BEST CONSTANTS IN MUCKENHOUPT'S INEQUALITY

ADAM OSĘKOWSKI

Abstract. The paper identifies optimal constants in weighted L^p inequalities for the dyadic maximal function. The proof rests on Bellman function technique: the estimates are deduced from the existence of certain special functions enjoying appropriate size conditions and concavity.

1. Introduction

The purpose of this paper is to study a sharp version of a very classical estimate of harmonic analysis, the weighted L^p bound for the dyadic maximal operator. Let us start with introducing the necessary background and notation. Recall that the dyadic maximal operator $\mathcal{M}$ on $\mathbb{R}^n$ is an operator acting on locally integrable functions $\phi : \mathbb{R}^n \to \mathbb{R}$ by the formula

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

Here the dyadic cubes are those formed by the grids $2^{-N}\mathbb{Z}^n$, $N = 0, 1, 2, \ldots$, and $\langle f \rangle_Q$ stands for $\frac{1}{|Q|}\int_Q f \mathrm{d}x$, the average of f over Q ($|Q|$ denotes the Lebesgue measure of Q). This maximal operator plays a fundamental role in analysis and PDEs, and in many applications it is of interest to control it efficiently, i.e., to have optimal or at least good bounds for its norms. For instance, $\mathcal{M}$ satisfies the weak-type $(1,1)$ inequality

$$\lambda \big| \{ x \in \mathbb{R}^n : \mathcal{M}\phi(x) \geq \lambda \} \big| \leq \int_{\{\mathcal{M}\phi \geq \lambda\}} |\phi(u)| \mathrm{d}u, \qquad \phi \in L^1(\mathbb{R}^n), \tag{1.1}$$

which, after integration, yields the corresponding L^p estimate

$$||\mathcal{M}\phi||_{L^p(\mathbb{R}^n)} \leq \frac{p}{p-1} ||\phi||_{L^p(\mathbb{R}^n)}, \qquad 1 < p \leq \infty. \tag{1.2}$$

Both estimates are sharp: the constant 1 in (1.1) and the constant $p/(p-1)$ in (1.2) cannot be decreased. These two results have been successfully extended in numerous directions and applied in various contexts of harmonic analysis. See e.g. [4, 5, 6, 7, 8, 13, 14] and the monograph [3], consult also references therein.

The primary goal of the present paper is to establish a sharp weighted version of (1.2). In what follows, the word 'weight' will refer to a nonnegative, integrable function on the underlying measure space (here, $\mathbb{R}^n$ with Lebesgue's measure). The following statement is a consequence of the classical work of Muckenhoupt [9].

2010 *Mathematics Subject Classification.* Primary: 42B25. Secondary: 46E30, 60G42.

Key words and phrases. Maximal, dyadic, Bellman function, best constants.

Research supported by Narodowe Centrum Nauki (Poland), grant no. DEC-2014/14/E/ST1/00532.

Suppose that $1 < p < \infty$ is given and fixed, and let w be a weight on $\mathbb{R}^n$. Then $\mathcal{M}$ is bounded as an operator on the weighted space

$$L^p(w) = \left\{ f : \mathbb{R}^n \to \mathbb{R} : ||f||_{L^p(w)} = \left(\int_{\mathbb{R}^n} |f|^p w \mathrm{d}x \right)^{1/p} < \infty \right\}$$

if and only if w belongs to the dyadic A_p class, i.e.,

$$[w]_{A_p} := \sup \langle w \rangle_Q \langle w^{-1/(p-1)} \rangle_Q^{p-1} < \infty,$$

where the supremum is taken over all dyadic cubes in $\mathbb{R}^n$. This result is a starting point for many interesting further questions. For example, one can ask about the dependence of $||\mathcal{M}||_{L^p(w) \to L^p(w)}$ on the size of the characteristic $[w]_{A_p}$. More precisely, for a given $1 < p < \infty$, the problem is to find the least number $\alpha = \alpha(p)$ such that

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

for some C_p depending only on p. This problem was solved in the nineties by Buckley [1], who showed that the optimal exponent $\alpha(p)$ is equal to $1/(p-1)$.

The contribution of this paper is the sharp upper bound for $||\mathcal{M}||_{L^p(w) \to L^p(w)}$ both in terms of p and $[w]_{A_p}$. Actually, we will work in the more general context of probability spaces equipped with a tree-like structure [4]. Here is the precise definition.

Definition 1.1. Suppose that (X, μ) is a nonatomic probability space. A set $\mathcal{T}$ of measurable subsets of X will be called a tree if the following conditions are satisfied:

(i) $X \in \mathcal{T}$ and for every $Q \in \mathcal{T}$ we have $\mu(Q) > 0$.
(ii) For every $Q \in \mathcal{T}$ there is a finite subset $C(Q) \subset \mathcal{T}$ containing at least two elements such that
  (a) the elements of $C(Q)$ are pairwise disjoint subsets of Q,
  (b) $Q = \bigcup C(Q)$.
(iii) $\mathcal{T} = \bigcup_{m \geq 0} \mathcal{T}^m$, where $\mathcal{T}^0 = \{X\}$ and $T^{m+1} = \bigcup_{Q \in \mathcal{T}^m} C(Q)$.
(iv) We have $\lim_{m \to \infty} \sup_{Q \in \mathcal{T}^m} \mu(Q) = 0$.

