Mathematical
Inequalities
& Applications
www.ele-math.com

AN EXTENSION OF A HARDY'S INEQUALITY AND ITS APPLICATIONS

ŁUKASZ KAMIŃSKI AND ADAM OSĘKOWSKI

Submitted to Math. Inequal. Appl.

Abstract. Let $p>0$ and let $s\leq q$ be fixed parameters. The paper contains the proof of the sharp Hardy-type inequality

$$\sum_{n=1}^{\infty}\lambda_n\left(\frac{\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^{-p+s}a_n^{-s}$$
$$\leq\left(1+\frac{1}{p}\right)^{q-s}\sum_{n=1}^{\infty}\lambda_n\left(\frac{\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^{-p+q}a_n^{-q}$$

for any sequences $(a_n)_{n=1}^{\infty}$, $(\lambda_n)_{n=1}^{\infty}$ of positive numbers. The approach exploits dynamic programming-type techniques and rests on the identification of the explicit Bellman function associated with the estimate. As applications, related estimates for Hardy operators in $\mathbb{R}^d$ and harmonic maximal operators on probability spaces are obtained.

1. Introduction

A celebrated Hardy's inequality [7] asserts that for any $1<p<\infty$ and any non-negative numbers a_1, a_2, $\ldots$ we have the estimate

$$\sum_{n=1}^{\infty}\left(\frac{a_1+a_2+\ldots+a_n}{n}\right)^p\leq\left(\frac{p}{p-1}\right)^p\sum_{n=1}^{\infty}a_n^p. \tag{1}$$

Furthermore, the constant $(p/(p-1))^p$ is the best possible. There are many different proofs of this fundamental inequality, this result has been extended in numerous directions and applied in various contexts. See e.g. the monograph [8] by Hardy, Littlewood and Pólya or consult the more recent books [10] by Kufner and Opic or [11] by Kufner and Persson. For example, one can study the following weighted version of (1): for any a_1, a_2, $\ldots$ as before and any positive λ_1, λ_2, $\ldots$ we have

$$\sum_{n=1}^{\infty}\left(\frac{\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n}{\lambda_1+\lambda_2+\ldots+\lambda_n}\right)^p\leq\left(\frac{p}{p-1}\right)^p\sum_{n=1}^{\infty}\lambda_na_n^p. \tag{2}$$

Again, the constant $(p/(p-1))^p$ cannot be improved, since the choice $\lambda_1=\lambda_2=\ldots=1$ reduces the estimate to (1), which is sharp (see [8]).

Mathematics subject classification (2010): 26D10, 26D15, 42B25.
Keywords and phrases: Hardy inequality; weight; best constant; Hardy operator; harmonic maximal operator.

We will be interested in a version of (2) for negative exponents. As proved by Nikolidakis in [16], for any $p \geq q > 0$ and any sequences $(a_n)_{n=1}^{\infty}$, $(\lambda_n)_{n=1}^{\infty}$ of positive numbers, we have the sharp bound

$$\begin{aligned} &\sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^{-p} \\ &\qquad \leq \left(1 + \frac{1}{p}\right)^q \sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^{-p+q} a_n^{-q}. \end{aligned} \tag{3}$$

This estimate was applied in [16] to obtain some new properties of Muckenhoupt weights on the positive halfline. We will generalize the inequality to the following.

THEOREM 1. *Let $p > 0$ and let $s < q$. Then for any sequences $(a_n)_{n=1}^{\infty}$, $(\lambda_n)_{n=1}^{\infty}$ of positive numbers, we have*

$$\begin{aligned} &\sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^{-p+s} a_n^{-s} \\ &\qquad \leq \left(1 + \frac{1}{p}\right)^{q-s} \sum_{n=1}^{\infty} \lambda_n \left(\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \right)^{-p+q} a_n^{-q}. \end{aligned} \tag{4}$$

The constant $\left(1 + \frac{1}{p}\right)^{q-s}$ is the best possible.

If $s > 0$, then (4) follows from (3) (by a simple use of the Hölder inequality), but for negative s this does not seem to be the case. A standard approximation leads to the following integral weighted bound.

THEOREM 2. *Let $p > 0$ and let $s < q$. In addition, suppose that f is a positive function on $[0,\infty)$ and let μ be a measure on $[0,\infty)$. Then we have the sharp estimate*

$$\int_0^{\infty} \left(\fint_{[0,t]} f d\mu \right)^{-p+s} f(t)^{-s} d\mu \leq \left(1 + \frac{1}{p}\right)^{q-s} \int_0^{\infty} \left(\fint_{[0,t]} f d\mu \right)^{-p+q} f(t)^{-q} d\mu, \tag{5}$$

where $\fint_{[0,t]} f d\mu = \dfrac{1}{\mu([0,t])} \displaystyle\int_0^t f d\mu$ denotes the weighted average of f on $[0,t]$.

The above inequalities can be applied in several directions. First, note that by a limiting argument, the estimate (5) leads to a classical result of Knopp [9]. Namely, take $s = 0$, $q = p$, substitute $g = |f|^{-p}$ and let $p \to 0$ to obtain

$$\int_0^{\infty} \exp\left(\fint_{[0,t]} \ln g \mathrm{d}\mu \right) \mathrm{d}\mu \leq e \int_0^{\infty} g \mathrm{d}\mu.$$

The constant e is the best possible already in the special case when μ is the Lebesgue measure [8]. To describe two further applications of (4) and (5), we need more definitions and notation, the details will be presented in later sections of this paper.

A few words about the proof are in order. Our approach rests on dynamic programming arguments [1] and can be easily modified to yield other inequalities of the above type. More specifically, we will link the validity of the estimate (4) to a certain special function which enjoys an appropriate monotonicity-type condition. From this point of view, the method seems to be related to the so-called Bellman function technique, a powerful tool used widely in probability and harmonic analysis (see e.g. [2, 15, 17] and consult the references therein).

The estimate (4) is established in the next section. In Section 3, using (5), we obtain sharp estimates for d-dimensional Hardy operators. The final part of the paper is devoted to another application: we obtain estimates for harmonic maximal operators in the context of probability spaces with tree-like structures.

2. Proof of Theorem 1

For the sake of brevity, we will use the notation $\Lambda_n = \sum_{k=1}^n \lambda_k$ and $A_n = \sum_{k=1}^n \lambda_k a_k$ for $n = 1, 2, \ldots$. Given $p > 0$ and $s < q$, we are interested in the optimal constant $C_{p,q,s}$ in the estimate

$$\sum_{n=1}^{\infty} \lambda_n \left(\frac{A_n}{\Lambda_n}\right)^{-p+s} a_n^{-s} \leq C_{p,q,s} \sum_{n=1}^{\infty} \lambda_n \left(\frac{A_n}{\Lambda_n}\right)^{-p+q} a_n^{-q}, \tag{6}$$

to be valid for any positive numbers $a_1, a_2, \ldots$ and $\lambda_1, \lambda_2, \ldots$. Our starting point is the introduction of a certain abstract special Bellman function on $(0,\infty)^2$ related to the above estimate. Namely, for a given $C > 0$, put

$$\mathbb{B}_C(a,\lambda) = \sup\left\{\sum_{n=1}^{\infty} \lambda_n \left(\frac{A_n}{\Lambda_n}\right)^{-p+s} a_n^{-s} - C\sum_{n=1}^{\infty} \lambda_n \left(\frac{A_n}{\Lambda_n}\right)^{-p+q} a_n^{-q}\right\} \in (-\infty,\infty].$$

Here the supremum is taken over all sequences $(a_n)_{n=1}^{\infty}$ of positive numbers such that $a_1 = a$, and all sequences $(\lambda_n)_{n=1}^{\infty}$ of *nonnegative* numbers, possessing only a finite number of nonzero terms, such that $\lambda_1 = \lambda$. Here the assumption on the vanishing of almost all terms of $(\lambda_n)_{n=1}^{\infty}$ guarantees that there is no problem with the convergence of the series under the supremum.

