Publ. Mat. **66** (2022), 631–681
DOI: 10.5565/PUBLMAT6622205

MULTIPLE VECTOR-VALUED, MIXED-NORM ESTIMATES FOR LITTLEWOOD–PALEY SQUARE FUNCTIONS

Cristina Benea and Camil Muscalu

Abstract: We prove that for any L^Q-valued Schwartz function f defined on $\mathbb{R}^d$, one has the multiple vector-valued, mixed-norm estimate

$$\|f\|_{L^P(L^Q)} \lesssim \|Sf\|_{L^P(L^Q)}$$

valid for every d-tuple P and every n-tuple Q satisfying $0 < P, Q < \infty$ componentwise. Here $S := S_{d_1} \otimes \cdots \otimes S_{d_N}$ is a tensor product of several Littlewood–Paley square functions S_{d_j} defined on arbitrary Euclidean spaces $\mathbb{R}^{d_j}$ for $1 \leq j \leq N$, with the property that $d_1 + \cdots + d_N = d$. This answers a question that came up implicitly in our recent works [**2**], [**3**], [**5**] and completes in a natural way classical results of Littlewood–Paley theory. The proof is based on the *helicoidal method* introduced by the authors in the aforementioned papers.

2020 Mathematics Subject Classification: 42B20.

Key words: multi-parameter Littlewood–Paley theory, multi-parameter Hardy spaces, mixed-norm estimates, weighted estimates and Littlewood–Paley theory.

1. Introduction

Let us start by recalling that a sequence of L^1-bounded Schwartz functions $(\psi_k)_{k\in\mathbb{Z}}$ defined on the Euclidean space $\mathbb{R}^m$ is called a Littlewood–Paley sequence, if its Fourier transform satisfies[1]

$$\text{(1)} \qquad \operatorname{supp} \widehat{\psi_k} \subseteq [2^{k-1}, 2^{k+1}], \quad |\partial^\alpha \widehat{\psi_k}(\xi)| \lesssim 2^{-|\alpha| k} \Big(1 + \frac{|\xi|}{2^k}\Big)^{-100\, m}$$

for every $\xi \in \mathbb{R}^m$ and sufficiently many multi-indices α, and if one also has

$$1 = \sum_{k\in\mathbb{Z}} \widehat{\psi_k}.$$

C. Muscalu is also a Member of the Simion Stoilow Institute of Mathematics of the Romanian Academy in Bucharest.

[1]Here and throughout the article we use the standard notation $A \lesssim B$, meaning that $A \leq CB$ for some constant $C > 0$ which can be universal or dependent on several implicit parameters derived from the specific context.

In particular, any Schwartz function f defined on $\mathbb{R}^m$ admits the Littlewood–Paley decomposition

$$f = \sum_{k\in\mathbb{Z}} f * \psi_k.$$

To any Littlewood–Paley sequence, one can also associate a Littlewood–Paley square function $S_m f$, defined by

$$S_m f(x) := \Big(\sum_{k\in\mathbb{Z}} |f * \psi_k(x)|^2\Big)^{1/2}. \tag{2}$$

Moreover, for any $N \geq 1$ such Littlewood–Paley sequences $(\psi_k^j)_{k\in\mathbb{Z}}$ defined on $\mathbb{R}^{d_j}$ for $1 \leq j \leq N$, one defines an N-parameter one $(\Psi_k)_{k\in\mathbb{Z}^N}$ on $\mathbb{R}^d := \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N}$ by

$$\Psi_k := \psi_{k_1}^1 \otimes \cdots \otimes \psi_{k_N}^N \tag{3}$$

for $k = (k_1, \ldots, k_N)$, where

$$\psi_{k_1}^1 \otimes \cdots \otimes \psi_{k_N}^N (x_1, \ldots, x_N) := \psi_{k_1}^1(x_1) \cdot \ldots \cdot \psi_{k_N}^N(x_N).$$

Here we think of the generic variable $x \in \mathbb{R}^d$ as being identified with the vector $(x_1, \ldots, x_N)$ with $x_j \in \mathbb{R}^{d_j}$ for $1 \leq j \leq N$.

In particular, any Schwartz function on $\mathbb{R}^d$ admits the decomposition

$$f = \sum_{k\in\mathbb{Z}^N} f * \Psi_k.$$

One can then also define the N-parameter square function Sf by the formula

$$Sf(x) := \Big(\sum_{k\in\mathbb{Z}^N} |f * \Psi_k(x)|^2\Big)^{1/2} \tag{4}$$

for $x \in \mathbb{R}^d$. This is the square function that will be studied in the present article.