An important example, which links this definition with the preceding considerations, is the cube $X = [0,1)^n$ endowed with Lebesgue measure and the tree of its dyadic subcubes. Any probability space equipped with a tree gives rise to the corresponding maximal operator $\mathcal{M}_\mathcal{T}$, acting on integrable functions $f : X \to \mathbb{R}$ by the formula

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

where this time $\langle \varphi \rangle_Q = \frac{1}{\mu(Q)} \int_Q \varphi \mathrm{d}\mu$ is the average of φ over Q with respect to the measure μ. In analogy to the dyadic setting described above, we say that a weight w on X satisfies Muckenhoupt's condition A_p (where $1 < p < \infty$ is a fixed parameter), if

$$[w]_{A_p} := \sup_{Q \in \mathcal{T}} \langle w \rangle_Q \langle w^{-1/(p-1)} \rangle_Q^{p-1} < \infty.$$

Furthermore, the weighted space $L^p(w)$ is given by

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

To formulate the main result of this paper, we need to introduce a certain special parameter d. Its geometric interpretation is explained on Figure 1 below. Let $c \geq 1$ and $1 < p < \infty$ be fixed. Then the line, tangent to the curve $wv^{p-1} = c$ at the point $(1, c^{1/(p-1)})$, intersects the curve $wv^{p-1} = 1$ at one point (if $c = 1$) or two points (if $c > 1$). Take the intersection point with larger w-coordinate, and denote this coordinate by $1 + d(p, c)$. Formally, $d = d(p, c)$ is the unique number in $[0, p - 1)$ satisfying the equation

$$c(1 + d)(p - 1 - d)^{p-1} = (p - 1)^{p-1}. \tag{1.3}$$

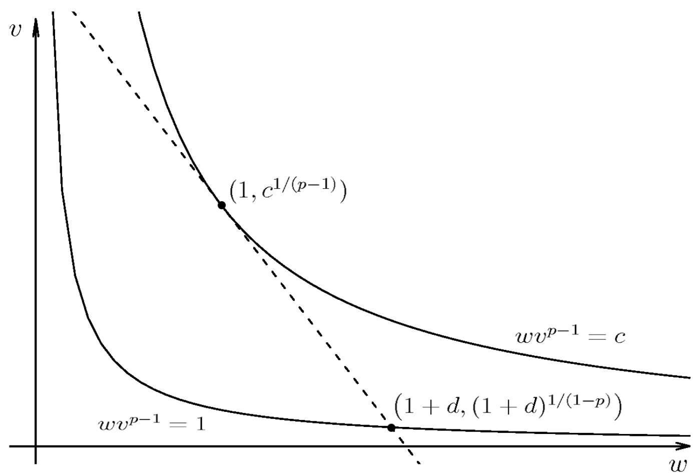

FIGURE 1. The geometric interpretation of the number $d = d(p, c)$.

We are ready to state the main result of the paper.

Theorem 1.2. *If $1 < p < \infty$ and w is an A_p weight on X, then we have the sharp bound*

$$||\mathcal{M}_\mathcal{T}||_{L^p(w)\to L^p(w)} \leq \frac{p}{p - 1 - d(p, [w]_{A_p})}. \tag{1.4}$$

Some remarks are in order. First, by sharpness we mean that for any $\varepsilon > 0$, any probability space (X, μ), any $1 < p < \infty$ and any $c \geq 1$ there is an A_p weight w with $[w]_{A_p} \leq c$ such that

$$||\mathcal{M}_\mathcal{T}||_{L^p(w)\to L^p(w)} > \frac{p}{p - 1 - d(p, [w]_{A_p})} - \varepsilon. \tag{1.5}$$

Thus, the above result is sharp also in the classical setting of $[0, 1)^n$ equipped with Lebesgue's measure and the tree of dyadic subcubes; by straightforward dilation and scaling, the result extends to the whole $\mathbb{R}^n$. Second, note that the above statement contains (1.2): indeed, setting $c = 1$ (which corresponds to the unweighted setting), we derive that $d(p, c) = 0$ and the optimal constant in (1.4) becomes equal to $p/(p - 1)$.

Our proof of (1.4) exploits the theory of two-weight inequalities. It follows from the results of Sawyer in [16] that if w, v are two weights on $\mathbb{R}^n$, then the (dyadic) maximal operator $\mathcal{M}$ is bounded as an operator from $L^p(v)$ to $L^p(w)$ if and only if the weights satisfy the so-called testing condition

$$\int_Q \big(\mathcal{M}(v^{-1/(p-1)}\chi_Q)\big)^p w\mathrm{d}x \leq C\int_Q v^{-1/(p-1)}\mathrm{d}x$$

for all dyadic cubes Q, where C depends only on p, w and v. We will study a sharp version of the testing condition for $w=v$, in the above context of probability spaces. Then we will combine this estimate with the weighted version of Carleson embedding theorem (cf. [11], [21]) and obtain the desired bound (1.4). Both these steps (i.e., sharp testing condition and Carleson imbedding theorem) will be established with the use of the so-called Bellman function method. The technique reduces the problem of proving a given inequality to the search for a certain special function, enjoying appropriate size conditions and concavity. The method originates from the theory of optimal stochastic control, and it has been studied intensively during the last thirty years. Its connection to the problems of martingale theory was firstly observed by Burkholder [2], who used it to identify the unconditional constant of the Haar system and related estimates for martingale transforms. This direction of research was further explored by Burkholder, his PhD students and other mathematicians (see [12] for the overview). In the nineties, Nazarov, Treil and Volberg (cf. [10], [11]) described the method from a wider perspective which allowed them to apply it to various problems of harmonic analysis. Since then, the technique has proved to be extremely efficient in various contexts; consult e.g. [15], [17], [18], [19], [20] and the references therein.

The rest of this paper is organized as follows. In the next section we provide the construction of an example showing that the constant in (1.4) cannot be smaller than $p/(p-1-d(p,[w]_{A_p}))$. The final part of the paper is devoted to the proof of (1.4).

2. An example

Throughout this section, $c\geq 1$ and $1<p<\infty$ are fixed parameters, and our goal here is to prove that for each $\varepsilon>0$ there is an A_p weight w with $[w]_{A_p}\leq c$ such that (1.5) holds true. We may exclude the trivial case $c=1$ from our considerations: the resulting constant in (1.5) is $p/(p-1)$, which is optimal in the unweighted setting. Thus, from now on, we assume that c is strictly bigger than 1.

It is convenient to split the reasoning into a few parts.

Step 1. Auxiliary geometrical facts and parameters. Pick $\tilde{c}\in(1,c)$. There are two lines passing through the point $K=(1,\tilde{c}^{1/(p-1)})$ which are tangent to the curve $wv^{p-1}=c$; pick the line ℓ which has smaller slope (equivalently: the w-coordinate of the tangency point is smaller than 1). This line intersects the curve $wv^{p-1}=1$ at two points: pick the point L with bigger w-coordinate and denote this coordinate by $1+d(\tilde{c})$. Furthermore, the line ℓ intersects the curve $wv^{p-1}=\tilde{c}$ at two points: one of them is K, while the second, denoted by M, is of the form $\big(1-\delta,(\tilde{c}(1-\delta))^{1/(1-p)}\big)$. See Figure 2 below.

