Annales Academiæ Scientiarum Fennicæ
Mathematica
Volumen 42, 2017, 889–904

# BEST CONSTANTS IN MUCKENHOUPT'S INEQUALITY

**Adam Osękowski**

University of Warsaw, Department of Mathematics, Informatics and Mechanics
Banacha 2, 02-097 Warsaw, Poland; ados@mimuw.edu.pl

**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 $\mathbf{R}^n$ is an operator acting on locally integrable functions $\phi\colon \mathbf{R}^n \to \mathbf{R}$ by the formula

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

Here the dyadic cubes are those formed by the grids $2^{-N}\mathbf{Z}^n$, $N = 0, 1, 2, \ldots$, 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

$$\lambda\left|\{x \in \mathbf{R}^n \colon \mathcal{M}\phi(x) \ge \lambda\}\right| \le \int_{\{\mathcal{M}\phi \ge \lambda\}} |\phi(u)|\,\mathrm{d}u, \quad \phi \in L^1(\mathbf{R}^n), \tag{1.1}$$

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

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

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, $\mathbf{R}^n$ with Lebesgue's measure). The following statement is a consequence of the classical work of Muckenhoupt [9]. Suppose that $1 < p < \infty$ is given and fixed, and let $w$ be a weight on $\mathbf{R}^n$. Then $\mathcal{M}$ is bounded as an operator on the weighted space

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

https://doi.org/10.5186/aasfm.2017.4256
2010 Mathematics Subject Classification: Primary 42B25; Secondary 46E30, 60G42.
Key words: Maximal, dyadic, Bellman function, best constants.
Research supported by Narodowe Centrum Nauki (Poland), grant no. DEC-2014/14/E/ST1/00532.

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 $\mathbf{R}^n$. 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 $(X, \mu)$ 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 $f\colon X \to \mathbf{R}$ by the formula

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

where this time $\langle \varphi \rangle_Q = \frac{1}{\mu(Q)} \int_Q \varphi \,\mathrm{d}\mu$ is the average of $\varphi$ over $Q$ with respect to the measure $\mu$. 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

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

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 $d(p,c)$. Formally, $d = d(p,c)$ is the unique number in $[0,p)$ satisfying the equation

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

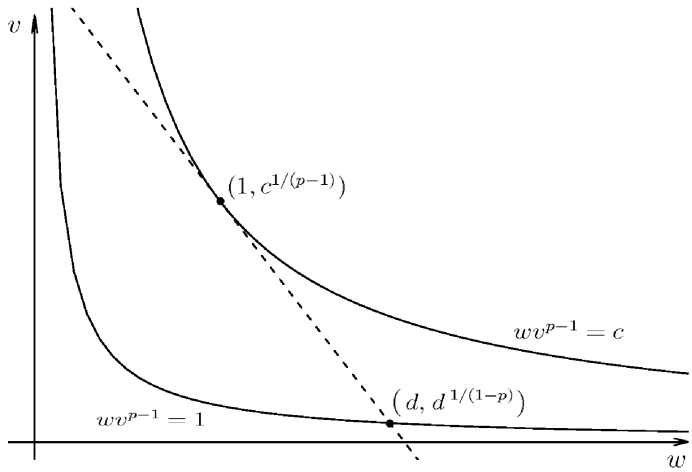


Figure 1. The geometric interpretation of the number $d = d(p,c)$.

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*

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

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

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

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 $\mathbf{R}^n$. Second, note that the above statement contains (1.2): indeed, setting $c = 1$ (which corresponds to the unweighted setting), we derive that $d(p,c) = 1$ and the optimal constant in (1.4) becomes equal to $p/(p-1)$. Finally, let us relate the above statement to the aforementioned result of Buckley. Since $d(p,c) \leq p$, we see that (1.3) yields

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

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 $\mathbf{R}^n$, 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 proof of (1.4). Section 3 is devoted to the construction of an example showing that the constant in (1.4) cannot be smaller than $p/(p - d(p, [w]_{A_p}))$. In the final part of the paper we explain how the Bellman function corresponding to the sharp version of Sawyer's condition was discovered.

## 2. Proof of (1.4)

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

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

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

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

We will prove that this object has the following properties.

