |y|^p - C_p^p|x|^p \quad \text{ if } |y| < \gamma_p|x|$$

and

$$U_p(x, y) = |y|^p - C_p^p|x|^p \quad \text{ if } |y| \geq \gamma_p|x|.$$

In the language of optimal stopping, we expect the continuation region to be of the form $C = \{(x, y) : |y| < \gamma_p|x|\}$. However, as it is always the case, on the continuation set the "concavity" part of condition 3° should be turned into "linearity". More precisely, this means that U_p should be harmonic on C. These observations bring us much closer to the discovery of U_p. On C, we write U_p in polar coordinates

$$U_p(x, y) = R^p g(\theta), \qquad \theta \in [-\theta_p, \theta_p],$$

where $\theta_p = \arctan \gamma_p$. By the symmetry of U with respect to the variable y, we see that g is an even function. Furthermore, the condition $\Delta U_p = 0$ on C is equivalent to saying that $g''(\theta) + p^2 g(\theta) = 0$ for $\theta \in [-\theta_p, \theta_p]$. We easily solve the differential equation, obtaining

$$g(\theta) = \alpha_p \cos(p\theta) + \beta_p \sin(p\theta),$$

for some unknown parameters α_p, β_p. Since g is an even function, we conclude that $\beta = 0$. Furthermore, the continuity of U_p implies

$$\alpha_p R^p \cos(p\theta_p) = R^p \left[\sin^p \theta_p - C_p^p \cos^p \theta_p \right],$$

or $\alpha_p \cos(p\theta_p) = \sin^p \theta_p - C_p^p \cos^p \theta_p$. Finally, the superharmonicity of U_p together with the majorization 2° enforce

$$U_{px}(1-, \gamma_p) = U_{px}(1+, \gamma_p),$$

or, equivalently,

$$-C_p^p \cos^{p-1} \theta_p = \alpha_p \cos((p-1)\theta_p).$$

Combining this with the preceding equation, we obtain

$$C_p^p = \tan^{p-1} \theta_p \cot((p-1)\theta_p).$$

It is natural to conjecture that the desired value of the parameter γ_p is such that the right-hand side is minimized. Denoting the expression on the right by $f(\theta_p)$, we compute that

$$f'(\theta) = \frac{(p-1)\tan^{p-2}\theta_p}{2\cos^2\theta_p \sin^2((p-1)\theta_p)} \Big(\sin(2(p-1)\theta_p) - \sin(2\theta_p) \Big).$$

Therefore f, considered as a function on $(0, \pi/2)$, attains its minimum for $\theta_p = \pi/(2p)$. Plugging this above, we get $C_p = \tan \frac{\pi}{2p}$, $\alpha_p = -\sin^{p-1} \frac{\pi}{2p} / \cos \frac{\pi}{2p}$ and the explicit formula for a candidate for the Bellman function. Now one can easily check that this function does satisfy the properties 1°, 2° and 3°.

The following observation will be useful later. Namely, using a simple localization argument, we can show that (3.33) holds true for any pair (X, Y) such that Y is differentially subordinate to X (i.e., we do not need to assume the boundedness of the processes). To see this, fix $t \geq 0$ and assume that $\mathbb{E}|X_t|^p < \infty$ (otherwise there is nothing to prove). Fix a huge positive number M and consider the stopping time $\tau = \inf\{s \geq 0 : |X_s| + |Y_s| \geq M\}$. Introduce the truncated martingales

$$X_s^M = X_{\tau\wedge s} 1_{\{\tau>0\}}, \qquad Y_s^M = Y_{\tau\wedge s} 1_{\{\tau>0\}}, \qquad s \geq 0.$$

Clearly, X and Y are bounded. Furthermore, Y^M is differentially subordinate to X^M, since

$$[X^M, X^M]_s = [X, X]_{\tau\wedge s} 1_{\{\tau>0\}} \quad \text{and} \quad [Y^M, Y^M]_s = [Y, Y]_{\tau\wedge s} 1_{\{\tau>0\}}.$$

Consequently, by (3.33) for bounded processes, which we have already established, we may write

$$\mathbb{E}|Y_t^M|^p \leq \tan^p \left(\frac{\pi}{2p}\right) \mathbb{E}|X_t^M|^p \leq \tan^p \left(\frac{\pi}{2p}\right) \mathbb{E}|X_t|^p.$$

Here in the last line we have exploited Doob's optional sampling theorem applied to the submartingale $(|X_s|^p)_{0\leq s\leq t}$. Letting $M \to \infty$, we see that $|Y_t^M|^p \to |Y_t|^p$ almost surely. Therefore it suffices to apply Fatou's lemma to get the desired estimate (3.33).

It remains to check that the constant $\tan \frac{\pi}{2p}$ cannot be decreased. To this end, let κ be a fixed positive number smaller than $\tan(\pi/(2p))$. Consider a two-dimensional Brownian motion $W = (W_t^1, W_t^2)_{t\geq 0}$ starting from $(1, 0)$ and let τ be the first exit time of W from the set $E = \{(x, y) : |y| \leq \kappa x\}$. By well-known properties of Brownian motion, we have $\tau < \infty$ almost surely. Since U_p is harmonic on E, we get

$$U_p(1, 0) = \mathbb{E}U_p(W_{\tau\wedge t}^1, W_{\tau\wedge t}^2) = \mathbb{E}\big[(W_{\tau\wedge t}^1)^p U_p(1, W_{\tau\wedge t}^2/W_{\tau\wedge t}^1)\big] \leq U_p(1, \kappa)\mathbb{E}(W_{\tau\wedge t}^1)^p.$$

Here we have used the fact that $U_p(1,\cdot)$ is increasing on $[0,\theta_p)$: a little calculation shows that

$$U_{py}(1,y) = -p\alpha_p R^{p-1}\sin((p-1)\arctan y) > 0.$$

Furthermore, $U_p(1,\kappa) < U_p(1,\gamma_p) = 0$, so the preceding estimate reads

$$\mathbb{E}(W^1_{\tau\wedge t})^p \le U_p(1,0)/U_p(1,\kappa). \tag{3.34}$$

Consider the (unbounded) martingales

$$X_t = W^1_{\tau\wedge t}, \qquad Y_t = W^2_{\tau\wedge t}, \qquad t \ge 0.$$

Then Y is differentially subordinate to X and both processes are orthogonal. Furthermore, we infer from (3.34) that X is bounded in L^p and hence, by (3.33), the same is true for Y. Therefore,

$$\lim_{t\to\infty} \frac{\mathbb{E}|Y_t|^p}{\mathbb{E}|X_t|^p} = \frac{\mathbb{E}|W^2_\tau|^p}{\mathbb{E}|W^1_\tau|^p} = \kappa^p,$$

so the best constant in (3.33) cannot be smaller than κ. It remains to observe that κ could be taken arbitrarily close to $\tan(\pi/(2p))$.

EXAMPLE 3.9. Our second example concerns the sharp weak-type (1,1) estimate for orthogonal martingales. We will find the best constant K in the estimate

$$\mathbb{P}(|Y_t| \ge 1) \le K\mathbb{E}|X_t|, \qquad t \ge 0,$$

where X, Y run over the class of all orthogonal martingales such that Y is differentially subordinate to X and $Y_0 = 0$.

Clearly, this inequality if of the form (3.31), with $G(x,y) = 1_{\{|y|\ge 1\}} - K|x|$. Let us first assume that both X and Y are bounded. In the light of Theorem 3.3, we need to construct a function $U : \mathbb{R}^2 \to \mathbb{R}$ satisfying the conditions 1°, 2° and 3°. As in the case of L^p estimates, it is useful to take a look at the weak-type (1,1) inequality for differentially subordinate martingales without the orthogonality: the corresponding Bellman function, corresponding to $\overline{G}(x,y) = 1_{\{|y|\ge 1\}} - 2|x|$, was

$$\overline{U}(x,y) = \begin{cases} y^2 - x^2 & \text{if } |x|+|y| \le 1, \\ 1 - 2|x| & \text{if } |x|+|y| > 1. \end{cases}$$

An important feature of this object is that the continuation set C was the strip $(\mathbb{R}\times(-1,1)) \setminus \{(0,0)\}$. A natural idea is to *assume* that in the orthogonal setting the continuation set C is the same. This immediately implies that we have

$$U(x,y) = 1 - K|x| \qquad \text{if } |y| \ge 1$$

and

$$U(x,y) > -K|x| \qquad \text{if } |y| < 1, \quad (x,y) \ne (0,0).$$

On the continuation set the function U must be harmonic: this gives that U is the harmonic lift of G on the strip. To identify the explicit formula for this function, we "translate" the problem into the more friendly context with the use of conformal mappings. Specifically, consider the conformal map

$$\phi(z) = i\exp(\pi z/2) = \left(e^{\pi x/2}\sin(\pi y/2), e^{\pi x/2}\cos(\pi y/2)\right),$$

which maps the strip $S=\mathbb{R}\times(-1,1)$ onto the upper halfplane $H=\mathbb{R}\times(0,\infty)$. This conformal map sends the boundary function $G|_{\partial S}$ onto the function

$$f(t)=1-K\cdot\left|\frac{2}{\pi}\log|t|\right|$$

on $\mathbb{R}=\partial H$. Consequently, if $(x,y)\in S$, then $U(x,y)=\mathcal{U}(\phi(x,y))$, where $\mathcal{U}$ is the harmonic extension of f obtain via the Poisson formula

$$\mathcal{U}(\alpha,\beta)=\frac{1}{\pi}\int_{\mathbb{R}}\frac{\beta f(t)}{(\alpha-t)^2+\beta^2}\mathrm{d}t=1-\frac{K}{\pi}\int_{\mathbb{R}}\frac{\beta\left|\frac{2}{\pi}\log|t|\right|}{(\alpha-t)^2+\beta^2}\mathrm{d}t.$$

Plugging the above formula for ϕ and substituting $t=se^{\pi x/2}$, we see that U admits the following formula

$$U(x,y)=\begin{cases}1-\dfrac{K}{\pi}\displaystyle\int_{\mathbb{R}}\frac{\cos(\frac{\pi y}{2})\left|\frac{2}{\pi}\log|s|+x\right|}{\left(\sin(\frac{\pi y}{2})+s\right)^2+\cos^2(\frac{\pi y}{2})}\mathrm{d}s & \text{if } |y|<1,\\ 1-K|x| & \text{if } |y|\geq 1.\end{cases}\tag{3.35}$$

To find K, let us exploit the initial condition 1°. We have

$$\begin{aligned}U(0,0)&=1-K\cdot\frac{4}{\pi^2}\int_0^\infty\frac{|\log s|}{s^2+1}\mathrm{d}s\\&=1-K\cdot\frac{4}{\pi^2}\int_{-\infty}^\infty\frac{|t|e^t}{e^{2t}+1}\mathrm{d}t\\&=1-K\cdot\frac{8}{\pi^2}\int_0^\infty te^{-t}\sum_{k=0}^\infty(-e^{-2t})^k\mathrm{d}t\\&=1-K\cdot\frac{8}{\pi^2}\sum_{k=0}^\infty\frac{(-1)^k}{(2k+1)^2},\end{aligned}$$

which leads us to the choice

$$K=\left(\frac{8}{\pi^2}\sum_{k=0}^\infty\frac{(-1)^k}{(2k+1)^2}\right)^{-1}=1.34\ldots.$$

Let us check that the function U defined in (3.35) indeed enjoys the properties 1°, 2° and 3° (since we assume that $Y_0=0$, the initial condition amounts to saying that $U(x,0)\leq 0$: see Remark 3.1 above). Let us first observe that for any fixed y, the function $x\mapsto U(x,y)$ is concave. This is evident for $|y|\geq 1$, while for remaining y it follows directly from the fact that $x\mapsto\left|x+\frac{2}{\pi}\log|s|\right|$ is convex. By harmonicity of U on the strip, we get $U_{yy}\geq 0$ there and hence, to check 1° and 2°, we need to verify that $U(x,0)\geq -K|x|$. Differentiating, we get

$$|U_x(x,0)|=\left|-\frac{K}{\pi}\int_{\mathbb{R}}\frac{\operatorname{sgn}\left(\frac{2}{\pi}\log|s|+x\right)}{s^2+1}\mathrm{d}s\right|\leq\frac{K}{\pi}\int_{\mathbb{R}}\frac{\mathrm{d}s}{s^2+1}=K,$$

so the aforementioned concavity of $x\mapsto U(x,0)$ and the equality $U(0,0)=0$ completes the proof of 1° and 2°. Finally, to show 3°, we only need to check superharmonicity of U. Clearly, if $x\in\mathbb{R}$ and $|y|\neq 1$, then U is obviously superharmonic in some neighborhood of (x,y). To check the remaining points, note that

$$U(x,y)\leq 1-K|x|\qquad\text{on } \mathbb{R}\times[-1,1],$$

because the function $(x, y) \mapsto 1 - K|x|$ is superharmonic and coincides with U on the boundary of the strip. Therefore, if $x \in \mathbb{R}$, then the superharmonicity in some neighborhood of $(x, \pm 1)$ follows directly from the superharmonicity of $(x, y) \mapsto 1 - K|x|$ and the mean value property.

