STUDIA MATHEMATICA 237 (1) (2017)

Two-weight L^p-inequalities for dyadic shifts and the dyadic square function

by

Emil Vuorinen (Helsinki)

Abstract. We consider two-weight $L^p \to L^q$-inequalities for dyadic shifts and the dyadic square function with general exponents $1 < p, q < \infty$. It is shown that if a so-called quadratic $\mathscr{A}_{p,q}$-condition related to the measures holds, then a family of dyadic shifts satisfies the two-weight estimate in an $\mathcal{R}$-bounded sense if and only if it satisfies the direct and the dual quadratic testing condition. In the case $p = q = 2$ this reduces to the result by T. Hytönen, C. Pérez, S. Treil and A. Volberg (2014).

The dyadic square function satisfies the two-weight estimate if and only if it satisfies the quadratic testing condition, and the quadratic $\mathscr{A}_{p,q}$-condition holds. Again in the case $p = q = 2$ we recover the result by F. Nazarov, S. Treil and A. Volberg (1999).

An example shows that in general the quadratic $\mathscr{A}_{p,q}$-condition is stronger than the Muckenhoupt type $A_{p,q}$-condition.

1. Introduction. The main purpose of this note is to consider two-weight norm inequalities for dyadic shifts and the dyadic square function. A *two-weight* $L^p \to L^q$*-inequality*, $1 < p, q < \infty$, for an operator T defined for a suitable class of functions is an inequality of the form

$$\Big(\int_{\mathbb{R}^n} |Tf|^q w\, dx\Big)^{1/q} \le C\Big(\int_{\mathbb{R}^n} |f|^p v\, dx\Big)^{1/p}, \tag{1.1}$$

where the constant $C > 0$ does not depend on f. Here v and w are *weights*, that is, non-negative Borel measurable functions. The two-weight inequality (1.1) can also be formulated a little differently, which will be done later.

Dyadic shifts are in a sense discrete models of Calderón–Zygmund singular integral operators. They are much simpler than a general Calderón–Zygmund operator, but they already have the complication that they are not *positive* integral operators.

2010 *Mathematics Subject Classification*: Primary 42B20; Secondary 42B25.
Key words and phrases: dyadic shift, dyadic square function, two-weight inequality, testing condition.
Received 13 May 2015; revised 1 July 2016.
Published online 27 January 2017.

It was shown in [5] that a general Calderón–Zygmund operator can be represented as an average over all dyadic systems on $\mathbb{R}^N$ of a rapidly convergent series of dyadic shifts. This representation was used to prove the so-called A_2-*conjecture* about sharp constants in *one-weight* estimates for Calderón–Zygmund operators.

Dyadic shifts fall also in the category of *well localized operators* as defined by F. Nazarov, S. Treil and A. Volberg [14] . They showed that a two-weight inequality holds for a well localized operator in L^2 if and only if the operator satisfies the so-called *Sawyer type testing conditions*. This means that it suffices to show that the operator and its formal adjoint satisfy the inequality with an arbitrary indicator of a (dyadic) cube, and hence Sawyer type testing may also be called *indicator testing.* Two-weight $L^p \to L^q$-inequalities for well localized operators were considered in [17].

The definition of a well localized operator depends on a parameter r which measures how "well" the operator is localized. The constant C in the two-weight inequality proved in [14] and [17] depends on r and the constants in Sawyer type testing conditions.

In [7] dyadic shifts were looked at from a little different perspective. There T. Hytönen, C. Pérez, S. Treil and A. Volberg proved a two-weight inequality in L^2 assuming the Sawyer type testing conditions and finiteness of the so-called A_2-constant related to the weights. This approach was related to the A_2-conjecture mentioned above, and this is the point of view that we take in this note. The main difference between this approach and the more general point of view of well localized operators is that this way one gets a better estimate depending on the *complexity* of the shift, which was crucial in the A_2-conjecture. The complexity of the shift is somewhat analogous to the "well localization" parameter in the definition of well localized operators.

The novelty here is that we characterize the two-weight inequality for dyadic shifts for general exponents $1 < p, q < \infty$, whereas it was only done before in the case $p = q = 2$. Despite the positive result in the case $p = q = 2$, F. Nazarov has constructed an example (unpublished) of a Haar multiplier (a special kind of dyadic shift) and a pair of weights such that the operator satisfies the Sawyer type testing conditions for some $1 < p = q < \infty$, $p \neq 2$, but still does not satisfy the (quantitative) two-weight estimate. See [17, Section 4] for a more precise statement of the example.

Knowing that there are problems with Sawyer type testing and general exponents $p \in (1, \infty)$, we generalize the testing conditions for $1 < p < \infty$ in the spirit of $\mathcal{R}$-bounded operator families, as used for example in [18]. We call these new testing conditions *quadratic testing conditions.* Similarly, we interpret the A_2-condition as a special case of a *quadratic* $\mathscr{A}_{p,q}$-*condition* (see Section 3 for the definition).

Now we state a special version of the main Theorem 5.1 for dyadic shifts. It is assumed here that we have some fixed underlying dyadic lattice $\mathscr{D}$ on $\mathbb{R}^N$ which is used in the definition of the shifts and the $\mathscr{A}_{p,q}$-condition.

THEOREM 1.1. *Fix $p, q \in (1, \infty)$, and assume that σ and w are two measures on $\mathbb{R}^N$ satisfying the quadratic $\mathscr{A}_{p,q}$-condition. Suppose T^σ is a dyadic shift with complexity κ, and let T^w be the formal adjoint of T^σ. Then there exists a constant C such that the inequality*

$$\|T^\sigma f\|_{L^q(w)} \le C\|f\|_{L^p(\sigma)} \tag{1.2}$$

holds for all $f \in L^p(\sigma)$ if and only if there exist constants C' and C'' such that for all sequences $(Q_i)_{i=1}^\infty \subset \mathscr{D}$ of dyadic cubes and all sequences $(a_i)_{i=1}^\infty$ of real numbers we have

$$\Big\|\Big(\sum_{i=1}^\infty (a_i 1_{Q_i} T^\sigma 1_{Q_i})^2\Big)^{1/2}\Big\|_{L^q(w)} \le C'\Big\|\Big(\sum_{i=1}^\infty a_i^2 1_{Q_i}\Big)^{1/2}\Big\|_{L^p(\sigma)}, \tag{1.3}$$

$$\Big\|\Big(\sum_{i=1}^\infty (a_i 1_{Q_i} T^w 1_{Q_i})^2\Big)^{1/2}\Big\|_{L^{p'}(\sigma)} \le C''\Big\|\Big(\sum_{i=1}^\infty a_i^2 1_{Q_i}\Big)^{1/2}\Big\|_{L^{q'}(w)}. \tag{1.4}$$

Moreover, if $\mathcal{T}^\sigma$ and $\mathcal{T}^w$ denote the best possible constants in (1.3) *and* (1.4), *respectively, and $[\sigma, w]_{p,q}$ is the quadratic $\mathscr{A}_{p,q}$-constant, then the best constant $\|T\|$ in* (1.2) *satisfies*

$$\|T\| \lesssim (1+\kappa)(\mathcal{T}^\sigma + \mathcal{T}^w) + (1+\kappa)^2[\sigma, w]_{p,q}. \tag{1.5}$$

If $p = q = 2$, quadratic testing is equivalent to indicator testing and the quadratic $\mathscr{A}_{2,2}$-condition is equivalent to the simple A_2-condition. Thus, when $p = q = 2$, the above theorem reduces to the one proved in [7].

As another novelty, in Theorem 5.1 we shall actually consider a family $\mathscr{T}$ of dyadic shifts with complexity at most a given κ. Then it is shown that under the quadratic $\mathscr{A}_{p,q}$-condition, the family is $\mathcal{R}$-bounded with the same quantitative bound as in (1.5) if and only if a quadratic testing condition for the whole family is satisfied. Our proof follows the broad outlines of L^2-theory but with additional complications coming from the general exponents. We also briefly outline the proof that if the dyadic shifts are of a special form that arises naturally in the representation theorem concerning general Calderón–Zygmund operators, then a certain weakening of the $\mathscr{A}_{p,q}$-condition is sufficient.

It will be shown that this quadratic $\mathscr{A}_{p,q}$-constant is comparable to the constant in the "two-weight Stein inequality" for conditional expectations from L^p into L^q in the same way as the usual A_2-constant is related to boundedness of conditional expectations in weighted L^2. We also construct an example showing that for $p > 2$ or $1 < q < 2$ the $\mathscr{A}_{p,q}$-condition is in general stronger than the simple $A_{p,q}$-condition. Since they are equivalent in the case $1 < p \leq 2 \leq q < \infty$, we deduce that the simple $A_{p,q}$-condition is sufficient for the two-weight Stein inequality to hold if and only if $1 < p \leq 2 \leq q < \infty$.

The two-weight inequality for the dyadic square function was characterized in L^2 in terms of Sawyer type testing and the A_2-condition in another paper by F. Nazarov, S. Treil and A. Volberg [13]. We use similar ideas to those for dyadic shifts and show that the two-weight inequality for the dyadic square function holds from L^p into L^q if and only if the quadratic testing condition and the quadratic $\mathscr{A}_{p,q}$-condition hold, and we get a similar quantitative estimate to the one for dyadic shifts. Here again we get the previous result as a special case when $p = q = 2$. Our approach to the dyadic square function is inspired by the strategy in [10], and similar steps appeared also in [13].

2. Set up and preliminaries. We begin by specifying the basic notation and concepts we use. Two Radon measures σ and w on $\mathbb{R}^N$ are fixed. Most of the definitions below are made with respect to the measure σ, but it will be clear that they are similar for any Radon measure.

For any $1 \leq p \leq \infty$ the usual L^p-space with respect to the measure σ is denoted by $L^p(\sigma)$. For a sequence $(f_i)_{i=1}^{\infty}$ of Borel measurable functions on $\mathbb{R}^N$ we define

$$\|(f_i)_{i=1}^{\infty}\|_{L^p(\sigma;l^2)} := \Big(\int \Big(\sum_{i=1}^{\infty} |f_i|^2\Big)^{p/2} d\sigma\Big)^{1/p},$$

and the space $L^p(\sigma; l^2)$ consists of those sequences $(f_i)_{i=1}^{\infty}$ for which this norm is finite. All our functions will be real-valued.

We fix a dyadic lattice $\mathscr{D}$ on $\mathbb{R}^N$. This means that $\mathscr{D} = \bigcup_{k\in\mathbb{Z}} \mathscr{D}_k$, where each $\mathscr{D}_k$ is a disjoint cover of $\mathbb{R}^N$ with cubes of the form $x+[0, 2^{-k})^N$, $x \in \mathbb{R}^N$, and for every $k \in \mathbb{Z}$ each cube $Q \in \mathscr{D}_k$ is a union of 2^N cubes in $\mathscr{D}_{k+1}$.

If $Q \in \mathscr{D}_k$, denote by $Q^{(1)}$ the unique cube in $\mathscr{D}_{k-1}$ that contains Q, and for any integer $r \geq 2$ define inductively $Q^{(r)} := (Q^{(r-1)})^{(1)}$. Write also $Q^{(0)} := Q$. For $m = 0, 1, 2, \ldots$ the collection $\mathrm{ch}^{(m)}(Q)$ consists of those $Q' \in \mathscr{D}$ such that $Q'^{(m)} = Q$, and we abbreviate $\mathrm{ch}^{(1)}(Q) =: \mathrm{ch}(Q)$. The side length of a cube $Q \in \mathscr{D}_k$ is $l(Q) := 2^{-k}$, and the volume $l(Q)^N$ is written as $|Q|$.

Martingale decomposition. If $Q \in \mathscr{D}$, then the average of a locally σ-integrable function f over Q is denoted by

$$\langle f\rangle_Q^{\sigma} := \frac{1}{\sigma(Q)} \int_Q f\, d\sigma$$

with the understanding that $\langle f\rangle_Q^\sigma = 0$ if $\sigma(Q) = 0$. For two functions f and g we write $\langle f, g\rangle_\sigma := \int fg\,d\sigma$ whenever the integral makes sense. The *averaging* or *conditional expectation operator* $\mathbb{E}_k$, $k \in \mathbb{Z}$, is defined as

$$\mathbb{E}_k^\sigma f := \sum_{Q\in\mathscr{D}_k} \langle f\rangle_Q^\sigma 1_Q.$$

The martingale difference related to a cube $Q \in \mathscr{D}$ is defined as

$$\varDelta_Q^\sigma f := \sum_{Q'\in\mathrm{ch}(Q)} \langle f\rangle_{Q'}^\sigma 1_{Q'} - \langle f\rangle_Q^\sigma 1_Q. \tag{2.1}$$

Let $(\varepsilon_i)_{i=1}^\infty$ be a sequence of independent random signs on some probability space $(\varOmega, \mathbb{P})$. This means that the sequence is independent and $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$ for all i. We will use the Kahane–Khinchin inequality [8] saying that for any Banach space X, any $1 \le p, q < \infty$ and any $x_1, \ldots, x_M \in X$ we have

$$\Big(\mathbb{E}\Big\|\sum_{i=1}^M \varepsilon_i x_i\Big\|_X^q\Big)^{1/q} \simeq_{p,q} \Big(\mathbb{E}\Big\|\sum_{i=1}^M \varepsilon_i x_i\Big\|_X^p\Big)^{1/p}, \tag{2.2}$$

where $\mathbb{E}$ refers to the expectation with respect to the random signs.

The notation $\simeq_{p,q}$ in (2.2) means that there exists a constant $C > 0$ depending only on p and q and not on M, X or the elements x_i such that if A and B denote the left and right hand sides of (2.2), respectively, then $C^{-1}B \le A \le CB$. The subscript on $\simeq$ indicates what the constant C depends on, and is sometimes omitted. We use this kind of notation only if the constant C does not depend on any relevant information in the situation, and no confusion should arise. Similarly $A \le CB$ will be written as $A \lesssim B$.

