University of Warsaw

Faculty of Mathematics, Informatics and Mechanics

Mateusz Rapicki

Weighted and unweighted estimates for maximal operators

PhD dissertation

Supervisor:
Dr hab. Adam Osękowski
Institute of Mathematics
University of Warsaw

January 2021

Author's declaration:
I hereby declare that this dissertation is my own work.

January 15, 2021 . . . .

Mateusz Rapicki

Supervisor's declaration:
The dissertation is ready to be reviewed.

January 15, 2021 . . . .

dr hab. Adam Osękowski

Abstract

The thesis is devoted to the study of various classes of inequalities for maximal operators, both in the weighted and the unweighted settings. It is distinguished by the following two features:

I. We put a particular emphasis on obtaining sharp results, i.e., deriving the optimal values of the constants involved;

II. We establish the results with the use of the Bellman function method.

The motivation for I. comes from the fact that maximal operators provide an efficient tool for the study of wide classes of operators appearing in harmonic analysis. In particular, maximal estimates can often be applied to obtain the appropriate boundedness properties of such operators. From this perspective, II. is a very natural direction: Bellman function method is a tool which enables the investigation of extremal problems and optimal constants. We expect that the special functions and the unified approach presented in the thesis can be further extended and exploited in the study of related sharp bounds in probability theory and harmonic analysis.

Throughout, we focus mainly on the dyadic maximal operators and their certain extensions, the so-called dyadic-like maximal operators, which have a direct interpretation in the probability theory. The material is organized as follows. Chapter 1 is of an introductory character and contains some motivation, necessary background information, and notation. Chapter 2 is devoted to the description of the Bellman function method, specified to the context of maximal estimates; this unified approach is a compilation of several works from the literature. The main contribution of the thesis has been placed in the next five chapters. In Chapter 3 we obtain a sharp weighted extension of the Kolmogorov inequality: a maximal L^p estimate for $p < 1$. Chapter 4 contains a proof of a transference theorem which enables the passage from a certain type of unweighted estimates for the dyadic maximal operator to their Fefferman-Stein counterparts in the dyadic-like context. Chapter 5 is devoted to a certain maximal weak-type estimate in the presence of Muckenhoupt's A_p weights, $p > 1$. Chapter 6 deals with a sharp Lorentz-norm inequality for maximal functions. In the final part of the paper, Chapter 7, we establish a sharp two-weight bound for a slightly different object, the so-called harmonic maximal operator.

The contents of the dissertation is quite technical, at least at some points. The search for the optimal constants and the use of the Bellman function method makes the calculations really involved in many cases.

AMS 2020 subject classification: Primary: 42B25. Secondary: 46E30, 60G42.

Keywords: maximal operator, dyadic, sharp inequality, weight, Lorentz space

Streszczenie

Rozprawa jest poświęcona badaniu różnych klas nierówności dla operatorów maksymalnych, zarówno w kontekście ważonym jak i bezwagowym. Wyróżniają ją następujące dwie cechy:

I. Kładziemy szczególny nacisk na otrzymywanie wyników z optymalną stałą;

II. Stosujemy metodę funkcji Bellmana.

Motywacja dla I. wynika z faktu, że operatory maksymalne stanowią efektywne narzędzie do studiowania szerokich klas operatorów występujących w analizie harmonicznej. W szczególności, oszacowania maksymalne często mogą być stosowane do otrzymywania odpowiednich własności ograniczoności takich operatorów. Z tej perspektywy II. jest bardzo naturalnym kierunkiem: metoda funkcji Bellmana jest narzędziem pozwalającym na badanie problemów ekstremalnych z optymalnymi stałymi. Mamy nadzieję, że funkcje specjalne oraz zunifikowane podejście zaprezentowane w niniejszej rozprawie będą mogły być dalej uogólniane i wykorzystywane w pracy nad pokrewnymi nierównościami z optymalną stałą w teorii prawdopodobieństwa i analizie harmonicznej.

W całej rozprawie skupiamy się głównie na diadycznych operatorach maksymalnych i na pewnym ich uogólnieniu, tak zwanych quasi-diadycznych operatorach maksymalnych, które ma bezpośrednią interpretację w probabilistyce. Praca jest zorganizowana w następujący sposób. Rozdział 1 ma charakter wprowadzający i zawiera motywację, potrzebne informacje wstępne i notację. Rozdział 2 jest poświęcony opisowi metody funkcji Bellmana w kontekście oszacowań maksymalnych; to zunifikowane podejście jest kompilacją kilku prac z literatury. Główny wkład niniejszej rozprawy umieściliśmy w następnych pięciu rozdziałach. W Rozdziale 3 otrzymujemy oszacowanie ważone z optymalną stałą będące uogólnieniem nierówności Kołmogorowa: nierówności maksymalnej w L^p dla $p < 1$. Rozdział 4 zawiera dowód twierdzenia pozwalającego na przejście od pewnego typu oszacowań bezwagowych dla diadycznego operatora maksymalnego do ich odpowiedników typu Feffermana-Steina w kontekście quasi-diadycznym. Rozdział 5 jest poświęcony pewnej nierówności maksymalnej słabego typu z wagą klasy A_p Muckenhoupta, $p > 1$. W Rozdziale 6 zajmujemy się oszacowaniem z optymalną stałą dla norm Lorentza funkcji maksymalnych. W ostatniej części pracy, Rozdziale 7, dowodzimy nierówności dwuwagowej z optymalną stałą dla nieco innego obiektu, tak zwanego harmonicznego operatora maksymalnego.

Zawartość rozprawy jest dość techniczna, przynajmniej w niektórych miejscach. Poszukiwanie optymalnych stałych i stosowanie metody funkcji Bellmana w wielu przypadkach wymaga skomplikowanych obliczeń.

Acknowledgements

First and foremost, I would like to thank my supervisor Adam Osękowski. He inspired me to enter the path of research and to pursue the doctoral degree, and devouted a great amount of time to help me along the way. I am grateful for his guidance and patience.

Secondly, I would like to thank my friends from the MIMUW Choir for all the time we spent together. Music and singing has been my refuge.

I would like to express my gratitude to the National Science Centre in Poland for supporting my research via grants no. DEC-2014/14/E/ST1/00532 and 2018/29/N/ST1/02840. I also acknowledge the financial support of Warsaw Center of Mathematics and Computer Science.

Contents

Chapter 1

Introduction

Maximal operators are fundamental objects in analysis and probability theory, and have far reaching applications in other areas of mathematics. In particular, they enable the study of the boundedness, in various function spaces, of wide families of classical operators (e.g., fractional or singular). This is a direct consequence of the fact that many such operators, or some of their components, can be controlled pointwise by an appropriate version of a maximal function. This gives rise to the question about the efficient handling of maximal estimates, and the purpose of the thesis is to present a number of techniques which can be exploited in this type of problems. Furthermore, we will put particular emphasis on the size of the constants involved.

We start with recalling some background and notation. For a locally integrable function $f : \mathbb{R}^d \to \mathbb{R}$ (where the integrability is with respect to the d-dimensional Lebesgue measure), the Hardy-Littlewood maximal operator is defined as follows:

$$Mf(x) = \sup \left\{ \langle |f| \rangle_Q : Q \in \mathcal{Q}_x \right\}.$$

Here $\langle f \rangle_Q$ stands for $\frac{1}{|Q|} \int_Q f(x) \mathrm{d}x$, the average of f over Q, $|Q|$ denotes the Lebesgue measure of Q and for each $x \in \mathbb{R}^d$, $\mathcal{Q}_x$ is a certain family of sets containing x. There are five most important choices for such families studied widely in the literature:

(i) the *centered maximal operator*, associated with the classes $\mathcal{Q}_x = \{B(x,r) : r > 0\}$ of open balls centered at x;

(ii) the *uncentered maximal operator*, which corresponds to $\mathcal{Q}_x$ being the family of all balls containing x;

(iii) and (iv) are the versions of (i) and (ii) with balls replaced with cubes having sides parallel to the axes;

(v) the *dyadic maximal operator*, which corresponds to $\mathcal{Q}_x$ being the class of all dyadic cubes containing x. (For the discussion on the dyadic lattice, see below).

It is worth saying here that the operators in the above contexts are essentially comparable: having proved an estimate in one setting, one immediately deduces the corresponding statement for the remaining operators (with a different constant, depending on the dimension d). While this is more or less obvious for the families (i)-(iv) (see e.g. Grafakos [19]), the interplay between the dyadic and non-dyadic context is a little more involved: consult [22, 32]. The subject becomes much more challenging for general metric spaces equipped with a non-doubling measure, but we will not touch this area in this thesis.

Maximal operators have been studied and applied in numerous problems of analysis and probability. In principle, any textbook on harmonic analysis begins with some more or less detailed presentation of this topic (for example, we refer the interested reader to the monographs [19] and [74]). Depending on the problem under investigation, it is often convenient to stick to one of the settings (i)-(v); each of them provides some additional geometric or combinatorial arguments which might be crucial in the study. Here is a short discussion on some important selected examples; for a more detailed presentation, see [19, 45, 68]. We will encounter further examples later in this section, when studying the dyadic context.

(a) Lebesgue's differentiation theorem asserts that if $f : \mathbb{R}^d \to \mathbb{R}$ is a locally integrable function, then for almost all $x \in \mathbb{R}^d$ the limit

$$\lim_{r\to 0} \langle f \rangle_{B(x,r)}$$

exists and is equal to $f(x)$. The proof of this important result exploits weak-type bounds for centered maximal operators and the density of continuous functions in L^p spaces. A similar argumentation leads to various extensions of Lebesgue's theorem, involving averages over other families of sets containing x. Such results have turned out to be extremely useful, for instance in the context of Calderón-Zygmund decompositions and their extensions.

(b) Maximal functions dominate (pointwise) a large class of convolution operators. For example, suppose that $k : [0,\infty) \to [0,\infty)$ is a decreasing continuous function such that the function $K : \mathbb{R}^d \to \mathbb{R}$ defined by $K(x) = k(|x|)$ is integrable. Define the ε-dilation of K by $K_\varepsilon(x) = \varepsilon^{-d}K(\varepsilon^{-1}x)$. Then for any locally integrable function f we have the estimate

$$\sup_{\varepsilon>0}(K_\varepsilon * |f|)(x) \leq \|K\|_{L^1(\mathbb{R}^d)}\, Mf(x),$$

where M is the uncentered maximal operator. Therefore, any estimate for M immediately yields the corresponding bound for the convolution operators given by $T_\varepsilon f = K_\varepsilon * f$. This observation can be exploited in the study of (maximal) Hilbert transforms and Poisson/heat semigroups, but the range of applications is much wider.

(c) Estimates for maximal operators form a crucial ingredient in the proofs of various types of extrapolation theorems. The simplest form of such a theorem asserts that if for some fixed $1 < p_0 < \infty$ a given operator T is bounded on the weighted space $L^{p_0}(w_0)$ for any A_{p_0} weight w_0, then automatically T is bounded on any weighted space $L^p(w)$ for any $1 < p < \infty$ and any A_p weight w. The proof exploits the so-called algorithm of Rubio de Francia, which depends heavily on estimates for uncentered maximal operators.

(d) Maximal estimates are also of fundamental importance in the study of various singular integral operators. In many cases, the analysis of such operators involves discretization or approximation of the underlying kernel, which gives rise to sums of various dyadic-type operators (e.g., the so-called sparse operators, a class which has been rapidly developed in the recent literature). These discrete structures are typically controlled by dyadic maximal functions or their appropriate modifications, and here the maximal estimates come into play.

(e) There are also numerous examples of maximal inequalities in the probability theory. These include estimates for sums of independent random variables as well as semimartingale inequalities, which have their further applications to the theory of stochastic integration.

1.1 The dyadic and the dyadic-like maximal operator, and their boundedness properties

Throughout the thesis, we will be mostly concerned with the dyadic context, and some of its extensions arising naturally in the probability theory. Recall the definition: the dyadic maximal operator acts on locally integrable function $f:\mathbb{R}^d\to\mathbb{R}$ by

$$Mf(x)=\sup\left\{\langle|f|\rangle_Q : Q\subset\mathbb{R}^d \text{ is a dyadic cube, } x\in Q\right\}.$$

Here the family of dyadic cubes in $\mathbb{R}^d$ is formed by the grids $(2^{-n}\mathbb{Z}^d)_{n=0,1,2,\ldots}$; in other words, the dyadic lattice is the collection of all cubes of the form

$$[a_1\cdot 2^{-n},(a_1+1)\cdot 2^{-n})\times[a_2\cdot 2^{-n},(a_2+1)\cdot 2^{-n})\times\ldots\times[a_d\cdot 2^{-n},(a_d+1)\cdot 2^{-n}),$$

where $a_1,\,a_2,\,\ldots,\,a_d$ are arbitrary integers and n is a nonnegative integer. One of the key features is that any two dyadic cubes are either disjoint, or one is contained in the other. It turns out that this property allows the use of certain inductive arguments in the study of the operator M. Moreover, there are nice and fruitful connections between the operator M and the probability theory. To see this, let us restrict ourselves to functions supported on the unit cube $[0,1)^d$ and denote by $\mathcal{D}^n$ the family of dyadic subcubes of $[0,1)^d$ having volume equal to 2^{-nd}. This immediately suggests the following important link to the theory of martingales. Namely, consider the probability space $([0,1)^d,\mathcal{B}([0,1)^d),|\cdot|)$ and, for any $n\geq 0$, define the σ-algebra $\mathcal{F}_n=\sigma(\mathcal{D}^n)$ and the function/random variable $f_n=\mathbb{E}(f|\mathcal{F}_n)$. Then $(\mathcal{F}_n)_{n\geq 0}$ forms a filtration, i.e., a nondecreasing family of sub-σ-algebras of $\mathcal{B}([0,1)^d)$. In addition, the sequence $(f_n)_{n\geq 0}$ is a martingale adapted to $(\mathcal{F}_n)_{n\geq 0}$: for any $n\geq 0$, the function/random variable f_n is measurable with respect to $\mathcal{F}_n$ and we have $\mathbb{E}(f_{n+1}|\mathcal{F}_n)=f_n$ almost surely. Finally, for each $n\geq 0$, one easily checks that $Mf_n=\max_{k\leq n}|f|_k$ is the truncated maximal function of the martingale $(|f|_n)_{n\geq 0}$, and similarly, the dyadic maximal operator can be written as $Mf=\sup_{n\geq 0}|f|_n$. In other words, the analysis of dyadic maximal operators and martingale maximal functions are parallel; this interplay enables us to employ various probabilistic tools and interpretations in the study of M.

From the viewpoint of applications discussed above, it is important to study the boundedness of M in various function spaces. This subject has been intensively investigated in the literature (see e.g. [19, 33, 34, 35, 38, 74], consult also the references in the more recent of these works), and the related sub-problem of obtaining sharp, or at least good bounds for the corresponding norms has gained considerable interest. This is one of the main themes in this thesis, so let us present more details on the topic. The first example is that M is bounded as an operator from L^1 to $L^{1,\infty}$; actually, it satisfies the slightly stronger estimate

$$\lambda\left|\left\{x\in\mathbb{R}^d : Mf(x)>\lambda\right\}\right|\leq\int_{\{Mf>\lambda\}}|f(x)|\mathrm{d}x \tag{1.1}$$

for any $f \in L^1(\mathbb{R}^d)$ and $\lambda > 0$. This in particular implies

$$\|Mf\|_{L^{1,\infty}(\mathbb{R}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)},$$

where, for $1 \leq p < \infty$, $\|f\|_{L^{p,\infty}(\mathbb{R}^d)} = \sup_{\lambda>0} \lambda |\{x \in \mathbb{R}^d : |f(x)| > \lambda\}|^{1/p}$. The estimate is sharp: there is a nontrivial function f for which equality is attained. Actually, the weak-type (p,p) inequality holds, with the unchanged constant 1, in the full range $1 \leq p < \infty$ (cf. [36]): we have the sharp bound

$$\|Mf\|_{L^{p,\infty}(\mathbb{R}^d)} \leq \|f\|_{L^p(\mathbb{R}^d)}.$$

The next example is the celebrated Hardy-Littlewood-Doob inequality

$$\|Mf\|_{L^p(\mathbb{R}^d)} \leq \frac{p}{p-1} \|f\|_{L^p(\mathbb{R}^d)}, \quad 1 < p \leq \infty, \tag{1.2}$$

for any $f \in L^p(\mathbb{R}^d)$, in which the constant $p/(p-1)$ is the best possible; see [5, 33]. This result gives rise to a number of interesting problems. For instance, the version of (1.2) does not hold with any finite constant if $p = 1$; as a substitute, one might consider the corresponding weak-type (1,1) estimate (which, as we have just seen above, holds with a constant 1). Motivated by the classical results of Zygmund, one can study a different end-point estimate expressed in terms of sharp LlogL-type estimates. It can be extracted from the works of Melas [34] (see also Osękowski [45]) that if $K > 1$, E is a measurable subset of $\mathbb{R}^d$ and a function $f : \mathbb{R}^d \to \mathbb{R}$ satisfies $\|f\|_{L\log L(\mathbb{R}^d)} = \int_{\mathbb{R}^d} (|f|+1)\log(|f|+1)\mathrm{d}x < \infty$, then there is a finite constant $L(K)$ such that

$$\int_E Mf \mathrm{d}x \leq K\|f\|_{L\log L(\mathbb{R}^d)} + L(K) \cdot |E|.$$

The subsequent work [35] concerns yet another extension of (1.2): the action of M, considered as an operator from $L^p(\mathbb{R}^d)$ to localized $L^q(\mathbb{R}^d)$ (for $1 \leq q < p$), is studied there. Specifically, among other things, Melas determined the best constant $C_{p,q}$ in the following inequality: for any measurable subset E of $\mathbb{R}^d$, we have

$$\left(\int_E (Mf)^q \mathrm{d}x\right)^{1/q} \leq C_{p,q}\|f\|_{L^p(\mathbb{R}^d)} |E|^{1/q-1/p}.$$

The paper [38] by Melas and Nikolidakis extends the above estimate to the wider range of parameters. It is devoted to the following sharp version of Kolmogorov's inequality: for any $0 < q < 1$ and any measurable $E \subset \mathbb{R}^d$,

$$\left(\int_E |Mf|^q \mathrm{d}x\right)^{1/q} \leq \left(\frac{1}{1-q}\right)^{1/q} \|f\|_{L^1(\mathbb{R}^d)} |E|^{1/q-1}.$$

Lorentz-norm estimates for M have also gathered a lot of interest. Let us first provide some necessary definitions and notation. Recall that if f is a measurable function on some measure space (Ω, μ), then its nonincreasing rearrangement $f^* : [0,\infty) \to [0,\infty)$ is given by

$$f^*(t) = \inf\Big\{s > 0 : \mu\big(\{x \in \Omega : |f(x)| > s\}\big) \leq t\Big\}.$$

Note that if $\mu(\Omega) < \infty$, then $f^*(t)$ vanishes for $t > \mu(\Omega)$. Given $0 < p, q < \infty$, we define the Lorentz space $L^{p,q} = L^{p,q}(\Omega, \mu)$ as the family of all (equivalence classes of) measurable functions f on Ω such that

$$\|f\|_{L^{p,q}} := \left(\int_0^\infty \left(t^{1/p} f^*(t)\right)^q \frac{\mathrm{d}t}{t}\right)^{1/q}$$

is finite. The space $L^{p,\infty} = L^{p,\infty}(\Omega, \mu)$ is defined similarly, with the use of the quasinorm

$$\|f\|_{L^{p,\infty}} := \sup_{t>0} t^{1/p} f^*(t).$$

Melas and Nikolidakis [38] proved that for any $1 < p, q < \infty$ we have

$$\|Mf\|_{L^{p,q}(\mathbb{R}^d)} \leq \frac{p}{p-1} \|f\|_{L^{p,q}(\mathbb{R}^d)}.$$

There is also a related estimate concerning the action of M between the spaces $L^{p,\infty} \to L^{q,r}$, see [38, 50, 49] for details. Chapter 6 of this thesis will also be devoted to a result in this direction.

It turns out that all the above results can be extended significantly: the maximal inequalities hold in a much more general context of measure spaces equipped with a tree structure. The following concept generalizes the notion of a dyadic lattice.

Definition 1.1.1 (Tree)**.** Assume that (Ω, μ) is a nonatomic measure space with $\mu(\Omega) < \infty$. A family $\mathcal{T}$ of measurable subsets of Ω is called a *tree* if the following conditions are satisfied:

(i) $\Omega \in \mathcal{T}$ and for every $Q \in \mathcal{T}$ we have $\mu(Q) > 0$.

(ii) For every $Q \in \mathcal{T}$ there is a finite partition $C(Q) \subset \mathcal{T}$ of Q (i.e., the elements of $C(Q)$ are pairwise disjoint subsets of Q and their union is Q).

(iii) $\mathcal{T} = \bigcup_{m \geq 0} \mathcal{T}^m$, where $\mathcal{T}^0 = \{\Omega\}$ and $\mathcal{T}^{m+1} = \bigcup_{Q \in \mathcal{T}^m} C(Q)$.

(iv) We have $\lim_{m\to\infty} \sup_{Q \in \mathcal{T}^m} \mu(Q) = 0$.

A natural example is the cube $\Omega = [0,1)^d$ with the Lebesgue measure μ and the tree of the dyadic cubes contained in $[0,1)^d$. Comparing the above definition to the previous context of $\mathbb{R}^d$ with its dyadic lattice, we see that now we impose the additional finiteness assumption $\mu(\Omega) < \infty$. This extra condition has a technical character and its main purpose is to provide the base point for induction arguments. It should be emphasized that it is not restrictive in most applications: having proved any estimate in the finite, "local" dyadic case, one performs rather standard translation and limiting arguments to obtain the result in the general dyadic case.

Definition 1.1.2 (Maximal operator)**.** Any measure space equipped with a tree gives rise to the corresponding maximal operator $\mathcal{M} = \mathcal{M}_{\Omega,\mu,\mathcal{T}}$, given by

$$\mathcal{M}f(x) = \sup \left\{ \langle |f| \rangle_{Q,\mu} : Q \in \mathcal{T},\ x \in Q \right\},$$

where $\langle f \rangle_{Q,\mu} = \frac{1}{\mu(Q)} \int_Q f \mathrm{d}\mu$ is the average of f over Q with respect to the measure μ. Such operator will be called the dyadic-like maximal operator associated with $\mathcal{T}$.

If $\mu(\Omega) = 1$, i.e., if (Ω, μ) is a probability space, then there is a direct correspondence between tree structures and atomic filtrations $(\sigma(\mathcal{T}^n))_{n \geq 0}$. In particular, all the results can be interpreted in terms of martingales and their maximal functions. The passage between the analytic and the probabilistic case is essentially the same as in the dyadic setup already discussed above, so we omit the details.

It can be shown that all the maximal estimates presented above hold true, with unchanged constants, if we replace the dyadic operator M with the dyadic-like counterpart $\mathcal{M}$ on an arbitrary measure space with tree-like structure. The reason for this is that the works cited above contain much more: they actually identify the explicit formulae for the associated Bellman functions. Let us say a few general words about this approach (for more on the subject, we refer the reader to [33, 42, 43, 46, 72, 73, 76, 77]; for the version of the technique, specialized to the maximal context, see Chapter 2 below). The Bellman function method links a given estimate under investigation to the existence of a certain special function, enjoying appropriate size and concavity requirements: once such an object is constructed, then the exploitation of its properties in an appropriate order yields the inequality. However, in many cases the interplay goes much deeper: the validity of the estimate is actually *equivalent* to the existence of the special function, so in particular, the method can be used to track the best constants involved. In addition, not only does the special function yield the estimate, but it also encodes the extremizers, i.e., the terms for which equality is attained, or almost attained.

To finish the discussion, let us point out here that the Bellman function approach, if applied appropriately, "does not recognize" (or rather: does not refer to) the dyadic splitting and works equally fine for any trees: this yields the maximal estimates in the general setup with no additional effort.

1.2 Weights

Actually, we will be interested in a wide class of estimates, in the presence of some additional objects, the so-called weights. Here and below, the word "weight" refers to a locally integrable, nonnegative function on the base space (X, μ) (e.g., $X = \mathbb{R}^d$ and $\mu = |\cdot|$, or $(X, \mu) = (\Omega, \mu)$ - this will be clear from the context). We will typically denote the weights using the letters u, w or v. Any weight w gives rise to the corresponding measure, also denoted by w, given by $w(E) = \int_E w\mathrm{d}\mu$ for all measurable sets E. Furthermore, the associated weighted L^p spaces are given by

$$L^p(w) = \left\{ f : X \to \mathbb{R} \,:\, \|f\|_{L^p(w)} = \left(\int_X |f|^p w \mathrm{d}\mu \right)^{1/p} < \infty \right\}, \qquad 0 < p < \infty,$$

with the usual identification of the functions which are equal μ-almost everywhere. The case $p = \infty$ is handled in a standard manner, with the use of essential supremum with respect to the measure w. Analogously, one can define the weighted weak-L^p spaces $L^{p,\infty}(w)$, or more generally, the weighted versions of Lorentz spaces, by the use of the decreasing rearrangements relative to the measure w.

One can study the action of maximal operators on various weighted Lorentz spaces. We will focus on the case $X = \mathbb{R}^d$, as the statements we will refer to mostly concern this case; to avoid confusion, the uncentered maximal operator will be denoted by $\mathscr{M}$, and its dyadic version by M. The starting point is the following result, which can be extracted from the work of Fefferman and Stein [16]. For any $f \in L^1(\mathbb{R}^d)$, $\lambda > 0$ and any weight w we have the weak-type bound

$$\lambda w\left(\left\{ x \in \mathbb{R}^d : \mathscr{M} f(x) > \lambda \right\} \right) \leq C_d \int_{\{\mathscr{M} f > \lambda\}} |f(x)| \mathscr{M} w(x) \mathrm{d}x \tag{1.3}$$

for some constant C_d depending only on the dimension. One can show an analogous bound for the dyadic maximal operator M, with the constant 1. This is an extension of (1.1): the classical estimate follows if we take $w \equiv 1$. By a straightforward interpolation argument, one obtains the weighted bounds

$$\|\mathscr{M} f\|_{L^p(w)} \leq \frac{pC_d}{p-1} \|f\|_{L^p(\mathscr{M} w)}, \quad 1 < p \leq \infty,$$

and

$$\|Mf\|_{L^p(w)} \leq \frac{p}{p-1} \|f\|_{L^p(Mw)}, \quad 1 < p \leq \infty. \tag{1.4}$$

Note that the latter estimate generalizes (1.2). This gives rise to the question about the extensions of other estimates from the previous section, to this two-weight ($w - \mathscr{M} w$ and $w - Mw$) context. We will study this problem in Chapters 3 and 4.

There is another, perhaps more natural problem, which concerns the boundedness of $\mathscr{M}$ and M as operators on the weighted space $L^p(w)$. More precisely, suppose that $1 < p < \infty$ is a fixed exponent. It is not difficult to construct a weight w such that $\|\mathscr{M}\|_{L^p(w) \to L^p(w)} = \infty$ and the problem is to provide the characterization of those w, for which the norm is finite. A similar question can be posed if one replaces $\mathscr{M}$ with the dyadic maximal operator. The former problem was successfully handled by Muckenhoupt [39] in the beginning of the seventies: the uncentered maximal operator is bounded as an operator on $L^p(w)$ if and only the weight w belongs to the class A_p (satisfies Muckenhoupt's condition A_p). The latter means that the A_p characteristic of w, given by

$$[w]_{A_p}^{general} := \sup\left\{ \langle w \rangle_Q \left\langle w^{1/(1-p)} \right\rangle_Q^{p-1} : Q \subset \mathbb{R}^d \text{ is a cube with sides parallel to the axes} \right\}$$

is finite. It turns out that the answer of the corresponding question concerning the dyadic maximal operator requires only some minor modifications: the dyadic maximal operator is bounded as an operator on $L^p(w)$ if and only if the weight w belongs to the dyadic A_p class, i.e.,

$$[w]_{A_p}^{dyadic} := \sup\left\{ \langle w \rangle_Q \left\langle w^{1/(1-p)} \right\rangle_Q^{p-1} : Q \subset \mathbb{R}^d \text{ is a dyadic cube} \right\} < \infty. \tag{1.5}$$

From now on, we will simply write $[w]_{A_p}$ for the characteristics, it should be clear which context we are working in.

Muckenhoupt's result is considered to be the cornerstone of the weighted theory and it has been subject to numerous extensions and generalizations. It turns out that the condition A_p characterizes the boundedness of other important classical operators in harmonic analysis. For example, the sufficiency of Muckenhoupt's condition for the boundedness of the Hilbert transform was proved by Hunt, Muckenhoupt and Wheeden [20], while the setting of Riesz transforms (actually, even a more wider class of singular integrals) was handled by Coifman and Fefferman [8]. (Proofs of the necessity can be found in [17, 20, 75].) The weighted estimates for fractional and Poisson integrals were investigated by Sawyer [70, 71], the analysis of square function operators can be found in the works of Buckley [3], Chanillo and Wheeden [7] and Lerner [26].

Another extension of Muckenhoupt's result, which has gained a lot of interest in the recent literature, concerns the optimal dependence of the norm $\|\mathscr{M}\|_{L^p(w)\to L^p(w)}$ and its dyadic counterpart $\|M\|_{L^p(w)\to L^p(w)}$ on the size of the characteristic $[w]_{A_p}$. More precisely, for a given $1 < p < \infty$, the problem is to find the least exponent $\alpha = \alpha(p)$ such that

$$\|\mathscr{M} f\|_{L^p(w)} \leq C_p[w]_{A_p}^{\alpha(p)} \|f\|_{L^p(w)} \tag{1.6}$$

where the constant C_p depends only on p. This question was posed and answered in the nineties by Buckley [3]: he showed that the optimal exponent $\alpha(p)$ is equal to $1/(p-1)$. Again, one can study analogous problems, replacing the maximal function with other important operators of harmonic analysis; we only mention here the works of Hÿtonen [21] and Lerner [27] for some information on Calderón-Zygmund singular integrals, Lacey et. al. [25] for the study of fractional integrals and Lerner [26] for the context of Littlewood-Paley square functions. Consult also the references in these works.

Coming back to the maximal functions, we would like to mention the further improvement to (1.6) obtained by Osękowski: the paper [54] contains, for any $1 < p < \infty$ and any $c \geq 1$, the identification of the optimal constant $C_{p,c}$ such that

$$\|M\|_{L^p(w)\to L^p(w)} \leq C_{p,[w]_{A_p}}. \tag{1.7}$$

In the thesis, we will obtain results related to this estimate.

We would like to point out that there are also versions of the A_p condition in the endpoint cases $p \in \{1, \infty\}$. We will focus here on the case $p = 1$, as the A_∞ condition will not appear in our considerations below. Namely, w is an A_1 weight if

$$[w]_{A_1} := \operatorname*{esssup}_{\mathbb{R}^d} \frac{\mathscr{M} w}{w}$$

is finite; an obvious modification leads to the dyadic A_1 weights. It seems natural to expect that in the boundary case $p = 1$ we should have some weighted weak-type bound for $\mathscr{M}$ and M. The Fefferman-Stein inequality (1.3) (and its dyadic version) shows that this is the case: it immediately yields the one-weight bound

$$\lambda w\left(\left\{x \in \mathbb{R}^d : \mathscr{M} f(x) > \lambda\right\}\right) \leq C_d[w]_{A_1} \int_{\{\mathscr{M} f>\lambda\}} |f(x)|w(x)\mathrm{d}x, \tag{1.8}$$

along with its dyadic version. Furthermore, it is easy to show that the linear dependence on the characteristic is the best.

The estimates (1.3) and (1.8) can be pushed in a very interesting direction, by means of appropriate duality arguments. Motivated by the works of Lerner et. al. [29, 30, 31], we consider the strong dual version

$$w\Big(\{x \in \mathbb{R}^d : \mathscr{M} f(x) \geq \mathscr{M} w(x)\}\Big) \leq C_d \int_{\mathbb{R}^d} |f|\mathrm{d}x, \tag{1.9}$$

where w is an arbitrary weight and C_d depends only on the dimension. The weak dual inequality concerns A_1 weights and reads

$$w\Big(\{x \in \mathbb{R}^d : \mathscr{M} f(x) \geq w(x)\}\Big) \leq C_d[w]_{A_1} \int_{\mathbb{R}^d} |f|\mathrm{d}x. \tag{1.10}$$

It is not difficult to establish these estimates in the uncentered and the dyadic contexts. The aforementioned works of Lerner et. al. contained the analysis of analogous inequalities for general Calderón-Zygmund singular integral operators, which are far more challenging. The contribution of the thesis in this direction, presented in Chapter 5, concerns a version of (1.10) in the context of dyadic A_p weights with $p > 1$.

All the results for the dyadic maximal operators formulated above can be studied in the more general setup of measure spaces (Ω, μ) with a tree-like structure $\mathcal{T}$. All the discussion can be carried out with no difficulty: one needs to treat a weight as a measurable and non-negative function on Ω, the only essential change concerns the Muckenhoupt's condition. The modification is straightforward: for a weight w and an exponent $1 < p < \infty$, we define

$$[w]_{A_p} := \sup \left\{ \langle w \rangle_{Q,\mu} \left\langle w^{\frac{-1}{p-1}} \right\rangle_{Q,\mu}^{p-1} : Q \in \mathcal{T} \right\} < \infty.$$

The boundary case $p = 1$ is handled in an obvious manner.