The passage from bounded to unbounded processes X, Y is performed with the same arguments as in the L^p case above. To see that the constant K is optimal, we consider the two-dimensional Brownian motion (X, Y) started at the origin and stopped upon leaving the strip $\mathbb{R} \times (-1, 1)$. Then X, Y are orthogonal and Y is differentially subordinate to X. Furthermore, U is harmonic in the interior of the strip $\mathbb{R} \times [-1, 1]$ (in which the pair (X, Y) takes its values), so Itô's formula implies that $(U(X_t, Y_t))_{t \geq 0}$ is a martingale starting from 0. However, as we will show now, this martingale is bounded in L^2. Indeed, first note that $\mathbb{E}|X_t|^2 = \mathbb{E}|Y_t|^2 \leq 1$ for all t: hence X is square integrable. Furthermore, as we have already shown above, we have

$$-K|x| \leq U(x, y) \leq 1 - K|x| \qquad \text{on } \mathbb{R} \times [-1, 1].$$

Thus $(U(X_t, Y_t))_{t \geq 0}$ is bounded in L^2 and hence in particular $\mathbb{E}U(X_\infty, Y_\infty) = 0$. Since $|Y_\infty| = |W_\tau^2| = 1$ and U coincides with G at the boundary of the strip $\mathbb{R} \times [-1, 1]$, this implies $\mathbb{P}(|Y_\infty| \geq 1) - K\mathbb{E}|X_\infty| = 0$. It remains to write

$$\lim_{t \to \infty} \frac{\mathbb{P}(|Y_t| \geq 1)}{\mathbb{E}|X_t|} = \frac{\mathbb{P}(|Y_\infty| \geq 1)}{\mathbb{E}|X_\infty|} = K,$$

thus proving the desired sharpness.

3.5. Problems

1. Find the smallest Bellman function corresponding to the inequality

$$\mathbb{P}(g_n \geq 1) \leq 2\mathbb{E}|f_n|, \qquad n = 0, 1, 2, \ldots.$$

2. (Burkholder [**2**], Suh [**20**]) Let $1 \leq p < \infty$. Prove that the best constant in the weak-type inequality

$$\mathbb{P}(|g_n| \geq 1) \leq C_p^p \mathbb{E}|f_n|^p, \qquad n = 0, 1, 2, \ldots,$$

is given by

$$C_p^p = \begin{cases} \dfrac{2}{\Gamma(p+1)} & \text{if } 1 \leq p \leq 2, \\ \dfrac{p^{p-1}}{2} & \text{if } p > 2. \end{cases}$$

3. (Burkholder [**3**]) Let $a < b$ are fixed real numbers. Find the best constant C in the weak-type inequality

$$\mathbb{P}(|g_n| \geq 1) \leq C\mathbb{E}|f_n|, \qquad n = 0, 1, 2, \ldots,$$

under the assumption that g is a transform of f by a predictable sequence with values in $[a, b]$.

4. (Osękowski [**15**]) For any $K > 1$, find the least constant $L(K)$ in the logarithmic estimate

$$\mathbb{E}|g_n| \leq K\mathbb{E}|f_n| \log |f_n| + L(K), \qquad n = 0, 1, 2, \ldots.$$

5. (Osękowski [**16**]) For any $1 < p < \infty$, find the least constant C_p in the inequality

$$\mathbb{E}|g_n| \leq C_p \left(\mathbb{E}|f_n|^p\right)^{1/p}, \qquad n = 0, 1, 2, \ldots,$$

where f, g are martingales such that g is differentially subordinate to f.

6. Find the best constants in the estimates

$$\mathbb{P}(|X_t|+|Y_t|\geq 1)\leq C_1\mathbb{E}|X_t|, \qquad t\geq 0,$$
$$\mathbb{P}(X_t^2+Y_t^2\geq 1)\leq C_2\mathbb{E}X_t^2, \qquad t\geq 0,$$

under the assumption that X, Y are orthogonal martingales such that Y is differentially subordinate to X.

7. (Tomaszewski [**21**]) Find the best constant C in the inequality

$$\mathbb{P}(X_t^2+Y_t^2\geq 1)\leq C\mathbb{E}|X_t|, \qquad t\geq 0,$$

under the assumption that X, Y are orthogonal martingales such that Y is differentially subordinate to X.

8. For a given $K>0$, find the least $L=L(K)$ such that if X, Y are orthogonal martingales, X is nonnegative and Y is differentially subordinate to X, then

$$\mathbb{E}|Y_t|\leq K\mathbb{E}X_t\log X_t+L(K), \qquad t\geq 0.$$

9. (Strzelecki [**19**]) Let X, Y be orthogonal martingales such that Y is differentially subordinate to X. Prove that the inequality

$$\mathbb{P}(Y_t\geq 1)\leq \mathbb{E}|X_t|^2, \qquad t\geq 0,$$

is sharp. Find the least Bellman function corresponding to this estimate.

10. (Osękowski [**17**]) Let X, Y be orthogonal martingales such that Y is differentially subordinate to X. Prove that for any $t\geq 0$ we have the sharp inequality

$$\mathbb{P}(Y_t\geq 1)\leq \mathbb{E}|X_t|.$$

CHAPTER 4

Inequalities for maximal functions

4.1. Definitions and notation

In this chapter we will study the boundedness properties of maximal operators. Such estimates play a fundamental role in various areas of mathematics, and it is often of interest to have a sharp, or at least tight information on the constants involved.

We start with a classical Euclidean setting. Let $d \geq 1$ be a fixed dimension. The (centered) Hardy-Littlewood maximal operator $\mathcal{M}_c$ acts on locally integrable functions $f : \mathbb{R}^d \to \mathbb{R}$ by the formula

$$\mathcal{M}_c f(x) = \sup \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)| \mathrm{d}y,$$

where the supremum is taken over $r > 0$ (here $B(x,r)$ is the ball of center x and radius x). The uncentered version of $\mathcal{M}_c$, denoted by $\mathcal{M}$ (i.e, without the lower index c), is given by the same formula but this time the supremum is taken over a larger family: we consider all balls $B \subset \mathbb{R}^d$ which just *contain* x inside.

Actually, from the viewpoint of boundedness properties, the choice of the maximal operator (i.e., centered or uncentered) is irrelevant, since both operators are equivalent up to a constant depending on the dimension. Specifically, we have $\mathcal{M}_c f \leq \mathcal{M} f$ (which follows trivially from the very definition) and $\mathcal{M} f \leq 2^d \mathcal{M}_c f$. To show the latter bound, note that for any $x \in \mathbb{R}^d$ and any ball $B \subset \mathbb{R}^d$ of radius r containing x we have $B \subset B(x, 2r)$, so

$$\frac{1}{|B|} \int_B |f(y)| \mathrm{d}y \leq \frac{2^d}{|B(x,2r)|} \int_{B(x,2r)} |f(y)| \mathrm{d}y \leq \mathcal{M}_c f(x)$$

and taking the supremum over all B gives the desired bound. We should also mention here that sometimes in the definitions of $\mathcal{M}_c$ and $\mathcal{M}$ one considers cubes with sides parallel to the axes instead of balls. Arguing as above, we get that these objects are essentially of the same size (pointwise), up to a constant depending on the dimension only.

EXAMPLE 4.1. Consider the following simple application. Let $\Phi : \mathbb{R}^d \to [0, \infty)$ be a C^1 function satisfying $\int_{\mathbb{R}^d} \Phi \, \mathrm{d}x = 1$ and let $(\Phi_t)_{t>0}$ be the associated approximation of identity given by $\Phi_t(x) = t^{-d} \Phi(x/t)$. In addition, we assume that Φ is radially decreasing, i.e., there exists a C^1 nonincreasing function $\varphi : [0, \infty) \to [0, \infty)$ such that $\Phi(x) = \varphi(|x|)$. Then for any locally integrable function $f : \mathbb{R}^d \to \mathbb{R}$ and any $x \in \mathbb{R}^d$ we have

$$\sup_{t>0} |(f * \Phi_t)(x)| \leq \mathcal{M}_c f(x). \tag{4.1}$$

To see this, fix $t > 0$ and observe that

$$\Phi(x) = \varphi(|x|) = -\int_{|x|}^{\infty} \varphi'(r)\mathrm{d}r = \int_0^{\infty} (-\varphi'(r))\chi_{B(0,r)}(x)\mathrm{d}r. \tag{4.2}$$

Consequently,

$$\begin{aligned}
|(f * \Phi_t)(x)| &= \left| \int_{\mathbb{R}^d} \int_0^{\infty} f(y) \cdot t^{-d}(-\varphi'(r))\chi_{B(0,r)}((x-y)/t)\mathrm{d}r\mathrm{d}y \right| \\
&\leq \int_0^{\infty} \int_{\mathbb{R}^d} \left| f(y) \cdot t^{-d}(-\varphi'(r))\chi_{B(0,r)}((x-y)/t) \right| \mathrm{d}y\mathrm{d}r \\
&= \int_0^{\infty} t^{-d}(-\varphi'(r)) \int_{\mathbb{R}^d} |f(y)|\chi_{B(x,rt)}(y)\mathrm{d}y\mathrm{d}r \\
&\leq \int_0^{\infty} t^{-d}(-\varphi'(r))|B(x,rt)|\mathcal{M}_c f(x)\mathrm{d}r \\
&= \mathcal{M}_c f(x) \int_0^{\infty} (-\varphi'(r))|B(0,r)|\mathrm{d}r \\
&= \mathcal{M}_c f(x) \cdot \int_{\mathbb{R}^d} \Phi(x)\mathrm{d}x,
\end{aligned}$$

where in the last line we have exploited (4.2) and Fubini's theorem. Since the integral of Φ is equal to 1, (4.1) follows.

Here is the further application. Suppose that $1 < p < \infty$ and let $f \in L^p(\mathbb{R}^d)$. Then for Φ and $(\Phi_t)_{t>0}$ as above, we have

$$f * \Phi_t \to f \quad \text{in } L^p \text{ as } t \to 0. \tag{4.3}$$

Let us first show the convergence for continuous functions having bounded support. Pick such a function and observe that it is uniformly continuous. Therefore, a standard argument shows that $f * \Phi_t \to f$ pointwise as $t \to 0$. But

$$|f * \Phi_t(x) - f(x)| \leq \sup_{t>0} |f * \Phi_t(x)| + |f(x)|$$

and both summands on the right are in L^p: here we use (4.1) and the fact that $\mathcal{M}_c$ is bounded on L^p. This proves (4.3) for continuous, compactly supported functions and the passage to the general case follows from a standard approximation argument.