Let us record here two important facts. First, the points K, L, M are colinear: some simple algebra allows to transform this observation into the equality

$$(\tilde{c}(1+d(\tilde{c})))^{1/(p-1)}\big(d(\tilde{c})+\delta-d(\tilde{c})(1-\delta)^{1/(1-p)}\big)=\delta, \tag{2.1}$$

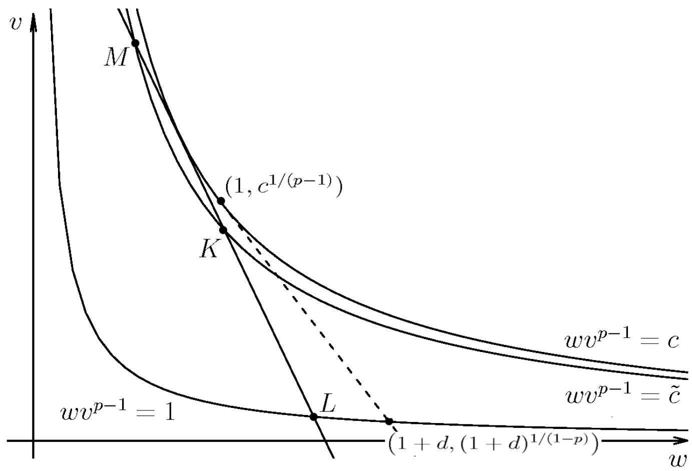

FIGURE 2. The crucial parameters and their geometric interpretation: $K = (1, \tilde{c}^{1/(p-1)})$, $L = \big(1 + d(\tilde{c}), (1 + d(\tilde{c}))^{1/(1-p)}\big)$ and $M = \big(1 - \delta, (\tilde{c}(1 - \delta))^{1/(1-p)}\big)$.

which will be useful later. Second, it follows immediately from the geometric interpretation of $d(p, c)$ and $d(\tilde{c})$ that

$$d(\tilde{c}) < d(p, c) < p - 1, \tag{2.2}$$

and $d(\tilde{c})$ can be made arbitrarily close to $d(p, c)$ by picking $\tilde{c}$ sufficiently close to c.

Finally, we introduce a parameter r, which is assumed to be a negative number satisfying $r > -1/p - 1/(pd(p, c))$. By the left estimate in (2.2), we see that for all $\tilde{c}$ we have $r > -1/p - 1/(pd(\tilde{c}))$, which combined with the right bound in (2.2) implies

$$1 + rd(\tilde{c}) > 0. \tag{2.3}$$

Step 2. Construction. Now, recall the following technical fact, which can be found in [4].

Lemma 2.1. *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*

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

We use this fact inductively, to construct an appropriate family $A_0 \supset A_1 \supset A_2 \supset \ldots$ of sets. Namely, we start with $A_0 = X$. Suppose we have successfully constructed A_n, which is a union of pairwise almost disjoint elements of $\mathcal{T}$, called the *atoms* of A_n (this condition is satisfied for $n = 0$: we have $A_0 = X \in \mathcal{T}$). Then, for each atom Q of A_n, we apply the above lemma with $\beta = d(\tilde{c})/(d(\tilde{c}) + \delta)$ and get a subfamily $F(Q)$. Put $A_{n+1} = \bigcup_Q \bigcup_{Q' \in F(Q)} Q'$, the first union taken over all atoms Q of A_n. Directly from the definition, this set is a union of the family $\{F(Q) : Q \text{ an atom of } A_n\}$, which consists of pairwise disjoint elements of $\mathcal{T}$. We call these elements the atoms of A_{n+1} and conclude the description of the induction step.

As an immediate consequence of the above construction, we see that if Q is an atom of A_m, then for any $n \geq m$ we have

$$\mu(Q \cap A_n) = \mu(Q) \left(\frac{d(\tilde{c})}{d(\tilde{c}) + \delta} \right)^{n-m}$$

and hence

$$\mu(Q \cap (A_n \setminus A_{n+1})) = \mu(Q) \left(\frac{d(\tilde{c})}{d(\tilde{c}) + \delta} \right)^{n-m} \frac{\delta}{d(\tilde{c}) + \delta}. \tag{2.4}$$

Now, introduce the weight w on X by the formula

$$w = \sum_{n=0}^{\infty} \chi_{A_n \setminus A_{n+1}} (1 + d(\tilde{c}))(1 - \delta)^n$$

and let $f : X \to \mathbb{R}$ be given by

$$f = \sum_{n=0}^{\infty} \chi_{A_n \setminus A_{n+1}} (1 + rd(\tilde{c}))(1 - r\delta)^n,$$

where r is the number fixed at the previous step.

Step 3. Verification of Muckenhoupt's condition. First we will check that w is an A_p weight satisfying $[w]_{A_p} \leq c$. To this end, we use (2.4) to obtain that for each atom Q of A_m we have

$$\langle w \rangle_Q = \sum_{n=m}^{\infty} \left(\frac{d(\tilde{c})}{d(\tilde{c}) + \delta} \right)^{n-m} (1 - \delta)^n (1 + d(\tilde{c})) \cdot \frac{\delta}{d(\tilde{c}) + \delta} = (1 - \delta)^m \tag{2.5}$$

and

$$\begin{aligned} \langle w^{-1/(p-1)} \rangle_Q &= \sum_{n=m}^{\infty} \left(\frac{d(\tilde{c})}{d(\tilde{c}) + \delta} \right)^{n-m} (1 - \delta)^{n/(1-p)} (1 + d(\tilde{c}))^{1/(1-p)} \cdot \frac{\delta}{d(\tilde{c}) + \delta} \\ &= \frac{(1 + d(\tilde{c}))^{1/(1-p)} \delta}{d(\tilde{c}) + \delta} (1 - \delta)^{m/(1-p)} \cdot \left(1 - \frac{d(\tilde{c})}{d(\tilde{c}) + \delta} (1 - \delta)^{1/(1-p)} \right)^{-1} \\ &= c^{1/(p-1)} (1 - \delta)^{m/(1-p)}, \end{aligned}$$

where in the last passage we have exploited (2.1). Suppose that R is an arbitrary element of $\mathcal{T}$. Then there is an integer m such that $R \subseteq A_{m-1}$ and $R \nsubseteq A_m$. We have

$$\begin{aligned} \langle w \rangle_R &= \frac{1}{\mu(R)} \int_{R \setminus A_m} w \mathrm{d}\mu + \frac{1}{\mu(R)} \int_{R \cap A_m} w \mathrm{d}\mu \\ &= \frac{1}{\mu(R)} \int_{R \setminus A_m} (1 + d(\tilde{c}))(1 - \delta)^{m-1} \mathrm{d}\mu + \frac{1}{\mu(R)} \int_{R \cap A_m} w \mathrm{d}\mu. \end{aligned}$$