**Lemma 2.1.** (i) *If* $\mathtt{u}^{p-1}\mathtt{w} \leq c$, *then*

$$\frac{\partial B}{\partial \mathtt{u}}(\mathtt{u}, \mathtt{w}, \mathtt{v}) \leq 0. \tag{2.1}$$

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

$$B(\mathtt{v}, \mathtt{w}, \mathtt{v}) \leq (c d(p,c))^{p/(p-1)} \mathtt{v}. \tag{2.2}$$

(iii) *We have*

$$B(\mathtt{u}, \mathtt{w}, \mathtt{v}) \geq \mathtt{u}^p \mathtt{w}. \tag{2.3}$$

*Proof.* (i) We easily compute that

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

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

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

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

$$\frac{c(p-1)d(p,c)}{d(p,c)-1}\left[\big(cd(p,c)\big)^{1/(p-1)} - 1\right] = \frac{c(p-1)d(p,c)}{p-d(p,c)} = (cd(p,c))^{p/(p-1)}.$$

(iii) The majorization is equivalent to

$$\mathtt{u}^p\mathtt{w} + (p-1)\big(cd(p,c)\big)^{p/(p-1)}\mathtt{v} - pcd(p,c)\mathtt{u} \ge 0.$$

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

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

We compute that $G'(\mathtt{v}) = (p-1)\big(-(\mathtt{u}/\mathtt{v})^p + (cd(p,c))^{p/(p-1)}\big)$ and hence $G$ attains its minimum at the point $\mathtt{v}=\mathtt{u}(cd(p,c))^{-1/(p-1)}$. 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 2.2.** *Suppose that a weight $w$ satisfies $[w]_{A_p}=c$. Then for any $R\in\mathcal{T}$,*

$$\int_R \big(\mathcal{M}_{\mathcal{T}}(w^{-1/(p-1)}\chi_R)\big)^p w\,d\mu \le \big(cd(p,c)\big)^{p/(p-1)}\int_R w^{-1/(p-1)}\,d\mu. \tag{2.4}$$

*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

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

In the probabilistic language, the functional sequences $(\mathtt{w}_n)_{n\ge m}$ and $(\mathtt{v}_n)_{n\ge m}$ are martingales corresponding to the terminal variables $w$ and $w^{-1/(p-1)}$, while $(\mathtt{u}_n)_{n\ge m}$ is the maximal function of $(\mathtt{v}_n)_{n\ge m}$. Note that for any $n\ge m$ 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

$$\int_Q \mathtt{w}_{n+1}\,d\mu = \mu(Q)\mathtt{w}_n|_Q,\qquad \int_Q \mathtt{v}_{n+1}\,d\mu = \mu(Q)\mathtt{v}_n|_Q. \tag{2.5}$$

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

$$\begin{aligned}\lim_{n\to\infty}\mathtt{u}_n(x) &= \sup_{n\ge m}\big\langle w^{-1/(p-1)}\big\rangle_{Q^n(x)} = \sup_{n\ge m}\big\langle w^{-1/(p-1)}\chi_R\big\rangle_{Q^n(x)} \\ &= \sup_{n\ge 0}\big\langle w^{-1/(p-1)}\chi_R\big\rangle_{Q^n(x)} = \mathcal{M}_{\mathcal{T}}(w^{-1/(p-1)}\chi_R)\end{aligned} \tag{2.6}$$

almost everywhere.

*Step 2. Monotonicity property.* The main part of the proof is to show that the sequence $\left(\int_R B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\,\mathrm{d}\mu\right)_{n\geq m}$ is nondecreasing. It follows from (2.5) that if $n \geq m$ and $Q$ is an element of $\mathcal{T}^n$, then

$$\int_Q B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\,\mathrm{d}\mu = \mu(Q) B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)|_Q = \int_Q B(\mathtt{u}_n, \mathtt{w}_{n+1}, \mathtt{v}_{n+1})\,\mathrm{d}\mu, \tag{2.7}$$

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

$$B(\mathtt{u}_n, \mathtt{w}_{n+1}, \mathtt{v}_{n+1}) \geq B(\mathtt{u}_{n+1}, \mathtt{w}_{n+1}, \mathtt{v}_{n+1}). \tag{2.8}$$

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 (2.1) immediately yields (2.8) and hence (2.7) 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 (2.3) to get

$$\int_R \mathtt{u}_n^p \mathtt{w}_n\,\mathrm{d}\mu \leq \int_R B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\,\mathrm{d}\mu \leq \int_R B(\mathtt{u}_m, \mathtt{w}_m, \mathtt{v}_m)\,\mathrm{d}\mu. \tag{2.9}$$

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 (2.2),

$$\begin{aligned}\int_R B(\mathtt{u}_m, \mathtt{w}_m, \mathtt{v}_m)\,\mathrm{d}\mu &\leq \mu(R)(cd(p,c))^{p/(p-1)}\mathtt{v}_m|_R \\ &= (cd(p,c))^{p/(p-1)} \int_R w^{-1/(p-1)}\,\mathrm{d}\mu.\end{aligned} \tag{2.10}$$

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 (2.6) and Lebesgue's monotone convergence theorem. Combining these observations with (2.9) yields (2.4). The sharpness of this estimate will follow immediately from the sharpness of (1.4). See Remark 2.4 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 2.3.** *Suppose that $w$ is an $A_p$ weight. Let $K$ be a positive constant and assume that nonnegative numbers $\alpha_Q$, $Q \in \mathcal{T}$, satisfy*

$$\frac{1}{\mu(R)} \sum_{Q\subseteq R} \alpha_Q \langle w^{-1/(p-1)}\rangle_Q^p \leq K \langle w^{-1/(p-1)}\rangle_R \tag{2.11}$$