Unfortunately, it seems that Bellman function technique does not work for the above maximal operators (i.e., there seems to be no modification of the method which would lead to sharp estimates). Roughly speaking, the reason is that given x, the class of all balls/cubes containing x, or the class of all balls/cubes having x as its center, does not possess any ordered structure which would enable the inductive argument which lies at the very heart of the Bellman function method. This is why we will work with yet another maximal operator, associated with the dyadic lattice in $\mathbb{R}^d$. Let $\mathcal{D}$ denote the class of all dyadic cubes in $\mathbb{R}^d$, i.e., all cubes of the form

$$Q = (2^n a_1, 2^n(a_1+1)] \times (2^n a_2, 2^n(a_2+1)] \times \ldots (2^n a_d, 2^n(a_d+1)], \tag{4.4}$$

where n, a_1, a_2, $\ldots$, a_d run over the set of all integers. Note that in the dyadic setting we have the following dichotomy: for any Q_1, $Q_2 \in \mathcal{D}$, the cubes are either disjoint, or one of them is contained in the other. This property induces an appropriate structure on the dyadic lattice which makes Bellman function method applicable.

The dyadic maximal operator M acts on locally integrable functions f on $\mathbb{R}^d$ by the formula

$$Mf(x) = \sup\left\{\frac{1}{|Q|}\int_Q |f(y)|\mathrm{d}y \,:\, x \in Q \in \mathcal{D}\right\}.$$

Actually, one can study the dyadic maximal operators in a yet more general context which arises naturally in probability theory. Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, equipped with some discrete-time filtration $(\mathcal{F}_n)_{n\geq 0}$. Given an integrable random variable f, let $(f_n)_{n\geq 0}$ be the associated martingale (with a slight abuse of notation, this martingale will again be denoted by f), given by $f_n = \mathbb{E}(f|\mathcal{F}_n)$ for all n. Then the maximal function of f is defined by

$$f^* = \sup_{n\geq 0} |f_n|.$$

We will also consider its truncated version, given by $f_N^* = \max_{0\leq n\leq N} f_n$. To see how the above probabilistic setup generalizes the preceding analytic setting, set $\Omega = (0,1]^d$, $\mathcal{F} = \mathcal{B}((0,1]^d)$ and let $\mathbb{P}$ be the Lebesgue measure on $(0,1]^d$. Next, for any $n \geq 0$, let $\mathcal{F}_n$ be the σ-algebra generated by dyadic cubes contained in $(0,1]^d$, with volume at least 2^{-nd}. Then $(\mathcal{F}_n)_{n\geq 0}$ is a nondecreasing family of sub-σ-algebras of $\mathcal{F}$. Furthermore, if f is an integrable function on $(0,1]^d$, then

$$\sup \frac{1}{|Q|}\left\{\int_Q |f|\mathrm{d}x \,:\, x \in Q \subseteq (0,1]^d,\ Q \text{ dyadic}\right\},$$

is precisely the maximal function of the martingale induced by the random variable $|f|$. Consequently, having proved any inequality for the martingale maximal function, we immediately recover its version (with the same constants involved) for the "localized" dyadic maximal operator, i.e., restricted to the unit cube $(0,1]^d$. Now the use of standard dilation and translation arguments allow to extend such an inequality to the "non-local" dyadic maximal operator, without any change in the constants.

Another important application of the probabilistic setup concerns the one-sided Hardy-Littlewood maximal operators on the nonnegative halfline (which can be regarded as a different approach to the integral inequalities studied in Chapter 1). Let $\Omega = (0,1]$, $\mathcal{F} = \mathcal{B}((0,1])$ and suppose that $\mathbb{P}$ is the Lebesgue measure on $(0,1]$. Fix a large positive integer N and consider the finite filtration $(\mathcal{F}_n)_{n=0}^N$, where $\mathcal{F}_n$ is generated by the intervals $(0, 1-n/N]$, $(1-n/N, 1-(n-1)/N]$, $(1-(n-1)/N, 1-(n-2)/N]$, …, $(1-1/N, 1]$. Extend the filtration by setting $\mathcal{F}_k = \mathcal{F}_N$ for $k > N$. Given an integrable and nonnegative function f on $(0,1]$, consider the associated martingale $(f_n)_{n\geq 0}$. This martingale evolves up to time N and its terminal variable f_N is constant on any interval of the form $(k/N, (k+1)/N]$ (where $k = 0, 1, 2, \ldots, N-1$), equal to $N\int_{k/N}^{(k+1)/N} f$ there. On the other hand, we easily compute that the maximal function f^* of the martingale satisfies

$$f^*(x) = f_N^*(x) = \max_{0\leq n\leq N} f_n(x) \geq \frac{1}{\lceil Nx\rceil/N}\int_0^{\lceil Nx\rceil/N} f(u)\mathrm{d}u.$$

Now, if we let $N \to \infty$, then the expression on the right converges to $\frac{1}{x}\int_0^x f(u)\mathrm{d}u$, the one-sided Hardy-Littlewood function of f evaluated at x. On the other hand, the terminal variable f_N converges almost everywhere to f, in the light of Lebesgue's differentiation theorem. In practice, this often means that if we manage to establish the martingale inequality

$$\mathbb{E}G(f_N, f_N^*) \leq 0,$$

then it automatically proves the analytic estimate

$$\int_0^1 G\left(f(x), \frac{1}{x}\int_0^x f(u)\mathrm{d}u\right) \leq 0.$$

The above discussion justifies why in our considerations below we restrict ourselves to the probabilistic setup and nonnegative martingales. We should point out that essentially all the estimates we will establish will be sharp even in the special analytic context of dyadic operators.

4.2. On the method

Now we will describe the modification of Bellman function method which allows the study of sharp estimates for the pair (f, f^*), where f is a *simple* nonnegative martingale and f^* is its maximal function. In all cases, the simplicity assumption can be imposed by some standard approximation arguments and, as in the previous chapters, it will enable us to avoid technical integrability issues. The pair (f, f^*) takes values in the domain

$$Dom = \{(x, y) \in \mathbb{R} : 0 \leq x \leq y\}.$$

Suppose that $G : Dom \to \mathbb{R}$ is a given function and assume that we are interested in showing the estimate

$$\mathbb{E}G(f_n, f_n^*) \leq 0, \qquad n = 0,\, 1,\, 2,\, \ldots, \tag{4.5}$$

for any pair (f, f^*) as above. The underlying idea of the approach is the same as in the previous chapters: for a given (f, f^*), we search for a supermartingale $(U_n)_{n\geq 0}$ majorizing the sequence $(G(f_n, f_n^*))_{n\geq 0}$ and satisfying $\mathbb{E}U_0 \leq 0$. Actually, we look for a function $U : Dom \to \mathbb{R}$, not depending on f and g, such that the supermartingale is of the form $U_n = U(f_n, g_n)$ for all n. Then the above requirements for U_n imply that U should satisfy the following properties.

1° (initial condition) For any $x \geq 0$, we have $U(x, x) \leq 0$.

2° (majorization) We have $U \geq G$ on Dom.

3° (concavity) For any $(x, y) \in D$ and any numbers $t_1,\, t_2 \geq -x$, $\alpha_1,\, \alpha_2 \in (0, 1)$ such that $\alpha_1 t_1 + \alpha_2 t_2 = 0$, we have

$$U(x, y) \geq \alpha_1 U(x + t_1, (x + t_1) \vee y) + \alpha_2 U(x + t_2, (x + t_2) \vee y). \tag{4.6}$$

Let us study precisely the relation between the validity of (4.5) and the existence of a function U possessing the above three properties.

LEMMA 4.1. *If U satisfies 1°, 2° and 3°, then* (4.5) *holds true.*

PROOF. This is done in the same manner as in the previous chapters. Fix a pair (f, f^*). Condition 3°, by standard induction, implies that if $(x, y) \in Dom$ and X is a simple mean-zero random variable taking values in $[-x, \infty)$, then $U(x, y) \geq$

$\mathbb{E}U(x+X,(x+X)\vee y)$. Applying this inequality conditionally yields

$$\begin{aligned}\mathbb{E}\big[U(f_{n+1},f_{n+1}^*)|\mathcal{F}_n\big] &= \mathbb{E}\big[U(f_{n+1},f_{n+1}\vee f_n^*)|\mathcal{F}_n\big]\\ &= \mathbb{E}\big[U(f_n+df_{n+1},(f_n+df_{n+1})\vee f_n^*)|\mathcal{F}_n\big]\\ &\leq \mathbb{E}\big[U(f_n,f_n^*)|\mathcal{F}_n\big] = U(f_n,f_n^*),\end{aligned}$$

i.e., the desired supermartingale property. Consequently, applying 2° and then 1°, we get

$$\mathbb{E}G(f_n,f_n^*) \leq \mathbb{E}U(f_n,f_n^*) \leq \mathbb{E}U(f_0,f_0^*) = \mathbb{E}U(f_0,f_0) \leq 0$$

and we are done. □

The implication of the above lemma can be reversed. The Bellman function $V: Dom \to \mathbb{R}$ associated with the problem (4.5) is defined by

$$V(x,y) = \sup \mathbb{E}G(f_n, f_n^* \vee y), \tag{4.7}$$

where the supremum is taken over all n and the class of all simple martingales f starting from x (the use of $f_n^*\vee y$ on the second variable guarantees that the maximal process on the second coordinate starts from y). The probability space is also vary, unless it is assumed to be nonatomic.

LEMMA 4.2. *If (4.5) is valid, then the function V satisfies 1°, 2° and 3°.*

PROOF. The condition 1° is a direct consequence of the validity of (4.5): for any $x \geq 0$ and any simple martingale f starting from x we have

$$\mathbb{E}G(f_n, f_n^*\vee x) = \mathbb{E}G(f_n,f_n^*) \leq 0,$$

and taking the supremum over all such f gives the inequality. The majorization 2° follows by considering the constant martingale $f\equiv x$ in the definition of $V(x,y)$. To show 3°, fix $(x,y)\in Dom$ and t_1, t_2, α_1, α_2 as in the statement of the condition. Let f^1, f^2 be simple martingales as in the definition of $V(x+t_1,(x+t_1)\vee y)$ and $V(x+t_2,(x+t_2)\vee y)$, respectively. Suppose that these processes are given on some probability spaces $(\Omega^1,\mathcal{F}^1,\mathbb{P}^1)$ and $(\Omega^2,\mathcal{F}^2,\mathbb{P}^2)$. We splice these probability spaces into one triple $(\Omega^1\cup\Omega^2,\sigma(\mathcal{F}^1,\mathcal{F}^2),\alpha_1\mathbb{P}^1+\alpha_2\mathbb{P}^2)$. Here, as in the setting of martingale transforms, the probability is given by

$$(\alpha_1\mathbb{P}^1+\alpha_2\mathbb{P}^2)(A_1\cup A_2) = \alpha_1\mathbb{P}^1(A_1)+\alpha_2\mathbb{P}^2(A_2)$$

for any $A_1\in\mathcal{F}^1$ and $A_2\in\mathcal{F}^2$. We splice the martingales f^1 and f^2 into one sequence $f=(f_n)_{n\geq 0}$ on the new probability space, setting $f_0\equiv x$ and, for $n\geq 1$,

$$f_n(\omega) = \begin{cases} f_{n-1}^1(\omega) & \text{if } \omega\in\Omega^1,\\ f_{n-1}^2(\omega) & \text{if } \omega\in\Omega^2.\end{cases}$$