Let $f \in L^p(\sigma)$ for some $1 < p < \infty$. Then for any $l \in \mathbb{Z}$ we have the martingale difference decomposition

$$f = \sum_{Q\in\mathscr{D}_l} \langle f\rangle_Q^\sigma 1_Q + \sum_{\substack{Q\in\mathscr{D}\\ l(Q)\le 2^{-l}}} \varDelta_Q^\sigma f, \tag{2.3}$$

where the series converges to f in any order (that is, unconditionally). Burkholder's inequality

$$\|f\|_{L^p(\sigma)} \simeq_p \Big\|\Big(\sum_{Q\in\mathscr{D}_l} |\langle f\rangle_Q^\sigma|^2 1_Q + \sum_{\substack{Q\in\mathscr{D}\\ l(Q)\le 2^{-l}}} |\varDelta_Q^\sigma f|^2\Big)^{1/2}\Big\|_{L^p(\sigma)} \tag{2.4}$$

implies that

$$\|f\|_{L^p(\sigma)} \simeq \mathbb{E}\Big\|\sum_{Q\in\mathscr{D}_l} \varepsilon_Q \langle f\rangle_Q^\sigma 1_Q + \sum_{\substack{Q\in\mathscr{D}\\ l(Q)\le 2^{-l}}} \varepsilon_Q \varDelta_Q^\sigma f\Big\|_{L^p(\sigma)}, \tag{2.5}$$

where $\{\varepsilon_Q\}_{Q\in\mathscr{D}}$ is a collection of independent random signs. Burkholder's inequality (2.4) was originally proved in [1] in a little different situation.

From (2.5) and the Kahane–Khinchin inequalities one can deduce the following lemma for $L^p(\sigma;l^2)$-norms. Below we shall also call (2.6) Burkholder's inequality.

Lemma 2.1. *Let $1<p<\infty$ and $(f_k)_{k=-\infty}^{\infty}\in L^p(\sigma;l^2)$. Then for any $l\in\mathbb{Z}$,*

$$\|(f_k)_{k=-\infty}^{\infty}\|_{L^p(\sigma;l^2)} \simeq_p \Big\|\Big(\sum_{k=-\infty}^{\infty}\sum_{Q\in\mathscr{D}_l}|\langle f_k\rangle_Q^\sigma|^2 1_Q + \sum_{k=-\infty}^{\infty}\sum_{\substack{Q\in\mathscr{D}\\ l(Q)\le 2^{-l}}}|\Delta_Q^\sigma f_k|^2\Big)^{1/2}\Big\|_{L^p(\sigma)}. \tag{2.6}$$

Proof. By monotone convergence we may assume that only finitely many functions f_k are non-zero. Furthermore, by martingale convergence, we can suppose that for every k there are only finitely many terms in the martingale decomposition of f_k. Thus the sums in the following computation are actually finite.

Let $\{\varepsilon_k\}_{k\in\mathbb{Z}}$ and $\{\varepsilon'_Q\}_{Q\in\mathscr{D}}$ be sequences of independent random signs on some distinct probability spaces, and $\mathbb{E}$ and $\mathbb{E}'$ the corresponding expectations. Then using the Kahane–Khinchin inequalities and (2.5) we compute

$$\begin{aligned}\|(f_k)_{k=-\infty}^{\infty}\|_{L^p(\sigma;l^2)}^p &= \Big\|\Big(\mathbb{E}\Big|\sum_{k=-\infty}^{\infty}\varepsilon_k f_k\Big|^2\Big)^{1/2}\Big\|_{L^p(\sigma)}^p \\ &\simeq \mathbb{E}\int_{\mathbb{R}^N}\Big|\sum_{k=-\infty}^{\infty}\varepsilon_k f_k\Big|^p\,d\sigma \\ &\simeq \mathbb{E}\mathbb{E}'\int_{\mathbb{R}^N}\Big|\sum_{k=-\infty}^{\infty}\sum_{Q\in\mathscr{D}_l}\varepsilon_k\varepsilon'_Q\langle f_k\rangle_Q^\sigma 1_Q + \sum_{k=-\infty}^{\infty}\sum_{\substack{Q\in\mathscr{D}\\ l(Q)\le 2^{-l}}}\varepsilon_k\varepsilon'_Q\Delta_Q^\sigma f_k\Big|^p\,d\sigma.\end{aligned} \tag{2.7}$$

If $\{c_{k,Q}\}_{k\in\mathbb{Z},\,Q\in\mathscr{D}}$ is any doubly indexed finitely non-zero set of real numbers, then

$$\begin{aligned}\mathbb{E}\mathbb{E}'\Big|\sum_{k=-\infty}^{\infty}\sum_{Q\in\mathscr{D}}\varepsilon_k\varepsilon'_Q c_{k,Q}\Big|^p &= \mathbb{E}\mathbb{E}'\Big|\sum_{Q\in\mathscr{D}}\varepsilon'_Q\sum_{k=-\infty}^{\infty}\varepsilon_k c_{k,Q}\Big|^p \\ \simeq \mathbb{E}\Big(\mathbb{E}'\Big|\sum_{Q\in\mathscr{D}}\varepsilon'_Q\sum_{k=-\infty}^{\infty}\varepsilon_k c_{k,Q}\Big|^2\Big)^{p/2} &= \mathbb{E}\Big(\mathbb{E}'\Big|\sum_{k=-\infty}^{\infty}\varepsilon_k\sum_{Q\in\mathscr{D}}\varepsilon'_Q c_{k,Q}\Big|^2\Big)^{p/2} \\ \simeq \Big(\mathbb{E}\mathbb{E}'\Big|\sum_{k=-\infty}^{\infty}\varepsilon_k\sum_{Q\in\mathscr{D}}\varepsilon'_Q c_{k,Q}\Big|^2\Big)^{p/2} &= \Big(\sum_{k=-\infty}^{\infty}\sum_{Q\in\mathscr{D}}|c_{k,Q}|^2\Big)^{p/2}.\end{aligned} \tag{2.8}$$

Using (2.8) in (2.7) we get the estimate we wanted. ∎

Principal cubes and Carleson's embedding theorem. We will need the construction of principal cubes. More precisely, suppose $f \in L^1_{\text{loc}}(\sigma)$ and take some cube $Q_0 \in \mathscr{D}$. Set $\mathscr{S}_0 = \{Q_0\}$, and assume that $\mathscr{S}_0, \dots, \mathscr{S}_k$ are defined for some non-negative integer k. Then, for $S \in \mathscr{S}_k$, let $\operatorname{ch}_{\mathscr{S}}(S)$ consist of the maximal cubes $S' \in \mathscr{D}$ such that $S' \subset S$ and

$$\langle |f| \rangle^\sigma_{S'} > 2\langle |f| \rangle^\sigma_S.$$

Set $\mathscr{S}_{k+1} := \bigcup_{S \in \mathscr{S}_k} \operatorname{ch}_{\mathscr{S}}(S)$ and

$$\mathscr{S} := \bigcup_{k=0}^{\infty} \mathscr{S}_k.$$

Now for every $Q \in \mathscr{D}$ with $Q \subset Q_0$, there exists a unique smallest $S \in \mathscr{S}$, denoted by $\pi_{\mathscr{S}} Q$, that contains Q, and it follows from the construction that $\langle |f| \rangle^\sigma_Q \le 2\langle |f| \rangle^\sigma_S$.

Let $\gamma \in (0,1)$. We say that a collection $\mathscr{D}_0 \subset \mathscr{D}$ is γ-*sparse* if there exist pairwise disjoint measurable sets $E(Q) \subset Q$, $Q \in \mathscr{D}_0$, such that $\sigma(E(Q)) \ge \gamma\sigma(Q)$ for all $Q \in \mathscr{D}_0$. The collection $\mathscr{S}$ of stopping cubes constructed above is 1/2-sparse, which is seen by defining $E(S) := S \setminus \bigcup_{S' \in \operatorname{ch}_{\mathscr{S}}(S)} S'$ for $S \in \mathscr{S}$. Related to these sparse families we shall use the following form of *Carleson's embedding theorem*:

Lemma 2.2. *Suppose $1 < p < \infty$, $\gamma \in (0,1)$ and $(f_k)_{k=1}^\infty \subset L^p(\sigma; l^2)$. For each k let $\mathscr{S}_k$ be any γ-sparse collection. Then*

$$\Big\| \Big(\sum_{k=1}^\infty \sum_{S \in \mathscr{S}_k} (\langle f_k \rangle^\sigma_S)^2 1_S \Big)^{1/2} \Big\|_{L^p(\sigma)} \lesssim_{\gamma,p} \Big\| \Big(\sum_{k=1}^\infty f_k^2 \Big)^{1/2} \Big\|_{L^p(\sigma)}. \tag{2.9}$$

Proof. Let M^d_σ be the dyadic maximal function defined for any Borel measurable f by

$$M^d_\sigma(f) = \sup_{Q \in \mathscr{D}} 1_Q \langle |f| \rangle^\sigma_Q.$$

For any k and $S \in \mathscr{S}_k$ denote again by $E_k(S)$ the measurable subset of S such that $\sigma(E_k(S)) \ge \gamma\sigma(S)$ and $E_k(S') \cap E_k(S) = \emptyset$ for any other $S' \in \mathscr{S}_k$.

To prove (2.9), assume without loss of generality that every f_k is non-negative. We want to argue by duality, and for that purpose let $\{g_{k,S} : k = 1, 2, \dots,\ S \in \mathscr{S}_k\}$ be any finitely non-zero collection of $L^{p'}(\sigma)$ functions (p' denotes the Hölder conjugate exponent to p). Then

$$\begin{aligned}\sum_{k=1}^\infty \sum_{S \in \mathscr{S}_k} \int \langle f_k \rangle^\sigma_S 1_S g_{k,S}\, d\sigma &\le \gamma^{-1} \sum_{k=1}^\infty \sum_{S \in \mathscr{S}_k} \langle f_k \rangle^\sigma_S \langle g_{k,S} \rangle^\sigma_S \sigma(E_k(S)) \\ &\le \gamma^{-1} \sum_{k=1}^\infty \sum_{S \in \mathscr{S}_k} \int M^d_\sigma(f_k) M^d_\sigma(g_{k,S}) 1_{E_k(S)}\, d\sigma\end{aligned}$$

$$\leq \gamma^{-1}\Big\|\Big(\sum_{k=1}^{\infty}\sum_{S\in\mathscr{S}_k}(M_\sigma^d(f_k))^2 1_{E_k(S)}\Big)^{1/2}\Big\|_{L^p(\sigma)}$$
$$\times\Big\|\Big(\sum_{k=1}^{\infty}\sum_{S\in\mathscr{S}_k}(M_\sigma^d(g_{k,S}))^2 1_{E_k(S)}\Big)^{1/2}\Big\|_{L^{p'}(\sigma)}.$$

Since for a fixed k the sets $E_k(S)$, $S\in\mathscr{S}_k$, are pairwise disjoint, the first factor on the right hand side satisfies

$$\Big\|\Big(\sum_{k=1}^{\infty}\sum_{S\in\mathscr{S}_k}(M_\sigma^d(f_k))^2 1_{E_k(S)}\Big)^{1/2}\Big\|_{L^p(\sigma)}$$
$$\leq \Big\|\Big(\sum_{k=1}^{\infty}(M_\sigma^d(f_k))^2\Big)^{1/2}\Big\|_{L^p(\sigma)} \lesssim_p \Big\|\Big(\sum_{k=1}^{\infty} f_k^2\Big)^{1/2}\Big\|_{L^p(\sigma)},$$

where in the last step we have used the dyadic Fefferman–Stein inequality [3]. In the second factor we may just omit the indicators $1_{E_k(S)}$ and apply the Fefferman–Stein inequality again. These estimates prove (2.9). ∎

Stein's inequality. Let $(f_k)_{k=-\infty}^{\infty}\in L^p(\sigma;l^2)$, $1<p<\infty$. *Stein's inequality*, which originally appeared in [16], says that

$$\|(E_k^\sigma f_k)_{k=-\infty}^{\infty}\|_{L^p(\sigma;l^2)} \lesssim_p \|(f_k)_{k=-\infty}^{\infty}\|_{L^p(\sigma;l^2)}. \tag{2.10}$$

This can equivalently be formulated by saying that for any set $\{f_Q\}_{Q\in\mathscr{D}}$, where each f_Q is a locally σ-integrable function, we have

$$\Big\|\Big(\sum_{Q\in\mathscr{D}}(\langle f_Q\rangle_Q^\sigma)^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)} \lesssim_p \Big\|\Big(\sum_{Q\in\mathscr{D}} f_Q^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)}. \tag{2.11}$$

Note that (2.10) also follows from the dyadic Fefferman–Stein inequality that was used in the proof of Carleson's embedding theorem.

3. The quadratic $\mathscr{A}_{p,q}$-condition. In this section we introduce the quadratic $\mathscr{A}_{p,q}$-condition and investigate its relation to the Muckenhoupt type $A_{p,q}$-condition. Here $1<p,q<\infty$. The quadratic $\mathscr{A}_{p,q}$-condition will be used in the characterization of two-weight inequalities for the dyadic square function and dyadic shifts.

The measures σ and w are said to satisfy the *simple* or *Muckenhoupt type $A_{p,q}$-condition* if

$$(\sigma,w)_{p,q} := \sup_{Q\in\mathscr{D}}\frac{\sigma(Q)^{1/p'}w(Q)^{1/q}}{|Q|}<\infty. \tag{3.1}$$

If $p=q$, we just write A_p.