Actually, it will be more convenient to work with a slightly different function

$$\mathbb{B}_C^0(a,\lambda) = \sup\left\{\sum_{n=2}^{\infty} \lambda_n \left(\frac{A_n}{\Lambda_n}\right)^{-p+s} a_n^{-s} - C\sum_{n=2}^{\infty} \lambda_n \left(\frac{A_n}{\Lambda_n}\right)^{-p+q} a_n^{-q}\right\},$$

where the supremum is taken over the same parameters as previously. The difference in comparison to $\mathbb{B}_C$ is that both sums under the supremum start with $n = 2$, not $n = 1$. It is easy to see that the functions $\mathbb{B}_C$ and $\mathbb{B}_C^0$ are linked by the identity

$$\mathbb{B}_C(a,\lambda) = \mathbb{B}_C^0(a,\lambda) + \lambda a^{-p}(1-C). \tag{7}$$

Now the general (and a little informal) philosophy behind the forthcoming arguments is the following. The key fact is a certain 'self-similarity' of the inequality (6), which translates into an appropriate monotonicity condition for $\mathbb{B}_C$. The optimal constant $C_{p,q,s}$, by the very definition, is the least C for which the function $\mathbb{B}_C$ takes values in the nonpositive halfline (in particular, then $\mathbb{B}_C$, and hence also $\mathbb{B}^0_C$, are finite). It is natural to expect that among all functions B_C, corresponding to differrent choices of $C \geq C_{p,q,s}$, the function $\mathbb{B}_{C_{p,q,s}}$ will play a distinguished role; for example, the aforementioned monotonicity condition should degenerate appropriately. This observation, combined with a homogeneity-type property of $\mathbb{B}_C$, will allow us to *guess* the explicit formula for the Bellman function. Then we will prove rigorously that the obtained candidates B_C and B^0_C coincide with $\mathbb{B}$ and $\mathbb{B}^0$.

For the sake of clarity, we have decided to split the reasoning into six separate parts.

Step 1. Homogeneity. Fix $C > 0$. We start with the identities

$$\mathbb{B}_C(a,\lambda) = \lambda a^{-p}\mathbb{B}_C(1,1) \quad \text{and} \quad \mathbb{B}^0_C(a,\lambda) = \lambda a^{-p}\mathbb{B}^0_C(1,1). \tag{8}$$

In the light of (7), it is enough to prove the first equality. Pick arbitrary sequences $(a_n)_{n=1}^\infty$, $(\lambda_n)_{n=1}^\infty$ as in the definition of $\mathbb{B}_C(1,1)$. Then $(a\cdot a_n)_{n=1}^\infty$ and $(\lambda\cdot\lambda_n)_{n=1}^\infty$ have all the properties listed at the definition of $\mathbb{B}_C(a,\lambda)$, so

$$\mathbb{B}_C(a,\lambda) \geq \lambda a^{-p}\sum_{n=1}^\infty \lambda_n\left(\frac{A_n}{\Lambda_n}\right)^{-p+s} a_n^{-s} - C\sum_{n=1}^\infty \lambda_n\left(\frac{A_n}{\Lambda_n}\right)^{-p+q} a_n^{-q},$$

so taking the supremum over all $(a_n)_{n=1}^\infty$ and $(\lambda_n)_{n=1}^\infty$ as above gives

$$\mathbb{B}_C(a,\lambda) \geq \lambda a^{-p}\mathbb{B}_C(1,1).$$

The proof of the reverse inequality is similar: one starts with arbitrary sequences $(a_n)_{n=1}^\infty$, $(\lambda_n)_{n=1}^\infty$ as in the definition of $\mathbb{B}_C(a,\lambda)$ and considers their scaled versions $(a_n/a)_{n=1}^\infty$ and $(\lambda_n/\lambda)_{n=1}^\infty$.

Step 2. Self-similarity and monotonicity. Pick arbitrary positive numbers a,λ and two auxiliary parameters a', λ' satisfying $\lambda' > \lambda$ and $a' > a\lambda/\lambda'$. Take any sequences $(a_n)_{n=1}^\infty$ and $(\lambda_n)_{n=1}^\infty$ with $a_1 = a$, $a_2 = (a'\lambda' - a\lambda)/(\lambda'-\lambda)$ and $\lambda_1 = \lambda$, $\lambda_2 = \lambda' - \lambda$. Then by the very definition of $\mathbb{B}^0_C$,

$$\mathbb{B}^0_C(a,\lambda) \geq \sum_{n=2}^\infty \lambda_n\left(\frac{A_n}{\Lambda_n}\right)^{-p+s} a_n^{-s} - C\sum_{n=2}^\infty \lambda_n\left(\frac{A_n}{\Lambda_n}\right)^{-p+q} a_n^{-q}. \tag{9}$$

Now rewrite the right-hand side as $I + II$, where

$$I = \sum_{n=3}^\infty \lambda_n\left(\frac{A_n}{\Lambda_n}\right)^{-p+s} a_n^{-s} - C\sum_{n=3}^\infty \lambda_n\left(\frac{A_n}{\Lambda_n}\right)^{-p+q} a_n^{-q}$$

and

$$\begin{aligned} II &= \lambda_2\left(\frac{A_2}{\Lambda_2}\right)^{-p+s} a_2^{-s} - C\lambda_2\left(\frac{A_2}{\Lambda_2}\right)^{-p+q} a_2^{-q} \\ &= (\lambda'-\lambda)(a')^{-p}\left[\left(\frac{a'\lambda'-a\lambda}{a'\lambda'-a'\lambda}\right)^{-s} - C\left(\frac{a'\lambda'-a\lambda}{a'\lambda'-a'\lambda}\right)^{-q}\right]. \end{aligned}$$

