Monographie Matematyczne
Instytut Matematyczny PAN
Vol. 72
New Series

Adam Osękowski

Sharp Martingale and Semimartingale Inequalities

Birkhäuser

Birkhäuser

Monografie Matematyczne
Instytut Matematyczny Polskiej Akademii Nauk (IMPAN)

Volume 72
(New Series)

Founded in 1932 by
S. Banach, B. Knaster, K. Kuratowski,
S. Mazurkiewicz, W. Sierpinski, H. Steinhaus

Volumes 31–62 of the series
Monografie Matematyczne were published by
PWN – Polish Scientific Publishers, Warsaw

Adam Osękowski

Sharp Martingale and Semimartingale Inequalities

Birkhäuser

Adam Osękowski
Institute of Mathematics
University of Warsaw
Warsaw
Poland

ISBN 978-3-0348-0369-4 ISBN 978-3-0348-0370-0 (eBook)
DOI 10.1007/978-3-0348-0370-0
Springer Basel Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012945527

Mathematics Subject Classification (2010): Primary: 60G42, 60G44, 60G46; secondary: 31B05, 31B20, 42B25, 46B20, 46C15

Printed on acid-free paper

Springer Basel AG is part of Springer Science+Business Media (www.birkhauser-science.com)

To the memory of my Father

Contents

Preface

The purpose of this monograph is to present a unified approach to a certain class of semimartingale inequalities, which have their roots at some classical problems in harmonic analysis. Preliminary results in this direction were obtained by Burkholder in 60s and 70s during his work on martingale transforms and geometry of UMD Banach spaces. The rapid development in the field occurred after the appearance of a large paper of Burkholder in 1984, in which he described a powerful method to handle martingale inequalities and used it to obtain a number of interesting results. Since then, the method has been extended considerably in many directions and successfully implemented in the study of related problems in various areas of mathematics.

The literature on the subject is very large. One of the objectives of this exposition is to put most of the existing results together, explain in detail the underlying concepts and point out some connections and similarities. This book contains also a number of new results as well as some open problems, which, we hope, will stimulate the reader's further interest in this field. The recent applications of the above results in the theory of quasiregular mappings (with deep implications in geometric function theory), Fourier multipliers as well as their connections to rank-one convexity and quasiconvexity indicate the need of further developing this area.

Acknowledgment

I would like to express my gratitude to Professor Stanisław Kwapień for many discussions and encouragement, which motivated me greatly during the work on the monograph. I am also indebted to Professor Quanhua Xu at Université de Franche-Comté in Besançon, France; this is where the research on the subject was initiated. Last but not least, I wish also to thank my whole family, especially my wife Joanna, for love and constant support.

This work was financially supported by MNiSW Grant N N201 422139.

Chapter 1

Introduction

Inequalities for semimartingales appear in both measure-based and noncommutative probability theory, where they play a distinguished role, and have numerous applications in many areas of mathematics. Before we introduce the necessary probabilistic background, let us start with a related classical problem which interested many mathematicians during the first part of the 20th century. The question is: how does the size of a periodic function control the size of its conjugate? To be more specific, let f be a trigonometric polynomial of the form

$$f(\theta) = \frac{a_0}{2} + \sum_{k=1}^{N} \big(a_k \cos(k\theta) + b_k \sin(k\theta)\big), \qquad \theta \in [0, 2\pi),$$

with real coefficients a_0, a_1, a_2, ..., a_N, b_1, b_2, ..., b_N. The polynomial conjugate to f is defined by

$$g(\theta) = \sum_{k=1}^{N} \big(a_k \sin(k\theta) - b_k \cos(k\theta)\big), \qquad \theta \in [0, 2\pi).$$

The problem can be stated as follows. For a given $1 \leq p \leq \infty$, is there a universal constant C_p (that is, not depending on the coefficients or the number N) such that

$$||g||_p \leq C_p ||f||_p\,? \tag{1.1}$$

Here $||f||_p$ denotes the L_p-norm of f, which is given by $\left[\int_0^{2\pi} |f(\theta)|^p \frac{d\theta}{2\pi}\right]^{1/p}$ when p is finite, and equals the essential supremum of f over $[0, 2\pi)$ when $p = \infty$. This question is very easy when $p = 2$: the orthogonality of the trigonometric system implies that the inequality holds with the constant $C_2 = 1$. Furthermore, this value is easily seen to be optimal. What about the other values of p? As shown by M. Riesz in [179] and [180], when $1 < p < \infty$, the estimate does hold with some absolute $C_p < \infty$; for $p = 1$ or $p = \infty$, the inequality does not hold with any finite constant. The best value of C_p was determined by Pichorides [171] and Cole

DOI 10.1007/978-3-0348-0370-0_1,

(unpublished; see Gamelin [79]): $C_p = \cot(\pi/(2p^*))$ is the optimal choice, where $p^* = \max\{p, p/(p-1)\}$.