To complete the presentation of the main notations that we will use, we also recall that given any $n \geq 1$ σ-finite measurable spaces (A_j, Σ_j, μ_j) for $1 \leq j \leq n$ and $R = (r_1, \ldots, r_n)$, an n-tuple of positive real numbers, one can define the iterated (or mixed-norm) Lebesgue space $L^R(A, \Sigma, \mu)$ to be the space containing those functions g which are measurable on the product space

$$(A, \Sigma, \mu) := \Big(\prod_{j=1}^n A_j, \prod_{j=1}^n \Sigma_j, \prod_{j=1}^n \mu_j\Big)$$

and for which the (quasi)-norm $\|g\|_R$ defined by

$$\|g\|_R := \|g\|_{L^R(A,\Sigma,\mu)} = \|\ldots\|g(a_1, \ldots, a_n)\|_{L^{r_n}(A_n,\Sigma_n,\mu_n)} \cdots \|_{L^{r_1}(A_1,\Sigma_1,\mu_1)}$$

is finite.

Classical Littlewood–Paley theory states that the inequalities

$$\|f\|_{L^p(\mathbb{R}^m)} \lesssim \|S_m f\|_{L^p(\mathbb{R}^m)} \lesssim \|f\|_{L^p(\mathbb{R}^m)} \tag{5}$$

are true, provided that $1 < p < \infty$ and that, in addition, the left-hand side of (5)

$$\|f\|_{L^p(\mathbb{R}^m)} \lesssim \|S_m f\|_{L^p(\mathbb{R}^m)} \tag{6}$$

is in fact available in the whole range $0 < p < \infty$; see for instance [**19**] and [**20**].

Standard duality and vector-valued arguments for singular integrals allow one to extend (5) very easily to the setting of mixed-norm spaces and N-parameter square functions. This implies that the inequalities

$$\|f\|_{L^P(L^Q)} \lesssim \|Sf\|_{L^P(L^Q)} \lesssim \|f\|_{L^P(L^Q)} \tag{7}$$

are true for L^Q-valued Schwartz functions defined in $\mathbb{R}^d$ for every n-tuple Q and d-tuple P satisfying $1 < P, Q < \infty$ componentwise.

To be more specific, the space L^P above is considered with respect to the product Lebesgue measure in $\mathbb{R}^d$, and as before, by $\|h\|_{L^P(L^Q)}$ one means the mixed (quasi)-norm given by

$$\|h\|_{L^P(L^Q)} := \| \; \|h(x,a)\|_{L^Q(A,\Sigma,\mu)} \, \|_{L^P(\mathbb{R}^d)}.$$

The main result of the present article proves an extension of the estimate (6).

Theorem 1.1. *The estimate*

$$\|f\|_{L^P(L^Q)} \lesssim \|Sf\|_{L^P(L^Q)} \tag{8}$$

is true, for every L^Q-valued Schwartz function f on $\mathbb{R}^d$, as long as the n-tuples Q and the d-tuples P satisfy the condition $0 < P, Q < \infty$ componentwise.

As we will see, unlike (7), the proof of Theorem 1.1 is far from being routine, and it is based on the *helicoidal method* developed by the authors in [**2**], [**3**], [**5**]. The question addressed and answered by Theorem 1.1 surfaced quite naturally in our recent works [**2**], [**3**] and it is related to an open problem of Kenig on mixed-norm estimates for paraproducts on polydisks. See also our recent expository work [**4**], in particular Theorem 5 there. We also refer the reader to Sections 5 and 6 of [**2**], where scalar and Banach-valued versions of Theorem 1.1 are used to prove mixed-norm estimates for multi-parameter operators. Thus the vector-valued extension in (8) for the multi-parameter square function is essential for widening the range of boundedness of those operators, although certain endpoints are still excluded.

Some particular cases of (8) were known in the scalar case, that is, when $L^Q = \mathbb{C}$. The case when all the entries of the d-tuple P are equal to each other is the well-known multi-parameter case studied by Gundy and Stein in [**12**]. More recently, Hart, Torres, and Wu proved the case when $N = 1$ and $d = 2$, again, in the scalar situation [**13**]. Even more recently, the case $N = 1$ was extended in [**14**] to arbitrary dimensions, also in an anisotropic setting.

The central point of the paper will be the proof of our main Theorem 1.1 based on techniques from [**2**], [**3**], and [**5**]. The $L^P(\mathbb{R}^d)$ mixed-norms, the iterated Lebesgue spaces $L^Q(A, \Sigma, \mu)$, and the N parameters could make the result seem convoluted, but several simplifications are attainable. We will first focus on the situation when $L^P(\mathbb{R}^d) = L^p(\mathbb{R}^d)$, leaving out the mixed-norms in a first instance; the same principle used for proving vector-valued extensions will be useful for proving the mixed-norm estimates as well, since both problems boil down to a change of (quasi)-norm. With this simplification, we are reduced to proving the vector-valued extension for the N-parameter square function

$$\|f\|_{L^p(L^Q)} \lesssim \|Sf\|_{L^p(L^Q)}, \tag{9}$$

for any $0 < p, Q < \infty$. Following the viewpoint of [**2**], multi-parameter operators can be treated using vector-valued extensions of operators depending on fewer parameters. Since our vector spaces are precisely iterated Lebesgue spaces, the treatment of the N-parameter square function is reduced eventually to multiple vector-valued estimates for the one-parameter square function.

