WEIGHTED INEQUALITIES FOR q-FUNCTIONS

TOMASZ GAŁĄZKA AND ADAM OSĘKOWSKI

Abstract. Let f be a martingale on an arbitrary atomic probability space equipped with a tree-like structure and let $S(f,q)$ denote the associated q-function. The paper is devoted to weighted L^p-estimates

$$c_{p,q,w}^{-1}\|S(f,q)\|_{L^p(w)} \leq \|f\|_{L^p(w)} \leq C_{p,q,w}\|S(f,q)\|_{L^p(w)}, \qquad 1 \leq p < \infty,$$

for Muckenhoupt weights. Using the combination of the theory of sparse operators, extrapolation and Bellman function method, we identify the optimal dependence of the constants $c_{p,q,w}$ and $C_{p,q,w}$ on the A_p characteristics of the weights involved.

1. Introduction

The purpose of this paper is to study weighted estimates for q-functions, which arise naturally in the abstract probabilistic and analytic context. Let us start with the necessary background and notation. Let d be a fixed dimension and let $\mathcal{D}(\mathbb{R}^d)$ denote the standard dyadic lattice in $\mathbb{R}^d$. This lattice gives rise to the natural dyadic filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}}$, where for each n, the σ-algebra $\mathcal{F}_n$ is generated by all cubes $Q \in \mathcal{D}(\mathbb{R}^d)$ satisfying $|Q| = 2^{-nd}$. Let $(\mathcal{E}_n)_{n\in\mathbb{Z}}$ be the associated sequence of conditional expectations. Given an integrable function f on $\mathbb{R}^d$, we denote by $(f_n)_{n\in\mathbb{Z}}$ the corresponding $(\mathcal{F}_n)_{n\in\mathbb{Z}}$-martingale, i.e., we set $f_n = \mathcal{E}_n(f)$ for all $n \in \mathbb{Z}$. The associated difference sequence $(df_n)_{n\in\mathbb{Z}}$ is defined by $df_n = f_n - f_{n-1}$, $n \in \mathbb{Z}$. Sometimes, to indicate the cube we are restricting to, we will use the notation $\langle f\rangle_Q = \frac{1}{|Q|}\int_Q f\,\mathrm{d}x$ for the average of f over Q and denote by $\Delta_Q f = \langle f\rangle_Q - \langle f\rangle_{Q'}$ the difference of f with respect to Q; here Q' stands for the direct dyadic parent of Q. So, for each $n \in \mathbb{Z}$ we have the identities

$$f_n = \sum_Q \langle f\rangle_Q \chi_Q \quad \text{ and } \quad df_n = \sum_Q \Delta_Q f \chi_Q,$$

where the summation is taken over all $Q \in \mathcal{D}(\mathbb{R}^d)$ of measure 2^{-nd}.

In the paper, we will be interested in the boundedness properties of the q-function of f, where $q \in (1,\infty)$ is a fixed parameter. This object is defined by

$$S(f,q) = \left(\sum_{n\in\mathbb{Z}} |df_n|^q\right)^{1/q} = \left(\sum_{Q\in\mathcal{D}(\mathbb{R}^d)} |\Delta_Q f|^q \chi_Q\right)^{1/q},$$

it can also be easily formulated in the general probabilistic context (see below). Note that in the special case $q = 2$, we obtain the classical dyadic square function associated with f. The inequalities between f and $S(f,q)$ have played an important role in probability and analysis. Of course, the case $q = 2$ is prominent here and is foundational to the whole

2010 *Mathematics Subject Classification.* Primary: 42B25; Secondary: 60G42.

Key words and phrases. Weight, sparse, q-function, Bellman function.

The research was supported by Narodowe Centrum Nauki (Poland), grant 2018/30/Q/ST1/00072.

harmonic and stochastic analysis, but the case of other q has also been investigated in the literature, as we briefly discuss now. The L^p inequalities

$$c_p^{-1}\|S(f,p)\|_{L^p} \leq \|f\|_{L^p} \leq C_p\|S(f,p)\|_{L^p}, \qquad 1<p<\infty, \tag{1.1}$$

with the particular emphasis put on the description of the optimal values of c_p and C_p, have been studied by many authors: see e.g. the works of Alsmeyer and Rösler [1], von Bahr and Esseen [2], Cox [7], Pinelis [27] and Pisier [30]. Such estimates have found applications in probability and statistics, e.g. they can be applied in the concentration of measure of separate Lipschitz functions on product spaces. They can also be used in the study of various geometric properties of L^p spaces (e.g., p-smoothness, see also below).

The estimate (1.1) can be improved. Another instance, in which the q-function appears, is one of the equivalent formulations of the Rosenthal inequality for martingales: for $p \geq 2$ and any $q \in (p,\infty)$,

$$\|f\|_{L^p} \asymp_p \left\| \left(\sum_{n\in\mathbb{Z}} \mathcal{E}_{n-1}(df_n^2) \right)^{1/2} \right\|_{L^p} + \|S(f,q)\|_{L^p},$$

where the symbol $A \asymp_p B$ means the existence of a constant $1 \leq c_p < \infty$ depending only on p such that $c_p^{-1}B \leq A \leq c_pB$. See [15, 19, 31] for details. Consult also the recent works [28, 29] by Pinelis which contain extensions to vector-valued context: the above definition of a q-function makes perfect sense for Banach-space valued martingales, one only needs to interpret $|\cdot|$ as the appropriate norm.

Finally, we would like to mention that the inequalities between $S(f,q)$ and f arise naturally in the context of superreflexivity. To recall the relevant definitions, suppose that X and Y are Banach spaces. We say that X is finitely representable in Y if for all finite-dimensional subspaces E of X and all $\lambda > 1$, there is a linear map $T : E \to Y$ such that $\lambda^{-1}\|x\|_X \leq \|Tx\|_Y \leq \lambda\|x\|_X$ for all $x \in E$. A Banach space X is superreflexive if it is reflexive and every Banach space that is finitely representable in X is also reflexive. This notion was introduced in 1972 by James, see the papers [17, 18] for some basic properties of superreflexive spaces. Pisier [30] extended James' work and found an equivalent formulation for superreflexivity: the Banach space X is superreflexive, if and only if there are $p \in (1,\infty)$ and $q \in [2,\infty)$ such that for all X-valued dyadic martingales f,

$$\|S(f,q)\|_{L^p} \leq c_{p,q}\|f\|_{L^p},$$

where $c_{p,q}$ is a finite constant depending only on the parameters indicated.

In this paper, we will be interested in the weighted context. Here and below, the word 'weight' will refer to a positive, locally integrable function, which will usually be denoted by w or v. Any weight w on $\mathbb{R}^d$ gives rise to the corresponding measure $w\mathrm{d}x$ and we will often use the notation $w(A) = \int_A w\mathrm{d}x$ for any Borel subset A of $\mathbb{R}^d$. The associated weighted L^p spaces, $0 < p < \infty$, are given by

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

There is an interesting problem to investigate the extensions of (1.1), in which $S(f,p)$ is replaced by the general q-function $S(f,q)$ and the space L^p is replaced by its weighted counterpart. This problem has been studied quite intensively in the special case $q = 2$ (see [3, 4, 11, 16, 22]). It is not difficult to see that such estimates cannot hold for arbitrary weights and some structural assumption needs to be imposed. Given $1 < p < \infty$, a weight w is said to satisfy Muckenhoupt's condition A_p (or to belong to the class A_p), if the characteristic

$$[w]_{A_p} := \sup_{Q\in\mathcal{D}(\mathbb{R}^d)} \left(\frac{1}{|Q|}\int_Q w\mathrm{d}x\right)\left(\frac{1}{|Q|}\int_Q w^{1/(1-p)}\mathrm{d}x\right)^{p-1}$$

is finite. There are also appropriate versions of this condition in the boundary case $p\in\{1,\infty\}$; we will only recall the case $p=1$, as the infinite case is irrelevant for our considerations below. We say that w belongs to the (dyadic) class A_1, if

$$[w]_{A_1} = \sup_{Q\in\mathcal{D}(\mathbb{R}^d)} \frac{\langle w\rangle_Q}{\operatorname{essinf}_Q w}$$

is finite. Then for any $1<p<\infty$ and any $w\in A_p$, there are finite constants $c_{p,w}$, $C_{p,w}$ depending only on the parameters indicated such that

$$c_{p,w}^{-1}\|S(f,2)\|_{L^p(w)} \leq \|f\|_{L^p(w)} \leq C_{p,w}\|S(f,2)\|_{L^p(w)}.$$