The measures σ and w are said to satisfy the *quadratic $\mathscr{A}_{p,q}$-condition* if for every collection $\{a_Q\}_{Q\in\mathscr{D}}$ of real numbers we have

$$\Big\|\Big(\sum_{Q\in\mathscr{D}}\Big(a_Q\frac{\sigma(Q)}{|Q|}\Big)^2 1_Q\Big)^{1/2}\Big\|_{L^q(w)} \le [\sigma,w]_{p,q}\Big\|\Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)}, \tag{3.2}$$

where $[\sigma,w]_{p,q}\in[0,\infty)$ is the best possible constant. We also write $[\sigma,w]_{p,q} < \infty$ to mean that the condition holds, and $[\sigma,w]_{p,q}=\infty$ to mean that it does not hold. It is clear that $(\sigma,w)_{p,q}\le[\sigma,w]_{p,q}$, which follows by taking only one term in the sums in (3.2).

LEMMA 3.1. *Let $1<p,q<\infty$. The quadratic $\mathscr{A}_{p,q}$-condition is symmetric in the sense that $[\sigma,w]_{p,q}\simeq[w,\sigma]_{q',p'}$.*

Proof. Choose any (finitely non-zero) collection $\{a_Q\}_{Q\in\mathscr{D}}$ of real numbers, and let $\{f_Q\}_{Q\in\mathscr{D}}$ be a collection of $L^p(\sigma)$ functions. Then

$$\begin{aligned}\int\sum_{Q\in\mathscr{D}} a_Q\frac{w(Q)}{|Q|}1_Q f_Q\,d\sigma &= \int\sum_{Q\in\mathscr{D}} a_Q\frac{\int_Q f_Q\,d\sigma}{|Q|}1_Q\,dw\\ &\le \Big\|\Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q\Big)^{1/2}\Big\|_{L^{q'}(w)}\Big\|\Big(\sum_{Q\in\mathscr{D}}\Big(\langle|f_Q|\rangle_Q^\sigma\frac{\sigma(Q)}{|Q|}\Big)^2 1_Q\Big)^{1/2}\Big\|_{L^q(w)}\\ &\le [\sigma,w]_{p,q}\Big\|\Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q\Big)^{1/2}\Big\|_{L^{q'}(w)}\Big\|\Big(\sum_{Q\in\mathscr{D}}(\langle|f_Q|\rangle_Q^\sigma)^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)}\\ &\lesssim [\sigma,w]_{p,q}\Big\|\Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q\Big)^{1/2}\Big\|_{L^{q'}(w)}\Big\|\Big(\sum_{Q\in\mathscr{D}}|f_Q 1_Q|^2\Big)^{1/2}\Big\|_{L^p(\sigma)},\end{aligned}$$

where in the last step we have used Stein's inequality. By duality this shows that $[w,\sigma]_{q',p'}\lesssim[\sigma,w]_{p,q}$. ■

For $1<p,q<\infty$ a two-weight version of Stein's inequality (2.11) can be formulated as

$$\Big\|\Big(\sum_{Q\in\mathscr{D}}\Big(\frac{\int_Q f_Q\,d\sigma}{|Q|}\Big)^2 1_Q\Big)^{1/2}\Big\|_{L^q(w)} \le \mathscr{S}\Big\|\Big(\sum_{Q\in\mathscr{D}} f_Q^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)}, \tag{3.3}$$

where $\{f_Q\}_{Q\in\mathscr{D}}$ is again a collection of locally σ-integrable functions, and $\mathscr{S}=\mathscr{S}(\sigma,w,p,q)$ denotes the smallest possible constant with the understanding that it may be infinite.

LEMMA 3.2. *The best constant $\mathscr{S}=\mathscr{S}(\sigma,w,p,q)$ in (3.3) satisfies $\mathscr{S}\simeq[\sigma,w]_{p,q}$.*

Proof. That $[\sigma,w]_{p,q}\le\mathscr{S}(\sigma,w,p,q)$ follows from (3.3) with the special functions $f_Q=a_Q1_Q$, where $a_Q\in\mathbb{R}$. To see that $\mathscr{S}(\sigma,w,p,q)\lesssim[\sigma,w]_{p,q}$, choose any set $\{f_Q\}_{Q\in\mathscr{D}}$ of locally σ-integrable functions. Then

$$\mathrm{LHS}(3.3) = \Big\| \Big(\sum_{Q\in\mathscr{D}} \Big(\langle f_Q\rangle_Q^\sigma \frac{\sigma(Q)}{|Q|}\Big)^2 1_Q \Big)^{1/2} \Big\|_{L^q(w)}$$
$$\le [\sigma,w]_{p,q} \Big\| \Big(\sum_{Q\in\mathscr{D}} (\langle f_Q\rangle_Q^\sigma)^2 1_Q \Big)^{1/2} \Big\|_{L^p(\sigma)} \lesssim [\sigma,w]_{p,q} \Big\| \Big(\sum_{Q\in\mathscr{D}} f_Q^2 1_Q \Big)^{1/2} \Big\|_{L^p(\sigma)},$$

where we have used Stein's inequality (2.11) in the last step. Hence also $[\sigma,w]_{p,q} \lesssim \mathscr{S}(\sigma,w,p,q)$. ∎

The next lemma shows that the quadratic $\mathscr{A}_{p,q}$-condition is actually equivalent to the simple $A_{p,q}$-condition if $1 < p \le 2 \le q < \infty$, and a similar remark will apply to the quadratic testing conditions below.

Lemma 3.3. *If $1 < p \le 2 \le q < \infty$, then $[\sigma,w]_{p,q} = (\sigma,w)_{p,q}$.*

Proof. This follows from the fact that L^p-spaces have certain *type* and *cotype* properties. For our purposes it is not necessary to define these in general, but it suffices to note that for any sequence $(f_k)_{k=1}^\infty \subset L^p(\sigma;l^2)$, $1 < p \le 2$,

$$\Big\| \Big(\sum_{k=1}^\infty f_k^2 \Big)^{1/2} \Big\|_{L^p(\sigma)} \ge \Big(\sum_{k=1}^\infty \|f_k\|_{L^p(\sigma)}^2 \Big)^{1/2}, \tag{3.4}$$

and for any sequence $(g_k)_{k=1}^\infty \subset L^q(\sigma;l^2)$, $2 \le q < \infty$,

$$\Big\| \Big(\sum_{k=1}^\infty g_k^2 \Big)^{1/2} \Big\|_{L^q(\sigma)} \le \Big(\sum_{k=1}^\infty \|g_k\|_{L^q(\sigma)}^2 \Big)^{1/2}. \tag{3.5}$$

Of course these inequalities are independent of the measure.

Suppose now that the simple $A_{p,q}$-condition holds with $1 < p \le 2 \le q < \infty$, and let $\{a_Q\}_{Q\in\mathscr{D}} \subset \mathbb{R}$. Then

$$\Big\| \Big(\sum_{Q\in\mathscr{D}} \Big(a_Q \frac{\sigma(Q)}{|Q|}\Big)^2 1_Q \Big)^{1/2} \Big\|_{L^q(w)} \le \Big(\sum_{Q\in\mathscr{D}} \Big\| a_Q \frac{\sigma(Q)}{|Q|} 1_Q \Big\|_{L^q(w)}^2 \Big)^{1/2} \tag{3.6}$$
$$\le (\sigma,w)_{p,q} \Big(\sum_{Q\in\mathscr{D}} \|a_Q 1_Q\|_{L^p(\sigma)}^2 \Big)^{1/2} \le (\sigma,w)_{p,q} \Big\| \Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q \Big)^{1/2} \Big\|_{L^p(\sigma)},$$

and thus $[\sigma,w]_{p,q} \le (\sigma,w)_{p,q}$. ∎

4. The dyadic square function. In this section we consider the dyadic square function. Let $\{b_Q\}_{Q\in\mathscr{D}}$ be a collection of real numbers. For a locally Lebesgue integrable function the *generalized dyadic square function* is defined by

$$S_b(f) := \Big(\sum_{Q\in\mathscr{D}} (b_Q \Delta_Q f)^2 \Big)^{1/2},$$

where $\Delta_Q f$ is the usual martingale difference related to the cube Q as in (2.1), but with respect to the Lebesgue measure. The "generalized" here refers to the coefficients b_Q, and the usual dyadic square function corresponds to $b_Q = 1$ for all $Q \in \mathscr{D}$.

Now we are interested in the two-weight estimate for this operator. Namely, we fix exponents $1 < p, q < \infty$ and want to characterize when there exists a constant $C \ge 0$ such that the inequality

$$\Big\| \Big(\sum_{Q\in\mathscr{D}} (b_Q \Delta_Q(f\sigma))^2 \Big)^{1/2} \Big\|_{L^q(w)} \le C \|f\|_{L^p(\sigma)} \tag{4.1}$$

holds for all $f \in L^p(\sigma)$. Here $\Delta_Q(f\sigma)$ is understood as

$$\Delta_Q(f\sigma) := \sum_{Q'\in \mathrm{ch}(Q)} \frac{\int_{Q'} f\, d\sigma}{|Q'|} 1_{Q'} - \frac{\int_Q f\, d\sigma}{|Q|} 1_Q.$$

Denote by S_b^σ the operator defined for locally σ-integrable functions f by

$$S_b^\sigma(f) := \Big(\sum_{Q\in\mathscr{D}} (b_Q \Delta_Q(f\sigma))^2 \Big)^{1/2},$$

and for all $Q \in \mathscr{D}$ define the localized version

$$S_{b,Q}^\sigma(f) := \Big(\sum_{\substack{Q'\in\mathscr{D}\\ Q'\subset Q}} (b_{Q'} \Delta_{Q'}(f\sigma))^2 \Big)^{1/2}.$$

If u and v are weight functions on $\mathbb{R}$, that is, positive Borel functions, and $p = q = 2$, a result from [13] says that

$$\|S_b(fu)\|_{L^2(v)} \le C \|f\|_{L^2(u)} \tag{4.2}$$

for all f if and only if there exists a constant C' such that

$$\|S_b(1_I u)\|_{L^2(v)} \le C' \|1_I\|_{L^2(u)} \tag{4.3}$$

for all $I \in \mathscr{D}$. Also in this case the best constants in (4.2) and (4.3) satisfy $C' \simeq C$. Actually a bit more was shown, namely that the two-weight inequality holds if and only if a Muckenhoupt type condition for the measures and a localized testing condition hold.

Here we are going to give a characterization for (4.1) with any $1 < p, q < \infty$. This will be done in terms of a quadratic testing condition and the quadratic $\mathscr{A}_{p,q}$-condition introduced in the last section, and for $p = q = 2$ the theorem reduces to the result from [13].

We say that the operator S_b^σ satisfies the *global quadratic testing condition* (with respect to p and q) if there exists a constant C such that for every collection $\{a_Q\}_{Q\in\mathscr{D}} \subset \mathbb{R}$ we have

$$\Big\| \Big(\sum_{Q\in\mathscr{D}} S_b^\sigma (a_Q 1_Q)^2 \Big)^{1/2} \Big\|_{L^q(w)} \le C \Big\| \Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q \Big)^{1/2} \Big\|_{L^p(\sigma)}. \tag{4.4}$$

The operator S_b^σ is said to satisfy the *local quadratic testing condition* if it similarly satisfies

$$\Big\| \Big(\sum_{Q\in\mathscr{D}} S_{b,Q}^\sigma (a_Q 1_Q)^2 \Big)^{1/2} \Big\|_{L^q(w)} \le C \Big\| \Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q \Big)^{1/2} \Big\|_{L^p(\sigma)}. \tag{4.5}$$

Of course it is equivalent to assume that these inequalities hold for all finitely non-zero collections $\{a_Q\}_{Q\in\mathscr{D}}$.

We shall modify the quadratic $\mathscr{A}_{p,q}$-conditions according to the coefficients b_Q. The measures satisfy the $\mathscr{A}_{p,q}^b$*-condition* if for every collection $\{a_Q\}_{Q\in\mathscr{D}}$ of real numbers we have

$$\Big\| \Big(\sum_{Q\in\mathscr{D}} \Big(a_Q b_Q \frac{\sigma(Q)}{|Q|} \Big)^2 1_Q \Big)^{1/2} \Big\|_{L^q(w)} \le [\sigma, w]_{p,q}^b \Big\| \Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_Q \Big)^{1/2} \Big\|_{L^p(\sigma)}, \tag{4.6}$$

where again $[\sigma, w]_{p,q}^b$ denotes the best possible constant.

Now we can state the two-weight theorem for the dyadic square function:

THEOREM 4.1. *Let $1 < p, q < \infty$. The dyadic square function S_b^σ satisfies the two-weight inequality* (4.1) *if and only if it satisfies the global quadratic testing condition* (4.4) *and if and only if it satisfies the local quadratic testing condition* (4.5) *and the quadratic $\mathscr{A}_{p,q}^b$-condition* (4.6) *holds.*

In this case the best constant $\|S_b^\sigma\|$ in (4.1) *satisfies $\|S_b^\sigma\| \simeq \mathfrak{S}_{\mathrm{glob}} \simeq \mathfrak{S}_{\mathrm{loc}} + [\sigma, w]_{p,q}^b$, where $\mathfrak{S}_{\mathrm{glob}}$ and $\mathfrak{S}_{\mathrm{loc}}$ are the best possible constants in* (4.4) *and* (4.5), *respectively.*

Let us discuss the case $p = q = 2$, or more generally $1 < p \le 2 \le q < \infty$. Similarly to what was noted above in Lemma 3.3, the $\mathscr{A}_{p,q}^b$-condition is equivalent to assuming

$$\sup_{Q\in\mathscr{D}} |b_Q| \frac{\sigma(Q)^{1/p'} w(Q)^{1/q}}{|Q|} \lesssim 1.$$

The same kind of computation shows that the quadratic testing conditions are equivalent to the corresponding Sawyer type testing conditions. For example considering the global testing (4.4), this means that it is enough to assume just

$$\|S_b^\sigma(1_Q)\|_{L^q(w)} \le C \sigma(Q)^{1/p}$$

uniformly for all $Q \in \mathscr{D}$.