By (2.5), applied to each atom Q of A_m contained in R, we get

$$\int_{R \cap A_m} w \mathrm{d}\mu = \mu(R \cap A_m)(1 - \delta)^m$$

and hence, setting $\eta := \mu(R \cap A_m)/\mu(R)$, we rewrite the preceding equality in the form

$$\langle w \rangle_R = (1 - \eta)(1 + d(\tilde{c}))(1 - \delta)^{m-1} + \eta(1 - \delta)^m.$$

A similar calculation shows that

$$\begin{aligned}&\langle w^{-1/(p-1)}\rangle_R\\&=(1-\eta)(1+d(\tilde c))^{1/(1-p)}(1-\delta)^{(m-1)/(1-p)}+\eta c^{1/(p-1)}(1-\delta)^{m/(1-p)}\end{aligned}$$

and therefore

$$\begin{aligned}&\langle w\rangle_R\langle w^{-1/(p-1)}\rangle_R^{p-1}\\&=\Big(\eta(1-\delta)+(1-\eta)(1+d(\tilde c))\Big)\Big(\eta(1-\delta)^{1/(1-p)}+(1-\eta)(1+d(\tilde c))^{1/(1-p)}\Big)^{p-1}.\end{aligned}$$

This number does not exceed c. To see this, rewrite the right-hand side in the form

$$(\eta M_w+(1-\eta)L_w)(\eta M_v+(1-\eta)L_v)^{p-1},$$

where M_w, M_v and L_w, L_v are the coordinates of the points M and L (see Figure 2). As η ranges from 0 to 1, the point $\eta M+(1-\eta)L$ runs over the line segment ML which is entirely contained in $\{(w,v) : wv^{p-1}\leq c\}$. Since R was arbitrary, we obtain the desired A_p condition: $[w]_{A_p}\leq c$.

Step 4. Completion of the proof. In the same manner as above, one verifies that if Q is an atom of $\mathcal{A}_m$, then

$$\langle f\rangle_Q=\sum_{n=m}^{\infty}\left(\frac{d(\tilde c)}{d(\tilde c)+\delta}\right)^{n-m}(1-r\delta)^n(1+rd(\tilde c))\cdot\frac{\delta}{d(\tilde c)+\delta}=(1-r\delta)^m$$

(the ratio of the geometric series, equal to $d(\tilde c)(1-r\delta)/(d(\tilde c)+\delta)$), is less than 1: this is equivalent to (2.3)). Consequently, we see that $\mathcal{M}_{\mathcal{T}}f\geq(1-r\delta)^m$ on A_m and hence, by the definition of f, we obtain $\mathcal{M}_{\mathcal{T}}f\geq(1+rd(\tilde c))^{-1}f$ on $A_m\setminus A_{m+1}$. The latter bound does not depend on m, so we can rewrite it uniformly as

$$\mathcal{M}_{\mathcal{T}}f\geq(1+rd(\tilde c))^{-1}f\qquad\text{on } X.$$

Now if we choose r sufficiently close to $-1/p-1/(pd(p,c))$ and then $\tilde c$ sufficiently close to c, then the number $(1+rd(\tilde c))^{-1}$ can be made arbitrarily close to $p/(p-1-d(p,c))$, the constant in (1.4). Thus, it is enough to show that for such choices, the function f belongs to $L^p(w)$. To this end, we compute that

$$||f||^p_{L^p(w)}=\sum_{n=0}^{\infty}(1+d(\tilde c))(1-\delta)^n(1+rd(\tilde c))^p(1-r\delta)^{np}\left(\frac{d(\tilde c)}{d(\tilde c)+\delta}\right)^n\frac{\delta}{d(\tilde c)+\delta}$$

and the ratio of this geometric series is equal to $(1-\delta)(1-r\delta)^pd(\tilde c)/(d(\tilde c)+\delta)$. Now recall that we have taken r close to (but larger than) $-1/p-1/(pd(p,c))$; hence $1+pr+1/d(p,c)>0$. If we make $\tilde c$ sufficiently close to c (then δ approaches 0: see Figure 2), we see that the ratio is

$$1-\delta\left(1+pr+\frac{1}{d(p,c)}\right)+o(\delta)<1.$$

This establishes the desired sharpness.

3. Proof of (1.4)

Throughout this section, $p \in (1,\infty)$ is given and fixed. For any $c \geq 1$, introduce the domain

$$\mathcal{D} = \mathcal{D}_{p,c} = \{(\mathtt{u},\mathtt{v},\mathtt{w}) \in (0,\infty)^3 : 1 \leq \mathtt{w}\mathtt{v}^{p-1} \leq c\}$$

and let $B : \mathcal{D}_{p,c} \to \mathbb{R}$ be the function given by the formula

$$B(\mathtt{u},\mathtt{w},\mathtt{v}) = \left(1+\frac{1}{d(p,c)}\right)\mathtt{u}^p\mathtt{w} + \frac{p-1}{d(p,c)}\big(c(1+d(p,c))\big)^{p/(p-1)}\mathtt{v} - \frac{p\mathtt{u}}{d(p,c)}c(1+d(p,c)).$$

We will prove that this object has the following properties.