Set $a_1' = a'$, $\lambda_1' = \lambda'$ and $a_n' = a_{n+1}$, $\lambda_n' = \lambda_{n+1}$ for $n \geq 2$. Furthermore, let $(A_n')_{n=1}^{\infty}$, $(\Lambda_n')_{n=1}^{\infty}$ be the sequences of partial sums, built on $(a_n')_{n=1}^{\infty}$ and $(\lambda_n')_{n=1}^{\infty}$. Then

$$I = \sum_{n=2}^{\infty} \lambda_n' \left(\frac{A_n'}{\Lambda_n'}\right)^{-p+s} (a_n')^{-s} - C \sum_{n=2}^{\infty} \lambda_n' \left(\frac{A_n'}{\Lambda_n'}\right)^{-p+q} (a_n')^{-q},$$

so taking the supremum over the sequences $(a_2', a_3', \ldots) = (a_3, a_4, \ldots)$ and $(\lambda_2', \lambda_3', \ldots) = (\lambda_3, \lambda_4, \ldots)$, one gets $\mathbb{B}_C^0(a', \lambda')$. Note that the expression II does not depend on these sequences, so (9) yields

$$\begin{aligned} &\mathbb{B}_C^0(a,\lambda) \\ &\geq \mathbb{B}_C^0\left(a',\lambda'\right) + (\lambda'-\lambda)(a')^{-p}\left[\left(\frac{a'\lambda'-a\lambda}{a'\lambda'-a'\lambda}\right)^{-s} - C\left(\frac{a'\lambda'-a\lambda}{a'\lambda'-a'\lambda}\right)^{-q}\right]. \end{aligned} \tag{10}$$

Let us invoke (8), divide both sides by λa^{-p} and substitute $u = \lambda'/\lambda > 1$, $v = a'/a > \lambda/\lambda'$. As the result, the above inequality becomes

$$\mathbb{B}_C^0(1,1) \geq \mathbb{B}_C^0(1,1)uv^{-p} + (u-1)v^{-p}\left[\left(\frac{uv-1}{uv-v}\right)^{-s} - C\left(\frac{uv-1}{uv-v}\right)^{-q}\right].$$

Summarizing, for any $C > 0$, any $u > 1$ and any $v > u^{-1}$ the function $\mathbb{B}_C^0$ satisfies

$$\mathbb{B}_C^0(1,1)v^p \geq \mathbb{B}_C^0(1,1)u + (u-1)\left[\left(\frac{uv-1}{uv-v}\right)^{-s} - C\left(\frac{uv-1}{uv-v}\right)^{-q}\right]. \tag{11}$$

Step 3. The guess for the best constant $C_{p,q,s}$. Now suppose that the inequality (6) holds true with some constant C. Then, as we have observed above, the function $\mathbb{B}_C^0$ is finite. Therefore, if we fix $u > 1$ and let $v = u^{1/p}$ in (11), the terms involving $\mathbb{B}_C^0(1,1)$ cancel out and we obtain

$$C \geq \left(\frac{uv-1}{uv-v}\right)^{q-s} = \left(\frac{u^{1+1/p}-1}{(u-1)u^{1/p}}\right)^{q-s}. \tag{12}$$

Now a direct differentiation shows that the function $u \mapsto (u^{1+1/p}-1)/((u-1)u^{1/p})$ is decreasing on $u \in (1,\infty)$. Indeed, the derivative is equal to

$$u \mapsto (u-1)^{-2}u^{-1/p}\Big(1 - (u^{-1})^{-1/p} - (-1/p)(1-u^{-1})\Big) \leq 0,$$

where the latter bound is due to the mean-value property of the convex function $t \mapsto t^{-1/p}$, $t > 0$. Consequently, the estimate (12) is optimized by letting $u \to 1$: as the result of this passage, we obtain

$$C \geq \left(1 + \frac{1}{p}\right)^{q-s}.$$

Thus, we have formally proved that the best constant in (6) is at least the right-hand side above. We *conjecture* that this expression is actually the optimal constant in (6):

$$C_{p,q,s}=\left(1+\frac{1}{p}\right)^{q-s}. \tag{13}$$

From now on, we will assume that the constant C is given by the right-hand side above.

Step 4. On the guess for $\mathbb{B}_C^0(1,1)$. Let us exploit (11) with $u=1+\varepsilon$ and $v=1+\kappa\varepsilon$, where ε is a small positive number and $\kappa>-1$ is an arbitrary parameter (note that $v\geq u^{-1}$, as required in (11)). If the inequality (6) holds true with the constant C, then $\mathbb{B}_C^0(1,1)$ is finite; hence putting $\mathbb{B}_C^0(1,1)u$ on the left, dividing both sides by ε and letting $\varepsilon\to 0$ gives

$$\mathbb{B}_C^0(1,1)(p\kappa-1)\geq(1+\kappa)^{-s}-(1+p^{-1})^{q-s}(1+\kappa)^{-q}.$$

Suppose that $p\kappa>1$, divide throughout by $p\kappa-1$ and let $\kappa\to 1/p$. After some simple manipulations, we obtain

$$\mathbb{B}_C^0(1,1)\geq\left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1}.$$

Again, let us *assume* that we have equality here. This leads us to the following *candidates* for the Bellman functions $\mathbb{B}_C^0$ and $\mathbb{B}_C$:

$$B^0(a,\lambda)=\left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1}\lambda a^{-p}$$

and

$$B(a,\lambda)=\left[\left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1}+1-\left(1+\frac{1}{p}\right)^{q-s}\right]\lambda a^{-p}.$$

Let us make an important comment here. Steps 3 and 4 are informal and their only purpose was to obtain the lower bound for the best constant and the explicit formulas for special functions B and B^0, basing on some more or less natural assumptions and guesses. Formally, these two functions need not coincide with $\mathbb{B}_C$ and $\mathbb{B}_C^0$, so we denote them using different letters. We would like to emphasize here that the rigorous proof of (4) is contained *solely* in Steps 5 and 6 below, and the reader could just skip the Steps 1-4 above. However, we believe that these steps are meaningful, since they indicate how to discover the special functions (which are key to the whole proof, as we will see).

Step 5. A key inequality. We will prove that the function B^0 satisfies the estimate (11). It reads

$$\left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1}\frac{v^p-u}{u-1}\geq\left(\frac{uv-1}{uv-v}\right)^{-s}-\left(1+\frac{1}{p}\right)^{q-s}\left(\frac{uv-1}{uv-v}\right)^{-q}. \tag{14}$$

By continuity, we may and do assume that $v \neq 1$. Introduce a new variable $\beta = (uv - v)/(uv-1)$; then $\beta \neq 1$ and a simple transformation shows that

$$u = \frac{\beta - v}{(\beta-1)v}, \qquad u - 1 = \frac{\beta(1-v)}{(\beta-1)v}. \tag{15}$$

This enables us to rewrite the desired bound (14) in the form

$$\left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1}\left(\frac{v^p-1}{v^{-1}-1}(1-\beta^{-1})-1\right) \geq \beta^s - \left(1+\frac{1}{p}\right)^{q-s}\beta^q. \tag{16}$$