This result can be further sharpened in the following direction: one can ask about the extraction of the optimal dependence of $c_{p,w}$ and $C_{p,w}$ on the characteristic $[w]_{A_p}$. Specifically, the problem is to determine, for any fixed p, the smallest exponents k_p, K_p such that $c_{p,w}\leq c_p[w]_{A_p}^{k_p}$ and $C_{p,w}\leq C_p[w]_{A_p}^{K_p}$, where this time c_p and C_p depend only on p. It turns out that the answer is $k_p=\max\{1/2,1/(p-1)\}$ and $K_p=1$ (if the filtration/tree is regular, then K_p can be decreased): see [3, 4, 11, 16, 22].

Our contribution is the study of analogous questions for an arbitrary q. Our result in the dyadic context can be formulated as follows.

Theorem 1.1. *Let f, w be a function and a weight on $\mathbb{R}^d$.*

(i) Suppose that $q\geq 2$. Then for any $1<p<\infty$ we have

$$\|S(f,q)\|_{L^p(w)} \leq C_{p,q}[w]_{A_p}^{\max\{\frac{1}{q},\frac{1}{p-1}\}}\|f\|_{L^p(w)}. \tag{1.2}$$

Here $C_{p,q}$ depends only on the parameters indicated and the exponent $\max\{\frac{1}{q},\frac{1}{p-1}\}$ is the best possible.

(ii) Suppose that $1<q\leq 2$. Then for any $1\leq p<\infty$ we have

$$\|f\|_{L^p(w)} \leq C_{p,q}[w]_{A_p}\|S(f,q)\|_{L^p(w)}. \tag{1.3}$$

Here $C_{p,q}$ depends only on the parameters indicated and the linear dependence on $[w]_{A_p}$ is the best possible.

We would like to emphasize here that the above estimates are dimension-free: the constants involved in (1.2) and (1.3) do not depend on d. As we have already mentioned above, if we allowed the dependence on the dimension, then the linear factor $[w]_{A_p}$ in (1.3) could be improved (see [11] for the case $q=2$).

Actually, the regular dyadic structure is not necessary for the validity of the above result. We will study the above statement in a much more general context of arbitrary probability spaces equipped with tree-like structures. Here is the precise definition.

Definition 1.2. Suppose that $(X,\mathcal{F},\mu)$ 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 $\lim_{m\to\infty} \sup_{Q \in \mathcal{T}^m} \mu(Q) = 0$.

All the definitions formulated in the dyadic context can easily be extended to the above probabilistic setting, as we briefly discuss now. Let $(X, \mathcal{F}, \mu)$ be a probability space with a tree $\mathcal{T}$. Then the notions of martingales, differences, q-functions, averages and weights make perfect sense, it suffices to simply replace $\mathbb{R}^d$ with X, the Lebesgue measure dx with μ and the dyadic lattice $\mathcal{D}(\mathbb{R}^d)$ with $\mathcal{T}$. The only essential difference is that now the filtration becomes a one-sided sequence: we take $\mathcal{F}_n = \sigma(\mathcal{T}_n)$ for each $n = 0, 1, 2, \ldots$. This forces us to modify the definition of the difference sequence, in which we set $df_0 = f_0$ and $\Delta_X f = \langle f \rangle_X$.

We will prove the following analogue of Theorem 1.1.

Theorem 1.3. *Let $(X, \mathcal{F}, \mu)$ be a probability space with a tree $\mathcal{T}$. Assume further that f, w are a random variable and a weight on X.*

(i) Suppose that $q \geq 2$. Then for any $1 < p < \infty$ we have

$$\|S(f,q)\|_{L^p(w)} \leq C_{p,q} [w]_{A_p}^{\max\{\frac{1}{q}, \frac{1}{p-1}\}} \|f\|_{L^p(w)}. \tag{1.4}$$

Here $C_{p,q}$ is a finite constant depending on p and q only, and the exponent $\max\{\frac{1}{q}, \frac{1}{p-1}\}$ *is the best possible.*

(ii) Suppose that $1 < q \leq 2$. Then for any $1 \leq p < \infty$ we have

$$\|f\|_{L^p(w)} \leq C_{p,q} [w]_{A_p} \|S(f,q)\|_{L^p(w)}. \tag{1.5}$$

Here $C_{p,q}$ depends only on the parameters indicated and the linear dependence on $[w]_{A_p}$ is the best possible.

In analogy to the dyadic setting discussed previously, we would like to emphasize that in the above statement we do not require the tree to satisfy any regularity properties. To see that the above probabilistic context does generalize the dyadic setting, observe first that the unit cube $[0,1]^d$ with its Borel subsets and the Lebesgue measure forms a probability space, and the dyadic subsets of $[0,1]^d$ form a tree. Now, the use of standard dilation and limiting arguments allow to deduce the validity of Theorem 1.1 from its probabilistic counterpart.

A few words about the organization of the paper and the proof are in order. In our considerations below, we exploit a variety of methods. The proof of the case $q \geq 2$, presented in Section 2, is a little simpler: the inequality (1.4) is deduced from extrapolation and the theory of sparse operators (which have become an independent research area in the recent literature: see e.g. [10, 20, 23, 24] and consult the references therein). The sharpness of the exponent is obtained via the combination of extrapolation again and the construction of appropriate examples. The analysis of the case $1 < q \leq 2$, contained in Section 3, is a little more involved. First, one can try to apply duality arguments to (1.4), but one obtains the suboptimal version of (1.5):

$$\|f\|_{L^p(w)} \leq C_{p,q} [w]_{A_p}^{\max\{1, 1/(q'(p-1))\}} \|S(f,q)\|_{L^p(w)},$$

in which the linear dependence appears only for $p \geq 2-q^{-1}$ (here and below, $q' = q/(q-1)$ is the dual exponent to q). The use of sparse operators does not seem to lead to the sharp inequality either. To overcome this difficulty, we will use a completely different approach and exploit the Bellman function technique. This method is a well-known tool, used widely in probability theory and harmonic analysis (cf. [5, 25, 26, 32, 33], consult also the references therein), which often leads to sharp or at least tight estimates. Roughly speaking, the argument allows to deduce a given estimate from the existence of a certain function enjoying appropriate size and majorization conditions. This method gives us a Fefferman-Stein-type L^1 estimate for general weights which is of independent interest and connections. Combined with extrapolation, it gives us a unified proof of (1.5) in the full range $1 \leq p < \infty$. The sharpness follows from the construction of examples.

2. The case $q \geq 2$

2.1. **Proof of** (1.4)**.** As we have mentioned above, the argument will rest on the theory of sparse operators. Let us recall the relevant notions.

Definition 2.1. Suppose that $(X, \mathcal{F}, \mu)$ is a probability space equipped with a tree structure $\mathcal{T}$. A collection $\mathcal{J} \subset \mathcal{T}$ is called sparse, if for any $Q \in \mathcal{J}$ there exists a measurable subset $E(Q) \subset Q$ with $\mu(E(Q)) \geq \mu(Q)/2$ and such that $E(Q) \cap E(Q') = \emptyset$ unless $Q = Q'$.

