A SURVEY: BELLMAN FUNCTION METHOD AND SHARP INEQUALITIES FOR MARTINGALES

ADAM OSĘKOWSKI

Abstract. Bellman function method is an efficient device which enables to relate certain types of estimates arising in probability and harmonic analysis to the existence of the associated special function satisfying appropriate majorization and concavity. This technique has gained considerable interest in the recent years and led to many interesting results concerning the boundedness of wide classes of singular integrals, Fourier multipliers, maximal functions and other related objects. The objective of this survey is to describe the Bellman function approach to certain classical results for martingale transforms. We present the detailed study of the weak-type and moment estimates, and develop some arguments which allow us to simplify and extend the statements, originally proved by Burkholder and others.

Contents

1. Introduction 2
2. Bellman function method 5
2.1. A basic version 5
2.2. An extension 9
2.3. Further extensions 12
3. Weak type $(1, 1)$ inequalities for martingale transforms 15
3.1. A toy example 15
3.2. Related one-sided bound 17
3.3. More exact information on the two-sided bound 22
4. Weak type (p, p) estimates for martingale transforms, $p > 1$ 27
4.1. The case $1 < p \leq 2$ 27
4.2. The case $p > 2$ 29
4.3. One sided bound, $p > 2$ 35
4.4. One sided bound, $1 < p \leq 2$ 36
4.5. More exact information on weak-type bounds 38
5. Strong-type inequalities for martingale transforms 39
5.1. On the search of extremal examples 39
5.2. Basic inequality 40
5.3. Burkholder's function 42
Acknowledgments 43
References 44

1991 *Mathematics Subject Classification.* Primary: 60G42. Secondary: 42A05, 60G46, 49K20.
Key words and phrases. Martingale, Bellman function, best constants.
Research supported by the NCN grant DEC-2012/05/B/ST1/00412.

1. Introduction

Bellman function method is a powerful tool in proving various types of inequalities arising in probability and harmonic analysis. The technique has its origins in the theory of stochastic optimal control, and its fruitful connection with other areas of mathematics was firstly observed by Burkholder in [8], during the study of certain sharp inequalities for martingale transforms. Since then, the approach has been extended and applied essentially in two directions. The first path is probabilistic: Burkholder's arguments from [8] were modified and exploited extensively to investigate numerous estimates for semimartingales. The literature here is quite large, we mention here only the subsequent works of Burkholder [10]-[13], K. P. Choi [14], [15], Suh [44], Wang [51]-[53] and the monograph [33] by the author, which contains the more complete bibliography on the subject. The second path, which pushed the method towards applications in harmonic analysis, started with the seminal paper [29] by Nazarov and Treil (inspired by the preprint version of [30]). This analytic approach has been continued in many papers, including the works of Dragičević and Volberg [20]-[22], Ivanishvili et. al. [23], Melas [26], Melas and Nikolidakis [27], Nazarov, Treil and Volberg [31], Slavin, Stokolos and Vasyunin [41], Slavin and Vasyunin [42], [43], Vasyunin [45], [46], Vasyunin and Volberg [47]-[49] and the author [35]. The results in the aforementioned papers have found many important applications, including tight bounds for wide classes of Fourier multipliers and stochastic integrals: see the papers cited above, consult also Bañuelos, Bielaszewski and Bogdan [1], Bañuelos and Wang [2], Borichev, Janakiraman and Volberg [3], [4], Nazarov and Volberg [32], the author [34], and the works of many other mathematicians.

The purpose of the current work is to present the refined study of some fundamental results for martingale transforms, with the use of both probabilistic and analytic aspects of the Bellman function method. The contents of this survey extends and complements the material contained in the monograph [33]. Our contribution will be of twofold nature. First, this mixed approach will allow us to strengthen some of the classical results: we will obtain a more exact information on the control of a martingale over its ±1-transforms. Second, we will present a certain simplification argument which, in a sense, splits the problem of identifying a given Bellman function into two easier steps. In some cases, this argument reduces significantly the technical difficulties involved in the search of the corresponding Bellman function and, as we hope, can be applied and further extended in many other problems of this type.

We start with some motivation and introduce some basic notation. Let $(h_n)_{n\geq 0}$ be the Haar system on $[0,1]$. Recall that this family of functions is given by

$$\begin{aligned} &h_0 = [0,1), && h_1 = [0,1/2) - [1/2,1),\\ &h_2 = [0,1/4) - [1/4,1/2), && h_3 = [1/2,3/4) - [3/4,1),\\ &h_4 = [0,1/8) - [1/8,1/4), && h_5 = [1/4,3/8) - [3/8,1/2),\\ &h_6 = [1/2,5/8) - [5/8,3/4), && h_7 = [3/4,7/8) - [7/8,1) \end{aligned}$$

and so on (here we have identified a set with its indicator function). As shown by Schauder [40], this collection forms a basis in $L^p(0,1)$ (endowed with Lebesgue measure) for $1 \leq p < \infty$. A classical result of Marcinkiewicz [25] (exploiting the earlier work of Paley [37]) asserts that this basis is unconditional if and only if

$1 < p < \infty$. That is, for any such p there is a finite constant β_p (depending only on p), such that the following holds: for any sequence $a_0, a_1, a_2, \ldots$ of real numbers and any sequence $\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots$ of signs, we have

$$\left|\left|\sum_{k=0}^{n} \varepsilon_k a_k h_k\right|\right|_{L^p(0,1)} \leq \beta_p \left|\left|\sum_{k=0}^{n} a_k h_k\right|\right|_{L^p(0,1)}, \qquad n = 0, 1, 2, \ldots. \tag{1.1}$$

This beautiful property of $(h_n)_{n\geq 0}$ is a starting point for various extensions, which greatly influenced the shape of contemporary mathematics and stimulated the development of many areas, including harmonic analysis, complex analysis, interpolation theory and the geometry of Banach spaces.

We will be interested in the probabilistic version of (1.1), obtained in 1966 by Burkholder [5]. Suppose 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 $f = (f_0, f_1, f_2, \ldots)$ be a real, adapted martingale with difference sequence $df = (df_0, df_1, df_2, \ldots)$. That is,

$$f_n = \sum_{k=0}^{n} df_k, \qquad n = 0, 1, 2, \ldots,$$

where for each k the variable $df_k : \Omega \to \mathbb{R}$ is integrable and $\mathcal{F}_k$-measurable with $\mathbb{E}(df_{k+1}|\mathcal{F}_k) = 0$ (the latter condition is equivalent to saying that for any bounded function $\varphi : \mathbb{R}^{k+1} \to \mathbb{R}$ we have $\mathbb{E}\big[df_{k+1}\varphi(f_0, f_1, \ldots, f_k)\big] = 0$). We say that $g = (g_0, g_1, g_2, \ldots)$ is a ± 1-transform of f, if there is a deterministic sequence ε_0, ε_1, $\varepsilon_2, \ldots$ of signs such that

$$g_n = \sum_{k=0}^{n} \varepsilon_k df_k, \qquad n = 0, 1, 2, \ldots.$$

That is, for any $n \geq 0$ we have $dg_n \equiv df_n$ or $dg_n \equiv -df_n$. Note that the sequence g is also an adapted martingale. The aforementioned result of Burkholder can be stated as follows.

Theorem 1.1. *For any $1 < p < \infty$ there is a constant β_p depending only on p, such that if f is a martingale and g is its ± 1-transform, then*

$$||g_n||_p \leq \beta_p ||f_n||_p, \qquad n = 0, 1, 2, \ldots. \tag{1.2}$$

Here we have used the notation $||f_n||_p = ||f_n||_{L^p(\Omega)}$. Actually, Burkholder proved this strong-type estimate in the more general case when $\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots$ is an arbitrary *predictable* sequence bounded in absolute value by 1. Here by predictability we mean that each ε_k is $\mathcal{F}_{(k-1)\vee 0}$-measurable (and in particular, it may be random). However, in our considerations below, we will be mainly concerned with the case when ε is a deterministic sequence of signs: in most situations, having proved an estimate in this extremal setting, one deduces the more general predictable version with the use of appropriate decomposition theorems (cf. Lemma A.1 in [10]).

Haar system forms a martingale difference sequence on the probability space $([0,1], \mathcal{B}(0,1), |\cdot|)$ (equipped with its natural, dyadic filtration), and hence so does $(a_n h_n)_{n\geq 0}$, for an arbitrary sequence $a_0, a_1, a_2, \ldots$ of real numbers. Consequently, the above statement does generalize (1.1). As a further illustration, let us provide another, closely related example.

Definition 1.2. A system $\{A_{n,i} : i = 1, 2, \ldots, 2^n, n = 0, 1, 2, \ldots\}$ of subsets of $[0,1]$ is called a *dyadic tree*, if for all n and $1 \leq i \leq 2^n$ we have

$$A_{n+1,2i-1} \cap A_{n+1,2i} = \emptyset$$

and

$$A_{n+1,2i-1} \cup A_{n+1,2i} = A_{n,i}.$$

Definition 1.3. Given a dyadic tree of sets satisfying $\mu(A_{n,i}) > 0$ for all n and i, we define the associated *generalized Haar sequence* $h = (h_k)_{k\geq 0}$ by $h_0 = h_{0,1} = \chi_{A_{0,1}}/||\chi_{A_{0,1}}||_1$ and

$$h_{2^{n-1}+i-1} = h_{n,i} = H_{n,i}/||H_{n,i}||_1,$$

where

$$H_{n,i} = \chi_{A_{n,2i-1}}/\mu(A_{n,2i-1}) - \chi_{A_{n,2i}}/\mu(A_{n,2i}), \qquad i \leq 2^n,\ n = 1, 2, \ldots.$$

As in the classical case, one easily verifies that the generalized Haar system $(h_n)_{n\geq 0}$ is a martingale difference sequence with respect to the filtration it generates. Thus, if $(h_n)_{n\geq 0}$ forms a basis of $L^p(0,1)$ (for some $1 < p < \infty$), then (1.2) implies that it is automatically unconditional. This statement is a particular case of a more general fact, which also follows from (1.2), that every monotone basis of $L^p(0,1)$ is unconditional; see Dor and Odell [18] and Pełczyński and Rosenthal [38]. Consult also the first of these papers and the work of Doust [19] for closely related results concerning contractive projections in L^p.

Let us say a few words about the proof of (1.2). In his original approach, Burkholder established first the related weak-type bound

$$\mathbb{P}(|g_n| \geq 1) \leq c||f_n||_1, \qquad n = 0, 1, 2, \ldots, \tag{1.3}$$

for some absolute constant c. Since $||g_n||_2 = ||f_n||_2$ (which is a consequence of the orthogonality of martingale differences), Marcinkiewicz interpolation theorem gives the L^p bound for $1 < p < 2$, and the case $2 < p < \infty$ follows from duality arguments. This reasoning, though simple and very natural, does not produce the best (i.e., the least possible) value of the constant β_p. From the viewpoint of applications (as well as for aesthetic reasons), it is desirable to identify this optimal number for each p. To accomplish this, Burkholder refined his proof and constructed in [8] a certain special object: the Bellman function associated with the inequality (1.2). The careful exploitation of the properties of this function yields that the best β_p in (1.2) equals $p^* - 1$, where $p^* = \max\{p, p/(p-1)\}$. In fact, the paper [8] contains a number of other estimates, proved by the Bellman function approach, including the sharp version of (1.3) as well as the more general sharp weak-type (p,p) estimate for $1 \leq p \leq 2$:

$$\mathbb{P}(|g_n| \geq 1) \leq \frac{2}{\Gamma(p+1)}||f_n||_p^p, \qquad n = 0, 1, 2, \ldots. \tag{1.4}$$

For the review of further results in this direction, we refer the interested reader to the monograph [33] and the references therein.

In what follows, we will mainly focus on the estimates (1.2), (1.3), (1.4) and their various improvements. The remainder of this paper is organized as follows. The next section contains the description of the Bellman function method in the probabilistic setting. In Section 3 we apply the technique to study various versions of weak-type inequalities for martingale transforms. The final part of the paper addresses the strong-type bound and its certain aspects.

Before we proceed, we would like to point out that in general, the guessing of the Bellman function is a difficult task, and in fact, in most cases it is the heart of the matter. Having constructed this special object, the verification that it enjoys all the necessary properties (and hence yields the desired bound) is just a question of some more or less complicated calculations. For the readers' convenience, in most of the estimates studied below, we have decided to present the detailed steps which lead to the discovery of the associated Bellman function. As we hope, this can be helpful during the study of other related problems which arise naturally in the area. On the other hand, to control the size of this survey, we have decided to skip some technicalities, referring instead to the papers where they were originally proved.

2. Bellman function method

The underlying concept of the Bellman function method, both in the probabilistic and analytic version, relates the validity of a certain given inequality to the existence of a certain special function, which possesses appropriate majorization and concavity-type properties. Actually, this special object often carries much more information concerning the problem: see below. The purpose of this section is to present the probabilistic version of the technique. The contents of this part of the survey is a refined version of Chapter 2 from [33], combined with some arguments taken from the works [8], [10], [29] and [49].

2.1. A basic version.

Assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with the filtration $(\mathcal{F}_n)_{n\geq 0}$, a nondecreasing sequence of sub-σ-algebras of $\mathcal{F}$. In what follows, $f = (f_n)_{n\geq 0}$, $g = (g_n)_{n\geq 0}$ will be two adapted martingales taking values in $\mathbb{R}$, with the corresponding difference sequences $(df_n)_{n\geq 0}$, $(dg_n)_{n\geq 0}$, respectively. We may and will assume that $(\mathcal{F}_n)_{n\geq 0}$ is the natural filtration of f and g, i.e., $\mathcal{F}_n = \sigma(f_0, g_0, f_1, g_1, \ldots, f_n, g_n)$ for each $n \geq 0$.

In the previous section we assumed that $dg_n = \varepsilon_n df_n$ for each n, but it will be convenient to work with a wider class of sequences. For any x, $y \in \mathbb{R}$, let $M(x, y)$ denote the class of all pairs (f, g) of adapted martingales satisfying $(f_0, g_0) \equiv (x, y)$ (that is, f starts from x and g starts from y), such that $dg_n \equiv df_n$ or $dg_n \equiv -df_n$ for any $n \geq 1$. Thus, we see that g need not be the ± 1-transform of f, but this can be violated only on the first difference (which happens if and only if $x \neq \pm y$). We stress here that the filtration can vary, as well as the underlying probability space (unless it is nonatomic). For technical reasons, we will assume throughout that f is a *simple* martingale: that is, for any nonnegative integer n the random variable f_n takes a finite number of values and there is a deterministic integer N such that $f_N = f_{N+1} = f_{N+2} = \ldots$. Of course, then g is also simple, and in a typical situation it suffices to deal with a given martingale inequality under such more restrictive assumption (the passage to the general case follows by standard approximation).

Next, let $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function, not necessarily Borel or even measurable. Suppose that we are interested in the numerical value of the associated *Bellman function*

$$\mathcal{B}^0(x, y) = \sup \left\{ \mathbb{E}V(f_n, g_n) \ : \ (f, g) \in M(x, y),\ n = 0,\ 1,\ 2,\ \ldots \right\}. \tag{2.1}$$

Of course, there is no problem with measurability or integrability of $V(f_n, g_n)$, since the sequences f and g are simple.

The relation between the function $\mathcal{B}^0$ and the bounds mentioned in the previous section is evident. For instance, the inequality (1.2) is equivalent to saying that $\mathcal{B}^0(x,\pm x)\leq 0$ for all $x\in\mathbb{R}$, where the underlying function V is given by $V(x,y)=|y|^p-\beta_p^p|x|^p$. The same applies to (1.3), (1.4), and it is clear that a large class of martingale inequalities (encoded with appropriate V) can be deduced from the corresponding upper bounds for $\mathcal{B}^0$. At the first glance it seems that only the values of $\mathcal{B}^0$ at the diagonals $y=\pm x$ are relevant, but this is not the case: the use of a certain inductive argument (see below) requires the knowledge of the values of $\mathcal{B}^0$ on the whole domain $\mathbb{R}\times\mathbb{R}$. As a by-product, if we successfully estimate $\mathcal{B}^0$ on the plane, we get more information about the underlying inequality: we obtain a related result for martingales starting from arbitrary points (but then evolving according to the transforming sequence ε_1, ε_2, $\ldots$).

The key idea during the search of an upper bound for $\mathcal{B}^0$ is to introduce a class of special functions. The class consists of all $B:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ which satisfy the following conditions 1° and 2°:

1° (Majorization property) For all x, $y\in\mathbb{R}$,

$$B(x,y)\geq V(x,y). \tag{2.2}$$

2° (Concavity-type property) For all x, $y\in\mathbb{R}$, $\varepsilon\in\{-1,1\}$ and any $\alpha\in(0,1)$, t_1, $t_2\in\mathbb{R}$ such that $\alpha t_1+(1-\alpha)t_2=0$, we have

$$\alpha B(x+t_1,y+\varepsilon t_1)+(1-\alpha)B(x+t_2,y+\varepsilon t_2)\leq B(x,y). \tag{2.3}$$

By a straightforward induction argument, the condition 2° is equivalent to the following: for any $(x,y)\in\mathbb{R}^2$, any $\varepsilon\in\{-1,1\}$ and any simple mean-zero variable ξ we have

$$\mathbb{E}B(x+\xi,y+\varepsilon\xi)\leq B(x,y). \tag{2.4}$$

That is to say, (2.3) means that the function B is *diagonally concave*, i.e., concave along the lines of slope ±1.

What is the connection between the above class and the Bellman function $\mathcal{B}^0$? The answer is contained in the two statements below, Theorem 2.1 and 2.2.

Theorem 2.1. *Suppose that B satisfies 1° and 2° and let f, g be two simple martingales such that $dg_n\equiv df_n$ or $dg_n\equiv -df_n$ for all $n\geq 1$. Then we have*

$$\mathbb{E}V(f_n,g_n)\leq\mathbb{E}B(f_0,g_0),\qquad n=0,\,1,\,2,\,\ldots. \tag{2.5}$$

In particular, this implies

$$\mathcal{B}^0(x,y)\leq B(x,y)\qquad \textit{for all } x,\ y\in\mathbb{R}. \tag{2.6}$$

Proof. The general fact, which is valid in essentially all versions of the Bellman method, is that the composition of the special function with the underlying processes forms a supermartingale. In our situation, the sequence $(B(f_n,g_n))_{n\geq0}$ has this property. Indeed, all the variables involved are integrable (by simplicity of f and g); furthermore, for any $n\geq1$ we have

$$\mathbb{E}\big[B(f_n,g_n)|\mathcal{F}_{n-1}\big]=\mathbb{E}\big[B(f_{n-1}+df_n,g_{n-1}+dg_n)|\mathcal{F}_{n-1}\big].$$

Applying (2.4) conditionally on $\mathcal{F}_{n-1}$, with $x=f_{n-1}$, $y=g_{n-1}$, $\varepsilon=\varepsilon_n$ and $\xi=df_n$, we get the supermartingale property. Combining this with the majorization 1°, we obtain

$$\mathbb{E}V(f_n,g_n)\leq\mathbb{E}B(f_n,g_n)\leq\mathbb{E}B(f_0,g_0) \tag{2.7}$$

and the proof is complete. □

As a corollary, we see that $\mathcal{B}^0(x,y) \le \inf B(x,y)$, where the infimum is taken over all B satisfying 1° and 2°. The remarkable feature of the method is that the reverse estimate is also valid. Namely, we have the following statement.