Lemma 3.1. *(i) If* $\mathtt{u}^{p-1}\mathtt{w} \leq c$, *then*

$$\frac{\partial B}{\partial \mathtt{u}}(\mathtt{u},\mathtt{w},\mathtt{v}) \leq 0. \tag{3.1}$$

(ii) For any positive $\mathtt{w},\mathtt{v}$ *satisfying* $\mathtt{w}\mathtt{v}^{p-1} \leq c$ *we have*

$$B(\mathtt{v},\mathtt{w},\mathtt{v}) \leq (c(1+d(p,c)))^{p/(p-1)}\mathtt{v}. \tag{3.2}$$

(iii) We have

$$B(\mathtt{u},\mathtt{w},\mathtt{v}) \geq \mathtt{u}^p\mathtt{w}. \tag{3.3}$$

Proof. (i) We easily compute that

$$\frac{\partial B}{\partial \mathtt{u}}(\mathtt{u},\mathtt{w},\mathtt{v}) = p\left(1+\frac{1}{d(p,c)}\right)\left(\mathtt{u}^{p-1}\mathtt{w} - c\right) \leq 0.$$

(ii) Plugging $\mathtt{u} = \mathtt{v}$ in the formula for B gives

$$\begin{aligned} &B(\mathtt{v},\mathtt{w},\mathtt{v}) \\ &= \left(1+\frac{1}{d(p,c)}\right)\mathtt{v}^p\mathtt{w} + \frac{p-1}{d(p,c)}\big(c(1+d(p,c))\big)^{p/(p-1)}\mathtt{v} - \frac{p\mathtt{v}}{d(p,c)}c(1+d(p,c)) \\ &\leq c\left(1+\frac{1}{d(p,c)}\right)\mathtt{v} + \frac{p-1}{d(p,c)}\big(c(1+d(p,c))\big)^{p/(p-1)}\mathtt{v} - \frac{p\mathtt{v}}{d(p,c)}c(1+d(p,c)) \\ &= c(p-1)\left(1+\frac{1}{d(p,c)}\right)\left[\big(c(1+d(p,c))\big)^{1/(p-1)} - 1\right]\mathtt{v}. \end{aligned}$$

It remains to apply (1.3): we have $\big(c(1+d(p,c))\big)^{1/(p-1)} = (p-1)/(p-1-d(p,c))$, so

$$\begin{aligned} c(p-1)&\left(1+\frac{1}{d(p,c)}\right)\left[\big(c(1+d(p,c))\big)^{1/(p-1)} - 1\right] \\ &= c(p-1)\left(1+\frac{1}{d(p,c)}\right)\frac{d(p,c)}{p-1-d(p,c)} \\ &= \frac{c(1+d(p,c))(p-1)}{p-1-d(p,c)} \\ &= (c(1+d(p,c)))^{p/(p-1)}. \end{aligned}$$

(iii) The majorization is equivalent to

$$\mathtt{u}^p\mathtt{w} + (p-1)\big(c(1+d(p,c))\big)^{p/(p-1)}\mathtt{v} - pc(1+d(p,c))\mathtt{u} \geq 0.$$

Let $\mathtt{u} > 0$ be fixed. Since $\mathtt{w}\mathtt{v}^{p-1} \geq 1$, the left-hand side above is not smaller than

$$G(\mathtt{v}) := \mathtt{u}^p\mathtt{v}^{1-p} + (p-1)\big(c(1+d(p,c))\big)^{p/(p-1)}\mathtt{v} - pc(1+d(p,c))\mathtt{u}.$$

We compute that $G'(\mathtt{v}) = (p-1)\big(-(\mathtt{u}/\mathtt{v})^p + (c(1+d(p,c)))^{p/(p-1)}\big)$ and hence G attains its minimum at the point $\mathtt{v} = \mathtt{u}(c(1+d(p,c)))^{-1/(p-1)}$. We easily check that this point is a root of G and hence the assertion follows. □

Now we will establish a sharp version of Sawyer's dyadic testing condition.

Theorem 3.2. *Suppose that a weight w satisfies $[w]_{A_p} = c$. Then for any $R \in \mathcal{T}$,*

$$\int_R \big(\mathcal{M}_{\mathcal{T}}(w^{-1/(p-1)}\chi_R)\big)^p w\, d\mu \leq \big(c(1+d(p,c))\big)^{p/(p-1)} \int_R w^{-1/(p-1)}\, d\mu. \tag{3.4}$$

The constant is the best possible.

Proof. We split the reasoning into three parts.

Step 1. Auxiliary notation. The set R belongs to some generation of the tree $\mathcal{T}$: say, $R \in \mathcal{T}^m$. For any n and any $x \in X$, let $Q^n(x)$ be the element of $\mathcal{T}^n$ which contains x; such a set is uniquely defined for almost all x. Next, introduce the notation

$$\mathtt{w}_n = \langle w \rangle_{Q^n(x)}, \qquad \mathtt{v}_n = \big\langle w^{-1/(p-1)} \big\rangle_{Q^n(x)}, \qquad \mathtt{u}_n = \max_{m \leq k \leq n} \mathtt{v}_k.$$

In the probabilistic language, the functional sequences $(\mathtt{w}_n)_{n\geq m}$ and $(\mathtt{v}_n)_{n\geq m}$ are martingales corresponding to the terminal variables w and $w^{-1/(p-1)}$, while $(\mathtt{u}_n)_{n\geq m}$ is the maximal function of $(\mathtt{v}_n)_{n\geq m}$. Note that for any $n \geq m$ and any $Q \in \mathcal{T}^n$, the functions $\mathtt{u}_n$, $\mathtt{w}_n$ and $\mathtt{v}_n$ are constant on Q and we have

$$\int_Q \mathtt{w}_{n+1} \mathrm{d}\mu = \mu(Q)\mathtt{w}_n|_Q, \qquad \int_Q \mathtt{v}_{n+1} \mathrm{d}\mu = \mu(Q)\mathtt{v}_n|_Q. \tag{3.5}$$

Furthermore, the sequence $(\mathtt{u}_n)_{n\geq m}$ is nondecreasing and satisfies

$$\begin{aligned} \lim_{n\to\infty} \mathtt{u}_n(x) &= \sup_{n\geq m} \big\langle w^{-1/(p-1)} \big\rangle_{Q^n(x)} \\ &= \sup_{n\geq m} \big\langle w^{-1/(p-1)}\chi_R \big\rangle_{Q^n(x)} \\ &= \sup_{n\geq 0} \big\langle w^{-1/(p-1)}\chi_R \big\rangle_{Q^n(x)} = \mathcal{M}_{\mathcal{T}}(w^{-1/(p-1)}\chi_R) \end{aligned} \tag{3.6}$$

almost everywhere.