1.3 The harmonic maximal operator

There is another interesting version of the maximal function, the so-called dyadic *harmonic* maximal operator $M^{\mathcal{H}}$, which is defined by the identity

$$M^{\mathcal{H}} f(x) = \sup \left\{ \langle |f|^{-1} \rangle_Q^{-1} : Q \subset \mathbb{R}^d \text{ is a dyadic cube, } x \in Q \right\}.$$

This definition generalizes easily to the context of a measure space (Ω, μ) with a tree $\mathcal{T}$: set

$$\mathcal{M}^{\mathcal{H}}_{\Omega,\mu,\mathcal{T}} f(x) = \sup \left\{ \langle |f|^{-1} \rangle_{Q,\mu}^{-1} : Q \in \mathcal{T},\ x \in Q \right\}.$$

Here and below, we use the convention $1/0 = \infty$ and $1/\infty = 0$. The joint behavior of (and the interplay between) M and $M^{\mathcal{H}}$ is similar to that of the arithmetic and the harmonic averages

$$\frac{|x_1| + |x_2| + \ldots + |x_n|}{n}, \qquad \left(\frac{|x_1|^{-1} + |x_2|^{-1} + \ldots + |x_n|^{-1}}{n} \right)^{-1},$$

where $x_1, x_2, \ldots, x_n$ are arbitrary real numbers. In particular, we have the pointwise estimate $Mf \geq M^{\mathcal{H}} f$ on $\mathbb{R}^d$. The harmonic maximal operators appeared for the first time in the works [9, 10, 11] in a slightly different form: the authors studied there the so-called minimal operator

$$\mathfrak{M} f(x) = \inf \left\{ \langle |f| \rangle_Q : Q \subset \mathbb{R}^d \text{ is a dyadic cube, } x \in Q \right\},$$

which is linked to $M^{\mathcal{H}}$ via the identity $M^{\mathcal{H}} f = \mathfrak{M}(|f|^{-1})^{-1}$. In a sense, the minimal operator $\mathfrak{M}$ controls f on the set where the function is small (while the maximal operator M controls f where the function is large). The minimal operator was used to study the fine structure of A_p weights in [10], further applications to weighted norm inequalities and differentiation theory can be found in [11].

One can ask about sharp versions of the estimates from the previous subsections, with $\mathscr{M}$ and M replaced with $\mathcal{M}^{\mathcal{H}}$. To the best of our knowledge, very little is known about this topic. The paper [23] contains the proof of sharp weak- and strong-type estimates for $M^{\mathcal{H}}$. The (unweighted) L^p norms of $\mathcal{M}^{\mathcal{H}}$ can be extracted from the appropriate general Φ-estimates for martingale maximal functions, see Chapter 7 in [45]. There is a natural question about the weighted inequalities for $\mathcal{M}^{\mathcal{H}}$. As shown in [11], for any fixed $0 < p < \infty$, the operator $\mathcal{M}^{\mathcal{H}}$ is bounded as an operator from $L^p(v)$ to $L^p(u)$ if and only if the pair (u, v) of weights satisfies

$$[u, v]_{A_{-p}} := \sup \left\{ \langle u \rangle_{Q,\mu} \left\langle v^{\frac{1}{p+1}} \right\rangle_{Q,\mu}^{-p-1} : Q \in \mathcal{T} \right\} < \infty$$

(with the convention $0 \cdot 0^{-p-1} = 0$). See also Duffee and Moen [15]. We will provide a certain sharp two-weight estimate in this direction.

1.4 The organization and the contribution of the thesis

The remaining part of this thesis has been divided into six separate chapters: let us briefly discuss their contents and indicate the main results. All the estimates are obtained in the context of dyadic-like operators on measure spaces equipped with trees.

Chapter 2 is devoted to the detailed description of the Bellman function method, which will play a distinguished role in our considerations. The material presented in this chapter is not new, it is a compilation of several texts from the literature, including [43, 45].

Chapter 3 handles the weighted Kolmogorov's inequalities for maximal functions. Given $p < 1$ and an arbitrary weight w, we have the sharp bound

$$\|\mathcal{M}f\|_{L^p(w)}^p \leq \frac{1}{1-p} \|f\|_{L^1}^p \|w\|_{L^1} + \frac{p^2}{1-p} E_{\mathcal{T}}(f, w),$$

where $E_{\mathcal{T}}(f, w)$ is an appropriate error term. This term is equal to zero when $w = \text{const}$, so it is indeed a generalization of the unweighted Kolmogorov's inequality. One of the very interesting additional features is the associated Bellman function, which has quite an unusual form. The contents of this chapter is taken from [64].

Chapter 4 shows that any integral inequality for $\mathcal{M}$ of a rather general form automatically self-improves into a weighted estimate with an arbitrary weight w on one side and $\mathcal{M}w$ on the other, in analogy to the Fefferman-Stein inequality (1.3) being an extension of (1.1). The contents of this chapter is based on [65].

Chapter 5 contains the proof of the weak-type inequality

$$w\left(\{x \in \Omega : \mathcal{M}f \geq \mathcal{M}w\}\right) \leq C_p [w]_{A_p} \int_\Omega f \mathrm{d}\mu,$$

where $p > 1$ and w is an A_p weight. The linear dependence on the A_p characteristic is optimal. The contents of this chapter is taken from [60].

Chapter 6 is dedicated to the study of $\mathcal{M}$ as an operator between unweighted Lorentz spaces. More precisely, we obtain an explicit formula for $\|\mathcal{M}\|_{L^{p,q_1} \to L^{p,q_2}}$, where $1 < p \leq q_1 < q_2 < \infty$. The contents of this chapter is taken from [59].

The final part of the thesis, Chapter 7, investigates the weighted inequalities for the harmonic maximal operator. More specifically, we will obtain the following sharp two-weight L^p estimate for $\mathcal{M}^{\mathcal{H}}$. Suppose that $p > 0$ and (u, v) is a pair of weights satisfying

$$[u, v]_{A_{-p}} := \sup \left\{ \langle u \rangle_{Q,\mu} \left\langle v^{\frac{1}{p+1}} \right\rangle_{Q,\mu}^{-p-1} : Q \in \mathcal{T} \right\} < \infty.$$

As we have mentioned above, this condition guarantees that the operator $\mathcal{M}^{\mathcal{H}}$ is bounded as an operator from $L^p(v)$ to $L^p(u)$. Motivated by (1.7), one study and answer the question about the optimal bound for $\|\mathcal{M}^{\mathcal{H}}\|_{L^p(v)\to L^p(u)}$ in terms of the joint characteristic $[u,v]_{A_{-p}}$. The contents of this chapter is taken from [61].

Chapter 2

Bellman function method

In the literature, estimates for maximal operators have been studied with the use of various techniques: these include, for example, covering theorems, Calderón-Zygmund-type decompositions, interpolation, and many more. One of the main tools exploited in the thesis is the Bellman function method. This technique has proved to be very powerful and efficient in the investigation of numerous problems of harmonic analysis and probability theory. From the historical point of view, the approach has its origins in the theory of optimal stochastic control developed by Bellman [1]. Probably the first results, which explored the connections of the method with other areas of mathematics, were those of Burkholder [4], who used it to show sharp estimates for martingale transforms and to identify the unconditional constants of the Haar system. Soon after the appearance of the seminal paper [4], Burkholder's arguments were extended from the martingale setting to the context of general semimartingales: see e.g. the monograph [45] for an overview. This can be regarded as a probabilistic direction for the Bellman function method. In the nineties, a decisive step was made by Nazarov and Treil [42], inspired by the preprint version of [43]. In that paper, the technique was pushed towards quite general applications in harmonic analysis; since then, the method has been used in many contexts, including the properties of BMO spaces, weighted estimates, properties of Carleson sequences, and many, many more.

Let us roughly explain the main idea behind the approach. The Bellman function method relates the validity of a given estimate to the existence of a certain special function which enjoys appropriate size and concavity requirements. Let us be a little more specific here. Typically, the study of a given inequality is equivalent to finding the supremum of some quantity under some fixed parameters. The maximal value of this quantity is just by definition the value of the Bellman function (for this extremal problem), and the parameters are the arguments of this function. A trivial choice for these parameters gives the corresponding lower bound for the function, which is typically called the obstacle condition (or the majorization property). The key fact is that in many cases the extremal problem is self-similar under scaling; therefore, the definition of the Bellman function immediately leads to some inequality of concavity type. The Bellman function appears to be the minimal possible among all the functions satisfying this concavity condition. This extremal property implies that the concavity must degenerate in some directions (here the so-called Bellman equation comes into play), which yields a nonlinear partial differential equation for the Bellman function. This equation sometimes can be solved explicitly, which gives the formula for the Bellman function and establishes the initial estimate.

Typically, the Bellman function is quite a complicated object. The discovery of the degenerate directions, as well as finding the solution to the corresponding partial differential equation, is in general a quite elaborate issue. The difficulty grows significantly with the dimension of the problem: there are many results in which a Bellman function of two variables was calculated successfully; there are only a few papers handling the three-dimensional context, and, to the best of our knowledge, no examples in higher dimensions.

In the case when the identification of the Bellman function is too difficult, one can study the less challenging problem of finding the so-called *supersolution*, i.e., a majorant which satisfies the same concavity and obstacle conditions, but not necessarily the underlying Bellman equation. Sometimes such sub-extremal objects are also called Bellman functions in the literature, although they do not satisfy the appropriate minimality. It is also worth pointing out that in many cases they also lead to sharp-constant estimates and have a simpler, less technical formula, so from the viewpoint of the estimate under investigation, they are more convenient to work with. However, in most situations the supersolutions do not yield the optimal constants, but, on the positive side, they can still give the asymptotic sharpness with respect to some parameter (e.g., the characteristic of a weight: see (1.6) above).

We restrain ourselves from the further general discussion and turn our attention to the specific context of maximal inequalities. We have decided to present the details in the analytic language, but it is easy to translate them all into the probabilistic context of martingales.

2.1 Bellman function method for unweighted estimates

For the sake of simplicity and clarity, we have decided to start with the presentation of the method in the case when the weighted ingredient does not appear. Suppose that (Ω, μ) is a fixed finite measure space equipped with a tree $\mathcal{T}$ and the associated dyadic-like maximal operator $\mathcal{M}$. Actually, we will restrict ourselves to probability spaces: in most cases, this normalization can be imposed with no loss of generality. A function f on $\mathcal{T}$ is called $\mathcal{T}$-simple, if it is measurable with respect to $\sigma(\mathcal{T}^N)$ for some integer N. Note that in particular, $\mathcal{T}$-simple functions have only finitely many values. Consider the angular domain $D = \{(x, y) \in [0, \infty) \times [0, \infty) : x \leq y\}$. Suppose that $V : D \to \mathbb{R}$ is a given function and assume that we are interested in showing the estimate

$$\int_\Omega V(f, \mathcal{M}f)\mathrm{d}\mu \leq 0 \tag{2.1}$$

for all nonnegative $\mathcal{T}$-simple functions f on Ω. Note that the pair $(f, \mathcal{M}f)$ takes values in D, so the integrand on the left makes sense. For example, the choice $V(x, y) = y^p - \left(\frac{p}{p-1}\right)^p x^p$, $1 < p < \infty$, corresponds to the strong-type estimates (1.2). Another example is $V(x, y) = \chi_{(\lambda,\infty)}(y)(\lambda - x)$, $\lambda > 0$, which leads to the weak-type bound (1.1). Note that the requirement $f \geq 0$ imposed above is not excessive: in all of the maximal estimates discussed in the previous chapter, the dependence on the function f was through $|f|$, in other words, the passage from f to $|f|$ does not change anything. Furthermore, the restriction to $\mathcal{T}$-simple functions removes all the inconvenient problems concerning the measurability or integrability on the left of (2.1). On the other hand, in all the relevant examples the passage from the simple to the general case is a matter of standard approximation arguments.

The underlying idea of the approach towards (2.1) is the following: we look for a function $B : D \to \mathbb{R}$ that satisfies the following three properties.

1° (Initial condition). For any $x \geq 0$ we have $B(x,x) \leq 0$.

2° (Majorization). We have $B \geq V$ on D.

3° (Concavity). For any $(x,y) \in D$, any numbers $x_1, x_2 \geq 0$ and $\lambda_1, \lambda_2 \in (0,1)$ such that $\lambda_1 + \lambda_2 = 1$ and $\lambda_1 x_1 + \lambda_2 x_2 = x$, we have

$$B(x,y) \geq \lambda_1 B(x_1, y \vee x_1) + \lambda_2 B(x_2, y \vee x_2), \tag{2.2}$$

where $a \vee b$ is the maximum of a and b. Let us study precisely the relation between the validity of (2.1) and the existence of a function B possessing the above three properties.

Theorem 2.1.1. *If there is a function B which satisfies* 1°, 2° *and* 3°, *then* (2.1) *holds true.*

Proof. We start with a simple generalization of the condition 3° to a finite number of variables. A straightforward induction implies that for any $m \geq 2$, any $(x,y) \in D$ and any numbers $x_k \in [0,\infty),\ \lambda_k \in (0,1),\ k = 1,2,\ldots,m$ satisfying $\sum_{k=1}^m \lambda_k = 1$ and $\sum_{k=1}^m \lambda_k x_k = x$, we have

$$B(x,y) \geq \sum_{k=1}^m \lambda_k B(x_k, y \vee x_k). \tag{2.3}$$

For any $n \geq 0$, we will use the notation $f_n = \sum_{Q \in \mathcal{T}^n} \langle f \rangle_{Q,\mu} \chi_Q$ (in the probabilistic language, $(f_n)_{n\geq 0}$ is the martingale associated with f). The key ingredient of the proof of the theorem is to show that the sequence $\left(\int_\Omega B(f_n, \mathcal{M}f_n)\mathrm{d}\mu\right)_{n\geq 0}$ is nonincreasing. To prove this fact, fix $n \geq 0$, an element $Q \in \mathcal{T}^n$ and denote the children of Q in $\mathcal{T}^{n+1}$ by $Q_1, Q_2, \ldots, Q_m$. The functions f_n and $\mathcal{M}f_n$ are constant on Q: denote the corresponding values by x and y. Similarly, f_{n+1} and $\mathcal{M}f_{n+1}$ are constant on each Q_k: denoting the value of $f_{n+1}|_{Q_k}$ by x_k, one easily checks that $\mathcal{M}f_{n+1} = y \vee x_k$ on Q_k. Let us check that the conditions listed above (2.3) are satisfied, with $\lambda_k = \mu(Q_k)/\mu(Q)$. The numbers λ_k sum up to 1 and

$$x = \frac{1}{\mu(Q)} \int_Q f \mathrm{d}\mu = \sum_{k=1}^m \frac{\mu(Q_k)}{\mu(Q)} \cdot \frac{1}{\mu(Q_k)} \int_{Q_k} f \mathrm{d}\mu = \sum_{k=1}^m \lambda_k x_k.$$

Hence we can apply (2.3) and obtain

$$\begin{aligned}\int_Q B(f_{n+1}, \mathcal{M}f_{n+1})\mathrm{d}\mu &= \sum_{k=1}^m \int_{Q_k} B(f_{n+1}, \mathcal{M}f_{n+1})\mathrm{d}\mu = \sum_{k=1}^m \mu(Q_k) B(x_k, y \vee x_k) \\ &= \mu(Q) \sum_{k=1}^m \lambda_k B(x_k, y \vee x_k) \leq \mu(Q) B(x,y) = \int_Q B(f_n, \mathcal{M}f_n)\mathrm{d}\mu.\end{aligned}$$

Summing over all $Q \in \mathcal{T}^n$ we get the aforementioned monotonicity property

$$\int_\Omega B(f_{n+1}, \mathcal{M}f_{n+1})\mathrm{d}\mu \leq \int_\Omega B(f_n, \mathcal{M}f_n)\mathrm{d}\mu.$$

To finish the proof, fix a large integer N such that f is $\sigma(\mathcal{T}^N)$-measurable. Applying 2°, then the above monotonicity, and finally 1°, we get

$$\begin{aligned}\int_\Omega V(f, \mathcal{M}f)\mathrm{d}\mu &= \int_\Omega V(f_N, \mathcal{M}f_N)\mathrm{d}\mu \\ &\leq \int_\Omega B(f_N, \mathcal{M}f_N)\mathrm{d}\mu \leq \int_\Omega B(f_0, \mathcal{M}f_0)\mathrm{d}\mu = B(f_0, f_0) \leq 0\end{aligned} \tag{2.4}$$

and the proof is finished. □

The implication of the above theorem can be reversed. The Bellman function $\mathfrak{B} : D \to \mathbb{R}$ associated with the problem (2.1) is defined by

$$\mathfrak{B}(x, y) = \sup \int_\Omega V(f, y \vee \mathcal{M}f)\mathrm{d}\mu, \tag{2.5}$$

where the supremum is taken over the class of all nonnegative, $\mathcal{T}$-simple functions f satisfying $\int_\Omega f\mathrm{d}\mu = x$. The tree $\mathcal{T}$ and the probability space (Ω, μ) are also allowed to vary in the above supremum.

Theorem 2.1.2. *If* (2.1) *holds true for any tree $\mathcal{T}$ on any probability space (Ω, μ) and any $\mathcal{T}$-simple, nonnegative function f, then $\mathfrak{B}$, given by* (2.5)*, satisfies $1°$, $2°$ and $3°$. Actually, it is the least function on D enjoying these conditions.*

Proof. The condition $1°$ is a direct consequence of the validity of (2.1): for any $x \geq 0$ and any $\mathcal{T}$-simple function $f \geq 0$ with $\int_\Omega f\mathrm{d}\mu = x$, we have $\mathcal{M}f_n \geq x$ and

$$\int_\Omega V(f_n, x \vee \mathcal{M}f_n)\mathrm{d}\mu = \int_\Omega V(f_n, \mathcal{M}f_n)\mathrm{d}\mu \leq 0.$$

Thus, taking the supremum over all such f gives the initial condition. The majorization $2°$ follows by considering the constant function $f \equiv x$ in the definition of $\mathfrak{B}(x, y)$. The main difficulty lies in showing the concavity condition $3°$. Fix $(x, y) \in D$ and numbers $\lambda_1, \lambda_2, x_1, x_2$ as in the statement of the condition. Let f^1, f^2 be nonnegative functions as in the definition of $\mathfrak{B}(x_1, y \vee x_1)$ and $\mathfrak{B}(x_2, y \vee x_2)$, respectively. We may assume that these functions are given on some disjoint probability spaces (Ω_1, μ_1) and (Ω_2, μ_2) endowed with the trees $\mathcal{T}_1, \mathcal{T}_2$, and that f^1 is $\mathcal{T}_1$-simple and f^2 is $\mathcal{T}_2$-simple. We splice these probability spaces into one space (Ω, μ), where $\Omega = \Omega_1 \cup \Omega_2$ and $\mu(A_1 \cup A_2) = \lambda_1\mu_1(A_1) + \lambda_2\mu_2(A_2)$ for any $A_1 \in \sigma(\mathcal{T}_1)$ and $A_2 \in \sigma(\mathcal{T}_2)$. This new probability space is equipped with the tree $\mathcal{T}$ defined by $\mathcal{T}^0 = \{\Omega\}$ and $\mathcal{T}^n = \mathcal{T}_1^{n-1} \cup \mathcal{T}_2^{n-1}$ for $n \geq 1$. Next, let us splice the functions f^1, f^2 into one function f on the new probability space, setting $f = f^1\chi_{\Omega_1} + f^2\chi_{\Omega_2}$. This new function is $\mathcal{T}$-simple and satisfies $f_0 = \lambda_1 x_1 + \lambda_2 x_2 = x$ and $f_n = f^1_{n-1}\chi_{\Omega_1} + f^2_{n-1}\chi_{\Omega_2}$ for $n \geq 1$. Denote the corresponding dyadic-like maximal operators by $\mathcal{M}$, $\mathcal{M}_1$ and $\mathcal{M}_2$. Since $x \leq y$, we see that for $n \geq 1$ we have

$$y \vee \mathcal{M}f_n = (y \vee \mathcal{M}_1 f^1_{n-1})\chi_{\Omega_1} + (y \vee \mathcal{M}_2 f^2_{n-1})\chi_{\Omega_2}$$

and hence $y \vee \mathcal{M}f = (y \vee \mathcal{M}_1 f^1)\chi_{\Omega_1} + (y \vee \mathcal{M}_2 f^2)\chi_{\Omega_2}$. Thus, by the very definition of $\mathfrak{B}$,

$$\mathfrak{B}(x, y) \geq \int_\Omega V(f, y \vee \mathcal{M}f)\mathrm{d}\mu = \lambda_1 \int_{\Omega_1} V(f^1, y \vee \mathcal{M}_1 f^1)\mathrm{d}\mu_1 + \lambda_2 \int_{\Omega_2} V(f^2, y \vee \mathcal{M}_2 f^2)\mathrm{d}\mu_2,$$

and taking the supremum over all f^1, f^2 as above, we get the desired concavity condition. To see that $\mathfrak{B}$ has the minimality property, fix an arbitrary function B enjoying $1°$, $2°$ and $3°$. For a given $(x, y) \in D$, we repeat the arguments leading to (2.4), replacing $\mathcal{M}f$ with $y \vee \mathcal{M}f$ in all the places. As the result, we obtain the estimate

$$\int_\Omega V(f, y \vee \mathcal{M}f)\mathrm{d}\mu \leq B(f_0, y \vee f_0) = B(x, y)$$

and taking the supremum over all f, we get the pointwise bound $\mathfrak{B}(x, y) \leq B(x, y)$. □

The above two theorems illustrate well the heart of the Bellman function method. The validity of the estimate (2.1) is indeed equivalent to the existence of the corresponding special function. Furthermore, it is also well-visible that the concavity of the Bellman function is intricately connected to the self-similarity of the estimate. It originates in the fact that if we split an element Q into its children Q_1, Q_2, $\ldots$, Q_m, then the analysis of the term $\int_{Q_k} V(f, \mathcal{M}f)\mathrm{d}\mu$ for each of Q_k is 'parallel' to that on Q (and, in turn, also on Ω). The parameter y plays the role of the 'memory' of the maximal function, which is key when passing from Q to Q_k.

Remark 2.1.3. Let us say a few words about a certain modification of the approach, which will become very important for our considerations in Chapter 6. For the sake of clarity, we will present the argument for the L^p inequality

$$\|\mathcal{M}f\|_{L^p} \leq C_p\|f\|_{L^p}, \qquad 1 < p < \infty. \tag{2.6}$$

This estimate can be studied with the above method (the choice $V(x,y) = y^p - C_p^p x^p$ reduces the estimate to (2.1)). But we can also proceed by considering a slightly different Bellman function $\mathfrak{B}'$, defined for $(x, y, z) \in [0,\infty)^3$ satisfying $x^p \leq z$ by

$$\mathfrak{B}'(x,y,z) = \sup\left\{\int_\Omega (y \vee \mathcal{M}f)^p \mathrm{d}\mu\right\}.$$

Here the supremum is taken over all nonnegative functions f satisfying $\int_\Omega f\mathrm{d}\mu = x$ and $\int_\Omega f^p\mathrm{d}\mu = z$. One can check that $\mathfrak{B}'$ satisfies

1° (Initial condition). For any $x \geq 0$ we have $\mathfrak{B}'(x,y,z) \leq C_p^p z$.

2° (Majorization). We have $\mathfrak{B}'(x,y,x^p) \geq y^p$ for all x, y.

3° (Concavity). For any (x,y,z), any numbers $x_1, x_2 \geq 0$, $z_1 \geq x_1^p$, $z_2 \geq x_2^p$ and $\lambda_1, \lambda_2 \in (0,1)$ such that $\lambda_1 + \lambda_2 = 1$, $\lambda_1 x_1 + \lambda_2 x_2 = x$ and $\lambda_1 z_1 + \lambda_2 z_2 = z$, we have

$$\mathfrak{B}'(x,y,z) \geq \lambda_1\mathfrak{B}'(x_1, y \vee x_1, z_1) + \lambda_2\mathfrak{B}'(x_2, y \vee x_2, z_2). \tag{2.7}$$

Furthermore, the existence of *some* function B which satisfies the above set of requirements yields the validity of (2.6). We omit the details, and nor we discuss the interesting interplay between the Bellman functions $\mathfrak{B}$ and $\mathfrak{B}'$, it would lead us beyond the scope of the thesis. We only would like to mention that, obviously, the function $\mathfrak{B}'$ is more difficult to handle: it involves more variables. The passage from this context to the simpler $\mathfrak{B}$ ("dropping the variable z") is achieved by moving (the p-th power of) the right-hand side of (2.6) to the left and inserting it into the function V. We should emphasize that such a maneuver might not always be possible (see Chapter 6), but instead the version described in this remark might be available.

Let us conclude with a simple variant which enables to handle the regularity of the tree. The above lemmas work with no structural properties on $\mathcal{T}$, but these can be included, which yields an interesting modification of method for special filtrations.

Definition 2.1.4. Let $\alpha \in (0, 1/2]$ be a fixed parameter. A tree $\mathcal{T}$ on (Ω, μ) is called α-regular, if for any $n \geq 0$, any $Q \in \mathcal{T}^n$ and its child $Q' \in \mathcal{T}^{n+1}$, we have $\mu(Q') \geq \alpha\mu(Q)$.

A classical example is the dyadic tree on $[0,1)^d$, which is 2^{-d}-regular. Here is the modification of the Bellman function method which takes into account the α-regularity. The proof is analogous to that above and is omitted.

Theorem 2.1.5. *Let $\alpha \in (0, 1/2]$ be a fixed parameter. The inequality* (2.1) *holds true for all α-regular trees if and only if there exists a function B satisfying 1°, 2° and a version of 3° in which the parameters λ_1 and λ_2 are assumed to be at least α.*

In particular, let us restrict ourselves to the one-dimensional dyadic setting. Consider the probability space $[0, 1)$ with the Lebesgue measure equipped with the dyadic tree $\mathcal{D}$. The Bellman function $\mathfrak{B} : D \to \mathbb{R}$ associated with the problem (2.1) is defined by

$$\mathfrak{B}(x, y) = \sup \int_0^1 V(f(t), y \vee Mf(t))\mathrm{d}t, \tag{2.8}$$

where the supremum is taken over the set of all nonnegative, $\mathcal{D}$-simple functions f with $\int_0^1 f(t)\mathrm{d}t = x$. We have the following fact.

Lemma 2.1.6. *If* (2.1) *holds true for the dyadic tree $\mathcal{D}$ on the probability space* $[0, 1)$ *with the Lebesgue measure and any $\mathcal{D}$-simple, nonnegative function f, then $\mathfrak{B}$ given by* (2.8) *satisfies* $1°$, $2°$, *and a weaker version of* $3°$, *where the numbers λ_1, λ_2 are fixed to be equal to* $1/2$.

2.2 Bellman function method for weighted estimates

Now we will show how to modify the above approach in the presence of the A_p weights. Suppose that (Ω, μ) is an arbitrary probability space equipped with a tree-like structure $\mathcal{T}$. Let $c \in [1, \infty)$, $p \in (1, \infty)$ be given parameters and let $V : [0, \infty)^3 \to \mathbb{R}$ be a fixed function. Assume further that we are interested in showing the estimate

$$\int_\Omega V\left(f, \mathcal{M}f, w\right) \mathrm{d}\mu \leq 0 \tag{2.9}$$

for any $\mathcal{T}$-simple function $f : \Omega \to [0, \infty)$ and any $\mathcal{T}$-simple A_p weight w on Ω satisfying $[w]_{A_p} \leq c$. The appearance of the weight increases the dimension of the problem from two to four. Consider the domain $D = D_{p,c} = \left\{(x, y, u, v) \in [0, \infty)^4 \,:\, x \leq y,\, 1 \leq uv^{p-1} \leq c\right\}$ and the class of all special functions $B : D \to \mathbb{R}$ which enjoy the following structural properties.

1° (Initial condition). We have

$$B(x, x, u, v) \leq 0 \qquad \text{if } (x, x, u, v) \in D. \tag{2.10}$$

2° (Majorization). If $0 \leq x \leq y$, then

$$B(x, y, u, u^{1/(1-p)}) \geq V(x, y, u). \tag{2.11}$$

3° (Concavity). Let λ_1, λ_2, …, $\lambda_m \geq 0$ be nonnegative numbers summing up to 1 and let (x, y, u, v), (x_1, y_1, u_1, v_1), …, (x_m, y_m, u_m, v_m) be elements of D enjoying the following conditions: we have $y_j = \max\{x_j, y\}$ for all $j = 1,\ 2,\ \ldots,\ m$ and

$$x = \sum_{k=1}^m \lambda_k x_k, \qquad u = \sum_{k=1}^m \lambda_k u_k, \qquad v = \sum_{k=1}^m \lambda_k v_k.$$

Then we have

$$B\left(x, y, u, v\right) \geq \sum_{k=1}^m \lambda_k B(x_k, y_k, u_k, v_k). \tag{2.12}$$

The existence of such a function is equivalent to the validity of (2.9). As previously, we study each implication separately.

Theorem 2.2.1. *Let $1 < p < \infty$ be fixed. If there is a function B satisfying 1°, 2° and 3°, then* (2.9) *holds true for any probability space (Ω, μ) with a tree $\mathcal{T}$, any $\mathcal{T}$-simple function $f : \Omega \to [0, \infty)$ and any $\mathcal{T}$-simple weight $w \in A_p$ satisfying $[w]_{A_p} \leq c$.*

Proof. The argument goes along the same lines as in the unweighted setting. Fix (Ω, μ), $\mathcal{T}$ and any f, w as in the statement, and suppose that f and w are $\sigma(\mathcal{T}^N)$-measurable. For any $n \geq 0$, define the functions f_n, g_n, w_n and z_n on Ω by $f_n = \sum_{Q \in \mathcal{T}^n} \langle f \rangle_{Q,\mu} \chi_Q$, $g_n = \mathcal{M} f_n$, $w_n = \sum_{Q \in \mathcal{T}^n} \langle w \rangle_{Q,\mu} \chi_Q$ and $z_n = \sum_{Q \in \mathcal{T}^n} \langle w^{1/(1-p)} \rangle_{Q,\mu} \chi_Q$. It is easy to see that (f_n, g_n, w_n, z_n) takes values in the domain D: this is the consequence of the inequality $[w]_{A_p} \leq c$.

As previously, the main part of the proof is to show the monotonicity of an appropriate sequence. Using the same arguments as in the unweighted context, one shows that the sequence $\left(\int_\Omega B(f_n, g_n, w_n, z_n) \mathrm{d}\mu\right)_{n \geq 0}$ is nonincreasing. Combining this with the initial condition, we get

$$\int_\Omega B(f_N, g_N, w_N, z_N) \mathrm{d}\mu \leq \int_\Omega B(f_0, g_0, w_0, z_0) \mathrm{d}\mu \leq 0,$$

since $f_0 \equiv g_0$. Furthermore, we have $f_N = f$, $g_N = \mathcal{M} f$, $w_N = w$ and $z_N = w^{1/(1-p)} = w_N^{1/(1-p)}$, so applying (2.11) to the left-hand side, we get the claim. □

Now we will handle the implication in the reverse direction. Introduce the abstract function $\mathfrak{B} : D \to \mathbb{R}$ by the formula

$$\mathfrak{B}(x, y, u, v) = \sup \left\{ \int_\Omega V\left(f, y \vee \mathcal{M} f, w\right) \mathrm{d}\mu \right\}. \tag{2.13}$$

Here the supremum is taken over all probability spaces Ω with a tree $\mathcal{T}$, all $\mathcal{T}$-simple functions $f : \Omega \to [0, \infty)$ satisfying $\int_\Omega f \mathrm{d}\mu = x$, all $\mathcal{T}$-simple A_p weights w on Ω satisfying $[w]_{A_p} \leq c$, $\int_\Omega w \mathrm{d}\mu = u$ and $\int_\Omega w^{1/(1-p)} \mathrm{d}\mu = v$.