Theorem 2.2. *If $\mathcal{B}^0$ is finite, then it is the least function satisfying 1° and 2°.*

Proof. The fact that $\mathcal{B}^0$ satisfies 1° is immediate: the deterministic constant pair (x,y) belongs to $M(x,y)$. To prove 2°, we make use the so called "splicing argument". Take x, y, ε, α, t_1, t_2 as in the statement of the condition. Pick two arbitrary pairs (f^j,g^j) from the class $M(x+t_j,y+\varepsilon t_j)$, $j=1$, 2. We may assume that these pairs are given on the Lebesgue's probability space $([0,1],\mathcal{B}([0,1]),|\cdot|)$, equipped with some filtration. By the simplicity, there is a deterministic integer T such that these pairs terminate before time T. Now we will "glue" these pairs into one using the number α. To be precise, let (f,g) be a pair on $([0,1],\mathcal{B}([0,1]),|\cdot|)$, given by $(f_0,g_0)\equiv(x,y)$,

$$(f_{2n-1},g_{2n-1})(\omega)=\begin{cases}(f^1_{n-1},g^1_{n-1})(\omega/\alpha) & \text{if } \omega\in[0,\alpha)\,,\\ (f_{2n-2},g_{2n-2})(\omega) & \text{if } \omega\in[\alpha,1)\end{cases}$$

and

$$(f_{2n},g_{2n})(\omega)=\begin{cases}(f_{2n-1},g_{2n-1})(\omega) & \text{if } \omega\in[0,\alpha),\\ (f^2_{n-1},g^2_{n-1})\left(\frac{\omega-\alpha}{1-\alpha}\right) & \text{if } \omega\in[\alpha,1)\,,\end{cases}$$

when $n=1,2,\ldots,T$. Finally, we let $df_n=dg_n\equiv 0$ for $n>2T$. Then it is straightforward to check that f, g are martingales with respect to natural filtration and $(f,g)\in M(x,y)$. Therefore, by the very definition of $\mathcal{B}^0$,

$$\begin{aligned}\mathcal{B}^0(x,y)&\ge \mathbb{E}V(f_{2T},g_{2T})\\ &=\int_0^\alpha V(f^1_{T-1},g^1_{T-1})\left(\frac{\omega}{\alpha}\right)\mathrm{d}\omega+\int_\alpha^1 V(f^2_{T-1},g^2_{T-1})\left(\frac{\omega-\alpha}{1-\alpha}\right)\mathrm{d}\omega\\ &=\alpha\mathbb{E}V(f^1_{T-1},g^1_{T-1})+(1-\alpha)\mathbb{E}V(f^2_{T-1},g^2_{T-1}).\end{aligned}$$

Taking supremum over the pairs (f^1,g^1) and (f^2,g^2) gives

$$\mathcal{B}^0(x,y)\ge\alpha\mathcal{B}^0(x+t_1,y+\varepsilon t_1)+(1-\alpha)\mathcal{B}^0(x+t_2,y+\varepsilon t_2),$$

which is 2°. To see that $\mathcal{B}^0$ is the *least* special function, simply apply (2.6). □

The above two facts give the following general method of proving inequalities for ±1-transforms. Let $V:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ be a given function and suppose we are interested in showing that

$$\mathbb{E}V(f_n,g_n)\le 0,\qquad n=0,1,2,\ldots, \tag{2.8}$$

for all simple f, g, such that $dg_n\equiv df_n$ or $dg_n\equiv -df_n$ for all n (in particular, also for $n=0$).

Theorem 2.3. *The inequality* (2.8) *is valid if and only if there exists $B:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ satisfying 1°, 2° and the* initial condition

3° $B(x,\pm x)\le 0$ *for all $x\in\mathbb{R}$.*

Proof. If there is a function B satisfying 1°, 2° and 3°, then (2.8) follows immediately from (2.5), since 3° guarantees that the term $\mathbb{E}B(f_0, g_0)$ is nonpositive. To get the reverse implication, we use Theorem 2.2: as we know from its proof, the function $\mathcal{B}^0$ satisfies 1° and 2°. It also enjoys 3°, directly from the definition of $\mathcal{B}^0$ combined with the inequality (2.8). The only thing which needs to be checked is the finiteness of $\mathcal{B}^0$, which is assumed in Theorem 2.2. Since $\mathcal{B}^0 \geq V > -\infty$, we only need to show the upper bound $\mathcal{B}^0(x, y) < \infty$ for every (x, y). The condition 3°, which we have already established, guarantees the inequality on the diagonals $y = \pm x$. Suppose that $|x| \neq |y|$ and let (f, g) be any pair from $M(x, y)$. Consider another martingale pair (f', g'), which starts from $((x+y)/2, (x+y)/2)$ and, in the first step, moves to (x, y) or to (y, x). If it jumped to (y, x), it stops; otherwise, we determine (f', g') by the assumption that the conditional distribution of $(f'_n, g'_n)_{n\geq 1}$ with respect to $\mathcal{F}_1$ coincides with the (unconditional) distribution of $(f_n, g_n)_{n\geq 0}$. We easily check that g' is a ± 1-transform of f', and hence, for any $n \geq 1$,

$$0 \geq \mathbb{E}V(f'_n, g'_n) = \frac{1}{2}V(y, x) + \frac{1}{2}\mathbb{E}V(f_{n-1}, g_{n-1}).$$

Consequently, taking supremum over f, g and n gives $\mathcal{B}^0(x, y) \leq -V(y, x)$ and we are done. □

Let us say a few words about the notation we plan to use throughout; we hope that the reader will find it helpful. Typically, the superscript "0" will be reserved for the "theoretical" Bellman functions as in (2.1): that is, for such objects, we will use the symbols $\mathcal{B}^0$, b^0, b^0_c, and so on. On the other hand, the lack of this superscript (e.g., B, b, b_c) will indicate that we work with the corresponding candidates.

We conclude this subsection by providing several important comments and observations.

(a) Suppose we want to show the inequality (2.8) for some given V. As we have already proved, if this estimate holds true, it can be established with the use of Theorem 2.3 above. A very natural question arises: is the special function B unique (i.e., does it necessarily equals $\mathcal{B}^0$)? The answer in general is no, and in some situations the choice of the right function does simplify the calculations involved. On the other hand, we would like to repeat (and stress) here that the knowledge of $\mathcal{B}^0$ is desirable: the discovery of this function brings much more information about the estimate (2.8).

(b) As previously, suppose we are given a function V and we want to prove (2.8). A natural idea in the search of the corresponding special function is to take a look at the definition (2.1). This formula shows that $\mathcal{B}^0$ inherits some types of properties from V. For instance, if V enjoys the symmetry property

$$V(x, y) = V(-x, y) \qquad \text{for all } x,\ y \in \mathbb{R}, \tag{2.9}$$

then so does $\mathcal{B}^0$. Indeed, we have $(f, g) \in M(x, y)$ if and only if $(-f, g) \in M(-x, y)$, so

$$\begin{aligned} \mathcal{B}^0(-x, y) &= \sup\left\{\mathbb{E}V(f_n, g_n)\ :\ (f, g) \in M(-x, y),\ n = 0,\ 1,\ 2,\ \dots\right\} \\ &= \sup\left\{\mathbb{E}V(-f_n, g_n)\ :\ (-f, g) \in M(x, y),\ n = 0,\ 1,\ 2,\ \dots\right\} = \mathcal{B}^0(x, y). \end{aligned}$$

Analogously, if V is homogeneous of order p, then the same is true for $\mathcal{B}^0$. This can be shown as above, with the use of the fact that $(f, g) \in M(x, y)$ if and only if $(\lambda f, \lambda g) = M(\lambda x, \lambda y)$ for all $\lambda \neq 0$. In other words, if V has a property of the above type, then we may search for Bellman function in the class of all functions which share this property.

(c) The above method concerns real-valued martingales f and g. 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 (2.8) for nonnegative f (but g *may* take negative values). Then all the above arguments remain valid (only some minor straightforward modifications are required). Namely, one needs to construct appropriate special functions on $[0,\infty)\times\mathbb{R}$ and the parameters x, t_i appearing in 2° must be assumed to satisfy $x+t_i\geq 0$. We leave the necessary changes to the reader. Analogously, one extends the method so that it worked for Hilbert or Banach-space valued sequences: see Subsection 2.3 below, and [10], [33] for the more detailed exposition.

(d) There is an iterative procedure which may be helpful in some situations, as it provides some approximation for Bellman function. This type of reasoning has its roots in the theory of moments: for the description of this theory, see e.g. Kemperman [24] and Cox [16], [17]. Suppose that $V:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is given and fixed, and we aim at solving (2.1). Consider the sequence $(V_n)_{n\geq 0}$ of real-valued functions on $\mathbb{R}\times\mathbb{R}$, given by $V_0=V$ and, for $n\geq 0$ and $(x,y)\in\mathbb{R}^2$,

$$V_{n+1}(x,y)=\sup\mathbb{E}V_n(x+\xi,y+\varepsilon\xi), \tag{2.10}$$

where the supremum is taken over all $\varepsilon\in\{-1,1\}$ and all two-point centered random variables ξ. This recurrence has a very nice geometrical interpretation: given V_n, let us consider all the intervals of slope ± 1, with endpoints lying on the graph of this function (here by slope ± 1 we mean that the endpoints are of the form $(x,y,V(x,y))$, $(x+t,y\pm t,V(x+t,y\pm t))$ for some x, y, $t\in\mathbb{R}$; in particular, we allow $t=0$). Then the graph of V_{n+1} is the upper boundary of the union of all such intervals. Clearly, the sequence $(V_n)_{n\geq 0}$ is nondecreasing; furthermore, if $\mathcal{B}^0$ is finite, then it coincides with the pointwise limit of $(V_n)_{n\geq 0}$. To prove the latter statement, observe that by a straightforward induction, the equality (2.10) holds for arbitrary simple centered random variable ξ. Therefore we have the following alternative definition of V_n:

$$V_n(x,y)=\sup\big\{\mathbb{E}V(f_n,g_n)\ :\ (f,g)\in M(x,y)\big\},$$

or, to put it in yet another words, V_n is a version of $\mathcal{B}^0$ in which only the martingales of length $n+1$ are considered. This clearly gives the pointwise convergence of $(V_n)_{n\geq 0}$ to $\mathcal{B}^0$. In particular, if the iteration (2.10) stabilizes after a finite number of steps, then the "fixed function" must coincide with $\mathcal{B}^0$.

2.2. **An extension.** Sometimes it is of interest to study the function (2.1) under more restrictive assumptions on the martingales f and g. We will consider this problem given some additional integral-norm bounds on f and/or g. To formulate the statement rigorously, assume that Φ, $\Psi:\mathbb{R}\to\mathbb{R}$ are two fixed functions and let x, y, t, s be four real numbers. We define $M(x,y,s,t)$ as the class of all pairs $(f,g)\in M(x,y)$ such that

$$\mathbb{E}\Phi(f_\infty)=s\quad\text{and}\quad\mathbb{E}\Psi(g_\infty)=t. \tag{2.11}$$

Here and below, f_∞, g_∞ denote the pointwise limits of f and g, which exist because of the simplicity of the martingales. Suppose that the class $M(x,y,s,t)$ is nonempty for all (x,y,s,t). In analogy with the preceding setting, assume that we are interested in the numerical value of

$$\mathcal{B}^0(x,y,s,t)=\sup\big\{\mathbb{E}V(f_n,g_n)\ :\ (f,g)\in M(x,y,s,t),\ n=0,\,1,\,2,\,\ldots\big\}. \tag{2.12}$$

Of course, this problem is more difficult that (2.1) and the solution to it gives much more information about the behavior of the pairs (f,g) (indeed: having successfully identified $\mathcal{B}^0$, we recover the Bellman function from the previous section by taking the supremum over all s and t). As previously, to study (2.12), we introduce a class of special functions. The modification of the method requires some changes in the conditions 1° and 2°, which become

1° (Majorization) For all $(x,y,s,t)\in\mathbb{R}^4$,

$$B(x,y,s,t)\geq V(x,y). \tag{2.13}$$

2° (Concavity) For all $(x,y,s,t)\in\mathbb{R}^4$, $\varepsilon\in\{-1,1\}$ and any $\alpha\in(0,1)$, d_1, d_2, s_1, s_2, t_1, $t_2\in\mathbb{R}$ such that

$$\alpha d_1+(1-\alpha)d_2=0,\ \ \alpha s_1+(1-\alpha)s_2=s,\ \ \alpha t_1+(1-\alpha)t_2=t,$$

we have

$$\alpha B(x+d_1,y+\varepsilon d_1,s_1,t_1)+(1-\alpha)B(x+d_2,y+\varepsilon d_2,s_2,t_2)\leq B(x,y,s,t). \tag{2.14}$$

Sometimes we will also refer to (2.14) as to diagonal concavity of B. Observe that by a simple induction argument, this property implies the following: for any x, y, s, $t\in\mathbb{R}$, $\varepsilon\in\{-1,1\}$ and any simple random variables ξ, S, T satisfying $\mathbb{E}\xi=0$, $\mathbb{E}S=s$ and $\mathbb{E}T=t$,

$$\mathbb{E}B(x+\xi,y+\varepsilon\xi,S,T)\leq B(x,y,s,t). \tag{2.15}$$

We have the following version of Theorems 2.1 and 2.2.

Theorem 2.4. *(i) Suppose that $B:\mathbb{R}^4\to\mathbb{R}$ is a function satisfying 1° and 2°. Then $\mathcal{B}^0\leq B$.*

(ii) If $\mathcal{B}^0$ is finite on $\mathbb{R}^4$, then it is the least function satisfying 1° and 2°.

Proof. (i) Pick $(f,g)\in M(x,y,s,t)$ and consider the auxiliary martingales $S_n=\mathbb{E}(\Phi(f_\infty)|\mathcal{F}_n)$, $T_n=\mathbb{E}(\Psi(g_\infty)|\mathcal{F}_n)$, $n=0,\,1,\,2,\,\ldots$. Then, arguing as in the proof of Theorem 2.1 (i.e., using the conditional version of (2.15)), we show that the sequence $(B(f_n,g_n,S_n,T_n))_{n\geq0}$ is a supermartingale. Therefore, by (2.13),

$$\mathbb{E}V(f_n,g_n)\leq\mathbb{E}B(f_n,g_n,S_n,T_n)\leq\mathbb{E}B(f_0,g_0,S_0,T_0).$$

But $\mathcal{F}_0$ is a trivial σ-field, so $S_0=\mathbb{E}\Phi(f_\infty)=s$, $T_0=\mathbb{E}\Psi(g_\infty)=t$ and hence

$$\mathbb{E}V(f_n,g_n)\leq B(x,y,s,t).$$

Taking the supremum over all $n\geq0$ and all (f,g) from $M(x,y,s,t)$ yields the claim.

(ii) We proceed exactly in the same manner as in the proof of Theorem 2.2. The function $\mathcal{B}^0$ clearly satisfies (2.13): we may always take $n=0$ in its definition. To prove (2.14), pick two pairs $(f^j,g^j)\in M(x+d_j,y+\varepsilon d_j,s_j,t_j)$ and splice them into one pair (f,g) as above. Then f, g are simple martingales starting from x and y, respectively, which satisfy $df_n\equiv dg_n$ or $df_n\equiv dg_n$ for all $n\geq1$. Furthermore, the distribution of (f_∞,g_∞) is an appropriate mixture of (f^1_∞,g^1_∞) and (f^2_∞,g^2_∞), so the inclusion $(f,g)\in M(x,y,s,t)$ is valid. Thus,

$$\mathcal{B}^0(x,y,s,t)\geq\mathbb{E}V(f_\infty,g_\infty)=\alpha\mathbb{E}V(f^1_\infty,g^1_\infty)+(1-\alpha)\mathbb{E}V(f^2_\infty,g^2_\infty),$$

and taking the supremum over all (f^1,g^1), (f^2,g^2) as above completes the proof. □

The above theorem covers only an exemplary situation. The modifications mentioned at the end of the previous subsection are valid (with some obvious alterations). Let us describe here some further changes which will be often used later.

(e) In practice, the class $M(x, y, s, t)$ may be empty for some choices of the parameters (x, y, s, t), and one has to restrict to appropriate subsets of $\mathbb{R}^4$. For instance, if one assumes that Φ, Ψ are nonnegative, then the requirement $(x, y, s, t) \in \mathbb{R}^4$ should be replaced by $(x, y, s, t) \in \mathbb{R}^2 \times \mathbb{R}^2_+$. Another important example concerns the case when Φ, Ψ are convex: then one has to consider the set $\{(x, y, s, t) : s \geq \Phi(x),\ t \geq \Psi(y)\}$.

(f) If we want to impose the restriction (2.11) on one martingale only, the corresponding boundary value problem simplifies to three-dimensional. This is evident: if we study (2.12) for all $(f, g) \in M(x, y)$ such that, say,

$$\mathbb{E}\Phi(f_\infty) = t,$$

then the function $\mathcal{B}^0$ depends only on three variables: x, y and t.

(g) Sometimes it is convenient to work with the modification of (2.11) in which equalities are replaced by inequalities

$$\mathbb{E}\Phi(f_\infty) \leq s, \quad \text{and} \quad \mathbb{E}\Psi(g_\infty) \leq t. \tag{2.16}$$

Then the whole methodology can be applied, since the class $M(x, y, s, t)$ (which this time consists of all $(f, g) \in M(x, y)$ which satisfy (2.16)) is monotone in the sense that it grows when s, t increase. The necessary change in the approach is as follows. In Theorem 2.4, instead of working with B which satisfy 1° and 2°, one considers the class of all functions which enjoy 1°, 2° and the additional property

2°' For any $s' \leq s$, $t' \leq t$ we have $B(x, y, s', t') \leq B(x, y, s, t)$.

Indeed, then the technique works: in the proof of (i), we write

$$\mathbb{E}V(f_n, g_n) \leq \mathbb{E}B(f_0, g_0, S_0, T_0) = B\big(x, y, \mathbb{E}\Phi(f_\infty), \mathbb{E}\Psi(g_\infty)\big) \leq B(x, y, s, t)$$

so the assertion holds true; in the proof of the second half, we observe that $\mathcal{B}^0$ satisfies 2°' directly from its definition and the monotonicity of the class M.

There is a very natural question which we want to address now. Namely, given V, how can we proceed in the search of the corresponding function $\mathcal{B}^0$? The remark (b) mentioned above shows that we can restrict ourselves to symmetric or homogeneous functions if only V enjoys these properties. We will present some further, intuitive observations which may be helpful. In many aspects, the search is similar to that arising in the optimal stopping problems. In that setting, one looks for the least superharmonic majorant (in the sense of an underlying Markov process) of the gain function, which in turn leads to the corresponding free-boundary problem (a convenient reference is [39]). To give some ideas, assume that we are in the setting of Theorem 2.4. The "state space" $\mathbb{R}^4$ can be split into two sets:

$$\mathcal{S} = \{(x, y, s, t) : \mathcal{B}^0(x, y, s, t) = V(x, y)\},$$
$$\mathcal{C} = \{(x, y, s, t) : \mathcal{B}^0(x, y, s, t) > V(x, y)\},$$

which in the theory of the optimal stopping, are called the stopping and the continuation region, respectively. This is due to the following interpretation: during the computation of $\mathcal{B}^0(x, y, s, t)$, on the set $\mathcal{S}$ the best choice is just to take any pair $(f, g) \in M(x, y, s, t)$ and evaluate it at its starting point; so, from the viewpoint of $\mathcal{B}^0$, the pair could just stop at the point (x, y) (its further evolution occurs only due to the assumptions on $M(x,y,s,t)$). On contrary, when $(x,y,s,t) \in \mathcal{C}$, then there is a nontrivial choice of (f,g) and n, that is, there are non-constant sequences which matter in the computation of $\mathcal{B}^0(x,y,s,t)$.

Thus, the problem reduces to finding $\mathcal{C}$ and the restriction of $\mathcal{B}^0$ to $\mathcal{C}$. Since $\mathcal{B}^0$ is *the least* diagonally concave majorant of V, it seems plausible to assume the following. For each $(x,y,s,t) \in \mathcal{C}$ there is a direction along which $\mathcal{B}^0$ is locally linear: otherwise, it would be possible to make $\mathcal{B}^0$ a little smaller. More precisely, for such (x,y,s,t), there are $\varepsilon \in \{-1,1\}$ and $m, n \in \mathbb{R}$ such that

$$d \mapsto \mathcal{B}^0(x+d, y+\varepsilon d, s+md, t+nd)$$

is linear for d lying in some neighborhood of 0. In other words, the whole set $\mathcal{C}$ can be "foliated" into line segments of appropriate slope along which the function $\mathcal{B}^0$ is linear. If $\mathcal{B}^0$ were twice differentiable on $\mathcal{C}$ (this is quite a reasonable expectation: this function is extremal, so it should possess some additional regularity), the latter condition yields the second-order differential equation for $\mathcal{B}^0$. It is more convenient to state it in terms of the"rotated" function

$$\mathcal{M}(x,y,s,t) = \mathcal{B}^0(x+y, x-y, s, t).$$

After this change of variables, we see that the condition on $\mathcal{B}^0$ becomes

the functions $(x,s,t) \mapsto \mathcal{M}(x,y,s,t)$, $(y,s,t) \mapsto \mathcal{M}(x,y,s,t)$ are concave,
and for each point (x,y,s,t) one of them is linear in some direction.