One may consider a non-periodic version of the problem above. To formulate the statement, we need to introduce the Hilbert transform on the real line. For $1 \le p < \infty$, let $f \in L^p(\mathbb{R})$ and put

$$Hf(x) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_{|y| \ge \varepsilon} \frac{f(x-y)}{y} \mathrm{d}y.$$

This limit can be shown to exist almost everywhere. The transform H is the non-periodic analogue of the harmonic conjugate operator and satisfies the following L^p bound: if $1 < p < \infty$, then

$$||Hf||_p \le C_p ||f||_p,$$

where C_p is the same constant as in (1.1): see [180] and [205].

Riesz's inequality has been extended in many directions. A significant contribution is due to Calderón and Zygmund [41], [42], who obtained the following result concerning singular integral operators. They showed that for a large class of kernels $K : \mathbb{R}^n \setminus \{0\} \to \mathbb{C}$, the limit

$$Tf(x) = \lim_{\varepsilon \to 0} \int_{|y| \ge \varepsilon} f(x-y) K(y) \mathrm{d}y$$

exists almost everywhere if $f \in L^p(\mathbb{R}^n)$ for some $1 \le p < \infty$ and

$$||Tf||_p \le C_p ||f||_p$$

when $1 < p < \infty$. Here the constant C_p may be different from the one in (1.1), but it depends only on K and p.

We may ask analogous questions in the martingale setting. First let us introduce the necessary notation: suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, filtered by $(\mathcal{F}_n)_{n \ge 0}$, a non-decreasing family of sub-σ-fields of $\mathcal{F}$. Assume that $f = (f_n)_{n \ge 0}$, $g = (g_n)_{n \ge 0}$ are adapted discrete-time martingales taking values in $\mathbb{R}$. Then $df = (df_n)_{n \ge 0}$, $dg = (dg_n)_{n \ge 0}$, the difference sequences of f and g, respectively, are defined by $df_0 = f_0$ and $df_n = f_n - f_{n-1}$ for $n \ge 1$, and similarly for dg. In other words, we have the equalities

$$f_n = \sum_{k=0}^{n} df_k \quad \text{and} \quad g_n = \sum_{k=0}^{n} dg_k, \qquad n = 0, 1, 2, \ldots.$$

Let us now formulate the corresponding version of Riesz's inequality. The role of a conjugate function in the probabilistic setting is played by a ± 1-transform, given as follows. Let $\varepsilon = (\varepsilon_k)_{k \ge 0}$ be a deterministic sequence of signs: $\varepsilon_k \in \{-1, 1\}$ for each k. If f is a given martingale, we define the transform of f by ε as

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

Of course, such a sequence is again a martingale. A fundamental result of Burkholder [17] asserts that for any f, ε and g as above we have

$$||g||_p \leq C_p||f||_p, \qquad 1 < p < \infty, \tag{1.2}$$

for some finite C_p depending only on p. Here and below, $||f||_p$ denotes the pth moment of f, given by $\sup_n(\mathbb{E}|f_n|^p)^{1/p}$ when $0 < p < \infty$ and $||f||_\infty = \sup_n \text{essup}|f_n|$. This result is a starting point for many extensions and refinements and the purpose of this monograph is to present a systematic and unified approach to this type of problems. Let us first say a few words about the proof of (1.2). The initial approach of Burkholder used a related weak-type estimate and some standard interpolation and duality arguments. Then Burkholder refined his proof and invented a method which can be used to study general estimates for a much wider class of processes. Roughly speaking, the technique reduces the problem of proving a given inequality to the existence of a certain special function or, in other words, to finding the solution to a corresponding boundary value problem. This is described in detail in Chapter 2 below, and will be illustrated on many examples in Chapter 3. The method can also be used in the case when the dominating process f is a sub- or supermartingale and, after some modifications, allows also to establish general inequalities for maximal and square functions. Another very important feature of the approach is that it enables us to derive the optimal constants: this, except for the elegance, provides some additional insight into the structure of the extremal processes and often leads to some further implications.

Coming back to the martingale setting described above, we shall be particularly interested in the following estimates.

(i) moment inequalities (or strong type (p,p) inequalities): see (1.2),

(ii) weak-type (p,p) inequalities:

$$||g||_{p,\infty} \leq c_{p,p}||f||_p, \qquad 1 \leq p < \infty,$$

where $||g||_{p,\infty} = \sup_{\lambda>0} \lambda\big(\mathbb{P}(\sup_n |g_n| \geq \lambda)\big)^{1/p}$ denotes the weak pth norm of g,

(iii) logarithmic estimates:

$$||g||_1 \leq K \sup_n \mathbb{E}|f_n| \log |f_n| + L(K),$$

(iv) tail and Φ-inequalities:

$$\mathbb{P}(\sup_n |g_n| \geq \lambda) \leq P(\lambda), \quad \sup_n \mathbb{E}\Phi(|g_n|) \leq C_\Phi,$$

under the assumption that $||f||_\infty \leq 1$.

We shall study these and other related problems in the more general setting in which the transforming sequence ε is predictable and takes values in $[-1,1]$. Here by predictability we mean that each term ε_n is measurable with respect to $\mathcal{F}_{(n-1)\vee 0}$ (in particular, it may be random): this does not affect the martingale property of the sequence g. This assumption will be further relaxed to the case when g is a martingale differentially subordinate to f, which amounts to saying that for any $n \geq 0$ we have $|dg_n| \leq |df_n|$ almost surely. We will succeed in determining the optimal constants in most of the aforementioned estimates.