The key ingredient of the proof of the L^p bound is the pointwise domination of q-functions by sparse operators. We will prove the following statement.

Theorem 2.2. *Let $q \geq 2$. There is a constant $C > 0$ such that for every integrable and nonnegative function f on X, there exists a sparse family $\mathcal{J} = \mathcal{J}_f$ such that*

$$|S(f,q)|^q \leq C \sum_{Q \in \mathcal{J}} \langle f \rangle^q \chi_Q.$$

Proof. The argument is rather standard and rests on an appropriate recursive procedure. Let M be the martingale maximal function, acting on integrable functions φ on X by $M\varphi = \sup_Q \langle |\varphi| \rangle_Q \chi_Q$. It is well-known that both M and $S(\cdot, q)$ are bounded as operators from L^1 to $L^{1,\infty}$, so there exists a constant C_0 such that the set

$$\mathcal{H} = \mathcal{H}(X) := \left\{ x \in X : \max\{Mf, S(f,q)\} > \frac{1}{2} C_0^{\frac{1}{q}} \langle f \rangle_X \right\}$$

satisfies $\mu(\mathcal{H}) \leq \frac{1}{2}\mu(X)$. Let $\mathcal{E}$ be the family of maximal elements in $\mathcal{T}$ contained in $\mathcal{H}$ (i.e., $Q \in \mathcal{E}$ if and only if $Q \in \mathcal{T}$, $Q \subset \mathcal{H}$ and $Q' \not\subset \mathcal{H}$, where Q' is the parent of Q in $\mathcal{T}$). We claim that

$$|S(f,q)(x)|^q \leq C_0 \langle f \rangle_X^q + \sum_{R \in \mathcal{E}} |S_R(f,q)(x)|^q, \qquad x \in X, \tag{2.1}$$

where $S_R(f,q) = \left(\langle f \rangle_R^q \chi_R + \sum_{Q \subset R} |\Delta_Q f|^q \right)^{\frac{1}{q}}$ is the version of the q-function restricted to the set R. Indeed, for $x \in X \setminus \mathcal{H}$ the above inequality reads

$$|S(f,q)(x)|^q \leq C_0 \langle f \rangle_X^q,$$

which is true from the very definition of $\mathcal{H}$. If $x \in \mathcal{H}$, then by the maximality of elements of $\mathcal{E}$, there exists a unique $R \in \mathcal{E}$ containing x. We have

$$\begin{aligned}|S(f,q)(x)|^q &= \sum_{Q:R' \subsetneq Q} |\Delta_Q f|^q(x) + |\langle f\rangle_R - \langle f\rangle_{R'}|^q + \sum_{Q:Q\subset R} |\Delta_Q f|^q(x)\\ &\leq \sum_{Q:R' \subsetneq Q} |\Delta_Q f|^q(x) + \langle f\rangle_{R'}^q + \langle f\rangle_R^q + \sum_{Q:Q\subset R} |\Delta_Q f|^q(x),\end{aligned}$$

where, as above, R' stands for the direct parent of R in $\mathcal{T}$. Here the estimate is due to the assumption $f \geq 0$. Again, by the maximality of elements of $\mathcal{E}$, each of the first two elements of the sum is bounded by $\frac{1}{2}C_0\langle f\rangle_X^q$. The last two elements sum up to $|S_R(f,q)|^q$. We put X into the desired sparse family $\mathcal{J}$ and repeat recursively the above reasoning. Namely, we use the argumentation with $S(f,q)$ replaced with $S_R(f,q)$ and X replaced with R, for any $R \in \mathcal{E}$ obtained above. Having done this, we obtain

$$|S(f,q)|^q \leq C_0\langle f\rangle_X^q + C_0 \sum_{R\in\mathcal{E}} \langle f\rangle_R^q + \sum_{\tilde{R}\in\mathcal{E}'} |S_{\tilde{R}}(f,q)|^q,$$

where $\mathcal{E}'$ is the collection of all maximal cubes contained in $\mathcal{H}(R)$, $R \in \mathcal{E}$. We continue this procedure with $S(f,q)$ replaced with $S_{\tilde{R}}(f,q)$ for all $\tilde{R} \in \mathcal{E}'$, and so on. The fact that $\mu(\mathcal{H}(R)) \leq \mu(R)/2$ guarantees that the procedure yields a convergent series on the right. This ends the proof. □

Therefore, the q-function associated with nonnegative f is controlled by the sparse operator $T_{q,\mathcal{J}}f = \left(\sum_{Q\in\mathcal{J}} \langle f\rangle^q \chi_Q\right)^{1/q}$, and the case of general f is handled by a standard decomposition into the positive and the negative parts: if $f = f_+ - f_-$, then $S(f,q) \leq 2^{1-1/q}(S(f_+,q) + S(f_-,q))$ and $\|f_\pm\|_{L^p(w)} \leq \|f\|_{L^p(w)}$. Thus, to show (1.4), it is enough to establish the corresponding weighted bound for sparse operators. Such estimates are well-known in the literature: see e.g. [23] or [11] in the dyadic context. In the general probabilistic setting the reasoning is similar: for the sake of convenience and completeness, we provide the details. Our starting point is the following extrapolation theorem (see Duoandikoetxea [13]).

Theorem 2.3. *Suppose that f, g are two given functions on X. Assume further that for some $1 \leq p_0 < \infty$, there exists $\alpha(p_0) > 0$ and a finite constant C such that*

$$\|g\|_{L^{p_0}(w)} \leq C[w]_{A_{p_0}}^{\alpha(p_0)} \|f\|_{L^{p_0}(w)}$$

for all weights $w \in A_{p_0}$. Then for every $1 < p < \infty$ and $w \in A_p$, we have

$$\|g\|_{L^p(w)} \leq C'[w]_{A_p}^{\alpha(p_0)\max\{1,\frac{p_0-1}{p-1}\}} \|f\|_{L^p(w)},$$

where the constant C' depends only on C, p_0 and p.

The next step is to establish the weighted bound for the sparse operators for a particular exponent.

Lemma 2.4. *Let $q \geq 2$. Then for any nonnegative f and any $w \in A_{q+1}$,*

$$\|T_{q,\mathcal{J}}f\|_{L^{q+1}(w)} \leq c[w]_{A_{q+1}}^{\frac{1}{q}} \|f\|_{L^{q+1}(w)}.$$

Proof. Fix a positive $h \in L^{q+1}(w)$ of norm 1 and an arbitrary weight $w \in A_{q+1}$. Let $v = w^{-\frac{1}{q}}$ be the dual weight to w. In what follows, we will also use the notation $\langle f\rangle_{w,Q} = \frac{1}{w(Q)}\int_Q fw\mathrm{d}x$ for the average of f over Q with respect to the measure $w\mathrm{d}x$. By the definition of A_{q+1} class, Hölder's inequality and the properties of sparse family, we have

