A sharp two-weight estimate for the maximal operator under a bump condition

Adam Osękowski

Abstract. Let $\mathcal{M}_{\mathcal{D}}$ be the dyadic maximal operator on $\mathbb{R}^n$. The paper contains the identification of the best constant in the two-weight estimate

$$\|\mathcal{M}_{\mathcal{D}} f\|_{L^p(w)} \leq C_{p,\sigma,w} \|f\|_{L^p(\sigma^{1-p})}$$

under the assumption that the pair (σ, w) of weights satisfies an appropriate bump condition. The result is shown to be true in the larger context of abstract probability spaces equipped with a tree-like structure.

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

Keywords. Maximal, dyadic, weight, bump condition, Bellman function.

1. Introduction

The purpose of this paper is to study a class of sharp two-weight L^p estimates for maximal operators, under the assumption that the weights satisfy a certain bump condition. To present the results from an appropriate perspective, we start with the dyadic setting. Suppose that $\mathcal{D} = \mathcal{D}(\mathbb{R}^n)$ is a class of dyadic cubes contained in $\mathbb{R}^n$ and let $\mathcal{M}_{\mathcal{D}}$ be the associated maximal operator, acting on locally integrable functions $f : \mathbb{R}^n \to \mathbb{R}$ by

$$\mathcal{M}_{\mathcal{D}} f = \sup_{Q \in \mathcal{D}(\mathbb{R}^n)} \langle |f| \rangle_Q \chi_Q.$$

Here we have used the notation $\langle f \rangle_Q$ for $\frac{1}{|Q|} \int_Q f$, the average of f over Q (with respect to the Lebesgue measure). This operator is a fundamental object in analysis and there is a huge literature devoted to the study of tight estimates for this object. For example, $\mathcal{M}_{\mathcal{D}}$ satisfies the weak-type $(1,1)$ inequality

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

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

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

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 are of fundamental importance to analysis and have been extended and applied in numerous directions. For some related sharp estimates, see e.g. [6, 7, 8, 9, 10, 13, 14, 16].

Let us turn our attention to the weighted versions of (1.1) and (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). A weight w gives rise to the corresponding L^p and weak L^p spaces, given by

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

and

$$L^{p,\infty}(w) = \left\{ f : \mathbb{R}^n \to \mathbb{R} : \|f\|_{L^{p,\infty}(w)} = \sup_{\lambda>0} \left(\lambda^p \int_{\{x:|f(x)|\geq\lambda\}} w \right)^{1/p} < \infty \right\}.$$

The following statement is due to Muckenhoupt [11]. Suppose that $1 \leq p < \infty$ is given and fixed, and let w be a weight on $\mathbb{R}^n$. Then $\mathcal{M}_{\mathcal{D}}$ is bounded as an operator from $L^p(w) \to L^{p,\infty}(w)$ 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 $\mathbb{R}^n$ (for $p = 1$, we need to pass to the limit: $[w]_{A_1} = \sup \langle w \rangle_Q \operatorname{esssup}_Q w^{-1}$). This condition also characterizes the boundedness of $\mathcal{M}_{\mathcal{D}}$ as an operator on $L^p(w)$, for a given $1 < p < \infty$.

In this paper, we will be interested in the two-weight context, in which the weights v and w in the base and in the target space are different. Following the customary convention, we will make the substitution $v = \sigma^{1-p}$. The argument of Muckenhoupt [11] shows that we have $\|\mathcal{M}_{\mathcal{D}}\|_{L^p(\sigma^{1-p}) \to L^{p,\infty}(w)} < \infty$ if and only if

$$[\sigma, w]_{A_p} = \sup_{Q \in \mathcal{D}(\mathbb{R}^n)} \langle w \rangle_Q \langle \sigma \rangle_Q^{p-1} < \infty. \tag{1.3}$$

Actually, we have the quantitative result $\|\mathcal{M}_{\mathcal{D}}\|_{L^p(\sigma^{1-p}) \to L^{p,\infty}(w)} = [\sigma, w]_{A_p}$. Quite unexpectedly - at least in the light of the aforementioned one-weight setting - the condition $[\sigma, w]_{A_p} < \infty$ does not characterize the boundedness of $\mathcal{M}_{\mathcal{D}}$ as an operator from $L^p(\sigma^{1-p})$ to $L^p(w)$. The correct characterization was given by Sawyer [18], by means of the so-called testing condition: we have $\|\mathcal{M}_{\mathcal{D}}\|_{L^p(\sigma^{1-p}) \to L^p(w)} < \infty$ if and only if

$$\int_Q \left(\mathcal{M}_{\mathcal{D}}(\sigma \chi_Q)\right)^p w \mathrm{d}x \leq C \int_Q \sigma \mathrm{d}x \qquad \text{for all } Q \in \mathcal{D}(\mathbb{R}^n),$$

where C depends only on p, w and σ. One of the drawbacks of this requirement is that it is hard to verify in practice, in contrast to the much simpler criterion $[\sigma, w]_{A_p} < \infty$. This observation leads to the very natural question about the modification of the latter simple criterion which would be *sufficient* for the L^p boundedness: the idea is to enlarge ("bump") one or two factors appearing under the supremum in (1.3). This problem has gained a lot of interest in the literature and has also been studied in the wider context of the boundedness of singular integral operators. The first result in this direction was the following theorem of Neugebauer [12].