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

$$\sum_{Q\in\mathcal{T}} \alpha_Q \langle f\rangle_Q^p \leq K \left(\frac{p}{p-1}\right)^p \int_X f^p w\,d\mu. \tag{2.12}$$

*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

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

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

$$\mathtt{y}_n \leq \mathtt{x}_n^{1/p}\mathtt{z}_n^{1-1/p} \quad \text{and} \quad \mathtt{t}_n \leq \mathtt{z}_n, \tag{2.13}$$

where the first estimate follows from the Hölder inequality and the second is due to (2.3). Introduce the function $B\colon [0,\infty)^2 \times (0,\infty)^2 \to \mathbf{R}$ 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

$$\begin{aligned} B(\mathtt{x}+\mathtt{h},\mathtt{y}+\mathtt{k},\mathtt{z}+\mathtt{l},\mathtt{t}+\mathtt{m}) &\leq B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t}) + \frac{\partial B}{\partial \mathtt{x}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{h} \\ &+ \frac{\partial B}{\partial \mathtt{y}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{k} + \frac{\partial B}{\partial \mathtt{z}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{l} + \frac{\partial B}{\partial \mathtt{t}}B(\mathtt{x},\mathtt{y},\mathtt{z},\mathtt{t})\mathtt{m}. \end{aligned} \tag{2.14}$$

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

$$\sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{h}_j = \sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{k}_j = \sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{l}_j = 0. \tag{2.15}$$

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

$$\sum_{j=1}^m \frac{\mu(Q_j)}{\mu(Q)}\mathtt{m}_j = -\frac{\alpha_Q\langle w^{-1/(p-1)}\rangle_Q^p}{\mu(Q)}. \tag{2.16}$$

Let us apply (2.14), 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 (2.15) and (2.16), we get

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

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 (2.13)), 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 (2.13), 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,

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

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 $\sigma$-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 (2.4) implies (2.11) with $K=(cd(p,c))^{p/(p-1)}$. 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 (2.11) follows. Consequently, (2.12) is also true and this is precisely the desired weighted bound (1.4). □

**Remark 2.4.** The inequality (2.4) is sharp. Indeed, otherwise we would be able to improve the constant in the estimate (1.4) which, as we will see in Section 3 below, is impossible.

## 3. 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 $\ell$ 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 $d(\tilde{c})$. Furthermore, the line $\ell$ 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.

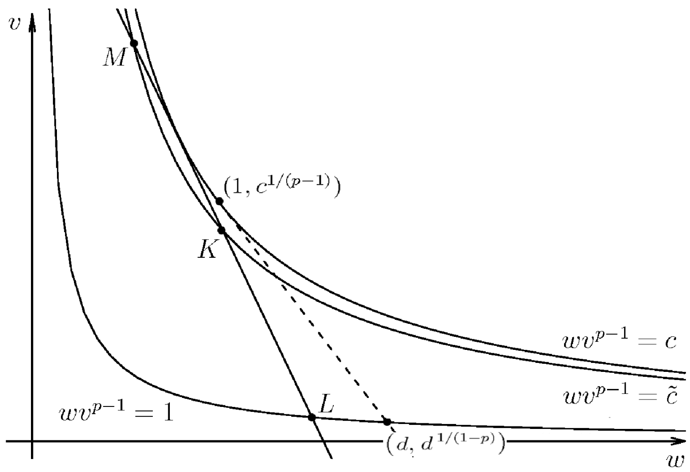


Figure 2. The crucial parameters and their geometric interpretation: $K = (1, \tilde{c}^{1/(p-1)})$, $L = \big(d(\tilde{c}), (d(\tilde{c}))^{1/(1-p)}\big)$ and $M = \big(1-\delta, (\tilde{c}(1-\delta))^{1/(1-p)}\big)$.

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

$$(\tilde{c}d(\tilde{c}))^{1/(p-1)}\big(d(\tilde{c}) - 1 + \delta - (d(\tilde{c}) - 1)(1-\delta)^{1/(1-p)}\big) = \delta, \tag{3.1}$$

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

$$d(\tilde{c}) < d(p, c) < p, \tag{3.2}$$

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 $r > -1/p - 1/(p(d(p,c)-1))$. By the left estimate in (3.2), we see that for all $\tilde{c}$ we have $r > -1/p - 1/(p(d(\tilde{c})-1))$, which combined with the right bound in (3.2) implies

$$1 + r(d(\tilde{c})-1) > 0. \tag{3.3}$$

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

**Lemma 3.1.** *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 $\{F(Q)\colon Q \text{ an atom of } A_n\}$, 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