In order to ease the presentation even more, we consider in the first part of the paper the case when all the square functions S_{d_j} for $1 \leq j \leq N$ are one-dimensional, that is, when $d_1 = \cdots = d_N = 1$; all the objects considered have unequivocal higher-dimensional analogues, and this reduction does not produce a loss of generality. Notice that this situation corresponds numerically to $N = d$. The proof of this case represents the core of the present article.

Under this assumption, we first show in Section 2 that the N-parameter estimate (8) follows easily, by induction, from the one-parameter case at the cost of increasing the complexity of the vector-valued spaces considered. Notice that in this situation, (8) becomes a multiple vector-valued extension of the well-known (scalar) inequality (6). Then, in Section 3, we explain how this multiple vector-valued case is implied by a certain discrete analogue of it.

Next, in Section 4, which is more involved, we describe the proof of this discrete case, by using ideas that lie at the heart of our helicoidal method in [**2**], [**3**], [**5**]. In Section 5 we explain how one can modify the proof presented up to now in order to deal with the general, mixed-norm case of Theorem 1.1.

Lastly, in the final Section 6, we will see how Theorem 1.1 can also be obtained through extrapolation from a weighted, scalar version of Theorem 1.1, which appeared in the context of weighted Hardy spaces in [**9**]. Since we are outside the Banach setting, the extrapolation needed concerns A_∞ weights and pairs of functions. For the mixed-norm estimates, we need to adapt a result of Kurtz [**15**].

That the vector-valued result of Theorem 1.1 allows also for a proof based on extrapolation and weighted theory should not be surprising: the helicoidal method yields vector-valued results that can be obtained also through extrapolation, once weighted estimates for the *correct* class of weights are known. This was the case also with the bilinear Hilbert transform (see [**2**], [**5**], [**6**], [**16**]). For completeness, in Subsection 6.2 we show how to deduce the weighted version of Theorem 1.1 by using the helicoidal method: the same maximal inequality used in Section 4 plays a central role, and only the stopping time algorithm changes.

Acknowledgements. C. Benea was partially supported by ERC Project FAnFArE no. 637510. C. Muscalu was partially supported by NSF Grant DMS 1500262. He also acknowledges partial support through a grant from the Ministry of Research and Innovation of Romania, CNCS - UEFISCDI, project PN-III-P4-ID-PCE-2016-0823 within PNCDI - III. During the spring semester of 2017, C. Muscalu was a member of the MSRI in Berkeley, as part of the Program in Harmonic Analysis, and during the fall semester of 2017 he was a visiting Simons Fellow at the Mathematics Department of Université Paris-Sud Orsay. He is grateful to both institutions for their hospitality, and to the Simons Foundation for their generous support.

The authors are grateful to Dachun Yang for pointing out the results in [**14**].

2. Reduction to the multiple vector-valued case

As mentioned above, we first study the case when $d_1 = \cdots = d_N = 1$. From now on, until Section 5, we work under this assumption.

And as also mentioned in the introduction, in this section we show that Theorem 1.1 follows by induction, from its particular case $d = 1$. Recall also that $d = N$ now. Let us therefore assume that Theorem 1.1 is true for dimensions smaller than or equal to $d - 1$ and we will explain how to deduce the d-dimensional case. The argument is based on the identity

$$Sf = \Big(\sum_{k\in\mathbb{Z}} |S_{(x_1,\dots,x_{d-1})}(f * \psi_k^d)|^2\Big)^{1/2}, \tag{10}$$

where $S_{(x_1,\dots,x_{d-1})}(g)$ denotes the $(d-1)$-dimensional part of the square function, taken with respect to the variables $x_1, \dots, x_{d-1}$, and explicitly given by

$$S_{(x_1,\dots,x_{d-1})}(g)(x) := \Big(\sum_{k_1,\dots,k_{d-1}} |g * (\psi^1_{k_1} \otimes \cdots \otimes \psi^{d-1}_{k_{d-1}})(x)|^2\Big)^{1/2} \tag{11}$$

for $x \in \mathbb{R}^d$. The first convolution in (10) is a one-dimensional one, taken with respect to the last variable x_d, while the convolution in (11) is a $(d-1)$-dimensional one, taken with respect to the first $d-1$ variables $x_1, \dots, x_{d-1}$. Using (10) one can write $\|Sf\|_{L^P(L^Q)}$ as

$$\begin{aligned}\Big\|\Big(\sum_{k\in\mathbb{Z}} |S_{(x_1,\dots,x_{d-1})}(f * \psi_k^d)|^2\Big)^{1/2}\Big\|_{L^P(L^Q)}&\\ &= \|(S_{(x_1,\dots,x_{d-1})}(f * \psi_k^d))_k\|_{L^P(L^Q(\ell^2))}\\ &= \|(S_{(x_1,\dots,x_{d-1})}(f * \psi_k^d))_k\|_{L^{\widetilde{P}}(L^{p_d}(L^Q(\ell^2)))},\end{aligned}$$

where $\widetilde{P} := (p_1, \dots, p_{d-1})$.