The next challenging problem is to study the above statements in the vector case, when the processes f, g take values from a certain separable Banach space $\mathcal{B}$. It turns out that when we deal with a Hilbert space, the passage from the real case is typically quite easy and does not require much additional effort. However, the extension of a given inequality to a non-Hilbert space setting is a much more difficult problem and the question about the optimal constants becomes hopeless in general. These and related martingale results are dealt with in Chapter 3. Sections 3.1–3.10 concern Hilbert-space-valued processes, while the next two treat the more general case in which the martingales take values in a separable Banach space $\mathcal{B}$. The final section of that chapter is devoted to some applications of these results to the Haar system on $[0, 1)$.

Chapter 4 contains analogous results in the case when the dominating process f is a sub- or a supermartingale. The notion of differential subordination, which is a very convenient condition in the martingale setting, becomes too weak and needs to be strengthened. We do this by imposing an extra conditional subordination and assume that the terms $|\mathbb{E}(dg_n|\mathcal{F}_{n-1})|$ are controlled by $|\mathbb{E}(df_n|\mathcal{F}_{n-1})|$, $n = 1, 2, \ldots$. This domination, called strong differential subordination, has a very natural counterpart in the theory of Itô processes and stochastic integrals, and is sufficient for our purposes.

In Chapter 5 we show how to extend the above inequalities to continuous-time processes and stochastic integrals. In fact, some of the estimates studied there can be regarded as a motivation to the results of Chapter 3 and Chapter 4. Consider the following example. Let $X = (X_t)_{t\geq 0}$ be a continuous-time real-valued martingale and let $H = (H_t)_{t\geq 0}$ be a predictable process taking values in $[-1, 1]$. As usual, we assume that the trajectories of both these processes are sufficiently regular, that is, are right-continuous and have limits from the left. Suppose that Y is the Itô integral of H with respect to X: that is, we have

$$Y_t = H_0 X_0 + \int_{0+}^{t} H_s \mathrm{d}X_s$$

for $t \geq 0$. Then the pair (X, Y) is precisely the continuous analogue of a martingale and its transform by a predictable process bounded in absolute value by 1. We may ask questions, similar to (i)–(iv) above, concerning the comparison of the sizes of X and Y. As we shall see, some approximation theorems allow one to carry over the inequalities from the discrete-time setting to that above, without changing the constants. In fact, we shall prove much more. The notions of differential subordination and strong differential subordination can be successfully generalized to the continuous-time case, and we shall present an appropriate modification of Burkholder's method, which can be applied to study the corresponding estimates in this wider setting. In one of the final sections, we apply the results to obtain some interesting estimates for harmonic functions on Euclidean domains, which can be regarded as extensions of Riesz's inequality.

In Chapter 6 we continue the line of research started in Chapter 5. Namely, we investigate there continuous-time processes under (strong) differential subordination, but this time we impose the additional orthogonality assumption. The martingales of this type turn out to be particularly closely related to periodic functions and their conjugates, mentioned at the beginning. The final section of that chapter is devoted to the description of the connections between these two settings.

Chapter 7 deals with another important class of estimates: the maximal ones. To be more specific, one can ask questions similar to (i)–(iv) above in the case when f (or g, or both) is replaced by the corresponding maximal function $|f|^* = \sup_{n\geq 0}|f_n|$ or one-sided maximal function $f^* = \sup_{n\geq 0} f_n$. For example, we shall present the proof the classical Doob's maximal inequality

$$||f^*||_p \leq \frac{p}{p-1}||f||_p, \qquad 1 < p \leq \infty,$$

where f is a nonnegative submartingale, as well as a number of related weak-type and logarithmic estimates. See Section 7.2.

Maximal inequalities involving two processes arise naturally in many situations and are of particular interest when the corresponding non-maximal versions are not valid. For instance, there is no universal $C_1 < \infty$ such that $||g||_1 \leq C_1||f||_1$ for all real-valued martingales f and their ± 1-transforms g; on the other hand, Burkholder [35] showed that for such f and g we have

$$||g||_1 \leq 2.536\ldots || \, |f|^*||_1,$$

and the bound is sharp. For other related results, see Sections 7.3–7.8.

Problems of this type can be successfully investigated using appropriate modification of Burkholder's method, which was developed in [35]. However, it should be stressed here that the corresponding boundary value problems become tricky due to the increase of the dimension. That is, the special functions we search for depend on three or four variables (each of the terms $|f|^*$, $|g|^*\ldots$ involves an additional parameter). Furthermore, in contrast with the non-maximal setting, the passage to the Hilbert-space-valued setting does require extra effort and the inequalities for Banach-space-valued semimartingales become even more difficult. The next interesting feature is that, unlike in the non-maximal case, the additional assumption on the continuity of paths does affect the optimal constants and Burkholder's method needs further refinement. See Sections 7.9–7.12.