Theorem 2.2.2. *If* (2.9) *holds, then $\mathfrak{B}$, given by* (2.13), *satisfies the properties 1°, 2° and 3°. Actually, it is the least function defined on D enjoying these conditions.*

Proof. The initial condition follows directly from (2.9): indeed, for any Ω, $\mathcal{T}$, f and w as in the definition of $\mathfrak{B}(x, x, u, v)$ we have

$$\int_\Omega V\big(f, x \vee \mathcal{M} f, w\big) \mathrm{d}\mu = \int_\Omega V\big(f, \mathcal{M} f, w\big) \mathrm{d}\mu \leq 0,$$

and the inequality remains valid if we take the supremum. The majorization is also very simple: pick arbitrary Ω, $\mathcal{T}$ and consider the constant function $f \equiv x$ and the constant weight $w \equiv u$. Then $[w]_{A_p} = 1 \leq c$ and $\int_\Omega w^{1/(1-p)} \mathrm{d}\mu = u^{1/(1-p)}$, so by the very definition of $\mathfrak{B}$,

$$\mathfrak{B}(x, y, u, u^{1/(1-p)}) \geq \int_\Omega V\bigg(f, x \vee \mathcal{M} f, w\bigg) \mathrm{d}\mu = V(x, y, u).$$

It remains to prove the concavity-type condition 3°. Fix an auxiliary number $\varepsilon > 0$ and pick parameters λ_j and points (x, y, u, v), (x_j, y_j, u_j, v_j) as in the statement of 3°. By the definition of $\mathfrak{B}$, there are probability spaces (Ω_j, μ_j) with a tree $\mathcal{T}_j$ each, as well as appropriate functions f_j and w_j on Ω_j such that

$$\mathcal{B}(x_j, y_j, u_j, v_j) \leq \int_{\Omega_j} V\bigg(f_j, \max\left\{\mathcal{M}_{\Omega_j} f_j, y_j\right\}, w_j\bigg) \mathrm{d}\mu_j + \varepsilon. \tag{2.14}$$

With no loss of generality, we may assume that Ω_j are pairwise disjoint. We splice them into one space $\Omega = \bigcup_{j=1}^m \Omega_j$ with the probability measure μ given by $\mu(A) = \sum_{j=1}^m \lambda_j \mu_j(A \cap \Omega_j)$ and the tree structure $\mathcal{T}$ such that $\mathcal{T}^0 = \{\Omega\}$ and $\mathcal{T}^n = \bigcup_{j=1}^m \mathcal{T}_j^{n-1}$ for $n \geq 1$. Next, we "splice" the functions and weights as follows: $f = \sum_{j=1}^m f_j \chi_{\Omega_j}$ and $w = \sum_{j=1}^m w_j \chi_{\Omega_j}$. Let us check that f and w satisfy the requirements in the definition of $\mathcal{B}(x, y, u, v)$. First, we have

$$\int_\Omega f \mathrm{d}\mu = \sum_{j=1}^m \int_{\Omega_j} f_j \mathrm{d}\mu = \sum_{j=1}^m \lambda_j \int_{\Omega_j} f_j \mathrm{d}\mu_j = \sum_{j=1}^m \lambda_j x_j = x$$

and similarly, $\int_\Omega w \mathrm{d}\mu = u$, $\int_\Omega w^{1/(1-p)} \mathrm{d}\mu = v$, so the averaging conditions are satisfied. Now we will verify that $[w]_{A_p} \leq c$. By the calculations we have just carried out, we see that $\langle w \rangle_{\Omega,\mu} \langle w^{1/(1-p)} \rangle_{\Omega,\mu} = uv^{p-1} \leq c$, where the latter bound follows from the inclusion $(x, y, u, v) \in D$. Next, if $Q \in \mathcal{T}$ is different than Ω, then Q belongs to $\mathcal{T}_j$ for some j; since $[w_j]_{A_p} \leq c$,

$$\langle w \rangle_{Q,\mu} \left\langle w^{1/(1-p)} \right\rangle_{Q,\mu} = \langle w_j \rangle_{Q,\mu} \left\langle w_j^{1/(1-p)} \right\rangle_{Q,\mu} \leq c.$$

This establishes the desired Muckenhoupt condition and hence, by the very definition of $\mathfrak{B}$,

$$\mathfrak{B}(x, y, u, v) \geq \int_\Omega V\Big(f, \max\{\mathcal{M}f, y\}, w\Big) \mathrm{d}\mu.$$

Now, since $x \leq y$, we have $\max\{\mathcal{M}f, y\} = \max\{\mathcal{M}_{\Omega_j} f_j, y\}$ on Ω_j and hence

$$\begin{aligned} \mathfrak{B}(x, y, u, v) &\geq \sum_{j=1}^m \lambda_j \int_{\Omega_j} V\Big(f_j, \max\{\mathcal{M}_{\Omega_j} f_j, y\}, w_j\Big) \mathrm{d}\mu_j \\ &\geq \sum_{j=1}^m \lambda_j \mathfrak{B}(x_j, y_j, u_j, v_j) - \varepsilon, \end{aligned}$$

where in the last passage we have exploited (2.14). Since ε was arbitrary, the concavity condition follows. The minimality of $\mathfrak{B}$ is handled as in the unweighted context. □

2.3 An alternative approach in the unweighted context

There is a different version of the method, also based on the construction of a certain special function, which can be used to study maximal estimates. We will not use this version in the thesis, but we have decided to provide some brief presentation for the sake of completeness. We will handle the unweighted case only (but one can develop the meaningful extension to the weighted context; see e.g. [43]). The starting point is the following simple observation, sometimes referred to as the linearization of the maximal operator. Namely, if f is $\sigma(\mathcal{T}^N)$-measurable, then, since $\mathcal{M}f = \max_{n \leq N} f_n$, for each $\omega \in \Omega$ there is an integer $n(\omega) \leq N$ such that $\mathcal{M}f(\omega) = f_{n(\omega)}(\omega)$ ('the maximal function is attained at some average'). Of course, such an n may not be unique; in such a case, we take $n(\omega)$ to be the smallest possible. For any $Q \in \mathcal{T}$ let

$$E(Q) = \{\omega \in \Omega : Q_{n(\omega)}(\omega) = Q\},$$

where $Q_n(\omega)$ denotes the unique element of $\mathcal{T}^n$ which contains ω. By the very definition, we have $E(Q) \subseteq Q$, the sets $(E(Q))_{Q \in \mathcal{T}}$ are pairwise disjoint and their union is the full Ω.

Furthermore, since f is $\sigma(\mathcal{T}^N)$-measurable, we have $E(Q) = \emptyset$ for all $Q \in \mathcal{T}^n$ with $n > N$. This allows us to write the maximal operator in the linear form

$$\mathcal{M}f = \sum_{Q \in \mathcal{T}} \langle f \rangle_{Q,\mu} \chi_{E(Q)}.$$

Define the functional sequence $(g_n)_{n \geq 0}$ with the formula

$$g_n = \sum_{R \in \mathcal{T}^n} \frac{\chi_R}{\mu(R)} \sum_{Q \subseteq R,\ Q \in \mathcal{T}} \mu(E(Q)). \tag{2.15}$$

Observe that g_n is $\sigma(\mathcal{T}^n)$-measurable and, for $R \in \mathcal{T}^n$, we have the identity

$$\int_R g_n \mathrm{d}\mu = \sum_{Q \subseteq R,\ Q \in \mathcal{T}} \mu(E(Q)).$$

The sequence $(g_n)_{n \geq 0}$ enjoys a certain decreasing integral property, namely, for each $R \in \mathcal{T}^n$ the sequence $\left(\int_R g_m \mathrm{d}\mu\right)_{m \geq n}$ is nonincreasing. To see this, fix $n \geq 0$, an element $R \in \mathcal{T}^n$, and denote the children of R in $\mathcal{T}^{n+1}$ by $R_1, R_2, ..., R_m$. Then we have

$$\int_R g_{n+1} \mathrm{d}\mu = \sum_{k=1}^m \sum_{Q \subseteq R_k,\ Q \in \mathcal{T}} \mu(E(Q)) = \sum_{Q \subsetneq R,\ Q \in \mathcal{T}} \mu(E(Q)) = \int_R g_n \mathrm{d}\mu - \mu(E(R)) \leq \int_R g_n \mathrm{d}\mu.$$

Now, suppose that we are given nondecreasing functions $\Phi, \Psi : [0, \infty) \to [0, \infty)$, and assume that we are interested in showing the estimate

$$\int_\Omega \Psi(\mathcal{M}f) \mathrm{d}\mu \leq \int_\Omega \Phi(f) \mathrm{d}\mu \tag{2.16}$$

for any $\mathcal{T}$-simple, nonnegative function f. We look for a function $B : [0, \infty) \times [0, 1] \to \mathbb{R}$ that satisfies the following properties.

1° For any $x \geq 0$ we have $B(x, 1) \leq 0$.

2° For any $x \geq 0$ we have $B(x, 0) \geq -\Phi(x)$.

3° For any $(x, y) \in [0, \infty) \times [0, 1]$ and any numbers $x_1, x_2 \geq 0$, $y_1, y_2 \in [0, 1]$ and $\lambda_1, \lambda_2 \in (0, 1)$ such that $\lambda_1 + \lambda_2 = 1$, $\lambda_1 x_1 + \lambda_2 x_2 = x$ and $\lambda_1 y_1 + \lambda_2 y_2 \leq y$ we have

$$B(x, y) \geq \lambda_1 B(x_1, y_1) + \lambda_2 B(x_2, y_2) + \Psi(x)(y - \lambda_1 y_1 - \lambda_2 y_2). \tag{2.17}$$

In particular, assuming $\lambda_1 y_1 + \lambda_2 y_2 = y$ in 3° we see that B must be concave on $[0, \infty) \times [0, 1]$. Furthermore, the inequality (2.17) implies that for any $m \geq 0$, if $(x, y) \in [0, \infty) \times [0, 1]$ and the numbers $x_k \in [0, \infty)$, $y_k \in [0, 1]$, $\lambda_k \in (0, 1)$, $k = 1, 2, \ldots, m$ satisfy $\sum_{k=1}^m \lambda_k = 1$, $\sum_{k=1}^m \lambda_k x_k = x$ and $\sum_{k=1}^m \lambda_k y_k \leq y$, then we have

$$B(x, y) \geq \sum_{k=1}^m \lambda_k B(x_k, y_k) + \Psi(x) \left(y - \sum_{k=1}^m \lambda_k y_k \right). \tag{2.18}$$

To see this, first apply (2.17) to the weights $\lambda_1, \lambda_2' = \sum_{k=2}^m \lambda_k$ and the points

$$x_1, \quad x_2' = \frac{1}{\lambda_2'}\sum_{k=2}^m \lambda_k x_k, \quad y_1, \quad y_2' = \frac{1}{\lambda_2'}\sum_{k=2}^m \lambda_k y_k.$$

As the result, one gets

$$B(x,y) \geq \lambda_1 B(x_1,y_1) + \lambda_2' B(x_2',y_2') + \Psi(x)\left(y - \sum_{k=1}^m \lambda_k y_k\right)$$

and it suffices to combine this with the estimate

$$B(x_2',y_2') \geq \sum_{k=2}^m \frac{\lambda_k}{\lambda_2'} B(x_k,y_k),$$

which follows from the concavity of B.

Let us study the relation between the validity of (2.16) and the existence of a function B possessing the above three properties. Again, we have the full equivalence, and we treat each implication separately.

Theorem 2.3.1. *If there is a function B which satisfies $1°$, $2°$ and $3°$, then* (2.16) *holds true for any tree $\mathcal{T}$ on any probability space (Ω,μ) and any $\mathcal{T}$-simple, nonnegative function f.*

Proof. Assume that f is $\sigma(\mathcal{T}^N)$-measurable. As before, let $f_n = \sum_{Q\in\mathcal{T}^n}\langle f\rangle_{Q,\mu}\chi_Q$ and let the sequence $(g_n)_{n\geq 0}$ be given by (2.15). Next, fix $n \geq 0$, an element $Q \in \mathcal{T}^n$ and denote its children in $\mathcal{T}^{n+1}$ by $Q_1, Q_2, \ldots, Q_m$. Use (2.18) with

$$\lambda_k = \frac{\mu(Q_k)}{\mu(Q)}, \quad x = \langle f\rangle_{Q,\mu}, \quad y = \langle g_n\rangle_{Q,\mu}, \quad x_k = \langle f\rangle_{Q_k,\mu}, \quad y_k = \langle g_{n+1}\rangle_{Q_k,\mu}$$

for $k = 1, 2, \ldots, m$. We obtain

$$\begin{aligned}
\int_Q B(f_n,g_n)\mathrm{d}\mu &= \mu(Q)B(x,y) \geq \mu(Q)\left(\sum_{k=1}^m \lambda_k B(x_k,y_k) + \Psi(x)\left(y - \sum_{k=1}^m \lambda_k y_k\right)\right)\\
&= \sum_{k=1}^m \mu(Q_k)B(x_k,y_k) + \Psi(x)\left(\mu(Q)y - \sum_{k=1}^m \mu(Q_k)y_k\right)\\
&= \sum_{k=1}^m \int_{Q_k} B(f_{n+1},g_{n+1})\mathrm{d}\mu + \Psi(x)\left(\int_Q g_n\mathrm{d}\mu - \sum_{k=1}^m \int_{Q_k} g_{n+1}\mathrm{d}\mu\right)\\
&= \int_Q B(f_{n+1},g_{n+1})\mathrm{d}\mu + \Psi(x)\int_Q (g_n - g_{n+1})\,\mathrm{d}\mu\\
&= \int_Q B(f_{n+1},g_{n+1})\mathrm{d}\mu + \Psi(x)\mu(E(Q))\\
&= \int_Q B(f_{n+1},g_{n+1})\mathrm{d}\mu + \int_{E(Q)} \Psi(f_n)\mathrm{d}\mu,
\end{aligned}$$

or

$$\int_Q B(f_n,g_n)\mathrm{d}\mu \geq \int_Q B(f_{n+1},g_{n+1})\mathrm{d}\mu + \int_{E(Q)} \Psi(\mathcal{M}f)\mathrm{d}\mu,$$

by the very definition of $E(Q)$. Summing over all atoms $Q \in \mathcal{T}^n$, we get

$$\int_\Omega B(f_n, g_n)\mathrm{d}\mu - \int_\Omega B(f_{n+1}, g_{n+1})\mathrm{d}\mu \geq \sum_{Q \in \mathcal{T}^n} \int_{E(Q)} \Psi(\mathcal{M}f)\mathrm{d}\mu.$$

Hence, by induction, we see that

$$\int_\Omega B(f_0, g_0)\mathrm{d}\mu - \int_\Omega B(f_{N+1}, g_{N+1})\mathrm{d}\mu \geq \int_\Omega \Psi(\mathcal{M}f)\mathrm{d}\mu,$$

because Ω is a disjoint union of $(E(Q))_{Q\in\mathcal{T}}$. However, we have $g_0 = 1$, $g_{N+1} = 0$ and $f_{N+1} = f$. Therefore the application of 1° and 2° gives us

$$\int_\Omega \Phi(f)\mathrm{d}\mu \geq \int_\Omega \Psi(\mathcal{M}f)\mathrm{d}\mu,$$

which is the desired estimate. □

Now we will reverse the above implication. Consider the abstract Bellman function $\mathfrak{B} : [0, \infty) \times [0, 1] \to \mathbb{R}$, associated with the problem (2.16), which is defined by

$$\mathfrak{B}(x, y) = \sup \left(\int_A \Psi(\mathcal{M}f)\mathrm{d}\mu - \int_\Omega \Phi(f)\mathrm{d}\mu \right). \tag{2.19}$$

Here the supremum is taken over the class of all nonnegative, $\mathcal{T}$-simple functions f such that $\int_\Omega f\mathrm{d}\mu = x$ and all sets $A \in \sigma(\mathcal{T})$ with $\mu(A) = y$. The tree $\mathcal{T}$ and the probability space (Ω, μ) are also allowed to vary in the above supremum.

Lemma 2.3.2. *If* (2.16) *holds true for any tree $\mathcal{T}$ on any probability space (Ω, μ) and any $\mathcal{T}$-simple, nonnegative function f, then $\mathfrak{B}$ given by* (2.19) *satisfies* 1°, 2° *and* 3°. *It is the least function on* $[0, \infty) \times [0, 1]$ *enjoying these conditions.*

Proof. The initial condition 1° is a direct consequence of (2.16), while 2° follows by considering the constant function $f \equiv x$. To establish the third property, fix the parameters as in the statement and let A_1, A_2, f^1, f^2 be sets and functions as in the definition of $\mathfrak{B}(x_1, y_1)$ and $\mathfrak{B}(x_2, y_2)$, respectively. Assume that these functions are given on disjoint probability spaces (Ω_1, μ_1), (Ω_2, μ_2) and perform the splicing procedure as in the proof of Theorem 2.1.2. Consider any set $A \in \sigma(\mathcal{T})$ with $\mu(A) = y$ and $A_1 \cup A_2 \subseteq A$. We have

$$\begin{aligned}
\mathfrak{B}(x, y) &\geq \int_A \Psi(\mathcal{M}f)\mathrm{d}\mu - \int_\Omega \Phi(f)\mathrm{d}\mu \\
&= \int_{A\setminus(A_1\cup A_2)} \Psi(\mathcal{M}f)\mathrm{d}\mu + \sum_{k=1}^{2} \left(\int_{A_k} \Psi(\mathcal{M}f)\mathrm{d}\mu - \int_{\Omega_k} \Phi(f)\mathrm{d}\mu \right) \\
&\geq \mu(A \setminus (A_1 \cup A_2))\Psi(x) + \sum_{k=1}^{2} \lambda_k \left(\int_{A_k} \Psi(\mathcal{M}_k f^k)\mathrm{d}\mu_k - \int_{\Omega_k} \Phi(f^k)\mathrm{d}\mu_k \right).
\end{aligned}$$

Taking the supremum over all A_1, A_2, f^1, f^2 as above, we get the desired claim. The minimality of $\mathfrak{B}$ is proved as previously. □

Remark 2.3.3. There is a natural question about the relation between the Bellman functions appearing in the two variants presented above. There seems to be no evident algebraic connection. Although some indication is hidden in the abstract formulas (2.5) and (2.19), the passage from one context to the other is nontrivial. The key difference between the two approaches is the role of the second variable y: in the former case it corresponds to the "memory" of the maximal function, while in the latter it measures the size of the set.

2.4 Other extensions

The above considerations cover just a part of the area of maximal estimates which can be studied with the use of the Bellman function method. For example, we have not included here the alternative approach for the weighted estimates; we have also not discussed the setting of Fefferman-Stein inequalities (see (1.3), (1.4)), which involves the use of both w and $\mathcal{M}w$, instead of the Muckenhoupt's condition. We believe that even the brief presentation of all the possible modifications would be quite extensive and rather tedious, as there is a quite big common part of the approach in each setting (we have already experienced a lot of repetitions above). Instead, we have decided to content ourselves with the material presented above, hoping that it provides the necessary insight into the technique, and to postpone the description of the relevant changes to the appropriate chapters.

2.5 Melas' Lemma

In the final section of this chapter, we discuss an issue which is not related to the Bellman function method, but is very useful in the construction of the extremizers (i.e., the functions for which both sides of an estimate under investigation become equal, or almost equal) on arbitrary probability spaces. Namely, there is a universal method of finding measurable subsets with a given, prescribed measure; here by the universality we mean that the structure of the tree does not play a role. Specifically, we have the following fact proved in [33].

Lemma 2.5.1 (Melas)**.** *For every $Q \in \mathcal{T}$ and every $\beta \in (0,1)$ there is a subfamily $F(Q) \subset \mathcal{T}$, consisting of pairwise disjoint subsets of Q, such that*

$$\mu\left(\bigcup_{R\in F(Q)} R\right) = \sum_{R\in F(Q)} \mu(R) = \beta\mu(Q).$$

Proof. This is rather straightforward and rests on an argument which can be roughly described as "at each step, take as much as you can". There is an integer m such that $Q \in \mathcal{T}^m$. We construct a nondecresing sequence $(A_n)_{n\geq m}$ by the following procedure. We start with $A_m = \emptyset$. Having constructed A_n, we define A_{n+1} to be the set which contains A_n, is the union of some elements of $\mathcal{T}^{n+1}$ and has the largest possible measure, satisfying $\mu(A_{n+1}) \leq \beta\mu(Q)$. It is easy to see that the limit set $\bigcup_{n\geq m} A_n$ has measure $\beta\mu(Q)$ and can be split into family $F(Q)$ as in the statement. $\square$

The above lemma can be applied recursively to obtain the whole family of sets, satisfying the prescribed measure requirements.

Lemma 2.5.2. *For any decreasing sequence $(a_n)_{n\geq 0}$ of real numbers from the interval $(0,1]$ there exists a sequence $(\mathcal{Q}^n)_{n\geq 0}$ of subfamilies of $\mathcal{T}$ and a decreasing sequence $(E_n)_{n\geq 0}$ of subsets of Ω with the following fractal structure.*

(i) $\mu(E_0) = a_0$.

(ii) *For all $n \geq 0$ we have $E_n = \bigcup \mathcal{Q}^n$.*

(iii) *For all $n \geq 0$ we have $E_{n+1} \subseteq E_n$.*

(iv) *If $Q \in \mathcal{Q}^m$, then for all $n \geq m$ we have $\mu(Q \cap E_n) = \mu(Q) \cdot \frac{a_n}{a_m}$.*

The elements of $\mathcal{Q}^n$ are called the atoms *of E_n.*

Proof. We use induction. The existence of the family $\mathcal{Q}^0$ follows directly from the previous lemma. Suppose that $n \geq 0$ and we have successfully constructed the family $\mathcal{Q}^n$. For each atom $Q \in \mathcal{Q}^n$ we use the previous lemma with $\beta = a_{n+1}/a_n$, obtaining a family $F(Q)$ of subsets of Q. Then we set $\mathcal{Q}^{n+1} = \bigcup_{Q\in\mathcal{Q}^n} F(Q)$ and $E_{n+1} = \bigcup \mathcal{Q}^{n+1}$. This completes the description of the induction step. All the properties follow easily from the construction. □

Chapter 3

A weighted inequality for $0 < p < 1$

3.1 Motivation and the statement of results

Suppose that (Ω, μ) is a probability space endowed with a tree structure $\mathcal{T}$ and $\mathcal{M}$ is the associated dyadic-like maximal operator. The motivation for the results discussed in this chapter comes from the weighted L^p estimates for $\mathcal{M}$: as we have mentioned in the introductory section, for each $1 < p < \infty$ and an arbitrary weight w we have the Fefferman-Stein estimate

$$\|\mathcal{M}f\|_{L^p(w)} \leq \frac{p}{p-1}\|f\|_{L^p(\mathcal{M}w)}. \tag{3.1}$$

Furthermore, if we additionally assume that $w \in A_p$, then

$$\|\mathcal{M}f\|_{L^p(w)} \leq C_{p,[w]_{A_p}}\|f\|_{L^p(w)} \tag{3.2}$$

(see (1.7)). This gives rise to the natural question about the case $0 < p < 1$. In the unweighted setting we have the Kolmogorov bound (cf. [37])

$$\|\mathcal{M}f\|_{L^p} \leq (1-p)^{-1/p}\|f\|_{L^1}. \tag{3.3}$$

Two important comments are in order. The first comment is that (3.3) involves two different function spaces on both sides: the more natural estimate

$$\|\mathcal{M}f\|_{L^p} \leq c_p\|f\|_{L^p}$$

simply does not hold with any finite constant c_p (if $0 < p < 1$). Second, Muckenhoupt condition becomes more and more restrictive as we decrease p, which implies that (3.2) becomes more "challenging for a weight". This in particular suggests that the A_1 condition might not be sufficient for the validity of the weighted version of Kolmogorov' estimate. Both these observations indicate that the appropriate extension of (3.3) to the weighted context might be of complicated form. Actually, it seems reasonable to expect that *there is no analogue* of (3.2) for $0 < p < 1$, and hence in our considerations below we will focus on the search for the appropriate version of (3.1).

The approach to the above problem, presented in this chapter, is based on the paper [64]. Our first step is to try to guess the appropriate "shape" of the weighted version of Kolmogorov's inequality. Motivated by the form in the unweighted setting, a plausible attempt is to separate the functions and assume that the right hand side depends on f and w through their L^1 norms only. Combining this conjecture with homogeneity restrictions, we arrive at the estimate

$$\|\mathcal{M}f\|^p_{L^p(w)} \le c_p \|f\|^p_{L^1} \|w\|_{L^1},$$

which, unfortunately, does not hold. Indeed, pick an arbitrary integrable function f for which $\mathcal{M}f \notin L^1$, and set $w = \mathcal{M}f^{1-p}$. Then $\|\mathcal{M}f\|_{L^p(w)} = \|\mathcal{M}f\|_{L^1} = \infty$, while $\|f\|_{L^1}$ and $\|w\|_{L^1}$ are finite.

However, a little experimentation reveals that if we introduce the additional error term, the inequality becomes valid. We define

$$E_{\mathcal{T}}(f, w) := \left(\left\| f \cdot (\mathcal{M}w)^{\frac{1}{p}} \right\|_{L^1} - \|f\|_{L^1} \|w\|^{\frac{1}{p}}_{L^1} \right) \left(\|f\|_{L^1} \|w\|^{\frac{1}{p}}_{L^1} \right)^{p-1}.$$

In the case when $\|f\|_{L^1}\|w\|_{L^1} = 0$, the error term is understood to be zero.

Our main result can be stated as follows:

Theorem 3.1.1. *Assume $f, w \in L^1(\Omega)$ are nonnegative. For $p \in (0,1)$ we have*

$$\|\mathcal{M}f\|^p_{L^p(w)} \le \frac{1}{1-p} \|f\|^p_{L^1} \|w\|_{L^1} + \frac{p^2}{1-p} E_{\mathcal{T}}(f, w). \tag{3.4}$$

Both constants $1/(1-p)$ and $p^2/(1-p)$ are the best possible.

The proof of the above statement will be based on a version of the Bellman function method, presented in a separate section below.

3.2 A modification of the Bellman function method

The estimate (3.4) involves four objects: f, $\mathcal{M}f$, w and $\mathcal{M}w$. Thus it is natural to expect that the special functions to be constructed will depend on four variables: x, y, u and v. Furthermore, since $f \le \mathcal{M}f$ and $w \le \mathcal{M}w$, we consider the domain $D = \{(x, y, u, v) \in \mathbb{R}^4 : 0 \le x \le y, 0 \le u \le v\}$. Motivated by the discussion from the previous chapter, we proceed as follows. Suppose that $V : D \to \mathbb{R}$ is a given function and assume that we are interested in the estimate

$$\int_\Omega V(f, \mathcal{M}f, w, \mathcal{M}w) \mathrm{d}\mu \le 0, \tag{3.5}$$

for all $\mathcal{T}$-simple and nonnegative f, w. Assume further that $B : D \to \mathbb{R}$ enjoys the following set of requirements:

1° (Initial condition). For any x, $u \ge 0$ we have $B(x, x, u, u) \le 0$.

2° (Majorization). We have $B \ge V$ on D.

3° (Concavity). For any $(x, y, u, v) \in D$, any numbers $x_1, x_2 \ge 0$, u_1, $u_2 \ge 0$ and $\lambda_1, \lambda_2 \in (0, 1)$ such that $\lambda_1 + \lambda_2 = 1$, $\lambda_1 x_1 + \lambda_2 x_2 = x$ and $\lambda_1 u_1 + \lambda_2 u_2 = u$, we have

$$B(x, y, u, v) \ge \lambda_1 B(x_1, y \vee x_1, u_1, v \vee u_1) + \lambda_2 B(x_2, y \vee x_2, u_2, v \vee u_2). \tag{3.6}$$

Remark 3.2.1. By a straightforward induction argument, the concavity condition is equivalent to the following statement. For any $m \geq 2$, any point (x, y, u, v) and any numbers $x_1, x_2, \ldots, x_m, u_1, u_2, \ldots, u_m \geq 0$, $\lambda_1, \lambda_2, \ldots, \lambda_m \in (0,1)$ satisfying

$$\sum_{j=1}^m \lambda_j = 1, \qquad \sum_{j=1}^m \lambda_j x_j = x, \qquad \sum_{j=1}^m u_j = u,$$

we have

$$B(x, y, u, v) \geq \sum_{j=1}^m \lambda_j B(x_j, y \vee x_j, u_j, v \vee u_j). \tag{3.7}$$

We have the following statement.

Theorem 3.2.2. *The inequality* (3.5) *holds true if and only if there is a function B satisfying the properties 1°, 2° and 3°.*

Proof. The argument is similar to that in the previous chapter, so we will be brief. First we show that the existence of B yields (3.5). For any $n \geq 0$, we let $f_n = \sum_{Q \in \mathcal{T}^n} \langle f \rangle_Q \chi_Q$, $g_n = \max_{0 \leq k \leq n} f_k$, $u_n = \sum_{Q \in \mathcal{T}^n} \langle w \rangle_Q \chi_Q$ and $v_n = \max_{0 \leq k \leq n} u_k$. The concavity condition (the enhanced version (3.7)) implies that the sequence

$$\left(\int_\Omega B(f_n, g_n, u_n, v_n) \mathrm{d}\mu \right)_{n \geq 0}$$

is nonincreasing. Combining this with the majorization and the initial condtion gives (3.5): by the $\mathcal{T}$-simplicity of f and w, for sufficiently large n we have

$$\begin{aligned} \int_\Omega V(f, \mathcal{M}f, w, \mathcal{M}w) \mathrm{d}\mu &= \int_\Omega V(f_n, g_n, u_n, v_n) \mathrm{d}\mu \\ &\leq \int_\Omega B(f_n, g_n, u_n, v_n) \mathrm{d}\mu \\ &\leq \int_\Omega B(f_0, g_0, u_0, v_0) \mathrm{d}\mu = \int_\Omega B(f_0, f_0, u_0, u_0) \mathrm{d}\mu \leq 0. \end{aligned} \tag{3.8}$$

To prove the reverse implication, we consider the function $\mathfrak{B} : D \to \mathbb{R}$, given by

$$\mathfrak{B}(x, y, u, v) = \sup \left\{ \int_\Omega V\Big(f, y \vee \mathcal{M}f, w, v \vee \mathcal{M}w\Big) \mathrm{d}\mu \right\},$$

where the supremum is taken over all $\mathcal{T}$-simple nonnegative functions f and w on Ω satisfying $\int_\Omega f \mathrm{d}\mu = x$ and $\int_\Omega w \mathrm{d}\mu = u$. Arguing as previously, one checks that $\mathfrak{B}$ enjoys 1°, 2° and 3°; actually, it is not difficult to show that it is the least function which satisfies these requirements. □

Remark 3.2.3. It is straightforward to modify the approach so that it enables the study of the estimates of the form

$$\int_\Omega V(f, \mathcal{M}f, w, \mathcal{M}w) \mathrm{d}\mu \leq G\left(\|f\|_1, \|w\|_1\right),$$

for some given V and G. Namely, the validity of such an inequality is equivalent to the existence of a function B which satisfies 2°, 3° and the following version of 1°:

1°′ For any x, $u \geq 0$ we have $B(x, x, u, u) \leq G(x, u)$.

To see one implication, simply modify the last passage in (3.8); to get the reverse, one considers the function $\mathfrak{B}$ given by the same formula.

3.3 A special function and its properties

We are ready for the study of the estimate (3.4). Let $B : D \to \mathbb{R}$ be given by

$$B(x,y,u,v) = y^p u + \frac{p}{1-p} y^{p-1} x v - \frac{p^2}{1-p} x v^{\frac{1}{p}}.$$

We will apply the Bellman function method with

$$V(x,y,u,v) = y^p u - \frac{p^2}{1-p} x v^{\frac{1}{p}}$$

and, in the terminology developed in Remark 3.2.3, with $G(x,u) = B(x,x,u,u)$. Obviously, with these choices of V and G, the initial and the majorization conditions hold true. Therefore, all we need is the concavity requirement. We will prove the following pointwise bound which, as we will show later, directly yields 3°.

Lemma 3.3.1. *If $(x,y,u,v) \in D$, $y \geq 1$, $v \geq 1$, $s \geq -x$ and $t \geq -u$ we have*

$$B(x,y,u,v) + s\left(\frac{p}{1-p} y^{p-1} v - \frac{p^2}{1-p} v^{\frac{1}{p}}\right) + t y^p \geq B(x+s, y \vee (x+s), u+t, v \vee (u+t)).$$

Proof. Fix variables $y, v \geq 1$ and consider the function $B_{y,v} : [0,\infty)^2 \to \mathbb{R}$ defined by the formula $B_{y,v}(x,u) = B(x, y \vee x, u, v \vee u)$. In explicit form, $B_{y,v}$ is given by

$$B_{y,v}(x,u) = \begin{cases} y^p u + \frac{p}{1-p} y^{p-1} x v - \frac{p^2}{1-p} x v^{\frac{1}{p}} & \text{for } x \leq y, u \leq v, \\ x^p \left(u + \frac{p}{1-p} v\right) - \frac{p^2}{1-p} x v^{\frac{1}{p}} & \text{for } x > y, u \leq v, \\ \left(y^p + \frac{p}{1-p} y^{p-1} x\right) u - \frac{p^2}{1-p} x u^{\frac{1}{p}} & \text{for } x \leq y, u > v, \\ \frac{1}{1-p} x^p u - \frac{p^2}{1-p} x u^{\frac{1}{p}} & \text{for } x > y, u > v. \end{cases}$$

We will first prove the claim for $x = y$ and $u = v$. We shall consider four cases, depending on signs of s and t. All of them are easy and require only the inequalities of the type $(y+s)^p \leq y^p + p y^{p-1} s$ or $-(v+t)^{\frac{1}{p}} \leq -\left(v^{\frac{1}{p}} + \frac{1}{p} v^{\frac{1}{p}-1} t\right)$, which follow directly from the mean-value theorem.

1. Assume $s \leq 0$ and $t \leq 0$. We have

$$\begin{aligned} B_{y,v}(y+s, v+t) &= y^p (v+t) \\ &\quad + \frac{p}{1-p} y^{p-1} (y+s) v - \frac{p^2}{1-p} (y+s) v^{\frac{1}{p}} \\ &= -\frac{p^2}{1-p} y v^{\frac{1}{p}} + y^p t + s\left(\frac{p}{1-p} y^{p-1} v - \frac{p^2}{1-p} v^{\frac{1}{p}}\right) + \frac{1}{1-p} y^p v \\ &= B(y,y,v,v) + s\left(\frac{p}{1-p} y^{p-1} v - \frac{p^2}{1-p} v^{\frac{1}{p}}\right) + t y^p. \end{aligned}$$

In this case we have equality, because the derivative of $U_{y,v}$ is constant on the rectangle $(0,y) \times (0,v)$.

2. Assume $s > 0$ and $t \le 0$. We have

$$\begin{aligned}
B_{y,v}(y+s,v+t) &= (y+s)^p\left(v+t+\frac{p}{1-p}v\right) - \frac{p^2}{1-p}(y+s)v^{\frac{1}{p}} \\
&\le (y^p + py^{p-1}s)\left(t + \frac{1}{1-p}v\right) - \frac{p^2}{1-p}yv^{\frac{1}{p}} - \frac{p^2}{1-p}sv^{\frac{1}{p}} \\
&= ty^p + s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + \frac{1}{1-p}y^pv - \frac{p^2}{1-p}yv^{\frac{1}{p}} + psty^{p-1} \\
&\le B(y,y,v,v) + s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + ty^p.
\end{aligned}$$

Here in the last inequality we used the estimate $psty^{p-1} \le 0$.

3. The next case we consider is $s \le 0$ and $t > 0$. We have

$$\begin{aligned}
B_{y,v}(y+s,v+t) &= \left(y^p + \frac{p}{1-p}y^{p-1}(y+s)\right)(v+t) - \frac{p^2}{1-p}(y+s)(v+t)^{\frac{1}{p}} \\
&\le \left(\frac{1}{1-p}y^p + \frac{p}{1-p}sy^{p-1}\right)(v+t) - \frac{p^2}{1-p}(y+s)\left(v^{\frac{1}{p}} + \frac{1}{p}tv^{\frac{1}{p}-1}\right) \\
&= \frac{1}{1-p}ty^p + \frac{p}{1-p}sy^{p-1}v - \frac{p}{1-p}tyv^{\frac{1}{p}-1} - \frac{p^2}{1-p}sv^{\frac{1}{p}} \\
&\quad + \frac{1}{1-p}y^pv + \frac{p}{1-p}sty^{p-1} - \frac{p^2}{1-p}yv^{\frac{1}{p}} - \frac{p}{1-p}stv^{\frac{1}{p}-1} \\
&= s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + B(y,y,v,v) \\
&\quad + t\left(\frac{1}{1-p}y^p - \frac{p}{1-p}yv^{\frac{1}{p}-1} + \frac{p}{1-p}sy^{p-1} - \frac{p}{1-p}sv^{\frac{1}{p}-1}\right) \\
&= s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + B(y,y,v,v) + ty^p \\
&\quad + \frac{p}{1-p}t\left(y^p - yv^{\frac{1}{p}-1} + sy^{p-1} - sv^{\frac{1}{p}-1}\right) \\
&= s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + B(y,y,v,v) + ty^p \\
&\quad + \frac{p}{1-p}t(y+s)\left(y^{p-1} - v^{\frac{1}{p}-1}\right).
\end{aligned}$$

But we have $y, v \ge 1$ and hence $y^{p-1} \le 1 \le v^{(1-p)/p}$, which means that the expression in the last bracket is nonpositive.