Theorem 1.1. *Let (σ, w) be a pair of weights and let $1 < p < \infty$ be a fixed exponent. If there is $r > 1$ for which*

$$\sup_{Q\in\mathcal{D}(\mathbb{R}^n)} \langle w^r\rangle_Q^{1/r} \langle \sigma^r\rangle_Q^{(p-1)/r} < \infty,$$

then $\mathcal{M}_\mathcal{D}$ is bounded as an operator from $L^p(\sigma^{1-p})$ to $L^p(w)$.

The paper of Lerner [5] contains the following related result. Let ψ be a positive function on $(0, \infty)$, with $\int_0^\infty \frac{dt}{t\psi(t)} < \infty$ (for example, one can take a function which behaves as $\log^{1+\varepsilon}(\varepsilon + 1/t)$ as $t \to 0$ and $\log^{1+e}(\varepsilon + t)$ as $t \to \infty$). Then the requirement

$$\sup_{Q\in\mathcal{D}(\mathbb{R}^n)} \langle w\rangle_Q \langle \sigma\rangle^{p-1} \psi(\langle\sigma\rangle_Q) < \infty, \tag{1.4}$$

is sufficient for the L^p boundedness of $\mathcal{M}_\mathcal{D}$. So, it is enough to bump the factor depending on σ. See also [2, 4] for related Orlicz bump conditions, [3, 17, 19] for the so-called entropy bounds and the monograph [1] for much more in the direction.

The principal purpose of this paper is to establish a sharp version of Lerner's result. For technical reasons, it will be convenient for us to denote the expression $\langle\sigma\rangle^{p-1}\psi(\langle\sigma\rangle_Q)$, appearing in (1.4), by $(\gamma(\langle\sigma\rangle_Q))^{-1}$. Here is our main result.

Theorem 1.2. *Let $\gamma : (0, \infty) \to (0, \infty)$ be a convex and decreasing function such that*

$$\int_0^\infty t^{p-2}\gamma(t)\, dt < \infty. \tag{1.5}$$

If the pair (σ, w) satisfies the condition

$$[\sigma, w]_\gamma := \sup_{Q\in\mathcal{D}(\mathbb{R}^n)} \langle w\rangle_Q \big(\gamma(\langle\sigma\rangle_Q)\big)^{-1} < \infty, \tag{1.6}$$

then we have the sharp estimate

$$\|\mathcal{M}_\mathcal{D}\|_{L^p(\sigma^{1-p})\to L^p(w)} \leq \left(p[\sigma, w]_\gamma \int_0^\infty t^{p-2}\gamma(t)\, dt \right)^{1/p}. \tag{1.7}$$

Here by sharpness we mean that for any $\varepsilon > 0$ and any function γ as in the statement, there is a pair (σ, w) of weights satisfying (1.6) such that

$$\|\mathcal{M}_\mathcal{D}\|_{L^p(\sigma^{1-p})\to L^p(w)} > \left(p[\sigma, w]_\gamma \int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t \right)^{1/p} - \varepsilon.$$

This is indeed a sharp form of the aforementioned result of Lerner: we see that under the substitution $\gamma(s) = s^{1-p}/\psi(s)$, the integrability condition (1.5) becomes $\int_0^\infty \frac{dt}{t\gamma(t)} < \infty$.

Actually, we will manage to obtain the extension of the above result in the context of abstract probability measures equipped with tree-like structures.

Definition 1.3. 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 almost 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 $\mathcal{T}^{m+1} = \bigcup_{Q \in \mathcal{T}^m} C(Q)$.

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

Here and in what follows, two measurable subsets A, $B \subseteq X$ are said to be almost disjoint, if $\mu(A \cap B) = 0$.

An important example, which links the above 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 $\langle f \rangle_Q = \frac{1}{\mu(Q)} \int_Q f \mathrm{d}\mu$ is the average of f over Q with respect to the measure μ. One extends the notions of weights and weighted L^p spaces to this new context in an obvious manner. Here is the version of Theorem 1.2.