Now we will optimize the left-hand side with respect to v. We consider two cases. If $\beta < 1$, then by the second equality in (15) we get $v - 1 = v(u-1)(\beta^{-1}-1) > 0$, so $v > 1$. On the other hand, the function

$$v \mapsto \frac{v^p - 1}{v^{-1}-1} = \frac{(v^{-1})^{-p}-1}{v^{-1}-1}$$

is negative and decreasing on $(0,1)\cup(1,\infty)$, by the convexity of the function $t \mapsto t^{-p}$. Consequently, the left-hand side of (16) is minimized by letting $v \downarrow 1$. A similar, 'symmetric' argument shows that if $\beta > 1$, then $v < 1$ and the left-hand side of (16) is minimized in the limit case $v \uparrow 1$. This limiting estimate is equivalent to

$$\left(\frac{p+1}{p}\beta\right)^{q+1} - \left(\frac{p+1}{p}\beta\right)^{s+1} \geq (q-s)\left(\frac{p+1}{p}\beta - 1\right),$$

which follows from the mean-value property for the function $t \mapsto ((p+1)\beta/p)^t$ and the elementary inequality $\ln x \geq (x-1)/x$. Indeed, there is $\eta \in (s,q)$ such that

$$\begin{aligned}
\left(\frac{p+1}{p}\beta\right)^{q+1} - \left(\frac{p+1}{p}\beta\right)^{s+1} &= (q-s)\left(\frac{p+1}{p}\beta\right)^{\eta+1}\ln\left(\frac{p+1}{p}\beta\right) \\
&\geq (q-s)\left(\frac{p+1}{p}\beta\right)^{\eta+1}\cdot\frac{\frac{p+1}{p}\beta - 1}{\frac{p+1}{p}\beta} \\
&\geq (q-s)\left(\frac{p+1}{p}\beta - 1\right),
\end{aligned}$$

(for the last passage, consider separately the cases $\frac{p+1}{p}\beta > 1$ and $\frac{p+1}{p}\beta < 1$).

Step 6. Proof of (4). Fix an arbitrary sequence $(a_n)_{n=1}^\infty$ of positive numbers and a sequence $(\lambda_n)_{n=1}^\infty$ of nonnegative numbers with $\lambda_1 > 0$. It is enough to establish (4) under the following additional assumption on $(\lambda_n)_{n=1}^\infty$: there is a positive integer N such that $\lambda_1, \lambda_2, \ldots, \lambda_N$ are strictly positive and $\lambda_{N+1} = \lambda_{N+2} = \ldots = 0$.

The inequality (14) implies that B^0 satisfies (10), i.e.,

$$\begin{aligned}
&B^0(a,\lambda) - B^0(a',\lambda') \\
&\quad \geq (\lambda'-\lambda)(a')^{-p}\left[\left(\frac{a'\lambda'-a\lambda}{a'\lambda'-a'\lambda}\right)^{-s} - \left(1+\frac{1}{p}\right)^{q-s}\left(\frac{a'\lambda'-a\lambda}{a'\lambda'-a'\lambda}\right)^{-q}\right],
\end{aligned}$$

provided $\lambda' > \lambda$ and $a'\lambda' > a\lambda$. Fix a positive integer k and apply this estimate to $\lambda = \Lambda_k$, $\lambda' = \Lambda_{k+1}$, $a = A_k/\Lambda_k$ and $a' = A_{k+1}/\Lambda_{k+1}$. We obtain

$$\begin{aligned} &B^0\left(\frac{A_k}{\Lambda_k},\Lambda_k\right) - B^0\left(\frac{A_{k+1}}{\Lambda_{k+1}},\Lambda_{k+1}\right)\\ &\geq \lambda_{k+1}\left(\frac{A_{k+1}}{\Lambda_{k+1}}\right)^{-p}\left[\left(\frac{A_{k+1}}{\Lambda_{k+1}}\right)^{s} a_{k+1}^{-s} - \left(1+\frac{1}{p}\right)^{q-s}\left(\frac{A_{k+1}}{\Lambda_{k+1}}\right)^{q} a_{k+1}^{-q}\right]. \end{aligned}$$

Summing over $k = 1, 2, \ldots, N-1$, we get

$$\begin{aligned} &B^0\left(\frac{A_1}{\Lambda_1},\Lambda_1\right) - B^0\left(\frac{A_N}{\Lambda_N},\Lambda_N\right)\\ &\geq \sum_{k=1}^{N-1}\lambda_{k+1}\left(\frac{A_{k+1}}{\Lambda_{k+1}}\right)^{-p}\left[\left(\frac{A_{k+1}}{\Lambda_{k+1}}\right)^{s} a_{k+1}^{-s} - \left(1+\frac{1}{p}\right)^{q-s}\left(\frac{A_{k+1}}{\Lambda_{k+1}}\right)^{q} a_{k+1}^{-q}\right]. \end{aligned}$$

However, we have $B^0(A_N/\Lambda_N, \Lambda_N) \geq 0$ and

$$\begin{aligned} B^0\left(\frac{A_1}{\Lambda_1},\Lambda_1\right) &= \left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1}\lambda_1 a_1^{-p}\\ &\leq -\lambda_1\left(\frac{A_1}{\Lambda_1}\right)^{-p+s} a_1^{-s} + \left(1+\frac{1}{p}\right)^{q-s}\lambda_1\left(\frac{A_1}{\Lambda_1}\right)^{-p+q} a_1^{-q}. \end{aligned}$$

Indeed, the latter bound is equivalent to

$$\left(\frac{q-s}{p}\right)\left(1+\frac{1}{p}\right)^{-s-1} \leq -1 + \left(1+\frac{1}{p}\right)^{q-s}$$

and follows from (16) which we have already established above (let $\beta \to 1$ there). Putting all the facts together, we get

$$\sum_{k=1}^{N}\lambda_k\left(\frac{A_k}{\Lambda_k}\right)^{-p+s} a_k^{-s} \leq \left(1+\frac{1}{p}\right)^{q-s}\sum_{k=1}^{N}\lambda_k\left(\frac{A_k}{\Lambda_k}\right)^{-p+q} a_k^{-q},$$

which is the desired claim.

3. Estimates for Hardy operators in $\mathbb{R}^d$

This section contains an application of the estimate (4) in the study of Hardy operator in an arbitrary dimension. We start with the necessary definitions and notation. For any positive integer d, we define Hardy operator $\mathbb{H}$ on $\mathbb{R}^d$, which acts on locally integrable functions f on $\mathbb{R}^d$ by

$$\mathbb{H}f(x) = \frac{1}{|B(0,|x|)|}\int_{B(0,|x|)}|f(y)|\mathrm{d}y, \qquad x \in \mathbb{R}^d \setminus \{0\}.$$

Here and in what follows, $B(x,r)$ denotes the ball in $\mathbb{R}^d$, of center x and radius r. The above operator is closely related to Hardy-Littlewood maximal operator M, given by

$$Mf(x) = \sup_{r>0} \frac{1}{|B(0,r)|} \int_{B(0,r)} |f(x+y)| \mathrm{d}y, \qquad x \in \mathbb{R}^d.$$

Namely, we have the obvious estimate $\mathbb{H}f \le Mf$. It is of interest to study the action of $\mathbb{H}$ on L^p spaces. The precise value of the norm $\|\mathbb{H}\|_{L^p \to L^p}$, $1 < p < \infty$, was identified by Christ and Grafakos [3]: we have $\|\mathbb{H}\|_{L^p \to L^p} = p/(p-1)$. See also [12, 20] for related results. We will study the analogues of this statement for negative exponents.