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

and hence

$$\mu(Q\cap(A_n\setminus A_{n+1})) = \mu(Q)\left(\frac{d(\tilde{c})-1}{d(\tilde{c})-1+\delta}\right)^{n-m}\frac{\delta}{d(\tilde{c})-1+\delta}. \tag{3.4}$$

Now, introduce the weight $w$ on $X$ by the formula

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

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

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

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 (3.4) to obtain that for each atom $Q$ of $A_m$ we have

$$\langle w\rangle_Q = \sum_{n=m}^{\infty}\left(\frac{d(\tilde{c})-1}{d(\tilde{c})-1+\delta}\right)^{n-m}(1-\delta)^n d(\tilde{c})\cdot\frac{\delta}{d(\tilde{c})-1+\delta} = (1-\delta)^m \tag{3.5}$$

and

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

where in the last passage we have exploited (3.1). 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

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

By (3.5), 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

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

A similar calculation shows that

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

and therefore

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

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 $\eta$ ranges from 0 to 1, the point $\eta M+(1-\eta)L$ runs over the line segment $ML$ which is entirely contained in $\{(w,v)\colon wv^{p-1}\le c\}$. Since $R$ was arbitrary, we obtain the desired $A_p$ condition: $[w]_{A_p}\le c$.

*Step 4. Completion of the proof.* In the same manner as above, one verifies that if $Q$ is an atom of $A_m$, then

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

(the ratio of the geometric series, equal to $(d(\tilde c)-1)(1-r\delta)/(d(\tilde c)-1+\delta)$), is less than 1: this is equivalent to (3.3)). Consequently, we see that $\mathcal{M}_{\mathcal{T}}f\ge(1-r\delta)^m$ on $A_m$ and hence, by the definition of $f$, we obtain $\mathcal{M}_{\mathcal{T}}f\ge(1+r(d(\tilde c)-1))^{-1}f$ on $A_m\setminus A_{m+1}$. The latter bound does not depend on $m$, so we can rewrite it uniformly as

$$\mathcal{M}_{\mathcal{T}}f\ge(1+r(d(\tilde c)-1))^{-1}f\quad\text{ on } X.$$

Now if we choose $r$ sufficiently close to $-1/p-1/(p(d(p,c)-1))$ and then $\tilde{c}$ sufficiently close to $c$, then the number $(1+r(d(\tilde{c})-1))^{-1}$ can be made arbitrarily close to $p/(p-d(p,c))$, 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

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

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

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

This establishes the desired sharpness.

## 4. On the special function corresponding to (2.4)

Now we will sketch an informal reasoning which has led us to the discovery of the best constant in (2.4) and the formula for the special function corresponding to this estimate. Let $1<p<\infty$ and $c>1$ be fixed. Suppose that the underlying probability space $(X,\mu)$ is the interval $[0,1]$ with Lebesgue's measure, and equip it with the tree of dyadic subintervals. Suppose we are interested in the least constant $\kappa(p,c)$ such that whenever $R$ is a dyadic subinterval of $[0,1]$ and $w$ is a weight on $[0,1]$ satisfying $[w]_{A_p}=c$, then

$$\int_R \left(\mathcal{M}_{\mathcal{T}}(w^{-1/(p-1)}\chi_R)\right)^p w\,\mathrm{d}\mu \le \kappa(p,c)\int_R w^{-1/(p-1)}\,\mathrm{d}\mu. \tag{4.1}$$

To this end, let us introduce the corresponding abstract Bellman function $\mathbf{B}\colon \mathcal{D}_{p,c}\to\mathbf{R}$ (where, as above, $\mathcal{D}_{p,c}=\{(\mathtt{u},\mathtt{v},\mathtt{w})\in(0,\infty)^3\colon 1\le \mathtt{w}\mathtt{v}^{p-1}\le c\}$), given by

$$\mathbf{B}(\mathtt{u},\mathtt{w},\mathtt{v})=\sup\left\{\frac{1}{|R|}\int_R \left(\max\left\{\mathcal{M}(w^{-1/(p-1)}),\mathtt{u}\right\}\right)^p w\,\mathrm{d}x\right\},$$

where $\mathcal{M}$ stands for the dyadic maximal operator restricted to $R$ and the supremum is taken over all $A_p$ weights $w$ on $R$ such that $[w]_{A_p}\le c$, $\langle w\rangle_R=\mathtt{w}$ and $\langle w^{-1/(p-1)}\rangle_R=\mathtt{v}$. At the first glance, the function $\mathbf{B}$ depends also on the interval $R$, however, this is not the case. Indeed, for any two dyadic intervals $R_1$ and $R_2$, an affine mapping of one interval onto another puts the $A_p$ weights in one-to-one correspondence, and such a change of the variable preserves the averages and the $A_p$ characteristics. On the other hand, considering different domains $R$ is crucial for the understanding of properties of $\mathbf{B}$.