In particular, this enforces the following "system" of Monge-Ampère equations: for each (x,y,s,t) we have

$$\det \begin{bmatrix} \mathcal{M}_{xx} & \mathcal{M}_{xs} & \mathcal{M}_{xt} \\ \mathcal{M}_{sx} & \mathcal{M}_{ss} & \mathcal{M}_{st} \\ \mathcal{M}_{tx} & \mathcal{M}_{ts} & \mathcal{M}_{tt} \end{bmatrix} (x,y,s,t) = 0$$

or

$$\det \begin{bmatrix} \mathcal{M}_{yy} & \mathcal{M}_{ys} & \mathcal{M}_{yt} \\ \mathcal{M}_{sy} & \mathcal{M}_{ss} & \mathcal{M}_{st} \\ \mathcal{M}_{ty} & \mathcal{M}_{ts} & \mathcal{M}_{tt} \end{bmatrix} (x,y,s,t) = 0.$$

Sometimes this system can be explicitly solved: see e.g. [8], [49], [50], and this brings the candidate for the Bellman function. It may not satisfy the assumed regularity (i.e., it need not be of class C^2 on $\mathcal{C}$), but this is not important: having obtained the candidate, one proves rigorously that the function has all the desired properties, and the problem is solved.

There is an alternative approach, which will also be of interest for us below. Namely, from more or less formal arguments one can guess the (approximate) shape of $\mathcal{C}$ and then try to get the formula for $\mathcal{B}^0$ by indicating the appropriate foliation of this set. Though this approach seems difficult and sometimes the reasoning does depend on luck, it has turned out to be very efficient. Actually, having seen several Bellman functions, in some cases the discovery of the new special function is not hard at all: the foliations often share some common features.

2.3. **Further extensions.** This subsection is irrelevant for our further considerations, but we decided to include it here to indicate some further refinements and extensions of the methodology. Our primary goal is to show here how to modify the technique so that it worked for wider class of processes.

Our first comment concerns the values of the transforming sequence. In our previous setting, we assumed that $(\varepsilon_n)_{n\geq 0}$ is deterministic and its terms are ± 1. However, it is easy to adjust the method to the less restrictive case. Namely, suppose that the sequence ε is simple, predictable and takes values in $[-1,1]$; furthermore, allow the martingales to be vector-valued. Then we have the following statement; the proof is the same as in the real-valued setting and hence is omitted.

Theorem 2.5. *Let X be a Banach space and let $V : X \times X \to \mathbb{R}$ be a given function. Consider the estimate*

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

where (f, g) runs over the class of all simple pairs of X-valued martingales such that g is a transform of f by a predictable sequence bounded in absolute value by 1. This inequality holds true if and only if there exists $B : X \times X \to \mathbb{R}$ satisfying the following three conditions.

1° *$B \geq V$ on $X \times X$.*

2° *For all $(x, y) \in X \times X$, any deterministic $a \in [-1, 1]$ and any $\alpha \in (0, 1)$, t_1, $t_2 \in \mathcal{B}$ such that $\alpha t_1 + (1-\alpha)t_2 = 0$ we have*

$$\alpha B(x + t_1, y + at_1) + (1-\alpha)B(x + t_2, y + at_2) \leq B(x, y).$$

3° *$B(x, y) \leq 0$ for all x, $y \in X$ such that $y = ax$ for some $a \in [-1, 1]$.*

Straightforward modifications lead to the similar vector-valued version of Theorem 2.4. The details are left to the reader.

The next extension concerns another very important class of martingale pairs. It is much wider than that considered in the previous two subsections and has many interesting applications. Assume that X is a Banach space with the norm $|\cdot|$.

Definition 2.6. Suppose that f, g are martingales taking values in X. Then g is *differentially subordinate* to f, if for any $n = 0, 1, 2, \ldots$ we have

$$|dg_n| \leq |df_n|$$

with probability 1.

If g is a transform of f by a predictable sequence bounded in absolute value by 1, then, obviously, g is differentially subordinate to f. Another very important example is related to martingale square function. Namely, suppose that f takes values in X and let g be $\ell^2(X)$-valued martingale, whose difference sequence is defined by $dg_n = (0, 0, \ldots, 0, df_n, 0, \ldots)$, $n = 0, 1, 2, \ldots$ (the term df_n appears on the n-th place). If we treat f as an $\ell^2(X)$-valued process, via the embedding $f_n \sim (f_n, 0, 0, \ldots)$, then g is differentially subordinate to f and f is differentially subordinate to g. However,

$$||g_n||_{\ell^2(X)} = \left(\sum_{k=0}^{n} |df_k|^2\right)^{1/2}, \qquad n = 0, 1, 2, \ldots,$$

is the square function of f. Thus, any inequality valid for differentially subordinated martingales with values in $\ell^2(X)$ leads to a corresponding estimate for the square function of a X-valued martingale. This observation has turned out to be particularly efficient when X is a separable Hilbert space (cf. [10]).

Let us formulate the version of Bellman function method when the considered martingales satisfy the differential subordination. Assume that $V : X \times X \to \mathbb{R}$ is a given Borel function and consider the class of all $B : X \times X \to \mathbb{R}$ such that

1° $B(x, y) \geq V(x, y)$ for all x, $y \in X$,

2° there are Borel a, $b : X \times X \to X^*$ such that for any x, $y \in X$ and any h, $k \in X$ with $|k| \leq |h|$, we have

$$B(x + h, y + k) \leq B(x, y) + \langle a(x, y), h\rangle + \langle b(x, y), k\rangle.$$

3° $B(x, y) \leq 0$ for all x, $y \in X$ with $|y| \leq |x|$.

Theorem 2.7. *Suppose that B satisfies 1°, 2° and 3°. Let f, g be $\mathcal{B}$-valued martingales such that g is differentially subordinate to f and*

$$\begin{aligned} &\mathbb{E}|V(f_n, g_n)| < \infty, \quad \mathbb{E}|B(f_n, g_n)| < \infty, \\ &\mathbb{E}\big(|a(f_n, g_n)||df_{n+1}| + |b(f_n, g_n)||dg_{n+1}|\big) < \infty, \end{aligned} \tag{2.17}$$

for all $n = 0, 1, 2, \ldots$. Then

$$\mathbb{E}V(f_n, g_n) \leq 0, \qquad n = 0, 1, 2, \ldots. \tag{2.18}$$

The proof is similar to that above: using 2°, we show that the composition $(B(f_n, g_n))_{n\geq 0}$ is a supermartingale, and then apply 1° and 3° to get

$$\mathbb{E}V(f_n, g_n) \leq \mathbb{E}B(f_n, g_n) \leq \mathbb{E}B(f_0, g_0) \leq 0.$$

The details are left to the reader. Analogously, one can extend Theorem 2.4 to this new setting.

Finally, let us mention that it is possible to extend the Bellman function method to other, less restrictive classes of semimartingales. For instance, one can handle the case in which the transformed sequence $(f_n)_{n\geq 0}$ is a submartingale, i.e., an adapted sequence of integrable variables satisfying $\mathbb{E}(df_{k+1}|\mathcal{F}_k) \geq 0$ for each $k \geq 0$. This setting is maybe not that important for analytic applications, but plays a distinguished role in stochastic analysis. Again, we focus on appropriate version of Theorems 2.1 and 2.2; the modification which leads to the analogue of Theorem 2.4 is left to the reader. For a given $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, let

$$\mathcal{B}^0(x, y) = \sup\big\{\mathbb{E}V(f_n, g_n) \ : \ (f, g) \in S(x, y),\ n = 0, 1, 2, \ldots\big\}.$$

Here $S(x, y)$ denotes the class of all simple pairs (f, g) starting from (x, y) such that f is a submartingale and for each $n \geq 1$ we have $dg_n \equiv df_n$ or $dg_n \equiv -df_n$. To study this object, consider the class which consists of all $B : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfying

1° We have $B \geq V$ on $\mathbb{R}^2$,

2° For any x, $y \in \mathbb{R}$, any $\varepsilon \in \{-1, 1\}$ and any $\alpha \in (0, 1)$, t_1, $t_2 \in \mathbb{R}$ such that $\alpha t_1 + (1 - \alpha)t_2 \geq 0$, we have

$$\alpha B(x + t_1, y + \varepsilon t_1) + (1 - \alpha)B(x + t_2, y + \varepsilon t_2) \leq B(x, y).$$

By induction, 2° implies that

$$\mathbb{E}B(x + \xi, y + \varepsilon\xi) \leq B(x, y)$$

for all x, $y \in \mathbb{R}$, $\varepsilon \in \{-1, 1\}$ and all simple random variables with *nonnegative* expectation. In other words, B is concave and nondecreasing (when going "from left to the right") on each line of slope ± 1.

Repeating the proof of Theorems 2.1 and 2.2, we get the following statement.

Theorem 2.8. *(i) Suppose that $B : \mathbb{R}^2 \to \mathbb{R}$ is a function satisfying 1° and 2°. Then $\mathcal{B}^0 \leq B$.*

(ii) If $\mathcal{B}^0$ is finite on $\mathbb{R}^2$, then it is the least function satisfying 1° and 2°.

For further discussion and many examples, we refer the reader to [11] and to the monograph [33].

3. Weak type $(1,1)$ inequalities for martingale transforms

The purpose of this section is to apply the above methodology in the study of sharp weak-type $(1,1)$ inequalities for martingale transforms.

3.1. **A toy example.** Suppose we are interested in the sharp estimate

$$\mathbb{P}(|g_n| \geq 1) \leq C_1 ||f_n||_1, \qquad n = 0,\, 1,\, 2,\, \ldots,$$

where f is a martingale and g is its ± 1-transform. By standard approximation, we may assume that f is simple and thus the problem can be rewritten in the form (2.8), with $V(x,y) = 1_{\{|y| \geq 1\}} - C_1|x|$. There are two objects to be determined: a priori unknown optimal value of the constant C_1 and an appropriate special function (which may be equal to $\mathcal{B}^0$, but need not: see Remark (a) in §2.1). To gain some intuition about the special function to be found, let us write down the definition (2.1) of $\mathcal{B}^0$:

$$\mathcal{B}^0(x,y) = \sup\left\{\mathbb{P}(|g_n| \geq 1) - C_1\mathbb{E}|f_n| \;:\; (f,g) \in M(x,y),\, n = 0,\, 1,\, 2,\, \ldots\right\}.$$

It is not difficult to determine the formula for $\mathcal{B}^0$, basing on the conditions 1°, 2° and 3°. We will present two different approaches to this problem: the first will be probabilistic, while the second will have analytic flavor.

Probabilistic approach. We will construct $\mathcal{B}^0$ by the very definition: for each (x,y) we will provide the corresponding extremal example. We split the reasoning into three steps.

Step 1. The case $|y| \geq 1$. Under this assumption, we have

$$\mathcal{B}^0(x,y) = 1 - C_1|x|. \tag{3.1}$$

Indeed, if $(f,g) \in M(x,y)$, then $\mathbb{P}(|g_n| \geq 1) \leq 1$ and $\mathbb{E}|f_n| \geq |x|$ for any n. This gives the inequality in one direction; letting $f \equiv x$ and $g \equiv y$ (or using 1°) yields the reverse estimate.

Step 2. The case $|x| + |y| \geq 1$. Next we show that for such (x,y),

$$\mathcal{B}^0(x,y) = 1 - C_1|x|. \tag{3.2}$$

We have already done this for $|y| \geq 1$, so let us assume that $|x| + |y| \geq 1 > |y|$. Furthermore, we may restrict ourselves to nonnegative x and y, since $\mathcal{B}^0(x,y) = \mathcal{B}^0(-x,y) = \mathcal{B}^0(x,-y)$ for all x, y: see Remark (b) in §2.1. Reasoning as in the previous step, we obtain $\mathcal{B}^0(x,y) \leq 1 - C_1|x|$. To get the reverse bound, we need an example. It is not difficult to find it: we must have $\mathbb{P}(|g_n| \geq 1) = 1$ and $\mathbb{E}|f_n| = |x|$ for some n; thus, we want to send g outside $(-1,1)$, and, on the other hand, f cannot change sign (otherwise, $\mathbb{E}|f_n|$ would increase). A little experimentation leads to the following element of $M(x,y)$: we set $(f_0,g_0) \equiv (x,y)$, assume that $df_1 = -dg_1$ is a centered random variable taking values in $\{-x, y+1\}$ (we do not need to specify the probabilities: they are uniquely determined by the requirement that the variable has mean zero) and put $df_n = dg_n \equiv 0$ for $n \geq 2$. Then $f \geq 0$, so $\mathbb{E}|f_n| = x$ for all n; furthermore, g_1 takes values in the set $\{-1, x+y\}$, so $|g_1| \geq 1$

almost surely. This implies the reverse inequality $\mathcal{B}^0(x,y) \geq 1 - C_1|x|$ and yields (3.2).

Step 3. The case $|x|+|y|<1$. How to construct an appropriate example in this case? A little thought and experimentation 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

$$\mathcal{B}^0(x,y) \geq 1 - C_1(1+|x|^2-|y|^2)/2. \tag{3.3}$$

If we apply the initial condition 3°, we get $1 - C_1/2 \leq \mathcal{B}^0(0,0) \leq 0$, or $C_1 \geq 2$. *Assuming* equality here and in (3.3), we obtain that $\mathcal{B}^0(x,y)$ should be equal to $|y|^2 - |x|^2$. Summarizing, the above three steps have led us to the candidate

$$B(x,y) = \begin{cases} |y|^2 - |x|^2 & \text{if } |x|+|y| \leq 1,\\ 1-2|x| & \text{if } |x|+|y|>1.\end{cases} \tag{3.4}$$

Now, one can easily check that this candidate satisfies 1°, 2° and 3°. Consequently, we get $\mathcal{B}^0 \leq B$ and $C_1 \leq 2$. However, we have already proved that $C_1 \geq 2$; hence we actually have $C_1 = 2$. Now the equality $\mathcal{B}^0 = B$ follows directly from Step 2 and (3.3). For a related reasoning, consult [6] and [8].

Analytic approach. Here we will use the iterative procedure described in Remark (d) from §2.1. We start with the function $V_0(x,y) = 1_{\{|y|\geq 1\}} - C_1|x|$, where C_1 is a constant to be found. Then some a bit lengthy, but straightforward calculations show that for $C_1 \leq 2$ we have

$$V_2(x,y) = V_3(x,y) = \ldots = \begin{cases} 1 - C_1|x| & \text{if } |x|+|y| \geq 1,\\ 1 - C_1(1+|x|^2-|y|^2)/2 & \text{if } |x|+|y|<1.\end{cases}$$

For $C_1 > 2$ the situation is more complicated: we will study this case in our later considerations. Therefore, we get that $\mathcal{B}^0 = V_2$ when $C_1 \leq 2$. In the limit case $C_1 = 2$, the function $\mathcal{B}^0$ satisfies 3° and hence 2 is the best constant in the weak-type $(1,1)$ inequality.

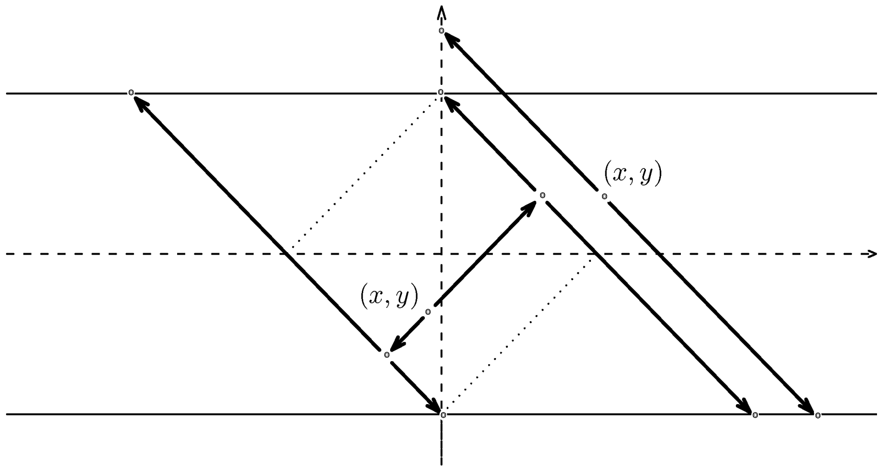

FIGURE 1. Examples arising in the study of $\mathcal{B}^0$. When $|y| \geq 1$, the examples are constant; when $|y| < 1$, they evolve as indicated.

Remark 3.1. There is a very natural question about the vector-valued version of the above weak-type estimate. Though we will not go further in this direction, let us say a few words about this more general setting, as they may shed some light on the differences between the scalar and vector cases. It turns out that if the transformed martingale f takes values in a Hilbert space X, then the best constant is still 2. This can be shown with the use of the Bellman function (3.4), with $|\cdot|$ being the norm of X. However, this function *does not* coincide with $\mathcal{B}^0$ (unless the Hilbert space is one-dimensional)! It can be shown (consult [9]) that

$$\mathcal{B}^0(x,y) = \begin{cases} 1-(1+2(x+y,x-y)+|x+y|^2|x-y|^2)^{1/2} & \text{if } |x+y| \vee |x-y| < 1, \\ 1-2|x| & \text{if } |x+y| \vee |x-y| \geq 1, \end{cases}$$

where $(\cdot,\cdot)$ denotes the scalar product of X. In the case when X is a Banach space which is not isomorphic to a Hilbert space, the situation is even more interesting (and much more difficult). Namely, it can be proved that the best constant for such X is strictly larger than 2, and it is finite if and only if X is the so-called UMD space (where UMD is the abbreviation for Unconditional for Martingale Differences). In this setting, the problem of identification of the weak-type constant becomes very difficult. See [7], [9] or [33] for more information on the subject.

3.2. **Related one-sided bound.** The estimate $\mathbb{P}(|g_n| \geq 1) \leq 2\mathbb{E}|f_n|$ we have just proved has its weaker, one-sided version, which is also of interest. Namely, we have

$$\mathbb{P}(g_n \geq 1) \leq 2\mathbb{E}|f_n|, \qquad n = 0, 1, 2, \ldots, \tag{3.5}$$

and it turns out that 2 is also the best here (modify slightly the above examples; see also [28], [33], [36]). The purpose of this section is to give an explicit formula for the related Bellman function

$$\mathcal{B}^0(x,y,t) = \sup\left\{\mathbb{P}(g_n \geq 1) \;:\; (f,g) \in M(x,y),\, \mathbb{E}|f_\infty| \leq t,\, n = 0, 1, 2, \ldots\right\},$$

defined on the set $\{(x,y,t) : t \geq |x|\}$ (it is obvious that for each (x,y,t) from this set, there is at least one pair (f,g) which satisfies the requirements under the supremum). By the facts presented above, we have $\mathcal{B}^0(x, \pm x, t) \leq 2t$, but for some points the inequality is strict. Thus, the identification of $\mathcal{B}^0$ can be regarded as a stronger, more exact version of (3.5). Furthermore, the formula for $\mathcal{B}^0$ will be useful for us later, when we study analogous Bellman function corresponding to the two-sided bound for g. Due to the appearance of control condition on $\mathbb{E}|f_\infty|$, we are forced to take the more difficult approach described in §2.2.

We present two solutions to this problem. In the first of them we will follow the path described at the end of §2.2: we will exploit a certain homogeneity property of $\mathcal{B}^0$ and point out the appropriate foliation. The second approach is different and is of independent interest, as it can be applied successfully in other Bellman function problems. Namely, we will show that the search for $\mathcal{B}^0$ can be split into two parts: first one searches for the whole family of *simpler* (less dimensional) Bellman functions, and then comes back to the original problem by appropriate optimization argument.

Approach 1. A direct use of Theorem 2.4. It is more convenient to work with

$$\mathbf{B}(x, y, t) = \sup\Big\{\mathbb{P}(g_n \geq 0)\ :\ (f, g) \in M(x, y),\ \mathbb{E}|f_\infty| \leq t,\ n = 0,\ 1,\ 2,\ \ldots\Big\},$$

which is related to $\mathcal{B}^0$ by the identity $\mathcal{B}^0(x, y, t) = \mathbf{B}(x, y-1, t)$ for all $(x, y, t) \in \mathbb{R}^3$ such that $t \geq |x|$. By Theorem 2.4, the function $\mathbf{B}$ is diagonally concave and satisfies the majorization

$$\mathbf{B}(x, y, t) \geq 1_{\{y \geq 0\}}. \tag{3.6}$$

The reason why we have turned to $\mathbf{B}$ is that this function enjoys the homogeneity-type property

$$\mathbf{B}(\pm\alpha x, \alpha y, \alpha t) = \mathbf{B}(x, y, t), \qquad \text{for all } \alpha > 0. \tag{3.7}$$

Indeed, we have $\mathbb{P}(g_n \geq 0) = \mathbb{P}(\alpha g_n \geq 0)$ and the equivalence $(f, g) \in M(x, y)$, $\mathbb{E}|f_\infty| \leq t$ if and only if $(\pm\alpha f, \alpha g) \in M(\alpha x, \alpha y)$, $\mathbb{E}|\alpha f_\infty| \leq \alpha t$.