With these facts Theorem 4.1 reduces to the result proved in [13] when $p = q = 2$.

Proof of Theorem 4.1. We begin by showing that the global, and hence also the local testing condition is a necessary consequence of the two-weight inequality (4.1). Then we show that global testing implies the quadratic $\mathscr{A}^b_{p,q}$-condition. The main part of the proof is to show that local testing and the $\mathscr{A}^b_{p,q}$-condition are also sufficient for (4.1).

Necessity of the testing conditions. This is very much like a classical theorem of Marcinkiewicz and Zygmund [12], which says that bounded linear operators in L^p-spaces have an extension to a vector-valued situation. Choose a sequence $(f_k)_{k=1}^l \subset L^p(\sigma)$ and let $(\varepsilon_k)_{k=1}^l$ be a sequence of independent random signs. Then using the Kahane–Khinchin inequalities we compute

$$
\begin{aligned}
(4.7)\qquad \Big\|\Big(\sum_{k=1}^{l}|S_b^\sigma(f_k)|^2\Big)^{1/2}\Big\|_{L^q(w)} &= \Big\|\Big(\sum_{Q\in\mathscr{D}}\sum_{k=1}^{l}|b_Q\varDelta_Q(f_k\sigma)|^2\Big)^{1/2}\Big\|_{L^q(w)}\\
&= \Big\|\Big(\sum_{Q\in\mathscr{D}}\mathbb{E}\Big|\sum_{k=1}^{l}\varepsilon_k b_Q\varDelta_Q(f_k\sigma)\Big|^2\Big)^{1/2}\Big\|_{L^q(w)}\\
&= \Big\|\Big(\mathbb{E}\Big\|\Big\{\sum_{k=1}^{l}\varepsilon_k b_Q\varDelta_Q(f_k\sigma)\Big\}_{Q\in\mathscr{D}}\Big\|_{l^2}^2\Big)^{1/2}\Big\|_{L^q(w)}\\
&\simeq \Big(\mathbb{E}\Big\|\Big\|\Big\{\sum_{k=1}^{l}\varepsilon_k b_Q\varDelta_Q(f_k\sigma)\Big\}_{Q\in\mathscr{D}}\Big\|_{l^2}\Big\|_{L^q(w)}^q\Big)^{1/q}\\
&\simeq \mathbb{E}\Big\|\Big\|\Big\{\sum_{k=1}^{l}\varepsilon_k b_Q\varDelta_Q(f_k\sigma)\Big\}_{Q\in\mathscr{D}}\Big\|_{l^2}\Big\|_{L^q(w)},
\end{aligned}
$$

where at the first "$\simeq$" we have used the Kahane–Khinchin inequality in l^2, and at the second in $L^q(w;l^2)$. Linearity of the martingale differences and the assumed two-weight inequality (4.1) imply

$$
\begin{aligned}
(4.8)\qquad \mathrm{RHS}(4.7) &= \mathbb{E}\Big\|S_b^\sigma\Big(\sum_{k=1}^{l}\varepsilon_k f_k\Big)\Big\|_{L^q(w)}\\
&\le \|S_b^\sigma\|\mathbb{E}\Big\|\sum_{k=1}^{l}\varepsilon_k f_k\Big\|_{L^p(\sigma)} \simeq \|S_b^\sigma\|\Big\|\Big(\sum_{k=1}^{l}f_k^2\Big)^{1/2}\Big\|_{L^p(\sigma)},
\end{aligned}
$$

where at "$\simeq$" we have used the Kahane–Khinchin inequality, first in $L^p(\sigma)$ and then in $\mathbb{R}$. From (4.7) and (4.8) it is seen that the two-weight inequality (4.1) implies the global quadratic testing condition (4.4).

Global testing implies the $\mathscr{A}^b_{p,q}$-condition. For any $Q \in \mathscr{D}$ let $\{Q_k\}_{k=1}^{2^N}$ be its dyadic children. If $Q \in \mathscr{D}$ and $k \in \{1,\dots,2^N\}$, then

$$\frac{\sigma(Q_k)}{|Q_k|} \lesssim |\Delta_Q(1_{Q_k}\sigma)(x)|$$

for any $x \in Q$, and thus

$$|a_Q b_Q| \frac{\sigma(Q_k)}{|Q_k|} 1_Q \lesssim S^\sigma_{b,Q}(a_Q 1_{Q_k}).$$

This leads to

$$\begin{aligned}\Big\| \Big(\sum_{Q\in\mathscr{D}} \Big(a_Q b_Q \frac{\sigma(Q_k)}{|Q_k|}\Big)^2 1_Q \Big)^{1/2} \Big\|_{L^q(w)} &\lesssim \Big\| \Big(\sum_{Q\in\mathscr{D}} S^\sigma_{b,Q}(a_Q 1_{Q_k})^2 \Big)^{1/2} \Big\|_{L^q(w)} \\ &\le \Big\| \Big(\sum_{Q\in\mathscr{D}} S^\sigma_{b}(a_Q 1_{Q_k})^2 \Big)^{1/2} \Big\|_{L^q(w)} \le \mathfrak{S}_{\mathrm{glob}} \Big\| \Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_{Q_k} \Big)^{1/2} \Big\|_{L^p(\sigma)} \\ &\le \mathfrak{S}_{\mathrm{glob}} \Big\| \Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_{Q} \Big)^{1/2} \Big\|_{L^p(\sigma)}.\end{aligned}$$

Since

$$\Big(\sum_{Q\in\mathscr{D}} \Big(a_Q b_Q \frac{\sigma(Q)}{|Q|}\Big)^2 1_Q \Big)^{1/2} \le \sum_{k=1}^{2^N} \Big(\sum_{Q\in\mathscr{D}} \Big(a_Q b_Q \frac{\sigma(Q_k)}{|Q_k|}\Big)^2 1_Q \Big)^{1/2},$$

we get $[\sigma, w]^b_{p,q} \lesssim \mathfrak{S}_{\mathrm{glob}}$.

Sufficiency of local testing and the $\mathscr{A}^b_{p,q}$-condition. Now we turn to the main part of the theorem, which consists in showing that local testing and the $\mathscr{A}^b_{p,q}$-condition are sufficient for the estimate (4.1). To this end, fix $f \in L^p(\sigma)$. We can assume here that there are only finitely many non-zero coefficients b_Q in the definition of S^σ_b, and we prove a bound that is independent of this finite number. Of course the original local testing condition implies the same condition for this "truncated" square function.

There are at most 2^N increasing sequences $Q^i_1 \subsetneq Q^i_2 \subsetneq \cdots$, $i = 1,\dots,j \le 2^N$, of dyadic cubes in $\mathscr{D}$ such that

$$\mathbb{R}^N = \bigcup_{i=1}^{j} \bigcup_{k=1}^{\infty} Q^i_k \tag{4.9}$$

and

$$\bigcup_{k=1}^{\infty} Q^i_k \cap \bigcup_{k=1}^{\infty} Q^{i'}_k = \emptyset \quad \text{for } i \ne i'.$$

It follows from the properties of dyadic systems that for every cube $Q \in \mathscr{D}$ there exists $i \in \{1,\dots,j\}$ such that $Q \subset \bigcup_{k=1}^{\infty} Q^i_k$.

Since there are only finitely many non-zero b_Qs, we can choose indices $k_1, \dots, k_j$ such that if $b_Q \neq 0$, then $Q \subset \bigcup_{i=1}^j Q^i_{k_i}$, and we write $\tilde{Q}_i := Q^i_{k_i}$. Thus we can assume that the function f is supported on $\bigcup_{i=1}^j \tilde{Q}_i$. Since $S^\sigma_b f = \sum_{i=1}^j S^\sigma_b(1_{\tilde{Q}_i} f)$, it is enough to bound each of these separately.

The choice of the cubes $\tilde{Q}_i$ implies that $S^\sigma_b(1_{\tilde{Q}_i}) = S^\sigma_{b,\tilde{Q}_i}(1_{\tilde{Q}_i})$, and thus

$$\begin{aligned}\|\langle f\rangle^\sigma_{\tilde{Q}_i} S^\sigma_b(1_{\tilde{Q}_i})\|_{L^q(w)} &= \|\langle f\rangle^\sigma_{\tilde{Q}_i} S^\sigma_{b,\tilde{Q}_i}(1_{\tilde{Q}_i})\|_{L^q(w)} \\ &\leq \mathfrak{S}_{\mathrm{loc}} \|\langle f\rangle^\sigma_{\tilde{Q}_i} 1_{\tilde{Q}_i}\|_{L^p(\sigma)} \leq \mathfrak{S}_{\mathrm{loc}} \|1_{\tilde{Q}_i} f\|_{L^p(\sigma)}.\end{aligned}$$

So finally it is enough to fix some $Q^i_{k_i} =: Q_0$, and assume that the function f is supported on Q_0 and has zero σ-average.

We use a splitting of the function inside the operator, similar to one in [10]; a corresponding step also appeared in [13]. Consider some $Q \in \mathscr{D}$. Since the martingale differences $\Delta^\sigma_Q f$ have σ-integral zero, the term $\Delta_Q(f\sigma)$ in the square function can be written as

$$\Delta_Q(f\sigma) = \Delta_Q\Big(\Big(\Delta^\sigma_Q f + \sum_{R\colon R\supsetneq Q} \Delta^\sigma_R f\Big)\sigma\Big) = \Delta_Q((\Delta^\sigma_Q f)\sigma) + \langle f\rangle^\sigma_Q \Delta_Q(1_Q\sigma).$$

Here we have used the fact that f has zero average to get $\sum_{R\colon R\supsetneq Q} \Delta^\sigma_R f 1_Q = \langle f\rangle^\sigma_Q 1_Q$. Accordingly we split the estimate for the square function into two parts as

$$\begin{aligned}\|S^\sigma_b(f)\|_{L^q(w)} \leq &\Big\|\Big(\sum_{Q\in\mathscr{D}} \big(b_Q \Delta_Q((\Delta^\sigma_Q f)\sigma)\big)^2\Big)^{1/2}\Big\|_{L^q(w)} \\ &+ \Big\|\Big(\sum_{Q\in\mathscr{D}} \big(b_Q \langle f\rangle^\sigma_Q \Delta_Q(1_Q\sigma)\big)^2\Big)^{1/2}\Big\|_{L^q(w)}.\end{aligned} \tag{4.10}$$

For the first term on the right hand side of (4.10) we estimate

$$|\Delta_Q((\Delta^\sigma_Q f)\sigma)| \lesssim \frac{\int |\Delta^\sigma_Q f|\, d\sigma}{|Q|} 1_Q = \langle |\Delta^\sigma_Q f|\rangle^\sigma_Q \frac{\sigma(Q)}{|Q|} 1_Q.$$

This together with the $\mathscr{A}^b_{p,q}$-condition gives

$$\begin{aligned}&\Big\|\Big(\sum_{Q\in\mathscr{D}} \big(b_Q \Delta_Q((\Delta^\sigma_Q f)\sigma)\big)^2\Big)^{1/2}\Big\|_{L^q(w)} \\ &\qquad\lesssim \Big\|\Big(\sum_{Q\in\mathscr{D}} \Big(b_Q \frac{\int |\Delta^\sigma_Q f|\, d\sigma}{|Q|}\Big)^2 1_Q\Big)^{1/2}\Big\|_{L^q(w)} \\ &\qquad\leq [\sigma, w]^b_{p,q} \Big\|\Big(\sum_{Q\in\mathscr{D}} (\langle |\Delta^\sigma_Q f|\rangle^\sigma_Q)^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)} \\ &\qquad\lesssim [\sigma, w]^b_{p,q} \Big\|\Big(\sum_{Q\in\mathscr{D}} (\Delta^\sigma_Q f)^2 1_Q\Big)^{1/2}\Big\|_{L^p(\sigma)} \simeq [\sigma, w]^b_{p,q} \|f\|_{L^p(\sigma)},\end{aligned}$$

where the second to last step follows from Stein's inequality (2.11), and the last step follows from Burkholder's inequality (2.4).

The last thing to do is to bound the second term in (4.10). Let $\mathscr{F}$ be the collection of principal cubes for the function f, constructed by beginning from the cube Q_0.

Note that $\Delta_Q(1_Q\sigma) = \Delta_Q(1_R\sigma)$ for every cube $\mathscr{D} \ni R \supset Q$. Using the principal cubes we estimate

$$\begin{aligned}
&\Big\| \Big(\sum_{Q\in\mathscr{D}} \big(b_Q \langle f\rangle_Q^\sigma \Delta_Q(1_Q\sigma)\big)^2\Big)^{1/2}\Big\|_{L^q(w)} \\
&\qquad\lesssim \Big\| \Big(\sum_{F\in\mathscr{F}} (\langle |f|\rangle_F^\sigma)^2 \sum_{\substack{Q\in\mathscr{D}\\ \pi_{\mathscr{F}}Q=F}} (b_Q\Delta_Q(1_F\sigma))^2\Big)^{1/2}\Big\|_{L^q(w)} \\
&\qquad\le \Big\| \Big(\sum_{F\in\mathscr{F}} (\langle |f|\rangle_F^\sigma)^2 S_{b,F}^\sigma(1_F)^2\Big)^{1/2}\Big\|_{L^q(w)} \\
&\qquad\le \mathfrak{S}_{\mathrm{loc}} \Big\| \Big(\sum_{F\in\mathscr{F}} (\langle |f|\rangle_F^\sigma)^2 1_F\Big)^{1/2}\Big\|_{L^p(\sigma)} \lesssim \mathfrak{S}_{\mathrm{loc}} \|f\|_{L^p(\sigma)},
\end{aligned}$$

where the last step follows from Carleson's embedding theorem (2.9).

Note that we have actually applied the quadratic testing condition only with a collection that is sparse with respect to the measure σ. This concludes the proof of Theorem 4.1. ∎

5. Dyadic shifts. Now we begin to consider dyadic shifts. First we give some basic definitions, and then we move on to characterize the two-weight inequality.