Directly from the definition of $\mathbf{B}$ and the validity of (4.1) (with some a priori unknown constant $\kappa(p,c)$), we have

$$\mathbf{B}(\mathtt{v},\mathtt{w},\mathtt{v})\le\kappa(p,c)\mathtt{v}\quad\text{for all } \mathtt{u},\ \mathtt{w},\ \mathtt{v}. \tag{4.2}$$

To see this, note that $\mathcal{M}(w^{-1/(p-1)})\ge\frac{1}{|R|}\int_R w^{-1/(p-1)}\,\mathrm{d}x=\mathtt{v}$; thus for $\mathtt{u}=\mathtt{v}$, we have $\max\left\{\mathcal{M}(w^{-1/(p-1)}),\mathtt{u}\right\}=\mathcal{M}(w^{-1/(p-1)})$ and (4.2) follows. Furthermore,

$$\frac{1}{|R|}\int_R\left(\max\left\{\mathcal{M}(w^{-1/(p-1)}),\mathtt{u}\right\}\right)^p w\,\mathrm{d}x\ge \mathtt{u}^p\frac{1}{|R|}\int_R w\,\mathrm{d}x$$

which implies

$$\mathbf{B}(\mathtt{u},\mathtt{w},\mathtt{v}) \geq \mathtt{u}^p\mathtt{w} \quad \text{for all } \mathtt{u},\ \mathtt{w},\ \mathtt{v}. \tag{4.3}$$

Similarly, since $\mathcal{M}(w^{-1/(p-1)}) \geq \frac{1}{|R|}\int_R w^{-1/(p-1)}\,\mathrm{d}x = \mathtt{v}$, we obtain

$$\mathbf{B}(\mathtt{u},\mathtt{w},\mathtt{v}) = \mathbf{B}(\max\{\mathtt{u},\mathtt{v}\},\mathtt{w},\mathtt{v}) \tag{4.4}$$

for all $\mathtt{u}$, $\mathtt{w}$, $\mathtt{v}$. The final property is the following concavity-type condition. Pick $(\mathtt{u},\mathtt{w},\mathtt{v}) \in \mathcal{D}_{p,c}$ such that $\mathtt{u} \geq \mathtt{v}$ and the positive numbers $\mathtt{w}_\pm$, $\mathtt{v}_\pm$ satisfying $1 \leq \mathtt{w}_\pm\mathtt{v}_\pm^{p-1} \leq c$ and such that

$$\mathtt{w} = \frac{\mathtt{w}_- + \mathtt{w}_+}{2}, \quad \mathtt{v} = \frac{\mathtt{v}_- + \mathtt{v}_+}{2}.$$

Furthermore, pick a weight $w_-$ on $I_- = [0,1/2)$ and a weight $w_+$ on $I_+ = [1/2,1)$ such that $[w_\pm] \leq c$, $\langle w_\pm\rangle_{I_\pm} = \mathtt{w}_\pm$ and $\langle w_\pm^{-1/(p-1)}\rangle_{I_\pm} = \mathtt{v}_\pm$ (it is not difficult to show that such weights exist). Then splice $w_-$ and $w_+$ into one weight on $[0,1)$ given by the formula $w = w_-\chi_{I_-} + w_+\chi_{I_+}$. Then $\langle w\rangle_{[0,1)} = \mathtt{w}$, $\langle w^{-1/(p-1)}\rangle_{[0,1)} = \mathtt{v}$ and hence

$$\begin{aligned}\mathbf{B}(\mathtt{u},\mathtt{w},\mathtt{v}) &\geq \int_{[0,1)} \left(\max\left\{\mathcal{M}(w^{-1/(p-1)}),\mathtt{u}\right\}\right)^p w\,\mathrm{d}x\\ &= \int_0^{1/2} \left(\max\left\{\mathcal{M}(w_-^{-1/(p-1)}),\mathtt{u}\right\}\right)^p w\,\mathrm{d}x + \int_{1/2}^1 \left(\max\left\{\mathcal{M}(w_-^{-1/(p-1)}),\mathtt{u}\right\}\right)^p w\,\mathrm{d}x.\end{aligned}$$