Theorem 1.4. *Let $\gamma : (0, \infty) \to (0, \infty)$ be a convex and decreasing function satisfying* (1.5)*. If the pair (σ, w) satisfies the condition*

$$[\sigma, w]_\gamma := \sup_{Q \in \mathcal{T}} \langle w \rangle_Q \big(\gamma(\langle \sigma \rangle_Q)\big)^{-1} < \infty, \tag{1.8}$$

then we have the sharp estimate

$$\|\mathcal{M}_\mathcal{T}\|_{L^p(\sigma^{1-p}) \to L^p(w)} \leq \left(p[\sigma, w]_\gamma \int_0^\infty t^{p-2} \gamma(t)\, dt \right)^{1/p}. \tag{1.9}$$

Here the sharpness of (1.9) is understood as in the dyadic context. Let us emphasize that the estimate is sharp for each individual probability space with a tree-like structure and each individual function γ satisfying the above requirements. From now on, we will assume that $[\sigma, w]_\gamma = 1$, which is allowed by a simple homogeneity argument: the function $\tilde{\gamma}(s) = \gamma(s) \cdot [\sigma, w]_\gamma$ inherits the structural properties and we have $[\sigma, w]_{\tilde{\gamma}} = 1$.

The remaining part of the paper is organized as follows. Section 2 contains the proofs of the two-weight estimates (1.7) and (1.9), which are obtained with the use of Bellman function method. The final part of the paper is devoted to the construction of extremal examples, which show that the constant cannot be improved.

2. Proof of (1.7) and (1.9)

The central role in this section is played by the function $B : (0,\infty)^4 \to \mathbb{R}$ given by

$$B(x,y,u,v) = y^p u - px^p v^{1-p} \int_0^{yv/x} t^{p-2}\gamma(t)\mathrm{d}t.$$

As we shall see, it enjoys certain concavity and size conditions which enable the extraction of the sharp weighted L^p bound. We start the analysis with the following statement.

Lemma 2.1. *For any parameter $y > 0$, the function*

$$\xi(x,v) = x^p v^{1-p} \int_0^{yv/x} t^{p-2}\gamma(t)\,dt$$

is convex on $(0,\infty)^2$.

Proof. By homogeneity, we may and do assume that $y = 1$. Note that if $\varphi : (0,\infty) \to [0,\infty)$ is convex and of class C^2, then the function $\zeta(x,v) = x\varphi(v/x)$ is convex on $(0,\infty)\times(0,\infty)$. Indeed, we compute that the Hessian matrix of ζ on $(0,\infty)^2$ is given by

$$D^2\zeta(x,v) = \begin{bmatrix} x^{-3}v^2\varphi''(v/x) & -x^{-2}v\varphi''(v/x) \\ -x^{-2}v\varphi''(v/x) & x^{-1}\varphi''(v/x) \end{bmatrix},$$

which is semipositive definite: the determinant of the full matrix is zero and the entry in the upper-left corner is nonnegative. Therefore, the assertion of the lemma will follow if we show that the function

$$\varphi(s) = s^{1-p} \int_0^s t^{p-2}\gamma(t)\mathrm{d}t$$

is convex on $(0,\infty)$. We compute directly that the second derivative of φ is given by

$$\varphi''(s) = p(p-1)s^{-p-1} \int_0^s t^{p-2}\gamma(t)\mathrm{d}t - ps^{-2}\gamma(s) + s^{-1}\gamma'(s).$$

Since γ is convex, we have $\gamma(t) \geq \gamma(s) + \gamma'(s)(t-s)$ for all $t > 0$. Applying this estimate under the above integral, we obtain the desired inequality $\varphi'' \geq 0$. □

We will also need the following monotonicity of B.

Lemma 2.2. *For any $(x,y,u,v) \in (0,\infty)^4$ such that $x \geq y$ and $u \leq \gamma(v)$, we have*

$$B(x,y,u,v) \geq B(x,x,u,v).$$

Proof. This is straightforward. We compute that

$$\begin{aligned}B_y(x,y,u,v) &= pxy^{p-2}\left(\frac{yu}{x}-\gamma\left(\frac{yv}{x}\right)\right)\\ &\leq pxy^{p-2}\left(\gamma(v)-\gamma\left(\frac{yv}{x}\right)\right)\leq pxy^{p-2}(\gamma(v)-\gamma(v))=0,\end{aligned}$$

since $y/x \leq 1$ and the function γ is nonincreasing. This gives the claim. □

Proof of (1.9). For the sake of clarity, we split the reasoning into a few separate parts.