Step 2. Monotonicity property. The main part of the proof is to show that the sequence $\big(\int_R B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\mathrm{d}\mu\big)_{n\geq m}$ is nondecreasing. It follows from (3.5) that if $n \geq m$ and Q is an element of $\mathcal{T}^n$, then

$$\int_Q B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\mathrm{d}\mu = \mu(Q)B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)|_Q = \int_Q B(\mathtt{u}_n, \mathtt{w}_{n+1}, \mathtt{v}_{n+1})\mathrm{d}\mu, \tag{3.7}$$

since the dependence of B on $\mathtt{w}$ and $\mathtt{v}$ is linear. Now we will show that

$$B(\mathtt{u}_n, \mathtt{w}_{n+1}, \mathtt{v}_{n+1}) \geq B(\mathtt{u}_{n+1}, \mathtt{w}_{n+1}, \mathtt{v}_{n+1}). \tag{3.8}$$

This is clear if $\mathtt{u}_n = \mathtt{u}_{n+1}$. On the other hand, the inequality $\mathtt{u}_{n+1} > \mathtt{u}_n$ implies $\mathtt{v}_{n+1} = \mathtt{u}_{n+1} > \mathtt{u}_n$ (since $\mathtt{u}_{n+1} = \mathtt{u}_n \vee \mathtt{v}_{n+1}$). Therefore, we have $\mathtt{u}_n^{p-1}\mathtt{w}_{n+1} \leq \mathtt{u}_n^{p-1} \cdot c\mathtt{v}_{n+1}^{1-p} < c$ and $\mathtt{u}_{n+1}^{p-1}\mathtt{w}_{n+1} = \mathtt{v}_{n+1}^{p-1}\mathtt{w}_{n+1} \leq c$, so for any $\mathtt{u} \in [\mathtt{u}_n, \mathtt{u}_{n+1}]$ we have the estimate $\mathtt{u}^{p-1}\mathtt{w}_{n+1} \leq c$. Combining this observation with (3.1) immediately yields (3.8) and hence (3.7) gives

$$\int_Q B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\mathrm{d}\mu \geq \int_Q B(\mathtt{u}_{n+1}, \mathtt{w}_{n+1}, \mathtt{v}_{n+1})\mathrm{d}\mu.$$

Summing over all $Q \in \mathcal{T}^n$ contained in R, we get the aforementioned monotonicity property of the sequence $\left(\int_R B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\mathrm{d}\mu\right)_{n\geq m}$.

Step 3. Completion of the proof. For a given $n \geq m$, let us apply (3.3) to get

$$\int_R \mathtt{u}_n^p \mathtt{w}_n \mathrm{d}\mu \leq \int_R B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\mathrm{d}\mu \leq \int_R B(\mathtt{u}_m, \mathtt{w}_m, \mathtt{v}_m)\mathrm{d}\mu. \tag{3.9}$$

Since $R \in \mathcal{T}^m$, the functions $\mathtt{w}_m$ and $\mathtt{v}_m$ are constant on R and $\mathtt{u}_m = \mathtt{v}_m$. Therefore, by (3.2),

$$\begin{aligned}\int_R B(\mathtt{u}_m, \mathtt{w}_m, \mathtt{v}_m)\mathrm{d}\mu &\leq \mu(R)(c(1+d(p,c)))^{p/(p-1)}\mathtt{v}_m|_R \\ &= (c(1+d(p,c)))^{p/(p-1)} \int_R w^{-1/(p-1)}\mathrm{d}\mu.\end{aligned}$$

On the other hand, $\mathtt{w}_n$ is the conditional expectation of w on $\mathcal{T}^n$, so $\int_R \mathtt{u}_n^p \mathtt{w}_n \mathrm{d}\mu = \int_R \mathtt{u}_n^p w \mathrm{d}\mu \xrightarrow{n\to\infty} \int_R (M(w^{-1/(p-1)}\chi_R))^p w \mathrm{d}\mu$, where in the last passage we have exploited (3.6) and Lebesgue's monotone convergence theorem. Combining these observations with (3.9) yields (3.4). The sharpness of this estimate will follow immediately from the sharpness of (1.4). See Remark 3.4 below. □

We are ready to establish our main result. It follows from the sharp weighted version of Carleson embedding theorem (cf. [21]), which we prove here for the sake of completeness.

Theorem 3.3. *Suppose that w is an A_p weight. Let K be a positive constant and assume that nonnegative numbers α_Q, $Q \in \mathcal{T}$, satisfy*

$$\frac{1}{\mu(R)} \sum_{Q\subseteq R} \alpha_Q \langle w^{-1/(p-1)}\rangle_Q^p \leq K\langle w^{-1/(p-1)}\rangle_R \tag{3.10}$$

for all $R \in \mathcal{T}$. Then for any integrable and nonnegative function f on X we have

$$\sum_{Q\in\mathcal{T}} \alpha_Q \langle f\rangle_Q^p \leq K\left(\frac{p}{p-1}\right)^p \int_X f^p w \, d\mu. \tag{3.11}$$

Proof. By homogeneity, we may and do assume that $K = 1$. Consider the functional sequences $(\mathtt{x}_n)_{n\geq 0}$, $(\mathtt{y}_n)_{n\geq 0}$, $(\mathtt{z}_n)_{n\geq 0}$ and $(\mathtt{t}_n)_{n\geq 0}$ given by

$$\mathtt{x}_n(x) = \langle f^p w\rangle_{Q^n(x)}, \qquad \mathtt{y}_n(x) = \langle f\rangle_{Q^n(x)}, \qquad \mathtt{z}_n = \langle w^{-1/(p-1)}\rangle_{Q^n(x)}$$

and

$$\mathtt{t}_n(x) = \frac{1}{\mu(Q_n(x))} \sum_{Q\subseteq Q_n(x), Q\in\mathcal{T}} \alpha_Q \langle w^{-1/(p-1)}\rangle_Q^p.$$