$$\begin{aligned}
\int_X (T_{q,\mathcal{J}}f)^q hwd\mu &= \sum_{Q\in\mathcal{J}} \langle f\rangle_Q^q \int_Q hw\mathrm{d}\mu \\
&= \sum_{Q\in\mathcal{J}} \frac{w(Q)v(Q)^q}{\mu(Q)^{q+1}}\mu(Q)\langle fv^{-1}\rangle_{v,Q}^q \langle h\rangle_{w,Q} \\
&\leq 2[w]_{A_{q+1}} \sum_{Q\in\mathcal{J}} \mu(E(Q)) \left(\frac{1}{v(Q)}\int_Q fw^{\frac{1}{q}}v\mathrm{d}\mu\right)^q \langle h\rangle_{w,Q} \\
&\leq 2[w]_{A_{q+1}} \sum_{Q\in\mathcal{J}} \int_{E(Q)} \left(M_v\left(fw^{\frac{1}{q}}\right)\right)^q M_w(h) w^{\frac{1}{q+1}} w^{-\frac{1}{q+1}}\mathrm{d}\mu \\
&\leq 2[w]_{A_{q+1}} \left(\int_X \left(M_v\left(fw^{\frac{1}{q}}\right)\right)^{q+1} v\mathrm{d}\mu\right)^{\frac{q}{q+1}} \left(\int_X (M_w(h))^{q+1} w\mathrm{d}\mu\right)^{\frac{1}{q+1}} \\
&= 2[w]_{A_{q+1}} \left\| M_v\left(fw^{\frac{1}{q}}\right)\right\|_{L^{q+1}(v)}^q \|M_w(h)\|_{L^{q+1}(w)} \\
&\leq 2\left(\frac{q+1}{q}\right)^{q+1} [w]_{A_{q+1}} \left\| fw^{\frac{1}{q}}\right\|_{L^{q+1}(v)}^q \|h\|_{L^{q+1}(w)} \\
&= 2\left(\frac{q+1}{q}\right)^{q+1} [w]_{A_{q+1}} \|f\|_{L^{q+1}(w)}^q .
\end{aligned}$$

The last inequality follows from the strong-type inequality for maximal operator. Taking supremum over all h as above, we obtain

$$\|T_{q,\mathcal{J}}f\|_{L^{q+1}(w)}^q \leq 2\left(\frac{q+1}{q}\right)^{q+1} [w]_{A_{q+1}} \|f\|_{L^{q+1}(w)}^q ,$$

and the claim follows. □

We are ready for the proof of the weighted inequality.

Proof of (1.4). It follows directly from the sparse domination (Theorem 2.2), extrapolation and Lemma 2.4, because of the identity $\frac{1}{q}\max\{1, \frac{(q+1)-1}{p-1}\} = \max\{\frac{1}{q}, \frac{1}{p-1}\}$. □

2.2. **Sharpness.** Now we will prove that the exponent $\max\{\frac{1}{q}, \frac{1}{p-1}\}$ appearing in (1.2) and (1.4) cannot be improved. The cases $p \leq q+1$ and $p > q+1$ are dealt with by different methods.

Sharpness for $1 < p \leq q+1$. Let $c > 1$ be a fixed parameter. Consider the probability space equal to the unit interval $(0,1]$ equipped with its Borel subsets and Lebesgue's measure. As the tree $\mathcal{T}$ we take the standard one-dimensional dyadic lattice. Introduce the function $f : (0,1] \to (0,\infty)$ by

$$f = \sum_{n=0}^{\infty} \left(2 - \frac{1}{c}\right)^n \chi_{(2^{-n-1},2^{-n}]}.$$

Note that

$$\begin{aligned}\Delta_{(2^{-n-1},2^{-n}]}f &= \langle f\rangle_{(2^{-n-1},2^{-n}]} - \langle f\rangle_{(0,2^{-n}]}\\ &= \left(2-\frac{1}{c}\right)^n - 2^n\sum_{k=n}^{\infty}\left(2-\frac{1}{c}\right)^k\cdot 2^{-k-1} = \left(2-\frac{1}{c}\right)^n(1-c),\end{aligned}$$

so $S(f,q)\geq (c-1)f$ almost everywhere. Our next observation is that f is an A_1 weight satisfying $[f]_{A_1}\leq c$: for any $x\in(0,1]$ and any dyadic cube $Q\subseteq[0,1)$ containing x we have $\langle f\rangle_Q\leq cf(x)$. Indeed, for any such x and Q, we have two possibilities. If Q is of the form $(0,2^{-n}]$ for some n, then $\langle f\rangle_Q = \left(2-\frac{1}{c}\right)^n\cdot c$, as we computed above; on the other hand, we have $x\leq 2^{-n}$, so $f(x)\geq\left(2-\frac{1}{c}\right)^n\geq c^{-1}\langle f\rangle_Q$. The second possibility is that Q is contained in some interval of the form $(2^{-n-1},2^{-n}]$ (such an interval is determined uniquely). But in such a case, the function f is constant on Q, so $\langle f\rangle_Q = f(x)$. The A_1 condition follows. Therefore, $w:=f^{1-p}$ is an A_p weight: we have $[w]_{A_p}=[f]_{A_{p'}}^{p-1}\leq[f]_{A_1}^{p-1}\leq c^{p-1}$. Putting all the above facts together, we see that for any exponent κ,

$$\frac{\|S(f,q)\|_{L^p(w)}}{[w]_{A_p}^{\kappa}\|f\|_{L^p(w)}}\geq\frac{c-1}{c^{\kappa(p-1)}}. \tag{2.2}$$

Now, if we had $\kappa<1/(p-1)$, then the right-hand side would converge to infinity as $c\to\infty$. Therefore, the optimal exponent in the weighted estimate must be at least $1/(p-1)$, which is exactly what we need in (1.4). Concerning the sharpness of (1.2), we extend the above f and w to the whole real line by setting $f\equiv 0$ and $w\equiv\langle w\rangle_{(0,1]}$ outside $(0,1]$. Then w is still an A_p weight and the characteristic does not change: indeed, if Q is an arbitrary dyadic cube not contained in $(0,1]$, then either $Q\cap(0,1]=\emptyset$ and then w is constant on Q (so $\langle w\rangle_Q\langle w^{1/(1-p)}\rangle_Q^{p-1}=1$), or $(0,1]\subset Q$ and then

$$\begin{aligned}\langle w\rangle_Q\langle w^{1/(1-p)}\rangle_Q^{p-1} &= \langle w\rangle_{(0,1]}\left(\frac{1}{|Q|}\langle w^{1/(1-p)}\rangle_{(0,1]}+\frac{|Q|-1}{|Q|}\langle w\rangle_{(0,1]}^{1/(1-p)}\right)^{p-1}\\ &\leq\langle w\rangle_{(0,1]}\langle w^{1/(1-p)}\rangle_{(0,1]}^{p-1}.\end{aligned}$$

It remains to note that (2.2) is still valid and the sharpness follows.

Sharpness for $p\in(q+1,\infty)$. We will present an adaptation of an argument which can be found in [6] and [21], and whose idea can be tracked back to the paper [14] by Fefferman and Pipher. Let $p_0\in(q+1,\infty)$ be a fixed parameter. Assume that there exists a nondecreasing function $\phi:[1,\infty)\to(0,\infty)$ with $\phi(t)/t^{1/q}\to 0$ as $t\to\infty$, such that

$$\|S(f,q)\|_{L^{p_0}(w)}\leq C\phi([w]_{A_{p_0}})\,\|f\|_{L^{p_0}(w)}\,. \tag{2.3}$$

Then for all $p>p_0$, we have the unweighted estimate

$$\|S(f,q)\|_{L^p(X)}\leq 2^{1/p_0}C\phi(p)\,\|f\|_{L^p(X)}\,, \tag{2.4}$$

which, as we show later, cannot hold. To prove that (2.3) implies (2.4), we use the standard procedure, known as the Rubio de Francia algorithm (a convenient reference is the monograph [8]). Fix $p>p_0$, a non-negative function $h\in L^{(p/p_0)'}(X)$ of norm 1 and define

$$Rh=\sum_{k=0}^{\infty}\frac{M^k h}{\left(2\,\|M\|_{L^{(p/p_0)'}\to L^{(p/p_0)'}}\right)^k}.$$