In particular, (3.7) implies that the function $x \mapsto \mathbf{B}(x, -x, x)$ is constant on $(0, \infty)$. On the other hand, this function is concave on $\mathbb{R}$, in view of the diagonal concavity of $\mathbf{B}$. Hence

$$\mathbf{B}(1/2, -1/2, 1/2) \geq \mathbf{B}(0, 0, 0) = 1, \tag{3.8}$$

where the latter equality follows from (3.6) and the obvious bound $\mathbf{B} \leq 1$. Next, let us introduce the function

$$b(x, y) = \mathbf{B}\left(\frac{x+1}{2}, \frac{x-1}{2}, y\right),$$

defined on the set $D = \big\{(x, y) \in \mathbb{R}^2 : y \geq |\frac{x+1}{2}|\big\}$. Using (3.7), we see that for $x \neq \pm y$,

$$b\left(\frac{x+y}{x-y}, \frac{t}{x-y}\right) = \mathbf{B}(x, y, t) = \mathbf{B}(-x, y, t) = b\left(\frac{x-y}{x+y}, -\frac{t}{x+y}\right),$$

and hence b satisfies

$$b\,(x, y) = b\left(\frac{1}{x}, -\frac{y}{x}\right). \tag{3.9}$$

Furthermore, since $\mathbf{B}$ is diagonally concave, we see that b is a concave function, and the majorization (3.6) implies that $b(x, y) \geq 1_{\{x \geq 1\}} \geq 0$. The condition (3.8) implies that $b(0, 1/2) \geq 1$; hence, using the concavity of b along the halflines starting from $(0,1/2)$ and contained in D, we infer that $b(x,y)\geq 1$ (and hence $b(x,y)=1$) provided $y\geq -x/2+1/2$. Therefore, all that remains is to identify the explicit formula for b on

$$\Omega=\{(x,y)\in D: y\leq -x/2+1/2\}.$$

It is easy to show that $b(-1,0)=\mathbf{B}(0,-1,0)=0$: indeed, the conditions $(f,g)\in M(0,-1)$, $\mathbb{E}|f_\infty|\leq 0$ are satisfied by only one, constant pair $(0,-1)$. The line segment which joins $(-1,0)$ and $(0,1/2)$ is a part of the boundary of Ω, so it seems plausible to guess that b is linear along this segment: $b(2y-1,y)=2y$ for $y\in[0,1/2]$.

Next, we impose some regularity on b: assume that b is of class C^1 in the interior of Ω. By (3.9), we may restrict our search to the triangle $\Omega\cap\{(x,y):x\geq -1\}$. Let us try to identify the foliation $\mathcal{F}$ of b restricted to this set (i.e., split the triangle into the union of maximal segments along which b is linear). We already know that the segment with the endpoints $(0,1/2)$ and $(-1,1)$, as well as the boundary segment with endpoints $(-1,0)$, $(0,1/2)$, belong to the foliation. Now pick a segment $I\in\mathcal{F}$ which contains the point $(-1,y)$ for a given $y\in(0,1)$. If I intersects one of the two boundary segments (call it J), at a point different from $(0,1/2)$, then b must be linear in the triangle spanned by I and J (i.e., the convex hull of $I\cup J$). In particular, this implies that b must be linear along the segment which joins $(-1,y)$ with $(0,1/2)$. Consequently, we see that there is only one reasonable foliation: the fan of segments from the vertex $(0,1/2)$. This implies

$$b(-1,y)-1=-b_x(-1,y)+b_y(-1,y)\left(y-\frac{1}{2}\right).$$

On the other hand, differentiating (3.9) with respect to x at the point $(-1,y)$, $y\in(0,1)$, yields

$$2b_x(-1,y)=yb_y(-1,y).$$

If we combine the two latter identities, we obtain the following differential equation: for $\varphi(y)=b(-1,y)$, $y\in[0,1]$, we have

$$\varphi(y)-1=\varphi'(y)\cdot\frac{y-1}{2}.$$

The solution is $\varphi(y)=K(y-1)^2+1$ for some parameter K. To determine this number, note that $\varphi(0)=\mathbf{B}(0,-1,0)=0$; this gives $K=-1$ and therefore

$$b(x,y)=(1+x)b\left(0,\frac{1}{2}\right)-xb\left(-1,\frac{1+x-2y}{2x}\right)=1-\frac{\left(\frac{x-1}{2}+y\right)^2}{x}$$

for $(x,y)\in\Omega$, $x\in[-1,0]$. By (3.9), the same formula is valid on the whole Ω. This yields the candidate

$$B(x,y,t)=\mathbf{B}(x,y-1,t)=b\left(\frac{x+y-1}{x-y+1},\frac{t}{x-y+1}\right),$$

which is given explicitly by

$$B(x,y,t)=\begin{cases}1 & \text{if } y+t\geq 1,\\ 1-\frac{(1-y-t)^2}{(1-y)^2-x^2} & \text{if } y+t<1.\end{cases}\tag{3.10}$$

We easily check that this function indeed satisfies the conditions 1° and 2°, which implies $\mathcal{B}^0\leq B$. Actually, it can be shown that it coincides with the desired function $\mathcal{B}^0$ (cf. [28]; one can also consult [33] and [36]). An alternative reasoning will be presented below.

Approach 2. Two-step procedure. Now we will present a very simple, yet powerful argument, which in some cases simplifies considerably the whole approach. Let us first note that the reasoning presented above rests on a direct search of a special function of three variables. The key is to reduce the problem to the simpler two-dimensional case described in §2.1. To accomplish this we take the restriction $\mathbb{E}|f_\infty| \le t$ and, in a sense, we move it into the optimized expression. More precisely, we consider a slightly different problem: let

$$b^0(x,y) = \sup\left\{\mathbb{P}(g_n \ge 0) - \mathbb{E}|f_n| \ : \ (f,g) \in M(x,y),\, n = 0,\, 1,\, 2,\, \dots\right\}. \tag{3.11}$$

We already know that b^0 is the least diagonally concave function which majorizes $V : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ given by $V(x,y) = 1_{\{y\ge 0\}} - |x|$. To find b^0, we split the reasoning into a few parts.

Step 1. If $|x| + y \ge 0$, then $b^0(x,y) = 1 - |x|$. Indeed, the estimate "$\le$" follows from the inequalities $\mathbb{P}(g_n \ge 0) \le 1$ and $\mathbb{E}|f_n| \ge |x|$ (valid for all $(f,g) \in M(x,y)$), while "$\ge$" can be shown with the use of the following example: take a large number $M > |x|$ and consider a pair (f,g) satisfying

(i) $(f_0, g_0) \equiv (x,y)$,
(ii) $df_1 = -\operatorname{sgn} x \cdot dg_1$ is a centered random variable taking values $(M-|x|)\operatorname{sgn} x$ and $-x$ only,
(iii) $df_2 = df_3 = \ldots = dg_2 = dg_3 = \ldots \equiv 0$.

Then f does not change its sign, so $\mathbb{E}|f_\infty| = |x|$. Furthermore, on the set where $df_1 = -x$ we have $g_1 = y + |x| \ge 0$, so

$$\mathbb{P}(|g_\infty| \ge 0) \ge \frac{M - |x|}{M},$$

which can be made arbitrarily close to 1. Hence $b^0(x,y) \ge 1 - |x|$.

Step 2. Consider the sets

$$\mathcal{S} = \{(x,y) : b^0(x,y) = V(x,y)\}, \qquad \mathcal{C} = \{(x,y) : b^0(x,y) > V(x,y)\},$$

which we have called *the stopping region* and *the continuation region*. We will study the "shape" of $\mathcal{S}$ and $\mathcal{C}$. Assume that $|x| + y < 0$ and observe that if $x \ne 0$, then $(x,y) \in \mathcal{C}$. This can be easily shown by the construction of an appropriate martingale pair: clearly, there is $(f,g) \in M(x,y)$ such that f does not change its sign and $\mathbb{P}(g_n \ge 0) > 0$ for some $n \ge 1$; thus, $b^0(x,y) \ge \mathbb{P}(g_n \ge 0) - \mathbb{E}|f_n| > -|x| = V(x,y)$. Next, note that for any fixed x, the function $b^0(x,\cdot)$ is nondecreasing. This follows directly from the fact that V has this property. Indeed, fix a pair $(f,g) \in M(x,y)$; then for any $d \ge 0$ we have $(f, g+d) \in M(x, y+d)$ and

$$\mathbb{P}(g_n \ge 0) - \mathbb{E}|f_n| \le \mathbb{P}(g_n + d \ge 0) - \mathbb{E}|f_n| \le b^0(x, y+d).$$

Taking the supremum over all $(f,g) \in M(x,y)$ gives the desired monotonicity of b^0. Thus, if $(0,y) \in \mathcal{S}$, then automatically the whole halfline $\{0\} \times (-\infty, y]$ is contained within $\mathcal{S}$. These considerations lead us to the following conjecture: we have

$$\mathcal{S} = (\mathbb{R} \times [0,\infty)) \cup \big(\{0\} \times (-\infty, y_0]\big) \quad \text{ and } \quad \mathcal{C} = \mathbb{R}^2 \setminus \mathcal{S},$$

for some $y_0 < 0$ to be found.

Step 3. Now we guess the right foliation and the formula for b^0. Let us first look at the set $\{|x|+y\leq y_0\}$. If we decomposed the set into halflines of slope -1, this would lead to the function $(x,y)\mapsto -|x|$ which, as we have already observed above, does not work (indeed, then we would get $(x,y)\in S$). Thus, it in natural to foliate the angle into line segments of slope -1. Actually, a little thought suggests to extend this foliation (i.e., all the segments) to the whole set $\{(x,y):|x|+y\leq 0,\ y\leq |x|+y_0\}$. We already know b^0 on $\{0\}\times(-\infty,y_0]$ and $\{(x,-|x|):x\in\mathbb{R}\}$, so we obtain

$$b^0(x,y)\geq \frac{2|x|}{|x|-y}\left(1-\frac{|x|-y}{2}\right)=\frac{2|x|}{|x|-y}-|x|.$$

Let us denote the right-hand side by $b(x,y)$. Before we proceed, let us specify the examples which lead to the value of $b(x,y)$. We will describe them for $x>0$, in the case when x is negative we proceed symmetrically. Let M be a large positive number and consider (f,g) such that

(i) $(f_0,g_0)\equiv(x,y)$,
(ii) $df_1=dg_1$ is a centered random variable taking values $-x$ and $(-x-y)/2$ (so, (f_1,g_1) moves along the line of slope 1, and ends at the line $x=0$ or at the line $y=-x$).
(iii) On the set where $df_1=-x$, put $df_2=dg_2\equiv 0$; on the set where $df_1=(-x-y)/2$, we have $(f_1,g_1)=\big((x-y)/2,(-x+y)/2\big)$, and we copy the example from the Step 1. That is, conditionally on this set, we assume that $df_2=-dg_2$ is a centered random variable taking values $(-x+y)/2$ and M only.
(iv) $df_3=df_4=\ldots=dg_3=dg_4=\ldots\equiv 0$.

Step 4. Finally, we turn our attention to the value of y_0 and the lower bound for b^0 on the remaining part of the domain. In many situations, it is natural to conjecture that the Bellman function has some additional regularity. For instance, in our case it is plausible to assume that the derivative $b^0_x(0,y)$, $y\in[y_0,0)$, exists (and thus is equal to zero, by the symmetry of b^0). On the other hand, by the previous step, $b^0_x(0,y_0)=-1-2/y_0$. This suggests $y_0=-2$, and thus all we need is to construct a candidate on the square $\{(x,y):|x|+y<0,\ y-|x|\geq -2\}$. This is done with the use of similar examples as in the two-sided case above, which we will describe in a Markovian language. Namely, for (x,y) from the square, consider a pair (f,g) satisfying

(i) $(f_0,g_0)\equiv(x,y)$,
(ii) (f_1,g_1) moves along the line of slope -1 and ends at the line $x+y=0$ or $x+y=-2$.
(iii) Depending on whether $f_1+g_1=0$ or $f_1+g_1=-2$, the process $(f_n,g_n)_{n\geq 2}$ evolves according to the rules listed in Step 1 or Step 3, respectively.

Computing $\mathbb{E}V(f_\infty,g_\infty)$, we get the candidate $b(x,y)=[(y+2)^2-x^2]/4$. Summarizing, we have constructed the following special function:

$$b(x,y)=\begin{cases}1-|x| & \text{if } |x|+y\geq 0,\\ \frac{1}{4}[(y+2)^2-x^2] & \text{if } |x|+y<0,\ y-|x|\geq -2,\\ \frac{2|x|}{|x|-y}-|x| & \text{if } |x|+y<0,\ y-|x|<-2.\end{cases}\tag{3.12}$$

So far, we can only say that $b^0 \geq b$, since the construction of b was based on examples. To prove the reverse, we check that b satisfies the conditions 1° and 2° (cf. [33]). Hence $b = b^0$, and the first part of our analysis is complete.

Remark 3.2. Alternatively, it is also quite easy to identify b^0 with the use of "iteration procedure" described in Remark (d) of §2.2. Namely, it can be shown that the sequence $(V_n)_{n\geq 0}$ stabilizes after three steps, and $V_4 = V_5 = \ldots$ is the function given by (3.12).

Now, we come back to the problem of identifying $\mathcal{B}^0$. We have just shown that for any $(f, g) \in M(x, y)$ we have

$$\mathbb{P}(g_n \geq 0) - \mathbb{E}|f_n| \leq b^0(x, y), \qquad n = 0,\, 1,\, 2, \ldots.$$

Now we apply a simple homogenization and translation argument: for any (f, g) as above and any $\alpha > 0$, the pair $(\alpha f, \alpha g - \alpha)$ belongs to $M(\alpha x, \alpha y - \alpha)$ and hence

$$\mathbb{P}(\alpha g_n - \alpha \geq 0) - \alpha\mathbb{E}|f_n| \leq b^0(\alpha x, \alpha y - \alpha), \qquad n = 0,\, 1,\, 2,\, \ldots,$$

or, in other words,

$$\mathbb{P}(g_n \geq 1) \leq b^0(\alpha x, \alpha y - \alpha) + \alpha\mathbb{E}|f_n|, \qquad n = 0,\, 1,\, 2,\, \ldots.$$

Therefore, if we additionally assume that $\mathbb{E}|f_\infty| \leq t$ (and hence also $\mathbb{E}|f_n| \leq t$), we obtain that

$$\mathcal{B}^0(x, y, t) \leq b^0(\alpha x, \alpha y - \alpha) + \alpha t$$

for all $\alpha > 0$. All that is left is to optimize over α. Actually, one easily checks that

$$B(x, y, t) = \inf_{\alpha>0} \left\{ b^0(\alpha x, \alpha y - \alpha) + \alpha t \right\}$$

gives precisely the function (3.10). This proves $\mathcal{B}^0 \leq B$. The reverse bound can be verified with the use of the examples from Steps 1, 3 and 4 above. More precisely: if $y + t \geq 1$, then the equality $\mathcal{B}^0(x, y, t) = 1$ can be proved with the use of a similar pair as in Step 1 (we only need to add 1 to y to ensure that $\mathbb{P}(|g_1| \geq 1)$ is arbitrarily close to 1). If $y + t < 1$, we consider all the examples of Steps 3 and 4, starting from the points of the form $(\alpha x, \alpha y - \alpha)$, where α runs over all positive numbers. From all these pairs (f, g), we pick one which satisfies $\mathbb{E}|f_\infty| = \alpha t$ (there are many such pairs; we pick the one corresponding to a large parameter M). Then $(\bar{f}, \bar{g}) = (f/\alpha, g/\alpha + 1)$ belongs to $M(x, y)$, $\mathbb{E}|\bar{f}_\infty| = t$ and, after some calculations, we check that $\mathbb{P}(\bar{g}_\infty \geq 1)$ can be made arbitrarily close to $B(x, y, t)$ (by taking M sufficiently large). This yields the desired lower bound $\mathcal{B}^0 \geq B$.

3.3. **More exact information on the two-sided bound.** Now we will provide the explicit formula for the function

$$\begin{aligned} &\mathcal{B}^0(x, y, t) \\ &= \sup \Big\{ \mathbb{P}(|g_n| \geq 1) \;:\; (f, g) \in M(x, y),\, \mathbb{E}|f_\infty| \leq t,\, n = 0,\, 1,\, 2,\, \ldots \Big\}. \end{aligned} \tag{3.13}$$

This will be accomplished by the technique described in Section 2, with $V(x, y) = 1_{\{|y|\geq 1\}}$.

Approach 1. In comparison to the one-sided case, the situation is more difficult since the function $\mathcal{B}^0$ does not seem to have any homogeneity-type property. Nevertheless, it majorizes the Bellman function corresponding to the one-sided estimate, which gives

$$\mathcal{B}^0(x,y,t) \geq \begin{cases} 1 & \text{if } |y|+t\geq 1, \\ 1-\frac{(1-|y|-t)^2}{(1-|y|)^2-x^2} & \text{if } |y|+t<1. \end{cases} \tag{3.14}$$

This, in particular, yields

$$\mathcal{B}^0(x,y,t)=1 \quad \text{provided } |y|+t\geq 1. \tag{3.15}$$

Next, we proceed as follows. Fix $a\in(0,1)$ and consider the function

$$b(x,y)=\mathcal{B}^0\left(\frac{x+a}{2},\frac{x-a}{2},y\right),$$

given on the set $\left\{(x,y)\in\mathbb{R}^2 : y\geq\left|\frac{x+a}{2}\right|\right\}$. This function is concave and, by (3.15), we have $b(x,y)=1$ for $y\geq 1-\left|\frac{x-a}{2}\right|$. Thus all we need is to determine the formula for b on the parallelogram $\mathcal{P}=\left\{(x,y):\left|\frac{x+a}{2}\right|\leq y<1-\left|\frac{x-a}{2}\right|\right\}$ (see Figure 2).

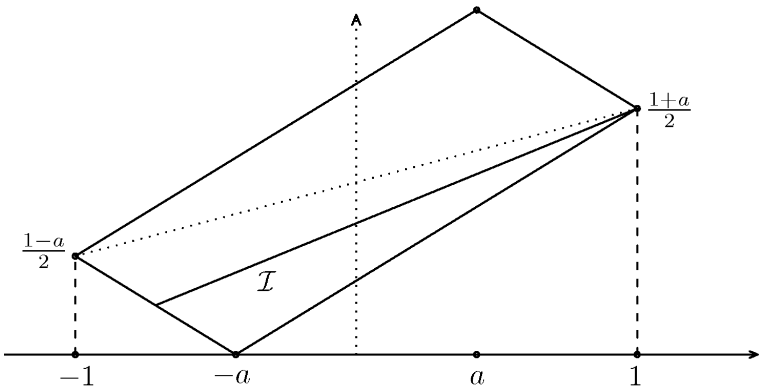

Figure 2. The parallelogram $\mathcal{P}$.