Then it is easy to see that f is a simple martingale with respect to its natural filtration $(\tilde{\mathcal{F}}_n)_{n\geq 0}$: the assumption $\alpha_1t_1+\alpha_2t_2=0$ implies that the first move of the sequence f is of martingale type (precisely: we have $\mathbb{E}(f_1|\tilde{\mathcal{F}}_0)=\mathbb{E}f_1=x$), and the martingale properties of f^i yield that f also enjoys this property for later steps. Furthermore, since $x\leq y$, we see that

$$f_n^*(\omega)\vee y = \max_{0\leq k\leq n} f_k(\omega) = \max_{1\leq k\leq n} f_k(\omega) = \begin{cases} (f_{n-1}^1)^*(\omega)\vee y & \text{if } \omega\in\Omega^1,\\ (f_{n-1}^2)^*(\omega)\vee y & \text{if } \omega\in\Omega^2.\end{cases}$$

Therefore,

$$V(x,y) \geq \mathbb{E}G(f_n,f_n^*\vee y) = \alpha_1\mathbb{E}^1G(f_{n-1}^1,(f_{n-1}^1)^*\vee y) + \alpha_2\mathbb{E}^2G(f_{n-1}^2,(f_{n-1}^2)^*\vee y)$$

and taking the supremum over all n and all f^1, f^2, we get the desired concavity condition. □

The above approach translates the problem of proving (4.5) into that of the search for an appropriate function U. How can we find such an object? The first step towards the answer to this question, in analogy to the context of martingale transforms, lies in considering the „finite-horizon" version of the problem. Namely, for any $n = 0, 1, 2, \ldots$, define $V^n : Dom \to \mathbb{R}$ by the formula

$$V^n(x,y) = \sup \mathbb{E}G(f_n, f_n^* \vee y),$$

where the supremum is taken over all simple martingales f starting from x. Here is the analogue of Theorem 3.1. As the proof requires only some minor changes, we leave it to the reader.

THEOREM 4.1. *The sequence $(V^n)_{n\geq 0}$ can be computed inductively from the relations $V^0 = G$ and, for $n \geq 1$,*

$$V^n(x,y) = \sup\left\{\alpha_1 V^{n-1}(x+t_1, (x+t_1)\vee y) + \alpha_2 V^{n-1}(x+t_2, (x+t_2)\vee y)\right\},$$

where the supremum is taken over all numbers α_1, $\alpha_2 \geq 0$, t_1, $t_2 \in \mathbb{R}$ such that $\alpha_1 + \alpha_2 = 1$ and $\alpha_1 t_1 + \alpha_2 t_2 = 0$.

Furthermore, a simple limiting argument shows that $V(x,y) = \lim_{n\to\infty} V^n(x,y)$. This, as in the contexts of optimal stopping or martingale transforms, gives us a method of searching for V, by trying to compute the sequence $(V^n)_{n\geq 0}$ explicitly. Unfortunately, in many cases the calculations arising during the study of such a sequence are technically involved and impossible to push through. An alternative approach rests on trying to find V directly, by understanding the condition 3°. It is easy to see that this condition implies that for any fixed $y > 0$, the section $x \mapsto U(x,y)$ must be a concave function on $[0,y]$, simply by restricting the inequality (4.6) to t_1, t_2 satisfying $x + t_1 \leq y$, $x + t_2 \leq y$. In addition, (4.6) imposes the following monotonicity requirement on U on the diagonal $x = y$. Suppose for a moment that U extends to a C^1 function on some open set $\mathfrak{D}$ containing Dom inside. Then applying (4.6) to $x = y$, $\alpha_1 = \alpha_2 = 1/2$ and $t_1 = -t_2 = \delta > 0$ gives

$$U(x,x) \geq \frac{1}{2}U(x+\delta, x+\delta) + \frac{1}{2}U(x-\delta, x).$$

Putting all the terms on the right, dividing throughout by δ and letting $\delta \to 0$ gives the partial differential inequality $U_y(x,x) \leq 0$.

We will continue the reasoning by studying a little more specific example. Suppose that Φ, Ψ are fixed C^1 functions on $[0,\infty)$, Ψ is strictly convex and assume that we are interested in proving the estimate

$$\mathbb{E}\Phi(f_n^*) \leq \mathbb{E}\Psi(f_n), \qquad n = 0, 1, 2, \ldots. \tag{4.8}$$

Obviously, this corresponds to the choice $G(x,y) = \Phi(y) - \Psi(x)$. Let us provide an informal reasoning which will lead us to an appropriate candidate for the special function. Let V be defined by (4.7). Introduce the continuation and the stopping sets by

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

What can be said about the shape of C and D? The first observation is that if $(x,y) \in D$, then also $(x',y) \in D$ for $x' < x$. The reason for this implication is the following. Directly from the formula for V, we see that when studying $V(x,y)$, we want to find a martingale f starting from x for which $\mathbb{E}\Phi(f_n^* \vee y)$ is relatively big while $\mathbb{E}\Psi(f_n)$ is relatively small. However, the process $f_n^* \vee y$ increases only if it gets on the diagonal $\{(z,z) : z \geq y\}$; otherwise it remains constant. On the other hand, since the function Ψ is strictly convex, the expectation $\mathbb{E}\Psi(f_n)$ increases at each step. This suggests that if the point (x,y) is too far from the diagonal, then it is not beneficial to send the process in the neighborhood of this set, since the potential gain obtained from the increase of $\mathbb{E}\Phi(f_n^*)$ would be totally consumed by the increase of $\mathbb{E}\Psi(f_n)$. Therefore, if $(x,y) \in D$ (the point (x,y) is already too far from the diagonal), then the whole set $[0,x] \times \{y\}$ should be contained in the stopping set.

The further observation is the following. As usual, we expect that on the continuation set, the concavity condition is actually the linearity condition. That is (we pass from the letter V to U, as usual), if $(x,y) \in C$ and $x < y$, then U should be linear along some small line segment $(x-\delta, x+\delta) \times \{y\}$, and if $(x,x) \in C$, then we expect the equality $U_y(x,x) = 0$ (or perhaps rather $U_y(x,x+) = 0$; since (x,x) lies on the diagonal, we should treat the partial derivative as the one-sided object). However, if Φ is strictly increasing (which is very often the case), then the whole diagonal $x = y$ lies in C: indeed, if for some x we have $U(x,x) = G(x,x)$, then for any $\delta \in [0,x]$ we may write

$$\begin{aligned} G(x,x) = U(x,x) &\geq \frac{1}{2}\left(U(x-\delta,x) + U(x+\delta,x+\delta)\right) \\ &\geq \frac{1}{2}\left(G(x-\delta,x) + G(x+\delta,x+\delta)\right), \end{aligned}$$

which by a similar limiting argument as above gives $G_y(x,x+) = \Phi'(x) \leq 0$, a contradiction.

Putting all the above facts together, we arrive at the following general method of handling (4.8). There should be a curve $\gamma : [0,\infty) \to [0,\infty)$ satisfying $\gamma(y) < y$, which is a boundary between C and D: see Figure 4.2 below. That is, we expect

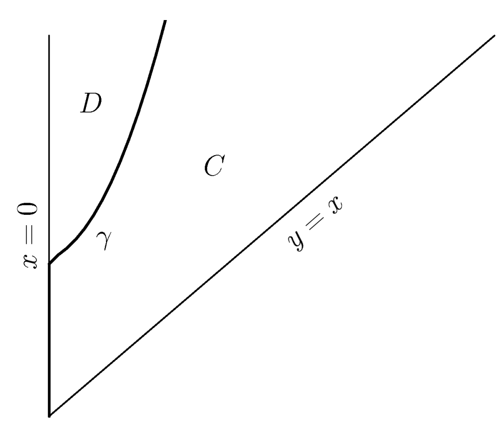

FIGURE 4.1. The structure of the continuation and the stopping regions.

that the continuation and the stopping regions should be of the form

$$C = \{(x,y) \in D : x > \gamma(y)\}, \qquad D = \{(x,y) \in D : x \leq \gamma(y)\}.$$

We put $U(x,y) = G(x,y)$ on D. To find V on C, we first note that for any $y > 0$ such that $\gamma(y) > 0$ we must have $U_x(\gamma(y), y) = G_x(\gamma(y), y)$. Indeed, otherwise the concavity along the variable x or the majorization 2° would not be satisfied at some point lying in the neighborhood of $(\gamma(y), y)$. Consequently, combining this observation with the linearity of $V(\cdot, y)$ on the continuation set, we see that necessarily

$$U(\gamma(y) + t, y) = G(\gamma(y), y) + G_x(\gamma(y), y)t$$

for $t \in [0, y - \gamma(y)]$. Now the inequality $U_y(x, x+)$ implies the differential equation for γ. In many cases this can be solved explicitly and we obtain the candidate for the special function: if it satisfies all the required conditions, then (4.8) follows.

4.3. Examples

EXAMPLE 4.2. Let $1 \leq p < \infty$. We will find the best constant in the weak type (p,p) inequality

$$\mathbb{P}(f_n^* \geq 1) \leq c^p \mathbb{E} f_n^p, \qquad n = 0, 1, 2, \ldots, \tag{4.9}$$

where $f = (f_n)_{n \geq 0}$ is a simple nonnegative martingale. We let $G(x,y) = 1_{\{y \geq 1\}} - c^p x^p$ and write down the formula for the Bellman function

$$V(x,y) = \sup \left\{ \mathbb{P}(f_n^* \vee y \geq 1) - c^p \mathbb{E} f_n^p \right\},$$

the supremum taken over all simple nonnegative martingales f starting from x. If $y \geq 1$, then the optimal choice is to take the constant martingale: then $\mathbb{E} f_n^p$ is the least, while $\mathbb{P}(f_n^* \vee y \geq 1) = 1$ is the largest possible. This shows that

$$V(x,y) = 1 - c^p x^p \qquad \text{if } y \geq 1.$$

On the other hand, if $y < 1$, then we may write

$$V(x,y) = \sup \left\{ \mathbb{P}(f_n^* \geq 1) - c^p \mathbb{E} f_n^p \right\},$$

and hence $V(x,y)$ depends only on x variable: $V(x,y) = v(x)$ for some $v : [0,1) \to \mathbb{R}$, with $v(0) = 0$ and, by the expected continuity of V, $v(1-) = 1 - c^p$. In the light of the above discussion, we expect the diagonal $\{(x,x) : x > 0\}$ to lie in the continuation region; this implies $v(x) > -c^p x^p$ for $x \in (0,1)$. This in turn leads to the estimate $V(x,y) > G(x,y)$ for all $0 < x \leq y < 1$ and consequently, v must be linear: $v(x) = ax$ for some a. By the initial condition, we must have $a \leq 0$; on the other hand, the majorization 2° enforces $a \geq 0$. Hence $v \equiv 0$ and $c = 1$: putting all the facts together, we obtain the candidate

$$U(x,y) = \begin{cases} 0 & \text{if } y < 1, \\ 1 - x^p & \text{if } y \geq 1. \end{cases}$$

It is very easy to check that U enjoys the conditions 1°, 2° and 3°, so (4.9) holds with the constant 1. Obviously, this choice is optimal, which can be immediately verified by considering the constant martingale $f \equiv 1$.