Then it is straightforward to check that $h(x) \le Rh(x)$ almost everywhere, $\|Rh\|_{L^{(p/p_0)'}(X)} \le 2\|h\|_{L^{(p/p_0)'}(X)} = 2$ and $[Rh]_{A_{p_0}} \le [Rh]_{A_1} \le 2\|M\|_{L^{(p/p_0)'}\to L^{(p/p_0)'}} \le p$. Thus, by (2.3) and Hölder's inequality,

$$\begin{aligned}\int_X (S(f,q)f)^{p_0}\, h\mathrm{d}\mu &\le \int_X (S(f,q))^{p_0}\, Rh\mathrm{d}\mu \\ &\le C^{p_0}\phi([Rh]_{A_{p_0}})^{p_0} \int_X |f|^{p_0} Rh\mathrm{d}\mu \\ &\le C^{p_0}\phi(p)^{p_0}\, \|f\|^{p_0}_{L^p(X)}\, \|Rh\|_{L^{(p/p_0)'}(X)} \le 2C^{p_0}\phi(p)^{p_0}\, \|f\|^{p_0}_{L^p(X)}\,.\end{aligned}$$

Taking the supremum over all h as above, we obtain (2.4). Now we construct an explicit example which contradicts the latter estimate. Consider the probability space $((0,1], \mathcal{B}(0,1), |\cdot|)$ with its dyadic tree, and distinguish the random variable

$$f(x) = \sum_{j=0}^{\infty} \chi_{(2^{-2j-1},2^{-2j}]}(x).$$

Then $\|f\|_{L^p(X)} \le \|f\|_{L^\infty(X)} = 1$. Setting $Q_n = (0, 2^{-n}]$, we easily compute that

$$\langle f\rangle_{Q_{2n}} = \frac{1}{|Q_{2n}|}\sum_{j=n}^{\infty}(2^{-2j} - 2^{-2j-1}) = \frac{2}{3}$$

and

$$\langle f\rangle_{Q_{2n-1}} = \frac{1}{|Q_{2n-1}|}\sum_{j=n}^{\infty}(2^{-2j} - 2^{-2j-1}) = \frac{1}{3}.$$

Consequently, for $2^{-2n-1} < x \le 2^{-2n}$, we have

$$(S(f,q)(x))^q \ge \sum_{j=1}^{n}(\langle f\rangle_{Q_{2j}} - \langle f\rangle_{Q_{2j-1}})^q = \frac{n}{3^q} \gtrsim \ln\left(\frac{1}{x}\right),$$

with a similar estimate for $2^{-2n} < x \le 2^{-2n+1}$. Here and below, the symbol '$\gtrsim$' refers to an inequality which holds up to a multiplicative constant depending only on q. Therefore, we obtain

$$\begin{aligned}\|S(f,q)\|_{L^p(X)} \gtrsim \left(\int_0^1 \left(\ln\left(\frac{1}{x}\right)\right)^{\frac{p}{q}}\right)^{\frac{1}{p}} &= \left(\sum_{k=0}^{\infty}\int_{2^{-k-1}}^{2^{-k}} \left(\ln\left(\frac{1}{x}\right)\right)^{\frac{p}{q}}\right)^{\frac{1}{p}} \\ &\ge \left(\sum_{k=0}^{\infty}\left(\ln\left(\frac{1}{2^{-k}}\right)\right)^{\frac{p}{q}}(2^{-k} - 2^{-k-1})\right)^{\frac{1}{p}} \\ &\gtrsim \left(\sum_{k=0}^{\infty} k^{\frac{p}{q}} 2^{-k}\right)^{\frac{1}{p}} \gtrsim p^{\frac{1}{q}},\end{aligned}$$

where the last inequality follows by taking just one summand corresponding to $k = \lceil p\rceil$. This contradicts (2.4) and finishes the proof in the probabilistic setup. Concerning (1.2), the same reasoning works: in the above extrapolation argument, we replace X with $\mathbb{R}$ throughout. When proving the sharpness in the reduced unweighted context, we extend the extremal function f by setting $f \equiv 0$ outside $(0,1]$.

3. The case $q < 2$

3.1. **Proof of** (1.5)**.** Here the reasoning will exploit the properties of a certain special function of four variables, enjoying appropriate size and concavity requirements. Let $\beta = (1+\sqrt{2})(\sqrt{2}+5+4\sqrt{2}\ln 2)$ and consider the domain

$$\mathcal{D} = [0,\infty) \times \mathbb{R} \times [0,\infty) \times (0,\infty).$$

Introduce the function $B : \mathcal{D} \to \mathbb{R}$ by the formula

$$B(x,y,u,v) = (x^2+y^2)^{\frac{1}{2}}u - \beta xv + 4xv\ln(uv^{-1}+1).$$

Lemma 3.1. *The function B enjoys the following properties.*

(i) *For any $y \in \mathbb{R}$ and $u > 0$,*

$$B(|y|,y,u,u) \leq 0. \tag{3.1}$$

(ii) *For any $(x,y,u,v) \in \mathcal{D}$ we have*

$$B(x,y,u,v) \geq |y|u - \beta xv. \tag{3.2}$$

Proof. The property (i) is straightforward: we have $B(|y|,y,u,u) = (2^{1/2}-\beta+4\ln 2)|y|u \leq 0$. The second majorization is also evident and follows immediately from the estimates $(x^2+y^2)^{1/2}u \geq |y|u$ and $4xv\ln(uv^{-1}+1) \geq 0$. □

The key property of B, which can be regarded as a certain type of concavity condition, is studied in a statement below.

Lemma 3.2. *For any $(x,y,u,v) \in \mathcal{D}$ and $h,d \in \mathbb{R}$ such that $u \leq v$ and $u+d \geq 0$, we have*

$$\begin{aligned} &B\left((x^2+h^2)^{\frac{1}{2}}, y+h, u+d, (u+d)\vee v\right) \\ &\qquad \leq B(x,y,u,v) + B_y(x,y,u,v)h + B_u(x,y,u,v)d. \end{aligned} \tag{3.3}$$

Proof. We start the proof with a simple observation which will be useful later on. For $\phi(s) = \left(x^2+s^2+(y+s)^2\right)^{\frac{1}{2}}$, we have

$$\phi'(s) = \frac{y+2s}{\sqrt{x^2+s^2+(y+s)^2}} \leq \frac{|y+2s|}{\sqrt{\frac{1}{2}(y+2s)^2}} = \sqrt{2}.$$

Consequently, by the mean-value theorem,

$$\left|\left(x^2+h^2+(y+h)^2\right)^{\frac{1}{2}} - \left(x^2+y^2\right)^{\frac{1}{2}}\right| = |\phi(h)-\phi(0)| \leq \sqrt{2}|h|. \tag{3.4}$$

Now, we split the reasoning into three separate parts.

Case I: $u+d \leq v, x \leq |h|$. The desired inequality reads

$$\begin{aligned} &(x^2+h^2+(y+h)^2)^{\frac{1}{2}}(u+d) - \beta(x^2+h^2)^{\frac{1}{2}}v + 4(x^2+h^2)^{\frac{1}{2}}v\ln((u+d)v^{-1}+1) \\ &\leq (x^2+y^2)^{\frac{1}{2}}u - \beta xv + 4xv\ln(uv^{-1}+1) + \frac{yuh}{(x^2+y^2)^{\frac{1}{2}}} + (x^2+y^2)^{\frac{1}{2}}d + \frac{4xd}{uv^{-1}+1}. \end{aligned}$$

This can be rewritten in the equivalent form $I_1 + I_2 + I_3 + I_4 \leq 0$, where

$$\begin{aligned} I_1 &= (x^2 + h^2 + (y+h)^2)^{1/2}(u+d) - (x^2+y^2)^{1/2}(u+d), \\ I_2 &= -\beta(x^2+h^2)^{\frac{1}{2}} v + \beta x v, \\ I_3 &= 4(x^2+h^2)^{\frac{1}{2}} v \ln\left((u+d)v^{-1} + 1\right), \\ I_4 &= -4xv \ln\left(uv^{-1}+1\right) - \frac{yuh}{(x^2+y^2)^{1/2}} - \frac{4xd}{uv^{-1}+1}. \end{aligned}$$