Step 1. Notation. Let f be an arbitrary function on X and let w, σ be weights on X satisfying $[\sigma, w]_\gamma = 1$. We introduce the auxiliary functional sequences $(f_n)_{n\geq 0}$, $(g_n)_{n\geq 0}$, $(w_n)_{n\geq 0}$ and $(\sigma_n)_{n\geq 0}$ as follows. For any $n \geq 0$ and any $\omega \in X$,

$$\begin{aligned}f_n(\omega) &= \langle f\rangle_{Q_n(\omega)}, \quad g_n(\omega) = \max_{0\leq k\leq n} f_k(\omega),\\ w_n(\omega) &= \langle w\rangle_{Q_n(\omega)}, \quad \sigma_n(\omega) = \langle \sigma\rangle_{Q_n(\omega)},\end{aligned}$$

where $Q_n(\omega)$ is the unique element of $\mathcal{T}^n$ which contains ω. One can interpret these sequences from the probabilistic point of view: one easily checks that $(f_n)_{n\geq 0}$, $(w_n)_{n\geq 0}$ and $(\sigma_n)_{n\geq 0}$ are martingales induced by the filtration $(\mathcal{T}^n)_{n\geq 0}$, with the terminal variables equal to f, w and σ, respectively; furthermore, $(g_n)_{n\geq 0}$ is the maximal process associated with $(f_n)_{n\geq 0}$. Note that by Lebesgue's differentiation theorem (or martingale covergence theorem), we obtain $g_n \to \mathcal{M}_\mathcal{T} f$ and $w_n \to w$ almost surely as $n \to \infty$.

Step 2. Monotonicity. Now we will prove that the four functional sequences above combine nicely with B. More precisely, we will show that the sequence

$$\left(\int_X B(f_n, g_n, w_n, \sigma_n)\mathrm{d}\mu\right)_{n\geq 0}$$

is nonincreasing. To see this, fix an integer $n \geq 0$, let Q be an arbitrary element of $\mathcal{T}^n$ and let Q_1, Q_2, ..., Q_m be the collection of all children of Q in $\mathcal{T}^{n+1}$. Observe that the functions f_n, g_n, w_n and σ_n are constant on Q, while the functions f_{n+1}, g_{n+1}, w_{n+1} and σ_{n+1} are constant on each Q_j. Next, we have the inequality

$$B(f_{n+1}, g_{n+1}, w_{n+1}, \sigma_{n+1}) \leq B(f_{n+1}, g_n, w_{n+1}, \sigma_{n+1})$$

on each Q_j. Indeed, if $g_{n+1} = g_n$ on Q_j, then there is nothing to prove; otherwise, by the definition of g, we must have $g_{n+1} = f_{n+1}$ and hence the above estimate follows from Lemma 2.2. Integrating over Q_j and summing over j, we obtain the estimate

$$\int_Q B(f_{n+1}, g_{n+1}, w_{n+1}, \sigma_{n+1})\mathrm{d}\mu \leq \int_Q B(f_{n+1}, g_n, w_{n+1}, \sigma_{n+1})\mathrm{d}\mu.$$

Observe that B depends linearly on u; furthermore, g_n is constant on Q. Consequently,

$$\int_Q B(f_{n+1}, g_n, w_{n+1}, \sigma_{n+1})\mathrm{d}\mu = \int_Q B(f_{n+1}, g_n, w_n, \sigma_{n+1})\mathrm{d}\mu.$$

It remains to note that the integral on the right does is not bigger than $\int_Q B(f_n, g_n, w_n, \sigma_n)\mathrm{d}\mu$. This is due to the identities $f_n|_Q = \langle f_{n+1}\rangle_Q$, $\sigma_n|_Q = \langle \sigma_{n+1}\rangle_Q$, Lemma 2.1 and the formula for B. Summing over all $Q \in \mathcal{T}^n$ we obtain the desired monotonicity.

Step 3. Completion of the proof. By the previous step, for any $n \geq 0$ we have

$$\int_X B(f_n, g_n, w_n, \sigma_n)\mathrm{d}\mu \leq \int_X B(f_0, g_0, w_0, \sigma_0)\mathrm{d}\mu. \tag{2.1}$$

But we have $g_0 = f_0$ and $w_0 \leq \gamma(\sigma_0)$, so the monotonicity of γ gives

$$\begin{aligned} B(f_0, g_0, w_0, \sigma_0) &\leq f_0^p\gamma(\sigma_0) - pf_0^p\sigma_0^{1-p}\int_0^{\sigma_0} t^{p-2}\gamma(t)\mathrm{d}t \\ &\leq f_0^p\gamma(\sigma_0) - pf_0^p\sigma_0^{1-p}\int_0^{\sigma_0} t^{p-2}\gamma(\sigma_0)\mathrm{d}t = -\frac{f_0^p\gamma(\sigma_0)}{p-1} \leq 0. \end{aligned}$$

Hence the right-hand side of (2.1) is nonpositive and the estimate yields

$$\begin{aligned} \int_X g_n^p w_n \mathrm{d}\mu &\leq p\int_X \left(f_n^p\sigma_n^{1-p}\int_0^{g_n\sigma_n/f_n} t^{p-2}\gamma(t)\mathrm{d}t\right)\mathrm{d}\mu \\ &\leq p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t \cdot \int_X f_n^p\sigma_n^{1-p}\mathrm{d}\mu \\ &\leq p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t \cdot \int_X f^p\sigma^{1-p}\mathrm{d}\mu. \end{aligned}$$