Directly from the concavity of b, we obtain that $b(x,y)=1$ if (x,y) lies on or above the dotted diagonal of $\mathcal{P}$ - precisely, the line segment with endpoints $\left(-1,\frac{1-a}{2}\right)$ and $\left(1,\frac{1+a}{2}\right)$ - due to the fact that b equals 1 when evaluated at the sides of $\mathcal{P}$ lying above this segment. For (x,y) lying below the diagonal we have, by (3.14),

$$b(x,y)\geq\zeta(x,y)=1-\frac{\left(1-\left|\frac{x-a}{2}\right|-y\right)^2}{(1-a)(1+x)}.$$

Let us search for the least concave majorant of ζ. Some experiments lead to the following idea. Take an interval $\mathcal{I}$ with endpoints $\left(1,\frac{1+a}{2}\right)$ and $\left(t,-\frac{t+a}{2}\right)$, where $t\in(-1,-a]$ (see Figure 2). It is easy to check that ζ is not concave along this interval and that the least concave majorant of $\zeta|_{\mathcal{I}}$ is given by

$$b_0(x,y)=\begin{cases} \zeta(x,y) & \text{if } (x,y)\in\mathcal{I},\ y<\frac{a}{2}-\left(\frac{1}{2}-a\right)x, \\ 2y-ax & \text{if } (x,y)\in\mathcal{I},\ y\geq\frac{a}{2}-\left(\frac{1}{2}-a\right)x. \end{cases}$$

Assuming $b = b_0$ for all (x, y) below the diagonal, we obtain the candidate for the Bellman function, given as follows. Consider the sets

$$
\begin{aligned}
\mathcal{D}_1 &= \{(x,y,t) : |x|+|y| \geq 1\} \cup \left\{(x,y,t) : |x|+|y|<1, t \geq \frac{1}{2}(x^2-y^2+1)\right\},\\
\mathcal{D}_2 &= \{(x,y,t) : |x|+|y|<1,\, t < x^2-y^2+|y|\},\\
\mathcal{D}_3 &= \left\{(x,y,t) : |x|+|y|<1,\, x^2-y^2+|y| \leq t < \frac{1}{2}(x^2-y^2+1)\right\}.
\end{aligned}
$$

Note that if $|x|+|y|<1$, then $x^2-y^2+|y| < \frac{1}{2}(x^2-y^2+1)$; thus the subsets are pairwise disjoint. The candidate B we obtain is given by

$$
B(x,y,t) = \begin{cases} 1 & \text{on } \mathcal{D}_1,\\ 1 - \frac{(1-|y|-t)^2}{(1-|y|)^2-x^2} & \text{on } \mathcal{D}_2,\\ 2t - x^2 + y^2 & \text{on } \mathcal{D}_3. \end{cases} \tag{3.16}
$$

It can be shown that this function satisfies 1° and 2° and hence $\mathcal{B}^0 \leq B$; on the other hand, the bound $\mathcal{B}^0 \geq B$ follows directly from the above construction. Thus $\mathcal{B}^0 = B$, as originally proved in [36].

Approach 2. As in the one-sided case, there is a question whether the function (3.16) can be discovered with the use of two-dimensional boundary value problem. The answer is positive, however, due to the fact that we have no additional homogeneity, we will actually need to study a whole *family* of auxiliary estimates. Namely, for $c \geq 0$, let $V_c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be given by $V_c(x,y) = 1_{\{|y|\geq 1\}} - c|x|$ and let

$$
b_c^0(x,y) = \sup\left\{\mathbb{E}V_c(f_n,g_n) \ :\ (f,g) \in M(x,y),\, n = 0,\,1,\,2,\,\ldots\right\}.
$$

We will find the formula for b_c^0 in several steps below.

Step 1. For $c \leq 2$, we proceed exactly in the same manner as in the search for (3.4) (or follow the analytic approach presented there). One way or another, we obtain that

$$
b_c^0(x,y) = \begin{cases} \frac{c}{2}(y^2-x^2) + 1 - \frac{c}{2} & \text{if } |x|+|y| \leq 1,\\ 1 - c|x| & \text{if } |x|+|y| > 1. \end{cases}
$$

Step 2. The situation gets more interesting when $c > 2$. The above formula does not work any more, since the majorization $b_c^0(0,0) \geq V_c(0,0)$ is no longer valid. How can we proceed? Again, some intuition can be gained from the one sided version of the problem. In a sense, we expect that b_c^0 will be a "symmetrized" modification of that Bellman function. A little thought leads to the following splitting of $\mathbb{R}^2$:

$$
\begin{aligned}
D_1 &= \{(x,y) : |x|+|y| \geq 1\},\\
D_2 &= \{(x,y) : |x|+|y| < 1, |y| \leq |x|\},\\
D_3 &= \{(x,y) : |x|+|y| < 1, |x| < |y| < |x|+\alpha\},\\
D_4 &= \mathbb{R}^2 \setminus (D_1 \cup D_2 \cup D_3).
\end{aligned}
$$

Here the parameter α, to be determined later, has the property that $b_c^0(0,y) = V(0,y) = 0$ for $|y| \leq \alpha$, and $b_c^0(0,y) > 0$ for $|y| > \alpha$. What about the formula for b_c^0? The same arguments as previously give that $b_c^0(x,y) = 1 - c|x|$ on D_1. Furthermore, both D_2 and D_4 consist of two squares; on each square, in analogy with the preceding considerations, we expect b_c^0 to be linear along line segments of slope ± 1 (in other words, it is quadratic there - see the computations below).

Finally, on $D_3 \cap [0,\infty)^2$, b_c^0 should be linear along the segments of slope 1 (the remaining part of D_3 is dealt with using symmetry).

Step 3. Now let us present some computations. For any $0 \leq y \leq \alpha$, the candidate b_c satisfies $b_c(0,y) = 0$, $b_c((1-y)/2,(1+y)/2) = 1-c|1-y|/2$ (because $((1-y)/2,(1+y)/2) \in D_1$) and is linear along the line segment joining the two evaluated points. Thus,

$$b_c(x,y) = \frac{2|x|}{1+|x|-|y|} - c|x| \qquad \text{on } D_3.$$

As in the one-sided case, we expect the equality $b_{cx}(0,\alpha) = 0$: this gives $\alpha = 1-2/c$. Now the discovery of the candidate b_c on D_2 and D_4 is just a mere repetition of the calculations from the one-sided setting. As the reader easily verifies, at the very end we obtain

$$b_c(x,y) = \begin{cases} 1 - c|x| & \text{if } (x,y) \in D_1, \\ y^2 - x^2 + 2|x| - c|x| & \text{if } (x,y) \in D_2, \\ \frac{2|x|}{1+|x|-|y|} - c|x| & \text{if } (x,y) \in D_3, \\ 1 - c(1-|y|) + \frac{c^2}{4}\big((|y|-1)^2 - x^2\big) & \text{if } (x,y) \in D_4. \end{cases}$$

As we have already seen a few times, all that we can claim so far is the bound $b_c \leq b_c^0$ (the construction of b_c rests on examples). Now we will verify that b_c satisfies 1° and 2°. We start with 2°. Fix $y \in \mathbb{R}$ and consider the function $G(t) = b_c(t, y-t)$, $t \in \mathbb{R}$. It suffices to prove that this function is concave (then, by the symmetry of b_c, we will obtain that the sections of the form $t \mapsto b_c(t, y+t)$ are also concave, and the property will follow). Note that if $(t, y-t)$ lies in the interior of one of the sets D_2 or D_4, we have $G''(t) = 0$. The same is true if $(t,y-t)$ belongs to the interior of D_1 or D_3 (for the latter set, simply compute the second derivative), unless $t = 0$. In this particular case the second derivative of G does not exist, but we easily check that $G'(0-) \geq G'(0+)$, so the concavity is preserved. Thus, by the continuity of G, all we need is to verify that there are appropriate inequalities between one sided derivatives when $(t,y-t)$ lies at the common boundary of some D_i's. By the symmetry, we may and do assume that $t \leq 0$. If $(t,y-t) \in \partial D_1 \cap \partial D_4$, then $G'(t-) = c \geq c - 2y - 2 = G'(t_+)$; if $(t, y-t) \in \partial D_1 \cap \partial D_2$, then $G'(t-) = c \geq c - 2(1-y)^{-1} = G'(t+)$. Next, if $(t,y-t) \in \partial D_1 \cap D_4$, then $G'(t-) = c$ and

$$G'(t+) = \frac{c^2}{2}(1-y) - c \leq c - c = 0,$$

since $y \geq 1 - 2/c$ (which follows from the assumption on $(t, y-t)$). The remaining cases (i.e., $(t,y-t) \in \partial D_2 \cap \partial D_3$ and $(t,y-t) \in \partial D_3 \cap \partial D_4$) are studied in a similar manner; we leave the details to the reader.

Finally, we turn our attention to 1°. We have equality for $|x| + |y| \geq 1$ and $b_c(0,|y|) \geq 0$ for $y \in (-1,1)$. All we need is to combine this with the fact that on the left and on the right halfplane, the function V_c is linear, while b_c is diagonally concave. This gives the majorization, and hence $b_c = b_c^0$.

Remark 3.3. As in the preceding settings, the function b_c can be identified with the use of the iterative approach described in Remark (d) in §2.1 (the recurrence stabilizes after three steps). The calculations are not very difficult, but there are several cases to be considered and one may find this path a little tedious.

Now we come back to the problem of identifying $\mathcal{B}^0$ defined in (3.13). For any $(f,g)\in M(x,y)$ satisfying $\mathbb{E}|f_\infty|\leq t$ we have

$$\mathbb{P}(|g_n|\geq 1)\leq b_c^0(x,y)+c\mathbb{E}|f_n|\leq b_c^0(x,y)+ct$$

and hence $\mathcal{B}^0(x,y,t)\leq \min_{c\geq 0}\big\{b_c^0(x,y)+ct\big\}$. Let us denote the right hand side by B and let us derive its explicit formula. Assume first that $|x|+|y|\geq 1$. Then $b_c(x,y)=1-c|x|$ and $b_c(x,y)+ct=1+c(t-|x|)$. Thus we must take $c=0$ and we get $B(x,y,t)=1$. Next, suppose that $|x|+|y|<1$, but $|x|\geq |y|$. Then

$$b_c^0(x,y)+ct=\begin{cases}\frac{c}{2}(y^2-x^2)+1-\frac{c}{2}+ct & \text{if } c\leq 2,\\ y^2-x^2+2|x|-c|x|+ct & \text{if } c>2.\end{cases}$$

If we compute the derivative with respect to c, we get

$$\frac{\mathrm{d}}{\mathrm{d}c}\big[b_c^0(x,y)+ct\big]=\begin{cases}t-\frac{1+x^2-y^2}{2} & \text{if } c<2,\\ t-|x| & \text{if } c>2.\end{cases}$$

So, we have two possibilities. If $t\geq (1+x^2-y^2)/2$, then the derivative is nonnegative on $(0,2)\cup(2,\infty)$, so the minimum is attained for $c=0$: this gives $B(x,y,t)=1$. On the other hand, if $t<(1+x^2-y^2)/2$, then the derivative is negative on $(0,2)$ and positive on $(2,\infty)$, so

$$B(x,y)=b_2^0(x,y)+2t=y^2-x^2+2t.$$

Next, assume that $|x|+|y|<1$ and $|y|>|x|$. Put $\alpha=1-2/c$; then

$$\begin{aligned}b_c^0(x,y)+ct&=\begin{cases}\frac{c}{2}(y^2-x^2)+1-\frac{c}{2}+ct & \text{if } c\leq 2,\\ 1-c(1-|y|)+\frac{c^2}{4}\big((|y|-1)^2-x^2\big)+ct & \text{if } c>2,\ |y|-|x|>\alpha,\\ \frac{2|x|}{1+|x|-|y|}-c|x|+ct & \text{if } c>2,\ |y|-|x|\leq\alpha\end{cases}\\ &=\begin{cases}\frac{c}{2}(y^2-x^2)+1-\frac{c}{2}+ct & \text{if } c\leq 2,\\ 1-c(1-|y|)+\frac{c^2}{4}\big((|y|-1)^2-x^2\big)+ct & \text{if } c\in(2,\frac{2}{1-|y|+|x|}),\\ \frac{2|x|}{1+|x|-|y|}-c|x|+ct & \text{if } c\geq \frac{2}{1-|y|+|x|}.\end{cases}\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}c}\big[b_c^0(x,y)+ct\big]=\begin{cases}t-\frac{1+x^2-y^2}{2} & \text{if } c<2,\\ t+|y|-1+\frac{c}{2}\big((|y|-1)^2-x^2\big) & \text{if } c\in(2,\frac{2}{1-|y|+|x|}),\\ t-|x| & \text{if } c>\frac{2}{1-|y|+|x|}.\end{cases}$$

This time we have three possibilities. If $t\geq (x^2-y^2+1)/2$, then for $c>2$,

$$\begin{aligned}t+|y|-1+\frac{c}{2}\big((|y|-1)^2-x^2\big)&\geq t+|y|-1+(|y|-1)^2-x^2\\ &\geq \frac{-x^2+(|y|-1)^2}{2}>0,\end{aligned}$$

so the minimum defining B is attained for $c=0$. Therefore, $B(x,y,t)=1$ in this case. Suppose then that $t<(x^2-y^2+1)/2$, but

$$t+|y|-1+(|y|-1)^2-x^2=t-x^2+y^2-|y|\geq 0.$$

Then the choice $c=2$ is optimal: $B(x,y,t)=b_2^0(x,y)+2t=y^2-x^2+2t$. Finally, if $t<(x^2-y^2+1)/2$ and $t-x^2+y^2-|y|<0$, then the above derivative vanishes for $c = 2(1-|y|-t)/((|y|-1)^2 - x^2)$, and then

$$B(x,y,t) = b_c^0(x,y) + ct = 1 - \frac{(1-|y|-t)^2}{(y-1)^2 - x^2}.$$

Thus, we end up with the function (3.16). So, we have shown that $\mathcal{B}^0 \leq B$, and the proof of the reverse bound rests on the construction of appropriate examples. For an alternative argument, see Approach 1 or consult [36].

4. Weak type (p,p) estimates for martingale transforms, $p > 1$

Now we will study the versions of the above weak-type bounds in the case $p > 1$. Precisely, we will determine the best constants C_p such that

$$\mathbb{P}(|g_n| \geq 1) \leq C_p^p \mathbb{E}|f_n|^p, \qquad n = 0, 1, 2, \ldots. \tag{4.1}$$

These best constants were originally identified by Burkholder in [8] for $1 < p \leq 2$, and by Suh [44] for remaining values of p. Our considerations below will strengthen the results from those papers. Quite unexpectedly, the arguments presented in the cases $1 < p \leq 2$ and $p > 2$ are entirely different. The two situations will be considered separately.

4.1. **The case $1 < p \leq 2$.** We proceed according to the methodology described in §2.1 and write down the formula (2.1):

$$\mathcal{B}^0(x,y) = \sup\left\{\mathbb{E}V_\beta(f_n,g_n) \ : \ (f,g) \in M(x,y),\ n = 0, 1, 2, \ldots\right\}.$$

Here $\beta \geq 0$ and $V_\beta(x,y) = 1_{\{|y|\geq 1\}} - \beta|x|^p$. The function $\mathcal{B}^0$ satisfies the symmetry condition

$$\mathcal{B}^0(x,y) = \mathcal{B}^0(-x,y) = \mathcal{B}^0(x,-y) \qquad \text{for all } x,\, y \in \mathbb{R}, \tag{4.2}$$

so it suffices to determine it in the first quadrant $[0,\infty) \times [0,\infty)$.

Step 1. The case $y \geq 1$. For these y we have

$$\mathcal{B}^0(x,y) = 1 - \beta x^p. \tag{4.3}$$

Indeed, for all f, g as above we may write $\mathbb{P}(|g_n| \geq 1) \leq 1$ and $\mathbb{E}|f_n|^p \geq x^p$, where the latter follows from Jensen's inequality. This gives the estimate in one direction, and the choice of constant f and g yields the reverse.

Step 2. Two key assumptions. Now we will impose two conditions on the candidate B for the Bellman function. The first of them concerns regularity on the set $\mathbb{R} \times [-1,1]$, while the second indicates the foliation on $[0,\infty) \times [0,1]$. Precisely, the first assumption reads

- (A1) B is continuous on $\{(x,y) : |y| \leq 1\}$ and of class C^1 in the interior of this set,

while the second is

- (A2) On $[0,\infty) \times [0,1]$, the function B is linear along the line segments of slope -1.

The latter condition comes out when one considers appropriate exemplary martingale pairs. Suppose, for instance, that $x > 0$, $y \in (0,1)$ with $x + y > 1$ and assume we are interested in identifying $(f,g) \in M(x,y)$ for which $\mathcal{B}^0(x,y)$ is (almost) attained. Intuitively speaking, we want to make $\mathbb{P}(|g_n| \geq 1)$ large, with keeping $\mathbb{E}|f_n|^p$ relatively small. Since the second derivative of $t \mapsto t^p$ goes to 0 as $t \to \infty$, the difference $\mathbb{E}|f_{n+1}|^p - \mathbb{E}|f_n|^p$ is insignificant, at least when f_n is large. Thus, it is reasonable to consider the pair $(f,g)\in M(x,y)$ which satisfies the following. Fix a small positive δ and assume that $(f_0,g_0)\equiv(x,y)$,

(i) $df_1=-dg_1$ is a centered random variable with values in $\{y-1,y\}$.
(ii) if $(f,g)=(x+y+2k\delta,0)$ for some nonnegative integer k, then at the next step (f,g) moves to $(x+y+2k\delta-1,-1)$ or to $(x+y+(2k+1)\delta,\delta)$.
(iii) if $(f,g)=(x+y+(2k+1)\delta,\delta)$ for some nonnegative integer k, then at the next step (f,g) moves to $(x+y+(2k+2)\delta-1,1)$ or to $(x+y+(2k+2)\delta,0)$.

Note that $\mathbb{P}(|g_n|\geq 1)\to 1$ as $n\to\infty$; furthermore, we "push" the evolution of f_n towards infinity, thus making the increase of the p-th moment smaller and smaller. The properties (i)-(iii) mean that on $[0,\infty)\times[0,1]$, (f,g) moves along the lines of slope -1, while on $[0,\infty)\times[-1,0]$, along the lines of slope 1. This is where (A2) originates from. For convenience, denote $a(t)=B(t,1)=1-\beta t^p$, $b(t)=B(t,0)$ and $c(t)=B(0,t)$. Then (A2) means that

$$(4.4)\qquad \begin{aligned} B(x,y)&=ya(x+y-1)+(1-y)b(x+y) && \text{if } x+y\geq 1\geq y\geq 0,\\ B(x,y)&=\frac{y}{x+y}c(x+y)+\frac{x}{x+y}b(x+y) && \text{if } x,\, y\geq 0,\ x+y<1.\end{aligned}$$

Step 3. Derivation of a candidate. Now we will see that the above conditions (A1) and (A2) determine uniquely the candidate B. The symmetry condition (4.2) implies

$$(4.5)\qquad B_x(0,y)=B_y(x,0)=0 \qquad \text{for } x\in\mathbb{R} \text{ and } y\in(-1,1).$$

Using this and (4.4), we obtain the differential equations

$$(4.6)\qquad c'(y)=\frac{c(y)-b(y)}{y} \qquad \text{for } y\in[0,1),$$

$$(4.7)\qquad b'(x)=\frac{b(x)-c(x)}{x} \qquad \text{for } x\in[0,1),$$

and

$$(4.8)\qquad b'(x)=b(x)-a(x-1) \qquad \text{for } x\geq 1.$$

By (4.6) and (4.7) and the condition $b(0)=c(0)=B(0,0)$ we have that $b(x)+c(x)=2c(0)$ on $[0,1]$. Plugging this into (4.6) yields $c(y)=c(0)+\alpha y^2$ for all $y\in[0,1]$ and some fixed $\alpha\in\mathbb{R}$. Since $c(1)=a(0)=1$, we see that $\alpha=1-c(0)$, so $b(x)=c(0)-(1-c(0))x^2$ and $c(y)=c(0)+(1-c(0))y^2$ for $x,\, y\in[0,1]$. Applying the second equality in (4.4), we get that

$$B(x,y)=c(0)+(1-c(0))(y^2-x^2) \qquad \text{if } |x|+|y|\leq 1.$$

Similarly, solving (4.8) (recall that $a(t)=1-\beta t^p$) gives

$$b(x)=1-\beta e^x\int_x^\infty e^{-t}(t-1)^p\mathrm{d}t+\gamma e^x$$

for $x>1$ and some $\gamma\in\mathbb{R}$. Note that $-\beta x^p\leq b(x)\leq 1-\beta x^p$, directly from the definition of b and $\mathcal{B}^0$; this implies $\gamma=0$, simply by letting $x\to\infty$ above. By continuity of b at 1, we derive that

$$1-\beta e\int_1^\infty e^{-t}(t-1)^p\mathrm{d}t=2c(0)-1,$$

or

$$\beta = \frac{2-2c(0)}{\Gamma(p+1)}.$$

By the first equation in (4.4), for $|x|+|y|>1$ we have

$$B(x,y) = 1 - |y|\beta(|x|+|y|-1)^p - \beta(1-|y|)e^{|x|+|y|-1}\int_{|x|+|y|-1}^\infty e^{-s}s^p \mathrm{d}s.$$

Now it can be checked that if $c(0)\in[0,1]$, then the above function B satisfies 1° and 2°, so $\mathcal{B}^0 = B$. Thus, we have successfully identified the Bellman function for $\beta \leq 2/\Gamma(p+1)$. How is it related to (4.1)? To answer this, we need to look at the condition 3°: one easily proves that $B(x,\pm x)\leq c(0)$, and hence the above reasoning yields the sharp bound

$$\mathbb{P}(|g_n|\geq 1) \leq \frac{2-2c(0)}{\Gamma(p+1)}\mathbb{E}|f_n|^p + c(0), \qquad n=0,\,1,\,2,\,\ldots,$$

for an arbitrary $c(0)\in[0,1]$. In particular, taking $c(0)=0$, we obtain that the best choice for C_p is $(2/\Gamma(p+1))^{1/p}$.