Chapter 8 is the final part of the exposition and is devoted to the study of square function inequalities. Recall that if $f = (f_n)_{n\geq 0}$ is an adapted sequence, then its square function is given by

$$S(f)=\left(\sum_{k=0}^{\infty}|df_k|^2\right)^{1/2}.$$

Again, we may ask about sharp moment, weak type and other related estimates between f and $S(f)$. In addition, we can consider maximal inequalities which involve a martingale, its square and maximal function. There are also continuous-time analogues of such estimates (the role of the square function is played by the square bracket or the quadratic covariance process), but they can be easily reduced to the discrete-time bounds by means of standard approximation. The problems of such type are classical and have numerous extensions in various areas of mathematics: see bibliographical notes at the end of Chapter 8. It turns out that many interesting estimates can be immediately deduced from related results concerning Hilbert-space-valued differentially subordinated martingales (see Chapter 2 below), but this approach does not always allow to keep track of the optimal constants. To derive the sharp estimates, one may apply an appropriate modification of Burkholder's method, invented in [38]. However, the corresponding boundary value problems are difficult in general and this technique has been successfully implemented only in a few of cases: see Sections 8.2, 8.3 and 8.4.

The situation becomes a bit easier when we restrict ourselves to conditionally symmetric martingales. Recall that f is conditionally symmetric when for each $n\geq 1$, the conditional distributions of df_n and $-df_n$ given $\mathcal{F}_{n-1}$ coincide. For example, consider the so-called *dyadic martingales*: take the probability space to be $([0,1],\mathcal{B}(0,1),|\cdot|)$ and put $f_n=\sum_{k=0}^{n}a_kh_k$, where $h=(h_k)_{k\geq 0}$ is the Haar system and a_1, a_2, … are coefficients, which may be real or vector valued. Then the corresponding boundary value problem can often (but not always) be dealt with by solving the heat equation on a part of the domain of the special function. See Sections 8.2, 8.3, 8.5 and 8.6.

The final part of Chapter 8 concerns the conditional square function. It is another classical object in the martingale theory, given by

$$s(f)=\left(\sum_{k=0}^{\infty}\mathbb{E}(|df_k|^2|\mathcal{F}_{(k-1)\vee 0})\right)^{1/2}.$$

We shall show how to modify Burkholder's method so that it yields estimates involving f and $s(f)$. The approach can be further extended to imply related estimates for sums of nonnegative random variables and their predictable projections: see Wang [199] and the author [134], but we shall not include this here.

A few words about the organization of the monograph. Typically, each section starts with the statement of a number of theorems, which contain results on semimartingale inequalities. In most cases, the proof of a given theorem is divided into three separate parts. In the first step we make use of Burkholder's method and establish the inequality contained in the statement. The next part deals with the optimality of the constants appearing in this estimate. In the final part we present some intuitive arguments which lead to the discovery of the special function, used in the first step. Wherever possible, we have tried to present new proofs and new reasoning which, as we hope, throws some additional light on the structure of the problems. We also did our best to keep the exposition as self-contained as possible, and omit argumentation or its part only when a similar reasoning has been presented earlier. Each chapter concludes with a section containing historical and bibliographical notes concerning the results studied in the preceding sections as well as some material for further reading.

Chapter 2

Burkholder's Method

We start by introducing the main tool which will be used in the study of semimartingale inequalities. For the sake of clarity, in this chapter we focus on the description of the method only for discrete-time martingales. The necessary modifications, leading to inequalities for wider classes of processes, will be presented in the further parts of the monograph.

2.1 Description of the technique

2.1.1 Inequalities for ±1-transforms

Burkholder's method relates the validity of a certain given inequality for semimartingales to a corresponding boundary value problem, or, in other words, to the existence of a special function, which has appropriate concave-type properties. To start, let us assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, which is filtered by $(\mathcal{F}_n)_{n\geq 0}$, a non-decreasing family of sub-σ-fields of $\mathcal{F}$. Consider adapted simple martingales $f = (f_n)_{n\geq 0}$, $g = (g_n)_{n\geq 0}$ taking values in $\mathbb{R}$, with the corresponding difference sequences $(df_n)_{n\geq 0}$, $(dg_n)_{n\geq 0}$, respectively. Here by simplicity of f we mean that 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 $f_N = f_{N+1} = f_{N+2} = \cdots$.

Let $D = \mathbb{R} \times \mathbb{R}$ and let $V : D \to \mathbb{R}$ be a function, not necessarily Borel or even measurable. Let x, $y \in \mathbb{R}$ be fixed and denote by $M(x, y)$ the class of all pairs (f, g) of simple martingales f and g starting from x and y, respectively, such that $dg_n \equiv df_n$ or $dg_n \equiv -df_n$ for any $n \geq 1$. Here the filtration may vary as well as the probability space, unless it is assumed to be non-atomic. Suppose that we are interested in the numerical value of

$$U^0(x, y) = \sup\big\{\mathbb{E}V(f_n, g_n)\big\}, \tag{2.1}$$

where the supremum is taken over $M(x, y)$ and all nonnegative integers n. Of course, there is no problem with measurability or integrability of $V(f_n, g_n)$, since