EXAMPLE 4.3. Now we will find the best constant in Doob's inequality

$$\mathbb{E}(f_n^*)^p \le C_p^p \mathbb{E} f_n^p, \qquad n = 0,\, 1,\, 2,\, \ldots, \tag{4.10}$$

where $1 < p < \infty$ is fixed. This inequality corresponds to the choice $G(x,y) = y^p - C_p^p x^p$ and the associated Bellman function is given by

$$V(x,y) = \sup \mathbb{E}\big\{(f_n^* \vee y)^p - C_p^p f_n^p\big\},$$

where the supremum is taken over all n and all simple martingales starting from x. Using a standard homogenization argument, we see that

$$V(\lambda x, \lambda y) = \lambda^p V(x,y)$$

for any $\lambda > 0$ and any $(x,y) \in Dom$. Consequently, we may write $V(x,y) = y^p v(x/y)$, for some unknown function $v : [0,1] \to \mathbb{R}$ to be found. Clearly, $v(0) = 1$. Furthermore, directly from the above abstract reasoning, we expect the existence of some number $\gamma_0 \in (0,1)$ such that

$$v(x) = \begin{cases} 1 - C_p^p x^p & \text{if} x \in [0, \gamma_0], \\ 1 - C_p^p \gamma_0^p - pC_p^p \gamma_0^{p-1}(x - \gamma_0) & \text{if } x \in [\gamma_0, 1]. \end{cases}$$

The requirement $V_y(1,1) = 0$ is equivalent to the equality $pv(1) - v'(1) = 0$ or, after some manipulations,

$$C_p^p = \frac{1}{(p-1)(\gamma_0^{p-1} - \gamma_0^p)}.$$

The right-hand side, considered as a function of γ_0, attains its minimal value equal to $(p/(p-1))^p$ at the point $\gamma_0 = (p-1)/p$. This suggests taking $C_p = p/(p-1)$ and the special function

$$U(x,y) = \begin{cases} y^p - \left(\dfrac{p}{p-1}\right)^p x^p & \text{if } x \le \dfrac{p-1}{p} y, \\ py^{p-1}\left(y - \dfrac{p}{p-1} x\right) & \text{if } x > \dfrac{p-1}{p} y. \end{cases}$$

One can verify readily that U has the desired properties and hence the inequality (4.10) holds with the constant $p/(p-1)$. Suppose that the best constant is equal to C_p and let V be the Bellman function associated with this sharp estimate. Fix $\delta > 0$ and use 3° with $x = y = 1$, $t_1 = \delta$, $t_2 = -1/p$ and $\alpha_1 = 1 - \alpha_2 = (1 + p\delta)^{-1}$ to get

$$\begin{aligned} V(1,1) &\ge \frac{1}{1+p\delta} V(1+\delta, 1+\delta) + \frac{p\delta}{1+p\delta} V(1 - 1/p, 1) \\ &\ge \frac{(1+\delta)^p}{1+p\delta} V(1,1) + \frac{p\delta}{1+p\delta} G(1 - 1/p, 1), \end{aligned} \tag{4.11}$$

where in the last passage we have used the homogeneity of V and the majorization property 2°. (The reason why we have chosen these particular constants is given at the end.) This estimate can be rewritten in the equivalent form

$$\frac{1 + p\delta - (1+\delta)^p}{p\delta} V(1,1) \ge 1 - C_p^p \left(\frac{p-1}{p}\right)^p.$$

Letting $\delta \to 0$ makes the left hand side vanish; this means $C_p \ge p/(p-1)$ and we are done.

It remains to comment on the parameters which have led us to (4.11). This choice actually encodes the move of the almost-optimal pair (f, f^*) starting from $(1, 1)$. It follows from the previous considerations that the continuation set is expected to be $\{(x, y) : \frac{p-1}{p} y \leq x \leq y\}$. Therefore, a natural candidate for the pair (f, f^*) is the following: in the first step, either go to the boundary between the stopping and the continuation set, i.e., to $((p-1)/p, 1)$, or move a little bit along the diagonal (i.e., jump to the point $(1+\delta, 1+\delta)$). If the first possibility occurs (which happens with probability α_2), then stop; otherwise, continue using the same pattern (i.e., go to $((1+\delta)^2, (1+\delta)^2)$ or to $\big((1+\delta)(p-1)/p, 1+\delta\big)$, and so on). Then the stopping when reaching $\partial C \cap \partial D$ is reflected in (4.11) by replacing V with G, while the phrase "continue using the same pattern" is precisely the replacement of $V(1+\delta, 1+\delta)$ with $(1+\delta)^p V(1, 1)$.

4.4. An alternative approach: Carleson sequences

There is a different method, also based on the construction of a certain special function, which can be used to study maximal estimates. Suppose that Φ, $\Psi : [0, \infty) \to [0, \infty)$ are given two nondecreasing functions and suppose we are interested in showing the estimate

$$\mathbb{E}\Phi(f_n^*) \leq \mathbb{E}\Psi(f_n), \qquad n = 0, 1, 2, \ldots, \tag{4.12}$$

for any n and any nonnegative simple martingale f. We may and will assume that the filtration has the following tree-type structure:

(i) $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

(ii) For any $k = 0, 1, 2, \ldots$ and any atom A of $\mathcal{F}_k$ with $\mathbb{P}(A) > 0$, there is a finite collection $A_1, A_2, \ldots, A_m$, $m \geq 2$, of pairwise disjoint elements of $\mathcal{F}_{k+1}$ such that $\mathbb{P}(A_j) > 0$ and $\bigcup_{j=1}^m A_j = A$.

The second condition above simply asserts that any nontrivial atom of any $\mathcal{F}_k$ is split, in $\mathcal{F}_{k+1}$, into at least two atoms of positive measure. Let $\mathtt{At}$ be the collection of all the atoms of the algebras $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$.

Consider a function $U : [0, \infty) \times [0, 1] \to \mathbb{R}$ which satisfies the following properties.

1° We have $U(x, 1) \leq 0$ for any $x \geq 0$.

2° We have $U(x, 0) \geq -\Psi(x)$ for all $x \geq 0$.

3° For any $(x, y) \in [0, \infty) \times [0, 1]$, any $\alpha_1, \alpha_2 \in (0, 1)$ and any $x_1, x_2, y_1, y_2 \in [0, 1]$ such that $\alpha_1 + \alpha_2 = 1$, $\alpha_1 x_1 + \alpha_2 x_2 = x$ and $\alpha_1 y_1 + \alpha_2 y_2 \leq y$ we have

$$U(x, y) \geq \alpha_1 U(x_1, y_1) + \alpha_2 U(x_2, y_2) + \Phi(x)(y - \alpha_1 y_1 - \alpha_2 y_2). \tag{4.13}$$

In particular, assuming $y = \alpha_1 y_1 + \alpha_2 y_2$ in 3°, we see that U must be concave on $[0, \infty) \times [0, 1]$. Furthermore, the inequality (4.13) implies that if $x \geq 0$, $y \in [0, 1]$, and $(\alpha_k)_{k=1}^m \subset (0, 1)$, $(x_k)_{k=1}^m \subset [0, \infty)$ and $(y_k)_{k=1}^m$ satisfy

$$\sum_{k=1}^m \alpha_k = 1, \qquad \sum_{k=1}^m \alpha_k x_k = x, \qquad \sum_{k=1}^m \alpha_k y_k \leq y,$$

then

$$U(x, y) \geq \sum_{k=1}^m \alpha_k U(x_k, y_k) + \Phi(x) \left(y - \sum_{k=1}^m \alpha_k y_k \right). \tag{4.14}$$

To see this, apply first (4.13) to the weights α_1, $\alpha_2' = \alpha_2 + \alpha_3 + \ldots + \alpha_m$ and the points

$$x_1, \quad x_2' = \frac{\alpha_2 x_2 + \ldots + \alpha_m x_m}{\alpha_2 + \ldots + \alpha_m}, \quad y_1, \quad y_2' = \frac{\alpha_2 y_2 + \ldots + \alpha_m y_m}{\alpha_2 + \ldots + \alpha_m}.$$

As the result, one gets

$$U(x,y) \geq \alpha_1 U(x_1, y_1) + \alpha_2' U(x_2', y_2') + \Phi(x)\left(y - \sum_{k=1}^m \alpha_k y_k\right)$$

and it suffices to combine this with the estimate

$$U(x_2', y_2') \geq \sum_{k=2}^m \frac{\alpha_k}{\alpha_2'} U(x_k, y_k),$$

which follows from the concavity of U.

The relation between the existence of the function U and the validity of (4.12) is described in the two lemmas below.

LEMMA 4.3. *If there is U satisfying the properties 1°, 2° and 3°, then* (4.12) *holds true.*

PROOF. As usual, we will show that the composition of U with certain processes yields a supermartingale-type process. Fix a martingale $f = (f_n)_{n\geq 0}$ and fix a nonnegative integer n; we will prove the estimate

$$\mathbb{E}\Phi(f_n^*) \leq \mathbb{E}\Psi(f_n).$$

We may modify $(f_k)_{k\geq 0}$ by setting $f_k = f_n$ for $k \geq n$; this will not affect the above inequality. The starting point is the following simple observation: since $f_n^* = \max_{0\leq k\leq n} f_k$, for each $\omega \in \Omega$ there is an integer $m = m(\omega) \leq n$ and an atom $A = A(\omega)$ of $\mathcal{F}_m$ such that

$$f_n^*(\omega) = f_m(\omega) = \frac{1}{\mathbb{P}(A)} \int_A f \mathrm{d}\mathbb{P}.$$

Of course, such an m may not be unique; in such a case, we take $m(\omega)$ to be smallest possible. For any $A \in \mathtt{At}$, let

$$E(A) = \{\omega \in \Omega : A(\omega) = A\}.$$

By the very definition, we have $E(A) \subseteq A$, the sets $(E(A))_{A\in\mathtt{At}}$ are pairwise disjoint and their union is the full Ω. Furthermore, since we consider the maximal function f_n^* up to time n, we must have $E(A) = \emptyset$ if A is an atom of $\mathcal{F}_k$, $k \geq n+1$. Define an adapted sequence $g = (g_k)_{k\geq 0}$ of random variables given as follows: for any k and any atom A of $\mathcal{F}_k$,

$$g_k|_A = \frac{1}{\mathbb{P}(A)} \sum_{B\subseteq A,\, B\in\mathtt{At}} \mathbb{P}(E(B)).$$

This sequence is a nonnegative supermartingale. Indeed, if $k \geq 0$, A is an atom of $\mathcal{F}_k$ and A_1, A_2, ..., A_m are its children from $\mathcal{F}_{k+1}$, then

$$\begin{aligned}\mathbb{E}(g_{k+1}|A) &= \sum_{j=1}^{m} \frac{\mathbb{P}(A_j)}{\mathbb{P}(A)} \cdot \frac{1}{\mathbb{P}(A_j)} \sum_{B\subseteq A_j,\, B\in \mathtt{At}} \mathbb{P}(E(B)) \\ &= \frac{1}{\mathbb{P}(A)} \sum_{B\subsetneq A,\, B\in\mathtt{At}} \mathbb{P}(E(B)) = g_k|_A - \frac{\mathbb{P}(E(A))}{\mathbb{P}(A)} \leq g_k|_A,\end{aligned}$$

so the supermartingale property is verified. Now, keeping the above notation, we apply (4.14) with

$$\alpha_j = \frac{\mathbb{P}(A_j)}{\mathbb{P}(A)}, \quad x = f_k|_A, \quad y = g_k|_A, \quad x_j = f_{k+1}|A_j, \quad y_j = g_{k+1}|_{A_j}$$

for $j = 1, 2, \ldots, m$. As the result, we get an estimate equivalent to

$$\int_A U(f_k, g_k)\mathrm{d}\mathbb{P} \geq \int_A \mathbb{E}\big[U(f_{k+1}, g_{k+1})|A\big]\mathrm{d}\mathbb{P} + \int_{E(A)} \Phi(f_k)\mathrm{d}\mathbb{P}, \tag{4.15}$$

or

$$\int_A U(f_k, g_k)\mathrm{d}\mathbb{P} \geq \int_A U(f_{k+1}, g_{k+1})\mathrm{d}\mathbb{P} + \int_{E(A)} \Phi(f_n^*)\mathrm{d}\mathbb{P},$$

by the very definition of $E(A)$. Summing over all atoms A of $\mathcal{F}_k$, we get

$$\mathbb{E}U(f_k, g_k) \geq \mathbb{E}U(f_{k+1}, g_{k+1}) + \sum_{A \text{ atom of } \mathcal{F}_k} \int_{E(A)} \Phi(f_n^*)\mathrm{d}\mathbb{P}.$$

Write such inequalities for $k = 0, 1, 2, \ldots, n$ and add them. As the result, we obtain

$$\mathbb{E}U(f_0, g_0) \geq \mathbb{E}U(f_{n+1}, g_{n+1}) - \mathbb{E}\Phi(f_n^*).$$

However, we have $g_0 = 1$, $g_{n+1} = 0$ and $f_{n+1} = f_n$. Therefore, the application of 1° and 2° gives

$$0 \geq -\mathbb{E}\Psi(f_n) + \mathbb{E}\Phi(f_n^*), \tag{4.16}$$

which is the desired estimate. □

REMARK 4.1. If we do not assume the validity of the condition 1°, the above approach yields the estimate

$$\mathbb{E}\Phi(f_n^*) \leq \mathbb{E}\Psi(f_n) + U(\mathbb{E}f_n, 1).$$

This can be used, for example, in the study of the estimates where the controlling term depends only on the first moment of f: see e.g. Exercise 4.7.