Here, one can use the induction hypothesis in the $(d-1)$-dimensional case to conclude that the above expression is larger than

$$\|(f * \psi_k^d)_k\|_{L^{\widetilde{P}}(L^{p_d}(L^Q(l^2)))} = \Big\|\Big(\sum_k |f * \psi_k^d|^2\Big)^{1/2}\Big\|_{L^{\widetilde{P}}(L^{p_d}(L^Q))}.$$

Finally, by using the one-dimensional case and Fubini, we see that this is also greater than

$$\|f\|_{L^{\widetilde{P}}(L^{p_d}(L^Q))} = \|f\|_{L^P(L^Q)},$$

which ends the argument.

3. The discrete multiple vector-valued case

Now that we know that Theorem 1.1 (in the special situation when $d_1 = \cdots = d_N = 1$) can be reduced to its $d = 1$ particular case, we show in this section that a further reduction is possible. The multiple vector-valued $d = 1$ case can be reduced to a *discrete* variant of it that will be described next.

Let us pause briefly and recall that a sequence of Schwartz functions $(\phi_I)_I$ on the real line, indexed by dyadic intervals I, is called an L^p normalized *lacunary* sequence (for some $p \in (0, \infty]$), if and only if the following estimates hold:

$$|\partial^\alpha \phi_I(x)| \lesssim \frac{1}{|I|^{1/p}} \frac{1}{|I|^\alpha} \Big(1 + \frac{\operatorname{dist}(x, I)}{|I|}\Big)^{-100} \tag{12}$$

for $x \in \mathbb{R}$, $0 \le \alpha \le 10$, and also if $\int_{\mathbb{R}} \phi_I(x)\, dx = 0$.

Now let $(\phi_I^1)_I$ and $(\phi_I^2)_I$ be two such L^2-normalized lacunary sequences, indexed by an arbitrary finite subset of dyadic intervals. The following discrete variant of the one-dimensional case of Theorem 1.1 is true.

Theorem 3.1. *For every $0 < p < \infty$ and tuple Q as before, one has*

$$\Big\| \sum_I \langle f, \phi_I^1 \rangle \phi_I^2 \Big\|_{L^p(L^Q)} \lesssim \Big\| \Big(\sum_I \frac{|\langle f, \phi_I^1\rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_{L^p(L^Q)}. \tag{13}$$

Observation 3.2. The function f above depends on the variables $(a_1, \ldots, a_n) \in A$ and on $x \in \mathbb{R}$. Sometimes we will write this explicitly as $f_{(a_1,\ldots,a_n)}(x)$. It is important to emphasize that, as we will see from the proof of Theorem 3.1, the estimate (13) holds also in the more general case when the families $(\phi_I^1)_I$ and $(\phi_I^2)_I$ depend on the variables $(a_1, \ldots, a_n) \in A$ as well, in a uniform manner, with respect to the implicit constants of (12).

We will now explain why Theorem 3.1 implies the one-dimensional case of Theorem 1.1. The argument is based on an idea that we learned from the article [**13**], and which goes back to the work of Frazier and Jawerth [**11**].

Proposition 3.3. *There exists a large universal constant N such that, given any sequence of intermediate points $x_I \in I$, there exists $(\tilde{\psi}_I)_I$, an L^∞ normalized lacunary sequence, so that every Schwartz function h on the real line can be decomposed as*

$$h = \sum_k \sum_{|I| = 2^{-k}} (h * \psi_{k-N})(x_I) \tilde{\psi}_I. \tag{14}$$

In (14), the sequence $(\psi_l)_l$ is any a priori fixed Littlewood–Paley sequence. We prove Proposition 3.3 in detail later on. In what follows, we describe how it helps reduce the $d = 1$ case of Theorem 1.1 to its discrete analogue from Theorem 3.1.

Fix $f(= f_{(a_1,\dots,a_n)}(x))$. For every $(a_1,\dots,a_n)\in A$ pick $x_I\in I$, a number with the property that

$$\inf_{y\in I}|(f_{(a_1,\dots,a_n)}*\psi_{k-N})(y)| = |(f_{(a_1,\dots,a_n)}*\psi_{k-N})(x_I)|,$$

where I is a dyadic interval with $|I|=2^{-k}$. Clearly, x_I depends on f and also, implicitly, on $(a_1,\dots,a_n)\in A$.