For any interval $I \subset \mathbb{R}$ write $h_I^0 := |I|^{-1/2}1_I$ and $h_I^1 := |I|^{-1/2}(1_{I_l} - 1_{I_r})$, where $|I|$ is the length of I, and I_l and I_r are the left and right halves of I. The function h_I^0 is called the *non-cancellative Haar function* and h_I^1 the *cancellative Haar function* related to the interval I.

For a cube $Q = I_1 \times \cdots \times I_N \in \mathscr{D}$, where each I_i is an interval in $\mathbb{R}$, define for $\eta \in \{0,1\}^N$ the Haar function related to the cube by

$$h_Q^\eta(x_1,\ldots,x_N) := \prod_{i=1}^N h_{I_i}^{\eta_i}(x_i).$$

If some η_i is non-zero, then h_Q^η is called *cancellative* since $\int h_Q^\eta\,dx = 0$; otherwise it is called *non-cancellative*. In any case $\int |h_Q^\eta|^2\,dx = 1$.

Fix two non-negative integers m and n. For every cube $K \in \mathscr{D}$ suppose we have a linear operator A_K^σ defined on locally σ-integrable functions by

$$A_K^\sigma f := \sum_{\substack{I,J\in\mathscr{D}\\ I^{(m)}=J^{(n)}=K}} a_{IJK}\langle f, h_I^J\rangle_\sigma h_J^I, \tag{5.1}$$

where h_I^J is a Haar function related to the cube (not interval) $I \in \mathscr{D}$ and h_J^I is a Haar function related to the cube $J \in \mathscr{D}$. The coefficients $a_{IJK} \in \mathbb{R}$ satisfy $|a_{IJK}| \le \sqrt{|I|\,|J|}/|K|$. Here the Haar functions are just some Haar functions, not any specific ones, and hence we do not specify them with the superscript η. Similarly define the corresponding dual operator

$$A_K^w g := \sum_{\substack{I,J\in\mathscr{D}\\ I^{(m)}=J^{(n)}=K}} a_{IJK}\langle g, h_J^I\rangle_w h_I^J$$

for locally w-integrable functions, where it should be noted that the functions h_I^J and h_J^I are in "opposite" places.

As a direct consequence of the size assumption on the coefficients we get, for any $f \in L^1_{\mathrm{loc}}(\sigma)$,

$$|A_K^\sigma f| \le \frac{1}{|K|}\int_K |f|\,d\sigma 1_K, \tag{5.2}$$

and a similar estimate holds for A_K^w.

We assume that there are only finitely many $K \in \mathscr{D}$ such that the coefficients a_{IJK} are non-zero. We make this assumption to have the dyadic shift well defined in the general two-weight setting, but all the bounds below will be independent of this number.

With the operators A_K^σ, the dyadic shift T^σ is defined by

$$T^\sigma f := \sum_{K\in\mathscr{D}} A_K^\sigma f, \qquad f \in L^1_{\mathrm{loc}}(\sigma), \tag{5.3}$$

and the shift T^w is defined analogously with the operators A_K^w. They are formal adjoints of each other in the sense that

$$\langle T^\sigma f, g\rangle_w = \langle f, T^w g\rangle_\sigma$$

for all $f \in L^1_{\mathrm{loc}}(\sigma)$ and $g \in L^1_{\mathrm{loc}}(w)$. The shift T^σ is said to have *parameters* (m,n), and correspondingly the shift T^w has parameters (n,m). The number $\max\{m,n\}$ is the *complexity* of the shift.

Instead of a single dyadic shift we are going to consider a family $\mathscr{T}$ of dyadic shifts with at most a given complexity. Let us first recall the definition of $\mathcal{R}$-bounded operator families as used for example in [18]. Suppose $(\varepsilon_k)_{k=1}^\infty$ is a sequence of independent random signs. If X and Y are two Banach spaces and $\mathscr{T}$ is a family of linear operators from X into Y, then $\mathscr{T}$ is said to be $\mathcal{R}$-*bounded* if there exists a constant C such that for all $U \in \{1,2,\dots\}$,

$(T_u)_{u=1}^U \subset \mathscr{T}$ and $(x_u)_{u=1}^U \subset X$,

$$\mathbb{E}\Big\| \sum_{u=1}^U \varepsilon_u T_u x_u \Big\|_Y \le C\mathbb{E}\Big\| \sum_{u=1}^U \varepsilon_u x_u \Big\|_X. \tag{5.4}$$

We denote the smallest possible constant C in (5.4) by $\mathcal{R}(\mathscr{T})$.

If $X = L^p(\sigma)$ and $Y = L^q(w)$ for some $1 \le p, q < \infty$, then similar computations with the Kahane–Khinchin inequality as above with the dyadic square function shows that in this case $\mathcal{R}$-boundedness can be equivalently defined as

$$\Big\| \Big(\sum_{u=1}^U (T_u f_u)^2 \Big)^{1/2} \Big\|_{L^q(w)} \lesssim \mathcal{R}(\mathscr{T}) \Big\| \Big(\sum_{u=1}^U f_u^2 \Big)^{1/2} \Big\|_{L^p(\sigma)}, \tag{5.5}$$

where $\mathcal{R}(\mathscr{T})$ is the constant as in (5.4). If $p = q = 2$, it is easily seen from (5.5) that in this case $\mathcal{R}$-boundedness is equivalent to uniform boundedness. On the other hand, from (5.4) one sees that if $\mathscr{T}$ consists of a single operator T, then $\mathcal{R}$-boundedness means just the boundedness of T.

Let $\mathscr{T} = \{T_\alpha^\sigma : \alpha \in \mathscr{A}\}$ be a collection of dyadic shifts. If $T_\alpha^\sigma \in \mathscr{T}$, then we write T_α^w for the corresponding formal adjoint. We say that the collection $\mathscr{T}$ of dyadic shifts satisfies the (*local*) *quadratic testing condition* (with respect to exponents $1 < p, q < \infty$) if for every $U \in \{1, 2, \dots\}$ and all sequences $(a_u)_{u=1}^U \subset \mathbb{R}$, $(T_u^\sigma)_{u=1}^U \subset \mathscr{T}$ and $(Q_u)_{u=1}^U \subset \mathscr{D}$ we have

$$\Big\| \Big(\sum_{u=1}^U (a_u 1_{Q_u} T_u^\sigma 1_{Q_u})^2 \Big)^{1/2} \Big\|_{L^q(w)} \le \mathcal{T}^\sigma \Big\| \Big(\sum_{u=1}^U a_u^2 1_{Q_u} \Big)^{1/2} \Big\|_{L^p(\sigma)}, \tag{5.6}$$

$$\Big\| \Big(\sum_{u=1}^U (a_u 1_{Q_u} T_u^w 1_{Q_u})^2 \Big)^{1/2} \Big\|_{L^{p'}(\sigma)} \le \mathcal{T}^w \Big\| \Big(\sum_{u=1}^U a_u^2 1_{Q_u} \Big)^{1/2} \Big\|_{L^{q'}(w)}, \tag{5.7}$$

where $\mathcal{T}^\sigma, \mathcal{T}^w < \infty$ are the best possible constants. Note that it is not forbidden that $T_u = T_{u'}$ for some $u \ne u'$. In particular, if $\mathscr{T}$ consists only of a single shift, then we get the corresponding quadratic testing condition for the dyadic square function as above.

The two-weight theorem for dyadic shifts is as follows:

THEOREM 5.1. *Let $1 < p, q < \infty$ and assume that the measures σ and w satisfy the quadratic $\mathscr{A}_{p,q}$-condition. Suppose $\mathscr{T}$ is a collection of dyadic shifts as in* (5.3) *with complexities at most κ. Then the collection $\mathscr{T}$ is $\mathcal{R}$-bounded from $L^p(\sigma)$ into $L^q(w)$ if and only if it satisfies the quadratic testing conditions* (5.6) *and* (5.7)*, and in this case*

$$\mathcal{R}(\mathscr{T}) \lesssim (1+\kappa)(\mathcal{T}^\sigma + \mathcal{T}^w) + (1+\kappa)^2 [\sigma, w]_{p,q}. \tag{5.8}$$

Again before proving the theorem we comment quickly on the case $1 < p \le 2 \le q < \infty$. Similar computations to (3.6) show that in this case

$\mathcal{R}$-boundedness is equivalent to uniform boundedness, the quadratic testing condition reduces to Sawyer type testing, and the quadratic $\mathscr{A}_{p,q}$-condition becomes the simple $A_{p,q}$-condition. Thus we find that a dyadic shift T^σ is bounded from $L^p(\sigma)$ into $L^q(w)$ if and only if the Sawyer type conditions

$$\|1_Q T^\sigma 1_Q\|_{L^q(w)} \leq \mathcal{T}^\sigma \sigma(Q)^{1/p}$$

and

$$\|1_Q T^w 1_Q\|_{L^{p'}(\sigma)} \leq \mathcal{T}^w w(Q)^{1/q'}$$

hold for all $Q \in \mathscr{D}$, and the measures satisfy the Muckenhoupt type $A_{p,q}$-condition

$$(\sigma, w)_{p,q} := \sup_{Q\in\mathscr{D}} \frac{\sigma(Q)^{1/p'} w(Q)^{1/q}}{|Q|} < \infty.$$

In this case

$$\|T^\sigma\|_{L^p(\sigma)\to L^q(w)} \lesssim (1+\kappa)(\mathcal{T}^\sigma + \mathcal{T}^w) + (1+\kappa)^2 (\sigma, w)_{p,q},$$

which is the result proved in [7] when $p = q = 2$.

Proof of Theorem 5.1. Suppose $\mathscr{T}$ is $\mathcal{R}$-bounded, whence clearly the quadratic testing condition (5.6) is satisfied. Using duality one sees that the collection of formal adjoints of the shifts in $\mathscr{T}$ is $\mathcal{R}$-bounded from $L^{q'}(w)$ into $L^{p'}(\sigma)$, and thus also (5.7) is satisfied. Hence it is enough to show the sufficiency of the testing conditions.

So we assume that we have a collection $\mathscr{T}$ of dyadic shifts with complexity at most κ satisfying the quadratic testing conditions (5.6) and (5.7). For any $U = 1, 2, \ldots$ suppose we have some sequences $(T_u^\sigma)_{u=1}^U \subset \mathscr{T}$ and $(f_u)_{u=1}^U \subset L^p(\sigma)$. To prove (5.8) it is enough to take an arbitrary sequence $(g_u)_{u=1}^U \subset L^{q'}(w)$ and show that

$$\Big|\sum_{u=1}^U \langle T_u^\sigma f_u, g_u\rangle_w\Big|$$
$$\lesssim \big((1+\kappa)(\mathcal{T}^\sigma + \mathcal{T}^w) + (1+\kappa)^2(\sigma, w)_{p,q}\big) \|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)} \|(g_u)_{u=1}^U\|_{L^{q'}(w;l^2)}.$$

For every u we write the corresponding shift as

$$T_u^\sigma f_u = \sum_{K\in\mathscr{D}} A_{u,K}^\sigma f_u = \sum_{K\in\mathscr{D}} \sum_{\substack{I,J\in\mathscr{D}\\ I^{(m)}=J^{(n)}=K}} a_{IJK}^u \langle f_u, h_{I,u}^J\rangle_\sigma h_{J,u}^I.$$

Let again $\bigcup_{k=1}^\infty Q_k^i$, $i = 1, \ldots, j \leq 2^N$, be the different "quadrants" of our dyadic system, as explained around (4.9). Because we have assumed that every shift consists of only finitely many operators A_K^σ, we can choose for every i a cube $Q_{k_i}^i := \tilde{Q}_i$ such that $a_{IJK}^u \neq 0$ implies $K \subset \bigcup_{i=1}^j \tilde{Q}_i$. Since the definition of the shift shows that $T_u^\sigma(f_u 1_{\tilde{Q}_i})$ is supported on $1_{\tilde{Q}_i}$, we have

$$\sum_{u=1}^{U}\langle T_u^\sigma f_u, g_u\rangle_w = \sum_{i=1}^{j}\sum_{u=1}^{U}\langle T_u^\sigma 1_{\tilde{Q}_i} f_u, 1_{\tilde{Q}_i} g_u\rangle_w,$$

and it is enough to estimate for each i separately.