A. Osękowski, *Sharp Martingale and Semimartingale Inequalities*, Monografie Matematyczne 72,
DOI 10.1007/978-3-0348-0370-0_2, © Springer Basel 2012

the sequences f and g are simple. Note that the definition of U^0 can be rewritten in the form

$$U^0(x,y) = \sup_{(f,g)\in M(x,y)} \big\{\mathbb{E}V(f_\infty, g_\infty)\big\},$$

where f_∞ and g_∞ stand for the pointwise limits of f and g (which exist due to the simplicity of the sequences). This is straightforward: for any $(f,g) \in M(x,y)$ and any nonnegative integer n, we have $(f_n, g_n) = (\overline{f}_\infty, \overline{g}_\infty)$, where the pair $(\overline{f}, \overline{g}) \in M(x,y)$ is just (f,g) stopped at time n.

In most cases, we will try to provide some upper bounds for U^0, either on the whole domain D, or on its part. The key idea in the study of such a problem is to introduce a class of special functions. The class consists of all $U : D \to \mathbb{R}$ satisfying the following conditions 1° and 2°:

1° (Majorization property) For all $(x,y) \in D$,

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

2° (Concavity-type property) For all $(x,y) \in D$, $\varepsilon \in \{-1,1\}$ 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

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

Using a straightforward induction argument, we can easily show that the condition 2° is equivalent to the following: for all $(x,y) \in D$, $\varepsilon \in \{-1,1\}$ and any simple mean-zero variable d we have

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

To put it in yet another words, (2.3) amounts to saying that the function U is *diagonally concave*, that is, concave along the lines of slope ± 1.

The interplay between the problem of bounding U^0 from above and the existence of a special function U satisfying 1° and 2° is described in the two statements below, Theorem 2.1 and Theorem 2.2.

Theorem 2.1. *Suppose that U satisfies 1° and 2°. Then for any simple f and g such that $dg_n \equiv df_n$ or $dg_n \equiv -df_n$ for $n \geq 1$ we have*

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

In particular, this implies

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

Proof. The key argument is that the process $(U(f_n, g_n))_{n\geq 0}$ is an $(\mathcal{F}_n)$-supermartingale. To see this, note first that all the variables are integrable, by the simplicity of f and g. Fix $n \geq 1$ and observe that

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

An application of (2.4) conditionally on $\mathcal{F}_{n-1}$, with $x = f_{n-1}$, $y = g_{n-1}$ and $d = df_n$ yields the supermartingale property. Thus, by 1°,

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

and the proof is complete. □

Therefore, we have obtained that $U^0(x, y) \le \inf U(x, y)$, where the infimum is taken over all U satisfying 1° and 2°. The remarkable feature of the approach is that the reverse inequality is also valid. To be more precise, we have the following statement.

Theorem 2.2. *If U^0 is finite on D, then it is the least function satisfying* 1° *and* 2°*.*

Proof. The fact that U^0 satisfies 1° is immediate: the deterministic constant pair (x, y) belongs to $M(x, y)$. To prove 2°, we will use the so-called "splicing argument". Take $(x, y) \in D$, $\varepsilon \in \{-1, 1\}$ and α, t_1, t_2 as in the statement of the condition. Pick 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 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 $(f_0, g_0) \equiv (x, y)$ and

$$(f_n, g_n)(\omega) = (f^1_{n-1}, g^1_{n-1})(\omega/\alpha), \qquad \text{if } \omega \in [0, \alpha)\,,$$

and

$$(f_n, g_n)(\omega) = (f^2_{n-1}, g^2_{n-1})\left(\frac{\omega - \alpha}{1 - \alpha}\right), \qquad \text{if } \omega \in [\alpha, 1)\,,$$

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

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

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

$$U^0(x, y) \ge \alpha U^0(x + t_1, y + \varepsilon t_1) + (1 - \alpha)U^0(x + t_2, y + \varepsilon t_2),$$

which is 2°. To see that U^0 is the least special function, simply look at (2.6). □

The above two facts give the following general method of proving inequalities for ± 1-transforms. Let $V : D \to \mathbb{R}$ 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* $U : D \to \mathbb{R}$ *satisfying* 1°, 2° *and the* initial condition

3° $U(x, y) \le 0$ *for all* x, y *such that* $y = \pm x$.

Proof. If there is a function U satisfying 1°, 2° and 3°, then (2.8) follows immediately from (2.5), since 3° guarantees that the term $\mathbb{E}U(f_0, g_0)$ is nonpositive. To get the reverse implication, we use Theorem 2.2: as we know from its proof, the function U^0 satisfies 1° and 2°. It also enjoys 3°, directly from the definition of U^0 combined with the inequality (2.8). The only thing which needs to be checked is the finiteness of U^0, which is assumed in Theorem 2.2. Since $U^0 \ge V$, we only need to show that $U^0(x, y) < \infty$ for every (x, y). The condition 3°, which we have already established, guarantees the inequality on the diagonals $y = \pm x$. Suppose that $|x| \ne |y|$ 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 $(f'_n, g'_n)_{n\ge 1}$ coincides with the (unconditional) distribution of $(f_n, g_n)_{n\ge 0}$. We easily check that g' is a ± 1-transform of f', and hence, for any $n \ge 1$,

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

Consequently, taking supremum over f, g and n gives $U^0(x, y) \le -V(y, x)$ and we are done. □

Remark 2.1. Suppose that V has the symmetry property

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

Then we may replace 3° by the simpler condition

3°' $U(0, 0) \le 0$.