Question 1: What is the formula for B when $\beta > 2/\Gamma(p+1)$?

4.2. **The case $p>2$.** As previously, we will study the more general setting in which the constant C_p^p in (4.1) is replaced by an arbitrary $\beta\geq 0$. That is, introduce the Bellman function

$$\mathcal{B}^0(x,y) = \sup\big\{\mathbb{E}V_\beta(f_n,g_n)\ :\ (f,g)\in M(x,y),\, n=0,\,1,\,2,\,\ldots\big\}, \tag{4.9}$$

where, as above, $V_\beta(x,y) = 1_{\{|y|\geq 1\}} - \beta|x|^p$. In analogy with the preceding case, we will denote by B the candidate for $\mathcal{B}^0$, which will be obtained after a number of assumptions. Actually, we will manage to find the Bellman function only for $\beta\geq 2^{p-1}/p$. The analysis in the case $0<\beta<2^{p-1}/p$ has eluded us.

Step 1. The case $|y|\geq 1$. The same argument as previously yields $\mathcal{B}^0(x,y) = 1-\beta|x|^p$.

Step 2. A special curve. The following intuitive observation is a key part of the construction. Let x be a large real number and let $y\in(-1,1)$. Suppose that we are interested in determining $\mathcal{B}^0(x,y)$. To do this, loosely speaking, we need to find such f, g and n, for which $\mathbb{P}(|g_n|\geq 1)$ is large and $\mathbb{E}|f_n|^p$ is relatively small. However, the "gain" we can get from the first term is at most 1. This may not be enough to compensate the "loss" coming from the growth of the p-th moment of f (at least if $|x|$ is sufficiently large). This is the consequence of the fact that the second derivative of $x\mapsto x^p$ grows to infinity as $x\to\infty$: this is where the condition $p>2$ plays the role. In other words, if $y\in(-1,1)$ and $|x|$ is large, it is natural to conjecture that the best pair $(f,g)\in M(x,y)$ is the constant one: hence $\mathcal{B}^0(x,y) = -\beta|x|^p$. This suggests that the following assumption:

(A1) There is $c\geq 0$ and an nondecreasing function $\gamma:[c,\infty)\to[0,1]$ of class C^1 such that if $|y|\leq\gamma(|x|)$, then $B(x,y) = -\beta|x|^p$.

As we shall see, the condition $\beta\geq 2^{p-1}/p$ enforces c to be not too large: $c\leq 1/2$. Thus, we will restrict ourselves to these values of this parameter. Now, what can be said about the value of $\gamma(c)$? Extending the function γ "as far as possible" leads to the following assumption:

(A2) We have $c=0$ or $\gamma(c)=0$.

See Figures 3, 4 and 5 below, which illustrate the possibilities that can occur.

Step 3. Further assumptions. So, it remains to determine B on the set $J = \{(x,y) : x > 0,\, \gamma(x) < y < 1\}$. Let $a(y) = B(0,y)$ and $b(x) = B(x,0)$ for $x,\, y \in \mathbb{R}$. It seems reasonable to impose the following regularity condition on B.

(A3) The function B is continuous on $\mathbb{R} \times [-1,1]$ and of class C^1 in the interior of this set.

By symmetry of V_β, we may restrict ourselves to the functions satisfying

$$B(x,y) = B(-x,y) = B(x,-y) \quad \text{for all } x,\, y \in \mathbb{R}. \tag{4.10}$$

By (A3), this gives

$$B_x(0,y) = 0 \quad \text{and} \quad B_y(x,0) = 0 \qquad \text{for } x,\, y \geq 0. \tag{4.11}$$

Now we will introduce an important structural condition on B. Instead of examples, we will just indicate the foliation along which they evolve (compare the assumption (A2) in the previous case, and the discussion following it). Suppose first that $c = 0$. Then the condition reads

(A4) There is $y_* \in [0,1]$ such that for $(x,y) \in J$,

$$B(x,y) = \frac{x}{x+t} B(x+t, y-t) + \frac{t}{x+t} a(x+y) \qquad \text{if } x+y \leq y_*, \tag{4.12}$$

where $t = t(x,y)$ is the unique positive number satisfying $y - t = \gamma(x+t)$, and

$$B(x,y) = \frac{x}{x+1-y} B(x+1-y, 1) + \frac{1-y}{x+1-y} a(y-x) \qquad \text{if } -x+y \geq y_*. \tag{4.13}$$

The condition (A4) enforces B to be linear along line segments of slope -1 contained in D_4 and line segments of slope 1 contained in D_5 (the regions D_4 and D_5 are as on Figure 3 and will be formally defined below).

In the case $c \in (0, 1/2]$, the assumption is slightly different:

(A4) There is $y_* \in [0,1]$ such that for $(x,y) \in J$,

$$B(x,y) = \frac{x}{x+t} B(x+t, y-t) + \frac{t}{x+t} a(x+y) \qquad \text{if } c \leq x+y \leq y_*, \tag{4.14}$$

where $t = t(x,y)$ is the unique positive number satisfying $y - t = \gamma(x+t)$. If $(x,y) \in J$ and $x + y < c$, then

$$B(x,y) = \frac{x}{x+y} b(x+y) + \frac{y}{x+y} a(x+y). \tag{4.15}$$

Finally,

$$B(x,y) = \frac{x}{x+1-y} B(x+1-y, 1) + \frac{1-y}{x+1-y} a(y-x) \qquad \text{if } -x+y \geq y_*. \tag{4.16}$$

Take a look at the Figures 3-5 below and compare the cases $c = 0$, $c > 0$.

Step 4. A special case. From now on, until we say otherwise, we assume that $c = \gamma(c) = 0$; the function corresponding to this choice of parameters will play a distinguished role in the considerations below (and will lead to the special functions of remaining cases via some simple transformations). Pick a nonnegative x satisfying $x + \gamma(x) \leq y^*$ and let $t \in [0,x]$. By (A3) and (A4) we have

$$B(x-t, \gamma(x)+t) = B(x,\gamma(x)) + (-B_x(x,\gamma(x)) + B_y(x,\gamma(x)))t,$$

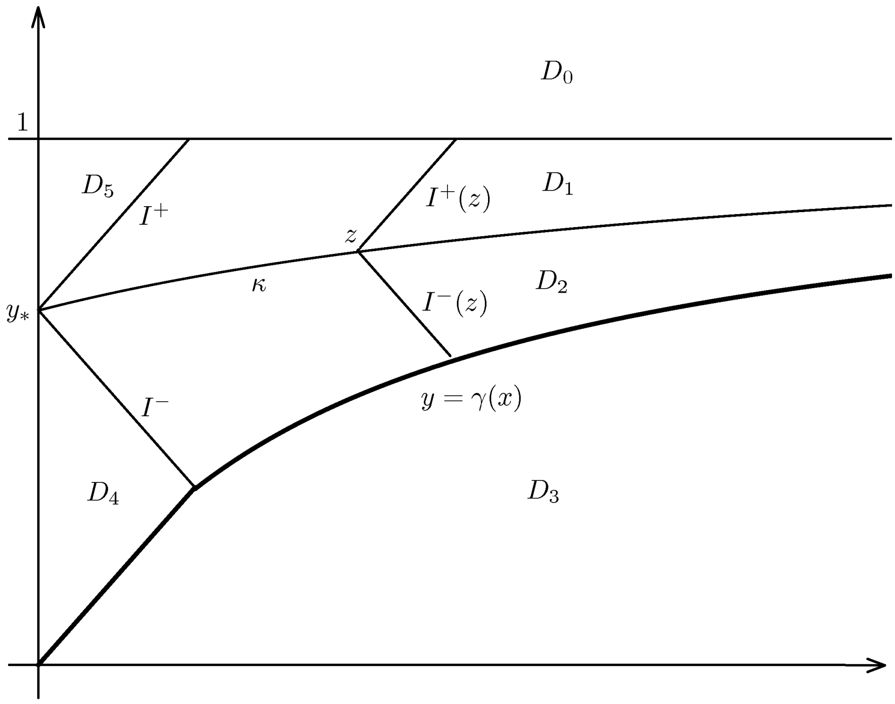

FIGURE 3. Bellman function for the weak-type (p,p) estimate in the case $c = \gamma(c) = 0$.

so, by (A1),

$$B(x-t, \gamma(x)+t) = -\beta x^p + p\beta x^{p-1}t. \tag{4.17}$$

Take $t = x$ and differentiate both sides over x. We get

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

On the other hand, differentiate in (4.17) over t, then let $t = x$ and use (4.11) to obtain

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

The two equations above give $\gamma'(x) = p-2$, and hence $\gamma(x) = (p-2)x$ provided $x + \gamma(x) \le y_*$.

By (4.12) and (4.13), the function B is linear on the line segments $I^\pm$ of slope ± 1 such that $(0, y^*) \in I^\pm \subset J$. Combining this with the symmetry condition $B(x,y) = B(-x,y)$, we get that the function

$$F(t) := B\left(\frac{y_*}{p-1} - t, \frac{(p-2)y_*}{p-1} + t\right)$$

is linear on $[0, 1-(p-2)y_*/(p-1)]$. Indeed, for t lying in this interval, the points

$$\left(\frac{y_*}{p-1} - t, \frac{(p-2)y_*}{p-1} + t\right)$$

fill out the line segment of slope -1, passing through $(0, y_*)$, with one endpoint lying on the line $y = 1$, and the other being the endpoint of the found linear piece

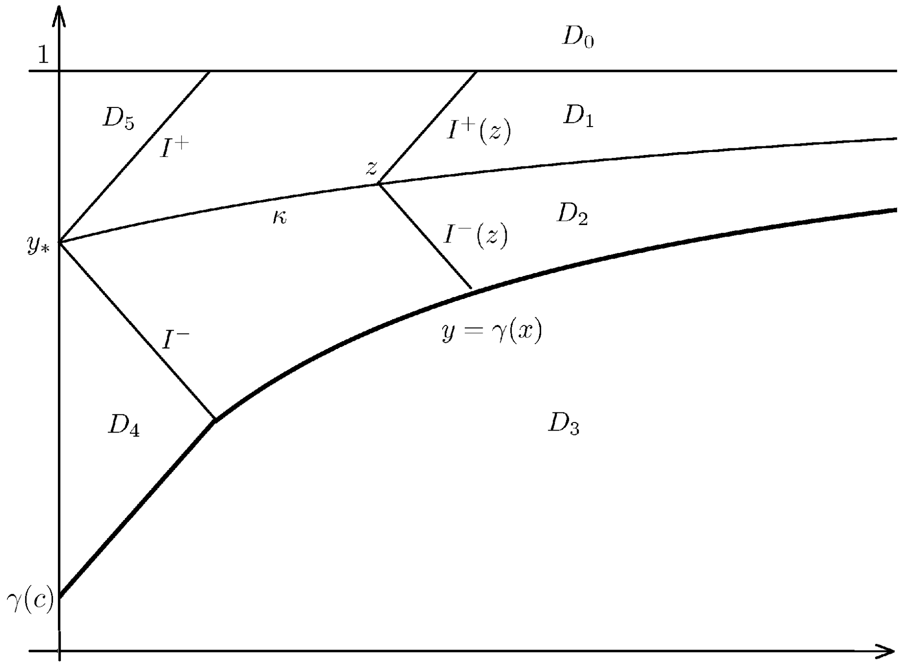

FIGURE 4. Bellman function for the weak-type (p, p) estimate in the case $c = 0 < \gamma(c)$.

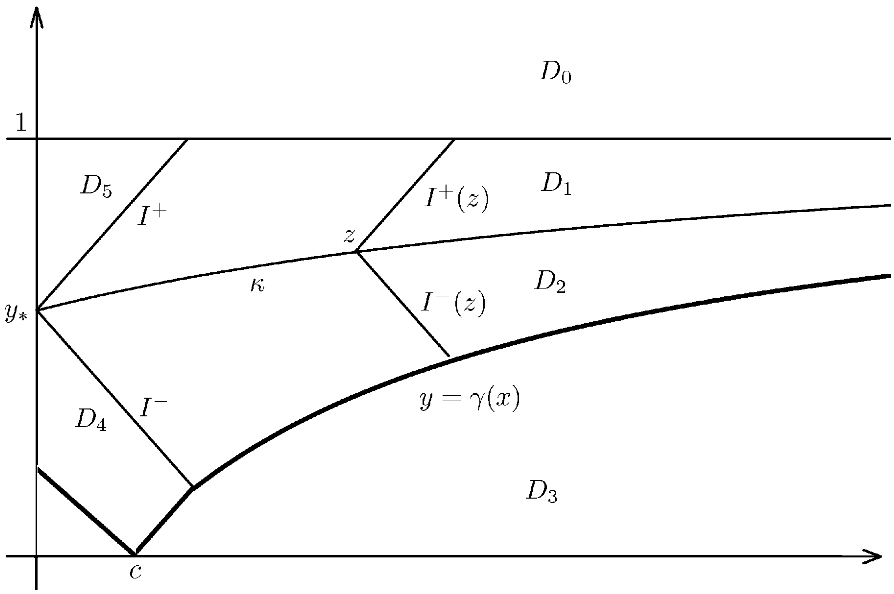

FIGURE 5. Bellman function for the weak-type (p, p) estimate in the case $\gamma(c) = 0 < c$.

of γ. Thus, we see that

$$B(y^*-1,1)=F\left(1-\frac{(p-2)y_*}{p-1}\right)=F(0)+F'(0)\left(1-\frac{(p-2)y^*}{p-1}\right).$$

However, by the assumptions (A1), (A3) and the fact that $(y_*/(p-1),(p-2)y_*/(p-1))$ lies on the linear piece of γ, we know what $F(0)$ and $F'(0)$ are. Namely, $F(0)=-\beta\,(y_*/(p-1))^p$ and

$$F'(0)=-B_x\left(\frac{y_*}{p-1},\frac{(p-2)y_*}{p-1}\right)+B_y\left(\frac{y_*}{p-1},\frac{(p-2)y_*}{p-1}\right)=p\beta\left(\frac{y_*}{p-1}\right)^{p-1}.$$

Therefore, the preceding equality can be rewritten in the form

$$\beta=\left[(1-y_*)^p-y_*^p(p-1)^{2-p}+py_*^{p-1}(p-1)^{1-p}\right]^{-1}.$$

Now, if we replace y_* by a certain number y (close to y_*), a similar argument yields

$$\beta\geq\left[(1-y)^p-y^p(p-1)^{2-p}+py^{p-1}(p-1)^{1-p}\right]^{-1}. \tag{4.18}$$

Indeed, this inequality is equivalent to saying that the line, tangent to the curve $t\mapsto V_\beta(\frac{y}{p-1}-t,\frac{(p-2)y}{p-1}+t)$ at the point $t=0$, must majorize $t\mapsto B_\beta(\frac{y}{p-1}-t,\frac{(p-2)y}{p-1}+t)$ for $t=1-(p-2)y/(p-1)$. Thus, we see that the derivative of the expression in the square brackets of (4.18) must vanish for $y=y_*$. This equality is equivalent to

$$(1-y_*)\left(\left(\frac{1-y_*}{y_*}\right)^{p-2}-(p-1)^{2-p}\right)=0,$$

and hence $y_*=1-p^{-1}$. This, in turn, implies $\beta=p^{p-1}/2$.

Step 5. A final assumption. We still work under the condition $c=\gamma(c)=0$. Observe that the segments I^+ and I^-, introduced in the previous step (see also Figure 3), have the same length. This suggests the final assumption (A5) below, to formulate which we need some notation. Introduce the curve

$$\kappa=\left\{\left(x-\frac{1-\gamma(x)}{2},\gamma(x)+\frac{1-\gamma(x)}{2}\right):x\geq p^{-1}\right\}$$

(for the better understanding of κ, see the geometric properties of $I^+(z)$ and $I^-(z)$ below). Let $D_1\subset J$ be the closed set bounded by the lines $y=1$, $-x+y=y_*$ and the curve κ; let $D_2\subset J$ be the closed set bounded by the line $x+y=y_*$, the curve κ and the graph of γ (see Figure 3). Note that D_1 and D_2 have the following property. Take any $z\in\kappa$ and let $I^+(z)\subset D_1$ (respectively, $I^-(z)\subset D_2$) denote the maximal line segment of slope 1 (respectively, -1), which contains z as one of its endpoints. Then $I^+(z)$ and $I^-(z)$ have the same length; so, in a sense, κ divides the set

$$\{(x,y):y_*-x\leq y\leq y_*+x,\ \gamma(x)\leq y\leq 1\}$$

into two halves. The assumption can be stated as follows.

(A5) We assume that

$$B\text{ is linear on each }I^+(z),\ z\in\kappa \tag{4.19}$$

and

$$B\text{ is linear on each }I^-(z),\ z\in\kappa. \tag{4.20}$$

Step 6. A formula for γ in the case $c=\gamma(c)=0$. So far, we have derived that

$$\gamma(x)=(p-2)x \qquad \text{for } x\in[0,p^{-1}]$$

and we need to determine this function on the remaining part of the positive halfline. By (A3) and (4.20), the equation (4.17) is valid for all $x\geq p^{-1}$ and $t\in[0,(1-\gamma(x))/2]$. This enables us to derive $B_x(z)+B_y(z)$ for any $z\in\kappa$: if $z=(x-(1-\gamma(x))/2,\gamma(x)+(1-\gamma(x))/2)$, then

$$B_x(z)+B_y(z)=-p\beta x^{p-1}+\frac{p(p-1)\beta x^{p-2}(1-\gamma(x))}{1+\gamma'(x)}.$$

On the other hand, by (A3) and (4.19), this must be equal to

$$\frac{B(x,1)-B(z)}{\frac{1-\gamma(x)}{2}}.$$

After some easy manipulations, this yields

$$\gamma'(x)+1=\frac{p^p(p-1)}{4}x^{p-2}(1-\gamma(x))^2. \tag{4.21}$$

Standard argumentation (cf. [44]) give the existence of a unique $\gamma:(p^{-1},\infty)\to[0,1]$ satisfying $\gamma(p^{-1}+)=1-2/p$; then $\gamma'(p^{-1}+)=p-2$. Thus we have obtained the desired function γ.

Step 7. The formula for B, the case $c=\gamma(c)=0$. We put all the things together: the equations (4.12), (4.13), (4.19) and (4.20) yield the candidate B, the one invented by Suh. Let $D_0=\{(x,y):x>0,\,y\geq 1\}$ and recall D_1 and D_2 introduced in Step 5. Moreover, let

$$\begin{aligned}
D_3&=\{(x,y):x\geq 0,\,0\leq y\leq\gamma(x)\},\\
D_4&=\{(x,y):x\geq 0,\,\gamma(x)\leq y\leq -x+y_*\},\\
D_5&=\{(x,y):\mathbb{R}_+\times\mathbb{R}_+\setminus(D_0\cup D_1\cup D_2\cup D_3\cup D_4)
\end{aligned}$$

(see Figure 3). Suppose that G is the inverse to the function $x\mapsto x+\gamma(x)$. Then

$$B(x,y)=\begin{cases}
1-\frac{p^{p-1}}{2}x^p & \text{on } D_0,\\
1-\frac{2(1-y)}{1-\gamma(x-y+1)} & \\
\quad -\frac{1}{2}p^{p-1}(x-y+1)^{p-1}\big(x-(p-1)(1-y)\big) & \text{on } D_1,\\
\frac{p^{p-1}}{2}(G(x+y))^{p-1}\left((p-1)G(x+y)-px\right) & \text{on } D_2,\\
-\frac{p^{p-1}}{2}x^p & \text{on } D_3,\\
\frac{1}{2}\left(\frac{p}{p-1}\right)^{p-1}(x+y)^{p-1}(y-(p-1)x) & \text{on } D_4,\\
\frac{p^{p-1}}{2}(1+x-y)^{p-1}\left(\frac{1-y}{p-1}-x\right)-\frac{p^2(1-y)}{2(p-1)}+1 & \text{on } D_5.
\end{cases} \tag{4.22}$$

The description of B is completed by the condition (4.10). One can now check that the function satisfies the conditions 1°, 2° as well as 3°: $B(x,\pm x)\leq 0$ for $x\in\mathbb{R}$. However, this requires a large amount of work and patience; for details, see [44].