Note that in the first of the three above integrals the maximal operator $\mathcal{M}$ on $[0,1)$ is considered, while in the last two integrals we may assume that $\mathcal{M}$ are maximal operators localized to the intervals $[0,1/2)$ and $[1/2,1)$, respectively. This is due to the assumption $\mathtt{u} \geq \mathtt{v}$. Therefore, taking the supremum over all weights $w_\pm$ as above, we get

$$\mathbf{B}(\mathtt{u},\mathtt{w},\mathtt{v}) \geq \frac{1}{2}\left(\mathbf{B}(\mathtt{u},\mathtt{w}_-,\mathtt{v}_-) + \mathbf{B}(\mathtt{u},\mathtt{w}_+,\mathtt{v}_+)\right) \tag{4.5}$$

for all $\mathtt{u}$, $\mathtt{w},\mathtt{v}$, $\mathtt{w}_\pm\,\mathtt{v}_\pm$ as above. There is also a structural, homogeneity-type property which $\mathbf{B}$ possesses. Namely, a weight $w$ satisfies $[w]_{A_p} \leq c$, $\langle w\rangle_{[0,1)} = \mathtt{w}$ and $\langle w^{-1/(p-1)}\rangle = \mathtt{v}$ if and only if for all $\lambda > 0$ the weight $\tilde{w} = \lambda w$ satisfies $[\tilde{w}]_{A_p} \leq c$, $\langle\tilde{w}\rangle_{[0,1)} = \lambda\mathtt{w}$ and $\langle\tilde{w}^{-1/(p-1)}\rangle_{[0,1)} = \lambda^{-1/(p-1)}\mathtt{v}$. Combining this fact with the definition of $\mathbf{B}$ gives

$$\mathbf{B}(\lambda^{-1/(p-1)}\mathtt{u},\lambda\mathtt{w},\lambda^{-1/(p-1)}\mathtt{v}) = \lambda^{-1/(p-1)}\mathbf{B}(\mathtt{u},\mathtt{w},\mathtt{v}) \tag{4.6}$$

for all $\lambda > 0$ and all $\mathtt{u}$, $\mathtt{w}$, $\mathtt{v}$.

The important observation in the reverse direction is the following. Namely, if we take some constant $\kappa(p,c)$ and manage to construct *some* function $B$ which satisfies the conditions (4.2), (4.3), (4.4) and (4.5), then the desired estimate (4.1) holds true (with the constant $\kappa(p,c)$ we have just picked). This can be easily shown by the same reasoning as in the proof of (2.4) in Section 2, the conditions (4.4) and (4.5) imply the monotonicity property of the sequence $\left(\int_0^1 B(\mathtt{u}_n,\mathtt{w}_n,\mathtt{v}_n)\,\mathrm{d}x\right)_{n\geq 0}$.

To find such a function $B$, we will make several guesses. Roughly speaking, we will take a look at the conditions (4.2), (4.3), (4.4) and (4.5) and assume that the inequalities they provide actually become equalities at some extremal instances. First look at the condition (4.5). This property means that for a fixed $\mathtt{u}$, the function $B$, considered as a function of $\mathtt{w}$ and $\mathtt{v}$ (satisfying $1 \leq \mathtt{w}\mathtt{v}^{p-1} \leq c$ and $\mathtt{v} \leq \mathtt{u}$), is concave.

There is a natural guess, which is fortunately successful. Namely, one assumes that $B$ is actually linear with respect to these two variables. In other words, we write

$$B(\mathtt{u},\mathtt{w},\mathtt{v}) = a_1(\mathtt{u})\mathtt{w} + a_2(\mathtt{u})\mathtt{v} + a_3(\mathtt{u}),$$

for some unknown functions $a_1$, $a_2$, $a_3$. Now, employing the homogeneity condition (4.6), we get that $a_1(\mathtt{u}) = \alpha\mathtt{u}^p$, $a_2(\mathtt{u}) = \beta$ and $a_3(\mathtt{u}) = \gamma\mathtt{u}$ for some real constants $\alpha$, $\beta$, $\gamma$. Now the property (4.3) implies $\alpha \geq 1$: indeed, apply the inequality to $\mathtt{v} = \mathtt{w}^{-1/(p-1)}$ and let $\mathtt{w} \to \infty$. Next, take $\mathtt{u} = \mathtt{v}$, $\mathtt{w} = \mathtt{w}_\pm < c\mathtt{v}^{1-p}$ and $\mathtt{v}_\pm = \mathtt{v} \pm \delta$, where $\delta$ is a small number (so that $1 \leq \mathtt{w}_\pm\mathtt{v}_\pm^{p-1} \leq c$), eventually sent to 0. Then (4.4) and (4.5) imply $\frac{\partial \mathbf{B}}{\partial \mathtt{u}}(\mathtt{v},\mathtt{w},\mathtt{v}) \leq 0$, or $p\alpha\mathtt{u}^{p-1}\mathtt{w} + \gamma \leq 0$. Since $\alpha$ is positive, this assumption gets most restrictive for the limit case $\mathtt{w} = c\mathtt{v}^{1-p} = c\mathtt{u}^{1-p}$ and becomes $p\alpha c + \gamma \leq 0$. We assume that we actually have equality here: so,

$$B(\mathtt{u},\mathtt{w},\mathtt{v}) = \alpha\mathtt{u}^p\mathtt{w} + \beta\mathtt{v} - p\alpha c\mathtt{u}.$$