Using Proposition 3.3, one can write

$$\|f\|_{L^p(L^Q)} = \Big\|\sum_k\sum_{|I|=2^{-k}}(f_{(a_1,\dots,a_n)}*\psi_{k-N})(x_I)\psi_{I,(a_1,\dots,a_n)}(x)\Big\|_{L^p(L^Q)}. \tag{15}$$

Using now the general form of Theorem 3.1 (see Observation 3.2 that followed it) one can majorize the above expression (15) further by

$$\Big\|\Big(\sum_k\sum_{|I|=2^{-k}}|(f_{(a_1,\dots,a_n)}*\psi_{k-N})(x_I)|^2\mathbf{1}_I(x)\Big)^{1/2}\Big\|_{L^p(L^Q)},$$

and using the definition of the sequence $(x_I)_I$ above, one can immediately see that this is smaller than

$$\Big\|\Big(\sum_k|(f_{(a_1,\dots,a_n)}*\psi_{k-N})(x)|^2\Big)^{1/2}\Big\|_{L^p(L^Q)} = \Big\|\Big(\sum_k|f*\psi_k|^2\Big)^{1/2}\Big\|_{L^p(L^Q)},$$

as desired.

3.1. Proof of Proposition 3.3. We now describe the proof of Proposition 3.3 using the ideas from [**11**].

Start by writing, for a generic function of a variable f:

$$f=\sum_k f*\psi_k=\sum_k f*\psi_{k-N}.$$

We will prove that for every $k\in\mathbb{Z}$, a family of functions $(\tilde\psi_I)_I$ as in Proposition 3.3 exists[2], so that

$$f*\psi_{k-N}=\sum_{|I|=2^{-k}}(f*\psi_{k-N})(x_I)\tilde\psi_I. \tag{16}$$

Clearly, this would be enough. Since the argument is scale-invariant, we will prove this in the particular case when $k=N$. In this case, (16) becomes

$$f*\psi_0=\sum_{|I|=2^{-N}}(f*\psi_0)(x_I)\tilde\psi_I. \tag{17}$$

[2]This time all the intervals I have the same length, $|I|=2^{-k}$.

Now consider $\widetilde{\psi}$, a Schwartz function such that $\widehat{\widetilde{\psi}} = 1$ on the support of $\widehat{\psi_0}$ and having the property that $\operatorname{supp} \widehat{\widetilde{\psi}} \subseteq [1/4, 8]$.

Then, one can write

$$
\begin{aligned}
f * \psi_0(x) &= (f * \psi_0) * \widetilde{\psi}(x) = \int_{\mathbb{R}} f * \psi_0(y) \widetilde{\psi}(x-y)\, dy \\
&= \sum_{|I|=2^{-N}} \int_{\mathbb{R}} f * \psi_0(y) \widetilde{\psi}(x-y) \mathbf{1}_I(y)\, dy \\
&= \sum_{|I|=2^{-N}} f * \psi_0(x_I) \int_{\mathbb{R}} \widetilde{\psi}(x-y) \mathbf{1}_I(y)\, dy + \mathrm{Rest}_1(x) \\
&= \sum_{|I|=2^{-N}} f * \psi_0(x_I) \phi_I^1(x) + \mathrm{Rest}_1(x),
\end{aligned} \tag{18}
$$

where $\phi_I^1 := \widetilde{\psi} * \mathbf{1}_I(x)$ and

$$
\begin{aligned}
\mathrm{Rest}_1(x) &=: \sum_{|I|=2^{-N}} \mathrm{Rest}_{1,I}(x) \\
&= \sum_{|I|=2^{-N}} \int_{\mathbb{R}} [f * \psi_0(y) - f * \psi_0(x_I)] \widetilde{\psi}(x-y) \mathbf{1}_I(y)\, dy.
\end{aligned} \tag{19}
$$

The above inner expression can be estimated by

$$
\begin{aligned}
f * \psi_0(y) - f * \psi_0(x_I) &= \int_{\mathbb{R}} f(z) [\psi_0(y-z) - \psi_0(x_I - z)]\, dz \\
&= \int_{\mathbb{R}} f(z) \psi_0'(\# - z)(y - x_I)\, dz,
\end{aligned}
$$

where # is a point lying inside the interval I and depending on y, x_I, and z. Since both y and x_I belong to I, it is easy to see that the above expression is at most $C\, 2^{-N} \|f\|_\infty$. Using this in (19) we obtain that

$$
|\mathrm{Rest}_{1,I}(x)| \leq C_{\tilde{M}} 2^{-N} \|f\|_\infty (1 + \operatorname{dist}(x, I))^{-\tilde{M}} |I|,
$$

which implies further

$$
|\mathrm{Rest}_1(x)| \leq C \|f\|_\infty 2^{-N}.
$$

We see these calculations as providing a first approximation towards the desired (17). To summarize, so far we have shown that

$$
f * \psi_0(x) = \sum_{|I|=2^{-N}} f * \psi_0(x_I) \phi_I^1(x) + \mathrm{Rest}_1(x), \tag{20}
$$

where $|\mathrm{Rest}_1(x)| \leq C\, \|f\|_\infty 2^{-N}$ and $(\phi_I^1)_I$ is a lacunary family.