This gives us the Bellman function (4.9) for $\beta=p^{p-1}/2$. Furthermore, we have obtained that the best C_p in (4.1) equals $(p^{p-1}/2)^{1/p}$.

Step 8. The formula for $\mathcal{B}^0$ for $\beta>p^{p-1}/2$. There is a very natural modification of Suh's function (4.22), which can be obtained by appropriate scaling. This object will correspond to $\beta > p^{p-1}/2$ and the case $c = 0$, $\gamma(c) > 0$. Namely, for any $\alpha \in (0,1)$, consider the function

$$B_\alpha(x,y) = \begin{cases} B\left(|x|/\alpha, (|y|-1+\alpha)/\alpha\right) & \text{if } |y| \geq 1-\alpha, \\ -\frac{p^{p-1}|x|^p}{2\alpha^p} & \text{if } |y| < 1-\alpha. \end{cases}$$

Obviously, this function is of class C^1 and diagonally concave (the rescaling factor α^{-1} occurs on both coordinates in the definition of B_α). Furthermore, we easily check that B_α majorizes V_β corresponding to $\beta = p^{p-1}/(2\alpha^p)$, and hence $\mathcal{B}^0 \leq B_\alpha$. However, it is also evident that the reverse holds true: if the function $\mathcal{B}^0$ was strictly smaller at some point (x,y), then the appropriate improvement of B would also be possible.

Step 9. The formula for $\mathcal{B}^0$ for $\beta < p^{p-1}/2$. In this case, the Bellman function is also closely related to Suh's function (4.22), but there is a little more to do. Again, the candidate comes into one's mind after taking a closer look at the picture: see Figure 5. Namely, consider the following scaling of B. Let $\alpha \in (2/p, 1)$ be a fixed parameter and, for any x, $y \in \mathbb{R}$ put

$$B_\alpha(x,y) = B\left(\alpha|x|, \alpha|y| + 1 - \alpha\right).$$

It follows at once from the majorization property of B that $B_\alpha(x,y) \geq 1_{\{|y|\geq 1\}} - \alpha^p p^{p-1}|x|^p/2$. Unfortunately, this function does not work, as the diagonal concavity fails to hold: for $|x| < \frac{1-\alpha}{(p-2)\alpha}$, the function $\xi(t) = B_\alpha(x+t,t)$ satisfies $\xi'(0-) < \xi'(0+)$. To overcome this problem, we modify B_α on the square

$$\left\{(x,y) \in \mathbb{R}^2 : |x| + |y| \leq \frac{1-\alpha}{(p-2)\alpha}\right\},$$

putting $B_\alpha(x,y) = \kappa_1(y^2 - x^2) + \kappa_2$ there. The parameters κ_1, κ_2 depend only on p and α, and we determine them by requiring that B_α is continuous. We obtain

$$\kappa_1 = \frac{p^p(1-\alpha)^{p-2}\alpha^2}{4(p-2)^{p-2}} \qquad \text{and} \qquad \kappa_2 = \frac{p^{p-1}(1-\alpha)^p}{4(p-2)^{p-1}}.$$

It turns out that if $\alpha \geq 2/p$, then the modified B_α has the required concavity (one easily verifies that the partial derivatives match appropriately at the boundary of the square). Thus B_α majorizes the Bellman function (4.9) corresponding to $\beta = \alpha^p p^{p-1}/2$. However, it is clear that we actually have equality (for instance, by considering appropriate examples). Let us also mention that, as a by-product, we get the sharp inequality

$$\mathbb{P}(|g_n| \geq 1) \leq \frac{1}{2}\alpha^p p^{p-1}\mathbb{E}|f_n|^p + \frac{p^{p-1}(1-\alpha)^p}{4(p-2)^{p-1}},$$

valid for $\alpha \in [2p^{-1}, 1]$. Now it is also clear that c must not exceed $1/2$; otherwise, the line segments $I^\pm$ would not fit into the picture: see Figure 5.

Question 2: What is the formula for $\mathcal{B}^0$ for $\beta < 2^{p-1}/p$?

4.3. **One sided bound, $p > 2$.** We turn our attention to the one-sided version of (4.9). Consider the function

$$\mathcal{B}^0(x,y) = \sup\left\{\mathbb{E}V(f_n, g_n) \ : \ (f,g) \in M(x,y),\, n = 0,\, 1,\, 2,\, \ldots\right\},$$

where $V(x,y)=1_{\{y\geq 0\}}-|x|^p$. It is more convenient to study first the situation when p is large; then the special function can be easily deduced from Suh's function, with the use of appropriate scaling. Namely, put

$$b(x,y)=\begin{cases} B\left(\left(\frac{2}{p^{p-1}}\right)^{1/p}x, \left(\frac{2}{p^{p-1}}\right)^{1/p}y+1\right) & \text{if } y\geq -(p^{p-1}/2)^{1/p}, \\ -|x|^p & \text{if } y< -(p^{p-1}/2)^{1/p}, \end{cases} \tag{4.23}$$

where B is given by (4.22). We easily check that b is of class C^1 on $\mathbb{R}\times(-\infty,0)$ and hence, by the properties of B, it is diagonally concave. Furthermore, $b(x,y)\geq V(x,y)$ for all $(x,y)\in\mathbb{R}^2$: this is obvious for $y<-(p^{p-1}/2)^{1/p}$, and for remaining points it follows immediately from the corresponding majorization for B. Thus, we have $\mathcal{B}^0\leq b$. On the other hand, if we had a strict inequality at some point (x,y), then necessarily we would have $y>-(p^{p-1}/2)^{1/p}$, and then the function B would not be the Bellman function for (4.9).

Clearly, the above arguments, combined with appropriate translation and homogenization, give the formula for the "shifted" function

$$(x,y)\mapsto \sup\left\{\mathbb{P}(|g_n|\geq 1)-c\mathbb{E}|f_n|^p\ :\ (f,g)\in M(x,y),\ n=0,\,1,\,2,\,\ldots\right\},$$

for any fixed c. Now it is easy to see that this new function satisfies 3° if and only if $c\geq p^{p-1}/2$, and thus we obtain the sharp one-sided bound

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

valid for all martingales f and their ± 1-transforms g.

4.4. **One sided bound,** $1<p\leq 2$**.** Now we deal with the more interesting case of small p. The discovery of the corresponding Bellman function is not difficult, especially in the light of the above considerations. Namely, a closer inspection of the arguments from the case $p=1$ and $p>2$ suggests the appropriate candidate. We will require the following auxiliary result, see [33] for the proof.

Theorem 4.1. *Let $1<p<2$. There is a continuous function $H=H_p:[0,\infty)\to[0,\infty)$ satisfying the differential equation*

$$H'(x)=\frac{p(p-1)}{2}x^{p-2}(H(x)-x)^2 \tag{4.24}$$

for $x>0$ and such that

$$H(0)=\frac{\left(\frac{2p}{p-1}\right)^{1/p}\Gamma\left(\frac{p+1}{p}\right)}{\Gamma\left(\frac{p-1}{p}\right)}. \tag{4.25}$$

We proceed as in the previous subsection: the desired Bellman function $\mathcal{B}^0$ is defined by the same formula. To define the candidate b, let us first distinguish certain subsets of $\mathbb{R}\times\mathbb{R}$. Here $H=H_p$ is the function studied in the preceding theorem and $h = h_p$ stands for its inverse.

$$\begin{aligned}
D_0 &= \{(x, y) : y \geq 0\},\\
D_1 &= \{(x, y) : h(|x| - y) - y < |x|,\ -|x| \leq y < 0\},\\
D_2 &= \{(x, y) : h(|x| - y) < |x| \leq h(|x| - y) - y,\ y - |x| \leq -H(0)\},\\
D_3 &= \{(x, y) : |x| - H(0) < y < -|x|\},\\
D_4 &= \{(x, y) : |x| \leq h(|x| - y)\}.
\end{aligned}$$

See Figure 4.4 below. What is the reason for such regions and what can be said

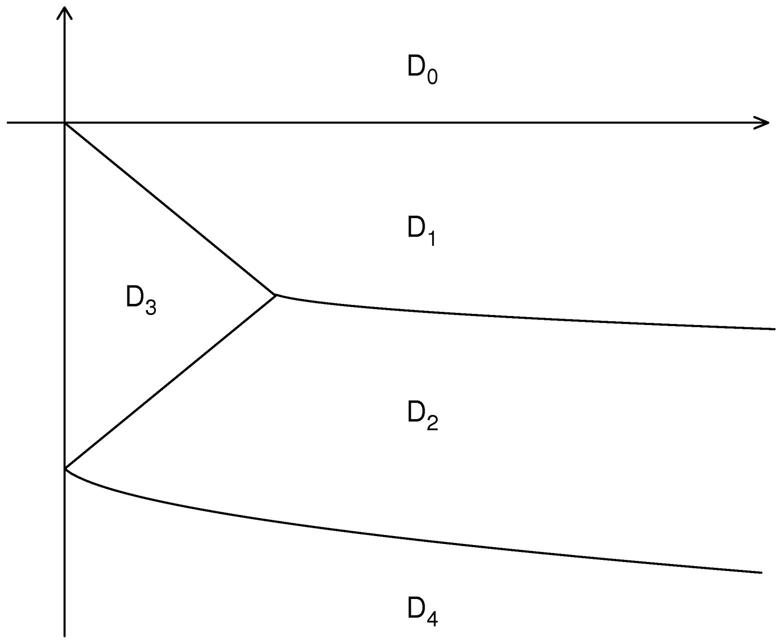

FIGURE 6. The regions D_0–D_4, intersected with $\{(x, y) : x \geq 0\}$.

about B? Arguing as above, we immediately get that $b(x, y) = 1 - |x|^p$ for $y \geq 0$ (this gives the justification for D_0). Let us describe the idea behind the shape of D_4. To compute $\mathcal{B}^0(x, y)$, we need pairs (f, g) for which $\mathbb{P}(g_n \geq 0)$ is large, while $\mathbb{E}|f_n|^p$ is small. For a given x, if y is sufficiently small, then the best pair is the constant one; for any other pair, the increase of the p-th moment of f is not compensated by the gain coming from pushing g on or above the x-axis. Thus we expect $\mathcal{B}^0(x, y) = -|x|^p$ for small y. On the other hand, fix such a y and start increasing x. Since p is smaller than 2, the second derivative of $x \mapsto x^p$ goes to 0; consequently, the p-th moment $\mathbb{E}|f_n|^p$ grows slower and slower. This means that for sufficiently large x, some nontrivial pairs (f, g) should be taken into consideration. Summarizing, the above reasoning suggests the existence of a certain nonincreasing function $\gamma : [0, \infty) \to (-\infty, 0)$, such that $\mathcal{B}^0(x, y) = -|x|^p$ for $y \leq \gamma(x)$ and $\mathcal{B}^0(x, y) > -|x|^p$ for $y > \gamma(x)$. To get the description of γ, we impose the following geometrical property on the sets D_1 and D_2 (see Figure 4.4). Pick a point $(x, y) \in \partial D_1 \cap D_2$, with $x > 0$, and consider the maximal line segment of slope -1 (respectively, $+1$) contained in D_1 (respectively, D_2) and having (x, y) as one of its endpoints. The requirement is that both these segments have the same length. Assuming that the Bellman function is continuous on $\mathbb{R}^2$ and of class C^1 on $\mathbb{R} \times (-\infty, 0)$, we obtain a differential equation for γ:

$$1 - \gamma'(x) = \frac{p(p-1)}{2} x^{p-2} \gamma(x)^2,$$

which should be compared to its counterpart (4.21) in the case $p \geq 2$. The substitution $H(x) = x - \gamma(x)$ transfers this equation to (4.24), and hence the function γ does exist. Furthermore, the solution satisfies

$$\gamma(0) = \frac{\left(\frac{2p}{p-1}\right)^{1/p} \Gamma\left(\frac{p+1}{p}\right)}{\Gamma\left(\frac{p-1}{p}\right)},$$

see (4.25). Where does this equality come from? It is a consequence of the fact that we want γ to be defined on the whole halfline $[0, \infty)$ and to satisfy $\gamma'(t) \to 0$ as $t \to \infty$ (all the other solutions γ, corresponding to different initial conditions, violate one of these two properties).

We come back to the search for the candidate b. It is clear what foliation we should use. On $D_1 \cap \{(x, y) : x > 0\}$, the function should be concave along the lines of slope -1. The set $D_2 \cap \{(x, y) : x > 0\}$ should be split into segments of slope 1. Finally, on D_3 we expect linearity along line of both slopes ± 1 (i.e., Bellman function should be quadratic there). A little calculation reveals the following candidate. Put $H(x) = x - \gamma(x)$ for $x > 0$, and let h be the inverse function to H. Then

$$(4.26) \qquad b(x, y) = \begin{cases} 1 - x^p & \text{if } (x, y) \in D_0, \\ 1 + (x+y)^{p-1}((p-1)y - x) + \frac{2y}{H(x+y) - x - y} & \text{if } (x, y) \in D_1, \\ [h(x-y)]^{p-1}[(p-1)h(x-y) - px] & \text{if } (x, y) \in D_2, \\ [(y + H(0))^2 - x^2](H(0))^{-2} & \text{if } (x, y) \in D_3, \\ -x^p & \text{if } (x, y) \in D_4. \end{cases}$$

Some lengthy calculations (cf. [33]) show that this function enjoys the conditions 1° and 2°, so it coincides with $\mathcal{B}^0$. As in the case $p > 2$, by appropriate translation and homogenization, we obtain the sharp one-sided bound

$$\mathbb{P}(g_n \geq 1) \leq \frac{\left(\frac{2p}{p-1}\right)^{1/p} \Gamma\left(\frac{p+1}{p}\right)}{\Gamma\left(\frac{p-1}{p}\right)} \mathbb{E}|f_n|^p, \qquad n = 0, 1, 2, \ldots,$$

valid for all $1 < p < 2$ and all martingale pairs (f, g) such that g is a ± 1-transform of f.

4.5. **More exact information on weak-type bounds.** The final part of this section concerns the Bellman functions

$$\mathcal{B}^0(x, y, t) = \sup\left\{\mathbb{P}(g_n \geq 1) \ : \ (f, g) \in M(x, y),\, \mathbb{E}|f_\infty|^p \leq t,\, n = 0, 1, 2, \ldots\right\}$$

and the two-sided version

$$\begin{aligned}&\mathcal{B}^0(x,y,t)\\ (4.27)\qquad &=\sup\left\{\mathbb{P}(|g_n|\geq 1)\ :\ (f,g)\in M(x,y),\ \mathbb{E}|f_\infty|^p\leq t,\ n=0,\,1,\,2,\,\ldots\right\}.\end{aligned}$$

We will be brief. The first of these functions can be easily extracted from the analysis of the above one-sided bounds. We simply repeat the translation/homogenization arguments from the case $p=1$. Unfortunately, the formula we obtain is complicated and non-explicit, so we have decided not to include it here. The function (4.27) could be handled similarly, but, unfortunately, in both cases $1<p<2$, $p>2$ the analysis of (4.9) is incomplete (there are some values of β for which the function has not been found) and this disables the identification of (4.27).

We omit the further details in this direction, and leave them to the reader.

5. Strong-type inequalities for martingale transforms

We turn to Burkholder's celebrated L^p-estimates for martingale transforms:

$$(5.1)\qquad ||g_n||_p\leq (p^*-1)||f_n||_p,\qquad 1<p<\infty,\ n=0,\,1,\,2,\,\ldots,$$

where $p^*=\max\{p,p/(p-1)\}$. Thus,

$$p^*-1=\begin{cases}(p-1)^{-1} & \text{if } 1<p\leq 2,\\ p-1 & \text{if } p\geq 2.\end{cases}$$

The primary goal of this section is to determine the explicit formula for the associated Bellman function

$$(5.2)\quad \mathcal{B}^0(x,y,t)=\sup\left\{\mathbb{E}|g_n|^p\ :\ (f,g)\in M(x,y),\ \mathbb{E}|f_\infty|^p\leq t,\ n=0,\,1,\,2,\,\ldots\right\},$$

originally invented by Burkholder in [8]. The reasoning we will present is a combination of probabilistic and analytic arguments. For an alternative, analytic approach which exploits the Monge-Ampère equation and the method of characteristics, see the recent work of Vasyunin and Volberg [50].

5.1. **On the search of extremal examples.**

We start our analysis from some natural examples, which will be helpful later. Let $1<p<\infty$ be fixed and suppose that $\beta_p\geq 1$ is the best constant in the inequality

$$(5.3)\qquad ||g_n||_p\leq \beta_p||f_n||_p,\qquad n=0,\,1,\,2,\,\ldots,$$

where f is a martingale and g is its ± 1-transform. The case $p=2$ is trivial: we have $||g_n||_2=||f_n||_2$ for each n and hence the best constant is equal to 1. Thus, suppose that $p\neq 2$. Since β_p cannot be improved, there must be martingales f, g for which both sides are equal, or asymptotically equal. A natural idea during the search for f and g is to let $n\to\infty$ and look for sequences which satisfy the *pointwise* equality $|g_\infty|\equiv\beta_p|f_\infty|$. The construction of such a pair is not difficult. Actually, from the viewpoint of our further reasoning, it will be more convenient to introduce the whole family of Markov martingales. Fix a small $\delta>0$ (eventually, we will let it go to 0), and consider the following transition function:

(i) The states lying in the set $\{(x,y):|y|\geq\beta_p|x|\}$ are absorbing.
(ii) The state (x,y) with $0<y<\beta_p x$, leads to $(x+y,0)$ or to $\left(\frac{x+y}{\beta_p+1},\frac{\beta_p(x+y)}{\beta_p+1}\right)$ (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 $\left(\frac{x}{\beta_p+1},-\frac{\beta_p x}{\beta_p+1}\right)$ (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).

Now if we assume that the pair (f,g) starts from $(1,1)$ and moves according to (i)-(iv), then indeed $|g_\infty|=\beta_p|f_\infty|$. At the first glance, this seems to show that the L^p bound does not hold with any β_p, since the ratio $|g_\infty|/|f_\infty|$ can be arbitrarily large. However, this is not the case: the above martingales are bounded in L^p if and only if $\beta_p<(p-1)^{-1}$, and we have

$$\lim_{\beta_p\uparrow(p-1)^{-1}}||f_\infty||_p=\infty$$

(cf. pages 55-56 in [33], or modify slightly the reasoning on pages 669-670 in [8]). That is, if the constant is at least $(p-1)^{-1}$, then both sides of (5.3) are infinite and hence the estimate holds true. On the other hand, the above example shows the lower bound $\beta_p\geq(p-1)^{-1}$. This is of value only for $1<p<2$, for $p>2$, the stronger bound $\beta_p\geq 1$ is given for free. How to modify the example when $p>2$? A good idea is to reflect the above transition function with respect to the diagonal $y=x$, that is, let (i)-(iv) describe the evolution of (g,f), not (f,g). Actually, a small modification is required: without it, the points of the diagonal are absorbing, so the above recipe would give us constant pairs (g,f) (for which we only have $||g||_p/||f||_p=1$). We modify the example as follows: we assume that (f,g) starts from (x,x) for some $x>0$, then make it move to $(2x,0)$ or to $(0,2x)$, and then require that (g,f) moves according to (i)-(iv). Then, as previously: $f\in L^p$ if and only if $\beta_p\geq(p-1)^{-1}$, and hence

$$\lim_{\beta_p\downarrow(p-1)^{-1}}||g||_p/||f||_p=p-1.$$

Thus, the best constant is at least $p-1$. Actually, we can consider the above Markov family for various choices of $\beta_p\in((p-1)^{-1},\infty)$. As we shall see, they will also play a role.

5.2. **Basic inequality.** In order to study the (difficult) Bellman function $\mathcal{B}^0$, we implement our "splitting procedure": as previously, we move the p-th moment $\mathbb{E}|f_\infty|^p$ from the assumption to the optimized expression and work with an appropriate family of estimates. More precisely, we will determine the explicit formula for the two-dimensional Bellman function

$$b^0(x,y)=\sup\left\{\mathbb{E}V_\beta(f_n,g_n)\;:\;(f,g)\in M(x,y),\,n=0,\,1,\,2,\,\dots\right\},\tag{5.4}$$

where $V_\beta(x,y)=|y|^p-\beta^p|x|^p$ and β is a given constant. Then, optimizing over β, we will obtain the desired function (5.2). Actually, as we shall see in a moment in §5.3, only the case $p\neq 2$ is of interest; when $p=2$, the Bellman function $\mathcal{B}^0$ can be derived with practically no effort.