4. The final case to consider is $s > 0$ and $t > 0$. We compute that

$$\begin{aligned}
B_{y,v}(y+s,v+t) &= \frac{1}{1-p}(y+s)^p(v+t) - \frac{p^2}{1-p}(y+s)(v+t)^{\frac{1}{p}} \\
&\le \frac{1}{1-p}\left(y^p + py^{p-1}s\right)(v+t) - \frac{p^2}{1-p}(y+s)\left(v^{\frac{1}{p}} + \frac{1}{p}v^{\frac{1}{p}-1}t\right)
\end{aligned}$$

and from this point we proceed exactly as in the previous case.

The above considerations establish the assertion of Lemma 3.3.1 in the particular case $x = y$ and $u = v$. Now we will investigate the general case, in which $x \leq y$ and $u \leq v$ are arbitrary. We have

$$\begin{aligned}
B_{y,v}(x+s, u+t) &= B_{y,v}(y + (x-y) + s, v + (u-v) + t) \\
&\leq B(y,y,v,v) + (s+x-y)\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + (t+u-v)y^p \\
&= \frac{1}{1-p}y^p v - \frac{p^2}{1-p}yv^{\frac{1}{p}} + \frac{p}{1-p}xy^{p-1}v - \frac{p^2}{1-p}xv^{\frac{1}{p}} - \frac{p}{1-p}y^p v \\
&\quad + \frac{p^2}{1-p}yv^{\frac{1}{p}} + y^p u - y^p v + s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + ty^p \\
&= y^p u + \frac{p}{1-p}y^{p-1}xv - \frac{p^2}{1-p}xv^{\frac{1}{p}} + s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + ty^p \\
&= B(x,y,u,v) + s\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + ty^p.
\end{aligned}$$

This is precisely the desired claim. □

The above lemma immediately implies that the function B satisfies the concavity condition 3°: roughly speaking, the pointwise estimate we have just established asserts that B lies below an appropriate tangent plane. To be more precise, pick the parameters (x, y, u, v), x_1, x_2, u_1, u_2, λ_1 and λ_2 as in the statement. For $i = 1,\ 2$, we have

$$B(x,y,u,v) + (x_i - x)\left(\frac{p}{1-p}y^{p-1}v - \frac{p^2}{1-p}v^{\frac{1}{p}}\right) + (u_i - u)y^p \geq B(x_i, y \vee x_i, u_i, v \vee u_i).$$

Multiplying both sides by λ_i and summing over $i = 1,\ 2$, we get (3.6).

3.4 Proof of (3.4)

Assume that f and w are nonnegative and $\mathcal{T}$-simple functions. Let us consider first the case in which $\|f\|_1 = \|w\|_1 = 1$. The application of the Bellman function method yields

$$\int_\Omega (\mathcal{M}f)^p w \mathrm{d}\mu - \frac{p^2}{1-p}\int_\Omega f(\mathcal{M}w)^{\frac{1}{p}}\mathrm{d}\mu \leq p+1.$$

Now, for the general case (in which $(f_n)_{n\geq 0}$ and $(w_n)_{n\geq 0}$ not necessarily start from 1), use the above bound with f/f_0 and w/w_0. As the result, we obtain

$$\int_\Omega \left(\frac{\mathcal{M}f}{f_0}\right)^p \frac{w}{w_0}\mathrm{d}\mu \leq \frac{p^2}{1-p}\int_\Omega \frac{f}{f_0}\left(\frac{\mathcal{M}w}{w_0}\right)^{\frac{1}{p}}\mathrm{d}\mu + p + 1,$$

which is equivalent to

$$\int_\Omega (\mathcal{M}f)^p w \mathrm{d}\mu \leq \frac{p^2}{1-p}\int_\Omega f(\mathcal{M}w)^{\frac{1}{p}}\mathrm{d}\mu \cdot f_0^{p-1}w_0^{1-\frac{1}{p}} + \frac{1-p^2}{1-p}f_0^p w_0,$$

or

$$\int_\Omega (\mathcal{M}f)^p w \mathrm{d}\mu \leq \frac{1}{1-p}\|f\|_1^p\|w\|_1 + \frac{p^2}{1-p}E_{\mathcal{T}}(f,w). \tag{3.9}$$

This completes the proof of (3.4) for all $\mathcal{T}$-simple functions f, w.

The general case is obtained by approximation and standard limiting theorems. More precisely, for any n the functions f_n and w_n are $\mathcal{T}$-simple, so the application of (3.9) gives

$$\int_\Omega (\mathcal{M}f_n)^p w_n \mathrm{d}\mu \leq \frac{1}{1-p}\|f_n\|_1^p \|w_n\|_1 + \frac{p^2}{1-p} E_{\mathcal{T}}(f_n, w_n).$$

But $\|f_n\|_1 = \|f\|_1$ and $\|w_n\|_1 = \|w\|_1$. Furthermore, notice that $\mathcal{M}w_n \leq \mathcal{M}w$ and the function $z \mapsto z^{\frac{1}{p}}$ is increasing: consequently, we have

$$\int_\Omega f_n(\mathcal{M}w_n)^{\frac{1}{p}} \mathrm{d}\mu = \int_\Omega f(\mathcal{M}w_n)^{\frac{1}{p}} \mathrm{d}\mu \leq \int_\Omega f(\mathcal{M}w)^{\frac{1}{p}} \mathrm{d}\mu,$$

which implies $E_{\mathcal{T}}(f_n, w_n) \leq E_{\mathcal{T}}(f, w)$. To handle the left-hand side, we let $n \to \infty$: we have $\int_\Omega (\mathcal{M}f_n)^p w_n \mathrm{d}\mu = \int_\Omega (\mathcal{M}f_n)^p w \mathrm{d}\mu \to \int_\Omega (\mathcal{M}f)^p w \mathrm{d}\mu$ by the Lebesgue monotone convergence theorem. Putting all the above facts together, we get the estimate (3.4) for general f, w.

3.5 Sharpness

Now we turn our attention to the optimality of the constants. The inequality (3.9) can be written as

$$\int_\Omega (\mathcal{M}f_n)^p w_n \mathrm{d}\mu \leq c_1 f_0^p w_0 + c_2 \left(\int_\Omega f_n (\mathcal{M}w_n)^{\frac{1}{p}} \mathrm{d}\mu - f_0 w_0^{\frac{1}{p}} \right) \cdot f_0^{p-1} w_0^{1-\frac{1}{p}},$$

with $c_1 = 1/(1-p)$ and $c_2 = p^2/(1-p)$. There is a natural question whether any of these two numbers can be improved. We will show that it is not possible.

The first constant, $c_1 = 1/(1-p)$, is the best, it cannot be decreased even at a cost of a significant enlargement of c_2. Indeed, this follows from the fact that $1/(1-p)$ is already optimal in the unweighted case [37], and $E_{\mathcal{T}}(f, w) = 0$ whenever w is constant.

The constant $c_2 = p^2/(1-p)$ is also the best possible, but the proof of this fact is a little more complicated. Take $\eta > 0$ and $N \in \mathbb{N}$, and introduce the auxiliary parameter $\delta = (1+\eta)^{1/N} - 1$. We use Lemma 2.5.2 with the sequence $a_n = (1+\delta)^{-n}$, $n = 0, 1, 2, \ldots$, to obtain the appropriate decreasing family $(E_n)_{n \geq 0}$ of subsets of Ω. If Q is an atom of E_k, then for any $n \geq k$ we have

$$\mu(Q \cap E_n) = \mu(Q) \left(\frac{1}{1+\delta} \right)^{n-k}. \tag{3.10}$$

For $n > N$, define f, w by the formulas $f = (1+\delta)^n \mathbb{1}_{E_n}$ and $w = (1+\delta)^N \mathbb{1}_{E_N}$. Then we have

$$\|f\|_1 = (1+\delta)^n \mu(E_n) = 1,$$

by (3.10) applied with $Q = \Omega$ and $k = 0$. Similarly, $\|w\|_1 = 1$. Let us now handle the maximal operator. By the very definition of $\mathcal{M}$, if $\omega \in E_n$, then $\mathcal{M}f(\omega) \geq f(\omega) = (1+\delta)^n$. Furthermore, if $\omega \in E_k \setminus E_{k+1}$ for some $k = 0, 1, 2, \ldots, n-1$, then ω belongs to some atom Q of E_k; therefore,

$$\mathcal{M}f(\omega) \geq \frac{1}{\mu(Q)} \int_Q f \mathrm{d}\mu = \frac{(1+\delta)^n \mu(Q \cap E_n)}{\mu(Q)} = (1+\delta)^k,$$

again by (3.10). Consequently, we have proved that

$$\mathcal{M}f \geq (1+\delta)^n \mathbb{1}_{E_n} + \sum_{k=0}^{n-1} (1+\delta)^k \mathbb{1}_{E_k \setminus E_{k+1}}$$

μ-almost everywhere. Actually, it is easy to see that equality holds here, but we will not need this. A similar analysis to that above shows that the integral $\int_\Omega f(\mathcal{M}w)^{1/p} \mathrm{d}\mu$ is equal to $(1+\delta)^{N/p} = (1+\eta)^{1/p}$ (the argument is even simpler: f vanishes outside E_n, we have $E_n \subset E_N$ and w is supported and constant on E_N; thus the identity follows at once). To see how the L^p norm of $\mathcal{M}f$ behaves for large n, we compute that

$$\begin{aligned}
\lim_{n\to\infty} \int_\Omega (\mathcal{M}f)^p w \mathrm{d}\mu &\geq \lim_{n\to\infty} \left((1+\delta)^{(p-1)n}(1+\eta) + \sum_{k=N}^{n-1} (1+\delta)^{(p-1)k}(1+\eta) \cdot \frac{\delta}{1+\delta} \right) \\
&= \sum_{k=N}^{\infty} (1+\delta)^{(p-1)k}(1+\eta) \cdot \frac{\delta}{1+\delta} \\
&= \frac{(1+\delta)^{(p-1)N}}{1-(1+\delta)^{p-1}}(1+\eta) \cdot \frac{\delta}{1+\delta} = \frac{(1+\eta)^p \left((1+\eta)^{1/N} - 1\right)}{(1+\eta)^{1/N} - (1+\eta)^{p/N}}.
\end{aligned}$$

Therefore we obtain a lower bound on c_2:

$$\begin{aligned}
c_2 &\geq \frac{\frac{(1+\eta)^p((1+\eta)^{1/N}-1)}{(1+\eta)^{1/N}-(1+\eta)^{p/N}} - \frac{1}{1-p}}{(1+\eta)^{1/p} - 1} \\
&= \frac{(1-p)(1+\eta)^p((1+\eta)^{1/N} - 1) - (1+\eta)^{1/N} + (1+\eta)^{p/N}}{(1-p)\left((1+\eta)^{1/N} - (1+\eta)^{p/N}\right)((1+\eta)^{1/p} - 1)} \\
&= \frac{(1-p)(1+\eta)^p((1+\eta)^{1/N} - 1) - (1+\eta)^{p/N}\left((1+\eta)^{(1-p)/N} - 1\right)}{(1-p)(1+\eta)^{p/N}\left((1+\eta)^{(1-p)/N} - 1\right)((1+\eta)^{1/p} - 1)}.
\end{aligned}$$

When we put $\eta = 1/N$ and let $N \to \infty$, we will obtain that the above expression has a limit of $p^2/(1-p)$, which proves that c_2 cannot be smaller.

Remark 3.5.1. The above reasoning shows that $c_2 = p^2/(1-p)$ is optimal if we take $c_1 = 1/(1-p)$. It is not clear to us what happens to the optimal value of c_2 if we allow c_1 to increase.

Remark 3.5.2. There is a natural question how we have discovered the Bellman function B exploited above, and we will provide some brief and informal discussion in this direction. We start with the observation that the search for the function was, essentially, parallel to the search of the appropriate weighted version of Kolmogorov's inequality. The first step is to take a look at the unweighted bound

$$\|\mathcal{M}f\|_{L^p(\Omega)} \leq \frac{1}{(1-p)^{1/p}} \|f\|_{L^1(\Omega)},$$

for which the associated Bellman function is (cf. [45])

$$b(x,y) = \alpha_p y^{p-1}\left(y + \frac{px}{1-p}\right), \qquad 0 < x \leq y,$$

for some α_p depending only on p. Note that this function is homogeneous of order p and linear in x. Together with the condition 3°, this suggests to search for a special function B which is homogeneous of order p with respect to variables x, y; furthermore, it should be (jointly) linear with respect to x and u. Finally, for $u = v$, one might expect the function to be close to b. All these observations do not leave too much freedom to the construction: we are more or less immediately led to

$$B(x, y, u, v) = \alpha_p y^{p-1} \left(yu + \frac{pxv}{1-p} \right).$$

Unfortunately, this function does not satisfy the condition 3° and some modification is required. We have discovered the additional term $xv^{1/p}$ after a series of experiments; this term is actually responsible for the appearance (and the shape) of the error term $E_{\mathcal{T}}$. There might be different modifications of B which can yield other weighted variants of Kolmogorov's inequality.

Chapter 4

General Fefferman-Stein inequalities

4.1 Motivation and the statement of results

We stick to the context of a probability space (Ω, μ) endowed with a tree structure $\mathcal{T}$ and the associated maximal operator $\mathcal{M}$. The purpose of this chapter is to inspect closer the following interplay we have encountered above. Namely, we have the sharp L^p estimate for the dyadic-like maximal operator:

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

and its variant of Fefferman and Stein:

$$\|\mathcal{M}f\|_{L^p(w)} \leq \frac{p}{p-1}\|f\|_{L^p(\mathcal{M}w)}, \qquad 1 < p < \infty.$$

We see that both estimates involve the same optimal constant $p/(p-1)$. The same phenomenon occurs for the logarithmic inequalities: as proved by Osękowski [53], for any f on Ω and any weight w, we have the bound

$$\int_\Omega \mathcal{M}fw\mathrm{d}\mu \leq K\int_\Omega |f|\log|f|\mathcal{M}w\mathrm{d}\mu + L(K)\int_\Omega \mathcal{M}w\mathrm{d}\mu. \tag{4.1}$$

Here, for a given $K > 1$, the constant $L(K)$ is given by $L(K) = \frac{K^2}{(K-1)e}$; if $K \leq 1$, the estimate does not hold with any finite constant $L(K)$. It turns out that this choice of $L(K)$ is already the best possible in the unweighted setting (i.e., for $w \equiv 1$), cf. Gilat [18]. See also [55] or below for a related exponential result.

The above discussion gives rise to the following very natural "transference" question. Namely, suppose that Φ, Ψ are two functions on $[0, \infty)$ satisfying the estimate

$$\int_\Omega \Psi(\mathcal{M}f)\mathrm{d}\mu \leq \int_\Omega \Phi(|f|)\mathrm{d}\mu.$$

Is it true that the analogous Fefferman-Stein estimate

$$\int_\Omega \Psi(\mathcal{M}f)w\mathrm{d}\mu \leq \int_\Omega \Phi(|f|)\mathcal{M}w\mathrm{d}\mu$$

holds in the weighted context? Basing on the contents of [65], we will give the affirmative answer to this question, under some mild regularity assumptions on Φ and Ψ. Here is the precise statement; the letter M stands for the dyadic maximal operator localized to $[0, 1)$.

Theorem 4.1.1. *Let $\Phi, \Psi : [0,\infty) \to [0,\infty)$ be nondecreasing functions such that Φ is convex and Ψ is left-continuous. Suppose in addition that the inequality*

$$||\Psi(Mf)||_{L^1(0,1)} \leq ||\Phi(|f|)||_{L^1(0,1)} \tag{4.2}$$

holds for all integrable functions $f : [0,1] \to \mathbb{R}$ (with the underlying Lebesgue measure). Then for any probability space (Ω,μ) with a tree structure $\mathcal{T}$, any weight w on Ω and any integrable function $f : \Omega \to \mathbb{R}$, we have

$$||\Psi(\mathcal{M}f)||_{L^1(w)} \leq ||\Phi(|f|)||_{L^1(\mathcal{M}w)}. \tag{4.3}$$

Thus, any "integral" inequality which is valid in the very special one-dimensional dyadic setting automatically extends to the general weighted context. A very nice feature is that the "shape" of the inequality is preserved, i.e., no additional multiplicative constants appear. This in particular implies that if the starting inequality (4.2) is sharp, then so is the weighted version (4.3). We will see several applications of the above result in the final subsection of this chapter.

It turns out that the Bellman function method is a very convenient tool for the proof of Theorem 4.1.1. The approach is quite natural: from the general theory, the validity of the assumed estimate (4.2) implies the existence of the associated Bellman function. Now it seems plausible to expect that carrying out some appropriate modifications should lead to the Bellman function corresponding to (4.3). This is indeed the case. We would like to mention that the analysis of the particular cases (iii) and (iv), presented in [53] and [55], was very helpful and indicated the direction in which the modifications should be made. A very nice feature of our result is that we treat a very general case and, in particular, do not need to invoke the explicit formula of the Bellman functions associated with (4.2) (its abstract form is sufficient).

4.2 A special function for the unweighted inequality

Our starting point is to associate a certain special abstract function with the assumed estimate (4.2). Throughout this subsection, we assume that the underlying probability space is the interval $[0,1)$ with Lebesgue's measure m, equipped with the dyadic tree $\mathcal{D}$.

We have already seen similar reasoning above, so we will be brief. We consider the domain $D = \{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq y\}$ and introduce the abstract function $U : D \to \mathbb{R} \cup \{\infty\}$ by the formula

$$U(x,y) = \sup\left\{\int_0^1 \Big(\Psi(Mf \vee y) - \Phi(f)\Big)\mathrm{d}m\right\}.$$

Here the supremum is taken over all $\mathcal{D}$-simple functions $f : [0,1) \to [0,\infty)$ satisfying $\int_0^1 f\mathrm{d}m = x$. The function U enjoys the following structural properties.

Lemma 4.2.1. *(i) We have $U(x,x) \leq 0$ for any $x \geq 0$.*

(ii) For any $(x,y) \in D$ we have $U(x,y) \geq \Psi(y) - \Phi(x)$.

(iii) The function U satisfies the following mid-concavity property: for any $(x,y) \in D$ and any $t \in [0,x]$ we have

$$U(x,y) \geq \frac{1}{2}U(x-t,y) + \frac{1}{2}U(x+t,(x+t)\vee y).$$

Proof. We omit the straightforward proof. We would only like to mention that the fact that we work with the special dyadic context implies that in the concavity condition (iii), one needs to consider the coefficients $1/2$ only. □

Properties (i) and (iii) imply that U is finite on D. We will also need the following additional technical fact.

Lemma 4.2.2. *For any $y > 0$ we have $\lim_{x\uparrow y} U(x,y) = U(y,y)$ and $\lim_{z\downarrow y} U(z,z) = U(y,y)$.*

Proof. Let f be an arbitrary nonnegative $\mathcal{D}$-simple function f on $[0,1)$ such that $\int_0^1 f\mathrm{d}m = y$. For any $z > y$, the function $f + (z-y)$ has integral z, so by the definition of $U(z,z)$,

$$\begin{aligned} U(z,z) &\geq \int_0^1 (\Psi(M(f+(z-y))\vee z) - \Phi(f+z-y))\mathrm{d}m \\ &= \int_0^1 (\Psi(M(f+(z-y))) - \Phi(f+z-y))\mathrm{d}m \\ &\geq \int_0^1 \big(\Psi(Mf) - \Phi(f+z-y)\big)\mathrm{d}m, \end{aligned}$$

where in the last line we have exploited the monotonicity of Ψ. Now, the function f takes a finite number of values, so using the convexity of Φ, the above estimate gives

$$\liminf_{z\downarrow y} U(z,z) \geq \int_0^1 \big(\Psi(Mf) - \Phi(f)\big)\mathrm{d}m.$$

Taking the supremum over all f, we get that

$$\liminf_{z\downarrow y} U(z,z) \geq U(y,y). \tag{4.4}$$

Next, observe that the condition (iii) implies in particular that for any $y > 0$, the function $x \mapsto U(x,y)$ is midpoint-concave on $[0,y]$. Since this function is bounded from below (by (ii)), it must be merely concave on $[0,y]$ and consequently,

$$\lim_{x\uparrow y} U(x,y) \geq U(y,y). \tag{4.5}$$

However, applying (iii) with $x = y$, we get

$$U(y,y) \geq \frac{1}{2}U(y-t,y) + \frac{1}{2}U(y+t,y+t),$$

and letting $t \to 0$ gives the claim, in the light of (4.4) and (4.5). □

Now we will modify the function U slightly, to ensure certain additional properties.

Lemma 4.2.3. *The function $\tilde{U} : D \to \mathbb{R}$ given by $\tilde{U}(x,y) = \min\{U(x,y), \Psi(y)\}$ satisfies (i), (ii) and the following two conditions.*

(iii') For any $(x,y) \in D$, $x \neq 0$, there exists a number $A(x,y)$ such that for any $s \geq -x$ we have

$$\tilde{U}(x+s,(x+s)\vee y) \leq \tilde{U}(x,y) + A(x,y)s. \tag{4.6}$$

(iv) For any $x,\ y \geq 0$ we have $\tilde{U}(x, x\vee y) \leq \Psi(y)$.

Note that (iii') is an extension of the concavity condition (iii), which can be seen by applying (4.6) to $s = -t$, $s = t$ and summing the obtained estimates.

Proof. The conditions (i) and (ii) are obviously preserved. The condition (iv) is also straightforward: if $x \geq y$, then $\tilde{U}(x, x \vee y) \leq U(x,x) \leq 0 \leq \Psi(y)$; on the other hand, if $x < y$, then the inequality follows directly from the definition of $\tilde{U}$. Thus, it remains to handle the property (iii'). As we have seen in the proof of the previous lemma, for any $y \geq 0$ the function $x \mapsto U(x,y)$ is concave on $[0,y]$ and continuous on $(0,y]$; it is clear that the function $\tilde{U}$ inherits both these properties. Set $A(x,y)$ to be the corresponding left-sided derivative $\tilde{U}_x(x-,y)$. Then the inequality (4.6) is obviously satisfied if $x+s \leq y$; on the other hand, if $x+s > y$, then we write

$$\begin{aligned} \tilde{U}(x,y) + A(x,y)s &= \tilde{U}(x,y) + A(x,y)(y-x) + A(x,y)(x+s-y) \\ &\geq \tilde{U}(y,y) + A(y,y)(x+s-y). \end{aligned}$$

Here in the last passage we have used (4.6) with $s = y - x$ (which we have just proved) and the inequality $A(x,y) \geq A(y,y)$ which is due to the concavity of $x \mapsto \tilde{U}(x,y)$. However, directly from the definition of $\tilde{U}$, the functions U and $\tilde{U}$ coincide on the diagonal $x = y$ and the left-sided derivative $A(y,y) = \tilde{U}_x(y-,y)$ is not smaller than the left-sided derivative $U_x(y-,y)$, so it is enough to prove that

$$U(x+s,x+s) \leq U(y,y) + U_x(y-,y)(x+s-y).$$

Suppose that the inequality is not satisfied: for some positive κ, we have

$$U(x+s,x+s) \geq U(y,y) + U_x(y-,y)(x+s-y) + \kappa. \tag{4.7}$$

An application of (iii) with $x = y := (x+s+y)/2$ and $t = (x+s-y)/2$ gives

$$U\left(\frac{x+s+y}{2}, \frac{x+s+y}{2}\right) \geq \frac{1}{2}U\left(y, \frac{x+s+y}{2}\right) + \frac{1}{2}U(x+s,x+s).$$

Directly from the definition of U and the monotonicity of Ψ, we see that $U(x,y)$ increases as y increases. Consequently, we have $U\left(y,(x+s+y)/2\right) \geq U(y,y)$, which combined with the previous estimate and (4.7) gives

$$U\left(\frac{x+s+y}{2}, \frac{x+s+y}{2}\right) \geq U(y,y) + U_x(y-,y)\frac{x+s-y}{2} + \frac{\kappa}{2}.$$

Note that this estimate is of the same form as (4.7) and we may iterate the above reasoning to obtain

$$U\left(y + \frac{x+s-y}{2^n}, y + \frac{x+s-y}{2^n}\right) \geq U(y,y) + U_x(y-,y)\frac{x+s-y}{2^n} + \frac{\kappa}{2^n}$$

for any nonnegative integer n. Now, if n is chosen sufficiently large, then $(x+s-y)/2^n \leq y$ and yet another application of (iii) gives

$$U(y,y) \geq \frac{1}{2}U\left(y - \frac{x+s-y}{2^n}, y\right) + \frac{1}{2}U\left(y + \frac{x+s-y}{2^n}, y + \frac{x+s-y}{2^n}\right),$$

which combined with the previous bound implies

$$U(y,y) - U\left(y - \frac{x+s-y}{2^n}, y\right) \geq U_x(y-,y)\frac{x+s-y}{2^n} + \frac{\kappa}{2^n}.$$

Dividing by $(x+s-y)/2^n$ and letting $n \to \infty$ yields $U_x(y-,y) \geq U_x(y-,y) + \kappa/(x+s-y)$, a contradiction to (4.7). □

4.3 A special function for the weighted inequality and a proof of Theorem 4.1.1

Now we will complicate the function $\tilde{U}$ constructed above to obtain the Bellman function corresponding to (4.3). First we introduce the extended domain $D_0 = \{(x,y,u,v) \in \mathbb{R}^4 : 0 \leq x \leq y,\ 0 \leq u \leq v\}$ and let $B : D_0 \to \mathbb{R}$ be given by

$$B(x,y,u,v) = (u-v)\Psi(y) + v\tilde{U}(x,y).$$

Observe that the properties (i) and (ii) enjoyed by $\tilde{U}$ imply

(i) $B(x,x,u,u) = u\tilde{U}(x,x) \leq 0$.

(ii) $B(x,y,u,v) \geq u\Psi(y) - v\Phi(x)$.

We will also need a property analogous to (iii').

Lemma 4.3.1. *The function B satisfies the following concavity-type property. Pick an arbitrary point $(x,y,u,v) \in D_0$ and two numbers $s \geq -x$, $t \geq -u$. If $x \neq 0$, then*

$$B(x+s, y \vee (x+s), u+t, v \vee (u+t)) \leq B(x,y,u,v) + svA(x,y) + t\Psi(y),$$

where A is the function guaranteed by Lemma 4.2.3. If $x = 0$, then

$$B(x,y,u+t,v \vee (u+t)) \leq B(x,y,u,v) + t\Psi(y).$$

Proof. The second estimate is straightforward. Indeed, if $u+t \leq v$, then both sides are equal, while for $u+t > v$ we use the bound $\tilde{U}(0,y) \leq \Psi(y)$ (the property (iv)) to get

$$\begin{aligned} B(x,y,u+t,v \vee (u+t)) = (u+t)\tilde{U}(x,y) &\leq (u+t-v)\Psi(y) + v\tilde{U}(x,y) \\ &= B(x,y,u,v) + t\Psi(y). \end{aligned}$$

To prove the first inequality of the lemma, we consider three cases.

The case $x+s \leq y$ and $u+t \leq v$. From the property (iii') we have $\tilde{U}(x+s,y) \leq \tilde{U}(x,y) + sA(x,y)$ and therefore

$$\begin{aligned} &B(x+s, y \vee (x+s), u+t, v \vee (u+t)) = B(x+s,y,u+t,v) \\ &\quad = (u+t-v)\Psi(y) + v\tilde{U}(x+s,y) \leq (u+t-v)\Psi(y) + v(\tilde{U}(x,y) + sA(x,y)) \\ &\quad = B(x,y,u,v) + svA(x,y) + t\Psi(y). \end{aligned}$$

The case $x+s > y$ *and* $u+t \le v$. The property (iii') yields $\tilde{U}(x+s, x+s) \le \tilde{U}(x, y)+sA(x, y)$. Moreover, notice that $\Psi(y) \le \Psi(x+s)$, since Ψ is nondecreasing. Therefore

$$\begin{aligned}
&B(x+s, y \vee (x+s), u+t, v \vee (u+t)) = B(x+s, x+s, u+t, v)\\
&\quad = (u+t-v)\Psi(x+s) + v\tilde{U}(x+s, x+s)\\
&\quad \le (u+t-v)\Psi(y) + v(\tilde{U}(x, y) + sA(x, y)) = B(x, y, u, v) + svA(x, y) + t\Psi(y).
\end{aligned}$$

The case $u+t > v$. Using the property (iii') and the inequality $\Psi(y) \ge \tilde{U}(x+s, y \vee (x+s))$ (guaranteed by the property (iv)), we have

$$\begin{aligned}
B(x, y, u, v) + svA(x, y) + t\Psi(y) &= (u+t-v)\Psi(y) + v(\tilde{U}(x, y) + sA(x, y))\\
&\ge (u+t-v)\Psi(y) + v\tilde{U}(x+s, y \vee (x+s))\\
&\ge (u+t)\tilde{U}(x+s, y \vee (x+s))\\
&= B(x+s, y \vee (x+s), u+t, v \vee (u+t)).
\end{aligned}$$

This completes the proof. $\square$

It remains to observe that the function B is the Bellman function corresponding to the estimate (4.3). Indeed, the properties (i) and (ii) are the appropriate size requirements, while (iii) gives the necessary concavity. See the previous chapter for a similar argumentation.

4.4 Applications

We conclude this chapter by the discussion on several interesting applications of the above statement. For the sake of completeness, we have decided to include the results discussed at the beginning.

(i) If we take $\Psi(x) = \lambda\chi_{(\lambda,\infty)}(x)$ and $\Phi(x) = x$ for a given $\lambda > 0$, then (4.2) becomes the unweighted weak-type bound (1.1), and the assertion (4.3) is precisely the aforementioned result of Fefferman and Stein (1.3), in the context of trees.

(ii) The choice $\Psi(x) = x^p$ and $\Phi(x) = (p/(p-1))^p x^p$ for a given $p > 1$ leads to the tree version of the weighted L^p-bound (1.4).

(iii) Let $K > 1$ be a fixed parameter. Setting $\Psi(x) = x$ and $\Phi(x) = Kx\log x + K^2/((K-1)e)$ corresponds to the sharp $L \log L$ bound (4.1).

(iv) Now we will present an application to a different type of maximal operator, the so-called geometric maximal operator $\mathcal{M}^{\mathcal{G}}$ associated with the tree $\mathcal{T}$. This object acts on log-integrable functions f on Ω by the formula

$$\mathcal{M}^{\mathcal{G}} f(x) = \sup\left\{\exp\left(\langle \log|f|\rangle_{Q,\mu}\right) : Q \in \mathcal{T},\ x \in Q\right\}.$$

It is well-known (see e.g. [22]) that for $p > 0$ we have $||\mathcal{M}^{\mathcal{G}} f||_{L^p(\Omega)} \le e^{1/p}||f||_{L^p(\Omega)}$ and the above theorem yields the following weighted version. Namely, the identity $\|\mathcal{M}^{\mathcal{G}}\|_{L^p \to L^p} = e^{1/p}$ implies the sharp estimate

$$\int_\Omega \exp(p\mathcal{M}g)\mathrm{d}\mu \le \int_\Omega \exp(1+p|g|)\mathrm{d}\mu$$

(indeed: this is $||\mathcal{M}^{\mathcal{G}} f||_{L^p(\Omega)} \leq e^{1/p}||f||_{L^p(\Omega)}$ with $f = e^{|g|}$). This inequality can be rewritten in the form (4.2), with $\Psi(x) = \exp(px)$ and $\Phi(x) = \exp(1+px)$. Consequently, Theorem 4.1.1 yields

$$\int_\Omega \exp(p\mathcal{M}g)w\mathrm{d}\mu \leq \int_\Omega \exp(1+p|g|)\mathcal{M}w\mathrm{d}\mu.$$

In particular, applying this bound to $g = \log|f|$, we obtain the estimate

$$\int_\Omega (\mathcal{M}^{\mathcal{G}} f)^p w\mathrm{d}\mu \leq \int_\Omega \exp(p\mathcal{M}(\log|f|))w\mathrm{d}\mu \leq \int_\Omega \exp(1+p|\log|f||)\mathcal{M}w\mathrm{d}\mu,$$

which is a bit worse than the desired weighted L^p bound for $\mathcal{M}^{\mathcal{G}}$. To overcome this problem, assume first that $\log|f|$ is bounded from below by some constant N and replace f by fe^{-N} in the latter estimate. As the result, we get an inequality equivalent to

$$\int_\Omega (\mathcal{M}^{\mathcal{G}} f)^p w\mathrm{d}\mu \leq e\int_\Omega |f|^p\mathcal{M}w\mathrm{d}\mu,$$

as needed. The general case follows by standard limiting arguments: we use the latter bound for $|f| \vee e^{-N_1}$ and finally let $N_1 \to \infty$.

(v) We would like to emphasize that the results obtained in the previous chapter do not quite fit into the picture described above. Kolmogorov's inequality studied there does not seem to transfer easily to the setup of Theorem 4.1.1.