In other words, if there is U which satisfies the conditions 1°, 2° and 3°', then there is $\bar{U}$ which satisfies 1°, 2° and 3°. To prove this, fix U as in the previous sentence. By (2.9), the functions $(x, y) \mapsto U(-x, y)$, $(x, y) \mapsto U(x, -y)$ and $(x, y) \mapsto U(-x, -y)$ also enjoy the properties 1°, 2° and 3°', and hence so does $\bar{U}$ given by

$$\bar{U}(x, y) = \min\{U(x, y), U(-x, y), U(x, -y), U(-x, -y)\}, \quad x,\ y \in \mathbb{R}.$$

But this function satisfies 3°: indeed, by 2°,

$$\bar{U}(x, \pm x) = \frac{\bar{U}(x,x) + \bar{U}(-x,-x)}{2} \le \bar{U}(0,0) \le 0,$$

for any $x \in \mathbb{R}$.

The approach described above concerns only real-valued processes. Furthermore, the condition of being a ± 1-transform is quite restrictive. There arises the natural question whether the methodology can be extended to a wider class of martingales and we will shed some light on it.

2.1.2 Inequalities for general transforms of Banach-space-valued martingales

Let us start with the following vector-valued version of Theorem 2.3. The proof is the same as in the real case and is omitted. Let $\mathcal{B}$ be a Banach space.

Theorem 2.4. *Let $V : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ be a given function. The inequality*

$$\mathbb{E}V(f_n, g_n) \le 0$$

holds for all n and all pairs (f, g) of simple $\mathcal{B}$-valued martingales such that g is a ± 1-transform of f if and only if there exists $U : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ satisfying the following three conditions.

1° *$U \ge V$ on $\mathcal{B} \times \mathcal{B}$.*

2° *For all x, $y \in \mathcal{B}$, $\varepsilon \in \{-1, 1\}$ 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*

$$\alpha U(x + t_1, y + \varepsilon t_1) + (1 - \alpha)U(x + t_2, y + \varepsilon t_2) \le U(x, y).$$

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

In the previous situation, we had $dg_n = v_n df_n$, $n = 0, 1, 2, \ldots$, where each v_n was deterministic and took values in the set $\{-1, 1\}$. Now let us consider the more general situation in which the sequence v is simple, predictable and takes values in $[-1, 1]$. Recall that predictability means that each v_n is measurable with respect to $\mathcal{F}_{(n-1)\vee 0}$ and, in particular, this allows random terms. The corresponding version of Theorem 2.4 can be stated as follows. We omit the proof, it requires no new ideas.

Theorem 2.5. *Let $V : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ be a given function. The inequality*

$$\mathbb{E}V(f_n, g_n) \le 0$$

holds for all n and all f, g as above if and only if there exists $U : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ satisfying the following three conditions.

1° *$U \geq V$ on $\mathcal{B} \times \mathcal{B}$.*

2° *For all $(x, y) \in \mathcal{B} \times \mathcal{B}$, 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*

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

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

Let us make here some important observations.

Remark 2.2. (i) Condition 2° of Theorem 2.5 extends to the following inequality: for all x, $y \in \mathcal{B}$, any deterministic $a \in [-1, 1]$ and any simple mean zero $\mathcal{B}$-valued random variable d we have

$$\mathbb{E}U(x + d, y + ad) \leq U(x, y).$$

(ii) Condition 2° can be rephrased as follows: for any x, y, $h \in \mathcal{B}$ and $a \in [-1, 1]$, the function $G = G_{x,y,h,a} : \mathbb{R} \to \mathbb{R}$ given by $G(t) = U(x + th, y + tah)$ is concave.

(iii) Arguing as in Remark 2.1, we can prove the following statement. If V satisfies $V(x, y) = V(-x, y) = V(x, -y)$ for all x, $y \in \mathcal{B}$, then we may replace the above initial condition 3° by

3°′ $U(0, 0) \leq 0$.

It is worth mentioning here that ± 1 transforms usually are the extremal sequences in the above class of transforms. To be more precise, we have the following decomposition.

Theorem 2.6. *Let g be the transform of a $\mathcal{B}$-valued martingale f by a real-valued predictable sequence v uniformly bounded in absolute value by 1. Then there exist $\mathcal{B}$-valued martingales $F^j = (F_n^j)_{n\geq 0}$ and Borel measurable functions $\phi_j : [-1, 1] \to \{-1, 1\}$ such that for $j \geq 1$ and $n \geq 0$,*

$$f_n = F_{2n+1}^j, \qquad \text{and} \qquad g_n = \sum_{j=1}^{\infty} 2^{-j}\phi_j(v_0)G_{2n+1}^j,$$

where G^j is the transform of F^j by $\varepsilon = (\varepsilon_k)_{k\geq 0}$ with $\varepsilon_k = (-1)^k$.