Here in the last passage we used the conditional version of Jensen's inequality, applied to the convex function $(x,u) \mapsto x^p u^{1-p}$. Letting $n \to \infty$ and exploiting Fatou's lemma, we get the claim (see the limiting behavior of the sequences $(g_n)_{n\geq 0}$ and $(w_n)_{n\geq 0}$, described at the end of Step 1 above). □

Proof of (1.7). This follows from a straightforward dilation argument. By (1.9), we get

$$\|\mathcal{M}_\mathcal{D}(f\chi_Q)\|_{L^p(w)} \leq \left(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p} \|f\chi_Q\|_{L^p(\sigma^{1-p})},$$

for an arbitrary cube $Q \in \mathcal{D}(\mathbb{R}^n)$. Thus, by Lebesgue's monotone convergence theorem,

$$\|\mathcal{M}_\mathcal{D}(f\chi_{[0,\infty)^n})\|_{L^p(w)} \leq \left(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p} \|f\chi_{[0,\infty)^n}\|_{L^p(\sigma^{1-p})}.$$

The same estimate holds if we replace $[0,\infty)$ by any product of the form $I_1 \times I_2 \times \ldots \times I_n$, where each I_j is either $[0,\infty)$ or $(-\infty, 0]$. Summing these estimates over all such products, we obtain the claim. □

3. Sharpness

Now we will show that the constants in (1.7) and (1.9) cannot be improved. Actually, as we will briefly explain now, we may restrict ourselves to the sharpness of the localized estimate (1.9). Indeed, suppose that we have shown that the constant $\big(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\big)^{1/p}$ is optimal for the probability space $(X,\mu)=([0,1)^n,|\cdot|)$ with the dyadic lattice. So, for any $\varepsilon>0$ there is a function $f:[0,1)^n\to\mathbb{R}$ and a pair (σ,w) of weights on $[0,1)^n$ satisfying $[\sigma,w]_\gamma=1$, for which

$$\|\mathcal{M}_\mathcal{T}f\|_{L^p(w)}>\left(\left(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p}-\varepsilon\right)\|f\|_{L^p(\sigma^{1-p})}.$$

We extend f, σ, w to functions $\tilde{f}$, $\tilde{\sigma}$, $\tilde{w}$ on $\mathbb{R}^n$, setting $\tilde{f}(x)=0$, $\tilde{\sigma}(x)=\langle\sigma\rangle_{[0,1)^n}$, $\tilde{w}(x)=\langle w\rangle_{[0,1)^n}$ for $x\notin[0,1)^n$. Then the condition $[\tilde{\sigma},\tilde{w}]_\gamma=1$ is preserved, since for $Q\not\subseteq[0,1)^n$ we have

$$\langle\tilde{w}\rangle_Q\big(\gamma(\langle\tilde{\sigma}\rangle_Q)\big)^{-1}=\langle w\rangle_{[0,1)^n}\big(\gamma(\langle\sigma\rangle_{[0,1)^n})\big)^{-1}\leq 1.$$

Furthermore,

$$\begin{aligned}\|\mathcal{M}_\mathcal{D}\tilde{f}\|_{L^p(\tilde{w})}\geq\|\mathcal{M}_\mathcal{T}f\|_{L^p(w)}&>\left(\left(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p}-\varepsilon\right)\|f\|_{L^p(\sigma^{1-p}))}\\&=\left(\left(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p}-\varepsilon\right)\|\tilde{f}\|_{L^p(\tilde{\sigma}^{1-p}))}.\end{aligned}$$

So, from now on we focus on (1.9). Fix an arbitrary probability space (X,μ), a tree-like structure $\mathcal{T}$ and an arbitrary function γ satisfying the conditions listed in the statement of Theorem 1.2. It is convenient to split the reasoning into a few parts.

Step 1. Construction. We start with the following technical fact, which can be found in Melas' paper [6].

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 almost 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. First, we let $A_0=X$. Next, suppose that we have constructed the set A_n and assume further, that this set is a union of pairwise almost disjoint elements of $\mathcal{T}$, called the *atoms* of A_n. (Note that 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 $\beta=(1+\delta)^{-1}$, obtaining a subfamily $F(Q)$. We define the next set of the sequence, putting $A_{n+1}=\bigcup_Q\bigcup_{Q'\in F(Q)}Q'$, 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 almost disjoint elements of $\mathcal{T}$. We call these elements the atoms of A_{n+1} and conclude the description of the induction step.