Finally, we split

$$\begin{aligned}(5.9)\qquad \langle T_u^\sigma 1_{\tilde{Q}_i} f_u, 1_{\tilde{Q}_i} g_u\rangle_w &= \langle T_u^\sigma(1_{\tilde{Q}_i}(f_u - \langle f_u\rangle_{\tilde{Q}_i}^\sigma)), 1_{\tilde{Q}_i}(g_u - \langle g_u\rangle_{\tilde{Q}_i}^w)\rangle_w \\ &\quad + \big\langle 1_{\tilde{Q}_i}(f_u - \langle f_u\rangle_{\tilde{Q}_i}^\sigma), \langle g_u\rangle_{\tilde{Q}_i}^w T_u^w 1_{\tilde{Q}_i}\big\rangle_\sigma + \big\langle \langle f_u\rangle_{\tilde{Q}_i}^\sigma T^\sigma 1_{\tilde{Q}_i}, 1_{\tilde{Q}_i} g_u\big\rangle_w,\end{aligned}$$

and the sum over u of the last two terms can be bounded directly by using the testing conditions. For example

$$\begin{aligned}&\Big|\sum_{u=1}^{U}\langle f_u\rangle_{\tilde{Q}_i}^\sigma \langle T_u^\sigma 1_{\tilde{Q}_i}, 1_{\tilde{Q}_i} g_u\rangle_w\Big| \\ &\qquad \le \Big\|\Big(\sum_{u=1}^{U}(\langle f_u\rangle_{\tilde{Q}_i}^\sigma 1_{\tilde{Q}_i} T_u^\sigma 1_{\tilde{Q}_i})^2\Big)^{1/2}\Big\|_{L^q(w)}\Big\|\Big(\sum_{u=1}^{U}|1_{\tilde{Q}_i} g_u|^2\Big)^{1/2}\Big\|_{L^{q'}(w)} \\ &\qquad \le \mathcal{T}^\sigma\Big(\sum_{u=1}^{U}(\langle f_u\rangle_{\tilde{Q}_i}^\sigma)^2\Big)^{1/2}\sigma(\tilde{Q}_i)^{1/p}\|(1_{\tilde{Q}_i} g_u)_{u=1}^U\|_{L^{q'}(w;l^2)},\end{aligned}$$

and using the fact that an l^2-sum of averages is less than the average of the l^2-sum we get

$$\begin{aligned}&\Big(\sum_{u=1}^{U}(\langle f_u\rangle_{\tilde{Q}_i}^\sigma)^2\Big)^{1/2}\sigma(\tilde{Q}_i)^{1/p} \\ &\qquad\qquad \le \Big\langle\Big(\sum_{u=1}^{U} f_u^2\Big)^{1/2}\Big\rangle_{\tilde{Q}_i}^\sigma \sigma(\tilde{Q}_i)^{1/p} \le \|(1_{\tilde{Q}_i} f_u)_{u=1}^U\|_{L^p(\sigma;l^2)}.\end{aligned}$$

After these reductions it is enough to fix one cube $Q_{k_i}^i =: Q_0$ and suppose that for every u the functions f_u and g_u are supported on Q_0 and have zero averages. Since the shifts T_u^σ are *a priori* bounded, by L^p-convergence of martingale differences we can assume that

$$f_u = \sum_{\substack{Q\in\mathscr{D}\\ Q\subset Q_0}} \Delta_Q^\sigma f_u, \qquad g_u = \sum_{\substack{Q\in\mathscr{D}\\ Q\subset Q_0}} \Delta_Q^w g_u,$$

where the sums are finite.

Using the martingale decomposition

$$(5.10)\qquad \sum_{u=1}^{U}\langle T_u^\sigma f_u, g_u\rangle_w = \sum_{u=1}^{U}\sum_{Q,R\in\mathscr{D}}\langle T_u^\sigma \Delta_Q^\sigma f_u, \Delta_R^w g_u\rangle_w,$$

we split the proof into parts depending on the relative positions of the cubes Q and R; this part of the proof follows the outlines in [4]. The cases "$l(Q) \le l(R)$" and "$l(Q) > l(R)$" are treated symmetrically, and here we concentrate on the first. Using the maximal possible complexity κ of the shifts, we further split into three cases "$Q \cap R = \emptyset$", "$Q^{(\kappa)} \subsetneq R$" and "$Q \subset R \subset Q^{(\kappa)}$", and these are treated separately using different properties of the shifts.

In the summations we understand that we are summing over dyadic cubes, and we will not always write "$Q \in \mathscr{D}$" in the summation condition. Moreover, since we have assumed the finite martingale decompositions of f and g, we can think that every $Q \in \mathscr{D}$ that appears below actually belongs to some sufficiently big *finite* collection $\mathscr{D}_0 \subset \mathscr{D}$. This way all the sums are actually finite, and one does not have to worry about any convergence issues.

At this point it is convenient to introduce the notation

$$\Delta_Q^{\sigma,i} f := \sum_{\substack{Q' \in \mathscr{D} \\ Q'^{(i)} = Q}} \Delta_{Q'}^{\sigma} f$$

for any $f \in L^1_{\mathrm{loc}}(\sigma)$, $Q \in \mathscr{D}$ and $i \in \{0, 1, 2, \dots\}$, and similarly for the measure w.

Disjoint cubes: $Q \cap R = \emptyset$ and $l(Q) \le l(R)$. Here we bound the part

$$\Big| \sum_{u=1}^{U} \sum_{\substack{l(Q) \le l(R) \\ Q \cap R = \emptyset}} \langle T_u^{\sigma} \Delta_Q^{\sigma} f_u, \Delta_R^w g_u \rangle_w \Big|. \tag{5.11}$$

Consider a fixed u first, and suppose the shift T_u^{σ} has parameters (m, n) with $m + n \le \kappa$. Fix two cubes $Q, R \in \mathscr{D}$ with $Q \cap R = \emptyset$ and suppose $K \in \mathscr{D}$ is such that $\langle A_{u,K}^{\sigma} \Delta_Q^{\sigma} f_u, \Delta_R^w g_u \rangle_w \ne 0$. We must have $Q \cap K \ne \emptyset \ne R \cap K$, which combined with $Q \cap R = \emptyset$ implies that $Q, R \subset K$. Also, since the functions $\Delta_Q^{\sigma} f$ and $\Delta_R^w g$ have zero σ- and w-averages, respectively, and a Haar function h_I is constant on the children of I, we have $K \subset Q^{(m)}$ and $K \subset R^{(n)}$. Thus the sum (5.11) is actually zero if $m = 0$ or $n = 0$. Hence we assume $m, n \ge 1$, rearrange the sum in question and estimate with (5.2) as

$$\begin{aligned} &\sum_{\substack{l(Q) \le l(R) \\ Q \cap R = \emptyset}} |\langle T_u^{\sigma} \Delta_Q^{\sigma} f_u, \Delta_R^w g_u \rangle_w| \\ &\qquad \le \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{K \in \mathscr{D}} \sum_{\substack{Q, R \in \mathscr{D} \\ Q^{(i)} = R^{(j)} = K}} |\langle A_{u,K}^{\sigma} \Delta_Q^{\sigma} f_u, \Delta_R^w g_u \rangle_w| \end{aligned} \tag{5.12}$$

$$\begin{aligned}&\leq \sum_{i,j=1}^{\kappa}\sum_{K\in\mathscr{D}}\sum_{\substack{Q,R\in\mathscr{D}\\Q^{(i)}=R^{(j)}=K}}\frac{\|\Delta_Q^\sigma f_u\|_{L^1(\sigma)}\|\Delta_R^w g_u\|_{L^1(w)}}{|K|}\\&=\sum_{i,j=1}^{\kappa}\sum_{K\in\mathscr{D}}\frac{\|\Delta_K^{\sigma,i}f_u\|_{L^1(\sigma)}\|\Delta_K^{w,j}g_u\|_{L^1(w)}}{|K|}.\end{aligned}$$

Note that this estimate does not depend on the parameters (m,n) of the shift.

Then for any fixed i and j, we sum over u, and continue with

$$\begin{aligned}(5.13)\qquad &\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}\frac{\|\Delta_K^{\sigma,i}f_u\|_{L^1(\sigma)}\|\Delta_K^{w,j}g_u\|_{L^1(w)}}{|K|}\\&=\int\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}\frac{\|\Delta_K^{\sigma,i}f_u\|_{L^1(\sigma)}}{|K|}|\Delta_K^{w,j}g_u|\,dw\\&\leq\Big\|\Big(\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}\Big(\frac{\|\Delta_K^{\sigma,i}f_u\|_{L^1(\sigma)}}{|K|}\Big)^2 1_K\Big)^{1/2}\Big\|_{L^q(w)}\\&\quad\cdot\Big\|\Big(\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}(\Delta_K^{w,j}g_u)^2 1_K\Big)^{1/2}\Big\|_{L^{q'}(w)}=:A\cdot B.\end{aligned}$$

Using the quadratic $\mathscr{A}_{p,q}$-condition we get

$$\begin{aligned}(5.14)\qquad A&=\Big\|\Big(\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}\Big(\langle|\Delta_K^{\sigma,i}f_u|\rangle_K^\sigma\frac{\sigma(K)}{|K|}\Big)^2 1_K\Big)^{1/2}\Big\|_{L^q(w)}\\&\leq[\sigma,w]_{p,q}\Big\|\Big(\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}(\langle|\Delta_K^{\sigma,i}f_u|\rangle_K^\sigma)^2 1_K\Big)^{1/2}\Big\|_{L^p(\sigma)}\\&\leq[\sigma,w]_{p,q}\Big\|\Big(\sum_{K\in\mathscr{D}}\Big(\Big\langle\Big(\sum_{u=1}^{U}(\Delta_K^{\sigma,i}f_u)^2\Big)^{1/2}\Big\rangle_K^\sigma\Big)^2 1_K\Big)^{1/2}\Big\|_{L^p(\sigma)}.\end{aligned}$$

Applying Stein's inequality (2.11) and then Burkholder's inequality (2.6) to the last term in (5.14) we obtain

$$\begin{aligned}\mathrm{RHS}(5.14)&\lesssim[\sigma,w]_{p,q}\Big\|\Big(\sum_{K\in\mathscr{D}}\sum_{u=1}^{U}(\Delta_K^{\sigma,i}f_u)^2 1_K\Big)^{1/2}\Big\|_{L^p(\sigma)}\\&\lesssim[\sigma,w]_{p,q}\|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)}.\end{aligned}$$

The factor B in (5.13) is estimated directly using Burkholder's inequality, and then it only remains to sum over the finite ranges of i and j, which produces a factor κ^2 in the final estimate. Hence we have shown that

$$(5.11) \lesssim \kappa^2 \cdot [\sigma, w]_{p,q} \|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)} \|(g_u)_{u=1}^U\|_{L^{q'}(w;l^2)}.$$

Deeply contained cubes: $Q^{(\kappa)} \subsetneq R$. We again consider a fixed T_u^σ with parameters (m,n) first. Assume $Q, R \in \mathscr{D}$ are such that $Q^{(\kappa)} \subsetneq R$. If $A_{u,K}^\sigma \Delta_Q^\sigma f_u$ is non-zero, we must have $K \subset Q^{(m)} \subset Q^{(\kappa)} \subsetneq R$. Since $A_{u,K}^\sigma \Delta_Q^\sigma f_u$ is supported on K and $\Delta_R^w g$ is constant on the children of R, we see that

$$\langle A_{u,K}^\sigma \Delta_Q^\sigma f_u, \Delta_R^w g_u \rangle_w = \big\langle A_{u,K}^\sigma \Delta_Q^\sigma f_u, \langle \Delta_R^w g_u \rangle_{Q^{(\kappa)}}^w 1_{Q^{(\kappa)}} \big\rangle_w,$$

and thus

$$\langle T_u^\sigma \Delta_Q^\sigma f_u, \Delta_R^w g_u \rangle_w = \big\langle T_u^\sigma \Delta_Q^\sigma f_u, \langle \Delta_R^w g_u \rangle_{Q^{(\kappa)}}^w 1_{Q^{(\kappa)}} \big\rangle_w.$$

Taking "$Q^{(\kappa)}$" as a new summation variable we can rewrite the sum to be estimated as

$$(5.15) \qquad \sum_{\substack{Q,R\in\mathscr{D} \\ Q^{(\kappa)}\subsetneq R}} \langle T_u^\sigma \Delta_Q^\sigma f_u, \Delta_R^w g_u \rangle_w = \sum_{Q\in\mathscr{D}} \sum_{\substack{R\in\mathscr{D} \\ R\supsetneq Q}} \sum_{\substack{Q'\in\mathscr{D} \\ Q'^{(\kappa)}=Q}} \big\langle T_u^\sigma \Delta_{Q'}^\sigma f_u, \langle \Delta_R^w g_u \rangle_Q^w 1_Q \big\rangle_w = \sum_{Q\in\mathscr{D}} \big\langle \Delta_Q^{\sigma,\kappa} f_u, \langle g_u \rangle_Q^w \Delta_Q^{\sigma,\kappa} T_u^w 1_Q \big\rangle_\sigma,$$

where we collapsed the sum $\sum_{R\in\mathscr{D},\, R\supsetneq Q} \langle \Delta_R^w g_u \rangle_Q^w 1_Q = \langle g_u \rangle_Q^w 1_Q$, and used the fact that the martingale difference operator $\Delta_Q^{\sigma,\kappa}$ can be put also on the other side of the pairing $\langle \cdot, \cdot \rangle_\sigma$. Now we have again an equation that is independent of the parameters (m,n), so it holds for all the shifts T_u^σ.

Then we sum over u and estimate

$$(5.16) \qquad \Big| \sum_{u=1}^U \sum_{Q\in\mathscr{D}} \big\langle \Delta_Q^{\sigma,\kappa} f_u, \langle g_u \rangle_Q^w \Delta_Q^{\sigma,\kappa} T_u^w 1_Q \big\rangle_\sigma \Big| = \Big| \int \sum_{u=1}^U \sum_{Q\in\mathscr{D}} \Delta_Q^{\sigma,\kappa} f_u \langle g_u \rangle_Q^w \Delta_Q^{\sigma,\kappa} T_u^w 1_Q \, d\sigma \Big| \le \Big\| \Big(\sum_{u=1}^U \sum_{Q\in\mathscr{D}} (\Delta_Q^{\sigma,\kappa} f_u)^2 \Big)^{1/2} \Big\|_{L^p(\sigma)} \cdot \Big\| \Big(\sum_{u=1}^U \sum_{Q\in\mathscr{D}} (\langle g_u \rangle_Q^w \Delta_Q^{\sigma,\kappa} T_u^w 1_Q)^2 \Big)^{1/2} \Big\|_{L^{p'}(\sigma)},$$

where Burkholder's inequality (2.6) implies that the first factor on the right hand side is dominated by $\|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)}$.

In the second factor we note that if φ is any locally w-integrable function, then $\Delta_Q^{\sigma,\kappa} A_{u,K}^w(1_{\complement Q}\varphi) = 0$ for any $Q, K \in \mathscr{D}$, which follows from the fact that the shift has complexity at most κ. This shows that

$$\Delta_Q^{\sigma,\kappa} T_u^w 1_Q = \Delta_Q^{\sigma,\kappa} T_u^w 1_P \tag{5.17}$$

for any $\mathscr{D} \ni P \supset Q$.