Here is the result in the reverse direction.

LEMMA 4.4. *If* (4.12) *is valid, then there is a function* U *satisfying the conditions 1°, 2° and 3°.*

PROOF. For any $x \geq 0$ and $y \in [0, 1]$, set

$$V(x, y) = \sup \left\{ \int_A \Phi(f_n^*)\mathrm{d}\mathbb{P} - \mathbb{E}\Psi(f_n) \right\}, \tag{4.17}$$

where the supremum is taken over all n, all simple nonnegative martingales $f = (f_n)_{n\geq 0}$ starting from x and all measurable sets A with $\mathbb{P}(A) = y$. Then the condition 1° is a direct consequence of (4.12), while 2° follows by considering the constant martingale $f \equiv x$. It remains to establish the third property. Fix the parameters as in the statement and let A^1, A^2, $f^{(1)}$, $f^{(2)}$ be arbitrary sets and simple positive martingales as in the definition of $V(x_1, y_1)$ and $V(x_2, y_2)$, respectively. We may assume that these processes are given on some disjoint probability spaces $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ and $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2)$, equipped with some filtrations $(\mathcal{F}_n^1)_{n\geq 0}$, $(\mathcal{F}_n^2)_{n\geq 0}$. As usual, we splice these two spaces into one: $\Omega = \Omega^1 \cup \Omega^2$, $\mathcal{F} = \sigma(\mathcal{F}^1, \mathcal{F}^2)$, while the probability is given by $\mathbb{P}(A_1 \cup A_2) = \alpha_1 \mathbb{P}(A_1) + \alpha_2 \mathbb{P}^2(A_2)$ if $A_i \in \mathcal{F}^i$, $i = 1, 2$. We also splice the filtration, setting $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(\mathcal{F}_{n-1}^1, \mathcal{F}_{n-1}^2)$. Then the process $f = (f_n)_{n\geq 0}$, given by

$$f_n(\omega) = \begin{cases} x & \text{if } n = 0, \\ f_{n-1}^{(1)}(\omega) & \text{if } \omega \in \Omega^1, \\ f_{n-1}^{(2)}(\omega) & \text{if } \omega \in \Omega^2, \end{cases}$$

is an adapted martingale. Consider any subset $A \in \mathcal{F}$ with $\mathbb{P}(A) = y$, such that $A^1 \cup A^2 \subseteq A$. Then for any $n \geq 1$,

$$\begin{aligned} V(x, y) &\geq \int_A \Phi(f_n^*) \mathrm{d}\mathbb{P} - \mathbb{E}\Psi(f_n) \\ &= \int_{A\setminus(A^1\cup A^2)} \Phi(f_n^*) \mathrm{d}\mathbb{P} + \sum_{i=1}^2 \left[\int_{A^i} \Phi(f_n^*) \mathrm{d}\mathbb{P} - \int_{\Omega^i} \Psi(f_n) \mathrm{d}\mathbb{P} \right] \\ &\geq \mathbb{P}(A \setminus (A^1 \cup A^2)) \cdot \Phi(x) + \sum_{i=1}^2 \alpha_i \left[\int_{A^i} \Phi(f_{n-1}^{(i)*}) \mathrm{d}\mathbb{P}^i - \int_{\Omega^i} \Psi(f_{n-1}^{(i)}) \mathrm{d}\mathbb{P}^i \right]. \end{aligned}$$

Taking the supremum over all $n \geq 1$, all A^i and all $f^{(i)}$ as above, we get the desired claim. □

Thus we have reduced the problem of proving (4.12) to that of finding a Bellman function U satisfying the conditions 1°, 2° and 3°. As usual, a successful treatment of this approach requires the understanding of the concavity-type condition 3°. To gain some intuition about this property, let us assume that the function U is of class C^2. We have already mentioned above that the choice of $y = \alpha_1 y_1 + \alpha_2 y_2$ in (4.13) implies that U must be concave on $\mathbb{R} \times [0, 1]$. Consequently, the Hessian matrix

$$D^2 U = \begin{bmatrix} U_{xx} & U_{xy} \\ U_{xy} & U_{yy} \end{bmatrix}$$

must be nonpositive-definite. The second observation is that if we take $x = x_1 = x_2$ and $y_1 = y_2 = y - \delta \geq 0$, the inequality (4.13) becomes $U(x, y) - U(x, y - \delta) \geq \Phi(x)\delta$ and hence in particular $U_y(x, y) \geq \Phi(x)$. It turns out that these implications can be reversed. Let us state this as a separate lemma.

LEMMA 4.5. *Suppose that U is of class C^2 and its Hessian matrix is nonpositive-definite. If $U_y(x, 1) \geq \Phi(x)$ for any x, then* (4.13) *is satisfied.*

PROOF. Fix the parameters x, y, α_i, x_i and y_i as in the statement of 3°. The assumption on the Hessian matrix implies the concavity of U and hence

$$\alpha_1 U(x_1, y_1) + \alpha_2 U(x_2, y_2) \leq U(x, \alpha_1 y_1 + \alpha_2 y_2).$$

Furthermore, the function $z \mapsto U(x,z)$ is also concave, so $U_y(x,z) \geq U_y(x,1) \geq \Phi(x)$ for all $z \in (0,1)$. Therefore,

$$U(x, \alpha_1 y_1 + \alpha_2 y_2) - U(x,y) = -\int_{\alpha_1 y_1 + \alpha_2 y_2}^{y} U_y(x,z)\mathrm{d}z \leq -\Phi(x)(y - \alpha_1 y_1 - \alpha_2 y_2),$$

and we are done. □

4.5. On solutions to Monge–Ampère equation in planar domains

As we have seen in Lemma 4.5 above, a successful treatment of maximal equations can be reduced to finding a function which satisfies the appropriate differential inequalities. As it is usually the case, when studying the optimal constants involved, one actually searches for those functions for which these differential inequalities actually become equalities. This, in particular, leads us to the question about the structure of the solutions to the equation

$$\det D^2U(x,y) = U_{xx}(x,y)U_{yy}(x,y) - (U_{xy}(x,y))^2 = 0, \tag{4.18}$$

the so-called Monge-Ampère equation. We will establish the following statement.

THEOREM 4.2. *Suppose that D is an open connected subset of $\mathbb{R}^2$ and assume that $U : D \to \mathbb{R}$ is a C^2 function satisfying* (4.18) *such that its Hessian matrix is non-degenerate for each $(x,y) \in D$. Then the domain D can be foliated, i.e., it can be written as a union of pairwise disjoint line segments along which the function U is linear. Furthermore, the partial derivatives of U are constant on the leaves of the foliation.*

PROOF. Fix a point $\mathtt{x} \in D$. For each point (x,y) belonging to some neighborhood $\mathcal{V}$ of $\mathtt{x}$ there is a nonzero vector $v = v(x,y)$ such that $D^2U(x,y)v = 0$; we can assume that v forms a continuous vector field on $\mathcal{V}$. Let $\varphi = (\varphi_1, \varphi_2)$ be the integral curve of this vector field, passing through (x,y): that is, φ is a function defined on some interval I containing 0 such that $\varphi(0) = (x,y)$ and $D^2U(\varphi(t))\varphi'(t) = 0$ for each $t \in I$. Observe that U_x and U_y are constant along φ: we have

$$\frac{d}{dt}U_x(\varphi(t)) = U_{xx}(\varphi(t))\varphi_1'(t) + U_{xy}(\varphi(t))\varphi_2'(t) = (D^2U(\varphi(t))\varphi'(t))_1 = 0 \tag{4.19}$$

and

$$\frac{d}{dt}U_y(\varphi(t)) = U_{xy}(\varphi(t))\varphi_1'(t) + U_{yy}(\varphi(t))\varphi_2'(t) = (D^2U(\varphi(t))\varphi'(t))_2 = 0.$$

Consequently, there is a function F such that $U_y(x,y) = F(U_x(x,y))$. A direct differentiation yields

$$U_{xy}(x,y) = F'(U_x(x,y))U_{xx}(x,y),$$

which gives that on any integral curve of v the ratio U_{xy}/U_{xx} is constant. Hence, by (4.19), the integral curves must be line segments. A similar calculation shows that the function $(x,y) \mapsto U(x,y) - xU_x(x,y) - yU_y(x,y)$ is also constant along φ. Indeed, if $\xi(t) = U(\varphi(t)) - \varphi_1(t)U_x(\varphi(t)) - \varphi_2(t)U_y(\varphi(t))$, then we easily check that

$$\xi'(t) = -\varphi_1(t) \cdot \frac{d}{dt}U_x(\varphi(t)) - \varphi_2(t) \cdot \frac{d}{dt}U_y(\varphi(t)) = 0.$$

Thus we have proved that

$$U(x,y) = U_x(x,y)x + U_y(x,y)y + c(x,y),$$

where c is some function which is constant on any integral curve of the vector field v. This is precisely the desired linearity property of U. □

4.6. Examples

4.6.1. Doob's inequality revisited. Fix $1 < p < \infty$ and let us see how the above approach yields the best constant C_p in the estimate

$$\mathbb{E}(f_n^*)^p \leq C_p^p \mathbb{E}(f_n)^p,$$

under the assumption that f is a nonnegative, simple martingale. We have $\Phi(x) = x^p$ and $\Psi(x) = C_p^p x^p$, and we need to find a function U on $[0, \infty) \times [0, 1]$ satisfying the conditions 1°, 2° and 3°. In this case, we will not need to exploit the structural properties of the solutions to Monge-Ampère equation studied in the preceding section. Instead, directly from the proof of Lemma 4.4, we infer that we may search for U of the form

$$U(x, y) = x^p \varphi(y),$$

for some $\varphi : [0, 1] \to \mathbb{R}$ to be found. Assuming that U is of class C^2, we apply Lemma 4.5. The inequality $U_y(x, 1) \geq \Phi(x)$ becomes

$$\varphi'(1) \geq 1. \tag{4.20}$$

Furthermore, the Hessian matrix is

$$\begin{bmatrix} p(p-1)x^{p-2}\varphi(y) & px^{p-1}\varphi'(y) \\ px^{p-1}\varphi'(y) & x^p\varphi''(y) \end{bmatrix},$$

so, by Sylvester's criterion, we need φ to be concave and satisfy

$$(p-1)\varphi\varphi'' \geq p(\varphi')^2.$$

It is natural to expect that we actually have *equality* here: solving this equation, we get

$$\varphi(y) = \frac{\alpha}{(y+\beta)^{p-1}},$$

for some $\alpha \in \mathbb{R}$, $\beta > 0$ to be found. Then the inequality (4.20) gives

$$-(p-1)\alpha(1+\beta)^{-p} \geq 1,$$

while the initial and the majorization conditions 1°, 2° yield

$$\alpha(1+\beta)^{1-p} = \varphi(1) \leq 0, \qquad \alpha\beta^{1-p} = \varphi(0) \geq -C_p^p.$$

Combining these properties, we obtain

$$C_p^p \geq -\alpha\beta^{1-p} \geq \frac{(\beta+1)^p}{(p-1)\beta^{p-1}}.$$

The right-hand side, considered as a function of $\beta > 0$ attains its minimal value $(p/(p-1))^p$ for $\beta = p-1$. Thus, if we take $\alpha = -p^p/(p-1)$ and $\beta = p-1$, then the above φ leads to the Bellman function

$$U(x, y) = -\frac{p^p x^p}{(p-1)(y+p-1)^{p-1}},$$

which gives Doob's estimate with the constant $C_p = p/(p-1)$.