Proof. First we consider the special case when each v_n takes values in the set $\{-1, 1\}$. Let

$$D_{2n} = \frac{1 + v_0 v_n}{2} d_n,$$
$$D_{2n+1} = \frac{1 - v_0 v_n}{2} d_n.$$

Then $D = (D_n)_{n\geq 0}$ is a martingale difference sequence with respect to its natural filtration. Indeed, for even indices,

$$\mathbb{E}(D_{2n}|\sigma(D_0, D_1, \ldots, D_{2n-1})) = \mathbb{E}\left[\frac{1+v_0v_n}{2}\mathbb{E}(d_n|\mathcal{F}_{n-1})\middle|\sigma(D_0, \ldots, D_{2n-1})\right] = 0.$$

Here $(\mathcal{F}_n)_{n\geq 0}$ stands for the original filtration. Furthermore, we have $D_{2n} = 0$ or $D_{2n+1} = 0$ for all n, so

$$\begin{aligned}\mathbb{E}(D_{2n+1}|\sigma(D_0, \ldots, D_{2n})) &= \mathbb{E}(D_{2n+1}1_{\{D_{2n}=0\}}|\sigma(D_0, \ldots, D_{2n})) \\ &= \mathbb{E}(D_{2n+1}1_{\{D_{2n}=0\}}|\sigma(D_0, \ldots, D_{2n-1})) = 0.\end{aligned}$$

Now, let F be the martingale determined by D and let G be its transform by ε. By the definition of D we have $d_n = D_{2n} + D_{2n+1}$ and $v_0v_nd_n = D_{2n} - D_{2n+1}$, so $f_n = F_{2n+1}$ and $g_n = v_0G_{2n+1}$.

In the general case when the terms v_n take values in $[-1, 1]$, note that there are Borel measurable functions $\phi_j : [-1, 1] \to \{-1, 1\}$ satisfying

$$t = \sum_{j=1}^{\infty} 2^{-j}\phi_j(t), \qquad t \in [-1, 1].$$

Now, for any j, consider the sequence $v^j = (\phi_j(v_n))_{n\geq 0}$, which is predictable with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$. By the previous special case there is a martingale F^j and its transform G^j satisfying

$$\begin{aligned} f_n &= F^j_{2n+1}, \\ \sum_{k=0}^{n} \phi_j(v_k) &= \phi_j(v_0)G^j_{2n+1}. \end{aligned}$$

It suffices to multiply both sides by 2^{-j} and sum the obtained equalities to get the claimed decomposition. □

2.1.3 Differential subordination

Now we shall introduce another very important class of martingale pairs. It is much wider than that considered in the previous two subsections and allows many interesting applications. Let $\mathcal{B}$ be a given separable Banach space with the norm $|\cdot|$.

Definition 2.1. Suppose that f, g are martingales taking values in $\mathcal{B}$. Then g is *differentially subordinate* to f, if for any $n = 0, 1, 2, \ldots,$

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

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. Suppose that f takes values in a given separable Banach space and let g be $\ell^2(\mathcal{B})$-valued process, defined by $dg_n = (0, 0, \ldots, 0, df_n, 0, \ldots)$, $n = 0, 1, 2, \ldots$ (where the difference df_n appears on the nth place). Let us treat f as an $\ell^2(\mathcal{B})$-valued process, via the embedding $f_n \sim (f_n, 0, 0, \ldots)$. Then, obviously, g is differentially subordinate to f and f is differentially subordinate to g. However,

$$||g_n||_{\ell^2(\mathcal{B})} = \left(\sum_{k=0}^{n} |df_k|^2\right)^{1/2}$$

is the square function of f. Thus, any inequality valid for differentially subordinate martingales with values in $\ell^2(\mathcal{B})$ leads to a corresponding estimate for the square function of a $\mathcal{B}$-valued martingale. This observation will be particularly efficient when $\mathcal{B}$ is a separable Hilbert space.

Let us formulate the version of Burkholder's method when the underlying domination is the differential subordination of martingales. Let $V : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ be a given Borel function. Consider $U : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ such that

1° $U(x, y) \geq V(x, y)$ for all $x, y \in \mathcal{B}$,

2° there are Borel $A, B : \mathcal{B} \times \mathcal{B} \to \mathcal{B}^*$ such that for any $x, y \in \mathcal{B}$ and any $h, k \in \mathcal{B}$ with $|k| \leq |h|$, we have

$$U(x + h, y + k) \leq U(x, y) + \langle A(x, y), h\rangle + \langle B(x, y), k\rangle.$$

3° $U(x, y) \leq 0$ for all $x, y \in \mathcal{B}$ with $|y| \leq |x|$.