THEOREM 3. *Suppose that $p > 0$ is a fixed exponent.*
(i) If $s \in (-1,0]$, then

$$\int_{\mathbb{R}^d} (\mathbb{H}f(x))^{-p+s} f(x)^{-s} dx \le \left(1 + \frac{1}{p}\right)^{-s} \int_{\mathbb{R}^d} (\mathbb{H}f(x))^{-p} dx. \tag{17}$$

(ii) If $q \ge 0$, then

$$\int_{\mathbb{R}^d} (\mathbb{H}f(x))^{-p} dx \le \left(1 + \frac{1}{p}\right)^{q} \int_{\mathbb{R}^d} (\mathbb{H}f(x))^{-p+q} f(x)^{-q} dx. \tag{18}$$

Both estimates are sharp.

In the proof of the above theorem, we will need a simple lemma. Here and in what follows, $w_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$ is the measure of the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$.

LEMMA 1. *For a locally integrable function f on $\mathbb{R}^d$, let*

$$g_f(x) = \frac{1}{\omega_{d-1}} \int_{\mathbb{S}^{d-1}} |f(|x|\xi)| d\xi, \qquad x \in \mathbb{R}^d.$$

Then $\mathbb{H}g_f = \mathbb{H}(|f|)$ and for any $p \in \mathbb{R}$, $\alpha \in (0,1]$,

$$\int_{\mathbb{R}^d} \big(\mathbb{H}f(x)\big)^p \big|f(x)\big|^{\alpha} dx \le \int_{\mathbb{R}^d} \big(\mathbb{H}g_f(x)\big)^p \big(g_f(x)\big)^{\alpha} dx.$$

If $\alpha < 0$, then the inequality is reversed.

Proof. The equation $\mathbb{H}g_f = \mathbb{H}(|f|)$ is the consequence of Fubini's theorem and the identity

$$\int_{r\mathbb{S}^{d-1}} g_f(\xi) \mathrm{d}\xi = \int_{r\mathbb{S}^{d-1}} f(\xi) \mathrm{d}\xi, \qquad r > 0$$

(which follows at once from the definition of g_f). The second part is due to Jensen's inequality and the fact that the function $\mathbb{H}f$ is radial: indeed, if $q \in (0,1)$, we have

$$\begin{aligned}
\frac{1}{\omega_{d-1} r^{d-1}} \int_{r\mathbb{S}^{d-1}} \big(\mathbb{H}f(x)\big)^p \big|f(x)\big|^{\alpha} \mathrm{d}x &= \big(\mathbb{H}f\big)^p\big|_{r\mathbb{S}^{d-1}} \cdot \frac{1}{\omega_{d-1} r^{d-1}} \int_{r\mathbb{S}^{d-1}} |f|^{\alpha} \mathrm{d}x \\
&\le \big(\mathbb{H}g_f\big)^p\big|_{r\mathbb{S}^{d-1}} \big(g_f\big)^{\alpha}\big|_{r\mathbb{S}^{d-1}} \\
&= \frac{1}{\omega_{d-1} r^{d-1}} \int_{r\mathbb{S}^{d-1}} \big(\mathbb{H}g_f(x)\big)^p \big(g_f(x)\big)^{\alpha} \mathrm{d}x.
\end{aligned}$$

For $\alpha < 0$, the function $s \mapsto s^\alpha$ is convex on $(0,\infty)$ and hence the inequality in the second line is reversed.

Proof of Theorem 3. By the above lemma, we see that we may restrict ourselves to functions f which are radial and nonnegative. For such a function, we easily compute

$$\begin{aligned}\mathbb{H}f(x) = \frac{1}{|B(0,1)||x|^d}\int_0^{|x|}\int_{\mathbb{S}^{d-1}} f(r\xi)r^{d-1}\mathrm{d}\xi\,\mathrm{d}r &= \frac{\omega_{d-1}}{d|B(0,1)||x|^d}\int_0^{|x|^d} f(r^{1/d}e_1)\mathrm{d}r \\ &= \frac{1}{|x|^d}\int_0^{|x|^d} f(r^{1/d}e_1)\mathrm{d}r,\end{aligned}$$

where e_1 is a fixed vector in $\mathbb{S}^{d-1}$. Therefore, passing to polar coordinates again, we get

$$\begin{aligned}\int_{\mathbb{R}^d} (\mathbb{H}f(x))^{-p+s} f(x)^{-s}\mathrm{d}x &= \omega_{d-1}\int_0^\infty \left(\frac{1}{u^d}\int_0^{u^d} f(r^{1/d}e_1)\mathrm{d}r\right)^{-p+s} f(ue_1)^{-s}u^{d-1}\mathrm{d}u \\ &= |B(0,1)|\int_0^\infty \left(\frac{1}{u}\int_0^{u} f(r^{1/d}e_1)\mathrm{d}r\right)^{-p+s} f(u^{1/d}e_1)^{-s}\mathrm{d}u.\end{aligned}$$

The same calculation holds for $s = 0$. Combining these identities with (5) (applied to the function $r \mapsto f(r^{1/d}e_1)$ and to μ equal to Lebesgue's measure) yields (17). The estimate (18) is proved analogously, just by replacing s with q.

It remains to show that both (17) and (18) are sharp. Consider the auxiliary parameter $\varepsilon > 0$, $\alpha > d/p$ and consider the function

$$f(x) = \chi_{B(0,\varepsilon)}(x) + |x|^\alpha \chi_{\mathbb{R}^d\setminus B(0,\varepsilon)}(x).$$

Then $\mathbb{H}f(x) = 1$ for $|x| \le \varepsilon$, while for remaining $x \in \mathbb{R}^d$ we compute, using polar coordinates, that

$$\begin{aligned}\mathbb{H}f(x) &= |B(0,|x|)|^{-1}\left(|B(0,\varepsilon)| + \omega_{d-1}\int_\varepsilon^{|x|} r^\alpha \cdot r^{d-1}\mathrm{d}r\right) \\ &= \frac{\varepsilon^d}{|x|^d} + \frac{d}{d+\alpha}\left(|x|^\alpha - \frac{\varepsilon^{d+\alpha}}{|x|^d}\right).\end{aligned}$$

Consequently, for any $s \in \mathbb{R}$ we have

$$\begin{aligned}&\int_{\mathbb{R}^d} (\mathbb{H}f)^{-p+s} f(x)^{-s}\mathrm{d}x \\ &= |B(0,\varepsilon)| + \omega_{d-1}\int_\varepsilon^\infty \left(\frac{\varepsilon^d}{r^d} + \frac{d}{d+\alpha}\left(r^\alpha - \frac{\varepsilon^{d+\alpha}}{r^d}\right)\right)^{-p+s} r^{\alpha s}\cdot r^{d-1}\mathrm{d}r.\end{aligned}$$