We will present the detailed reasoning in the case $1<p<2$ only. As we have proved above, only the case $\beta\geq(p-1)^{-1}$ is of interest; for smaller β, the Bellman function b^0 is infinite. So, assume that $\beta\geq(p-1)^{-1}$ and suppose that b^0 is finite. This function satisfies 1° and 2°; moreover, directly from (5.4), we have that it is homogeneous of order p (since V has this property):

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

Consider the function $w : \mathbb{R} \to \mathbb{R}$, given by

$$w(x) = b^0(x, 1-x). \tag{5.5}$$

It is enough to determine w on $[0, 1]$; then, by homogeneity, we will obtain b^0 on its whole domain. Note that w is concave and satisfies

$$w(x) = b^0(x, 1-x) = b^0(x, x-1) = (2x-1)^p w\left(\frac{x}{2x-1}\right) \qquad \text{for } x > 1. \tag{5.6}$$

Furthermore, it majorizes $u : \mathbb{R} \to \mathbb{R}$ given by $u(x) = V_\beta(x, 1-x)$. By the direct differentiation, we see that

$$u''(x) = p(p-1)\big[(1-x)^{p-2} - \beta^p x^{p-2}\big], \qquad x \in (0,1),$$

and hence u is concave on $(0, x_0)$ and convex on $(x_0, 1)$ for some $x_0 \in (0, 1)$. This suggests to consider the following candidate for w:

$$w(x) = \begin{cases} u(x) & \text{if } x \in (0, x_1), \\ u'(x_1)(x - x_1) + u(x_1) & \text{if } x \in [x_1, 1), \end{cases}$$

where $x_1 < x_0$ is a parameter which needs to be specified. To find x_1, we compare the behavior of the left- and the right-hand derivative of w at 1. Namely, note that (5.6) implies

$$\frac{w(x) - w(1)}{x-1} + (2x-1)^{p-1} \frac{w(1) - w\left(\frac{x}{2x-1}\right)}{1 - x/(2x-1)} = 2w(1)\frac{(2x-1)^p - 1}{(2x-1) - 1}$$

for $x > 1$, so letting $x \downarrow 1$, we get $w'(1+) + w'(1-) = 2pw(1)$ and thus

$$u'(x_1) = w'(1-) \geq \frac{w'(1+) + w'(1-)}{2} = pw(1) = pu'(x_1)(1 - x_1) + pu(x_1).$$

Now, it will be convenient to switch to the parameter γ, given by $x_1 = (1 + \gamma)^{-1}$. Then the above inequality is equivalent to $(p-1)\beta^p \geq (2-p)\gamma^{p-1} + \gamma^{p-2}$. We *assume* that we actually have equality here, so,

$$\beta^p = \frac{(2-p)\gamma^{p-1} + \gamma^{p-2}}{p-1}. \tag{5.7}$$

Note that the expression on the right of (5.7), considered as a function of γ, is decreasing on $(0, (p-1)^{-1})$ and increasing on $((p-1)^{-1}, \infty)$. Furthermore, its minimal value equals $(p-1)^{-p}$. Thus, if β is strictly larger than $(p-1)^{-1}$, we have two choices for γ, and in such a case we pick the smaller one, i.e., $\gamma < (p-1)^{-1}$.

Coming back to b^0, we obtain the following candidate for the Bellman function:

$$b(x, y) = \begin{cases} \left(\dfrac{\gamma}{\gamma+1}\right)^{p-2} (|x| + |y|)^{p-1} \left(|y| - \dfrac{|x|}{p-1}\right) & \text{if } |y| \leq \gamma|x|, \\ |y|^p - \beta^p |x|^p & \text{if } |y| > \gamma|x|. \end{cases}$$

Now, it is straightforward to check that b satisfies 1° (this follows directly from the homogeneity and the inequality $w \geq u$). Furthermore, some tedious calculations show that the condition 2° is also satisfied: see e.g. page 17 in [10]. Therefore, $b^0 \leq b$. Actually, we have equality $b^0(x, y) = b(x, y)$ for all (x, y), which can be verified with the use of examples from the previous section, with $\beta_p := \gamma$ (or just follows from the above construction). This is the place where we use the fact that we have chosen smaller γ in (5.7): for the larger choice, the computation of $\sup_{n \geq 0} \mathbb{E}V(f_n, g_n)$ would not lead to $b(x, y)$, but to a strictly smaller constant. We have yet another analytic explanation for the smaller γ's: the function b corresponding to larger choice would lead to a strict majorant of b^0.

In the case $p > 2$ we proceed similarly; let us briefly describe the main steps of the analysis. Exploitation of the examples of §5.1 leads us to the inequality $\beta \geq p-1$. The function u (given by the same formula as above), restricted to $[0,1]$, has dual convexity/concavity regions: that is, it is convex for small arguments and concave for large ones (i.e., close to 1). Thus, it is natural to conjecture that the restriction $w|_{[0,1]}$ (where w is given by (5.5)) is of the form

$$w(x) = \begin{cases} u'(x_1)(x-x_1) + u(x_1) & \text{if } x \in (0, x_1], \\ u(x) & \text{if } x \in (x_1, 1), \end{cases}$$

for some x_1. Passing to $\gamma = 1 - x_1^{-1}$ as above and analyzing the behavior of w' in the neighborhood of 0, we obtain $\beta^p \geq (p-1)\gamma^{p-1}/(\gamma+2-p)$. We *assume* that both sides are equal: again, for a given $\beta > p-1$ there are two γ's which satisfy the equation, and this time we pick the larger one. Putting all these facts together, we obtain the candidate

$$b(x,y) = \begin{cases} \dfrac{\gamma^{p-1}(\gamma+1)^{2-p}}{\gamma+2-p}(|x|+|y|)^{p-1}\big(|y|-(p-1)|x|\big) & \text{if } |y| \geq \gamma|x|, \\ |y|^p - \beta^p|x|^p & \text{if } |y| < \gamma|x|, \end{cases} \tag{5.8}$$

where γ is the larger positive root of the equation

$$\beta^p = \frac{(p-1)\gamma^{p-1}}{\gamma+2-p}.$$

Similar arguments to those above give $b = b^0$, which completes the analysis in the case $p > 2$.

5.3. **Burkholder's function.** We are ready to compute the formula for the general Bellman function $\mathcal{B}^0$ introduced in (5.2). Pick a point (x,y,t) and a pair (f,g) as in the definition of $\mathcal{B}^0(x,y,t)$. In the case $p=2$, everything is straightforward: we have

$$\mathbb{E}|g_n|^2 = y^2 + \sum_{k=1}^n |dg_k|^2 = y^2 + \sum_{k=1}^n |df_k|^2 = y^2 - x^2 + \mathbb{E}|f_n|^2 \leq y^2 - x^2 + t,$$

so $\mathcal{B}^0(x,y,t) \leq y^2 - x^2 + t$. On the other hand, if we pick any simple pair $(f,g) \in M(x,y)$ for which $\mathbb{E}|f_\infty|^2 = t$, we get equality. Thus, for $p=2$, we have

$$\mathcal{B}^0(x,y,t) = |y|^2 - |x|^2 + t.$$

Next, let us assume that $1 < p < 2$. Directly from the above considerations, we can write that

$$\mathcal{B}^0(x,y,t) \leq b^0(x,y) + \beta_p^p t.$$

Therefore, in the light of (5.7), $\mathcal{B}^0(x,y,t)$ does not exceed

$$\begin{cases} (1+\gamma^{-1})^{2-p}(|x|+|y|)^{p-1}\left(|y| - \dfrac{|x|}{p-1}\right) + \dfrac{(2-p)\gamma^{p-1}+\gamma^{p-2}}{p-1}t & \text{if } |y| \leq \gamma|x|, \\ |y|^p + \dfrac{(2-p)\gamma^{p-1}+\gamma^{p-2}}{p-1}(t-|x|^p) & \text{if } |y| > \gamma|x|. \end{cases}$$

Denote this expression by $B(x,y,t,\gamma)$ and let us minimize it over all $\gamma \leq (p-1)^{-1}$. If $|y| \geq (p-1)^{-1}|x|$, then

$$B(x,y,t,\gamma) = |y|^p + \frac{(2-p)\gamma^{p-1} + \gamma^{p-2}}{p-1}(t - |x|^p)$$

for all γ. We easily see that

$$\frac{\partial B}{\partial \gamma}(x,y,t,\gamma) = (2-p)(p-1)^{-1}\gamma^{p-3}(-1+(p-1)\gamma) \leq 0$$

and thus the choice $\gamma = (p-1)^{-1}$ is optimal. This calculation also shows that if $|y| < (p-1)^{-1}|x|$, then $\gamma \mapsto B(x,y,t,\gamma)$ is nonincreasing on $(0, |y|/|x|)$, and hence the minimum is attained on $[|y|/|x|, (p-1)^{-1}]$. A direct differentiation gives that γ_0 which minimizes B is the unique number γ from that interval, satisfying the equation

$$\frac{(|x|+|y|)^{p-1}}{t}\left(|y| - \frac{|x|}{p-1}\right) - (1+\gamma)^p + \frac{p}{p-1}(1+\gamma)^{p-1} = 0.$$

Plugging the above optimal choices of γ into the definition of B, we finally get that

$$\mathcal{B}^0(x,y,t) \leq \begin{cases} |y|^p + (p-1)^{-p}(t-|x|^p) & \text{if } |y| \geq (p-1)^{-1}|x|, \\ \gamma_0^p t & \text{if } |y| < (p-1)^{-1}|x|. \end{cases}$$

We actually have equality here, which can be verified with the examples considered in the previous subsection (see also [8]). This gives us the desired formula for $\mathcal{B}^0$, which was originally discovered by Burkholder.

It remains to deal with the case $p > 2$. We proceed analogously and exploit the function defined by (5.8). In comparison to the case $1 < p < 2$, there are no additional arguments, so we leave the details to the reader. Let us only write down the result: having carried out all the computations, we end up with

$$\mathcal{B}^0(x,y,t) = \begin{cases} |y|^p + (p-1)^p(t-|x|^p) & \text{if } |y| \leq (p-1)|x|, \\ \gamma_0^p t & \text{if } |y| > (p-1)|x|. \end{cases}$$

where γ_0 is the unique $\gamma \in [p-1, \infty)$ satisfying the equation

$$\frac{(|x|+|y|)^{p-1}}{t}\big(|y| - (p-1)|x|\big) + p(\gamma+1)^{p-1} - (\gamma+1)^p = 0.$$

This completes the analysis.

Acknowledgments

The author would like to thank anonymous Referees for the careful reading of the first version of the paper and many helpful comments and suggestions. This survey is an extended version of the mini-course given at Chebyshev Laboratory in St. Petersburg, Russia. The author would like to express his sincerest gratitude to A. Logunov, L. Slavin, D. Stolyarov, V. Vasyunin and P. Zatitskiy for warm hospitality and enjoyable meeting.

References

[1] R. Bañuelos, A. Bielaszewski and K. Bogdan, *Fourier multipliers for non-symmetric Lévy processes*, Marcinkiewicz centenary volume, Banach Center Publ. **95** (2011), pp. 9–25, Polish Acad. Sci. Inst. Math., Warsaw.

[2] R. Bañuelos and G. Wang, *Sharp inequalities for martingales with applications to the Beurling–Ahlfors and Riesz transformations*, Duke Math. J. **80** (1995), pp. 575–600.

[3] A. Borichev, P. Janakiraman and A. Volberg, *Subordination by conformal martingales in L^p and zeros of Laguerre polynomials*, Duke Math. J. **162** (2013), pp. 889–924.

[4] A. Borichev, P. Janakiraman and A. Volberg, *On Burkholder function for orthogonal martingales and zeros of Legendre polynomials*, Amer. J. Math. **135** (2012), pp. 207–236.

[5] D. L. Burkholder, *Martingale transforms*, Ann. Math. Statist. **37** (1966), pp. 1494–1504.

[6] D. L. Burkholder, *A sharp inequality for martingale transforms*, Ann. Probab. **7** (1979), pp. 858–863.

[7] D. L. Burkholder, *A geometrical characterization of Banach spaces in which martingale difference sequences are unconditional*, Ann. Probab. 9 (1981), 997–1011.

[8] D. L. Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab. **12** (1984), pp. 647–702.

[9] D. L. Burkholder, *Martingales and Fourier analysis in Banach spaces*, Probability and Analysis (Varenna, 1985) Lecture Notes in Math. **1206**, Springer, Berlin (1986), pp. 61–108.

[10] D. L. Burkholder, *Explorations in martingale theory and its applications*, École d'Ete de Probabilités de Saint-Flour XIX—1989, pp. 1–66, Lecture Notes in Math., **1464**, Springer, Berlin, 1991.

[11] D. L. Burkholder, *Strong differential subordination and stochastic integration*, Ann. Probab. **22** (1994), pp. 995–1025.

[12] D. L. Burkholder, *Sharp norm comparison of martingale maximal functions and stochastic integrals,* Proceedings of the Norbert Wiener Centenary Congress, 1994 (East Lansing, MI, 1994), pp. 343–358, Proc. Sympos. Appl. Math. **52**, Amer. Math. Soc., Providence, RI, 1997.

[13] D. L. Burkholder, *The best constant in the Davis' inequality for the expectation of the martingale square function*, Trans. Amer. Math. Soc. **354** (2002), pp. 91–105.

[14] K. P. Choi, *Some sharp inequalities for martingale transforms*, Trans. Amer. Math. Soc. **307** (1988), pp. 279–300.

[15] K. P. Choi, *A sharp inequality for martingale transforms and the unconditional basis constant of a monotone basis in $L^p(0,1)$*, Trans. Amer. Math. Soc. **330** (1993), pp. 509–529.

[16] D. C. Cox, *The best constant in Burkholder's weak-L^1 inequality for the martingale square function*, Proc. Amer. Math. Soc. **85** (1982), pp. 427–433.

[17] D. C. Cox, *Some sharp martingale inequalities related to Doob's inequality*, Inequalities in Statistics and Probability, IMS Lecture Notes-Monograph Series **5** (1984), pp. 78–83.

[18] L. E. Dor and E. Odell, *Monotone bases in L_p*, Pacific J. Math. **60** (1975), pp. 51–61.

[19] I. Doust, *Contractive projections on Banach spaces*, Proc. Centre for Math. Anal., Australian National University **20** (1988), pp. 50–58.

[20] O. Dragičević and A. Volberg, *Bellman function, Littlewood–Paley estimates, and asymptotics of the Ahlfors–Beurling operator in $L^p(\mathbb{C})$, $p > 1$,* Indiana Math. J. **54** (2005), pp. 971–995.

[21] O. Dragičević and A. Volberg, *Bellman function and dimensionless estimates of classical and Ornstein–Uhlenbeck Riesz transforms*, J. of Oper. Theory **56** (2006), pp. 167–198.

[22] O. Dragičević and A. Volberg, *Linear dimension-free estimates in the embedding theorem for Schrödinger operators*, J. Lond. Math. Soc. **85** (2012), pp. 191–222.

[23] P. Ivanishvili, N. N. Osipov, D. M. Stolyarov, V. Vasyunin, P. Zatitskiy, *On Bellman function for extremal problems in BMO*, C. R. Math. Acad. Sci. Paris **350** (2012), pp. 561–564.

[24] J. H. B. Kemperman, *The general moment problem, a geometric approach*, Ann. Math. Statist. **39** (1968), pp. 93–122.

[25] J. Marcinkiewicz, *Quelques théorèmes sur les séries orthogonales*, Ann. Soc. Polon. Math. **16** (1937), pp. 84–96.

[26] A. D. Melas, *Sharp general local estimates for dyadic-like maximal operators and related Bellman functions*, Adv. Math. **220** (2009), pp. 367–426.

[27] A. D. Melas and E. N. Nikolidakis, *Dyadic-like maximal operators on integrable functions and Bellman functions related to Kolmogorov's inequality*, Trans. Amer. Math. Soc. **362** (2010), pp. 1571–1597.

[28] F. Nazarov, A. Reznikov, V. Vasyunin and A. Volberg, A_1 conjecture: weak norm estimates of weighted singular operators and Bellman function, http://sashavolberg.files.wordpress.com/2010/11/a11_7loghilb11_21_2010.pdf.

[29] F. L. Nazarov and S. R. Treil, *The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis,* St. Petersburg Math. J. **8** (1997), pp. 721–824.

[30] F. L. Nazarov, S. R. Treil and A. Volberg, *The Bellman functions and two-weight inequalities for Haar multipliers*, J. Amer. Math. Soc., **12** (1999), pp. 909–928.

[31] F. L. Nazarov, S. R. Treil and A. Volberg, *Bellman function in stochastic optimal control and harmonic analysis (how our Bellman function got its name),* Oper. Theory: Adv. Appl. **129** (2001), pp. 393–424.

[32] F. Nazarov and A. Volberg, *Heating of the Ahlfors–Beurling operator and estimates of its norm*, St. Petersburg Math. J. **15** (2004), pp. 563–573.

[33] A. Osękowski, *Sharp martingale and semimartingale inequalities*, Monografie Matematyczne **72** (2012), Birkhäuser Basel.

[34] A. Osękowski, *Sharp logarithmic inequalities for Riesz transforms*, J. Funct. Anal. **263** (2012), pp. 89–108.

[35] A. Osękowski, *Sharp inequalities for the dyadic square function in the BMO setting*, Acta Math. Hungarica **139** (2013), pp. 85–105.

[36] A. Osękowski, *Some sharp estimates for the Haar system and other bases in* $L^1(0,1)$, to appear in Math. Scand.

[37] R. E. A. C. Paley, *A remarkable series of orthogonal functions*, Proc. London Math. Soc. **34** (1932), pp. 241–264.

[38] A. Pełczyński and H. P. Rosenthal, *Localization techniques in* L^p *spaces*, Studia Math. **52** (1975), pp. 263–289.

[39] G. Peskir and A. Shiryaev, *Optimal Stopping and Free-Boundary Problem*, Lecture in Mathematics, ETH Zurich

[40] J. Schauder, *Eine Eigenschaft des Haarschen Orthogonalsystems*, *Math. Z.* **28** (1928), pp. 317–320.

[41] L. Slavin, A. Stokolos and V. Vasyunin, *Monge–Ampère equations and Bellman functions: The dyadic maximal operator*, C. R. Acad. Sci. Paris, Ser. I **346** (2008), pp. 585–588.

[42] L. Slavin and V. Vasyunin, *Sharp results in the integral-form John–Nirenberg inequality*, Trans. Amer. Math. Soc. **363** (2011), pp. 4135–4169.

[43] L. Slavin and V. Vasyunin, *Sharp* L^p *estimates on BMO*, Indiana Math. J. **61** (2012), pp. 1051–1110.

[44] Y. Suh, *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.

[45] V. Vasyunin, *The exact constant in the inverse Hlder inequality for Muckenhoupt weights (Russian)*, Algebra i Analiz **15** (2003), 73–117; translation in St. Petersburg Math. J. **15** (2004), pp. 49–79.

[46] V. Vasyunin, *Mutual estimates for* L^p*-norms and the Bellman function* (in Russian), Zap. Nauchn. Sem. S.-Petersburg. Otdel. Mat. Inst. Steklov. (POMI) **355** (2008), Issledovaniya po Lineinym Operatoram i Teorii Funktsii. **36**, pp. 81–138, 237–238. Translation in J. Math. Sci. (N. Y.) **156** (2009), no. 5, pp. 766–798.

[47] V. Vasyunin and A. Volberg, *Bellman functions technique in Harmonic Analysis*, sashavolberg.wordpress.com

[48] V. Vasyunin and A. Volberg, *The Bellman function for certain two weight inequality: the case study*, St. Petersburg Math. J. **18** (2007), pp. 201–222.

[49] V. Vasyunin and A. Volberg, *Monge–Ampère equation and Bellman optimization of Carleson Embedding Theorems*, Amer. Math. Soc. Transl. (2), vol. **226**, "Linear and Complex Analysis", 2009, pp. 195-238.

[50] V. Vasyunin and A. Volberg, *Burkholder's function via Monge–Ampère equation*, Illinois J. Math. **54** (2010), pp. 1393–1428.

[51] G. Wang, *Sharp Square-Function Inequalities for Conditionally Symmetic Martingales*, Trans. Amer. Math. Soc., **328** (1991), pp. 393–419.

[52] G. Wang, *Sharp inequalities for the conditional square function of a martingale*, Ann. Probab. **19** (1991), pp. 1679–1688.

[53] G. Wang, *Differential subordination and strong differential subordination for continuous time martingales and related sharp inequalities*, Ann. Probab. **23** (1995), pp. 522–551.

Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

E-mail address: `ados@mimuw.edu.pl`