The first term is not bigger than $\sqrt{2}|h|v$, by (3.4) and the assumption $u + d \leq v$. Next, we have the estimates

$$I_2 = -\frac{\beta h^2 v}{\sqrt{x^2+h^2}+x} \leq -\frac{\beta |h| v}{\sqrt{2}+1},$$

and $I_3 \leq 4\sqrt{2}\ln 2|h|v$, by the assumption $x \leq |h|$. Finally, to handle I_4, we use the obvious bounds $-4xv\ln\left(uv^{-1}+1\right) \leq 0$,

$$-\frac{yuh}{(x^2+y^2)^{1/2}} \leq |h|u \leq |h|v \quad \text{and} \quad -\frac{4xd}{uv^{-1}+1} \leq \frac{4xu}{uv^{-1}+1} \leq 4xu \leq 4|h|v$$

to obtain that $I_4 \leq 5|h|v$. Putting all these observations together, we get

$$I_1 + I_2 + I_3 + I_4 \leq |h|v\left(\sqrt{2} - \frac{\beta}{\sqrt{2}+1} + 4\sqrt{2}\ln 2 + 5\right) = 0$$

and the claim follows.

Case II: $u + d \leq v, x \geq |h|$. Here the reasoning rests on the combination of two intermediate estimates. We start with the following inequality

$$B\left((x^2+h^2)^{\frac{1}{2}}, y+h, u+d, v\right) + \frac{vh^2}{x} \leq B(x, y+h, u+d, v). \tag{3.5}$$

To prove this, consider the auxiliary function $F(s) = B\left((x^2+s^2)^{\frac{1}{2}}, y, u, v\right) + \frac{vs^2}{x}$. It is enough to show that $F(s)$ is decreasing on $[0,h]$ if $h > 0$, and increasing on $[h,0]$ if $h < 0$ (indeed, then (3.5) follows, up to the substitution $y := y+h$ and $u := u+d$). If h is positive, then

$$\begin{aligned} F'(s) &= \frac{s}{\sqrt{x^2+s^2}}\left(-\beta v + 4v\ln\left(uv^{-1}+1\right) + \frac{u\sqrt{x^2+s^2}}{\sqrt{x^2+s^2+y^2}}\right) + \frac{2vs}{x} \\ &\leq \frac{vs}{\sqrt{x^2+s^2}}\left(-\beta + 4\ln 2 + 1\right) + \frac{2vs}{x} < -\frac{2\sqrt{2}vs}{\sqrt{x^2+s^2}} + \frac{2vs}{x} < 0, \end{aligned}$$

by the assumption $s \leq |h| \leq x$. The same calculation shows the monotonicity of F for $h < 0$ (then $s \in [h, 0]$ is also negative, so the above inequalities reverse). Having established (3.5), we move to the second step and consider the continuous function $G = G_{x,y,u,v,h,d} : [0,1] \to \mathbb{R}$ given by $G(t) = B(x, y+th, u+td, v)$. We will prove that $G''(t) \leq \frac{2vh^2}{x}$ for $t \in (0,1)$, and this will yield the claim. Indeed, if we manage to bound the second derivative, then by the mean-value theorem we will obtain

$$B(x, y+h, u+d, v) = G(1) \leq G(0) + G'(0) + \frac{vh^2}{x}.$$

Combining this with (3.5), we will get the desired inequality (3.3).

So, we turn our attention to the estimate for $G''(t)$. Simple calculations yield

$$\begin{aligned}
G''(t) &= B_{yy}(x, y+th, u+td, v)h^2 + 2B_{yu}(x, y+th, u+td, v)hd \\
&\quad + B_{uu}(x, y+th, u+td, v)d^2 \\
&= \frac{(u+td)x^2h^2}{(x^2+(y+th)^2)^{\frac{3}{2}}} + \frac{2(y+th)hd}{(x^2+(y+th)^2)^{\frac{1}{2}}} - \frac{4xd^2}{v(1+(u+td)v^{-1})^2} \\
&\leq \frac{vh^2}{x} + 2|h||d| - \frac{xd^2}{v} \leq \frac{2vh^2}{x}.
\end{aligned}$$

This completes the proof in the case $u+d \leq v$.

Case III: $u+d>v$. The assertion reads

$$\begin{aligned}
&(x^2+h^2+(y+h)^2)^{\frac{1}{2}}(u+d) - \beta(x^2+h^2)^{\frac{1}{2}}(u+d) + 4(x^2+h^2)^{\frac{1}{2}}(u+d)\ln 2 \\
&\leq (x^2+y^2)^{\frac{1}{2}}u - \beta xv + 4xv\ln(uv^{-1}+1) + \frac{yuh}{(x^2+y^2)^{\frac{1}{2}}} + (x^2+y^2)^{\frac{1}{2}}d + \frac{4xd}{uv^{-1}+1}.
\end{aligned}$$

If we put all the terms on the left-hand side, we immediately observe that the obtained expression depends linearly on d. Denoting this expression by $L(d)$, we see that it is enough to show that $L'(d) \leq 0$: then the claim will follow from the previous two cases. The estimate $L'(d) \leq 0$ is equivalent to

$$(x^2+h^2+(y+h)^2)^{\frac{1}{2}} - (x^2+y^2)^{\frac{1}{2}} - \beta(x^2+h^2)^{\frac{1}{2}} + 4(x^2+h^2)^{\frac{1}{2}}\ln 2 - \frac{4x}{uv^{-1}+1} \leq 0.$$

However, by (3.4), the left-hand side is bounded from above by

$$\sqrt{2}|h| - (x^2+h^2)^{\frac{1}{2}}(\beta - 4\ln 2) \leq |h|(\sqrt{2} - \beta + 4\ln 2) \leq 0,$$

which completes the proof of the lemma. □

The above special function B leads to the following result, which is of independent interest. We would like to emphasize here that the weight w below is not assumed to belong to any Muckenhoupt class.

Theorem 3.3. *Let w be an arbitrary weight on X (i.e., a nonnegative, integrable random variable) and let $f \in L^1(w)$. Then we have*

$$\|f\|_{L^1(w)} \leq \beta \, \|S(f,2)\|_{L^1(Mw)} \tag{3.6}$$

Proof. We may assume that $\|S(f,2)\|_{L^1(Mw)} < \infty$, since otherwise there is nothing to prove. Also, we may assume that w is strictly positive, by considering the modified weight $w+\varepsilon$ and letting $\varepsilon \to 0$ at the end. Consider the four dimensional process $H_n = (S_n(f,2), f_n, w_n, M_n w)$, $n = 0, 1, 2, \ldots$, where

$$S_n(f,2) = \left(\sum_{k=0}^{n} |df_k|^2\right)^{1/2} \quad \text{and} \quad M_n w = \max_{0\leq k\leq n} w_k$$

are the truncated versions of the square function of f and the (martingale) maximal function of w. Then the sequence $(B(H_n))_{n\geq 0}$ is a supermartingale. To see this, we apply Lemma 3.2 to obtain the pointwise estimate

$$\begin{aligned}
B(H_{n+1}) &= B\Big(((S_n(f,2))^2 + df_{n+1}^2)^{\frac{1}{2}}, f_n + df_{n+1}, w_n + dw_{n+1}, (w_n + dw_{n+1}) \vee M_n w\Big) \\
&\leq B(H_n) + B_y(H_n)df_{n+1} + B_u(H_n)dw_{n+1}.
\end{aligned}$$