Beginning from the cube Q_0, construct the sets $\mathscr{G}_u$ of principal cubes for the functions g_u with respect to the measure w. Since the functions g_u have finite martingale difference decompositions, and are accordingly constant on sufficiently small cubes $Q \in \mathscr{D}$, the collections $\mathscr{G}_u$ are finite.

Using the remark (5.17) we proceed with

$$\begin{aligned}
&\Big\| \Big(\sum_{u=1}^{U} \sum_{Q \in \mathscr{D}} (\langle g_u \rangle_Q^w \Delta_Q^{\sigma,\kappa} T_u^w 1_Q)^2 \Big)^{1/2} \Big\|_{L^{p'}(\sigma)} \\
&\qquad \lesssim \Big\| \Big(\sum_{u=1}^{U} \sum_{G \in \mathscr{G}_u} (\langle |g_u| \rangle_G^w)^2 \sum_{\substack{Q \in \mathscr{D} \\ \pi_{\mathscr{G}_u} Q = G}} (\Delta_Q^{\sigma,\kappa} T_u^w 1_G)^2 \Big)^{1/2} \Big\|_{L^{p'}(\sigma)} \\
&\qquad \lesssim \Big\| \Big(\sum_{u=1}^{U} \sum_{G \in \mathscr{G}_u} (\langle |g_u| \rangle_G^w 1_G T_u^w 1_G)^2 \Big)^{1/2} \Big\|_{L^{p'}(\sigma)} \\
&\qquad \le \mathcal{T}^w \Big\| \Big(\sum_{u=1}^{U} \sum_{G \in \mathscr{G}_u} (\langle |g_u| \rangle_G^w 1_G)^2 \Big)^{1/2} \Big\|_{L^{q'}(w)} \lesssim \mathcal{T}^w \| (g_u)_{u=1}^U \|_{L^{q'}(w;l^2)},
\end{aligned}$$

where we have used Burkholder's inequality (2.6) in the second step and Carleson's embedding theorem (2.9) in the last step. This concludes the proof for the part "$Q^{(\kappa)} \subsetneq R$".

Contained cubes of comparable size: $Q \subset R \subset Q^{(\kappa)}$. For a fixed u, the sum to be estimated in this last subsection can be written as

$$\begin{aligned}
&\sum_{i=0}^{\kappa} \sum_{R \in \mathscr{D}} \sum_{\substack{Q \in \mathscr{D} \\ Q^{(i)} = R}} \langle T_u^\sigma \Delta_Q^\sigma f_u, \Delta_R^w g_u \rangle_w \\
&\qquad = \sum_{i=0}^{\kappa} \sum_{k=1}^{2^N} \sum_{R \in \mathscr{D}} \big\langle \Delta_R^{\sigma,i} f_u, \langle \Delta_R^w g_u \rangle_{R_k}^w T_u^w 1_{R_k} \big\rangle_\sigma \\
&\qquad = \sum_{i=0}^{\kappa} \sum_{k=1}^{2^N} \sum_{R \in \mathscr{D}} \big\langle 1_{R_k} \Delta_R^{\sigma,i} f_u, \langle \Delta_R^w g_u \rangle_{R_k}^w T_u^w 1_{R_k} \big\rangle_\sigma \\
&\qquad\quad + \sum_{i=0}^{\kappa} \sum_{k=1}^{2^N} \sum_{R \in \mathscr{D}} \big\langle 1_{\complement R_k} \Delta_R^{\sigma,i} f_u, \langle \Delta_R^w g_u \rangle_{R_k}^w T_u^w 1_{R_k} \big\rangle_\sigma,
\end{aligned} \tag{5.18}$$

where the cubes R_k are the dyadic children of R.

Consider the first sum on the right side of (5.18). We fix some i and k, sum over u and use testing to deduce that

$$\begin{aligned}\Big|\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}\langle 1_{R_k}\Delta_R^{\sigma,i}f_u,\langle\Delta_R^w g_u\rangle_{R_k}^w T_u^w 1_{R_k}\rangle_\sigma\Big| &\leq \Big\|\Big(\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}(1_{R_k}\Delta_R^{\sigma,i}f_u)^2\Big)^{1/2}\Big\|_{L^p(\sigma)} \\ &\quad\cdot\Big\|\Big(\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}(\langle\Delta_R^w g_u\rangle_{R_k}^w 1_{R_k}T_u^w 1_{R_k})^2\Big)^{1/2}\Big\|_{L^{p'}(\sigma)} \\ &\lesssim \mathcal{T}^w\|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)}\Big\|\Big(\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}|\langle\Delta_R^w g\rangle_{R_k}^w 1_{R_k}|^2\Big)^{1/2}\Big\|_{L^{q'}(w)} \\ &\lesssim \mathcal{T}^w\|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)}\|(g_u)_{u=1}^U\|_{L^{q'}(w;l^2)}.\end{aligned}$$

Now turn to the other sum in (5.18) to be estimated. With the same notation as there, we have $1_{\complement R_k}A_{u,K}^w 1_{R_k}\neq 0$ only if $K\supset R$. Hence, using (5.2), we get

$$\begin{aligned}\big|\big\langle 1_{\complement R_k}\Delta_R^{\sigma,i}f_u,\langle\Delta_R^w g_u\rangle_{R_k}^w T_u^w 1_{R_k}\big\rangle_\sigma\big| &\leq \sum_{\substack{K\in\mathscr{D}\\ K\supset R}}\frac{\|1_{\complement R_k}\Delta_R^{\sigma,i}f_u\|_{L^1(\sigma)}\|1_{R_k}\Delta_R^w g_u\|_{L^1(w)}}{|K|} \\ &\simeq \frac{\|1_{\complement R_k}\Delta_R^{\sigma,i}f_u\|_{L^1(\sigma)}\|1_{R_k}\Delta_R^w g_u\|_{L^1(w)}}{|R|}.\end{aligned}$$

Summing this over k, and then over $R\in\mathscr{D}$ and $u\in\{1,\dots,U\}$, leads, as in (5.13) and (5.14), to

$$\begin{aligned}\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}\sum_{k=1}^{2^N}\frac{\|1_{\complement R_k}\Delta_R^{\sigma,i}f_u\|_{L^1(\sigma)}\|1_{R_k}\Delta_R^w g_u\|_{L^1(w)}}{|R|} &\leq \Big\|\Big(\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}\Big(\frac{\|\Delta_R^{\sigma,i}f_u\|_{L^1(\sigma)}}{|R|}\Big)^2 1_R\Big)^{1/2}\Big\|_{L^q(w)} \\ &\quad\cdot\Big\|\Big(\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}(\Delta_R^w g_u)^2\Big)^{1/2}\Big\|_{L^{q'}(w)} \\ &\lesssim [\sigma,w]_{p,q}\|(f_u)_{u=1}^U\|_{L^p(\sigma;l^2)}\|(g_u)_{u=1}^U\|_{L^{q'}(w;l^2)}.\end{aligned}\tag{5.19}$$

Summing over $i\in\{0,\dots,\kappa\}$ produces the factor $1+\kappa$ in the final estimate.

This finishes the proof of the case "$Q \subset R \subset Q^{(\kappa)}$", and hence also of Theorem 5.1. ∎

Lemma 5.2. *Let $1 < p, q < \infty$ and suppose $\mathscr{T}$ is a family of dyadic shifts containing all shifts with parameters (m, n). If $\mathscr{T}$ is $\mathcal{R}$-bounded from $L^p(\sigma)$ into $L^q(w)$, then*

$$[\sigma, w]_{p,q} \le 2^{N \min(m,n)} \mathcal{R}(\mathscr{T}).$$

Proof. Suppose for example that $m \le n$. The case $m > n$ is similar. For every $I \in \mathscr{D}$ define the shift

$$T_I^\sigma := \sum_{\substack{J \in \mathscr{D} \\ J^{(n-m)} = I}} \frac{\sqrt{|I|\,|J|}}{|I^{(m)}|} \langle \cdot, h_I \rangle_\sigma h_J,$$

where the functions h_I and h_J are some fixed Haar functions related to the cubes I and J. Define also $f_I := h_I \sqrt{|I|}$.

With these definitions we have $|T_I^\sigma f_I| = \frac{\sigma(I)}{2^{Nm}|I|} 1_I$, and clearly $|f_I| = 1_I$. Thus, if $\{a_I\}_{I \in \mathscr{D}}$ is any finitely non-zero set of real numbers, then

$$\begin{aligned} 2^{-Nm} \Big\| \Big(\sum_{I \in \mathscr{D}} \Big(a_I \frac{\sigma(I)}{|I|} 1_I \Big)^2 \Big)^{1/2} \Big\|_{L^q(w)} &= \Big\| \Big(\sum_{I \in \mathscr{D}} (a_I T_I^\sigma f_I)^2 \Big)^{1/2} \Big\|_{L^q(w)} \\ &\le \mathcal{R}(\mathscr{T}) \Big\| \Big(\sum_{I \in \mathscr{D}} (a_I f_I 1_I)^2 \Big)^{1/2} \Big\|_{L^p(\sigma)} = \mathcal{R}(\mathscr{T}) \Big\| \Big(\sum_{I \in \mathscr{D}} a_I^2 1_I \Big)^{1/2} \Big\|_{L^p(\sigma)}, \end{aligned}$$

which shows that $[\sigma, w]_{p,q} \le 2^{Nm} \mathcal{R}(\mathscr{T})$. ∎

Corollary 5.3. *Suppose $1 < p, q < \infty$. The family $\mathscr{T}$ of all shifts with parameters (m, n) is $\mathcal{R}$-bounded from $L^p(\sigma)$ into $L^q(w)$ if and only if the family satisfies the quadratic testing conditions* (5.6) *and* (5.7)*, and the quadratic $\mathscr{A}_{p,q}$-condition holds. Moreover, we have the quantitative estimate*

$$2^{-N \min(m,n)} [\sigma, w]_{p,q} + \mathcal{T}^\sigma + \mathcal{T}^w \lesssim \mathcal{R}(\mathscr{T}) \lesssim (1+\kappa)(\mathcal{T}^\sigma + \mathcal{T}^w) + (1+\kappa)^2 [\sigma, w]_{p,q},$$

where $\mathcal{T}^\sigma$ and $\mathcal{T}^w$ are the testing constants and $\kappa = \max\{m, n\}$.

Dyadic shifts of a specific form. We look at the case when all the operators A_K^σ in the definition of the dyadic shifts are of the form

$$A_K^\sigma f := \sum_{\substack{I, J \colon I^{(m)} = J^{(n)} = K \\ I \vee J = K}} a_{IJK} \langle f, h_I^J \rangle_\sigma h_J^I, \tag{5.20}$$

where $I \vee J$ denotes the smallest cube (if it exists) in $\mathscr{D}$ containing both I and J. Thus $I \vee J = K$ is equivalent to I and J being subcubes of different children of K. This kind of dyadic shifts arise naturally when representing general Calderón–Zygmund operators with dyadic shifts as in [5]. Note that in this case if A_K^σ is to be non-zero then $m, n \ge 1$.

In this situation a weaker form of the quadratic $\mathscr{A}_{p,q}$-condition is sufficient in Theorem 5.1. Namely, let again Q_k, $k \in \{1,\dots,2^N\}$, denote the dyadic children of a cube $Q \in \mathscr{D}$. We do not have any special ordering in mind, and in fact the ordering need not be the same for different cubes. Thus, if $Q, Q' \in \mathscr{D}$ and $Q \neq Q'$, then Q_k and Q'_k need not be in symmetrical places with respect to the parents Q and Q'. We say that the measures σ and w satisfy the *quadratic $\mathscr{A}^*_{p,q}$-condition* if for any $k, l \in \{1,\dots,2^N\}$, $k \neq l$, and any collection $\{a_Q\}_{Q\in\mathscr{D}}$ of real numbers we have

$$\Big\|\Big(\sum_{Q\in\mathscr{D}}\Big(a_Q\frac{\sigma(Q_k)}{|Q_k|}\Big)^2 1_{Q_l}\Big)^{1/2}\Big\|_{L^q(w)} \leq [\sigma,w]^*_{p,q}\Big\|\Big(\sum_{Q\in\mathscr{D}} a_Q^2 1_{Q_k}\Big)^{1/2}\Big\|_{L^p(\sigma)}, \tag{5.21}$$

where again $[\sigma,w]^*_{p,q}$ denotes the best possible constant. Similarly to the case of the quadratic $\mathscr{A}_{p,q}$-condition, we have $[\sigma,w]^*_{p,q} \simeq [w,\sigma]^*_{q',p'}$.

The two-weight inequality for the Hilbert transform was characterized by M. Lacey, E. Sawyer, C.-Y. Shen and I. Uriarte-Tuero [11] and M. Lacey [9] in the case when the measures σ and w do not have common point masses. This restriction was lifted by T. Hytönen [6], and a key new component was a similar kind of weakening to we have here of the *Poisson A_2 conditions* used in [11] and [9].

THEOREM 5.4. *Let $1 < p, q < \infty$ and assume that the measures σ and w satisfy the quadratic $\mathscr{A}^*_{p,q}$-condition. Suppose $\mathscr{T}$ is a collection of dyadic shifts with complexities at most κ, and suppose every shift in $\mathscr{T}$ is of the specific form* (5.20)*. Then the collection $\mathscr{T}$ is $\mathcal{R}$-bounded from $L^p(\sigma)$ into $L^q(w)$ if and only if it satisfies the quadratic testing conditions* (5.6) *and* (5.7)*, and in this case*

$$\mathcal{R}(\mathscr{T}) \lesssim (1+\kappa)(\mathcal{T}^\sigma + \mathcal{T}^w) + (1+\kappa)^2[\sigma,w]^*_{p,q}. \tag{5.22}$$

We outline the proof of Theorem 5.4, which is probably known to specialists.