(vi) Finally, we will show how the above approach leads to the sharp $L^{p,\infty} \to L^{p,\infty}$ estimate. The starting point is the unweighted bound for the dyadic maximal operator M on $[0,1)$:

$$|\{Mf > 1\}| \leq \int_0^1 (p(f-1)+1)^+\mathrm{d}m. \tag{4.8}$$

Let us decompose the set $\{Mf > 1\}$ into the union of pairwise disjoint maximal dyadic intervals. In other words, for any $x \in [0,1)$ let $\tau(x) = \inf\{n : \langle|f|\rangle_{Q^n(x)} > 1\}$, where $Q^n(x)$ is the unique dyadic interval with the measure 2^{-n} containing x. Here we use the convention $\inf\emptyset = \infty$. Then the collection $\{Q^\tau(x) : x \in [0,1),\, \tau(x) < \infty\}$ is the aforementioned family of maximal elements from which the set $\{Mf > 1\}$ is built. Now, for any such maximal element Q, we have

$$|Q| = \int_Q \mathrm{d}m \leq \int_Q (p(|f|-1)+1)\mathrm{d}m \leq \int_Q (p(|f|-1)+1)^+\mathrm{d}m,$$

and summing over all Q yields (4.8) (actually, even a slightly stronger form). Therefore, Theorem 4.1.1 gives the weighted variant of this bound:

$$w(\{\mathcal{M}f > 1\}) \leq \int_\Omega (p(|f|-1)+1)^+\mathcal{M}w\mathrm{d}\mu.$$

Now we proceed as follows: we have

$$\begin{aligned}\int_\Omega (p(|f|-1)+1)^+\mathcal{M}w\mathrm{d}\mu &= \int_\Omega p\left(|f| - \frac{p-1}{p}\right)^+ \mathcal{M}w\mathrm{d}\mu \\ &= p\int_{(p-1)/p}^\infty \mathcal{M}w(\{|f| > t\})\mathrm{d}t \\ &\leq p\int_{(p-1)/p}^\infty t^{-p}\|f\|^p_{L^{p,\infty}(\mathcal{M}w)}\mathrm{d}t = \left(\frac{p}{p-1}\right)^p \|f\|^p_{L^{p,\infty}(\mathcal{M}w)}.\end{aligned}$$

It remains to note that by homogeneity, this implies $\|\mathcal{M}f\|_{L^{p,\infty}(w)} \leq \frac{p}{p-1}\|f\|_{L^{p,\infty}(\mathcal{M}w)}$. The constant $\frac{p}{p-1}$ is the best possible in the unweighted setting: consult e.g. [37].

Chapter 5

A weak-type inequality with an A_p weight

5.1 Motivation and the statement of results

We turn our attention to a certain maximal estimate in the presence of Muckenhoupt's A_p weights. The starting point is the classical weak-type $(1,1)$ inequality

$$\lambda(\{Mf > \lambda\}) \leq \int_{\mathbb{R}^d} |f| \mathrm{d}x, \qquad \lambda > 0,$$

where M is the dyadic maximal operator on $\mathbb{R}^d$. The following problem has gained a lot of interest in the literature: can we replace the number λ with a weight? In other words, given a weight w, is the estimate

$$w(\{Mf > w\}) \leq C_w \int_{\mathbb{R}^d} |f| \mathrm{d}x \tag{5.1}$$

valid for all integrable functions f, where the constant C_w depends on the weight only? It is easy to see that in such a generality, the answer is negative. Indeed, let f be a nonzero integrable function. Then Mf is not integrable, and hence the choice $w = Mf/2$ provides us with a counterexample. Therefore, if we want (5.1) to hold, we need some assumptions on the weight. For example, Muckenhoupt's condition A_1 is sufficient, as was proved in [29], and then the estimate holds with $C_w = [w]_{A_1}$ (this is even true for non-dyadic A_1 context). Let us briefly show an extension of this result to the context of probability space (Ω, μ) with a tree structure $\mathcal{T}$. Namely, we have the stronger bound

$$w(\{\mathcal{M}f > \mathcal{M}w\}) \leq \|f\|_{L^1}$$

for arbitrary weight; this yields

$$w(\{\mathcal{M}f > w\}) \leq [w]_{A_1} \|f\|_{L^1} \tag{5.2}$$

be a simple homogeneity argument. To prove the above estimate, we use the following argument: let $\mathcal{Q}$ be the collection of maximal elements $Q \in \mathcal{T}$ such that $\langle |f| \rangle_{Q,\mu} \geq \langle w \rangle_{Q,\mu}$. Such elements are pairwise disjoint and the set $\{\mathcal{M}f > \mathcal{M}w\}$ is contained in the union $\bigcup \mathcal{Q}$. Consequently, we obtain

$$w(\{\mathcal{M}f > \mathcal{M}w\}) \leq \sum_{Q \in \mathcal{Q}} w(Q) \leq \sum_{Q \in \mathcal{Q}} \int_Q |f| \mathrm{d}\mu \leq \|f\|_{L^1},$$

and the estimate is established. It is not difficult to see that the linear dependence on the A_1 characteristic is optimal: for any $\kappa < 1$, the inequality

$$w(\{\mathcal{M}f > w\}) \leq C[w]_{A_1}^{\kappa} \|f\|_{L^1}$$

does not hold with any finite constant C (see examples below). There is an interesting question whether the estimate (5.2) extends to the more general setting in which the weight w is assumed to belong to the class A_p. Furthermore, if the answer is positive, what is the optimal dependence on the A_p characteristic? We will answer both these questions, basing on the results obtained in [60]. Here is the precise formulation of the main statement.

Theorem 5.1.1. *Let (Ω, μ) be a probability space with a tree $\mathcal{T}$. If $1 < p < \infty$ and w is an A_p weight on Ω, then for any integrable function $f : \Omega \to \mathbb{R}$ we have*

$$w\left(\{\mathcal{M}f(x) > w(x)\}\right) \leq 2ep[w]_{A_p} \int_\Omega f dx. \tag{5.3}$$

The linear dependence on the A_p characteristic is optimal.

It is worth stressing here that we do not impose any regularity condition on $\mathcal{T}$: for any element Q of $\mathcal{T}$ and any child Q' of Q, the ratio $\mu(Q')/\mu(Q)$ need not be bounded away from 0 or 1. On the other hand, the result is new even in the context of dyadic filtrations.

The following example shows that the linear independence on $[w]_{A_p}$ is optimal already in the one-dimensional dyadic context (i.e., for $[0, 1)$ with the dyadic tree structure). Fix an arbitrary integer $N > 1$. Define $f = 2^N \chi_{[0;2^{-N})}$ and $w = \mathcal{M}f/2$. Directly from the definition of the maximal operator, we easily compute that

$$\mathcal{M}f = 2^N \chi_{[0;2^{-N})} + \sum_{k=0}^{N-1} 2^k \chi_{[2^{-1-k};2^{-k})}.$$

We have $w(\mathcal{M}f \geq w) = w([0, 1)) = 1/2 + N/4$. After some easy calculations we see that $(N + 2)/2 > [w]_{A_p} > 2^{-p}(N + 2)$ and hence

$$\frac{w(\mathcal{M}f \geq w)}{[w]_{A_p} \int_0^1 f(x)\mathrm{d}x} > 1/2.$$

Since $[w]_{A_p} \to \infty$ as $N \to \infty$, this proves that we cannot have an inequality of the form $w(\mathcal{M}f \geq w) \leq C_p[w]_{A_p}^{\kappa} \int_\Omega f \mathrm{d}\mu$ with any $\kappa < 1$.

Thus, all we need is to establish (5.3), which will be handled with the use of the Bellman function method. We have already seen in Chapter 2 the appropriate modification of the approach: we need to construct a certain special function of four variables. Our initial considerations revealed that such an object probably has a quite complicated formula, and its discovery, as well as the verification of the required properties, seemed to be quite an elaborate issue. Fortunately, we have invented a shortcut which enables to overcome these technical difficulties. Namely, we will make a heavy use of an *abstract*, non-explicit formula for the Bellman function corresponding to (5.3). We have already encountered such an argument in the previous chapter, which also exploited such non-explicit Bellman functions; the novel and a little unexpected thing is that here we explore this path while studying a very particular estimate (and in the previous chapter a general inequality was investigated). The argument is motivated by analogous phenomenon which occurs in the context of weak-type estimates for the Haar system (see below).

5.2 A motivating example

To present the idea behind our approach, let us study, for a moment, the modification of the Bellman function method to the context of Haar multipliers, following the work of Burkholder [4]. Let $(h_n)_{n\geq 0}$ be the standard Haar system on $[0,1)$, i.e., a collection of functions given by $h_0 = \chi_{[0,1)}$, $h_1 = \chi_{[0,1/2)} - \chi_{[1/2,1)}$, $h_2 = \chi_{[0,1/4)} - \chi_{[1/4,1/2)}$, $h_3 = \chi_{[1/2,3/4)} - \chi_{[3/4,1)}$, and so on. Suppose that $V : \mathbb{R}^2 \to \mathbb{R}$ is a given function and assume that we are interested in showing the inequality

$$\int_{[0,1)} V\left(\sum_{k=0}^n a_k h_k, \sum_{k=0}^n \epsilon_k a_k h_k\right) \mathrm{d}x \leq 0 \qquad n = 0,\, 1,\, 2,\, \ldots, \tag{5.4}$$

for any sequence $(a_k)_{k\geq 0}$ of integers and any sequence $(\epsilon_k)_{k\geq 0}$ of signs. For instance, for the choice $V(x,y) = |y|^p - C_p^p|x|^p$ (where $1 < p < \infty$) the above estimate is related to the unconditionality of the Haar system. The key to handle this problem is to consider the class of all functions $B : \mathbb{R}^2 \to \mathbb{R}$ which enjoy the following properties:

1° (Initial condition) $B(x, \pm x) \leq 0$ for all $x \in \mathbb{R}$;
2° (Majorization) $B \geq V$ on $\mathbb{R}^2$;
3° (Concavity-type property) B is concave along any line of slope ± 1.

The existence of a function B with the above properties implies the validity of (5.4). Indeed, the third condition implies that for any $n \geq 0$ we have

$$\int_0^1 B\left(\sum_{k=0}^{n+1} a_k h_k, \sum_{k=0}^{n+1} \epsilon_k a_k h_k\right) \mathrm{d}x \leq \int_0^1 B\left(\sum_{k=0}^{n} a_k h_k, \sum_{k=0}^{n} \epsilon_k a_k h_k\right) \mathrm{d}x$$

(this is just the conditional Jensen's inequality), so by 2° and finally 1°, we obtain

$$\begin{aligned}\int_0^1 V\left(\sum_{k=0}^{n} a_k h_k, \sum_{k=0}^{n} \epsilon_k a_k h_k\right) \mathrm{d}x &\leq \int_0^1 B\left(\sum_{k=0}^{n} a_k h_k, \sum_{k=0}^{n} \epsilon_k a_k h_k\right) \mathrm{d}x \\ &\leq \int_0^1 B(a_0, \epsilon a_0)\mathrm{d}x \leq 0.\end{aligned}$$

Probably the simplest inequality which can be studied with the above approach is the L^2 bound

$$\left\|\sum_{k=0}^n \epsilon_k a_k h_k\right\|_{L^2}^2 \leq \left\|\sum_{k=0}^n a_k h_k\right\|_{L^2}^2,$$

$n = 0,\, 1,\, 2,\, \ldots$ (which, of course, follows at once from the orthogonality of the Haar system). The corresponding function V, i.e., the one which transforms the L^2 bound into (5.4), is given by $V(x,y) = y^2 - x^2$, and it turns out that $B = V$ is the corresponding special function. Let us see what happens for the weak-type $(1,1)$ estimate

$$\left|\left\{x \in [0,1) : \left|\sum_{k=0}^n \epsilon_k a_k h_k(x)\right| > 1\right\}\right| \leq C\left\|\sum_{k=0}^n a_k h_k\right\|_{L^1},$$

for $n = 0, 1, 2, \ldots$. This inequality is of the form (5.4), with $V(x,y) = \chi_{\{|y|>1\}} - C|x|$, and using the above approach, Burkholder showed the estimate with the optimal constant $C = 2$. The special function B is slightly more complicated:

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

For some more or less formal arguments which lead to the discovery of this function, see e.g. [45, 46]. For the sake of our further considerations concerning the estimate (5.3), let us make here some important observations. We see that B is built of two components: if (x,y) is close to $(0,0)$, then it coincides with the special function corresponding to the L^2 estimate; for remaining (x,y), it is an affine expression (in $|x|$), which is almost equal to V. One easily checks 1° and 2°; to verify 3°, we rewrite the above formula as

$$B(x,y) = \begin{cases} \min\left\{y^2 - x^2, 1 - 2|x|\right\} & \text{if } |x| \le 1, \\ 1 - 2|x| & \text{if } |x| > 1 \end{cases} \tag{5.5}$$

and now it is clear that the concavity holds: both $(x,y) \mapsto y^2 - x^2$ and $(x,y) \mapsto 1 - 2|x|$ are concave along the lines of slope ± 1, and hence so is B, being essentially the minimum of the two.

As we will see in Section 5.4, the inequality (5.3) can be efficiently studied in a similar manner: it will be handled with a certain Bellman function given as the minimum of special functions associated with L^2 estimates and the appropriate affine expressions. More precisely, we will proceed as follows: first we will prove *directly* a certain weighted L^2 estimate for a yet another class of important class of operators in harmonic analysis, the so-called dyadic shifts; this will give us the existence of the associated Bellman function B. Then we will take an appropriate modification of the formula (5.5), with the term $y^2 - x^2$ replaced with B, to obtain the function for the weak-type estimate.

5.3 Bellman function method for maximal operators

We return to the context of arbitrary probability space (Ω, μ) equipped with a tree-like structure $\mathcal{T}$. Let $c \in [1, \infty)$, $p \in (1, \infty)$ be given parameters and let $V : [0,\infty)^3 \to \mathbb{R}$ be a fixed function. Let us briefly recall the material presented in Chapter 2. Suppose we are interested in showing the estimate

$$\int_\Omega V\left(f, \mathcal{M}f, w\right) \mathrm{d}\mu \le 0 \tag{5.6}$$

for any integrable function $f : \Omega \to [0,\infty)$ and any A_p weight w on Ω satisfying $[w]_{A_p} \le c$. To this end, we consider the four-dimensional domain

$$D = D_{p,c} = \left\{(x,y,u,v) \in [0,\infty)^4 \, : \, x \le y, \, 1 \le uv^{p-1} \le c\right\}.$$

Then the validity of (5.6) for $\mathcal{T}$-simple functions and weights is equivalent to the existence of a special function $B : D \to \mathbb{R}$, which enjoys the following structural properties.

1° (Initial condition) We have

$$\mathcal{B}(x,x,u,v) \le 0 \qquad \text{if } (x,x,u,v) \in D. \tag{5.7}$$

2° (Majorization) If $0 \le x \le y$, then

$$\mathcal{B}(x, y, u, u^{1/(1-p)}) \ge V(x, y, u). \tag{5.8}$$

3° (Concavity-type property) Let λ_1, λ_2, ..., $\lambda_m \ge 0$ be nonnegative numbers summing up to 1 and let (x, y, u, v), (x_1, y_1, u_1, v_1), ..., (x_m, y_m, u_m, v_m) be elements of D enjoying the following conditions: we have $y_j = \max\{x_j, y\}$ for all $j = 1, 2, \ldots, m$ and

$$x = \sum_{k=1}^m \lambda_k x_k, \qquad u = \sum_{k=1}^m \lambda_k u_k, \qquad v = \sum_{k=1}^m \lambda_k v_k.$$

Then we have

$$\mathcal{B}\,(x, y, u, v) \ge \sum_{k=1}^m \lambda_k \mathcal{B}(x_k, y_k, u_k, v_k). \tag{5.9}$$

Furthermore, if the inequality (5.6) holds, then the smallest special function $\mathfrak{B} : D \to \mathbb{R}$ is given by the formula

$$\mathfrak{B}(x, y, u, v) = \sup\left\{\int_\Omega V\left(f, \max\left\{\mathcal{M}f, y\right\}, w\right) \mathrm{d}\mu\right\}. \tag{5.10}$$

Here the supremum is taken over all probability spaces Ω with a tree $\mathcal{T}$, all $\mathcal{T}$-simple functions $f : \Omega \to [0, \infty)$ satisfying $\int_\Omega f \mathrm{d}\mu = x$, all $\mathcal{T}$-simple A_p weights w on Ω satisfying $[w]_{A_p} \le c$, $\int_\Omega w \mathrm{d}\mu = u$ and $\int_\Omega w^{1/(1-p)} \mathrm{d}\mu = v$.

5.4 Proof of Theorem 5.1.1

Our starting point is the sharp dimension-free weighted L^p estimate for maximal operators established in [54]. Namely, for any $1 < p < \infty$ and any probability space (Ω, μ) with the tree structure $\mathcal{T}$ and any A_p weight w on Ω, we have

$$\|\mathcal{M}\|_{L^p(w)\to L^p(w)} \le \frac{p}{p - 1 - d(p, [w]_{A_p})}.$$

Here, for a given $1 < p < \infty$ and $c \ge 1$, the constant $d(p, c)$ is the unique number in $[0, p-1)$ satisfying the equation

$$c(1 + d)(p - 1 - d)^{p-1} = (p - 1)^{p-1}.$$

We will need the more explicit bound

$$\begin{aligned} \|\mathcal{M}\|_{L^p(w)\to L^p(w)} &\le \frac{p}{p - 1 - d(p, [w]_{A_p})} \\ &= \frac{p}{p-1}\big((1 + d(p, [w]_{A_p}))[w]_{A_p}\big)^{1/(p-1)} \\ &\le \frac{p}{p-1} p^{1/(p-1)} [w]_{A_p}^{1/(p-1)} \le \frac{pe}{p-1}[w]_{A_p}^{1/(p-1)}. \end{aligned} \tag{5.11}$$

Let q be the harmonic conjugate to p and consider the weight w^{1-q} dual to w. Since $[w^{1-q}]_{A_q} = [w]_{A_p}^{q-1}$, the aforementioned theorem implies that

$$\|\mathcal{M}\|_{L^q(w^{1-q})\to L^q(w^{1-q})} \le \frac{qe}{q-1}[w]_{A_p} = pe[w]_{A_p}.$$

Equivalently, for any A_p weight w with $[w]_{A_p} \le c$ and any $f \in L^q(w^{1-q})$ we have

$$\int_\Omega V(f, \mathcal{M}f, w)\mathrm{d}\mu \le 0,$$

for $V(x,y,u) = y^q u^{1-q} - (pecx)^q u^{1-q}$. In particular, the above estimate holds for all $\mathcal{T}$-simple functions f. Therefore, by the Bellman function method, the function $\mathfrak{B}$ given by (5.10) enjoys the properties 1°, 2° and 3°. We will need the following enhanced version of the majorization.

Lemma 5.4.1. *For all $(x,y,u,v) \in D$ we have*

$$\mathfrak{B}(x,y,u,v) \ge y^q v - (pecx)^q u^{1-q}. \tag{5.12}$$

Proof. Let us go back to the definition (5.10) of $\mathfrak{B}(x,y,u,v)$ (with $V(x,y,u) = y^q u^{1-q} - (pecx)^q u^{1-q}$). Take there an arbitrary weight w with the appropriate conditions on characteristic and averages, and put $f = xw/u$. Since $\int_\Omega f\mathrm{d}\mu = x$, we have

$$\begin{aligned}
\mathfrak{B}(x,y,u,v) &\ge \int_\Omega \Big[\max\{\mathcal{M}f, y\}\Big]^q w^{1-q}\mathrm{d}\mu - (pec)^q \int_\Omega f^q w^{1-q}\mathrm{d}\mu \\
&\ge \int_\Omega y^q w^{1-q}\mathrm{d}\mu - (pecx)^q u^{-q}\int_\Omega w\mathrm{d}\mu \\
&= y^q v - (pecx)^q u^{1-q}.
\end{aligned}$$

□

Now we will modify $\mathfrak{B}$ to obtain the Bellman corresponding to the weak-type estimate (5.3). Define $\mathcal{B} : D \to \mathbb{R}$ by

$$\mathcal{B}(x,y,u,v) = \begin{cases} \min\Big\{\mathfrak{B}(x,y,u,v), u - 2pecx\Big\} & \text{if } pecx < u, \\ u - 2pecx & \text{if } pecx \ge u \end{cases} \tag{5.13}$$

and $\overline{V} : [0;\infty)^3 \to \mathbb{R}$ by $\overline{V}(x,y,u) = u\chi_{\{y \ge u\}} - 2pecx$. Obviously, we have

$$\mathcal{B}(x,y,u,v) \le u - 2pecx \qquad \text{on } D. \tag{5.14}$$

Furthermore, by (5.12), if $pecx = u$, then

$$\mathfrak{B}(x,y,u,v) \ge y^q v - pecx \cdot (pecxu^{-1})^{q-1} \ge -pecx = u - 2pecx, \tag{5.15}$$

so we also have

$$\mathcal{B}(x,y,u,v) = \begin{cases} \min\Big\{\mathfrak{B}(x,y,u,v), u - 2pecx\Big\} & \text{if } pecx \le u, \\ u - 2pecx & \text{if } pecx > u \end{cases}$$

(in comparison to the formula (5.13), the inequalities $pecx < u$ and $pecx \ge u$ have become non-strict and strict, respectively). We will need the following additional property of $\mathcal{B}$.

Lemma 5.4.2. *For any point $(x,y,u,v) \in D$ and any $x' > x$ we have*

$$\mathcal{B}(x', \max\{x', y\}, u, v) \ge \mathcal{B}(x,y,u,v) - 2pec(x' - x).$$

Proof. We split the reasoning into a few parts.

Step 1. An easy case. If $\mathcal{B}(x', \max\{x', y\}, u, v) = u - 2pecx'$, then the claim follows immediately from (5.14):

$$\mathcal{B}(x', \max\{x', y\}, u, v) = u - 2pecx - 2pec(x' - x) \geq \mathcal{B}(x, y, u, v) - 2pec(x' - x).$$

Hence, from now on, we assume that $\mathcal{B}(x', \max\{x', y\}, u, v) < u - 2pecx'$; this in particular implies that $\mathfrak{B}(x', \max\{x', y\}, u, v) = \mathcal{B}(x', \max\{x', y\}, u, v)$ and $pecx' < u$, by the definition of $\mathcal{B}$.

Step 2. Monotonicity of $\mathfrak{B}$ with respect to y. Fix $(x, y, u, v) \in D$. Observe that if $y' > y$, then

$$\mathfrak{B}(x, y, u, v) \leq \mathfrak{B}(x, y', u, v), \tag{5.16}$$

which follows from the definition of $\mathfrak{B}$. Indeed, if (Ω, μ), $\mathcal{T}$ is an arbitrary probability space with a tree, and f, w are functions on Ω as in the definition of $\mathfrak{B}(x, y, u, v)$, then

$$\begin{aligned}&\int_\Omega \Big[\max\big\{\mathcal{M}f, y\big\}\Big]^q w^{1-q} \mathrm{d}\mu - (pec)^q \int_\Omega f^q w^{1-q} \mathrm{d}\mu\\ &\leq \int_\Omega \Big[\max\big\{\mathcal{M}f, y'\big\}\Big]^q w^{1-q} \mathrm{d}\mu - (pec)^q \int_\Omega f^q w^{1-q} \mathrm{d}\mu \leq \mathfrak{B}(x, y', u, v).\end{aligned}$$

Taking the supremum over all f and w yields (5.16).

Step 3. An additional concavity. We have $pecx' < u$ (see the end of Step 1 above), so x' belongs to the interval $(x, u/(pec))$ and hence there is $\lambda \in (0, 1)$ such that $x' = \lambda x + (1 - \lambda)u/(pec)$. Therefore, an application of the concavity property of $\mathfrak{B}$ yields

$$\begin{aligned}&\mathcal{B}(x', \max\{x', y\}, u, v)\\ &= \mathfrak{B}(x', \max\{x', y\}, u, v)\\ &\geq \lambda \mathfrak{B}(x, \max\{x', y\}, u, v) + (1 - \lambda)\mathfrak{B}(u/(pec), \max\{u/(pec), y\}, u, v).\end{aligned} \tag{5.17}$$

However, by (5.16) and the inequality $pecx < pecx' < u$ we have

$$\mathfrak{B}(x, \max\{x', y\}, u, v) \geq \mathfrak{B}(x, y, u, v) \geq \mathcal{B}(x, y, u, v). \tag{5.18}$$

Furthermore, by (5.15) and the definition of $\mathcal{B}$, we see that

$$\mathfrak{B}(u/(pec), \max\{u/(pec), y\}, u, v) \geq \mathcal{B}(u/(pec), \max\{u/(pec), y\}, u, v),$$

so by Step 1 above,

$$\mathfrak{B}(u/(pec), \max\{u/(pec), y\}, u, v) \geq \mathcal{B}(x', \max\{x', y\}, u, v) - 2pec\left(\frac{u}{pec} - x'\right). \tag{5.19}$$

Plugging (5.18) and (5.19) into (5.17) yields the claim. □

We are ready for the main ingredient of Theorem 5.1.1.

Theorem 5.4.3. *The function $\mathcal{B}$ satisfies the conditions 1°, 2° and 3° (with respect to $\overline{V}$).*

Proof. The property 1° is easy to check: by the initial property of $\mathfrak{B}$, if $pecx \leq u$, then $\mathcal{B}(x,x,u,v) \leq \mathfrak{B}(x,x,u,v) \leq 0$; on the other hand, if $pecx \geq u$, then $\mathcal{B}(x,x,u,v) = u - 2pecx \leq -pecx \leq 0$.

We proceed to the majorization condition 2°. If $pecx \geq u$, then there is nothing to prove, so from now on we may assume that the reverse estimate holds. Suppose first that $y \geq u$. Then, by the definition of $\mathcal{B}$, the majorization is equivalent to $\mathfrak{B}(x,y,u,v) \geq u - 2pecx$. However, applying (5.12) (and using the estimate $u^{q-1}v \geq 1$), we get

$$\mathfrak{B}(x,y,u,v) \geq y^q v - pecx \cdot \left(pecxu^{-1}\right)^{q-1} \geq u - pecx \geq u - 2pecx.$$

So, it remains to verify 2° for $y < u$; then the desired bound becomes

$$\mathcal{B}(x,y,u,v) \geq -2pecx.$$

This is obvious if $\mathcal{B}(x,y,u,v) = u - 2pecx$; otherwise, again by (5.12),

$$\begin{aligned}\mathcal{B}(x,y,u,v) &= \mathfrak{B}(x,y,u,v)\\ &\geq y^q v - (pecx)^q u^{1-q} \geq -pecx \cdot \left(pecxu^{-1}\right)^{q-1} \geq -2pecx.\end{aligned}$$

It remains to establish 3°. If $\mathcal{B}(x,y,u,v) = u - 2pecx$, then the condition follows directly from (5.14). So, suppose that $\mathcal{B}(x,y,u,v) = \mathfrak{B}(x,y,u,v) < u - 2pecx$. In particular this implies $pecx < u$ and hence we have $pecx_j \leq u_j$ for at least one j; relabelling the points if necessary, we may and do assume that there is an integer k such that $pecx_1 \leq u_1$, $pecx_2 \leq u_2$, $\ldots$, $pecx_k \leq u_k$ and $pecx_{k+1} > u_{k+1}$, $pecx_{k+2} > u_{k+2}$, $\ldots$, $pecx_m > u_m$. Now we will run a backward induction with respect to k. First, if $k = m$, then the claim follows from the concavity property 3° of $\mathfrak{B}$:

$$\mathcal{B}(x,y,u,v) = \mathfrak{B}(x,y,u,v) \geq \sum_{j=1}^{m} \lambda_j \mathfrak{B}(x_j,y_j,u_j,v_j) \geq \sum_{j=1}^{m} \lambda_j \mathcal{B}(x_j,y_j,u_j,v_j).$$

We proceed to the induction step. Assume that $pecx_1 \leq u_1$, $pecx_2 \leq u_2$, $\ldots$, $pecx_{k-1} \leq u_{k-1}$ and $pecx_k > u_k$, $pecx_{k+1} > u_{k+1}$, $\ldots$, $pecx_m > u_m$. The idea is to modify x_j, but keeping their average $\sum_{j=1}^m \lambda_j x_j$ fixed. More specifically, we may increase x_1, x_2, $\ldots$, x_{k-1} a little bit (so that the estimates $pecx_j \leq u_j$ remain valid) and decrease x_k to make $pecx_k > u_k$ into equality; the points x_{k+1}, x_{k+2}, $\ldots$, x_m remain unchanged. For notational convenience, denote these new values by x_1', x_2', $\ldots$, x_m'. Then, by the induction assumption, we have

$$\mathcal{B}(x,y,u,v) \geq \sum_{j=1}^{m} \lambda_j \mathcal{B}(x_j', \max\{x_j', y\}, u_j, v_j). \tag{5.20}$$

Now, by the previous lemma and (5.16), for any $j \leq k-1$ we have

$$\begin{aligned}\mathcal{B}(x_j', \max\{x_j', y\}, u_j, v_j) &\geq \mathcal{B}(x_j, \max\{x_j', y\}, u_j, v_j) - 2pec(x_j' - x_j)\\ &\geq \mathcal{B}(x_j, \max\{x_j, y\}, u_j, v_j) - 2pec(x_j' - x_j).\end{aligned}$$

Furthermore, by (5.15),

$$\mathcal{B}(x_k', \max\{x_k', y\}, u_k, v_k) \geq u_k - 2pecx_k' = u_k - 2pecx_k - 2pec(x_k - x_k').$$

Plugging the last two estimates into (5.20), we complete the proof of the induction step: we obtain

$$\mathcal{B}(x, y, u, v) \geq \sum_{j=1}^{m} \lambda_j \mathcal{B}(x_j, y_j, u_j, v_j).$$

Thus, $\mathcal{B}$ has the desired concavity property. □

The properties of $\mathcal{B}$ immediately yield our main weighted estimate, by the Bellman function method described in the precious section.

Chapter 6

Sharp unweighted inequalities in Lorentz spaces

6.1 Motivation and the statement of results

In contrast to the other chapters, this part of the thesis is devoted to a certain estimate for the maximal function without the presence of weight. Suppose that (Ω, μ) is a measure space endowed with a tree structure $\mathcal{T}$ and $\mathcal{M}$ is the associated dyadic-like maximal operator. As we have seen in Chapter 1, the operator $\mathcal{M}$ enjoys a number of interesting sharp unweighted estimates, in particular, the strong- and weak-type bounds. Motivated by interpolation theory, one might ask about the norm of $\mathcal{M}$ as an operator between other Lorentz spaces. As proved by Melas and Nikolidakis [38], if $1 < p < \infty$ and $1 \leq q \leq \infty$, then we have $\|\mathcal{M}\|_{L^{p,q}(\Omega) \to L^{p,q}(\Omega)} = p/(p-1)$. We will study the question about the extension of this result to the case of different parameters q in the base and target Lorentz space. That is, we will be interested in the explicit formula for the norm of $\mathcal{M}$ as an operator from $L^{p,q_1}(\Omega)$ to $L^{p,q_2}(\Omega)$. This part of thesis is based on the contents of [59].

Let us first discuss some preliminary results for general choice of parameters: $0 < p < \infty$ and $0 < q_1, q_2 < \infty$. We start with the observation that if $p < 1$, then

$$\|\mathcal{M}\|_{L^{p,q_1} \to L^{p,q_2}} = \infty, \tag{6.1}$$

no matter what q_1 and q_2 are: to see this, test $\mathcal{M}$ on the function $f = \chi_A$, where $A \in \mathcal{T}$. Then we have $f^* = \chi_{(0,\mu(A)]}$ and hence

$$\|f\|_{L^{p,q_1}} = \left(\int_0^{\mu(A)} t^{q_1/p-1} \mathrm{d}t \right)^{1/q_1} = \left(\frac{p}{q_1} \right)^{1/q_1} \mu(A)^{1/p}.$$

On the other hand, we have $\mathcal{M}f \geq \langle f \rangle_{\Omega,\mu} = \mu(A)/\mu(\Omega)$ almost surely, so $(\mathcal{M}f)^* \geq \mu(A)/\mu(\Omega)$ on $(0, \mu(\Omega)]$ and hence

$$\|\mathcal{M}f\|_{L^{p,q_2}} \geq \left(\int_0^{\mu(\Omega)} t^{q_2/p-1} (\mu(A)/\mu(\Omega))^{q_2} \mathrm{d}t \right)^{1/q_2} = \left(\frac{p}{q_2} \right)^{1/q_2} \mu(A) \mu(\Omega)^{\frac{1-p}{p}}.$$

Letting $\mu(A) \to 0$ shows that the ratio $\|\mathcal{M}f\|_{L^{p,q_2}} / \|f\|_{L^{p,q_1}}$ cannot be bounded.

For $p = 1$ the identity (6.1) holds as well, unless $q_1 = 1$ and $q_2 = \infty$ (but this special case has been already discussed in (1.1)). Therefore, from now on we only consider the case $p > 1$. Of course, if $q_1 > q_2$, then there are functions satisfying $\|f\|_{L^{p,q_1}} < \infty$ and $\|\mathcal{M}f\|_{L^{p,q_2}} \geq \|f\|_{L^{p,q_2}} = \infty$, so in this case (6.1) holds as well. Thus, the only notrivial cases left correspond to $1 < p < \infty$ and $0 < q_1 \leq q_2 < \infty$.

Our approach will allow us to study the case $1 < p \leq q_1 < q_2$, from now on we assume that this condition is satisfied. Set $\alpha = q_1/p - 1$, $\beta = q_2/p - 1$, $\gamma = q_1(p-1)/(p(q_1-1))$ and define

$$C_{p,q_1,q_2} = q_1^{\frac{1}{q_2}} (q_2(q_1-1))^{-\frac{1}{q_1}} \gamma^{\frac{q_2-q_1}{q_1q_2}-1} \left(\frac{(q_2-q_1)\Gamma\left(\frac{q_1q_2}{q_2-q_1}\right)}{\Gamma\left(\frac{q_2(q_1-1)}{q_2-q_1}\right)\Gamma\left(\frac{q_2}{q_2-q_1}\right)} \right)^{\frac{q_2-q_1}{q_1q_2}}.$$

Our main result can be formulated as follows.

Theorem 6.1.1. *Suppose that $1 < p \leq q_1 < q_2$ are fixed parameters. Then for any integrable function f on Ω,*

$$\|\mathcal{M}f\|_{L^{p,q_2}(\Omega)} \leq C_{p,q_1,q_2} \|f\|_{L^{p,q_1}(\Omega)} \tag{6.2}$$

and the constant on the right-hand side is the best possible for each individual tree.