Both sides are integrable, since the filtration is atomic and hence all the variables f_n, f_{n+1}, H_n, H_{n+1}, etc. take values in a finite set. Therefore we may apply the conditional expectation with respect to $\mathcal{F}_n$ and, as the result, we obtain $\mathcal{E}_n B(H_{n+1}) \leq B(H_n)$. Indeed, this follows from the fact that $B(H_n), B_y(H_n), B_u(H_n)$ are $\mathcal{F}_n$-measurable and $\mathcal{E}_n(df_{n+1}) = \mathcal{E}_n(dw_{n+1}) = 0$. The supermartingale property together with the majorization conditions (3.1) and (3.2) imply

$$\begin{aligned}\int_X |f_n| w_n \mathrm{d}\mu - \beta \int_X S_n(f,2) M_n w \mathrm{d}\mu &\leq \int_X B(H_n)\mathrm{d}\mu \\ &\leq \int_X B(H_0)\mathrm{d}\mu = \int_X B(|f_0|, f_0, w_0, w_0)\mathrm{d}\mu \leq 0.\end{aligned}$$

Therefore, we get $\int_X |f_n| w \mathrm{d}\mu \leq \beta \int_X S_n(f,2) M_n w \mathrm{d}\mu \leq \beta \int_X S(f,2) M w \mathrm{d}\mu$. Taking the supremum over all $n \geq 0$ finishes the proof. □

Now we are ready for the proof of the weighted inequality for q-functions.

Proof of (1.5). For any $1 < q \leq 2$ we have the pointwise estimate $S(f,2) \leq S(f,q)$ and hence the previous theorem gives

$$\|f\|_{L^1(w)} \leq \beta \|S(f,q)\|_{L^1(Mw)} \leq \beta [w]_{A_1} \|S(f,q)\|_{L^1(w)}.$$

Consequently, by Theorem 2.3, we obtain that for any $p > 1$ we have

$$\|f\|_{L^p(w)} \leq C[w]_{A_p} \|S(f,q)\|_{L^p(w)}.$$

This is the desired claim. □

3.2. **Sharpness.** We will actually prove a stronger statement: for any $1 < p < \infty$ and any $c > 2$ there is a dimension d and a dyadic A_1 weight w on $\mathbb{R}^d$ satisfying

$$[w]_{A_1} = c \qquad \text{and} \qquad \|f\|_{L^p(w)} > \frac{1}{3}[w]_{A_1} \|S(f,q)\|_{L^p(w)}. \tag{3.7}$$

We start with a certain inductive procedure, which leads to a certain class of dyadic cubes contained in $[0,1)^d$. The algorithm is as follows: first, we set $\mathcal{J}_0 = \{[0,1)^d\}$. Next, if $\mathcal{J}_n$ has been defined, then each element Q of $\mathcal{J}_n$ is split into 2^d children and one of the children, denoted by $C(Q)$, is distinguished. We let $\mathcal{I}_n = \bigcup_{Q \in \mathcal{J}_n} C(Q)$ and call $C(Q)$ the atoms of $\mathcal{I}_n$; the remaining (unselected) children of elements of $\mathcal{J}_n$ are put into $\mathcal{J}_{n+1}$. Directly from this fractal-type construction, we see that the sets $\mathcal{I}_0$, $\mathcal{I}_1$, $\mathcal{I}_2$, ... are pairwise disjoint and $|\mathcal{I}_n| = (1-2^{-d})^n 2^{-d}$; hence, in particular, the union of all $\mathcal{I}_n$'s covers almost the whole $[0,1)^d$.

Next, we define the weight $w : [0,1)^d \to [0,\infty)$ by

$$w = \sum_{n=0}^{\infty} \left(\frac{c2^d - 1}{c2^d - c}\right)^n \chi_{\mathcal{I}_n}.$$

Then w belongs to the class A_1 and $[w]_{A_1} = c$. To see this, pick an arbitrary dyadic cube Q contained in $[0,1)^d$ and let n be the smallest integer such that $Q \cap \mathcal{I}_n \neq \emptyset$. Then $w \geq \left(\frac{c2^d-1}{c2^d-c}\right)^n$ on Q, with equality on $Q \cap \mathcal{I}_n$. Now, if Q is contained in $\mathcal{I}_n$, then w is constant on Q and hence $\langle w \rangle_Q = w$ there. Otherwise, we compute that

$$\langle w \rangle_Q = \frac{1}{|Q|} \sum_{k=n}^{\infty} \left(\frac{c2^d - 1}{c2^d - c}\right)^k |Q \cap \mathcal{I}_k| = \sum_{k=n}^{\infty} \left(\frac{c2^d - 1}{c2^d - c}\right)^k \cdot (1-2^{-d})^{k-n} 2^{-d} = c\left(\frac{c2^d - 1}{c2^d - c}\right)^n,$$

so $\langle w\rangle_Q \le cw$ on Q and equality holds on $Q\cap\mathcal{I}_n$. This proves that $[w]_{A_1}=c$.

Next, pick a parameter $r>p$ and define the function

$$f=\sum_{n=0}^{\infty}\left(1+\frac{1}{rc(2^d-1)}\right)^n\chi_{\mathcal{I}_n}.$$

Then we have

$$\|f\|_{L^p(w)}^p=\sum_{n=0}^{\infty}\left(1+\frac{1}{rc(2^d-1)}\right)^{pn}\left(\frac{c2^d-1}{c2^d-c}\right)^n\cdot(1-2^{-d})^n2^{-d}.$$

The ratio of the above geometric series is equal to

$$\left(1+\frac{1}{rc(2^d-1)}\right)^p\left(1+\frac{c-1}{c(2^d-1)}\right)(1-2^{-d})=1+\frac{p-r}{rc(2^d-1)}+o(2^{-d})$$

as $d\to\infty$. There are two important consequences: first, since $r>p$, we have $f\in L^p(w)$ for large dimensions. Second, no matter how large the parameter d is, if we chose r close to p, then the ratio is close to 1 and we can make $\|f\|_{L^p(w)}$ arbitrarily large.

To study the properties of the q-function, pick an arbitrary dyadic subcube R of $[0,1)^d$ and let n be the smallest integer such that $R\cap\mathcal{I}_n\neq\emptyset$. Arguing as above, we check that

$$\langle f\rangle_R=\begin{cases}\left(1+\dfrac{1}{rc(2^d-1)}\right)^n & \text{if } R\subseteq\mathcal{I}_n,\\ \left(1+\dfrac{1}{rc(2^d-1)}\right)^n\cdot\dfrac{rc}{rc-1} & \text{otherwise.}\end{cases}$$

Consequently, if we have an atom Q of $\mathcal{I}_n$ and $[0,1)^d=Q_0\supset Q_1\supset Q_2\supset\ldots\supset Q_n\supset Q$ is the sequence of dyadic cubes decreasing to Q, then $\Delta_{Q_0}f=\frac{rc}{rc-1}$,

$$\Delta_{Q_j}f=\langle f\rangle_{Q_j}-\langle f\rangle_{Q_{j-1}}=\frac{1}{(rc-1)(2^d-1)}\left(1+\frac{1}{rc(2^d-1)}\right)^{j-1}\qquad j=1,2,\ldots,n,$$

and

$$\Delta_Q f=\langle f\rangle_Q-\langle f\rangle_{Q_n}=-\frac{1}{rc-1}\left(1+\frac{1}{rc(2^d-1)}\right)^n.$$