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

$$\begin{aligned} &\mathbb{E}|V(f_n, g_n)| < \infty, \quad \mathbb{E}|U(f_n, g_n)| < \infty \\ &\mathbb{E}\big(|A(f_n, g_n)||df_{n+1}| + |B(f_n, g_n)||dg_{n+1}|\big) < \infty, \end{aligned} \tag{2.10}$$

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

$$\mathbb{E}V(f_n, g_n) \leq 0 \tag{2.11}$$

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

Proof. Note that this result goes beyond the scope of Burkholder's method described so far, since the processes f, g are no longer assumed to be simple. This is why we have assumed the Borel measurability of U, V, A and B; this is also why we have imposed condition (2.10): it guarantees the integrability of the random variables appearing below. However, the underlying idea is the same: we show that for f, g as above the process $(U(f_n, g_n))_{n\geq 0}$ is a supermartingale. To prove this, we use 2° to obtain, for any $n \geq 1$,

$$U(f_n, g_n) \leq U(f_{n-1}, g_{n-1}) + \langle A(f_{n-1}, g_{n-1}), df_n\rangle + \langle B(f_{n-1}, g_{n-1}), dg_n\rangle$$

with probability 1. By (2.10), both sides above are integrable. Taking the conditional expectation with respect to $\mathcal{F}_{n-1}$ yields

$$\mathbb{E}(U(f_n, g_n)|\mathcal{F}_{n-1}) \leq U(f_{n-1}, g_{n-1})$$

and, consequently,

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

This completes the proof. □

Remark 2.3. (i) Condition 2° seems quite complicated. However, if U is of class C^1, it is easy to see that the only choice for A and B is to take the partial derivatives U_x and U_y, respectively. Then 2° is equivalent to saying that

$2^{\circ\prime}$ for any x, y, h, $k \in \mathcal{B}$ with $|k| \leq |h|$, the function $G = G_{x,y,h,k} : \mathbb{R} \to \mathbb{R}$, given by

$$G(t) = U(x + th, y + tk),$$

is concave.

In a typical situation, U is piecewise C^1, and then $2^{\circ\prime}$ still implies 2°: one takes $A(x, y) = U_x(x, y)$, $B(x, y) = U_y(x, y)$ for (x, y) at which U is differentiable and, for remaining points, one defines A and B as appropriate limits of U_x and U_y.

(ii) Condition $2^{\circ\prime}$ can be simplified further. Obviously, it is equivalent to

$$G''(t) \leq 0 \tag{2.12}$$

at the points where G is twice differentiable, and

$$G'(t-) \leq G'(t+) \tag{2.13}$$

for the remaining t. However, the family $(G_{x,y,h,k})_{x,y,h,k}$ enjoys the following translation property:

$$G_{x,y,h,k}(t + s) = G_{x+th,y+tk,h,k}(s) \quad \text{for all} \quad s,\ t.$$

Hence, it suffices to check (2.12) and (2.13) for $t = 0$ only (but, of course, for all appropriate x, y, h and k).

2.2 Further remarks

Now let us make some general observations, some of which will be frequently used in the later parts of the monograph.

(i) The technique can be applied in the situation when the pair (f, g) takes values in a set D different from $\mathcal{B} \times \mathcal{B}$. For example, one can work in $D = \mathbb{R}_+ \times \mathbb{R}$ or $D = \mathcal{B} \times [0, 1]$, and so on. This does not require any substantial changes in the methodology; one only needs to ensure that all the points (x, y), $(x + t_i, y + \varepsilon t_i)$, and so on, appearing in the statements of 1°, 2° and 3°, belong to the considered domain D.

(ii) A remarkable feature of Burkholder's method is its efficiency. Namely, if we know a priori that a given estimate

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

or, more generally,

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

is valid, then it can be established using the above approach. In particular, the technique can be used to derive the optimal constants in the inequalities under investigation.

(iii) Formula (2.1) can be used to narrow the class of functions in which we search for the suitable majorant. Here is a typical example. Suppose we are interested in showing the strong-type inequality

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

for all real martingales f and their ± 1-transforms g. This corresponds to the choice $V(x, y) = |y|^p - C^p|x|^p$, x, $y \in \mathbb{R}$. We have that V is homogeneous of order p and this property carries over to the function U^0. It follows from the fact that $(f, g) \in M(x, y)$ if and only if $(\lambda f, \lambda g) \in M(\lambda x, \lambda y)$ for any $\lambda > 0$. Thus, we may search for U in the class of functions which are homogeneous of order p. This reduces the dimension of the problem. Indeed, we need to find an appropriate function u of only one variable and then let $U(x, y) = |y|^p u(|x|/|y|)$ for $|y| \neq 0$ and $U(x, 0) = c|x|^p$ for some c. As another example, suppose that V satisfies the symmetry condition $V(x, y) = V(x, -y)$ for all x, y. Then we may search for U in the class of functions which are symmetric with respect to the second variable.

(iv) A natural way of showing that the constant in a given inequality is the best possible is to construct appropriate examples. However, this can be shown by the use of the reverse implication of Burkholder's method. That is, one assumes the validity of an estimate with a given constant C and then exploits the properties 1°, 2° and 3° of the function U^0 to obtain the lower bound for C. This approach is often much simpler and less technical, and will be frequently used in the considerations below.