Our approach will rest on a novel modification of the Bellman function method. As we have seen in Chapter 2, the classical version of the technique allows the study of inequalities which can be expressed in the integral form $\int_\Omega V(f, \mathcal{M}f)\mathrm{d}\mu \leq 0$. However, the Lorentz-norm estimates cannot be rewritten in such a form and hence some new splitting argument (leading to some concavity-type condition) is required. In the next section we introduce the abstract special function $\mathfrak{B}$ corresponding to (6.2): then the identification of the explicit formula for this object becomes our major task. To handle this problem, in Section 6.3 we establish an appropriate concavity-type (or rather monotone-type) property of $\mathfrak{B}$. Roughly speaking, this condition gives an indication how the function $\mathfrak{B}$ should look like: we present an informal reasoning which leads to an explicit *candidate* for $\mathfrak{B}$: we use a different symbol B for this object. In Section 6.4 we prove that this candidate satisfies $B \geq \mathfrak{B}$ (that is, B is a supersolution to our problem, in the terminology introduced in Chapter 2), which in particular yields the inequality (6.2). The final part of this chapter contains the proof of the reverse estimate $B \leq \mathfrak{B}$, which, in particular, allows us to show that the constant C_{p,q_1,q_2} in (6.2) is indeed the best possible.

As we shall work with different measure spaces, we will sometimes use the notation $\mathcal{M}_\Omega$ to emphasize that we study the action of the maximal operator on functions on Ω. One the other hand, for the sake of brevity, we have decided not to indicate the underlying tree structure. We believe that this should not lead to any confusion.

6.2 An abstract Bellman function

Suppose that x, y are nonnegative numbers and $T > 0$. Assume further that $q_1 \geq p$, so that $\alpha \geq 0$. Consider the class $\mathcal{C}(x, y, T)$, which consists of all nonnegative measurable functions f given on some measure space (Ω, μ) with $\mu(\Omega) = T$, such that

$$\frac{1}{T}\int_0^T f^*(t)\mathrm{d}t = x, \qquad \frac{1}{T}\int_0^T t^\alpha \left[f^*(t)\right]^{q_1}\mathrm{d}t \leq y.$$

Note that we have inequality in the second requirement. We emphasize that the measure space (Ω, μ) and the tree structure are allowed to vary. A simple application of Hölder's inequality shows that if the class $\mathcal{C}(x, y, T)$ is nonempty, then

$$T^\alpha x^{q_1} \leq \gamma^{1-q_1} y \tag{6.3}$$

(recall that $\gamma = q_1(p-1)/(p(q_1-1)))$. Actually, the reverse implication is also true, which can be seen by taking any measure space (Ω, μ) and any function $f : \Omega \to [0, \infty)$ satisfying $f^*(t) = \gamma x (T/t)^{\alpha/(q_1-1)}$ (for the existence of a function with a prescribed nonincreasing rearrangement, see [19, p. 65] or Lemma 2.3 in [50]). Note that if equality holds in (6.3), then this is the only choice for f^*.

The abstract Bellman function related to the estimate (6.2) is given by

$$\mathfrak{B}(x, y, T) = \sup \left\{ \int_0^T t^\beta \big[(\mathcal{M}f)^*(t)\big]^{q_2} \mathrm{d}t \, : \, f \in \mathcal{C}(x, y, T) \right\}$$

for $(x, y, T) \in [0, \infty)^2 \times (0, \infty)$ satisfying (6.3). We see a novel feature, which has already been discussed in Remark 2.1.3. The problem is that the estimate (6.2) cannot be rewritten in the integral form. To overcome this, we fix the right-hand side, which is hidden in the condition $\frac{1}{T}\int_0^T t^\alpha \big[f^*(t)\big]^{q_1} \mathrm{d}t \leq y$. This gives rise to the additional variable of the Bellman function and increases the difficulty (dimension) of the problem. At this cost, the method is still applicable.

In the next three sections, we will identify the explicit formula for $\mathfrak{B}$. We would like to emphasize here that our proof will yield a stronger fact. One might consider the above definition of $\mathcal{C}(x, y, T)$ and $\mathfrak{B}$ for a *fixed* measure space (Ω, μ) and a tree structure $\mathcal{T}$. We will actually show that for any such individual choice, the resulting Bellman function is the same. However, as it will be useful for us to switch the measure spaces and trees at some points of the proof, we have decided to work under the above definitions.

6.3 A candidate for the Bellman function

Throughout, we assume that $q_1 \geq p$. We start our search by proving the following estimate, which can be regarded as a version of the concavity-type property 3°.

Lemma 6.3.1. *For any S, $T > 0$ and any x, y, $c \geq 0$ we have*

$$\begin{aligned} &\mathfrak{B}\left(\frac{Tx + Sc}{T+S}, (T+S)^{-1}\left(Ty + c^{q_1} \cdot \frac{(T+S)^{\alpha+1} - T^{\alpha+1}}{\alpha+1}\right), T+S\right) \\ &\geq \mathfrak{B}(x, y, T) + \left(\frac{Tx+Sc}{T+S}\right)^{q_2} \frac{(T+S)^{\beta+1} - T^{\beta+1}}{\beta+1}. \end{aligned} \tag{6.4}$$

Proof. Take arbitrary measure spaces (Ω, μ_Ω), (Λ, μ_Λ) satisfying $\Omega \cap \Lambda = \emptyset$, $\mu_\Omega(\Omega) = T$, $\mu_\Lambda(\Lambda) = S$, equipped with some tree structures $\mathcal{T}_\Omega$, $\mathcal{T}_\Lambda$, respectively. Let $\mu_{\Omega\cup\Lambda}$ be the measure on the space $\Omega \cup \Lambda$, given by $\mu_{\Omega\cup\Lambda}(A \cup B) = \mu_\Omega(A) + \mu_\Lambda(B)$ for all measurable $A \subseteq \Omega$, $B \subseteq \Lambda$. Let $c \geq 0$ be a positive number. Suppose that $f : \Omega \to [0, \infty)$ satisfies

$$\frac{1}{T}\int_0^T f^*(t)\mathrm{d}t = x, \qquad \frac{1}{T}\int_0^T t^\alpha \big[f^*(t)\big]^{q_1} \mathrm{d}t \leq y \tag{6.5}$$

and consider its extension $\tilde{f} = f\mathbb{1}_\Omega + c\mathbb{1}_\Lambda$, a nonnegative function on the measure space $(\Omega \cup \Lambda, \mu_{\Omega\cup\Lambda})$. We compute directly that

$$\frac{1}{\mu_{\Omega\cup\Lambda}(\Omega\cup\Lambda)}\int_{\Omega\cup\Lambda}\tilde{f}\mathrm{d}\mu_{\Omega\cup\Lambda} = \frac{Tx+Sc}{T+S} \tag{6.6}$$

and, since $\alpha \geq 0$ (here is the place where we use the assumption $q_1 \geq p$),

$$\begin{aligned}
&\frac{1}{\mu_{\Omega\cup\Lambda}(\Omega\cup\Lambda)}\int_0^{T+S} t^\alpha[\tilde{f}^*(t)]^{q_1}\mathrm{d}t \\
&\leq \frac{1}{\mu_{\Omega\cup\Lambda}(\Omega\cup\Lambda)}\left[\int_0^T t^\alpha[f^*(t)]^{q_1}\mathrm{d}t + \int_T^{T+S} t^\alpha c^{q_1}\mathrm{d}t\right] \\
&= (T+S)^{-1}\left(Ty + c^{q_1}\cdot\frac{(T+S)^{\alpha+1}-T^{\alpha+1}}{\alpha+1}\right).
\end{aligned} \tag{6.7}$$

In other words, we have the inclusion

$$\tilde{f} \in \mathcal{C}\left(\frac{Tx+Sc}{T+S}, (T+S)^{-1}\left(Ty + c^{q_1}\cdot\frac{(T+S)^{\alpha+1}-T^{\alpha+1}}{\alpha+1}\right), T+S\right).$$

Now let us study the appropriate Lorentz norm of the maximal function of $\tilde{f}$. To this end, we equip the space $(\Omega\cup\Lambda, \mu_{\Omega\cup\Lambda})$ with the tree $\mathcal{T}_{\Omega\cup\Lambda}$ given by $\mathcal{T}^0_{\Omega\cup\Lambda} = \{\Omega\cup\Lambda\}$ and $\mathcal{T}^n_{\Omega\cup\Lambda} = \mathcal{T}^{n-1}_\Omega \cup \mathcal{T}^{n-1}_\Lambda$ for $n \geq 1$. To avoid confusion, we will denote by $\mathcal{M}_\Omega$ and $\mathcal{M}_{\Omega\cup\Lambda}$ the maximal operators on (Ω, μ_Ω) and $(\Omega\cup\Lambda, \mu_{\Omega\cup\Lambda})$. Of course, we may write

$$\begin{aligned}
&\int_0^{T+S} t^\beta\left[(\mathcal{M}_{\Omega\cup\Lambda}\tilde{f})^*(t)\right]^{q_2}\mathrm{d}t \\
&\quad = \int_0^T t^\beta\left[(\mathcal{M}_{\Omega\cup\Lambda}\tilde{f})^*(t)\right]^{q_2}\mathrm{d}t + \int_T^{T+S} t^\beta\left[(\mathcal{M}_{\Omega\cup\Lambda}\tilde{f})^*(t)\right]^{q_2}\mathrm{d}t.
\end{aligned}$$

Next, observe that on Ω,

$$\mathcal{M}_{\Omega\cup\Lambda}\tilde{f} = \max\left\{\mathcal{M}_\Omega f, \frac{1}{T+S}\int_{\Omega\cup\Lambda}\tilde{f}\mathrm{d}\mu_{\Omega\cup\Lambda}\right\} \geq \mathcal{M}_\Omega f.$$

Hence $(\mathcal{M}_{\Omega\cup\Lambda}\tilde{f})^* \geq (\mathcal{M}_\Omega f)^*$ on $(0, T]$ and the first integral on the right is not smaller than $\int_0^T t^\beta\left[(\mathcal{M}f)^*(t)\right]^{q_2}\mathrm{d}t$. To deal with the second integral, note that

$$\mathcal{M}_{\Omega\cup\Lambda}\tilde{f} \geq \frac{1}{\mu(\Omega\cup\Lambda)}\int_{\Omega\cup\Lambda}\tilde{f}\mathrm{d}\mu_{\Omega\cup\Lambda} = \frac{Tx+Sc}{T+S} \qquad \text{on } \Omega\cup\Lambda,$$

and hence

$$\int_T^{T+S} t^\beta\left[(\mathcal{M}_{\Omega\cup\Lambda}\tilde{f})^*(t)\right]^{q_2} \geq \left(\frac{Tx+Sc}{T+S}\right)^{q_2}\frac{(T+S)^{\beta+1}-T^{\beta+1}}{\beta+1}.$$

Thus, taking into account the above estimates for $\mathcal{M}_{\Omega\cup\Lambda}\tilde{f}$ and the conditions (6.6), (6.7), we obtain, by the very definition of $\mathfrak{B}$,

$$\begin{aligned}
&\mathfrak{B}\left(\frac{Tx+Sc}{T+S}, (T+S)^{-1}\left(Ty + c^{q_1}\cdot\frac{(T+S)^{\alpha+1}-T^{\alpha+1}}{\alpha+1}\right), T+S\right) \\
&\qquad \geq \int_0^T t^\beta\left[(\mathcal{M}f)^*(t)\right]^{q_2}\mathrm{d}t + \left(\frac{Tx+Sc}{T+S}\right)^{q_2}\frac{(T+S)^{\beta+1}-T^{\beta+1}}{\beta+1}.
\end{aligned}$$

Since (Ω, μ_Ω) was an arbitrary measure space and f was an arbitrary nonnegative function on Ω satisfying (6.5), we get the claim. $\square$

In what follows, we will also need a certain homogeneity-type property of $\mathfrak{B}$.

Lemma 6.3.2. *We have*

$$\mathfrak{B}(x,y,T) = x^{q_2} T^{\beta+1} \varphi\left(\frac{y}{x^{q_1} T^\alpha}\right), \tag{6.8}$$

where $\varphi(s) = \mathfrak{B}(1,s,1)$.

Proof. Fix an arbitrary measure space (Ω, μ) satisfying $\mu(\Omega) = T$ and an arbitrary function $f : \Omega \to [0,\infty)$ satisfying

$$\frac{1}{T}\int_0^T f^*(t)\mathrm{d}t = x, \qquad \frac{1}{T}\int_0^T t^\alpha \left[f^*(t)\right]^{q_1}\mathrm{d}t \le y.$$

Then for any $\lambda > 0$, the function $\tilde{f} = \lambda f$ satisfies

$$\frac{1}{T}\int_0^T \tilde{f}^*(t)\mathrm{d}t = \lambda x, \qquad \frac{1}{T}\int_0^T t^\alpha \left[\tilde{f}^*(t)\right]^{q_1}\mathrm{d}t \le \lambda^{q_1} y$$

and

$$\int_0^T t^\beta \left[(\mathcal{M}\tilde{f})^*(t)\right]^{q_2}\mathrm{d}t = \lambda^{q_2}\int_0^T t^\beta \left[(\mathcal{M}f)^*(t)\right]^{q_2}\mathrm{d}t,$$

so by the very definition of $\mathfrak{B}$ we obtain

$$\mathfrak{B}(\lambda x, \lambda^{q_1} y, T) \ge \lambda^{q_2}\int_0^T t^\beta \left[(\mathcal{M}f)^*(t)\right]^{q_2}\mathrm{d}t.$$

Since Ω and f were arbitrary, this gives $\mathfrak{B}(\lambda x, \lambda^{q_1} y, T) \ge \lambda^{q_2}\mathfrak{B}(x,y,T)$. Replacing x, y, λ with λx, $\lambda^{q_1} y$ and λ^{-1}, respectively, we get the reverse bound. Consequently, we may write

$$\mathfrak{B}(x,y,T) = x^{q_2}\mathfrak{B}(1, y/x^{q_1}, T). \tag{6.9}$$

Next, consider the space $(\Omega, \tilde{\mu}) := (\Omega, \mu/\lambda)$ with the same tree structure and let f be as above. We compute that

$$\frac{1}{\tilde{\mu}(\Omega)}\int_\Omega f \mathrm{d}\tilde{\mu} = x$$

and

$$\int_0^{T/\lambda} t^\alpha (f^*_{\tilde{\mu}}(t))^{q_1}\mathrm{d}t = \frac{\lambda^{-\alpha}}{T}\int_0^{T/\lambda} t^\alpha (f^*_{\tilde{\mu}}(t/\lambda))^{q_1}\mathrm{d}t = \frac{\lambda^{-\alpha}}{T}\int_0^T t^\alpha (f^*(t))^{q_1}\mathrm{d}t \le \lambda^{-\alpha} y.$$

Since $\mathcal{M}$ acts identically on the spaces (Ω, μ) and $(\Omega, \tilde{\mu})$, we have

$$\int_0^{T/\lambda} t^\beta ((\mathcal{M}f)^*_{\tilde{\mu}}(t))^{q_2}\mathrm{d}t = \lambda^{-\beta-1}\int_0^T t^\beta ((\mathcal{M}f)^*_\mu(t))^{q_2}\mathrm{d}t$$

and therefore, by the definition of $\mathfrak{B}$,

$$\mathfrak{B}(x, y/\lambda^\alpha, T/\lambda) \ge \lambda^{-\beta-1}\int_0^T t^\beta ((\mathcal{M}f)^*_\mu(t))^{q_2}\mathrm{d}t.$$

Since f was arbitrary, we get $\mathfrak{B}(x, y/\lambda^\alpha, T/\lambda) \geq \lambda^{-\beta-1}\mathfrak{B}(x, y, T)$. Replacing y, T, λ with $y\lambda^{-\alpha}$, T/λ and λ^{-1}, we obtain the reverse estimate. Combining this with (6.9), we finally arrive at

$$\mathfrak{B}(x, y, T) = x^{q_2}\mathfrak{B}(1, y/x^{q_1}, T) = x^{q_2}T^{\beta+1}\mathfrak{B}(1, x^{-q_1}yT^{-\alpha}, 1),$$

which is the desired identity. $\square$

To find the candidate for $\mathfrak{B}$, we will exploit the "infinitesimal" version of the concavity/monotonicity (6.4), which combined with the identity (6.8) will yield a certain ordinary differential inequality for φ. From now on we *assume* that $\mathfrak{B}$ is of class C^1. We would like to stress that at this point we may impose any regularity assumption, since our main purpose is to *guess* the explicit formula; the rigorous verification will be postponed to the next section.

Lemma 6.3.3. *The function* $\varphi = \mathfrak{B}(1, \cdot, 1)$ *satisfies*

$$\varphi\left(\gamma^{q_1-1}\right) = \frac{q_1}{q_2\gamma} \tag{6.10}$$

and the differential inequality

$$(q_1 - 1)\left[\gamma - \left(\frac{s\varphi'(s) - \frac{q_2}{q_1}\varphi(s)}{\varphi'(s)}\right)^{1/(q_1-1)}\right]\left(s\varphi'(s) - \frac{q_2}{q_1}\varphi(s)\right) \geq 1. \tag{6.11}$$

Proof. To show (6.10), note that the class $\mathcal{C}(1, \gamma^{q_1-1}, 1)$ contains only one element: see the discussion below (6.3) (formally: all the elements from the class have the same nonincreasing rearrangements) and hence the Bellman function can be directly evaluated. We turn our attention to the differential inequality. Put $T = x = 1$ and rewrite (6.4) in the form

$$\begin{aligned}\frac{1}{S}\left[\mathfrak{B}\left(\frac{1+Sc}{1+S}, (1+S)^{-1}\left(y + c^{q_1}\cdot\frac{(1+S)^{\alpha+1}-1}{\alpha+1}\right), 1+S\right) - \mathfrak{B}(1, y, 1)\right]&\\ \geq \left(\frac{1+Sc}{1+S}\right)^{q_2}\frac{(1+S)^{\beta+1}-1}{(\beta+1)S}&.\end{aligned}$$

Letting $S \to 0$ (and using the assumption that $\mathfrak{B}$ is of class C^1), we get the partial differential inequality

$$(-1 + c)\mathfrak{B}_x(1, y, 1) + (-y + c^{q_1})\mathfrak{B}_y(1, y, 1) + \mathfrak{B}_T(1, y, 1) \geq 1, \tag{6.12}$$

or equivalently

$$(q_2\varphi(y) - q_1y\varphi'(y))(c - 1) + \varphi'(y)(c^{q_1} - y) + (\beta + 1)\varphi(y) - \alpha y\varphi'(y) \geq 1.$$

Since $q_2/q_1 = (1 + \beta)/(1 + \alpha)$, this can be rewritten in the form

$$(q_2\varphi(y) - q_1y\varphi'(y))\left(c + \frac{\alpha+1}{q_1} - 1\right) + \varphi'(y)c^{q_1} \geq 1.$$

This estimate holds for all c, we may optimize over this parameter. Putting

$$c = \left(y - \frac{q_2\varphi(y)}{q_1\varphi'(y)}\right)^{1/(q_1-1)},$$

we obtain the desired differential inequality. $\square$

Now, let us *assume* that the differential inequality (6.11) is actually an equality. This leads us to the following candidate for the Bellman function. Namely, let φ be the solution of the differential *equation* (6.11) with the initial condition (6.10) (of course, we need to show that such a solution exists; this will be done below). Then the candidate B is obtained via the identity (6.8), i.e.,

$$B(x, y, T) = x^{q_2} T^{\beta+1} \varphi\left(\frac{y}{x^{q_1} T^{\alpha}}\right). \tag{6.13}$$

6.4 Proof of $\mathfrak{B} \leq B$

We start the formal analysis by showing that B is well-defined. To this end, we need the rigorous definition of φ. This will be proved with the help of the following statement.

Lemma 6.4.1. *For any $s > \gamma^{q_1-1}$ there is a unique $u = u(s) \in (0, \gamma)$ which satisfies the identity*

$$\begin{aligned} \frac{q_2(q_1-1)}{q_2-q_1} \int_u^{\gamma} & (\gamma - w)^{q_1/(q_2-q_1)}\, w^{q_1(q_1-1)/(q_2-q_1)+q_1-2} dw \\ &= \left(s - u^{q_1-1}\right)(\gamma-u)^{q_1/(q_2-q_1)}\, u^{q_1(q_1-1)/(q_2-q_1)}. \end{aligned} \tag{6.14}$$

Furthermore, $\lim_{s\to\gamma^{q_1-1}} u(s) = \gamma$ and $\lim_{s\to\infty} u(s) = 0$.

Proof. For a fixed s, consider the difference of the left- and the right-hand side as a function of $u \in (0, \gamma)$ and denote it by $F(u)$. A bit lengthy computation shows that

$$F'(u) = \frac{q_1}{q_2-q_1} (\gamma-u)^{q_2/(q_2-q_1)-1}\, u^{q_1(q_1-1)/(q_2-q_1)-1} G(u),$$

where $G(u) = s(q_1 u - q_1 + 1 + \alpha) - u^{q_1}$. Since $G'(u) = q_1(s - u^{q_1-1})$, the function G is increasing on the interval $(0, \gamma)$. Note that $G(0) = s(-q_1 + 1 + \alpha) < 0$ and

$$G(\gamma) = \gamma\left(s - \gamma^{q_1-1}\right) > 0,$$

so there is a unique u_0 such that the function G is negative on $(0, u_0)$ and positive on (u_0, γ). This implies that F decreases on $(0, u_0)$ and increases on (u_0, γ); since $F(0) > 0$ and $F(\gamma) = 0$, the existence of $u(s)$ is proved. The limiting behavior of this function as $s \to \gamma^{q_1-1}$ or $s \to \infty$ follows quickly from the definition (6.14). $\square$

Letting $s \to \infty$ in (6.14) and using the fact that $u(s) \to 0$, we see that

$$\begin{aligned} \frac{q_2(q_1-1)}{q_2-q_1} \int_0^{\gamma} & (\gamma - w)^{q_1/(q_2-q_1)}\, w^{q_1(q_1-1)/(q_2-q_1)+q_1-2} \mathrm{d}w \\ &= \gamma^{q_1/(q_2-q_1)} \lim_{s\to\infty} s u(s)^{q_1(q_1-1)/(q_2-q_1)}, \end{aligned}$$

or equivalently,

$$\begin{aligned} &\lim_{s\to\infty} s^{q_2/q_1-1} u(s)^{q_1-1} \\ &\quad = \left[\frac{q_2(q_1-1)}{q_2-q_1} \cdot \frac{\Gamma\left(\frac{q_2}{q_2-q_1}\right)\Gamma\left(\frac{q_2(q_1-1)}{q_2-q_1}\right)}{\Gamma\left(\frac{q_1 q_2}{q_2-q_1}\right)}\right]^{q_2/q_1-1} \gamma^{q_2(q_1-1)/q_1}, \end{aligned} \tag{6.15}$$

by the properties of beta function. We are ready for the proof of the existence of the function φ.

Lemma 6.4.2. *There is an increasing function $\varphi : [\gamma^{q_1-1}, \infty) \to \mathbb{R}$, satisfying the differential equation*

$$(q_1-1)\left[\gamma - \left(\frac{s\varphi'(s) - \frac{q_2}{q_1}\varphi(s)}{\varphi'(s)}\right)^{1/(q_1-1)}\right]\left(s\varphi'(s) - \frac{q_2}{q_1}\varphi(s)\right) = 1 \tag{6.16}$$

for $s > \gamma^{q_1-1}$ and the initial condition $\varphi\left(\gamma^{q_1-1}\right) = \frac{q_1}{q_2\gamma}$. Furthermore, we have $\varphi(s) \le C_{p,q_1,q_2}^{q_2} s^{q_2/q_1}$ for all s.

Proof. Define φ by the formula

$$\varphi(s) = \frac{q_1(s - u^{q_1-1}(s))}{q_2(q_1-1)u^{q_1-1}(s)(\gamma - u(s))}, \qquad s > \gamma^{q_1-1},$$

where u comes from the previous lemma. Some lengthy calculations show that

$$\varphi'(s) = \frac{1}{(q_1-1)u^{q_1-1}(s)(\gamma-u(s))} = \frac{\frac{q_2}{q_1}\varphi(s)}{s - u^{q_1-1}(s)}.$$

Consequently, we have $u(s) = (s - \frac{q_2}{q_1}\varphi(s)/\varphi'(s))^{1/(q_1-1)}$ and (6.16) follows. To prove the initial condition, recall that by the previous lemma,

$$\lim_{s\to\gamma^{q_1-1}} u(s) = \gamma$$

and hence, by the definitions of φ and u,

$$\lim_{s\to\gamma^{q_1-1}} \varphi(s) = \lim_{s\to\gamma^{q_1-1}} \frac{q_1\int_u^\gamma (\gamma-w)^{q_1/(q_2-q_1)}\, w^{q_1(q_1-1)/(q_2-q_1)+q_1-2}\mathrm{d}w}{(q_2-q_1)\,(\gamma-u(s))^{q_2/(q_2-q_1)}\,(u(s))^{q_2(q_1-1)/(q_2-q_1)}} = \frac{q_1}{q_2\gamma},$$

where in the last line we have used de l'Hospital rule. Finally, to establish the majorization $\varphi(s) \le C_{p,q_1,q_2}^{q_2} s^{q_2/q_1}$, one easily shows that the function $s \mapsto \varphi(s)/s^{q_2/q_1}$ is increasing and converges to $C_{p,q_1,q_2}^{q_2}$ as $s \to \infty$. Indeed, by differentiation, the monotonicity follows from the estimate $\varphi'(s)s \ge \frac{q_2}{q_1}\varphi(s)$ (which obviously holds), and the formula for the limit is a consequence of the definition of φ and the identity (6.15). □

Thus we have shown that the candidate B given by (6.13) is well-defined. We turn our attention to its properties.

Lemma 6.4.3. *We have*

$$B_x(x,y,T)\cdot\frac{c-x}{T} + B_y(x,y,T)\cdot\frac{c^{q_1}T^\alpha - y}{T} + B_T(x,y,T) \ge x^{q_2}T^\beta. \tag{6.17}$$

Proof. We will use certain formulas obtained in the previous section. First, note that we have the following analogue of (6.12):

$$(-1+c)B_x(1,s,1) + (-s+c^{q_1})B_y(1,s,1) + B_T(1,s,1) \ge 1. \tag{6.18}$$

To show this, observe that $B_y(1,s,1) > 0$ (since φ is an increasing function) and

$$B_x(1,s,1) = q_2\varphi(s) - q_1 s\varphi'(s) \le 0.$$

Hence the expression on the left of (6.18), considered as a function of $c \geq 0$, attains its minimum at $c = ((-B_x(1,s,1)/(q_1 B_y(1,s,1))))^{1/(q_1-1)}$. But this minimal value is equal to 1: this is equivalent to the differential equation (6.16), as we have already checked in the previous section. Hence (6.18) holds; replacing c with cx, we get

$$(-1+cx)B_x(1,s,1) + (-y + c^{q_1}x^{q_1})B_y(1,s,1) + B_T(1,s,1) \geq 1. \tag{6.19}$$

Put $s = x^{-q_1}yT^{-\alpha}$. It follows directly from the definition of B that

$$B_x(x,y,T) = x^{q_2-1}T^{\beta+1}B_x(1, x^{-q_1}yT^{-\alpha}, 1),$$

$$B_y(x,y,T) = x^{q_2-q_1}T^{\beta+1-\alpha}B_y(1, x^{-q_1}yT^{-\alpha}, 1)$$

and

$$B_T(x,y,T) = x^{q_2}T^{\beta}B_T(1, x^{-q_1}yT^{-\alpha}, 1).$$

Combining these identities with (6.19) yields the claim. Let us also record that if

$$c = \left(-\frac{B_x(x,y,T)}{q_1 T^{\alpha} B_y(x,y,T)}\right)^{1/(q_1-1)}, \tag{6.20}$$

then both sides if (6.17) are equal. This follows from the proof above. □

Now we will show that B satisfies the following main inequality.

Lemma 6.4.4. *For any S, $T > 0$, any x, y and $c \in [0,x]$ we have*

$$\begin{aligned} &B\left(\frac{Tx+Sc}{T+S}, (T+S)^{-1}\left(Ty + c^{q_1}\cdot\frac{(T+S)^{\alpha+1}-T^{\alpha+1}}{\alpha+1}\right), T+S\right) \\ &\geq B(x,y,T) + \left(\frac{Tx+Sc}{T+S}\right)^{q_2}\frac{(T+S)^{\beta+1}-T^{\beta+1}}{\beta+1}. \end{aligned} \tag{6.21}$$

Proof. Define auxiliary functions X, $Y : [T, S+T] \to [0,\infty)$ by the formulas

$$X(t) = \frac{Tx + (t-T)c}{t}, \qquad Y(t) = \frac{1}{t}\left(Ty + c^{q_1}\cdot\frac{t^{\alpha+1}-T^{\alpha+1}}{\alpha+1}\right).$$

We compute that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}B(X(t),Y(t),t) \\ &= B_x(X(t),Y(t),t)\cdot\frac{T(c-x)}{t^2} \\ &\quad + B_y(X(t),Y(t),t)\cdot\left(-\frac{Ty + (\alpha+1)^{-1}c^{q_1}(t^{\alpha+1}-T^{\alpha+1})}{t^2} + c^{q_1}t^{\alpha-1}\right) \\ &\quad + B_T(X(t),Y(t),t). \end{aligned} \tag{6.22}$$

However, by (6.17), the expression

$$B_x(X(t),Y(t),t)\cdot\frac{c-X(t)}{t} + B_y(X(t),Y(t),t)\cdot\frac{c^{q_1}t^{\alpha}-Y(t)}{t} + B_T(X(t),Y(t),t)$$

is not smaller than $X(t)^{q_2}t^\beta$. In addition, we have

$$\frac{c-X(t)}{t}=\frac{T(c-x)}{t^2}$$

and

$$\frac{c^{q_1}t^\alpha-Y(t)}{t}=-\frac{Ty+(\alpha+1)^{-1}c^{q_1}(t^{\alpha+1}-T^{\alpha+1})}{t^2}+c^{q_1}t^{\alpha-1},$$

so by (6.22), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}B(X(t),Y(t),t)\geq X(t)^{q_2}t^\beta\geq\left(\frac{Tx+Sc}{T+S}\right)^{q_2}t^\beta.$$

Here in the last line we have used the inequality $X(t)\geq(Tx+Sc)/(T+S)$, which is a direct consequence of the assumption $c\leq x$. This proves that

$$B(X(T+S),Y(T+S),T+S)\geq B(X(T),Y(T),T)+\int_T^{T+S}\left(\frac{Tx+Sc}{T+S}\right)^{q_2}t^\beta\mathrm{d}t,$$

and it remains to use the identities $(X(T),Y(T),T)=(x,y,T)$ and

$$\begin{aligned}&(X(T+S),Y(T+S),T+S)\\&=\left(\frac{Tx+Sc}{T+S},(T+S)^{-1}\left(Ty+c^{q_1}\cdot\frac{(T+S)^{\alpha+1}-T^{\alpha+1}}{\alpha+1}\right),T+S\right).\end{aligned}$$

The claim is established. □

Remark 6.4.5. Later on, we will need to know when both sides of (6.21) are almost equal. Of course, this is true if we let $S\to 0$, but actually the reasoning from the previous section allows to extract an „infinitesimal" version of this statement: namely, if $S\to 0$ and we let

$$c=\left(\frac{-B_x(X(T),Y(T),T)}{q_1T^\alpha B_y(X(T),Y(T),T)}\right)^{1/(q_1-1)},$$

then the difference of the left- and the right-hand side is of order $o(S)$. This follows from the proof of Lemma 6.4.3.

We are ready for the proof of the key estimate.

Proof of $\mathfrak{B}\leq B$. Let (Ω,μ) be an arbitrary measure space with $\mu(\Omega)=T$ and let $f:\Omega\to[0,\infty)$ be a measurable function belonging to the class $\mathcal{C}(x,y,T)$.

Step 1. Reductions. If equality holds in (6.3), then there is nothing to prove: we already know that $\mathfrak{B}=B$ at such point. So, suppose that we have strict inequality in (6.3); then by a simple approximation, we may assume that $\frac{1}{T}\int_0^T t^\alpha[f^*(t)]^{q_1}\mathrm{d}t<y$. Next, we replace f by an appropriate simple function. To this end, let N be a huge integer and let $g=g^N$ be the conditional expectation of f with respect to $\mathcal{T}^N$: that is, g is constant on each element Q of $\mathcal{T}^N$ and equal to $\frac{1}{\mu(Q)}\int_Q f\mathrm{d}\mu$ there. Clearly, g has the same average as f; furthermore, by Doob's martingale convergence theorem (and the assumption (iv) on the tree), we have $g^N\to f$ μ-almost everywhere and hence also $\|g^N\|_{L^{p,q_1}}\to\|f\|_{L^{p,q_1}}$; thus in particular $g^N\in\mathcal{C}(x,y,T)$ provided N is large enough. Furthermore, $\mathcal{M}g^N\leq\mathcal{M}f$ and $\mathcal{M}g^N\uparrow\mathcal{M}f$. Thus,

an upper estimate for $\|\mathcal{M}g^N\|_{L^{p,q_2}}$ will also imply the same bound for $\|\mathcal{M}f\|_{L^{p,q_2}}$. So, let N be fixed. Our final reduction is that we may assume that $g^N = \sum_{k=1}^M a_k \mathbb{1}_{A_k}$ for some pairwise disjoint sets A_k of the same measure $\mu(\Omega)/M = T/M$: this can be seen by modifying the generation $\mathcal{T}^N$ so that its elements have (almost) equal measures and discarding the generations $\mathcal{T}^{N+1}, \mathcal{T}^{N+2}, \ldots$. From now on, we will write g instead of g^N. We need to prove that

$$\int_0^T t^\beta [(\mathcal{M}g)^*(t)]^{q_2} \mathrm{d}t \le B(x, y, T). \tag{6.23}$$

This will be done by induction.