We are ready to show the sharpness of (17). By the above identity, we may write

$$\begin{aligned}\frac{\int_{\mathbb{R}^d} (\mathbb{H}f)^{-p+s} f(x)^{-s}\mathrm{d}x}{\int_{\mathbb{R}^d} (\mathbb{H}f)^{-p}\mathrm{d}x} &\ge \frac{\omega_{d-1}\int_\varepsilon^\infty \left(\frac{d}{d+\alpha}r^\alpha\right)^{-p+s} r^{\alpha s}\cdot r^{d-1}\mathrm{d}r}{\frac{\omega_{d-1}}{d}\varepsilon^d + \omega_{d-1}\int_\varepsilon^\infty \left(\frac{\varepsilon^d}{r^d} + \frac{d}{d+\alpha}r^\alpha\right)^{-p}\cdot r^{d-1}\mathrm{d}r} \\ &\xrightarrow{\varepsilon\to 0} \left(\frac{d}{d+\alpha}\right)^{-s}\end{aligned}$$

(the assumption $\alpha > d/p$ makes the integrals in the first line convergent). Taking α sufficiently close to d/p, we may make the constant $\left(\frac{d}{d+\alpha}\right)^{-s}$ as close to $(1+1/p)^{-s}$ as we wish. This yields the desired sharpness. The argument for (18) is essentially the same; we omit the details.

4. Inequalities for harmonic maximal operator

In this part of the paper we will derive some sharp estimates for the maximal harmonic operator in the general context of probability spaces equipped with tree-like structures. There is a huge literature devoted to various types of estimates in this setting; see e.g. [13, 14, 18, 19] and consult the references therein. Let us start with definitions.

DEFINITION 1. Suppose that (X,μ) is a nonatomic probability space. A set $\mathscr{T}$ of measurable subsets of X will be called a tree if the following conditions are satisfied:

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

(ii) For every $Q \in \mathscr{T}$ there is a finite subset $C(Q) \subset \mathscr{T}$ such that

(a) the elements of $C(Q)$ are pairwise disjoint subsets of Q,

(b) $Q = \bigcup C(Q)$.

(iii) $\mathscr{T} = \bigcup_{n\geq 0} \mathscr{T}_n$, where $\mathscr{T}_0 = \{X\}$ and $T_{n+1} = \bigcup_{Q\in\mathscr{T}_n} C(Q)$.

(iv) We have $\lim_{n\to\infty} \sup_{Q\in\mathscr{T}_n} \mu(Q) = 0$.

The crucial example the reader should have in mind is the unit cube $[0,1)^d$ with Lebesgue's measure and the dyadic tree: for any $n \geq 0$, the class $\mathscr{T}_n$ consists of all dyadic cubes contained in $[0,1)^d$ and having measure 2^{-nd}. We would also like to mention that instead of trees, one can think about special atomic filtrations of (X,μ). Indeed, setting $\mathscr{F}_n = \sigma(\mathscr{T}_n)$ for $n = 0, 1, 2, \ldots$, one obtains an increasing family $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \ldots$ of σ-algebras of subsets of X, satisfying $\mathscr{F}_0 = \{\emptyset, X\}$ and such that $\mathscr{F}_n$ contains only a finite number of elements.

Given $(X,\mathscr{T},\mu)$, the associated maximal operator $\mathrm{M}_{\mathscr{T}}$ is defined an operator acting on integrable functions (random variables) $\varphi : X \to \mathbb{R}$ by the formula

$$\mathrm{M}_{\mathscr{T}}\varphi(x) = \sup\left\{ \fint_Q |\varphi| \mathrm{d}\mu \,:\, x \in Q,\, Q \in \mathscr{T} \right\}.$$

In the particular dyadic setting described above this is just the classical dyadic maximal operator. We will be interested in the related object, the so-called *harmonic* maximal operator $\mathscr{M}_{\mathscr{T}}$ on $(X,\mathscr{T},\mu)$. This operator is defined by the identity

$$\mathscr{M}_{\mathscr{T}}\varphi(x) = \sup\left\{ \left(\fint_Q |\varphi|^{-1} \mathrm{d}\mu \right)^{-1} \,:\, x \in Q,\, Q \in \mathscr{T} \right\},$$

with the convention $1/0=\infty$ and $1/\infty=0$. The joint behavior of M and $\mathscr{M}$ is similar to that of the arithmetic and the harmonic averages

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

where x_1, x_2, ..., x_n are arbitrary real numbers. In particular, one easily verifies the pointwise estimate $\mathrm{M}_{\mathscr{T}}\varphi \geq \mathscr{M}_{\mathscr{T}}\varphi$ on X. In a sense, the harmonic maximal operator controls φ on the set where the function is small (while $\mathrm{M}_{\mathscr{T}}$ controls φ where the function is large). The harmonic maximal operators appeared for the first time in the works [4, 5, 6] in a slightly different form. They were used to study the fine structure of A_p weights in [4], further applications to weighted norm inequalities and differentiation theory can be found in [6].

Our contribution is the following sharp bound.

THEOREM 4. *Let (X,μ) be a probability space with a tree $\mathscr{T}$. For any $p,\alpha>0$ and any p-integrable random variable φ we have*

$$\int_X (\mathscr{M}_{\mathscr{T}}\varphi)^{\alpha}|\varphi|^{p-\alpha}d\mu \leq \left(1+\frac{1}{p}\right)^{\alpha}\int_X |\varphi|^p d\mu. \tag{19}$$

The inequality is sharp for each individual tree $\mathscr{T}$. In particular, setting $\alpha=p$, we obtain $\|\mathscr{M}_{\mathscr{T}}\|_{L^p(\mu)\to L^p(\mu)}=(1+1/p)^p$.

Proof of (19) We split the reasoning into a few steps.

Step 1. Truncation. Pick an arbitrary probability space (X,μ) equipped with a tree-like structure $\mathscr{T}$ and let N be a fixed nonnegative integer. Let φ be a random variable belonging to L^p. Consider the truncated harmonic maximal function $\mathscr{M}^N_{\mathscr{T}}\varphi$, given by

$$\mathscr{M}^N_{\mathscr{T}}\varphi(x)=\max\left\{\left(\fint_Q |\varphi|^{-1}\mathrm{d}\mu\right)^{-1} : x\in Q,\, Q\in\mathscr{T}_n,\, n\leq N\right\}.$$

In other words, when computing $\mathscr{M}^N_{\mathscr{T}}\varphi$ one proceeds as in the case of the (full) harmonic maximal operator, but the supremum is taken over sets Q belonging to the first $N+1$ generations of $\mathscr{T}$. Obviously, we have $\mathscr{M}^N_{\mathscr{T}}\varphi\uparrow\mathscr{M}_{\mathscr{T}}\varphi$ almost surely as $N\to\infty$ and therefore, by Lebesgue's monotone convergence theorem, it is enough to show that

$$\int_X (\mathscr{M}^N_{\mathscr{T}}\varphi)^{\alpha}|\varphi|^{p-\alpha}\mathrm{d}\mu \leq \left(1+\frac{1}{p}\right)^{\alpha}\int_X |\varphi|^p \mathrm{d}\mu. \tag{20}$$

By straightforward approximation (and enlarging N if necessary), we may assume that φ is constant on each element of $\mathscr{T}^N$.