Now, let us return to the requirement (4.3). If $\beta \geq 0$, as we temporarily assume, and we vary $\mathtt{v}$, then this requirement becomes most restrictive for $\mathtt{v} = \mathtt{w}^{-1/(p-1)}$. In this case, the inequality reads

$$(\alpha - 1)\mathtt{u}^p\mathtt{w} + \beta\mathtt{w}^{1/(1-p)} - p\alpha c\mathtt{u} \geq 0.$$

The left-hand side, considered as a function of $\mathtt{w}$ attains its minimal value at the point

$$\mathtt{w} = \left[\frac{\beta}{(p-1)\mathtt{u}^p(\alpha-1)}\right]^{(p-1)/p}, \tag{4.7}$$

and this minimal value is equal to

$$\frac{p(\alpha-1)^{1/p}\beta^{1-1/p}}{(p-1)^{1-1/p}} - p\alpha c.$$

Assuming that this value is 0, we get the formula for $\beta$:

$$\beta = \frac{(\alpha c)^{p/(p-1)}(p-1)}{(\alpha-1)^{1/(p-1)}}$$

and the formula (4.7) for extremal $w$ becomes

$$\mathtt{w} = c\mathtt{u}^{1-p} \cdot \frac{\alpha}{\alpha - 1}. \tag{4.8}$$

Finally, we turn our attention to (4.2): it reads

$$\alpha\mathtt{w}\mathtt{v}^p + \frac{(\alpha c)^{p/(p-1)}(p-1)}{(\alpha-1)^{1/(p-1)}}\mathtt{v} - p\alpha c\mathtt{u} \leq \kappa(p,c)\mathtt{v}.$$

Plugging $\mathtt{w} = c\mathtt{v}^{p-1}$ makes the left-hand side the largest possible (with respect to all the choices of the variable $\mathtt{w}$) and the estimate becomes

$$c(1-p)\alpha + \frac{(\alpha c)^{p/(p-1)}(p-1)}{(\alpha-1)^{1/(p-1)}} \leq \kappa(p,c).$$

Now we maximize the left-hand side with respect to $\alpha$ and *assume* that $\kappa(p,c)$ is equal to this extremal value. A simple analysis shows that the maximum is attained at $\alpha$ satisfying

$$c^{1/(p-1)}\left(\frac{\alpha}{\alpha-1}\right)^{1/(p-1)}\left(p - \frac{\alpha}{\alpha-1}\right) = p - 1.$$

Substituting $d = \alpha/(\alpha-1)$ we recover (1.3); then the maximal value is precisely $(cd(p,c))^{p/(p-1)}$ and the function $B$ coincides with that studied in Section 2.

Let us comment on the (almost) extremal weights in (2.4), i.e., those weights for which the equality is almost attained. These weights have appeared explicitly in Section 3, but the interesting point here is that they can be extracted from the function $B$ constructed above. Our reasoning will be a little informal: our purpose is to indicate the *idea* behind the construction of the extremal weights (the formal analysis of which have been conducted separately above). Let us begin by the trivial observation that the weight $w$ is extremal if all the intermediate inequalities appearing in the proof of (2.4) are actually (almost) equalities. This gives us the following three (informal) conditions:

1° We have $B(\mathtt{v}_0, \mathtt{w}_0, \mathtt{v}_0) = (cd(p,c))^{p/(p-1)}\mathtt{v}_0$ (see the first inequality in (2.10) for $m = 0$ and $R = X$).

2° We have $B(\mathtt{u}_\infty, \mathtt{w}_\infty, \mathtt{v}_\infty) = u_\infty^p w_\infty$ (see the first inequality in (2.9)).

3° The sequences $(\mathtt{u}_n)_{n\geq 0}$, $(\mathtt{w}_n)_{n\geq 0}$ and $(\mathtt{v}_n)_{n\geq 0}$ have the property that the sequence $\left(\int_X B(\mathtt{u}_n, \mathtt{w}_n, \mathtt{v}_n)\,\mathrm{d}\mu\right)_{n\geq 0}$ is (almost) constant.