Since f is constant on Q, all the remaining differences vanish on this set. These identities imply that on Q, we have the following inequality (for brevity, we set $\delta=(rc(2^d-1))^{-1}$)

$$\begin{aligned}S(f,q)^q&=\sum_{j=0}^{n}|\Delta_{Q_j}f|^q+|\Delta_Q f|^q\\&=\left(\frac{rc}{rc-1}\right)^q+\sum_{j=1}^{n}\left(\frac{rc}{rc-1}\right)^q(1+\delta)^{(j-1)q}\,\delta^q+\frac{(1+\delta)^{nq}}{(rc-1)^q}\\&\le\left(\frac{rc}{rc-1}\right)^q+\left(\frac{rc}{rc-1}\right)^q\cdot\frac{(1+\delta)^{nq}\delta^q}{(1+\delta)^q-1}+\frac{(1+\delta)^{nq}}{(rc-1)^q}\\&\le\left(\frac{rc}{rc-1}\right)^q+\left(\frac{rc}{rc-1}\right)^q\cdot(1+\delta)^{nq}\delta^{q-1}+\frac{(1+\delta)^{nq}}{(rc-1)^q}.\end{aligned}$$

Therefore, we have the pointwise estimate

$$S(f,q)\le\frac{rc}{rc-1}+\left(\frac{rc}{rc-1}\delta^{(q-1)/q}+\frac{1}{rc-1}\right)f$$

and hence

$$\frac{\|S(f,q)\|_{L^p(w)}}{\|f\|_{L^p(w)}} \leq \frac{\frac{rc}{rc-1}\|w\|_{L^1}^{1/p} + \left(\frac{rc}{rc-1}\delta^{(q-1)/q} + \frac{1}{rc-1}\right)\|f\|_{L^p(w)}}{\|f\|_{L^p(w)}}$$
$$= \frac{rc^{1+1/p}}{(rc-1)\|f\|_{L^p(w)}} + \frac{rc}{rc-1}\delta^{(q-1)/q} + \frac{1}{rc-1}.$$

Now we need to perform a limiting procedure. Let us keep $c > 2$ fixed, then pick r sufficiently close to p and finally let d be a large number (so that δ is close to zero). If this is done appropriately, then the norm $\|f\|_{L^p(w)}$ can be made as large as we wish, in particular, we can make the first term $rc^{1+1/p}/\big((rc-1)\|f\|_{L^p(w)}\big)$ smaller than $1/c$. In addition, the sum of the remaining two terms can be made smaller than $2/c$. Putting all these facts together, we see that we have constructed a pair f, w for which (3.7) is satisfied. This is precisely the desired claim in the probabilistic context of (1.5). Concerning (1.3), we extend f and w to the whole $\mathbb{R}^d$ by setting $f \equiv 0$ and $w \equiv \langle w\rangle_{(0,1]^d}$ outside $(0,1]^d$. Then, arguing as in the proof of (1.2), one checks that w is still an A_1 weight and the characteristic is preserved. When studying the q-function, a little more effort is needed since we are interested in the *upper* bound for $\|S(f,q)\|_{L^p(w)}$. However, it is not difficult to see that the surplus in $\|S(f,q)\|_{L^p(w)}$ can be bounded from above by a quantity that does not depend on r or d. This shows that the above argument extends with no changes and yields the optimality of the exponent.

Acknowledgments

The authors would like to thank an anonymous Referee for the careful reading of the paper and several helpful suggestions.

References

[1] G. Alsmeyer and U. Rösler, *The best constant in the Topchii-Vatutin inequality for martingales*, Statist. Probab. Lett. **65** (2003), 199–206.

[2] B. von Bahr and C.-G. Esseen, *Inequalities for the rth absolute moment of a sum of random variables*, $1 \leq r \leq 2$, Ann. Math. Statist. **36** (1965), 299–303.

[3] R. Bañuelos and A. Osękowski, *Sharp Weighted L^2 inequalities for square functions*, Trans. Amer. Math. Soc. **370** (2018), 2391–2422.

[4] R. Bañuelos and A. Osękowski, *Weighted square function estimates*, to appear in Bull. Sci. Math.

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

[6] D. C. Cox, *The best constant in Burkholder's weak-L^1 inequality for the martingale square function.* Proc. Amer. Math. Soc. **85** (1982), 427–433.

[7] D. C. Cox, *Note on a martingale inequality of Pisier*, Math. Proc. Cambridge Philos. Soc. **92** (1982), no. 1, 163–165.

[8] D. Cruz-Uribe, J. M. Martell, and C. Pérez. *Weights, extrapolation and the theory of Rubio de Francia.* Birkhäuser, Basel, 2011.

[9] D. Cruz-Uribe, J. M. Martell and C. Pérez, *Sharp weighted estimates for classical operators*, Adv. Math. 229 (2012), 408–441.

[10] A. Culiuc, F. Di Plinio, Y. Ou, *Uniform sparse domination of singular integrals via dyadic shifts*, Math. Res. Lett. **25** (2018), 21–42.

[11] K. Domelevo, P. Ivanisvili, S. Petermichl, S. Treil and A. Volberg, *On the failure of lower square function estimates in the non-homogeneous weighted setting*, Math. Ann. **374** (2019), No. 3-4, 1923–1952.

[12] O. Dragičević, L. Grafakos, M. Pereyra, and S. Petermichl. *Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces.* Publ. Mat., 49(1):73—91, 2005.

[13] J. Duoandikoetxea, *Extrapolation of weights revisited: new proofs and sharp bounds*, J. Funct. Anal. 260 (2011), no. 6, 1886–1901.

[14] R. Fefferman and J. Pipher, *Multiparameter Operators and Sharp Weighted Inequalities* , Amer. J. Math. **119** (1997), 337–369.

[15] P. Hitczenko, *Best constants in martingale version of Rosenthal's inequality*, Ann. Probab. **18** (1990), no. 4, 1656–1668.

[16] S. Hukovič, S. Treil and A. Volberg, *The Bellman functions and sharp weighted inequalities for square functions*, Complex analysis, operators, and related topics, 97–113, Oper. Theory Adv. Appl., 113, Birkhauser, Bäsel, 2000.

[17] R. C. James, *Superreflexive spaces with bases*, Pacific J. Math. **41** (1972), 309–419.

[18] R. C. James, *Superreflexive Banach spaces*, Can. J. Math. **24** (1972), 896–904.

[19] S. Kwapień, W. A. Woyczyński, *Random series and stochastic integral: single and multiple*, 1992 Birkhäuser.

[20] M. T. Lacey, *An elementary proof of the A_2 bound*. Isr. J. Math. 217, 181–195 (2017).

[21] A. Lerner. *On some sharp weighted norm inequalities*. J. Funct. Anal., 232(2):477–494, 2006.

[22] A. K. Lerner, Sharp weighted norm inequalities for Littlewood-Paley operators and singular integrals, Adv. Math. 226 (2011), 3912–3926.

[23] A. Lerner, *On an estimate of Calderón-Zygmund operators by dyadic positive operators*, J. Anal. Math. **121** (2013), 141–161.

[24] K. Li, *Sparse Domination Theorem for Multilinear Singular Integral Operators with L_r-Hörmander Condition*, Michigan Math. J. **67** (2018), 253–265.

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

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

[27] I. Pinelis, *Best possible bounds of the von Bahr-Esseen type*, Ann. Funct. Anal. **6** (2015), no. 4, 1–29.

[28] I. Pinelis, *Rosenthal-type inequalities for martingales in 2-smooth Banach spaces*, Theory Probab. Appl. **59** (2015), no. 4, 699–706.

[29] I. Pinelis, *Exact Rosenthal-type bounds*, Ann. Probab. **43** (2015), no. 5, 2511–2544.

[30] G. Pisier, *Martingales with values in uniformly convex spaces*, Israel J. Math. **20** (1975), 326–350.

[31] H. P. Rosenthal, *On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables*, Israel J. Math. **8** (1970) 273–303.

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

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

[34] G. Wang. *Sharp inequalities for the conditional square function of a martingale*. Ann. Probab., 19(4):1679–1688, 1991.

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

Email address: `T.Galazka@mimuw.edu.pl`

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

Email address: `A.Osekowski@mimuw.edu.pl`