Step 2. Some special elements of $\mathscr{T}_N$. We will show that there exists $Q^0\in\mathscr{T}_N$ on which $\mathscr{M}^N_{\mathscr{T}}\varphi=\left(\int_X|\varphi|^{-1}\mathrm{d}\mu\right)^{-1}$ (i.e., on Q^0, the maximum defining the maximal operator is attained at the full space X). Otherwise, the space X could be covered by a finite number of elements U_1, U_2, $\ldots$, U_k of $\mathscr{T}$ such that $\left(\int_X |\varphi|^{-1}\mathrm{d}\mu\right)^{-1} < \left(\int_{U_j} |\varphi|^{-1}\mathrm{d}\mu\right)^{-1}$ for each j. By the tree structure, we could assume that these sets are pairwise disjoint: indeed, if $U_i \cap U_j \neq \emptyset$ for some i, j, then one of the sets must be contained in the other and hence can be discarded. This would give

$$\int_X |\varphi|^{-1}\mathrm{d}\mu = \sum_{j=1}^k \int_{U_j} |\varphi|^{-1} < \sum_{j=1}^k \mu(U_j) \int_X |\varphi|^{-1}\mathrm{d}\mu = \int_X |\varphi|^{-1}\mathrm{d}\mu,$$

a contradiction. So, the aforementioned extremal set $Q^0 \in \mathscr{T}_N$ exists; let us record here the inequality

$$\left(\fint_{Q^0} |\varphi|^{-1}\mathrm{d}\mu\right)^{-1} \leq \int_X |\varphi|^{-1}\mathrm{d}\mu, \tag{21}$$

which follows from the very definition of the maximal operator.

We will use the above observation inductively: put $X^0 = X$, $\mu^0 = \mu$, $\mathscr{T}^{(0)} = \mathscr{T}$. Suppose we have successfully constructed X^j, μ^j, $\mathscr{T}^{(j)}$ and Q^j. Consider the modified space $X^{j+1} = X^j \setminus Q^j$ with the probability measure $\mu^{j+1} = \mu/\mu(X^{j+1})$ and the tree structure $\mathscr{T}^{(j+1)}$ such that $\mathscr{T}_n^{(j+1)}$ consists of all elements of the form $A \setminus Q^j$, $A \in \mathscr{T}_n^{(j)}$, $n \geq 0$. Applying the above reasoning to this new space, we obtain the existence of $Q^{j+1} \in \mathscr{T}_N^{(j+1)} \subset \mathscr{T}_N$ on which $\mathscr{M}^N_{\mathscr{T}^{(j+1)}}\varphi = \left(\int_{X^{j+1}} |\varphi|^{-1}\mathrm{d}\mu^{j+1}\right)^{-1}$. We continue the procedure until we use all the elements Q^0, Q^1, $\ldots$, Q^K of $\mathscr{T}_N$.

Step 3. Completion of the proof. Directly from the construction in the above step,

$$\left(\fint_{Q^j} |\varphi|^{-1}\mathrm{d}\mu\right)^{-1} \leq \left(\int_{X^j} |\varphi|^{-1}\mathrm{d}\mu^j\right)^{-1} \tag{22}$$

(which is due to (21)) and hence in particular

$$\left(\int_{X^j} |\varphi|^{-1}\mathrm{d}\mu^j\right)^{-1} \leq \left(\int_{X^{j+1}} |\varphi|^{-1}\mathrm{d}\mu^{j+1}\right)^{-1}. \tag{23}$$

Consequently, we see that for $x \in Q^j$ we have

$$\mathscr{M}^N_{\mathscr{T}}\varphi(x) \leq \left(\int_{X^j} |\varphi|^{-1}\mathrm{d}\mu^j\right)^{-1}. \tag{24}$$

Indeed, by the very definition of the truncated operator, there is $R \in \mathscr{T}_0 \cup \mathscr{T}_1 \cup \ldots \mathscr{T}_N$ containing x such that $\mathscr{M}^N_{\mathscr{T}}\varphi(x) = \left(\fint_R |\varphi|^{-1}\mathrm{d}\mu\right)^{-1}$. Set $R_+ = R \setminus X^j$ and $R_- = R \cap X^j$. Then R_+ is a union of some Q^k with $k < j$, so by (22) and (23),

$$\left(\fint_{R_+} |\varphi|^{-1}\mathrm{d}\mu\right)^{-1} \leq \left(\int_{X^j} |\varphi|^{-1}\mathrm{d}\mu^j\right)^{-1}.$$

On the other hand, R_- belongs to $\mathscr{T}^0_{(j)} \cup \mathscr{T}^1_{(j)} \cup \ldots \mathscr{T}^N_{(j)}$ and contains x, so by the definition of Q^j,

$$\left(\fint_{R_-} |\varphi|^{-1} \mathrm{d}\mu^j\right)^{-1} \leq \left(\int_{X^j} |\varphi|^{-1} \mathrm{d}\mu^j\right)^{-1}.$$

The last two estimates yield (24). Now we apply (4) with $q = p$ and $s = p - \alpha$, setting $\lambda_j = \mu(Q^{K+1-j})$ and $a_j = |\varphi|^{-1}\big|_{Q^{K+1-j}}$ (recall that we assumed that φ is constant on each Q^k) for $j = 1, 2, \ldots, K+1$; for remaining j, we put $\lambda_j = 0$, $a_j = 1$. Since for each $n \leq K+1$ we have

$$\frac{\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} = \fint_{X^{K+1-n}} |\varphi|^{-1} \mathrm{d}\mu,$$

the inequality (4) implies

$$\sum_{j=1}^{K} \int_{Q^j} \left(\fint_{X^j} |\varphi|^{-1} \mathrm{d}\mu\right)^{-\alpha} |\varphi|^{p-\alpha} \mathrm{d}\mu \leq \left(1 + \frac{1}{p}\right)^{\alpha} \int_X |\varphi|^p \mathrm{d}\mu.$$

It remains to apply (24) to get the claim.

Note that in the above reasoning, the property (iv) of the trees was not necessary. It will, however, be needed in the proof of sharpness of (19).

We will use the following statement which can be found in [13].

LEMMA 2. *For every $Q \in \mathscr{T}$ and every $\beta \in (0,1)$ there is a subfamily $F(Q) \subset \mathscr{T}$ consisting of pairwise disjoint subsets of Q such that*

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

Equipped with the above fact, we are ready to show that the constant $(1 + 1/p)^{\alpha}$ is optimal in (19).

Sharpness of (19). Fix an arbitrary probability space (X, μ) with a tree structure $\mathscr{T}$. Let $\varepsilon > 0$ be an auxiliary parameter. We split the argumentation into two separate parts.