We now iterate this fact carefully. Fix J with $|J| = 2^{-N}$ and recall the expression

$$\text{Rest}_{1,J}(x) = \int_{\mathbb{R}} [f * \psi_0(y) - f * \psi_0(x_J)]\widetilde{\psi}(x-y)\mathbf{1}_J(y)\,dy. \tag{21}$$

Using (20) for $x = y$ and $x = x_J$ in (21) we obtain a decomposition of $\text{Rest}_{1,J}(x)$ of the type

$$\sum_{|I|=2^{-N}} f * \psi_0(x_I) \int_{\mathbb{R}} [\phi_I^1(y) - \phi_I^1(x_J)]\widetilde{\psi}(x-y)\mathbf{1}_J(y)\,dy \\ + \sum_{|I|=2^{-N}} \int_{\mathbb{R}} [\text{Rest}_{1,I}(y) - \text{Rest}_{1,I}(x_J)]\widetilde{\psi}(x-y)\mathbf{1}_J(y)\,dy.$$

Summing over $|J| = 2^{-N}$, we obtain the formula

$$\text{Rest}_1(x) = \sum_{|I|=2^{-N}} f * \psi_0(x_I)\phi_I^2 + \text{Rest}_2(x),$$

where

$$\phi_I^2(x) := \sum_{|J|=2^{-N}} \int_{\mathbb{R}} [\phi_I^1(y) - \phi_I^1(x_J)]\widetilde{\psi}(x-y)\mathbf{1}_J(y)\,dy,$$

while

$$\text{Rest}_2(x) = \sum_{|I|=2^{-N}} \text{Rest}_{2,I}(x)$$

and

$$\text{Rest}_{2,I}(x) := \sum_{|J|=2^{-N}} \int_{\mathbb{R}} [\text{Rest}_{1,I}(y) - \text{Rest}_{1,I}(x_J)]\widetilde{\psi}(x-y)\mathbf{1}_J(y)\,dy.$$

Arguing exactly as before, given that both y and x_J belong to the interval J, it is not difficult to see that $(\phi_I^2)_I$ is a lacunary family satisfying

$$\|\phi_I^2\|_\infty \le C\,2^{-N}, \quad \text{while} \quad \|\,\text{Rest}_2\,\|_\infty \le C^2\,2^{-2N}\|f\|_\infty,$$

where, as always, C is a universal constant. In other words, at our second approximation step, we obtain the decomposition

$$f * \psi_0(x) = \sum_{|I|=2^{-N}} f * \psi_0(x_I)(\phi_I^1(x) + \phi_I^2(x)) + \text{Rest}_2(x).$$

Iterating this an arbitrary number of times, we obtain that $f * \psi_0(x)$ can be written as

$$f * \psi_0(x) = \sum_{|I|=2^{-N}} f * \psi_0(x_I)(\phi_I^1(x) + \cdots + \phi_I^l(x)) + \text{Rest}_l(x), \tag{22}$$

where $(\phi_I^j)_I$ is a lacunary family satisfying

$$\|\phi_I^j\|_\infty \le C^{j-1}2^{-(j-1)N} \quad \text{while} \quad \|\text{Rest}_l\|_\infty \le C^l 2^{-lN}\|f\|_\infty.$$

Thus, if N is large enough so that $C\,2^{-N} < 1$, by letting l go to ∞ in (22), we obtain the desired decomposition (17) with $\tilde{\psi}_I$ given by

$$\tilde{\psi}_I(x) := \sum_{l=1}^{\infty} \phi_I^l(x).$$

Strictly speaking, the families $(\phi_I^l)_I$ are naturally associated to intervals of length 1, not 2^{-N}, but since N is a fixed universal constant, it is not difficult to see that they satisfy the estimates (12) as well, at the expense of losing a harmless constant of the type 2^{1000N}. This completes the proof of Proposition 3.3.

4. Proof of Theorem 3.1

Recall that our goal now is to prove that

$$\Big\|\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big\|_{L^p(L^Q)} \lesssim \Big\|\Big(\sum_{I\in\mathcal{I}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p(L^Q)} \tag{23}$$

for every $0 < p < \infty$ and every n-tuple Q of positive real numbers. Also, $\mathcal{I}$ is a fixed finite collection of dyadic intervals. Of course, the implicit constant in (23) is meant to be independent of the cardinality of $\mathcal{I}$. We also denote by $\overline{\mathcal{I}}$ the collection of all dyadic intervals J having the property that there exists $I \in \mathcal{I}$ such that $I \subseteq J$ and satisfying $|J| \leq 2_0^M$ for some large fixed positive integer M_0. Sometimes we refer to the intervals in $\overline{\mathcal{I}}$ as being the *relevant* dyadic intervals.