We will construct the sequences $(\mathtt{u}_n)_{n\geq 0}$, $(\mathtt{w}_n)_{n\geq 0}$, $(\mathtt{v}_n)_{n\geq 0}$ and obtain the desired weight $w$ as the pointwise limit $w_\infty$. We start with 1°: it follows from the above analysis that this condition holds if $w_0 v_0^{p-1} = c$; so, for instance, let us take $w_0 \equiv 1$, $u_0 = v_0 \equiv c^{1/(p-1)}$ on $X$. On the other hand, the condition 2° means that the triple $(\mathtt{u}_\infty, \mathtt{w}_\infty, \mathtt{v}_\infty) = (\mathcal{M}_\mathcal{T}(w^{-1/(p-1)}), w, w^{1/(1-p)})$ must take values in the set $\mathcal{S} = \{(\mathtt{u}, \mathtt{w}, \mathtt{v}) : \mathtt{w} = \mathtt{v}^{1-p} = cd\mathtt{u}^{1-p}\}$: see (4.8) and recall that $\alpha/(\alpha-1) = d$. It is important to understand the geometry of this set. Namely, if we take the point $(c\mathtt{u}^{1-p}, \mathtt{u})$ lying on the curve $\mathtt{w}\mathtt{v}^{p-1} = c$, and draw the tangent line through this point. As we already know, this tangent intersects the curve $\mathtt{w}\mathtt{v}^{p-1} = 1$ at two points, one of which is $P_\mathtt{u} = (cd\mathtt{u}^{1-p}, (cd)^{1/(1-p)}\mathtt{u})$: the key observation is that $(\mathtt{u}, P_\mathtt{u})$ belongs to the set $\mathcal{S}$. This suggests the following construction: take a small positive $\delta$ and consider the halfline starting from $L_1 = (1-\delta, ((1-\delta)c)^{1/(1-p)})$ and passing through $(1, c^{1/(1-p)})$. This halfline intersects the curve $\mathtt{w}\mathtt{v}^{p-1} = 1$ at some point $R_1$ close to $(d(p,c), d(p,c)^{1/(1-p)})$. Then define functions $\mathtt{u}_1$, $\mathtt{w}_1$ and $\mathtt{v}_1$ by the requirement that $(\mathtt{w}_1, \mathtt{v}_1)$ takes values in the set $\{L_1, R_1\}$ (and $\mathtt{u}_1 = \max\{\mathtt{v}_0, \mathtt{v}_1\}$). On the set $\mathcal{A}_1$ where the value $R_1$ is taken, take $w = \mathtt{w}_1$ and finish construction. To continue on the compliment of this set, consider the halfline starting from $L_2 = ((1-\delta)^2, ((1-\delta)^2 c)^{1/(1-p)})$ and passing through $(1-\delta, ((1-\delta)c)^{1/(1-p)})$. This halfline intersects the curve $\mathtt{w}\mathtt{v}^{p-1} = 1$ at some point $R_2$. Define $\mathtt{u}_2$, $\mathtt{w}_2$ and $\mathtt{v}_2$ on $X \setminus A_1$ by the requirement that $(\mathtt{w}_2, \mathtt{v}_2)$ takes values in the set $\{L_2, R_2\}$ (and $\mathtt{u}_1 = \max\{\mathtt{v}_0, \mathtt{v}_1, \mathtt{v}_2\}$). On the set $A_2 \subset X \setminus A_1$ where the value $R_2$ is taken, take $w = \mathtt{w}_2$ and finish construction. On the set $X \setminus (A_1 \cup A_2)$, consider the halfline starting from $L_3 = ((1-\delta)^3, ((1-\delta)^3 c)^{1/(1-p)}$, and so on. It is easy to see that the weight $w$ obtained satisfies 1°, and almost satisfies 2° and 3°. Sometimes the tree is not rich enough for the above construction: having successfully constructed the weight $w$ on the set $A_k$, we may need several steps for the pair $(\mathtt{w}_n, \mathtt{v}_n)$ to reach the set $\{L_{k+1}, R_{k+1}\}$; but the remaining argumentation remains unchanged. See Section 3.

*Acknowledgment.* The author would like to thank an anonymous referee for the careful reading of the first version of the paper and many helpful suggestions.

## References


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

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

[3] Grafakos, L.: Classical and modern Fourier analysis. - Pearson Education, Inc., New Jersey, 2004.

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

[5] Melas, A. D.: Dyadic-like maximal operators on $L \log L$ functions. - J. Funct. Anal. 257, 2009, 1631–1654.

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

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

[8] Melas, A. D., 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] Muckenhoupt, B.: Weighted norm inequalities for the Hardy maximal function. - Trans. Amer. Math. Soc. 165, 1972, 207–226.

[10] Nazarov, F. L., 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.

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

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

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

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

[15] Petermichl, S.: The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical $A_p$-characteristic. - Amer. J. Math. 129:5, 2007, 1355–1375.

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

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

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

[19] Slavin, L., and A. Volberg: Bellman function and the $H^1$-BMO duality. - In: Harmonic analysis, partial differential equations, and related topics, Contemp. Math. 428, 2007, 113–126.

[20] Vasyunin, V.: The exact constant in the inverse Hölder inequality for Muckenhoupt weights. - Algebra i Analiz 15, 2003, 73–117; English transl. in St. Petersburg Math. J. 15, 2004, 49–79.

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


Received 1 December 2016 • Accepted 24 February 2017