Step 1. A special sets. We begin with an inductive use of Lemma 2 to obtain an increasing family $A_0 \supset A_1 \supset A_2 \supset \ldots$ of subsets of X. We start with putting $A_0 = X$. Suppose that we have constructed the set A_n and assume additionally that this set can be expressed as a union of pairwise disjoint elements of $\mathscr{T}$: these elements will be called the *atoms* of A_n. Obviously, such a decomposition holds for $n = 0$: we have $A_0 = X \in \mathscr{T}$. For each atom Q of A_n, we use Lemma 2 with $\beta = \varepsilon$, obtaining the appropriate subfamily $F(Q)$ of subsets of Q. Then we set $A_{n+1} = \bigcup_Q \bigcup_{Q' \in F(Q)} Q'$, where the first union is taken over all atoms Q of A_n. This set has the desired decomposition property: obviously, it is a union of the family $\{F(Q) : Q \text{ an atom of } A_n\}$, which consists of pairwise disjoint elements of $\mathscr{T}$. These elements are the required atoms of A_{n+1}. This completes the description of the induction step.

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

$$\mu(Q \cap (A_n \setminus A_{n+1})) = \mu(Q)\varepsilon^{n-m}(1-\varepsilon). \tag{25}$$

Step 2. The calculation. We take $\varphi = \sum_{n=0}^{\infty} \varepsilon^{\kappa n} \chi_{A_n \setminus A_{n+1}}$, where $\kappa > -1/p$ is an auxiliary parameter. Note that φ belongs to L^p: indeed, by (25) applied to $m = 0$ and $Q = X$,

$$\int_X |\varphi|^p \mathrm{d}\mu = \sum_{n=0}^{\infty} \varepsilon^{p\kappa n} \mu(A_n \setminus A_{n+1}) = (1-\varepsilon) \sum_{n=0}^{\infty} \varepsilon^{(1+p\kappa)n} < \infty.$$

Next, for any $m \geq 0$ and any atom Q of A_m we have, by (25),

$$⨍_Q |\varphi|^{-1} \mathrm{d}\mu = \sum_{n \geq m} \varepsilon^{-\kappa n + n - m}(1-\varepsilon) = \varepsilon^{-\kappa m} \cdot \frac{1-\varepsilon}{1-\varepsilon^{1-\kappa}}.$$

In particular, this implies the estimate $\mathscr{M}_{\mathscr{T}} \varphi \geq \varepsilon^{\kappa m} \cdot \dfrac{1-\varepsilon^{1-\kappa}}{1-\varepsilon}$ on A_m and hence also $\mathscr{M}_{\mathscr{T}} \varphi \geq \dfrac{1-\varepsilon^{1-\kappa}}{1-\varepsilon} \varphi$ on X. Consequently, we see that

$$\int_X (\mathscr{M}_{\mathscr{T}} \varphi)^{\alpha} |\varphi|^{p-\alpha} \mathrm{d}\mu \geq \left(\frac{1-\varepsilon^{1-\kappa}}{1-\varepsilon}\right)^{\alpha} \int_X |\varphi|^p \mathrm{d}\mu.$$

Consequently, the optimal constant in (19) must be at least $\big((1-\varepsilon^{1-\kappa})/(1-\varepsilon)\big)^{\alpha}$. Taking ε sufficiently close to 1 and κ sufficiently close to $-1/p$, we may make this constant arbitrarily close to $(1+1/p)^{\alpha}$. The proof of the sharpness is complete.

Acknowledgments

The authors would like to thank an anonymous referee for the careful reading of the first version of the paper and for several helpful corrections. The results were obtained during the realization of the tutoring program, supported by Polish Ministry of Science and Higher Education project MNiSW/2019/394/DIR/KH.

REFERENCES

[1] R. BELLMAN, *Dynamic programming*, Reprint of the 1957 edition. Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 2010.

[2] D. L. BURKHOLDER, *Explorations in martingale theory and its applications*, École d'Été de Probabilités de Saint-Flour XIX—1989, 1–66, Lecture Notes in Math., **1464**, Springer, Berlin.

[3] M. CHRIST AND L. GRAFAKOS, *Best constants for two nonconvolution inequalities*, Proc. Amer. Math. Soc. **123** (1995), 1687–1693.

[4] D. CRUZ-URIBE, SFO, AND C. J. NEUGEBAUER, *The structure of the reverse Hölder classes*, Trans. Amer. Math. Soc. **347** (1995), 2941–2960.

[5] D. CRUZ-URIBE,, *The minimal operator and the geometric maximal operator in* $\mathbb{R}^n$, Studia Math. **144**, 1 (2001), 1–37.

[6] D. CRUZ-URIBE, SFO, C.J. NEUGEBAUER, V. OLESEN, *Norm inequalities for the minimal and maximal operator, and differentiation of the integral*, Publ. Mat. **41** (1997), 577–604.
[7] G. HARDY, *Note on a theorem of Hilbert*, Math. Z. **6**, 3-4 (1920), 314–317.
[8] G. H. HARDY, J. E. LITTLEWOOD AND G. PÓLYA, *Inequalities*, 2nd edn., Cambridge University Press, 1952.
[9] K. KNOPP, *Über Reihen mit positiven Gliedern*, J. London Math. Soc. **3** (1928), 205–211.
[10] A. KUFNER, B. OPIC, *Hardy-type inequalities*, Longman Scientific and Technical, Harlow, 1990.
[11] A. KUFNER AND L. E. PERSSON, *Weighted Inequalities of Hardy Type*, World Sci., River Edge, NJ, 2003.
[12] S. Z. LU, D. Y. YAN AND F. Y. ZHAO, *Sharp bounds for Hardy type operators on higher-dimensional product spaces*, J. Inequal. Appl., 2013, 1 (2013), 1–11.
[13] A. D. MELAS, *The Bellman functions of dyadic-like maximal operators and related inequalities*, Adv. Math. **192** (2005), 310–340.
[14] A. D. MELAS AND E. N. NIKOLIDAKIS, *Sharp Lorentz estimates for dyadic-like maximal operators and related Bellman functions*, J. Geom. Anal. **27**, 4 (2017), 2644–2657.
[15] F. L. NAZAROV AND S.R. TREIL, *The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis*, St. Petersburg Math. J. **8** (1997), 721–824.
[16] E. NIKOLIDAKIS, T. STAVROPOULOS, *A refinement of a Hardy type inequality for negative exponents, and sharp applications to Muckenhoupt weights on* $\mathbb{R}$, Colloq. Math. **157** (2018), 295–308.
[17] A. OSĘKOWSKI, *Sharp martingale and semimartingale inequalities*, Monografie Matematyczne **72** (2012), Birkhäuser, 462 pp.
[18] A. OSĘKOWSKI, *Sharp* $L^{p,\infty} \to L^q$ *estimates for the dyadic-like maximal operators*, J. Fourier Anal. Appl. **20** (2014), 911–933.
[19] A. OSĘKOWSKI, *Sharp weak-type estimates for the dyadic-like maximal operators*, Taiwanese J. Math. **19** (2015), 1031–1050.
[20] A. OSĘKOWSKI, *Sharp logarithmic inequalities for Hardy operators*, Z. Anal. Anwend. **35** (2016), 1–20.

Łukasz Kamiński, Faculty of Mathematics, Informatics and Mechanics
University of Warsaw
Banacha 2, 02-097 Warsaw
Poland,
e-mail: `l.kaminski5@student.uw.edu.pl`

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

Corresponding Author: Adam Osękowski