Note that

$$\mathtt{y}_n \leq \mathtt{x}_n^{1/p} \mathtt{z}_n^{1-1/p} \quad \text{ and } \quad \mathtt{t}_n \leq \mathtt{z}_n, \tag{3.12}$$

where the first estimate follows from the Hölder inequality and the second is due to (3.3). Introduce the function $B : [0,\infty)^2 \times (0,\infty)^2 \to \mathbb{R}$ by

$$B(\mathtt{x}, \mathtt{y}, \mathtt{z}, \mathtt{t}) = \left(\frac{p}{p-1}\right)^p \left[\mathtt{x} - \mathtt{y}^p \left(\mathtt{z} + \frac{\mathtt{t}}{p-1}\right)^{1-p}\right].$$

This function is concave: it is easy to check that the Hessian D^2B is nonpositive-definite in the interior of the domain. Therefore for any nonnegative numbers $\mathtt{x}$, $\mathtt{y}$, any positive numbers $\mathtt{z}$, $\mathtt{t}$ and any $\mathtt{h} \geq -\mathtt{x}$, $\mathtt{k} \geq -\mathtt{y}$, $\mathtt{l} > -\mathtt{z}$ and $\mathtt{m} > -\mathtt{t}$ we have

$$\begin{aligned} &B(\mathtt{x}+\mathtt{h},\mathtt{y}+\mathtt{k},\mathtt{z}+\mathtt{l},\mathtt{t}+\mathtt{m})\\ &\leq B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t}) + \frac{\partial B}{\partial \mathtt{x}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{h} + \frac{\partial B}{\partial \mathtt{y}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{k} + \frac{\partial B}{\partial \mathtt{z}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{l}\\ &\qquad + \frac{\partial B}{\partial \mathtt{t}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{m}. \end{aligned} \tag{3.13}$$

Now we will show that the sequence $(\int_X B(\mathtt{x}_n,\mathtt{y}_n,\mathtt{z}_n,\mathtt{t}_n)\mathrm{d}\mu)_{n\geq 0}$ enjoys a certain monotonicity property. To this end, fix $n \geq 0$, $Q \in \mathcal{T}^n$ and pairwise disjoint elements Q_1, Q_2, …, Q_m of $\mathcal{T}^{n+1}$ whose union is Q. Put $\mathtt{x} = \mathtt{x}_n|_Q$, $\mathtt{y} = \mathtt{y}_n|_Q$, $\mathtt{z} = \mathtt{z}_n|_Q$ and $\mathtt{t} = \mathtt{t}_n|_Q$. Furthermore, for any $j = 1, 2, \ldots, m$, let $\mathtt{h}_j$, $\mathtt{k}_j$, $\mathtt{l}_j$ and $\mathtt{m}_j$ be given by $\mathtt{x}+\mathtt{h}_j = \mathtt{x}_{n+1}|_{Q_j}$, $\mathtt{y}+\mathtt{k}_j = \mathtt{y}_{n+1}|Q_j$, $\mathtt{z}+\mathtt{l}_j = \mathtt{z}_{n+1}|Q_j$ and $\mathtt{t}+\mathtt{m}_j = \mathtt{t}_{n+1}|_Q$. It is easy to check that

$$\sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{h}_j = \sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{k}_j = \sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{l}_j = 0. \tag{3.14}$$

Concerning the dynamics of the sequence $(\mathtt{t}_n)_{n\geq 0}$, we see that

$$\begin{aligned} \mathtt{t} &= \frac{1}{\mu(Q)} \sum_{R\subseteq Q,\, R\in\mathcal{T}} \alpha_R \langle w^{-1/(p-1)}\rangle_R^p\\ &= \frac{\alpha_Q\langle w^{-1/(p-1)}\rangle_Q^p}{\mu(Q)} + \sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\cdot\frac{1}{\mu(Q_j)} \sum_{R\subseteq Q_j,\, R\in\mathcal{T}} \alpha_R\langle w^{-1/(p-1)}\rangle_R^p\\ &= \frac{\alpha_Q\langle w^{-1/(p-1)}\rangle_Q^p}{\mu(Q)} + \sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}(\mathtt{t}+\mathtt{m}_j), \end{aligned}$$

which is equivalent to

$$\sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{m}_j = -\frac{\alpha_Q\langle w^{-1/(p-1)}\rangle_Q^p}{\mu(Q)}. \tag{3.15}$$

Let us apply (3.13), with $\mathtt{h} = \mathtt{h}_j$, $\mathtt{k} = \mathtt{k}_j$, $\mathtt{l} = \mathtt{l}_j$ and $\mathtt{m} = \mathtt{m}_j$, multiply throughout by $\mu(Q_j)/\mu(Q)$ and sum the obtained estimates over j. By (3.14) and (3.15), we get

$$\begin{aligned} &\sum_{j=1}^n \frac{\mu(Q_j)}{\mu(Q)} B(\mathtt{x}+\mathtt{h}_j,\mathtt{y}+\mathtt{k}_j,\mathtt{z}+\mathtt{l}_j,\mathtt{t}+\mathtt{m}_j)\\ &\qquad\qquad \leq B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t}) - \frac{\partial B}{\partial \mathtt{t}}(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\cdot\frac{\alpha_Q\langle w^{-1/(p-1)}\rangle_Q^p}{\mu(Q)}. \end{aligned}$$

However, we have

$$\frac{\partial B}{\partial \mathtt{t}}(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t}) = \left(\frac{p}{p-1}\right)^p \mathtt{y}^p \left(\mathtt{z}+\frac{\mathtt{t}}{p-1}\right)^{-p} \geq \frac{\mathtt{y}^p}{\mathtt{z}^p}$$

(in the last passage we have exploited the second estimate in (3.12)), so the preceding estimate implies

$$\frac{1}{\mu(Q)}\int_Q B(\mathtt{x}_{n+1},\mathtt{y}_{n+1},\mathtt{z}_{n+1},\mathtt{t}_{n+1})\mathrm{d}\mu \leq \frac{1}{\mu(Q)}\int_Q B(\mathtt{x}_n,\mathtt{y}_n,\mathtt{z}_n,\mathtt{t}_n)\mathrm{d}\mu - \frac{\alpha_Q\langle f\rangle_Q^p}{\mu(Q)}.$$