Now let $E \subseteq \mathbb{R}$ be a measurable subset. To prove (23) it is necessary to prove a more refined version of it given by

$$\begin{aligned}\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\mathbf{1}_E\Big\|_{L^p(L^Q)} \\ \lesssim \Big\|\Big(\sum_{I\in\mathcal{I}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p(L^Q)}\cdot(\mathrm{size}_{\mathcal{I}}\,\mathbf{1}_E)^{1/p-\epsilon},\end{aligned} \tag{24}$$

where $\epsilon > 0$ is arbitrarily small while

$$\mathrm{size}_{\mathcal{I}}\,\mathbf{1}_E := \sup_{I\in\overline{\mathcal{I}}}\frac{1}{|I|}\int_{\mathbb{R}}\mathbf{1}_E(x)\Big(1+\frac{\mathrm{dist}(x,I)}{|I|}\Big)^{-100}dx \tag{25}$$

is essentially the supremum over all L^1 averages of $\mathbf{1}_E(x)$ over the intervals of $\overline{\mathcal{I}}$. The ϵ in the right-hand side of (24) represents a small loss in the information corresponding to the localizing function $\mathbf{1}_E$, which will be traded later in (47) for the overall summability of the information corresponding to the level sets. The reader familiar with our earlier “helicoidal papers” [**2**], [**3**], and [**5**] will find our desire to prove (24) natural.

Clearly, (24) implies (23) since one can take E to be the whole real line $\mathbb{R}$.

Using interpolation arguments (see Proposition 4.1), it is enough to prove a weaker version of (24), namely

$$(26)\quad \Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \mathbf{1}_E \Big\|_{L^{p,\infty}(L^Q)} \lesssim \Big\| \Big(\sum_{I \in \mathcal{I}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_{L^p(L^Q)} \cdot (\text{size}_{\mathcal{I}} \, \mathbf{1}_E)^{1/p - \epsilon}.$$

Such interpolation arguments will in fact be freely used throughout the section, until the end of it, when they will be proved in detail.

Let us denote by $\mathbb{P}(n)$ the statement which says that (26) holds in full generality, for $0 < p < \infty$ and Q n-tuple of positive real numbers. We will prove $\mathbb{P}(n)$ by induction for every $n \geq 0$.

4.1. Proof of $\mathbb{P}(0)$. This is the scalar case, which now reads as

$$(27)\quad \Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \mathbf{1}_E \Big\|_{L^{p,\infty}} \lesssim \Big\| \Big(\sum_{I \in \mathcal{I}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_{L^p} \cdot (\text{size}_{\mathcal{I}} \, \mathbf{1}_E)^{1/p - \epsilon}.$$

Let s be any positive real number with the property $s \leq \min(1, p)$. To estimate the left-hand side of (27) we dualize the expression through L^s, as explained in [**3**]. Given also the scale invariance of the inequality, this amounts to proving that for every $F \subseteq \mathbb{R}$ measurable set with $|F| = 1$, there exists a subset of it $\widetilde{F} \subseteq F$ with $|\widetilde{F}| > 1/2$ such that

$$(28)\quad \Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \mathbf{1}_E \mathbf{1}_{\widetilde{F}} \Big\|_{L^s} \lesssim \Big\| \Big(\sum_{I \in \mathcal{I}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_{L^p} \cdot (\text{size}_{\mathcal{I}} \, \mathbf{1}_E)^{1/p - \epsilon}.$$

To construct the subset $\widetilde{F}$, we start by defining an exceptional set Ω as follows.

First, for every integer $k \geq 0$ we define

$$\Omega_k := \{x : Sf(x) > C 2^{10k/p} \|Sf\|_p\}.$$

Here, and from now on, by $Sf(x)$ we mean the"discrete" Littlewood–Paley square function given by

$$Sf(x) := \Big(\sum_{I\in\mathcal{I}} \frac{|\langle f, \phi_I^1\rangle|^2}{|I|}\mathbf{1}_I(x)\Big)^{1/2}. \tag{29}$$

When we need to emphasize that the square function above depends on the collection $\mathcal{I}$, we write $S_{\mathcal{I}}$.

It is not difficult to see that

$$|\Omega_k| \le \frac{1}{2^{10k}}\frac{1}{C^p}.$$

After that we set

$$\widetilde{\Omega_k} := \{x : M(\mathbf{1}_{\Omega_k})(x) > 1/2^k\},$$

where M is the Hardy–Littlewood maximal operator, and finally

$$\Omega := \bigcup_{k=0}^{\infty} \widetilde{\Omega_k}. \tag{30}$$

Clearly,

$$|\widetilde{\Omega_k}| \le \widetilde{C}\, 2^k |\Omega_k| \le \widetilde{C}\, 2^k \frac{1}{2^{10k}}\frac{1}{C^p} = \frac{\widetilde{C}}{C^p}\frac{1}{2^{9k}}$$

and in particular this implies that $|\Omega| < 1/10$ if C is a large enough constant[3].