It follows at once from the above construction that if Q is an atom of A_n, then for any $m \geq n$ we have $\mu(Q \cap A_m) = \mu(Q)(1+\delta)^{n-m}$ and hence

$$\mu(Q\cap(A_m \setminus A_{m+1})) = \mu(Q\cap A_m) - \mu(Q\cap A_{m+1}) = \mu(Q)(1+\delta)^{n-m-1}\delta. \quad (3.1)$$

Now we introduce some additional geometric objects. Suppose that $\tilde{\gamma} : (0,\infty) \to (0,\infty)$ is a C^1 convex function lying strictly below γ. Next, fix parameters $S > s > 0$, $\varepsilon > 0$ and let N be a large positive integer. Let $\delta > 0$ be uniquely determined by the requirement $s(1+\delta)^N = S$. For $n = 0, 1, 2, \ldots, N$ we define $s_n = s(1+\delta)^n + \varepsilon$; furthermore, let

$$t_n = \tilde{\gamma}(s_n) + \frac{\tilde{\gamma}(s_n) - \tilde{\gamma}(s_{n+1})}{\delta}, \qquad n = 0, 1, 2, \ldots, N-1.$$

Note that if N is taken sufficiently large, then δ can be made as small as we wish and hence we may assume that the piecewise linear curve joining the points $(s_n, \tilde{\gamma}(s_n))$, $n = 0, 1, 2, \ldots, N$, lies entirely below the graph of γ. See Figure 1 below.

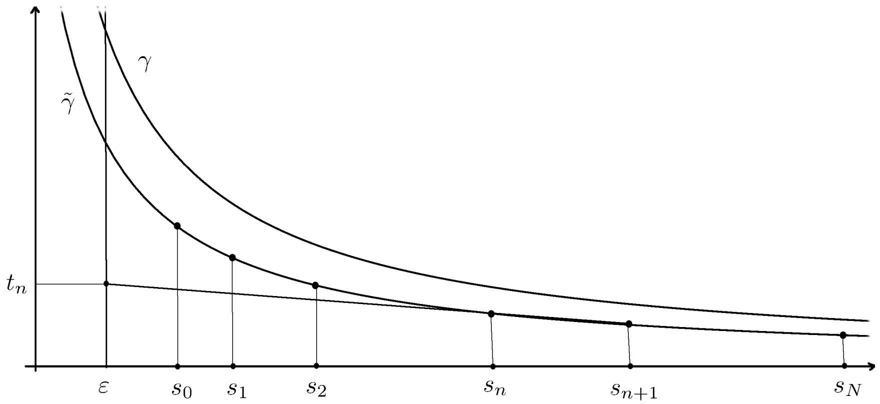

FIGURE 1. The geometric parameters used in the construction.

We are ready to define w, σ and f. Set

$$w = \tilde{\gamma}(s_N)\chi_{A_N} + \sum_{n=0}^{N-1} t_n \chi_{A_n \setminus A_{n+1}} \qquad \text{and} \qquad f = \sigma = \tilde{\gamma}(s_N)\chi_{A_N} + \varepsilon\chi_{X\setminus A_N}.$$

Step 2. The verification of the bump condition. We start with the observation that if Q is an atom of A_n, then, by (3.1),

$$\begin{aligned}\langle w\rangle_Q &= \tilde{\gamma}(s_N)(1+\delta)^{N-n}+\sum_{m=n}^{N-1} t_m(1+\delta)^{n-m-1}\delta\\ &= \tilde{\gamma}(s_N)(1+\delta)^{N-n}+\sum_{m=n}^{N-1}\big(\tilde{\gamma}(s_m)\delta+\tilde{\gamma}(s_m)-\tilde{\gamma}(s_{m+1})\big)(1+\delta)^{n-m-1}\\ &= \tilde{\gamma}(s_N)(1+\delta)^{N-n}+\sum_{m=n}^{N-1}\tilde{\gamma}(s_m)(1+\delta)^{n-m}-\sum_{m=n}^{N-1}\tilde{\gamma}(s_{m+1})(1+\delta)^{n-m-1}\\ &= \tilde{\gamma}(s_n)\end{aligned} \tag{3.2}$$

and

$$\langle\sigma\rangle_Q=\frac{1}{\mu(Q)}\int_{Q\cap A_N} s(1+\delta)^N\mathrm{d}\mu+\varepsilon = s(1+\delta)^n+\varepsilon=s_n. \tag{3.3}$$

We are ready for the verification of the bump condition $[\sigma,w]_\gamma\leq 1$: we will show that $\langle w\rangle_Q\leq\gamma(\langle\sigma\rangle_Q)$ for all $Q\in\mathcal{T}$. For an arbitrary Q, we have three possibilities.