It is very instructive to study how to extract extremal examples from the above function. The starting point is the proof of Lemma 4.3. If we want to obtain an (almost) equality in (4.16), we need to make sure that all the intermediate inequalities actually become (almost) equalities. The key property of the sequence

(f_n, g_n) considered there is that it must evolve along the "linearity regions" of U: this will guarantee that in (4.15) both sides will be (almost) equal. Let us be more precise here: we will construct an appropriate filtration and an adapted martingale on the probability space $([0,1], \mathcal{F}, |\cdot|)$. We study the sharpness of the estimate

$$\mathbb{E}(f_n^*)^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}f_n^p,$$

in which $(f_n^*)^p$ is integrated over Ω, the set of full measure; thus we take $g_0 \equiv 1$. Since $U(f_0, g_0) = U(f_0, 1) \leq 0$ no matter what f_0 is, we can take f_0 to be an arbitrary number; let us set $f_0 \equiv 1$, for example. To construct the further variables f_m and g_m, we must split the space appropriately; the information about this splitting is encoded in the function U. The general rule is the following: suppose that we have successfully constructed the variables f_k and g_k, and we aim at constructing f_{k+1} and g_{k+1} at some atom A of $\mathcal{F}_k$. Let $x = f_k|_A$ and $y = g_k|_A$.

Step 1. The linearity paths of U. First let us find parameters α_1, α_2, x_1, x_2, y_1, y_2 for which both sides of (4.13) are equal, or almost equal. If $y = 0$, then we set $f_m = f_k$ and $g_m = 0$ for all $m \geq k$, and the construction is over. So, suppose that $y > 0$. Observe that U is linear along any line segment of the form $I_a = \{(s,t) : s = a(t+p-1)\}$, where a is an arbitrary positive parameter. So, if

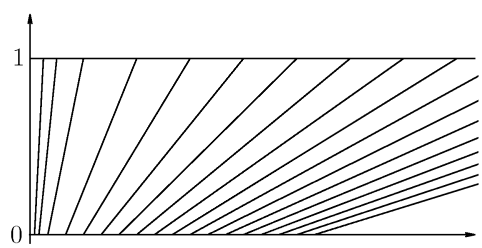

FIGURE 4.2. The strip $[0,\infty) \times [0,1]$ is "foliated", i.e., split into the line segments $I_a = \{(s,t) : s = a(t+p-1),\ t \in [0,1]\}$, $a \geq 0$, along which U is linear.

$y < 1$, then we can take for (x_1, y_1) and (x_2, y_2) the endpoints of the line segment I_a passing through (x, y), and for α_1, α_2 the weights for which

$$\alpha_1(x_1, y_1) + \alpha_2(x_2, y_2) = (x, y).$$

On contrary, if $y = 1$, then such a maneuver is impossible, since then (x, y) is already the endpoint of some I_a. However, it follows from the above construction that $U_y(x, 1) = x^p$, so if we take $x_1 = x_2 = x$, $y_1 = y_2 = 1 - \delta$ for some small positive δ and $\alpha_1 = 1$, $\alpha_2 = 0$, then

$$U(x, 1) \approx U(x, 1-\delta) + x^p\delta.$$

Here we make an error of size $o(\delta)$; the sum of such terms will be irrelevant as it will go to 0 when we let $\delta \to 0$. See below.

Step 2. Splitting. Having found the parameters x_1, x_2, y_1, y_2 and α_1, α_2, we split $A = A_1 \cup A_2$ so that $\mathbb{P}(A_i) = \alpha_i \mathbb{P}(A)$ (if $\alpha_2 = 0$, then take $A_2 = \emptyset$). Next, set $f_{k+1}|_{A_1} = x_1$, $f_{k+1}|_{A_2} = x_2$, $g_{k+1}|_{A_1} = y_1$ and $g_{k+1}|_{A_2} = y_2$. After performing this construction for all $A \in \mathcal{F}_k$, we obtain the variables f_{k+1}, g_{k+1} and the σ-algebra $\mathcal{F}_{k+1}$ generated by them.

Summarizing, the construction is the following. We start from the point $(1,0)$ and then move along the leaves of the foliation so that the moves are of martingale type. Furthermore, the points of the form $(x,0)$ are absorbing - the evolution stops here; the points of the form $(x,1)$ lead to the point $(x,1-\delta)$ (where δ is a small parameter sent eventually to 0). Then the x-variable of the obtained process is precisely the extremal or almost extremal martingale we search for.

This construction can be expressed in a very convenient form in the Markovian language. The process (f,g) is a Markov process on the state space $[0,\infty)\times[0,1]$ with a distribution uniquely determined by the following conditions.

(i) We have $(f_0,g_0)\equiv(1,1)$.
(ii) Any state of the form $(x,1)$ leads to $(x,1-\delta)$.
(iii) Any state of the form (x,y) with $y\in(0,1)$ leads to the endpoints of the line segment I_a passing through (x,y), i.e., to one of the points

$$\left(\frac{(p-1)x}{y+p-1},0\right),\quad \left(\frac{px}{y+p-1},1\right)$$

with probabilities $1-y$ and y, respectively.
(iv) The states of the form $(x,0)$ are absorbing.

Step 3. Final calculation. Now, directly from the above construction, we see that the process (f_n) evolves only at even moves. It starts from 1 and then it jumps (at time 2) to $(p-1)/(p-\delta)$ or to $p/(p-\delta)$. If the first possibility occurs, then the process stops (and then $f^*=1$). In the second case, the evolution continues: the process goes (at time 4) to $\frac{p}{p-\delta}\cdot\frac{p-1}{p-\delta}$ (and stays there forever; then $f^*=p/(p-\delta)$), or goes to $(p/(p-\delta))^2$, and the pattern of the movement is repeated. Summarizing, we see that the random variable f_{2n} takes values in the set

$$\left\{\frac{p-1}{p-\delta},\frac{p}{p-\delta}\cdot\frac{p-1}{p-\delta},\left(\frac{p}{p-\delta}\right)^2\frac{p-1}{p-\delta},\dots,\left(\frac{p}{p-\delta}\right)^{n-1}\frac{p-1}{p-\delta},\left(\frac{p}{p-\delta}\right)^n\right\}$$

and

$$\mathbb{P}\left(f_{2n}=\left(\frac{p}{p-\delta}\right)^k\frac{p-1}{p-\delta}\right)=(1-\delta)^k\delta,\qquad k=0,\,1,\,2,\,\dots,\,n-1.$$

Furthermore, we have $f_{2n}^*=\left(\frac{p}{p-\delta}\right)^k=\frac{p-\delta}{p-1}f_{2n}$ on the set where $f_{2n}=\left(\frac{p}{p-\delta}\right)^k\frac{p-1}{p-\delta}$, $k=0,\,1,\,2,\,\dots,\,n-1$, and $f_{2n}^*=\left(\frac{p}{p-\delta}\right)^n$ when $f_{2n}=\left(\frac{p}{p-\delta}\right)^n$. Consequently,

$$\begin{aligned}\mathbb{E}(f_{2n}^*)^p&\geq\sum_{k=0}^{n-1}\left(\frac{p}{p-\delta}\right)^{pk}(1-\delta)^k\delta\\&=\left(\frac{p-\delta}{p-1}\right)^p\mathbb{E}f_{2n}^p-\left(\frac{p-\delta}{p-1}\right)^p\left(\frac{p}{p-\delta}\right)^{pn}(1-\delta)^n.\end{aligned}\tag{4.21}$$

We easily check that the function $f(\delta)=\left(\frac{p}{p-\delta}\right)^p(1-\delta)$ satisfies $f'(0)=0$ and $f''(0)=(1-p)/p<0$, so it is negative for arguments sufficiently close to 0. Pick such a δ and let $n\to\infty$ to see that the last term in (4.21) tends to 0 as $n\to\infty$. Since $\mathbb{E}f_{2n}^p\geq\mathbb{E}f_0^p=1$, we see that the ratio $||f_{2n}^*||_p/||f_{2n}||_p$ can be made larger than $(p-\delta)/(p-1)-\delta$ if n is sufficiently big. Since δ can be as small as we wish, the constant $p/(p-1)$ is indeed the best.

4.6.2. Weak-type estimates revisited. Suppose that $1 \leq p < \infty$ and let us try to apply the approach to the study of

$$\mathbb{P}(f_n^* \geq 1) \leq c_p^p \mathbb{E} f_n^p, \qquad n = 0, 1, 2, \ldots,$$

where, as usual, we assume that $f = (f_n)_{n\geq 0}$ is a simple and nonnegative martingale. We will actually restrict ourselves to the case $p > 1$: the boundary case $p = 1$ follows from a simple limiting argument.

The inequality is of the form (4.12), with $\Phi(x) = 1_{\{x\geq 1\}}$ and $\Psi(x) = c^p x^p$, and hence we need to construct a function U which satisfies the conditions 1°, 2° and 3°. The condition 2° gives $U(x, 0) \geq -cx^p$. Let us now try to exploit the third condition. To this end, we will apply Lemma 4.5: *assume* for a moment that U is of class C^2 (as we shall see, the function we will end up with will not have this property). The concavity condition will be satisfied if the Hessian matrix of U is nonpositive-definite and $U_y(x, 1) \geq 1_{\{x\geq 1\}}$. Let us suppose that both these requirements hold true with equality, so

$$\det D^2 U = 0 \qquad \text{on } (0,\infty) \times (0, 1), \tag{4.22}$$

$$U_y(x, 1) = 1_{\{x\geq 1\}} \qquad \text{for } x \geq 0. \tag{4.23}$$

We will also assume that the partial derivatives U_x and U_y can be extended to continuous functions on $([0,\infty) \times [0, 1]) \setminus \{1, 1\}$ (the reason for the exclusion of the point $(1, 1)$ is evident, see (4.23)). By the argumentation from the previous section, the set $[0,\infty) \times [0, 1]$ can be foliated, i.e., written as a union of pairwise disjoint line segments, along which U is linear. Furthermore, the partial derivatives of U are constant along the line segments of the foliation. Let us now split the further reasoning into a few intermediate steps.

Step 1. Special parts of the domain. Directly from the abstract formula (4.17), we see that we can take $U(x, y) = y - c_p^p x^p$ when $x \geq 1$, since this is true for the optimal Bellman function V. A similar argument shows that we may assume that $U(0, y) = 0$ for all $y \in [0, 1]$ and $U(x, 0) = -c_p^p x^p$ for all $x \geq 0$. Thus, the set $(\{0\} \times [1,\infty)) \times [0, 1]$ is foliated into the line segments $(\{x\} \times [0, 1])_{x\in\{0\}\times[1,\infty)}$ and obviously, along each such segment, the function U is linear and the partial derivatives of U are constant. Note that by 1°, we must have $1 - c_p^p = U(1, 1) \leq 0$, so $c_p \geq 1$.