All we need to do is to look at the previous proof, consider the places where the quadratic $\mathscr{A}_{p,q}$-condition was applied, and show that in this special case it is enough to assume the weaker condition. The quadratic $\mathscr{A}_{p,q}$-condition was applied in two places: first at the end of the subsection dealing with the case "$Q \cap R = \emptyset$", and then at the end of the case "$Q \subset R \subset Q^{(\kappa)}$".

Assume that $K \in \mathscr{D}$ and we have an operator A^σ_K of the form (5.20). Then for $f \in L^1_{\mathrm{loc}}(\sigma)$ and $g \in L^1_{\mathrm{loc}}(w)$ we have

$$(5.23)\qquad |\langle A_K^\sigma f, g\rangle_w| = \Big|\sum_{\substack{k,l\in\{1,\dots,2^N\}\\ k\neq l}} \sum_{\substack{I^{(m-1)}=K_k\\ J^{(n-1)}=K_l}} a_{IJK}\langle f, h_I^J\rangle_\sigma \langle g, h_J^I\rangle_w\Big| \le \sum_{\substack{k,l\in\{1,\dots,2^N\}\\ k\neq l}} \frac{\|1_{K_k} f\|_{L^1(\sigma)}\|1_{K_l} g\|_{L^1(w)}}{|K|}.$$

If we use (5.23) in (5.12), we end up with the term

$$\sum_{i,j=1}^{\kappa}\sum_{K\in\mathscr{D}}\sum_{k\neq l} \frac{\|1_{K_k}\Delta_K^{\sigma,i} f_u\|_{L^1(\sigma)}\|1_{K_l}\Delta_K^{w,j} g_u\|_{L^1(w)}}{|K|}.$$

If one continues as in (5.13) with fixed $k\neq l$, the result is

$$\Big\|\Big(\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}\big(\frac{\|1_{K_k}\Delta_K^{\sigma,i} f_u\|_{L^1(\sigma)}}{|K|}\big)^2 1_{K_l}\Big)^{1/2}\Big\|_{L^q(w)} \cdot \Big\|\Big(\sum_{u=1}^{U}\sum_{K\in\mathscr{D}}(1_{K_l}\Delta_K^{w,j} g_u)^2\Big)^{1/2}\Big\|_{L^{q'}(w)}.$$

The factor related to g is directly handled by using Burkholder's inequality, and the other related to f is estimated via the $\mathscr{A}^*_{p,q}$-condition as in (5.14). In the end one can sum over the finite ranges of k and l. This takes care of the first application of the $\mathscr{A}^*_{p,q}$-condition.

The other application is even easier, since there the functions are already in the right form. If we look at the first term in (5.19), we see that it can be written as

$$\sum_{u=1}^{U}\sum_{R\in\mathscr{D}}\sum_{k=1}^{2^N} \frac{\|1_{\complement R_k}\Delta_R^{\sigma,i} f_u\|_{L^1(\sigma)}\|1_{R_k}\Delta_R^{w} g_u\|_{L^1(w)}}{|R|} = \sum_{\substack{k,l\\ k\neq l}}\sum_{u=1}^{U}\sum_{R\in\mathscr{D}} \frac{\|1_{R_l}\Delta_R^{\sigma,i} f_u\|_{L^1(\sigma)}\|1_{R_k}\Delta_R^{w} g_u\|_{L^1(w)}}{|R|},$$

and for a fixed pair $k\neq l$ this can again be estimated by using the $\mathscr{A}^*_{p,q}$-condition.

6. Examples related to the quadratic $\mathscr{A}_{p,q}$-condition. Consider the one-weight case with $p=q\in(1,\infty)$, where we have an almost everywhere (in the Lebesgue sense) positive Borel measurable function $w:\mathbb{R}^N\to\mathbb{R}$. With the same symbol we also denote the Borel measure

$$w(E) := \int_E w\,dx,$$

where $E \subset \mathbb{R}^N$ is any Borel set. The dual weight to w is $\sigma := w^{-1/(p-1)}$, and we again use σ for the corresponding measure. The *Muckenhoupt A_p characteristic* is defined as

$$[w]_p := \sup_{Q\in\mathscr{D}} \frac{\sigma(Q)^{p-1}w(Q)}{|Q|^p},$$

and the *Muckenhoupt A_p class* consists of those weights that have $[w]_p < \infty$.

In this one-weight case the weighted Stein inequality (3.3) can be equivalently written as

$$\Big\|\Big(\sum_{Q\in\mathscr{D}}\Big(\frac{\int_Q f_Q\,dx}{|Q|}\Big)^2 1_Q\Big)^{1/2}\Big\|_{L^p(w)} \le \mathscr{S}\Big\|\Big(\sum_{Q\in\mathscr{D}} f_Q^2 1_Q\Big)^{1/2}\Big\|_{L^p(w)}. \tag{6.1}$$

It can quite easily be seen that if $p = 2$ then the constant $\mathscr{S}$ in the weighted Stein inequality is $[w]_2^{1/2}$, that is, the inequality (6.1) holds with a finite constant if and only if the weight is in the Muckenhoupt A_2 class. A quantitative form of the extrapolation theorem of Rubio de Francia [15] by O. Dragičević, L. Grafakos, M. Pereyra and S. Petermichl [2] then implies that the best constant $\mathscr{S}(w,p)$ in (6.1) satisfies

$$\mathscr{S}(w,p) \lesssim \begin{cases} [w]_p^{1/(2(p-1))}, & 1 < p \le 2,\\ [w]_p^{1/2}, & 2 \le p < \infty.\end{cases}$$

Since Lemma 3.2 shows that the quadratic $\mathscr{A}_{p,q}$-constant is equivalent to the best constant in the two-weight Stein inequality, we get the quantitative estimates

$$\begin{cases} [w]_p^{1/p} \le [\sigma,w]_{p,p} \lesssim [w]_p^{1/(2(p-1))}, & 1 < p \le 2,\\ [w]_p^{1/p} \le [\sigma,w]_{p,p} \lesssim [w]_p^{1/2}, & 2 \le p < \infty.\end{cases}$$

On the other hand, in the general two-weight setting the quadratic $\mathscr{A}_{p,q}$-condition is strictly stronger than the simple $A_{p,q}$-condition if $p > 2$ or $q < 2$:

Lemma 6.1. *Let $p,q \in (1,\infty)$.*

(a) *If $1 < p \le 2 \le q < \infty$, then $(\sigma,w)_{p,q} = [\sigma,w]_{p,q}$ for all Radon measures σ and w.*
(b) *If $2 < p < \infty$ or $1 < q < 2$, then there exist Radon measures σ and w such that $(\sigma,w)_{p,q} < \infty$ but $[\sigma,w]_{p,q} = \infty$.*

Proof. Case (a) is just Lemma 3.3, so we need to prove only the other assertion. Let $1 < p,q < \infty$ and choose a cube $Q_0 \in \mathscr{D}$ with $|Q_0| = 1$. Then we simply set the measure σ to be $1_{Q_0}dx$, that is, the Lebesgue measure restricted to Q_0.

The measure w that we next construct must satisfy

$$w(Q) \leq C \frac{|Q|^q}{\sigma(Q)^{q/p'}}, \qquad Q \in \mathscr{D},$$

for some constant C. Keeping this in mind we set

$$w := \sum_{k=1}^{\infty} |Q_0^{(k)}|^{q-1} 1_{Q_0^{(k)} \setminus Q_0^{(k-1)}} \, dx.$$

To see that the pair (σ, w) satisfies the simple $A_{p,q}$-condition, first note that since the measures are supported on Q_0 and $\complement Q_0$, respectively, then $\sigma(Q)w(Q) = 0$ for all cubes $Q \in \mathscr{D}$ with $l(Q) \leq 1$. Also if $Q \in \mathscr{D}$ is such that $l(Q) > 1$ and $\sigma(Q) \neq 0$, there exists an $l \in \{1, 2, \dots\}$ such that $Q = Q_0^{(l)}$. But then

$$w(Q_0^{(l)}) = \sum_{k=1}^{l} |Q_0^{(k)}|^{q-1} |Q_0^{(k)} \setminus Q_0^{(k-1)}| \simeq \sum_{k=1}^{l} |Q_0^{(k)}|^q \simeq |Q_0^{(l)}|^q,$$

and this shows that

$$\frac{\sigma(Q_0^{(l)})^{1/p'} w(Q_0^{(l)})^{1/q}}{|Q_0^{(l)}|} \lesssim 1.$$

Thus $(\sigma, w)_{p,q} \lesssim 1$.

On the other hand, consider the quadratic $\mathscr{A}_{p,q}$-condition, and choose some $K \in \{1, 2, \dots\}$. We set $a_k = 1$ for $k \in \{1, \dots, K\}$ and $a_k = 0$ for $k > K$. Then the construction of the measures shows that

$$\begin{aligned} (6.2) \qquad & \Big\| \Big(\sum_{k=1}^{K} \Big(a_k \frac{\sigma(Q_0^{(k)})}{|Q_0^{(k)}|} \Big)^2 1_{Q_0^{(k)}} \Big)^{1/2} \Big\|_{L^q(w)}^q \\ &= \sum_{k=1}^{K} \Big(\sum_{m=k}^{K} |Q_0^{(m)}|^{-2} \Big)^{q/2} |Q_0^{(k)}|^{q-1} |Q_0^{(k)} \setminus Q_0^{(k-1)}| \simeq \sum_{k=1}^{K} |Q_0^{(k)}|^{-q+q} = K, \end{aligned}$$

where in the second to last step we have used the fact that a geometric sum is about as large as its largest term.

For the quadratic $\mathscr{A}_{p,q}$-condition to hold, this should be dominated by

$$(6.3) \qquad [\sigma, w]_{p,q}^q \Big\| \Big(\sum_{k=1}^{K} 1_{Q_0^{(k)}} \Big)^{1/2} \Big\|_{L^p(\sigma)}^q = [\sigma, w]_{p,q}^q K^{q/2}.$$

Comparing (6.2) and (6.3), we see that since K was arbitrary, (6.3) can dominate (6.2) only if $q \geq 2$.

So if $q < 2$, we can construct a pair (σ, w) of weights such that $(\sigma, w)_{p,q} < \infty$ but $[\sigma, w]_{p,q} = \infty$. On the other hand, if $p > 2$, then $p' < 2$, and we can construct measures such that $(\sigma, w)_{q',p'} = (w, \sigma)_{p,q} < \infty$ and $[\sigma, w]_{q',p'} \simeq [w, \sigma]_{p,q} = \infty$. ∎

Combining Lemmas 3.2 and 6.1 we get the following corollary:

COROLLARY 6.2. *If $p, q \in (1, \infty)$, then the simple $A_{p,q}$-condition is sufficient for the two-weight Stein inequality* (3.3) *if and only if* $1 < p \leq 2 \leq q < \infty$.

Acknowledgements. I am a member of the Finnish Centre of Excellence in Analysis and Dynamics Research. I am grateful to my PhD advisor Tuomas Hytönen for showing me the idea of quadratic testing and for suggesting the problem. This work is part of my PhD project. I also wish to thank Timo Hänninen for teaching me many facts about dyadic shifts.

References

[1] D. L. Burkholder, *Martingale transforms*, Ann. Math. Statist. 37 (1966), 1494–1504.

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

[3] C. Fefferman and E. M. Stein, *Some maximal inequalities*, Amer. J. Math. 93 (1971), 107–115.

[4] T. P. Hytönen, *Representation of singular integrals by dyadic operators, and the A_2-theorem*, arXiv:1108.5119v1 [math.CA] (2011).

[5] T. P. Hytönen, *The sharp weighted bound for general Calderón–Zygmund operators*, Ann. of Math. (2) 175 (2012), 1473–1506.

[6] T. P. Hytönen, *The two-weight inequality for the Hilbert transform with general measures*, arXiv:1312.0843 [math.CA] (2013).

[7] T. Hytönen, C. Pérez, S. Treil and A. Volberg, *Sharp weighted estimates for dyadic shifts and the A_2 conjecture*, J. Reine Angew. Math. 687 (2014), 43–86.

[8] J.-P. Kahane, *Some Random Series of Functions*, 2nd ed., Cambridge Stud. Adv. Math. 5, Cambridge Univ. Press, Cambridge, 1985.

[9] M. T. Lacey, *Two-weight inequality for the Hilbert transform: a real variable characterization, II*, Duke Math. J. 163 (2014), 2821–2840.

[10] M. T. Lacey and K. Li, *Two weight norm inequalities for the g function*, Math. Res. Lett. 21 (2014), 521–536.

[11] M. T. Lacey, E. T. Sawyer, C.-Y. Shen and I. Uriarte-Tuero, *Two-weight inequality for the Hilbert transform: a real variable characterization, I*, Duke Math. J. 163 (2014), 2795–2820.

[12] J. Marcinkiewicz and A. Zygmund, *Quelques inégalités pour les opérations linéaires*, Fund. Math. 32 (1939), 115–121.

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

[14] F. Nazarov, S. Treil and A. Volberg, *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Res. Lett. 15 (2008), 583–597.

[15] J. L. Rubio de Francia, *Factorization theory and A_p weights*, Amer. J. Math. 106 (1984), 533–547.

[16] E. M. Stein, *Topics in Harmonic Analysis Related to the Littlewood–Paley Theory*, Ann. of Math. Stud. 63, Princeton Univ. Press, Princeton, NJ, and Univ. of Tokyo Press, Tokyo, 1970.

[17] E. Vuorinen, $L^p(\mu) \to L^q(\nu)$ *characterization for well localized operators*, J. Fourier Anal. Appl. 22 (2016), 1059–1075.
[18] L. Weis, *Operator-valued Fourier multiplier theorems and maximal* L_p*-regularity*, Math. Ann. 319 (2001), 735–758.

Emil Vuorinen
Department of Mathematics and Statistics
P.O.B. 68 (Gustaf Hällströmin Katu 2b)
FI-00014 University of Helsinki, Finland
E-mail: emil.vuorinen@helsinki.fi