1° We have $Q\subseteq A_N$, up to a set of measure zero (that is, $\mu(Q\setminus A_N)=0$). Then $\langle w\rangle_Q=\tilde{\gamma}(s_N)=\tilde{\gamma}(\langle\sigma\rangle_Q)\leq\gamma(\langle\sigma\rangle_Q)$, as desired.

2° There is $n\leq N$ such that $Q\subseteq A_{n-1}$ and $Q\cap A_n=\emptyset$, up to a set of measure zero (precisely, $\mu(Q\setminus A_{n-1})=\mu(Q\cap A_n)=0$). Then $\langle w\rangle_Q=t_{n-1}<\gamma(\varepsilon)=\gamma(\langle\sigma\rangle)$, so the required condition holds.

3° There is $n<N$ such that $Q\subseteq A_{n-1}$ (up to a set of measure zero) and $\mu(Q\setminus A_n)<\mu(Q)$: this corresponds to the case in which Q has nontrivial intersections with A_n and $A_{n-1}\setminus A_n$. Then, by the very definition of w and σ, we have

$$\int_{Q\setminus A_n} w\mathrm{d}\mu=t_{n-1}\mu(Q\setminus A_n)\qquad\text{and}\qquad\int_{Q\setminus A_n}\sigma\mathrm{d}\mu=\varepsilon\mu(Q\setminus A_n).$$

Furthermore, if Q' is any atom of A_n contained in Q, then by (3.2) and (3.3) we have $\langle w\rangle_{Q'}=\tilde{\gamma}(s_n)$ and $\langle\sigma\rangle_{Q'}=s_n$. Summing over all such Q', we get

$$\int_{Q\cap A_n} w\mathrm{d}\mu=\tilde{\gamma}(s_n)\mu(Q\cap A_n)\qquad\text{and}\qquad\int_{Q\cap A_n}\sigma\mathrm{d}\mu=s_n\mu(Q),$$

which implies

$$\int_Q w\mathrm{d}\mu=\int_{Q\setminus A_n} w\mathrm{d}\mu+\int_{Q\cap A_n} w\mathrm{d}\mu=t_{n-1}\mu(Q\setminus A_n)+\tilde{\gamma}(s_n)\mu(Q\cap A_n)$$

and

$$\int_Q\sigma\mathrm{d}\mu=\int_{Q\setminus A_n}\sigma\mathrm{d}\mu+\int_{Q\cap A_n}\sigma\mathrm{d}\mu=\varepsilon\mu(Q\setminus A_n)+s_n\mu(Q\cap A_n)$$

That is, the point $(\langle\sigma\rangle_Q, \langle w\rangle_Q)$ lies on the line segment with endpoints (ε, t_{n-1}) and $(s_n, \tilde{\gamma}(s_n))$. However, this line segment lies below the graph of the function γ: this is guaranteed by taking N sufficiently large (see the discussion above Figure 1). This implies $\langle w\rangle_Q \leq \gamma(\langle\sigma\rangle_Q)$, and hence the bump condition holds true. Before we proceed, let us note that by (3.2) and (3.3) applied to $Q = X$, we have $\langle w\rangle_Q = \tilde{\gamma}(\langle\sigma\rangle_Q)$. Hence, if $\tilde{\gamma}$ is chosen sufficiently close to γ, then the quantity $[\sigma, w\rangle]_\gamma$ can be made arbitrarily close to 1.

Step 3. Completion of the proof. We proceed to the study the behavior of the ratio $\|\mathcal{M}_\mathcal{T} f\|_{L^p(w)}/\|f\|_{L^p(\sigma^{1-p})}$. The denominator equals

$$\left(\int_X f^p\sigma^{1-p}\mathrm{d}\mu\right)^{1/p} = \left(\int_X \sigma\mathrm{d}\mu\right)^{1/p} = (s+\varepsilon)^{1/p}. \tag{3.4}$$

The main technical difficulty lies in the analysis of $\|\mathcal{M}_\mathcal{T} f\|_{L^p(w)}$. We start with the observation that if Q is an atom of A_n, then, by (3.3), $\langle f\rangle_Q = \langle\sigma\rangle_Q = s_n$. Hence, by the definition of the maximal operator, we have

$$\mathcal{M}_\mathcal{T} f \geq s_N\chi_{A_N} + \sum_{n=0}^{N-1} s_n\chi_{A_n\setminus A_{n+1}}.$$