Step 2. Proof of (6.23) *for* $M = 1$. Then both g and $\mathcal{M}g$ are constant and equal to x on Ω and, in addition,

$$y \ge \frac{1}{T} \int_0^T t^\alpha (g^*(t))^{q_1} \mathrm{d}t = c^{q_1} T^\alpha / (\alpha + 1). \tag{6.24}$$

Note that $B \ge 0$, so (6.21) implies

$$\begin{aligned} &B\left(\frac{Tx + Sc}{T + S}, (T+S)^{-1}\left(Ty + c^{q_1} \cdot \frac{(T+S)^{\alpha+1} - T^{\alpha+1}}{\alpha + 1}\right), T + S\right) \\ &\ge \left(\frac{Tx + Sc}{T + S}\right)^{q_2} \frac{(T+S)^{\beta+1} - T^{\beta+1}}{\beta + 1}. \end{aligned}$$

So, letting $T \to 0$ we get, by the continuity of B,

$$B(c, c^{q_1} S^\alpha / (\alpha + 1), S) \ge c^{q_2} S^{\beta+1} / (\beta + 1).$$

Now replace c with x, S with T and use the inequality (6.24) together with the monotonicity of B with respect to the variable y to get

$$x^{q_2} T^{\beta+1} / (\beta + 1) \le B(x, y, T).$$

This is precisely (6.23) (for $M = 1$).

Step 3. Induction step. It follows from the weak-type inequality for $\mathcal{M}$ that there exists $k \in \{1, 2, \ldots, M\}$ such that $\mathcal{M}g = \frac{1}{\mu(\Omega)} \int_\Omega f \mathrm{d}\mu = x$ on A_k. Consider the space $\tilde{\Omega} = \Omega \setminus A_k$ equipped with the restricted measure μ and the tree $\tilde{\mathcal{T}}$ which consists of all sets of the form $A \setminus A_k$, $A \in \mathcal{T}$, provided that the difference is nonempty. Denote the associated maximal operator by $\tilde{\mathcal{M}}$.

Obviously, there is an index m such that $g = \min g$ on A_m. If $k \ne m$, then we replace g with

$$\tilde{g} = a_k \mathbb{1}_{A_m} + a_m \mathbb{1}_{A_k} + \sum_{r \notin \{k, m\}} a_r \mathbb{1}_{A_r},$$

i.e., we switch the values of g at the sets A_k and A_m. Since $\mu(A_k) = \mu(A_m)$, this modification does not change the nonincreasing rearrangement of g. On the other hand, note that on A_k we have

$$\mathcal{M}\tilde{g} \ge \frac{1}{\mu(\Omega)} \int_\Omega \tilde{g} \mathrm{d}\mu = \frac{1}{\mu(\Omega)} \int_\Omega g \mathrm{d}\mu = \mathcal{M}g.$$

Furthermore, we have

$$\tilde{\mathcal{M}}\tilde{g} \ge \mathcal{M}g \quad \text{on } \Omega \setminus A_k. \tag{6.25}$$

Indeed, suppose that $u \in \Omega \setminus A_k$ and let A be the element of $\mathcal{T}$ containing u such that

$$\mathcal{M}g = \frac{1}{\mu(A)} \int_A g \mathrm{d}\mu. \tag{6.26}$$

There may be many sets A with this property; if this is the case, we choose A which belongs to $\mathcal{T}^j$ with j as small as possible. If $A \cap A_k = \emptyset$, then $\tilde{g} \geq g$ on A and hence $\tilde{\mathcal{M}}\tilde{g}(u) \geq \frac{1}{\mu(A)} \int_A \tilde{g} \mathrm{d}\mu \geq \mathcal{M}g(u)$. On the other hand, if $A_k \subset A$, then $\frac{1}{\mu(A)} \int_A g \mathrm{d}\mu \leq x$, by the very definition of A_k, and hence we must actually have equality: see (6.26). Hence

$$\tilde{\mathcal{M}}\tilde{g}(u) \geq \frac{1}{\mu(\Omega \setminus A_k)} \int_{\Omega \setminus A_k} \tilde{g} \mathrm{d}\mu \geq \frac{1}{\mu(\Omega)} \int_\Omega g \mathrm{d}\mu = x = \mathcal{M}g(u)$$

and the desired majorization is established. Note that we may apply induction hypothesis to $\tilde{g}$, obtaining

$$\int_0^{T(M-1)/M} t^\beta \left[(\tilde{\mathcal{M}}\tilde{g})^*(t)\right]^{q_2} \mathrm{d}t \leq B\left(\tilde{x}, \tilde{y}, T(M-1)/M\right),$$

where

$$\tilde{x} = \frac{1}{\mu(\tilde{\Omega})} \int_{\tilde{\Omega}} \tilde{g} \mathrm{d}\mu, \qquad \tilde{y} = \frac{1}{\mu(\tilde{\Omega})} \int_0^{\mu(\tilde{\Omega})} t^\alpha (\tilde{g}^*(t))^{q_1} \mathrm{d}t.$$

Hence

$$\begin{aligned}
&\int_0^T t^\beta \left[(\mathcal{M}g)^*(t)\right]^{q_2} \mathrm{d}t \\
&= \int_0^{T(M-1)/M} t^\beta \left[(\mathcal{M}g)^*(t)\right]^{q_2} \mathrm{d}t + \int_{T(M-1)/M}^T t^\beta \left[(\mathcal{M}g)^*(t)\right]^{q_2} \mathrm{d}t \\
&\leq \int_0^{T(M-1)/M} t^\beta \left[(\tilde{\mathcal{M}}\tilde{g})^*(t)\right]^{q_2} \mathrm{d}t + x^{q_2} (\beta+1)^{-1} T^{\beta+1} \left(1 - \left(\frac{M-1}{M}\right)^{\beta+1}\right) \\
&\leq B(\tilde{x}, \tilde{y}, T(M-1)/M) + x^{q_2} (\beta+1)^{-1} T^{\beta+1} \left(1 - \left(\frac{M-1}{M}\right)^{\beta+1}\right).
\end{aligned} \tag{6.27}$$

In the light of (6.21) (applied with $x := \tilde{x}$, $y := \tilde{y}$, $S := T/M$, $T := T(M-1)/M$ and $c := a_m = \min g$), the latter expression is not bigger than $B(x, y, T)$. This completes the proof of (6.23) and the inequality $\mathfrak{B} \leq B$ follows. □

Proof of (6.2). Take any measure space (Ω, μ), any measurable function $f : \Omega \to \mathbb{R}$ and set

$$T = \mu(\Omega), \qquad x = \frac{1}{T} \int_0^T f^*(t) \mathrm{d}t, \qquad y = \frac{1}{T} \int_0^T t^\alpha (f^*(t))^{q_1} \mathrm{d}t.$$

Then by Lemma 6.4.2,

$$\begin{aligned}
\|\mathcal{M}f\|_{L^{p,q_2}(\Omega,\mu)}^{q_2} &= \int_0^T t^\beta [(\mathcal{M}f(t))^*]^{q_2} \mathrm{d}t \\
&\leq B(x, y, T) \leq x^{q_2} T^{\beta+1} \cdot C_{p,q_1,q_2}^{q_2} \left(\frac{y}{x^{q_1} T^\alpha}\right)^{q_2/q_1} = C_{p,q_1,q_2}^{q_2} \|f\|_{L^{p,q_1}}^{q_2}.
\end{aligned}$$

This completes the proof. □

6.5 The inequality $\mathfrak{B} \geq B$

It is convenient to split the reasoning into two parts.

6.5.1 On the search of the extremizer

First we will sketch some steps which lead to the discovery of extremal function. Let us emphasize here that the argumentation will be informal and brief, its purpose is to discover the formula for the nonincreasing rearrangement of the appropriate function. From the formal point of view, the reader might skip this subsection and proceed to the next one; however, we believe that the contents of this subsection is helpful as it explains the origins of the complicated formulas which will appear later. The idea is very simple: we will inspect carefully the above proof of the inequality $\mathfrak{B} \leq B$ and try to find a function g for which all the inequalities become (almost) equalities. Fix a huge integer N (it will be sent to infinity in a moment). First, we will consider a special measure space (Ω, μ): the interval $(0, 1]$ with the Lebesgue measure, and equip it with the tree $\mathcal{T}$, where for any $0 \leq n \leq N$, the family $\mathcal{T}^n$ contains the intervals $(0, (N-n)/N]$, $((N-n)/N, (N-n+1)/N]$, ..., $(1-1/N, 1]$. In what follows, we will assume that g is a nonincreasing function. Then $\mathcal{M}g$ also has this property, and hence the function $\tilde{g}$, appearing in the proof of $\mathfrak{B} \leq B$, coincides with g on its domain (therefore in (6.25) we will have equality). Thus the only inequalities which must be turned into (almost) equalities is the last passage in (6.27) and the fact that the final expression in (6.27) is not bigger than $B(x, y, T)$. Let us look at the second estimate: to see when both sides become almost equal, we go back to Remark 6.4.5. This statement suggests that on the interval $(m/N, (m+1)/N]$, g should equal

$$\left(\frac{-B_x(X(m/N), Y(m/N), m/N)}{q_1(m/N)^\alpha B_y(X(m/N), Y(m/N), m/N)}\right)^{1/(q_1-1)}$$
$$= X(m/N)\left(s - \frac{q_2\varphi(s)}{q_1\varphi'(s)}\right)^{1/(q_1-1)} = X(m/N)u(s),$$

where

$$X(m/N) = \frac{1}{m/N}\int_0^{m/N} g(t)\mathrm{d}t, \qquad Y(m/N) = \frac{1}{m/N}\int_0^{m/N} t^\alpha g(t)^{q_1}\mathrm{d}t$$

and $s = X^{-q_1}(m/N)Y(m/N)(m/N)^{-\alpha}$. Now let $N \to \infty$: we obtain that for any $t \in (0, 1]$, we should have

$$\xi(t) := \frac{g(t)}{\frac{1}{t}\int_0^t g(r)\mathrm{d}r} = u\left(\left(\frac{1}{t}\int_0^t g(r)\mathrm{d}r\right)^{-q_1}\left(\frac{1}{t}\int_0^t r^\alpha g(r)^{q_1}\mathrm{d}r\right)t^{-\alpha}\right). \tag{6.28}$$

Plug this into the definition of u: we get

$$\frac{q_2(q_1-1)}{q_2-q_1}\int_{\xi(t)}^{\gamma}(\gamma - w)^{q_1/(q_2-q_1)}\, w^{q_1(q_1-1)/(q_2-q_1)+q_1-2}\mathrm{d}w$$
$$= \left(\frac{Y(t)}{t^\alpha X(t)} - \xi(t)^{q_1-1}\right)(\gamma - \xi(t))^{q_1/(q_2-q_1)}\,\xi(t)^{q_1(q_1-1)/(q_2-q_1)}.$$

Now we differentiate both sides with respect to t. After some lengthy and tedious computations, we get the equivalent equality $I \cdot II = 0$, where

$$I = \xi'(t) + \frac{q_2 - q_1}{q_1} \cdot \frac{\xi(t)\,(\gamma - \xi(t))}{t}$$

and II is a certain complicated expression. Assuming that the term I vanishes, we obtain a simple differential equation for ξ, whose general solution is

$$\xi(t) = \gamma \left(1 + dt^{\gamma \frac{q_2 - q_1}{q_1}}\right)^{-1}.$$

Here d is an arbitrary real number. Having identified ξ, we easily find X and g: since $X'(t) = g(t)/t - X(t)/t$, (6.28) implies

$$\frac{\mathrm{d}}{\mathrm{d}t} X(t) = -\frac{X(t)}{t} + \frac{X(t)\xi(t)}{t}.$$

This is easily solved:

$$X(t) = ct^{-\alpha/(q_1-1)} \left(1 + dt^{\gamma \frac{q_2 - q_1}{q_1}}\right)^{-q_1/(q_2-q_1)},$$

(where c is an arbitrary number) and hence we obtain the following candidate for the extremizer:

$$g(t) = ct^{-\alpha/(q_1-1)} \left(1 + dt^{\gamma \frac{q_2 - q_1}{q_1}}\right)^{-q_2/(q_2-q_1)}. \tag{6.29}$$

Now, we can choose c and d so that

$$\int_0^1 g(t)\mathrm{d}t = x \quad \text{and} \quad \int_0^1 t^{\alpha}(g(t))^{q_1}\mathrm{d}t = y. \tag{6.30}$$

Indeed: we compute that

$$R(d) := \frac{\int_0^1 t^{\alpha}(g(t))^{q_1}\mathrm{d}t}{\left(\int_0^1 g(t)\mathrm{d}t\right)^{q_1}} = \frac{\int_0^1 t^{-\alpha/(q_1-1)} \left(1 + dt^{\gamma \frac{q_2 - q_1}{q_1}}\right)^{-q_2 q_1/(q_2-q_1)} \mathrm{d}t}{\left(\int_0^1 t^{-\alpha/(q_1-1)} \left(1 + dt^{\gamma \frac{q_2 - q_1}{q_1}}\right)^{-q_2/(q_2-q_1)} \mathrm{d}t\right)^{q_1}}$$

is a continuous function of $d \in [0, \infty)$ and

$$R(0) = \gamma, \qquad \lim_{d \to \infty} R(d) = \infty.$$

Therefore, there is d for which $R(d) = y/x^{q_1}$, and then we choose c so that $\int_0^1 g = x$.

6.5.2 A formal verification

Now we can prove rigorously the bound $\mathfrak{B}(x, y, T) \geq B(x, y, T)$. By homogeneity, we may assume that $T = 1$: that is, we assume that (Ω, μ) is a probability space. We repeat the above arguments in the reverse direction. Let g be given by (6.29), where c, d are chosen so that (6.30) holds. Then a careful inspection of the above arguments (or a direct calculation) shows that the function

$$G(t) := B\left(\frac{1}{t}\int_0^t g(r)\mathrm{d}r, \frac{1}{t}\int_0^t r^{\alpha}(g(r))^{q_1}\mathrm{d}r, t\right) + \int_t^1 r^{\beta}\left(\frac{1}{r}\int_0^r g(w)\mathrm{d}w\right)^{q_2}\mathrm{d}r$$

is constant. We have $G(1) = B(x, y, 1)$; let us check how G behaves in the neighborhood of 0. Note that

$$B\left(\frac{1}{t}\int_0^t g(r)\mathrm{d}r, \frac{1}{t}\int_0^t r^{\alpha}(g(r))^{q_1}\mathrm{d}r, t\right) = \left(\frac{1}{t}\int_0^t g(r)\mathrm{d}r\right)^{q_2} t^{\beta+1}\varphi(s),$$

where

$$s = \left(\frac{1}{t}\int_0^t r^{\alpha}g(r)^{q_1}\mathrm{d}r\right)\left(\frac{1}{t}\int_0^t g(r)\mathrm{d}r\right)^{-q_1} t^{-\alpha}.$$

Now if we let $t \to 0$, then $s \to \gamma^{q_1-1}$ as $t \to 0$, and the factor

$$\left(\frac{1}{t}\int_0^t g(r)\mathrm{d}r\right)^{q_2} t^{\beta+1}$$

converges to zero. Therefore

$$\lim_{t\to 0} G(t) = \int_0^1 r^{\beta}\left(\frac{1}{r}\int_0^r g(w)\mathrm{d}w\right)^{q_2}\mathrm{d}r$$

and hence we have proved that

$$\int_0^1 r^{\beta}\left(\frac{1}{r}\int_0^r g(w)\mathrm{d}w\right)^{q_2}\mathrm{d}r = B(x, y, 1).$$

Now we return to the general context. Let (Ω, μ) be a nonatomic probability space equipped with an arbitrary tree structure $\mathcal{T}$. The idea is very simple: we will construct a random variable f such that the distributions of f and g coincide, while the distributions of $\mathcal{M}f$ and the function $t \mapsto \frac{1}{t}\int_0^t g$ are arbitrarily close. Let us recall a notion which is frequently used in probability theory.

Definition 6.5.3. Suppose that f_1, f_2 are two measurable functions on some measure spaces (Ω_i, μ_i) with $\mu_i(\Omega_i) > 0$, $i = 1, 2$.

(i) Suppose that $\mu_1(\Omega_1) = \mu_2(\Omega_2)$. The measurable functions $f_1 : \Omega_1 \to \mathbb{R}$ and $f_2 : \Omega_2 \to \mathbb{R}$ are said to have the same distribution, if their nonincreasing rearrangements coincide: $f^* = g^*$.

(ii) Without the assumption $\mu_1(\Omega_1) = \mu_2(\Omega_2)$, the functions f_1 and f_2 are said to have the same *conditional* distribution, if their nonincreasing rearrangements, with respect to the normalized measures $\mu_1/\mu_1(\Omega_1)$, $\mu_2/\mu_2(\Omega_2)$, coincide.

We will freely use the fact that if (Ω_1, μ_1), (Ω_2, μ_2) are nonatomic measure spaces with $\mu_i(\Omega_i) > 0$, $i = 1, 2$, then for any measurable function f_1 on Ω_1, there exists a measurable function f_2 on Ω_2 with the same conditional distribution. See [19, p. 65] or Lemma 2.3 in [50].

Now we proceed to the construction. Let $N \geq 2$ be a fixed integer.

Step 1. First we use Lemma 2.5.2 with the sequence $a_j = (j - N)/N$, $j = 0, 1, ..., N - 1$ to obtain the families $(\mathcal{Q}^j)_{j=0}^{N-1}$ and the corresponding decreasing sequence of sets $(E_j)_{j=0}^{N-1}$. Additionally, denote $E_N = \emptyset$. This sequence corresponds to the sequence $[0, 1) \supset [0, 1 - N^{-1}) \supset [0, 1 - 2N^{-1}) \supset \ldots \supset [0, N^{-1}) \supset \emptyset$ which appears in the above analysis of g. By the lemma, for any $Q \in \mathcal{Q}^j$ we have $\mu(Q \cap E_{j+1}) = \mu(Q) \cdot (1 - (N - j)^{-1})$ and, for each $k \geq j$,

$$\frac{\mu(Q \cap E_k)}{\mu(Q)} = \frac{N - k}{N - j} = \frac{|[0, 1 - k/N)|}{|[0, 1 - j/N)|}. \tag{6.31}$$

Step 2. For each j the set $E_j \setminus E_{j+1}$ is the union of pairwise almost disjoint sets $Q \setminus E_{j+1}$, $Q \in \mathcal{Q}^j$. Let $f : \Omega \to \mathbb{R}$ be a function whose distribution is uniquely determined by the following requirement: for any j and any $Q \in \mathcal{Q}^j$, the function f restricted to $Q \setminus E_{j+1}$ and the function g restricted to $[1 - (j + 1)/N, 1 - j/N)$ have the same conditional distributions. Hence, if we fix j and sum over all $Q \in \mathcal{Q}^j$, we see that the distribution of f restricted to $E_j \setminus E_{j+1}$ and the distribution of g restricted to $[1-(j+1)/N, 1-j/N)$ coincide. Consequently, f and g have the same distribution and hence $f \in \mathcal{C}(x, y, 1)$.

It remains to handle the maximal function $\mathcal{M}f$, and this is the place where the fractal properties will be used. An important observation is that for any j and any $Q \in \mathcal{Q}^j$ the distribution of f restricted to Q and the distribution of g restricted to $[0, 1 - j/N)$ conditionally coincide; this follows from (6.31). So, in particular,

$$\frac{1}{\mu(Q)} \int_Q f \mathrm{d}\mu = \frac{N}{N - j} \int_{[0,1-j/N)} g(r) \mathrm{d}r.$$

Consequently, by the definition of the maximal function, we obtain

$$\mathcal{M}f \geq \frac{N}{N - j} \int_{[0,1-j/N)} g(r) \mathrm{d}r \qquad \text{on } Q,$$

and since $Q \in \mathcal{Q}^j$ was arbitrary, the above estimate holds on the whole E_j. By the very definition of the nonincreasing rearrangement, this yields

$$(\mathcal{M}f)^*(t) \geq \frac{1}{t + N^{-1}} \int_0^{t+N^{-1}} g(r) \mathrm{d}r,$$

since $\mu(E_j) = 1 - j/N$. Therefore,

$$\int_0^1 t^\beta \left[(\mathcal{M}f)^*(t)\right]^{q_2} \mathrm{d}t \geq \int_0^1 t^\beta \left(\frac{1}{t + N^{-1}} \int_0^{t+N^{-1}} g(r) \mathrm{d}r \right)^{q_2} \mathrm{d}r.$$

By Lebesgue's monotone convergence theorem, the expression on the right converges, as $N \to \infty$, to

$$\int_0^1 t^\beta \left(\frac{1}{t} \int_0^t g(r) \mathrm{d}r \right)^{q_2} \mathrm{d}t = B(x, y, 1).$$

This, by the very definition of $\mathfrak{B}$, shows that $\mathfrak{B}(x, y, 1) \geq B(x, y, 1)$ and completes the proof.

6.6 On an alternative proof of (6.2)

It was pointed out by the Referee of [59] that the Lorentz-norm estimate (6.2) can be established directly, without the use of the Bellman function method. The purpose of this section is to sketch briefly the main steps of the argumentation.

We start from an observation concerning the weak-type bound (1.1). Namely, it is well-known that the probabilistic version of this estimate is equivalent to

$$(\mathcal{M}f)^*(t) \leq \frac{1}{t}\int_0^t f^*(s)\mathrm{d}s \qquad \text{for all } t \in (0,1].$$

This inequality is extremely sharp: as we have seen above, for any nonincreasing and integrable function g and any probability space (Ω, μ) equipped with a tree $\mathcal{T}$, there exists a random variable f such that the distributions of f and g coincide, while the distributions of $\mathcal{M}f$ and $t \mapsto \frac{1}{t}\int_0^t g$ are as close as we wish. (For the probabilistic version of this sharpness, see Dubins and Gilat [14]). This observation allows to reduce the problem of finding the sharp constant in (6.2) to the question about the best constant in a modified Hardy's inequality

$$\left(\int_0^1 t^{1/p}\left(\frac{1}{t}\int_0^t g(s)\mathrm{d}s\right)^{q_1}\frac{\mathrm{d}t}{t}\right)^{1/q_1} \leq C_{p,q_1,q_2}\left(\int_0^1 t^{1/p}g^{q_2}(t)\frac{\mathrm{d}t}{t}\right)^{1/q_2} \tag{6.32}$$

tested against non-increasing functions g; the two constants coincide. The complete analysis of the latter estimate, for the full range of parameters p, q_1, q_2 can be found, for example, in the paper by Persson and Samko [63]. Interestingly, they studied the inequality for general (i.e., not necessarily monotone) functions and it turns out that the extremizers are nonincreasing if and only if $q_1 \leq q_2$. In other words, both approaches - exploiting the Bellman function and that above - allow to obtain the sharp version of the estimate (6.2) only in this limited range of q_1 and q_2.

Several comments are in order. The proof of the estimate (6.32) presented in [63] rests on a number of clever observations and substitutions which reduce the claim to the classical Bliss' inequality

$$\left(\int_0^\infty \left(\int_0^x g(t)\mathrm{d}t\right)^q x^{-q/p'-1}\mathrm{d}x\right)^{1/q} \leq c_{p,q}\left(\int_0^\infty g^p(x)\mathrm{d}x\right)^{1/p},$$

for $1 < p \leq q < \infty$. This estimate was established in [2] with the use of the calculus of variations (see also [51] for an alternative proof). We strongly believe that our approach to (6.2), which depends on the Bellman function method, is of independent interest and connections. One of its remarkable features is its flexibility, which possibly enables the unified treatment of Lorentz-norm estimates in related contexts.

Chapter 7

Sharp inequalities for the harmonic maximal operator

7.1 Motivation and the statement of results

In our considerations above, we have encountered two modifications of the dyadic maximal function: the geometric maximal operator $M^{\mathcal{G}}$ and the harmonic maximal operator $M^{\mathcal{H}}$. Recall that these objects are given by

$$M^{\mathcal{G}}f(x) = \sup\Big\{\exp(\langle \log|f|\rangle_Q) : Q \in \mathcal{D}(\mathbb{R}^d),\ x \in Q\Big\}$$

and

$$M^{\mathcal{H}}f(x) = \sup\Big\{\langle |f|^{-1}\rangle_Q^{-1} : Q \in \mathcal{D}(\mathbb{R}^d),\ x \in Q\Big\},$$

where $\mathcal{D}(\mathbb{R}^d)$ is the lattice of dyadic cubes in $\mathbb{R}^d$. The behavior of the triple M, $M^{\mathcal{G}}$ and $M^{\mathcal{H}}$ is similar to that of the arithmetic, geometric and the harmonic averages:

$$\frac{|x_1|+|x_2|+\ldots+|x_n|}{n}, \quad |x_1x_2\ldots x_n|^{1/n} \quad \text{and} \quad \left(\frac{|x_1|^{-1}+|x_2|^{-1}+\ldots+|x_n|^{-1}}{n}\right)^{-1},$$

where x_1, x_2, ..., x_n are arbitrary real numbers. In particular, we have the pointwise bound $Mf \geq M^{\mathcal{G}}f \geq M^{\mathcal{H}}f$ and hence all the estimates which hold true for M, are also valid for the remaining two maximal functions. However, the optimal constants are different in general, and one might ask for the explicit derivation of their values.

We will be interested in a certain two-weight estimate for the harmonic maximal operator. Our motivation comes from the following important question. As we have discussed in the introductory section, given $1 < p < \infty$, the class of the weights w for which the norm $\|M\|_{L^p(w)\to L^p(w)}$ is finite, is characterized by Muckenhoupt's condition A_p. One may ask about the counterpart of this result in the case when the L^p spaces are based on two different weights. That is, the problem is to characterize those pairs (u, v) of weights, for which the norm $\|M\|_{L^p(v)\to L^p(u)}$ is finite. Motivated by the form of the Muckenhoupt's result (which corresponds to the case $u = v$), it seems reasonable to expect that the condition reads

$$[u,v]_{A_p} := \sup\Big\{\langle u\rangle_Q \langle v^{1/(1-p)}\rangle_Q^{p-1} : Q \in \mathcal{D}(\mathbb{R}^d)\Big\} < \infty. \tag{7.1}$$

But this is not the case: the above condition is equivalent to $\|M\|_{L^p(v)\to L^{p,\infty}(u)} < \infty$, which is slightly weaker. The problem of the boundedness on L^p remained open for a few years and was finally solved by Sawyer [69]: the characterization can be expressed by the so-called testing condition

$$\int_Q \left(M(v^{-1/(p-1)}\chi_Q)\right)^p u\mathrm{d}x \leq c\int_Q v^{-1/(p-1)}\mathrm{d}x. \tag{7.2}$$

In other words, given $1 < p < \infty$ and a pair (u, v) of weights, the estimate $\int_{\mathbb{R}^d}(Mf)^p u\mathrm{d}x \leq C^p \int_{\mathbb{R}^d} f^p v\mathrm{d}x$ holds true for all functions f if and only if it holds (possibly, with a different constant) for special "test functions" of the form $f = v^{-1/(p-1)}\chi_Q$. Of course, the implication '$\Rightarrow$' is trivial, the main difficulty lies in the passage from the test functions to general f's. In fact, we have the following quantitative statement: if the estimate (7.2) holds true, then we have $\|M\|_{L^p(v)\to L^p(u)} \leq \frac{pc}{p-1}$ and the multiplicative factor $p/(p-1)$ is the best (see [43, 54]).

A similar result is true for the geometric maximal operator (cf. [56]) and it is natural to ask about the analogous statement for the harmonic maximal function. It turns out that in contrast to the context of the maximal operator, here the characterization *is* given by an appropriate version of (7.1). Namely, it follows from [11] that for any fixed $0 < p < \infty$, the operator $M^{\mathcal{H}}$ is bounded as an operator from $L^p(v)$ to $L^p(u)$ if and only if the pair (u, v) of weights satisfies

$$[u, v]_{A_{-p}} := \sup_{Q\in\mathcal{D}(\mathbb{R}^d)} \langle u\rangle_Q \langle v^{1/(p+1)}\rangle_Q^{-p-1} < \infty.$$

We will extend this result to the more general context of probability spaces with tree structures and provide the appropriate sharp quantitative statement. Recall that the associated dyadic-like harmonic maximal operator is given by

$$\mathcal{M}^{\mathcal{H}} f(x) = \sup\left\{\langle |f|^{-1}\rangle_{Q,\mu}^{-1} : Q\in\mathcal{T},\ x\in Q\right\}.$$

For a pair (u, v) of weights on Ω, the quantity $[u, v]_{A_{-p}}$ is given by the same formula as above, but the averages are calculated over elements of $\mathcal{T}$ and with respect to μ.

We will prove the following statement, basing on the paper [59].

Theorem 7.1.1. *If $0 < p < \infty$ and (u, v) is a pair of weights on Ω satisfying $[u, v]_{A_{-p}} < \infty$, then we have*

$$\|\mathcal{M}^{\mathcal{H}}\|_{L^p(v)\to L^p(u)} \leq \frac{(p+1)^{\frac{p+1}{p}}}{p}[u, v]_{A_{-p}}^{1/p}. \tag{7.3}$$

The estimate is sharp: for each individual triple $(\Omega, \mu, \mathcal{T})$, any $0 < p < \infty$, any $c > 0$ and any $\varepsilon > 0$ there is a pair (u, v) with $[u, v]_{A_{-p}} \leq c$ such that

$$\|\mathcal{M}^{\mathcal{H}}\|_{L^p(v)\to L^p(u)} > \frac{(p+1)^{\frac{p+1}{p}}}{p}c^{1/p} - \varepsilon.$$

Our approach, in a sense, follows that used by Sawyer in [69]. Namely, we will prove a sharp version of the testing condition: see Theorem 7.3.3. This will enable us to apply a certain change-of-measure argument, which will allow to deduce the result from its unweighted counterpart, established in the earlier Theorem 7.3.1. Fortunately, sharp constants obtained in both these theorems combine into the best constant in (7.3).

7.2 Bellman function method for harmonic maximal operators

The modification of the approach is based on the following relation between $\mathcal{M}^{\mathcal{H}}$ and the minimal operator $\mathfrak{M}$, mentioned in the introductory chapter:

$$\mathfrak{M}f(x) = \inf \left\{ \langle |f| \rangle_{Q,\mu} : Q \subset \mathcal{T},\ x \in Q \right\}.$$

The relation is $\mathcal{M}^{\mathcal{H}} f = (\mathfrak{M}(|f|^{-1}))^{-1}$. Hence it is enough to develop the version of the Bellman function method for $\mathfrak{M}$. As usual, we may restrict ourselves to nonnegative functions, and then the modification is straightforward: the definition of $\mathfrak{M}$ differs from that of $\mathcal{M}$ just by the use of infimum instead of supremum. Hence, it is enough to replace the maxima appearing in the technique by the minima. Formally, we proceed as follows. Suppose that $V : \{(x, y) : 0 \leq y \leq x\} \to \mathbb{R}$ is a fixed function and we aim at establishing the estimate

$$\int_\Omega V(f, \mathfrak{M}f) \mathrm{d}\mu \leq 0 \tag{7.4}$$

for all probability spaces with tree structures $\mathcal{T}$ and all $\mathcal{T}$-simple positive functions f. The validity of this inequality is equivalent to the existence of a function $B : \{(x, y) : 0 < y \leq x\} \to \mathbb{R}$, enjoying the following conditions:

1° (Initial condition). For any $x > 0$ we have $B(x, x) \leq 0$.

2° (Majorization). We have $B \geq V$ on $\{(x, y) : 0 < y \leq x\}$.

3° (Concavity). For any $0 < y \leq x$, any numbers $x_1, x_2 > 0$ and $\lambda_1, \lambda_2 \in (0, 1)$ such that $\lambda_1 + \lambda_2 = 1$ and $\lambda_1 x_1 + \lambda_2 x_2 = x$, we have

$$B(x, y) \geq \lambda_1 B(x_1, y \wedge x_1) + \lambda_2 B(x_2, y \wedge x_2),$$

where $a \wedge b = \min\{a, b\}$.

Furthermore, if (7.4) holds true, then the least special function satisfying the above requirements is given by

$$\mathfrak{B}(x, y) = \sup \int_\Omega V(f, y \wedge \mathfrak{M}f) \mathrm{d}\mu.$$

Here the supremum is taken over the class of all nonnegative, $\mathcal{T}$-simple functions f satisfying $\int_\Omega f \mathrm{d}\mu = x$, the tree $\mathcal{T}$ and the probability space (Ω, μ) are also allowed to vary. The proof of the above statements is the mere repetition of the arguments used in Chapter 2.

However, the above modification concerns the unweighted setting, while we are interested in the two-weight setting. As we will see in the next section, we will need to use (at least in one of our partial results) the version of the technique for weights satisfying $[u, v]_{A_{-p}} < \infty$. We have decided not to present the abstract formulation of the method here (in the spirit of the above considerations): it might look quite confusing. Instead, we have decided to illustrate it on a particular example (the estimate (7.6)), the reader will easily recognize the main features of the approach in the proof.

7.3 Proof of (7.3)

Throughout, p and c are given positive numbers and $(\Omega, \mu, \mathcal{T})$ is a fixed probability space with a tree structure. Our first result is the sharp unweighted inequality for the harmonic maximal operator.

Theorem 7.3.1. *We have* $\|\mathcal{M}^{\mathcal{H}}\|_{L^p(\Omega)\to L^p(\Omega)} = (p+1)/p$.

Proof. The inequality $\|\mathcal{M}^{\mathcal{H}}\|_{L^p(\Omega)\to L^p(\Omega)} \leq (p+1)/p$ will follow if we show that for any $\mathcal{T}$-simple and positive function f on Ω we have

$$\int_\Omega (\mathfrak{M}f)^{-p} \mathrm{d}\mu \leq \left(\frac{p+1}{p}\right)^p \int_\Omega f^{-p} \mathrm{d}\mu.$$

This estimate is of the form (7.4), with $V(x,y) = y^{-p} - \left(\frac{p+1}{p}\right)^p x^{-p}$. The associated Belmman function B is given by the formula

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

Some steps which lead to the discovery of this object are presented in Remark 7.3.2 below. It remains to verify the properties 1°, 2° and 3°. The initial condition is trivial. The majorization follows at once from the mean-value theorem for the convex function $t \mapsto t^{-p}$, $t > 0$ (simply multiply both sides of $B \geq V$ by y^p and substitute $t = x/y$). To check the concavity, we prove the stronger pointwise bound

$$B(x+h, y \wedge (x+h)) \leq B(x,y) + B_x(x,y)h. \tag{7.5}$$

This is equivalent to the estimate

$$(x+h)\big(y \wedge (x+h)\big)^{-p-1} - \frac{p+1}{p}\big(y \wedge (x+h)\big)^{-p} \leq (x+h)y^{-p-1} - \frac{p+1}{p}y^{-p}.$$

Now if $x + h \geq y$, then both sides are equal; if $x + h < y$, then multiplying both sides by y^p and substituting $t = (x+h)/y$ transforms the inequality into

$$\frac{p+1}{p} \leq t + \frac{t^{-p}}{p},$$

which again follows from the mean-value theorem for the convex function $t \mapsto t^{-p}$.