Step 2. The case $c_p = 1$. Under this assumption, we have $U(1, 1) = 0$ and hence, by concavity of U, the equality $U(x, 1) = 0$ must hold for all $x \in (0, 1)$ as well. Take such an x and consider the line segment I of the foliation having $(x, 1)$ as one of its endpoints. Then the second endpoint of I cannot be of the form $(x', 0)$ for some $x' > 0$. Indeed, by Theorem 4.2, we have $U_y(\mathtt{x}) = 0$ for any point $\mathtt{x}$ lying on the left of I. This in particular would mean that the function $y \mapsto U(x', y)$ would be constant: a contradiction, since $U(x', 0) = -(x')^p$ and $U(x', 1) = 0$.

Thus, the second endpoint of I is of the form $(0, y)$, $y \in [0, 1]$. Take the leaf of the foliation starting from $(0, 0)$; it follows from the above argumentation that the second endpoint must be equal to $(1, 1)$. This, in turn, implies that the remaining leaves of the foliation must be the collection of all line segments joining $(1, 1)$ with some $(x', 0)$.

Having identified the foliation, we immediately find the formula for U. Indeed, U vanishes on $\{0\} \times [0,1]$ and $[0,1] \times \{1\}$, which implies that U is zero on the triangle with vertices $(0,0)$, $(1,1)$, $(0,1)$. Furthermore, if $0 \leq y < x < 1$, then by the above considerations,

$$U(x,y) = (1-y)U\left(\frac{x-y}{1-y}, 0\right) + yU(1,1) = -(x-y)^p(1-y)^{1-p}.$$

Summarizing, we have discovered the Bellman function

$$U(x,y) = \begin{cases} 0 & \text{if } 0 \leq x < y \leq 1, \\ -(x-y)^p(1-y)^{1-p} & \text{if } 0 \leq y < x < 1, \\ y - x^p & \text{if } x \geq 1. \end{cases}$$

The conditions 1° and 2° hold obviously. To check 3°, note that U is concave on each of the sets $\{(x,y) : 0 \leq x \leq y \leq 1\}$, $\{(x,y) : 0 < y < x < 1\}$ and $[1,\infty) \times [0,1]$ (the claim for the second set follows from the observation that $\det D^2U = 0$ and $U_{xx} \leq 0$ there). However, U is continuous on $[0,\infty) \times [0,1]$ and of class C^1 in the interior of this set, so it is concave on its full domain. Finally, we have $U_x(x,1) = 1_{\{x \geq 1\}}$ and hence, repeating the proof of Lemma 4.5, the property 3° is satisfied. This proves the weak-type bound with the constant 1. This estimate is sharp, which can be easily seen by considering the constant martingale $f \equiv 1$.

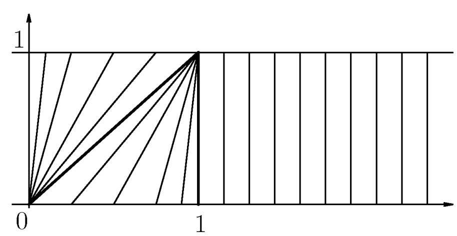

FIGURE 4.3. The foliation in the case $c_p = 1$.

Step 3. The case $c_p > 1$. Though the reasoning presented in the previous step fully answers the question about the best constant in the weak-type estimate, it is instructive to look at the "shape" of the Bellman function in the case of bigger c_p. The first observation is that in such a case there is no line segment of the foliation which would join points of the from $(x,1)$ and $(0,y)$. Suppose on contrary that such a leaf exists and let $I_{x'}$ be a line segment of the foliation starting from $(0,0)$: it must end at some point $(x',1)$ with $x' \in (x,1]$. Then U must vanish on the triangle with vertices $(0,0)$, $(x',1)$ and $(0,1)$. Indeed, by (4.23), the partial derivative U_y is zero in the triangle. Furthermore, we have $U_x \geq 0$ on $\{0\} \times [0,1]$, since otherwise we would have $U(x,y) < -c_p^p x^p$ for some (x,y) from the triangle. Thus, again by the fact that the partial derivatives are constant along the leaves of foliation, we have $U_x \geq 0$ on the triangle and hence $U(x,1) \geq 0$ for each $x \in [0,x']$. Applying 1°, we get that we actually have equality here, which implies that U_x vanishes on $\{0\} \times [0,1]$ and hence U is zero on the triangle. This, in turn, gives that $x' < 1$, since otherwise the concavity along the horizontal half-line $[0,\infty) \times \{1\}$ would be violated. So, there is a line segment $I_{x''}$ starting from some point $(x'',0)$, $0 < x'' < x'$, and ending at some point $(x''',1)$ with $x''' \in (x',1)$. But U_y is constant along the leaves of the foliation, so U_y is zero on the part of $[0,1]\times[0,1]$ contained between the segments $I_{x'}$ and $I_{x''}$. This is a contradiction: we would have $U(x'',y)=-c_p^p(x'')^p$ for small y and $U(x'',y)=0$ for large y.

Step 4. The case $c_p>1$, continuation. In the previous step we have shown that for each $(x,1)$, $x\in(0,1)$, there is a point $x_0=x_0(x)\in(0,1)$ such that the leaf starting from $(x,1)$ ends at $(x_0,0)$. Let $x_1=\sup_{x\in(0,1)}x_0(x)$. Then, by a similar argument to that used above, we have $U_y=0$ on $[0,x_1)\times[0,1]$ and hence $U(x,y)=-c_p^px^p$ there. This, in turn, implies $x_1<1$, since otherwise the concavity of U would not be true. Consequently, the set $[0,x_1]\times[0,1]$ is foliated into vertical line segments $\{x\}\times[0,1]$ which implies that $x_0(x)=x_1$ for all $x\in[x_1,1)$. Therefore, $U_y=0$ and $U_x=-c_p^px_1^{p-1}$ on the triangle with vertices $(x_1,0)$, $(1,1)$ and $(x_1,1)$ and hence

$$U(x,y)=-c_p^px_1^p-pc_p^px_1^{p-1}(x-x_1)$$

there. However, U must be continuous at $(1,1)$, which enforces the equality

$$(p-1)c_p^px_1^p-pc_p^px_1^{p-1}=1-c_p^p. \tag{4.24}$$

A straightforward analysis shows that the left-hand side, considered as a function of $x_1\in[0,1]$, is decreasing and contains $1-c_p^p$ in its range, so the point x_1 is uniquely determined.

Step 5. Final part of the domain. It is clear that on the triangle with vertices $(x_1,0)$, $(1,0)$, $(1,1)$, the only choice for the foliation is the family of line segments joining $(x,0)$ with $(1,1)$, $x\in(x_1,1)$. Therefore, for any (x,y) from the triangle,

$$U(x,y)=(1-y)U\left(\frac{x-y}{1-y},0\right)+yU(1,1)=-c_p^p(x-y)^p(1-y)^{1-p}.$$

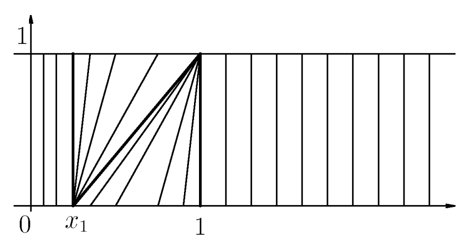

FIGURE 4.4. The foliation in the case $c_p>1$.

So, we have obtained the following candidate for the Bellman function:

$$U(x,y)=\begin{cases}-c_p^px^p & \text{if } x\le x_1,\\ -c_p^px_1^p-pc_p^px_1^{p-1}(x-x_1) & \text{if } x\ge x_1,\ \frac{x-x_1}{1-x_1}\le y\le 1,\\ -c_p^p(x-y)^p(1-y)^{1-p} & \text{if } 0\le y\le \frac{x-x_1}{1-x_1}<1,\\ y-c_p^px^p & \text{if } x\ge 1.\end{cases}$$

As in the case $c_p=1$, we easily check that this function satisfies the conditions 1°, 2° and 3°. Therefore, the application of the Bellman function method yields the estimate

$$\mathbb{P}(f_n^*\ge 1)\le c_p^p\mathbb{E}f_n^p+U(\mathbb{E}f_n,1). \tag{4.25}$$

This bound is sharp, no matter what $\mathbb{E}f_n$ is. Indeed, if $a := \mathbb{E}f_n \in [0, x_1]$ or $a \geq 1$, then the constant martingale $f \equiv a$ gives equality in (4.25). Of course this optimal choice is encoded in the foliation. The point $(\mathbb{E}f_n, 1)$ is the endpoint of the vertical segment and the martingale evolution along this segment is constant in the x-variable (see the similar reasoning described in the study of the L^p estimates). If $a \in (x_1, 1)$, then the foliation illustrated at Figure 4.6.2 suggests introducing the martingale which behaves as follows. It starts from a and then, at each step, it jumps to x_1 and stops ultimately, or moves a little bit to the right, until it reaches 1. More precisely , the distribution of the process is given by

$$\mathbb{P}(f_n = a(1+\delta)^n) = 1 - \mathbb{P}(f_n = x_1) = \frac{a - x_1}{a(1+\delta)^n - x_1}, \qquad n = 0, 1, 2, \ldots, N,$$

where δ and N are chosen so that $a(1+\delta)^N = 1$. This martingale satisfies

$$\mathbb{P}(f_N^* \geq 1) = \frac{a - x_1}{1 - x_1}, \qquad \mathbb{E}f_N^p = x_1^p \cdot \frac{1-a}{1-x_1} + 1 \cdot \frac{a - x_1}{1 - x_1}$$

and hence, by (4.24),

$$\mathbb{P}(f_N^* \geq 1) - c_p^p \mathbb{E}f_n^p = U(a, 1).$$

4.7. Problems

1. For a given $K > 1$, find the best constant $L(K)$ in the logarithmic inequality

$$||f_n^*||_1 \leq K\mathbb{E}f_n \log f_n + L(K), \qquad n = 0, 1, 2, \ldots,$$

where f is assumed to be a nonnegative martingale.

2. For a fixed $1 < p < \infty$, find the best constant in the inequality

$$||f_n^*||_1 \leq C_p ||f_n||_p, \qquad n = 0, 1, 2, \ldots.$$

3. For a fixed $0 < p < 1$, find the best constant C_p in the inequality

$$||f_n^*||_p \leq C_p ||f_n||_1, \qquad n = 0, 1, 2, \ldots.$$

4. (Coifman and Rochberg [**7**]) A nonnegative random variable w is said to be an A_1 weight, if there is a finite constant c such that $w^* \leq cw$ almost surely. The smallest c with this property is called the A_1 characteristics of w and denoted by $[w]_{A_1}$. Prove that if $0 < p < 1$ and f is a nonnegative random variable, then $(f^*)^p$ is an A_1 weight satisfying $[(f^*)^p]_{A_1} \leq (1-p)^{-1}$.

5. (Nikolidakis [**14**]) For a fixed $1 < p < \infty$, find the best constant c_p in the inequality

$$||f_n^*||_{p,\infty} \leq c_p ||f_n||_{p,\infty}, \qquad n = 0, 1, 2, \ldots.$$

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] Robbins, H. *Optimal stopping*, Amer. Math. Mothly **77** (1970), pp. 333-343.

[19] Strzelecki, M. *A note on sharp one-sided bounds for the Hilbert transform*, Proc. Amer. Math. Soc. **144** (2016), pp. 1171–1181.

[20] 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.

[21] Tomaszewski, B. *The best constant in a weak-type H^1-inequality, complex analysis and applications*, Complex Variables, Vol. 4, 1984, pp. 35–38.