Consequently,

$$\begin{aligned}
&\|\mathcal{M}_\mathcal{T} f\|^p_{L^p(w)}\\
&\geq s_N^p\tilde{\gamma}(s_N)\mu(A_N) + \sum_{n=0}^{N-1} s_n^p t_n\mu(A_n\setminus A_{n+1})\\
&= s_N^p\tilde{\gamma}(s_N)(1+\delta)^{-N} + \sum_{n=0}^{N-1} s_n^p\left(\tilde{\gamma}(s_n) + \frac{\tilde{\gamma}(s_n)-\tilde{\gamma}(s_{n+1})}{\delta}\right)\delta(1+\delta)^{-n-1}\\
&= (S+\varepsilon)^p\tilde{\gamma}(S+\varepsilon)\cdot\frac{s}{S} + \sum_{n=0}^{N-1} s_n^p\left(\tilde{\gamma}(s_n) + \frac{\tilde{\gamma}(s_n)-\tilde{\gamma}(s_{n+1})}{\delta}\right)\delta(1+\delta)^{-n-1}.
\end{aligned}$$

Now we perform a limiting procedure and let $N\to\infty$; then the parameter δ goes to zero. Since $s_{n+1}-s_n = s(1+\delta)^n\delta = (s_n-\varepsilon)\delta$ and

$$\frac{\tilde{\gamma}(s_n)-\tilde{\gamma}(s_{n+1})}{\delta} = -\tilde{\gamma}'(s_n)(s_n-\varepsilon) + O(\delta),$$

we easily check that the above expression converges to

$$(S+\varepsilon)^p\tilde{\gamma}(S+\varepsilon)\cdot\frac{s}{S} + s\int_s^S (\tilde{\gamma}(t+\varepsilon)-\tilde{\gamma}'(t+\varepsilon)t)\cdot\frac{(t+\varepsilon)^p}{t^2}\mathrm{d}t.$$

Next, if we let $\varepsilon\to 0$, this tends further to

$$\begin{aligned}
sS^{p-1}\tilde{\gamma}(S) + s\int_s^S \frac{\tilde{\gamma}(t)-\tilde{\gamma}'(t)t}{t^2}\cdot t^p\mathrm{d}t &= s^p\tilde{\gamma}(s) + ps\int_s^S t^{p-2}\tilde{\gamma}(t)\mathrm{d}t\\
&> ps\int_s^S t^{p-2}\tilde{\gamma}(t)\mathrm{d}t,
\end{aligned}$$

where the latter equality follows from integration by parts. Thus, taking ε small enough and then N sufficiently large, we may make $\|\mathcal{M}_{\mathcal{T}}f\|_{L^p(w)}$ bigger than $\left(ps\int_s^S t^{p-2}\tilde{\gamma}(t)\mathrm{d}t\right)^{1/p}$. Combining this with (3.4) and noting that the auxiliary C^1 convex function $\tilde{\gamma}$ was chosen arbitrarily (which in particular implies that $[\sigma, w]_\gamma$ is as close to 1 as we wish, see the previous step), the best constant in the L^p estimate cannot be smaller than

$$\left(p\int_s^S t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p}.$$

It remains to observe that if we let $s \to 0$ and $S \to \infty$, then the latter expression converges to $\left(p\int_0^\infty t^{p-2}\gamma(t)\mathrm{d}t\right)^{1/p}$. This proves the desired sharpness.

References

[1] D. V. Cruz-Uribe, J. M. Martell, C. Pérez, Weights, extrapolation and the theory of Rubio de Francia. Operator Theory: Advances and Applications, 215. Birkhäuser/Springer Basel AG, Basel, 2011.

[2] M. T. Lacey, *On the separated bumps conjecture for Calderón-Zygmund operators*, Hokkaido Math. J. **45** (2016), 223–242.

[3] M. T. Lacey and S. Spencer, *On entropy bumps for Calderón-Zygmund operators*, Concr. Oper. **2** (2015), 47–52.

[4] K. Li, *Two weight inequalities for bilinear forms*, Collect. Math. **68** (2017), 129–144.

[5] A. Lerner, *On separated bump conditions for Calderon-Zygmund operators*, available at https://arxiv.org/abs/2008.05866.

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

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

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

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

[10] 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.

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

[12] C. J. Neugebauer, *Inserting A_p-weights*, Proc. Amer. Math. Soc. **87** (1983), 644–648.

[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] A. Osękowski, *Best constants in Muckenhoupt's inequality*, Ann. Acad. Sci. Fenn. Math. **42** (2017), 889–904.

[16] A. Osękowski, M. Rapicki, *Sharp Lorentz-norm estimates for dyadic-like maximal operators*, Studia Math. **257** (2021), 87–110.

[17] R. Rahm and S. Spencer, *Entropy bumps and another sufficient condition for the two-weight boundedness of sparse operators*, Israel J. Math. **223** (2018), no. 1, 197–204.

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

[19] S. Treil and A. Volberg, *Entropy conditions in two weight inequalities for singular integral operators*, Adv. Math. **301** (2016), 499–548.

Adam Osękowski
Faculty of Mathematics, Informatics and Mechanics
University of Warsaw
Banacha 2, 02-097 Warsaw
Poland
e-mail: ados@mimuw.edu.pl