Thus, the application of the Bellman function method yields $\|\mathcal{M}^{\mathcal{H}}\|_{L^p(\Omega)\to L^p(\Omega)} \leq (p+1)/p$. The reverse bound is postponed: we will prove below that a more general two-weight estimate is also sharp. □

Remark 7.3.2. Let us sketch briefly how the function B above was constructed (and how the optimal constant $(p+1)/p$ was discovered). The function V is homogeneous of order $-p$, so it is natural to expect that B should also have this property; actually, it can be proved rigorously that the least function, $\mathfrak{B}$, must enjoy it. The second hint is indicated by the inequality (7.5): a little thought leads to the guess that for a fixed y, the function B should be linear with respect to x. These two observations lead to

$$B(x,y) = axy^{-p-1} + by^{-p}.$$

Now we verify the conditions 1° and 2° with $V(x,y) = y^{-p} - C^p x^{-p}$. The optimization with respect to a and b reveals that the least C for which these conditions can be guaranteed, is $(p+1)/p$. Furthermore, for such a C, we need to take $a = p^2/(p+1)$ and $b = -p$.

The next step of our analysis is the appropriate version of the testing condition. Consider the domain

$$D = D_{p,c} = \{(x,y,z) \in (0,\infty)^3 : x \le c y^{p+1}\}$$

(whose form captures the condition $[u,v]_{A_{-p}} \le c$) and let $B : D \to \mathbb{R}$ be defined by

$$B(x,y,z) = xz^{-p} + cpz.$$

This function is a key tool in the proof of the following statement.

Theorem 7.3.3. *Suppose that a pair (u,v) of weights on Ω satisfies $[u,v]_{A_{-p}} \le c$. Then for any $R \in \mathcal{T}$ we have*

$$\int_R \left(\mathcal{M}^{\mathcal{H}}\left(v^{\frac{-1}{p+1}}\chi_R\right)\right)^p u d\mu \le (p+1)[u,v]_{A_{-p}} \int_R v^{\frac{1}{p+1}} d\mu. \tag{7.6}$$

The constant $(p+1)[u,v]_{A_{-p}}$ is the best possible.

Proof. It is convenient to split the argumentation into three parts.

Step 1. Since $R \in \mathcal{T}$, there is an integer m such that $R \in \mathcal{T}^m$. Consider the functional sequences $(x_n)_{n\ge m}$, $(y_n)_{n\ge m}$ and $(z_n)_{n\ge m}$ given by

$$x_n = \sum_{Q\in\mathcal{T}^n} \langle u\rangle_{Q,\mu}\chi_Q, \qquad y_n = \sum_{Q\in\mathcal{T}^n} \left\langle v^{1/(p+1)}\right\rangle_{Q,\mu}\chi_Q, \qquad z_n = \min_{m\le k\le n} y_k.$$

Note that $(z_n)_{n\ge m}$ corresponds to the localized minimal operator applied to $v^{1/(p+1)}$. Obviously, for any $n \ge m$ and any $Q \in \mathcal{T}^n$, the functions x_n, y_n and z_n are constant on Q and we have

$$\int_Q x_{n+1}\mathrm{d}\mu = \mu(Q)x_n|_Q, \qquad \int_Q y_{n+1}\mathrm{d}\mu = \mu(Q)y_n|_Q. \tag{7.7}$$

In addition, the sequence $(z_n)_{n\ge m}$ is nonincreasing and, as we discussed above,

$$\lim_{n\to\infty} z_n = \mathcal{M}^{\mathcal{H}}(v^{-1/(p+1)}\mathbb{1}_R)^{-1} \tag{7.8}$$

almost everywhere. Finally, by the definition of $(x_n)_{n\ge m}$, $(y_n)_{n\ge m}$, $(z_n)_{n\ge m}$ and the assumption $[u,v]_{A_{-p}} \le c$, it follows at once that $(x_n, y_n, z_n) \in D_{p,c}$.

Step 2. Now we make use of the standard argument: we prove that the sequence $\left(\int_R B(x_n,y_n,z_n)\mathrm{d}\mu\right)_{n\ge m}$ is nondecreasing. It follows from (7.7) that if $n \ge m$ and Q is an element of $\mathcal{T}^n$, then

$$\int_Q B(x_n,y_n,z_n)\mathrm{d}\mu = \mu(Q)B(x_n,y_n,z_n)|_Q = \int_Q B(x_{n+1},y_{n+1},z_n)\mathrm{d}\mu, \tag{7.9}$$

since the dependence of B on x (and y) is linear. Now, observe that

$$B(x_{n+1},y_{n+1},z_n) \ge B(x_{n+1},y_{n+1},z_{n+1}). \tag{7.10}$$

Indeed, if $z_n = z_{n+1}$, there is nothing to prove; on the other hand, if $z_n > z_{n+1}$, then necessarily $y_{n+1} = z_{n+1} < z_n$ (since $z_{n+1} = \min\{z_n, y_{n+1}\}$) and

$$\begin{aligned} B(x_{n+1}, y_{n+1}, z_n) - B(x_{n+1}, y_{n+1}, z_{n+1}) &= \int_{z_{n+1}}^{z_n} B_z(x_{n+1}, y_{n+1}, s)\mathrm{d}s \\ &= p\int_{z_{n+1}}^{z_n} \big(- x_{n+1}s^{-p-1} + c\big)\mathrm{d}s \\ &\geq p\int_{z_{n+1}}^{z_n} \big(- x_{n+1}y_{n+1}^{-p-1} + c\big)\mathrm{d}s \geq 0, \end{aligned}$$

where the last estimate follows from the condition $[u,v]_{A_{-p}} \leq c$. This completes the proof of (7.10). Plugging this into (7.9) and summing over all $Q \in \mathcal{T}^n$ which are contained in R, we obtain the desired monotonicity of the sequence $\big(\int_R B(x_n, y_n, z_n)\mathrm{d}\mu\big)_{n\geq m}$.

Step 3. We are ready for the proof of (7.6). Note that

$$\int_R x_n z_n^{-p}\mathrm{d}\mu \leq \int_R B(x_n, y_n, z_n)\mathrm{d}\mu \leq \int_R B(x_m, y_m, z_m)\mathrm{d}\mu,$$

where in the second passage we have used the previous step. But $R \in \mathcal{T}^m$, so the functions x_m, y_m and z_m are constant on R; actually, $z_m = y_m$, by the very definition of z_m. Since $x_m \leq c y_m^{p+1}$ (which is due to $[u,v]_{A_{-p}} \leq c$), we get $B(x_m, y_m, z_m) = x_m y_m^{-p} + cpy_m \leq cy_m + cpy_m = c(p+1)y_m$ and hence

$$\int_R B(x_m, y_m, z_m)\mathrm{d}\mu \leq \mu(R)B(x_m, y_m, y_m)\big|_R \leq c(p+1)\int_R v^{1/(p+1)}\mathrm{d}\mu.$$

On the other hand, x_n is the conditional expectation of u on $\mathcal{T}^n$, so

$$\int_R x_n z_n^{-p}\mathrm{d}\mu = \int_R z_n^{-p} u\mathrm{d}\mu \xrightarrow{n\to\infty} \int_R (\mathcal{M}^{\mathcal{H}}(v^{-1/(p+1)}\mathbb{1}_R))^p u\mathrm{d}\mu,$$

by virtue of (7.8) and Lebesgue's monotone convergence theorem (recall that the sequence z_n^{-1} is nondecreasing). Putting all the above facts together, we get the desired estimate (7.6). The sharpness of this inequality will follow immediately from the sharpness of (7.3). See Remark 7.3.4 below. □

Proof of (7.3). Take an arbitrary pair (u,v) with $[u,v]_{A_{-p}} = c$ and an arbitrary integrable function $f > 0$. By a simple approximation argument, we may assume that $\varphi = f^{-1}$ is measurable with respect to a σ-algebra generated by some generation $\mathcal{T}^N$. Then we have $\mathcal{M}^{\mathcal{H}} f = \max_{Q\in\mathcal{T}^n, n\leq N}\langle\varphi\rangle_{Q,\mu}^{-1}\mathbb{1}_Q$ and hence for each $\omega \in \Omega$ there is an element $Q = Q(\omega)$ belonging to $\bigcup_{n\leq N}\mathcal{T}^n$ such that $\mathcal{M}^{\mathcal{H}} f(\omega) = \langle\varphi\rangle_{Q,\mu}^{-1}$. Such a Q may not be unique: in such a case we pick the set belonging to $\mathcal{T}^n$ with n as small as possible. For any $Q \in \mathcal{T}$, take $E(Q) = \{\omega \in Q : Q(\omega) = Q\}$. Note that $\{E(Q)\}_{Q\in\mathcal{T}}$ are pairwise disjoint and we have the linearization

$$\mathcal{M}^{\mathcal{H}} f = \sum_{Q\in\mathcal{T}} \langle\varphi\rangle_{Q,\mu}^{-1}\chi_{E(Q)}.$$

So, we may write

$$\int_\Omega (\mathcal{M}^{\mathcal{H}} f)^p u\mathrm{d}\mu = \sum_{Q\in\mathcal{T}} \langle\varphi\rangle_{Q,\mu}^{-p} u(E(Q)).$$

Now we perform a change-of-measure argument: we have

$$\langle\varphi\rangle_{Q,\mu} = \frac{1}{\mu(Q)}\int_Q \varphi v^{-1/(p+1)}\cdot v^{1/(p+1)}\mathrm{d}\mu = \langle \varphi v^{-1/(p+1)}\rangle_{Q,v^{1/(p+1)}}\langle v^{1/(p+1)}\rangle_{Q,\mu},$$

where $\langle\cdot\rangle_{Q,v^{1/(p+1)}}$ is the average over Q with respect to the measure $v^{1/(p+1)}d\mu$. Plugging this above, we obtain

$$\int_\Omega (\mathcal{M}^{\mathcal{H}}f)^p u\mathrm{d}\mu = \sum_{Q\in\mathcal{T}} \langle\varphi v^{-1/(p+1)}\rangle^{-p}_{Q,v^{1/(p+1)}}\langle v^{1/(p+1)}\rangle^{-p}_{Q,\mu}u(E(Q)). \tag{7.11}$$

The crucial observation is that there exists a collection $\{E'(Q)\}_{Q\in\mathcal{T}}$ of pairwise disjoint sets such that

$$\langle v^{1/(p+1)}\rangle^{-p}_{Q,\mu}u(E(Q)) = c(p+1)v^{1/(p+1)}(E'(Q)). \tag{7.12}$$

This follows from the inductive application of (7.6). For $n > N$ and $Q\in\mathcal{T}^n$ we have $E(Q) = \emptyset$: this follows directly from the definition of $E(Q)$ and the fact that φ is measurable with respect to $\sigma(\mathcal{T}^N)$. Consequently, we may also take $E'(Q) = \emptyset$ for such Q. To see the (backward) induction step, suppose that we defined $E'(Q)$ for all $Q\in\mathcal{T}^{n+1}$ and pick $R\in\mathcal{T}^n$. By (7.6), we have

$$\sum_{Q\in\mathcal{T},Q\subseteq R}\langle v^{1/(p+1)}\rangle^{-p}_{Q,\mu}u(E(Q)) \le \int_R\left(\mathcal{M}^{\mathcal{H}}\left(v^{\frac{-1}{p+1}}\chi_R\right)\right)^p u\mathrm{d}\mu \le (p+1)c\int_R v^{\frac{1}{p+1}}\mathrm{d}\mu,$$

and hence, by the inductive assumption,

$$\langle v^{1/(p+1)}\rangle^{-p}_{R,\mu}u(E(R)) \le (p+1)c\int_R v^{\frac{1}{p+1}}\mathrm{d}\mu - (p+1)c\int_{R'} v^{\frac{1}{p+1}}\mathrm{d}\mu = (p+1)c\int_{R\setminus R'} v^{\frac{1}{p+1}}\mathrm{d}\mu,$$

where $R' = \bigcup\{Q\in\mathcal{T} : Q\subsetneq R\}$. So, we may pick an appropriate $E'(R)\subset R\setminus R'$ and the inductive step is described. Now, plugging (7.12) into (7.11) yields

$$\begin{aligned}\int_\Omega (\mathcal{M}^{\mathcal{H}}f)^p u\mathrm{d}\mu &= c(p+1)\sum_{Q\in\mathcal{T}}\langle\varphi v^{-1/(p+1)}\rangle^{-p}_{Q,v^{1/(p+1)}}v^{1/(p+1)}(E'(Q))\\ &\le \|\mathcal{M}^{\mathcal{H}}_{v^{1/(p+1)}}(fv^{1/(p+1)})\|^p_{L^p(v^{1/(p+1)})},\end{aligned}$$

where $\mathcal{M}^{\mathcal{H}}_{v^{1/(p+1)}}$ is the harmonic maximal operator, but calculated with respect to the measure $v^{1/(p+1)}$. This measure need not be probabilistic, but we still can apply the estimate '$\le$' of Theorem 7.3.1, performing an appropriate normalization if necessary. As the result, we get

$$\|\mathcal{M}^{\mathcal{H}}f\|_{L^p(u)} \le c(p+1)\cdot\left(\frac{p+1}{p}\right)^p\int_\Omega (fv^{1/(p+1)})^p v^{1/(p+1)}\mathrm{d}\mu = c(p+1)\cdot\left(\frac{p+1}{p}\right)^p\|f\|^p_{L^p(v)},$$

which is the desired claim. □

Remark 7.3.4. The inequality (7.6) is sharp, for each individual probability space (Ω,μ) with a tree $\mathcal{T}$. Indeed, otherwise we would be able to improve the constant in the estimate (7.3); however, we will see in the next section that this is impossible.

7.4 Sharpness

Throughout this section, p and c are given positive parameters and $(\Omega, \mu, \mathcal{T})$ is a fixed probability space with a tree. We will show that for each $\varepsilon > 0$ there is a pair (u, v) of weights on Ω satisfying $[u, v]_{A_{-p}} \leq c$ and

$$\|\mathcal{M}^{\mathcal{H}}\|_{L^p(v) \to L^p(u)} > \frac{(p+1)^{\frac{p+1}{p}}}{p} c^{1/p} - \varepsilon.$$

It is convenient to split the reasoning into a few parts.

Step 1. Auxiliary geometrical facts and parameters. Pick $\tilde{c} \in (0, c)$ and $\delta, \eta > 0$. If δ is chosen small enough, then the line ℓ passing through the points $K = (1-\delta, \tilde{c}(1-\delta)^{p+1})$ and $L = (1, \tilde{c})$ lies below the curve $y = cx^{p+1}$. Fix such a δ and distinguish the point

$$M = \left(1 + \eta, \tilde{c}\left(1 + \eta \cdot \frac{1 - (1-\delta)^{p+1}}{\delta}\right)\right), \tag{7.13}$$

which is easily seen to lie on ℓ. See Figure 7.1 below. Note that if we let $\tilde{c} \to c$, then δ converges to 0.

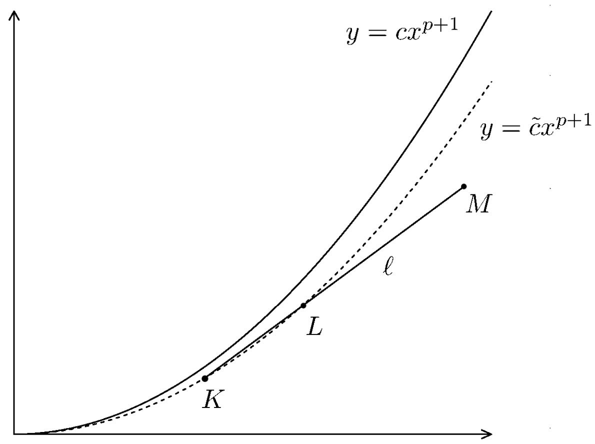

Figure 7.1: The crucial points and their geometric interpretation: $K = (1-\delta, \tilde{c}(1-\delta)^{p+1})$ and $L = (1, \tilde{c})$ lie on the curve $y = \tilde{c}x^{p+1}$, the point $M = \left(1 + \eta, \tilde{c}\left(1 + \eta \cdot \frac{1-(1-\delta)^{p+1}}{\delta}\right)\right)$ lies on the line ℓ.

Step 2. Construction. We use Lemma 2.5.2 with the sequence $a_n = \left(\frac{\eta}{\eta+\delta}\right)^n$ to obtain the decreasing sequence $(E_n)_{n \geq 0}$ of subsets of Ω. If Q is an atom of E_m, then for any $n \geq m$ we have

$$\mu(Q \cap E_n) = \mu(Q) \left(\frac{\eta}{\eta + \delta}\right)^{n-m}$$

and hence in particular,

$$\mu(Q \cap (E_n \setminus E_{n+1})) = \mu(Q) \left(\frac{\eta}{\eta + \delta}\right)^{n-m} \frac{\delta}{\eta + \delta}. \tag{7.14}$$

Recall the point M defined in (7.13) and denote its coordinates by M_x and M_y. Introduce the weights u, v on Ω by

$$u = M_y \sum_{n=0}^{\infty} (1-\delta)^{n(p+1)} \mathbb{1}_{E_n \setminus E_{n+1}}, \qquad v = M_x^{p+1} \sum_{n=0}^{\infty} (1-\delta)^{n(p+1)} \mathbb{1}_{E_n \setminus E_{n+1}}$$

and let $f : \Omega \to \mathbb{R}$ be given by

$$f = \sum_{n=0}^{\infty} (1+r\delta)^{-n} \mathbb{1}_{E_n \setminus E_{n+1}},$$

where r is an auxiliary parameter satisfying $-(p+1)/p < r < 0$.

Step 3. Verification of the condition $[u,v]_{A_{-p}} \le c$. By (7.14), if Q is an atom of E_m, then

$$\begin{aligned}\langle u \rangle_{Q,\mu} &= M_y \sum_{n=m}^{\infty} (1-\delta)^{n(p+1)} \left(\frac{\eta}{\eta+\delta} \right)^{n-m} \frac{\delta}{\eta+\delta} \\ &= \frac{M_y \delta}{\eta + \delta - (1-\delta)^{p+1}\eta} \cdot (1-\delta)^{m(p+1)} = \tilde{c}(1-\delta)^{m(p+1)}\end{aligned} \tag{7.15}$$

and

$$\langle v^{1/(p+1)} \rangle_{Q,\mu} = M_x \sum_{n=m}^{\infty} (1-\delta)^{n} \left(\frac{\eta}{\eta+\delta} \right)^{n-m} \frac{\delta}{\eta+\delta} = (1-\delta)^m. \tag{7.16}$$

Now, suppose that R is an arbitrary element of $\mathcal{T}$. Then there is an integer m such that $R \subseteq E_{m-1}$ and $R \nsubseteq E_m$. We have

$$\langle u \rangle_{R,\mu} = \frac{1}{\mu(R)} \int_{R \setminus E_m} u \mathrm{d}\mu + \frac{1}{\mu(R)} \int_{R \cap E_m} u \mathrm{d}\mu.$$

But $u = M_y(1-\delta)^{(m-1)(p+1)}$ on $R \setminus E_m$; furthermore, by (7.15), applied to each atom Q of E_m contained in R, we get

$$\int_{R \cap E_m} u \mathrm{d}\mu = \mu(R \cap E_m) \cdot \tilde{c}(1-\delta)^{m(p+1)}.$$

Therefore, setting $\kappa := \mu(R \cap E_m)/\mu(R) \in [0,1]$, we rewrite the preceding equality in the form

$$\langle u \rangle_{R,\mu} = (1-\delta)^{(m-1)(p+1)} \Big[\kappa K_y + (1-\kappa) M_y \Big].$$

(In analogy to the above notation, K_y stands the second coordinate of the point K; the number K_x, which will appear below, is the first coordinate of this point). A similar calculation shows that

$$\langle v^{1/(p+1)} \rangle_{R,\mu} = (1-\delta)^{m-1} \Big[\kappa K_x + (1-\kappa) M_x \Big]$$

and therefore

$$\langle u \rangle_{R,\mu} \langle v^{1/(p+1)} \rangle_R^{-p-1} = \Big[\kappa K_y + (1-\kappa) M_y \Big] \Big[\kappa K_x + (1-\kappa) M_x \Big]^{-p-1}.$$

This number does not exceed c. Indeed, as κ ranges from 0 to 1, the point $\kappa K + (1-\kappa)M$ runs over the line segment KM which lies below the curve $y = c|x|^{p+1}$ (see Step 1). Since R was arbitrary, the inequality $[u,v]_{A_{-p}} \leq c$ follows.

Step 4. Completion of the proof. In the same manner as above, one verifies that if Q is an atom of E_m, then

$$\langle f^{-1}\rangle_{Q,\mu} = \sum_{n=m}^{\infty}(1+r\delta)^n\left(\frac{\eta}{\eta+\delta}\right)^{n-m}\frac{\delta}{\eta+\delta} = \frac{(1+r\delta)^m}{1-r\eta}.$$

This immediately yields $\mathcal{M}^{\mathcal{H}}f \geq (1-r\eta)(1+r\delta)^{-m}$ on E_m and hence, by the definition of u, v and f, we obtain

$$(\mathcal{M}^{\mathcal{H}}f)^p u \geq \frac{(1-r\eta)^p M_y}{M_x^{p+1}} f^p v \qquad \text{on } E_m \setminus E_{m+1}.$$

The latter bound does not depend on m, so we can rewrite it uniformly as

$$(\mathcal{M}^{\mathcal{H}}f)^p u \geq \frac{(1-r\eta)^p M_y}{M_x^{p+1}} f^p v \qquad \text{on } \Omega.$$

Consequently, $(1-r\eta)^p M_y/M_x^{p+1}$ is the lower bound for the norm $\|\mathcal{M}^{\mathcal{H}}\|_{L^p(v)\to L^p(u)}$, as long as we have $\|f\|_{L^p(v)} < \infty$. Let us study the latter estimate. Note that

$$\int_\Omega f^p v \mathrm{d}\mu = (1+\eta)^{p+1}\sum_{n=0}^{\infty}(1+r\delta)^{-np}(1-\delta)^{n(p+1)}\left(\frac{\eta}{\eta+\delta}\right)^n\frac{\delta}{\eta+\delta}$$

and observe that the ratio of the above geometric series is equal to

$$(1+r\delta)^{-p}(1-\delta)^{p+1}\cdot\frac{\eta}{\eta+\delta} \leq 1 - pr\delta - (p+1)\delta + o(\delta).$$

Therefore for any r as above (i.e., satisfying $r > -(p+1)/p$), any $\eta > 0$ and $\tilde{c}$ sufficiently close to c (so that δ is close enough to 0) we have $\|f\|_{L^p(v)} < \infty$. Rewrite the constant $(1-r\eta)^p M_y/M_x^{p+1}$ explicitly as

$$\frac{(1-r\eta)^p M_y}{M_x^{p+1}} = \frac{(1-r\eta)^p\cdot\tilde{c}\left(1+\eta\delta^{-1}\big(1-(1-\delta)^{p+1}\big)\right)}{(1+\eta)^{p+1}}.$$

Now, we choose η to be very large, then δ is made small, and finally, we pick r close to $-(p+1)/p$. Then the above expression can be made as close to $c(p+1)\left(\frac{p+1}{p}\right)^p$ as we wish. This establishes the desired sharpness.

Bibliography

[1] R. Bellman, *Dynamic programming*, Princeton University Press, Princeton, NJ, 1957

[2] G. Bliss, *An integral inequality*, J. London. Math. Soc. **5** (1930), 40–46.

[3] S. M. Buckley, *Estimates for operator norms on weighted spaces and reverse Jensen inequalities*, Trans. Amer. Math. Soc. **340** (1993), 253–272.

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

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

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

[7] S. Chanillo and R. L. Wheeden, *Some weighted norm inequalities for the area integral*, Indiana Univ. Math. J. **36** (1987), pp. 277–294.

[8] R. Coifman and C. Fefferman, *Weighted norm inequalities for maximal functions and singular integrals*, Studia Math. **51** (1974), 241–250.

[9] D. Cruz-Uribe, SFO *The minimal operator and the geometric maximal operator in $\mathbb{R}^n$*, Studia Math. **144** (2001), no. 1, 1–37.

[10] D. Cruz-Uribe, SFO and C. J. Neugebauer, *The structure of the reverse Hölder classes*, Trans. Amer. Math. Soc. **347** (1995), 2941–2960.

[11] D. Cruz-Uribe, SFO, C.J. Neugebauer and V. Olesen, *Norm inequalities for the minimal and maximal operator, and differentiation of the integral*, Publ. Mat. **41** (1997), 577–604.

[12] D. Cruz-Uribe, SFO and C. Pérez, *Two-weight extrapolation via the maximal operator*, J. Funct. Anal. **174** (2000), pp. 1–17.

[13] J. L. Doob, *Stochastic processes*. Reprint of the 1953 original, Wiley Classics Library, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1990.

[14] L. E. Dubins and D. Gilat, *On the distribution of maxima of martingales*, Proc. Amer. Math. Soc. **68** (1978), 337–338.

[15] L. A. Duffee and K. Moen, *On the harmonic and geometric maximal operators*, Math. Inequal. Appl. **21** (2018), 265–286.

[16] C. Fefferman and E. M. Stein, *Some maximal inequalities*, Amer. J. Math. **93** (1971), 107–115.

[17] J. Garcia-Cuerva and J. L. Rubio de Francia, *Weighted norm inequalities and related topics*, volume 116 of North-Holland Mathematics Studies. North-Holland Publishing Co., Amsterdam 1985.

[18] D. Gilat, *The best bound in the LlogL inequality of Hardy and Littlewood and its martingale counterpart*, Proc. Amer. Math. Soc. **97** (1986), no. 3, pp. 429–436.

[19] L. Grafakos, *Classical Fourier Analysis*, Pearson Education, Inc. New Jersey, 2004.

[20] R. Hunt, B. Muckenhoupt and R. L. Wheeden, *Weighted norm inequalities for the conjugate function and Hilbert transform*, Trans. Amer. Math. Soc. **176** (1973), 227–251.

[21] T. P. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. Math. **175** (2012), 1473–1506.

[22] T. Hytönen and C. Pérez, *Sharp weighted bounds involving A_∞*, Anal. PDE **6** (2013), 777–818.

[23] Ł. Kamiński and A. Osękowski, *Best constants in some estimates for the harmonic maximal operator on the real line*, Colloq. Math., to appear.

[24] M. Lacey, *An elementary proof of the A_2 bound*, Israel J. Math. **217** (2017), 181–195.

[25] M. T. Lacey, K. Moen, C. Pérez and R. H. Torres, *Sharp weighted bounds for fractional integral operators*, J. Funct. Anal. **259** (2010), 1073–1097.

[26] A. K. Lerner, *On some weighted norm inequalities for Littlewood-Paley operators*, Illinois J. Math. **52** (2009), 653–666.

[27] A. K. Lerner, *On an estimate of Calderón-Zygmund operators by dyadic positive operators*, J. Anal. Math. **121** (2013), 141–161.

[28] A. K. Lerner and F. Nazarov, *Intuitive dyadic calculus: the basics*, available at arXiv:1508.05639.

[29] A. K. Lerner, S. Ombrosi and C. Pérez, *Sharp A_1 bounds for Calderón-Zygmund operators and the relationship with a problem of Muckenhoupt and Wheeden*, Int. Math. Res. Not. IMRN 6 (2008), Art. ID rnm161, 11 p.

[30] A. K. Lerner, S. Ombrosi and C. Pérez, *Weak type estimates for singular integrals related to a dual problem of Muckenhoupt-Wheeden*, J. Fourier Anal. Appl. **15** (2009), pp. 394–403.

[31] A. K. Lerner, S. Ombrosi and C. Pérez, *A_1 bounds for Calderón-Zygmund operators related to a problem of Muckenhoupt and Wheeden*, Math. Res. Lett. **16** (2009), pp. 149–156.

[32] T. Mei, *BMO is the intersection of two translates of dyadic BMO*, C. R. Acad. Sci. Paris, Ser. I **336** (2003), 1003–1006.

[33] A. D. Melas, *The Bellman functions of dyadic-like maximal operators and related inequalities*, Adv. Math. **192** (2005), 310–340.

[34] A. D. Melas, *Dyadic-like maximal operators on LlogL functions*, J. Funct. Anal. **257** (2009), 1631–1654.

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

[36] A. D. Melas and E. N. Nikolidakis, *On weak type inequalities for dyadic maximal functions*, J. Math. Anal. Appl. **348** (2008), 404–410.

[37] 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), 1571–1597.

[38] A. D. Melas and E. N. Nikolidakis, *Sharp Lorentz estimates for dyadic-like maximal operators and related Bellman functions*, J. Geom. Anal. **27** (2017), no. 4, 2644–2657.

[39] B. Muckenhoupt, *Weighted norm inequalities for the Hardy maximal function*, Trans. Amer. Math. Soc. **165** (1972), 207–226.

[40] F. Nazarov, A. Reznikov, V. Vasyunin and A. Volberg, *A Bellman function counterexample to the A_1 conjecture: the blow-up of weak norm estimates of weighted singular operators*, arXiv:1506.04710v1, 1–23.

[41] F. Nazarov, A. Reznikov, V. Vasyunin and A. Volberg, *On weak weighted estimates of the martingale transform and a dyadic shift*, Anal. PDE **11**, Number 8 (2018), 2089–2109.

[42] 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), 721–824.

[43] 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), 909–928.

[44] E. N. Nikolidakis, *Extremal sequences for the Bellman function of the dyadic maximal operator*, Rev. Mat. Iberoam. **33** (2017), no. 2, 489–508.

[45] A. Osękowski, *Sharp martingale and semimartingale inequalities*, Monografie Matematyczne **72** (2012), Birkhäuser, 462 pp.

[46] A. Osękowski, *Survey article: Bellman function method and sharp inequalities for martingales*, Rocky Mountain J. Math. **43** (2013), 1759–1823.

[47] A. Osękowski, *Weighted maximal inequalities for martingales*, Tohoku Math. J. **65** (2013), 75–91.

[48] A. Osękowski, *Maximal weak-type inequality for stochastic integrals*, Electronic Communications in Probability **19** (2014), Article 25, 1–13.

[49] A. Osękowski, *Sharp $L^{p,\infty} \to L^q$ estimates for the dyadic-like maximal operators*, J. Fourier Anal. Appl. **20** (2014), 911–933.

[50] A. Osękowski, *Weak-type inequalities for maximal operators acting on Lorentz spaces*, Calculus of Variations and PDEs, Banach Center Publ. **101** (2014), pp. 145–162.

[51] A. Osękowski, *A new approach to Hardy-type inequalities*, Arch. Math. **104** (2015), 165–176.

[52] A. Osękowski, *Sharp weak-type estimates for the dyadic-like maximal operators*, Taiwanese J. Math. **19** (2015), 1031–1050.

[53] A. Osękowski, *Sharp weighted logarithmic bound for maximal operators*, Archiv der Mathematik **107** (2016), 635–644.

[54] A. Osękowski, *Best constants in Muckenhoupt's inequality*, Ann. Acad. Sci. Fenn. Math. **42** (2017), 889–904.

[55] A. Osękowski, *Sharp weighted bounds for geometric maximal operators*, Glasgow Math. J. **59** (2017), 533–547.

[56] A. Osękowski, *Two-weight inequalities for geometric maximal operators*, Math. Ineq. Appl. **20** (2017), 1121-1138.

[57] A. Osękowski, *Sharp L^p-bounds for the martingale maximal function*, Tohoku Math. J. **70** (2018), 121–138.

[58] A. Osękowski, *A weighted weak-type bound for Haar multipliers*, Royal Soc. Edinburgh Proc A (Mathematics) **148A** (2018), 643–658.

[59] A. Osękowski and M. Rapicki, *Sharp Lorentz-norm estimates for dyadic-like maximal operators*, accepted for publication in Studia Mathematica.

[60] A. Osękowski and M. Rapicki, *A weighted maximal weak-type inequality*, Mathematika **67** (2021), 145–157.

[61] A. Osękowski and M. Rapicki, *Sharp weighted inequalities for harmonic maximal operators*, submitted.

[62] C. Pérez, *Weighted norm inequalities for singular integral operators*, J. Lond. Math. Soc. **49** (2) (1994), pp. 296–308.

[63] L.-E. Persson and S. G. Samko, *A note on the best constants in some Hardy inequalities*, J. Math. Ineq. **9** (2015), 437–447.

[64] M. Rapicki, *A weighted inequality for a dyadic-like maximal operator*, Anal. Math. **44** (2018), 577–585.

[65] M. Rapicki, *Fefferman-Stein inequalities for the dyadic-like maximal operators*, Arch. Math. **113** (2019), 81–93.

[66] M. C. Reguera, *On Muckenhoupt-Wheeden conjecture*, Adv. Math. **227** (2011), no. 4, pp. 1436–1450.

[67] M. C. Reguera and C. Thiele, *The Hilbert transform does not map $L^1(Mw)$ to $L^{1,\infty}(w)$*, Math. Res. Lett. **19** (2012), no. 1, pp. 1–7.

[68] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, 3rd edition, Springer Verlag, 1999.

[69] E. T. Sawyer, *A characterization of a two-weight norm inequality for maximal operators*, Studia Math. **75** (1982), 1–11.

[70] E. T. Sawyer, *A two weight weak type inequality for fractional integrals*, Trans. Amer. Math. Soc. **281** (1984), 339–345.

[71] E. T. Sawyer. *A characterization of two weight norm inequalities for fractional and Poisson integrals*, Trans. Amer. Math. Soc. **308** (1988), 533–545.

[72] 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), 585–588.

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

[74] E. M. Stein, *Singular integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, 1970.

[75] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of Princeton Mathematical Series. Princeton University Press, Princeton, NJ, 1993. With the assistance of T. S. Murphy, Monographs in Harmonic Analysis, III.

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

[77] 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", 195–238 (2009).