Multiply both sides by $\mu(Q)$ and sum over all $Q\in\mathcal{T}^n$ to obtain

$$\int_X B(\mathtt{x}_{n+1},\mathtt{y}_{n+1},\mathtt{z}_{n+1},\mathtt{t}_{n+1})\mathrm{d}\mu \leq \int_X B(\mathtt{x}_n,\mathtt{y}_n,\mathtt{z}_n,\mathtt{t}_n)\mathrm{d}\mu - \sum_{Q\in\mathcal{T}^n}\alpha_Q\langle f\rangle_Q^p$$

and hence for each n we have

$$\int_X B(\mathtt{x}_{n+1},\mathtt{y}_{n+1},\mathtt{z}_{n+1},\mathtt{t}_{n+1})\mathrm{d}\mu \leq \int_X B(\mathtt{x}_0,\mathtt{y}_0,\mathtt{z}_0,\mathtt{t}_0)\mathrm{d}\mu - \sum_{Q\in\mathcal{T}^k,\,k\leq n}\alpha_Q\langle f\rangle_Q^p.$$

Now, by the first inequality in (3.12), we have

$$B(\mathtt{x}_{n+1},\mathtt{y}_{n+1},\mathtt{z}_{n+1},\mathtt{t}_{n+1}) \geq \left(\frac{p}{p-1}\right)^p(\mathtt{x}_{n+1}-\mathtt{y}_{n+1}^p\mathtt{z}_{n+1}^{1-p})\geq 0$$

and, obviously,

$$B(\mathtt{x}_0,\mathtt{y}_0,\mathtt{z}_0,\mathtt{t}_0)\leq \left(\frac{p}{p-1}\right)^p\mathtt{x}_0 = \left(\frac{p}{p-1}\right)^p\int_X f^p w\mathrm{d}\mu.$$

Combining these observations with the previous estimate and letting $n\to\infty$ yields the assertion. □

Proof of (1.4). Take an arbitrary A_p weight w and a function f, and set $c=[w]_{A_p}$. We may assume that f is nonnegative, since the passage from f to $|f|$ does not change the L^p norm of the function and may only increase the the maximal function $\mathcal{M}_\mathcal{T}f$. Furthermore, by a simple approximation argument, we may assume that f is measurable with respect to a σ-algebra generated by some generation $\mathcal{T}^N$. Then we have $\mathcal{M}_\mathcal{T}f=\max_{Q\in\mathcal{T}^n,\,n\leq N}\langle f\rangle_Q\chi_Q$ and hence for each $x\in X$ there is an element $Q=Q(x)$ belonging to $\bigcup_{n\leq N}\mathcal{T}^n$ such that $\mathcal{M}_\mathcal{T}f(x)=\langle f\rangle_Q$. 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 $E(Q)=\{x\in Q: Q(x)=Q\}$ and put $\alpha_Q=w(E(Q))$. We will prove that the inequality (3.4) implies (3.10) with $K=(c(1+d(p,c)))^{p/(p-1)}$. To this end, observe that for any R we have

$$\frac{1}{\mu(R)}\sum_{Q\subset R}\alpha_Q\langle w^{-1/(p-1)}\rangle_Q^p=\frac{1}{\mu(R)}\int_R\sum_{Q\in R}\chi_{E(Q)}\langle w^{-1/(p-1)}\rangle_Q^p w\mathrm{d}\mu.$$

Notice that the sets $E(Q)$ are pairwise disjoint and $E(Q)\subset Q$; therefore, from the very definition of $\mathcal{M}_\mathcal{T}$, we have the pointwise bound $\sum_{Q\in R}\chi_{E(Q)}\langle w^{-1/(p-1)}\rangle_Q^p\leq \mathcal{M}_\mathcal{T}(w^{-1/(p-1)}\chi_R)^p$ on R and hence (3.10) follows. Consequently, (3.11) is also true and this is precisely the desired weighted bound (1.4). □

Remark 3.4. It is now evident that the inequality (3.4) is sharp. Indeed, otherwise we would be able to improve the constant in the estimate (1.4) which, as we have seen in Section 2, is impossible.

References

[1] S. M. Buckley, *Estimates for operator norms on weighted spaces and reverse Jensen inequalities*, Trans. Amer. Math. Soc. **340** (1993), 253–272.

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

[3] L. Grafakos, *Classical and Modern Fourier Analysis*, Pearson Education, Inc. New Jersey, 2004.

[4] A. D. Melas, *The Bellman functions of dyadic-like maximal operators and related inequalities*, Adv. Math. 192 (2005), 310–340.

[5] A. D. Melas, *Dyadic-like maximal operators on LlogL functions*, J. Funct. Anal. 257 (2009), 1631–1654.

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

[7] A. D. Melas and E. N. Nikolidakis, *On weak-type inequalities for dyadic maximal functions*, J. Math. Anal. Appl. 367 (2008), 404–410.

[8] 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.

[9] B. Muckenhoupt, *Weighted norm inequalities for the Hardy maximal function*, Trans. Amer. Math. Soc **165** (1972), 207–226.

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

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

[12] A. Osękowski, *Sharp martingale and semimartingale inequalities*, Monografie Matematyczne **72**, Birkhäuser, 2012.

[13] A. Osękowski, *Sharp* $L^{p,\infty} \to L^q$ *estimates for the dyadic-like maximal operators*, J. Fourier Anal. Appl. **20** (2014), pp. 911–933.

[14] A. Osękowski, *Sharp weak-type estimates for the dyadic-like maximal operators*, Taiwanese J. Math. **19** (2015), pp. 1031–1050.

[15] S. Petermichl, *The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical* A_p*-characteristic*, Amer. J. Math. **129** (2007), no. 5, 1355-1375.

[16] E. T. Sawyer, *A characterization of a two-weight norm inequality for maximal operators*, Studia Math. **75** (1982), 1–11.

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

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

[19] L. Slavin and A. Volberg, *Bellman function and the* H^1*-BMO duality*, Harmonic analysis, partial differential equations, and related topics, 113?126, Contemp. Math., 428, Amer. Math. Soc., Providence, RI, 2007.

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

[21] V. Vasyunin and A. Volberg, *Monge-Ampére equation and Bellman optimization of Carleson embedding theorems*, Linear and complex analysis, pp. 195?238, Amer. Math. Soc. Transl. Ser. 2, 226, Amer. Math. Soc., Providence, RI, 2009.

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

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