In the end we set $\widetilde{F} := F\backslash\Omega$, which is a major subset of F, in the sense that it satisfies $|\widetilde{F}| \sim 1$. Now, using a result from [**18**], we decompose the functions ϕ_I^2 as

$$\phi_I^2 = \sum_{\ell=0}^{\infty} 2^{-\tilde{M}\,\ell}\phi_{I,\ell}^2, \tag{31}$$

where $\tilde{M}$ is arbitrarily large and for each $\ell \ge 0$, $(\phi_{I,\ell}^2)_I$ is still a lacunary family with the additional property that

$$\operatorname{supp}\phi_{I,\ell}^2 \subseteq 2^\ell I.$$

In particular, one can estimate the left-hand side of (28) by

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_s^s \lesssim \sum_{\ell=0}^{\infty} 2^{-\tilde{M}s\,\ell}\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_{I,\ell}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_s^s. \tag{32}$$

[3]The constant $\widetilde{C}$ is the boundedness constant of $M\colon L^1 \to L^{1,\infty}$.

The right-hand side of (32) can also be rewritten as

$$\sum_{\ell=0}^{\infty} 2^{-\tilde{M}s\ell/2}\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,\ell}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_s^s,$$

where $\widetilde{\phi}_{I,\ell}^2 := 2^{-\tilde{M}\,\ell/2}\phi_{I,\ell}^2$. We will see in what follows that for each $\ell \geq 0$ one has

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,\ell}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_s^s \lesssim 2^{L\ell}\|Sf\|_p^s\cdot(\operatorname{size}_{\mathcal{I}}\mathbf{1}_E)^{(1/p-\epsilon)s}, \tag{33}$$

where L is some constant depending on s and p. However, because of the large constant $\tilde{M}$ in (32), this will be enough to complete our proof. We will prove (33) in detail in the main case when $\ell = 0$ and then we will explain how to modify the argument to obtain (33) in general.

In other words, the goal for us now is to prove that

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_s^s \lesssim \|Sf\|_p^s\cdot(\operatorname{size}_{\mathcal{I}}\mathbf{1}_E)^{(s/p-\epsilon)}. \tag{34}$$

Recall that now, since

$$\operatorname{supp}\widetilde{\phi}_{I,0}^2 \subseteq I,$$

one must have $I\cap\Omega^c \neq \emptyset$, which in particular implies that $I\cap\Omega_0^c\neq\emptyset$. From the definition of Ω_0, one can see that this set admits a natural decomposition as a disjoint union of maximal dyadic intervals denoted by $I_{\max}$. In particular, our dyadic intervals I have the property that they are either disjoint from all these $I_{\max}$ or they contain strictly at least one of them. In either case, it is not difficult to see that one has the pointwise estimate

$$\Big(\sum_{I\in\mathcal{I}:I\cap\Omega_0^c\neq\emptyset}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I(x)\Big)^{1/2}\leq\widetilde{C}\|Sf\|_p, \tag{35}$$

where $\widetilde{C}$ is a universal constant. To prove (34) we will combine two stopping time arguments, one performed with the help of averages of the type

$$\frac{1}{|I_0|^{1/p}}\Big\|\Big(\sum_{I\subseteq I_0}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p} \tag{36}$$

and the other one with the help of averages of the type

$$\frac{1}{|I_0|}\int_{\mathbb{R}}\mathbf{1}_{E\cap\widetilde{F}}(x)\Big(1+\frac{\operatorname{dist}(x,I_0)}{|I_0|}\Big)^{-100}dx. \tag{37}$$

The latter will be denoted from now on by $\mathrm{ave}^1_{I_0}(\mathbf{1}_{E\cap\widetilde{F}})$. Clearly, because of the pointwise bound (35), averages such as the ones in (36) cannot be larger than $\widetilde{C}\,\|Sf\|_p$, while averages of the type (37) cannot be larger than $\mathrm{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}})$.

We now describe in detail the first stopping time.

We start by selecting maximal dyadic intervals $I_0\in\overline{\mathcal{I}}$ with the property that $I_0\cap\Omega_0^c\neq\emptyset$ and such that

$$\frac{1}{|I_0|^{1/p}}\Big\|\Big(\sum_{I\subseteq I_0}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p}\geq\frac{\widetilde{C}}{2}\|Sf\|_p. \tag{38}$$

Of course, as pointed out before, we implicitly assume that all the intervals I that participate in the summation above have the property $I\cap\Omega_0^c\neq\emptyset$. It is also important to observe that these selected intervals I_0 are all disjoint, as a consequence of their maximality. Then, we disregard all the relevant dyadic intervals that lie inside one of these selected intervals and consider only those that are left. They are either disjoint from the selected ones or they contain at least one of the selected ones.

After this, among those that are left, we pick those maximal ones, still denoted by I_0, for which

$$\frac{1}{|I_0|^{1/p}}\Big\|\Big(\sum_{I\subseteq I_0}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p}\geq\frac{\widetilde{C}}{2^2}\|Sf\|_p \tag{39}$$

and so forth. The maximal intervals selected at the first step are collected in $\mathcal{I}_1^{(1)}$, those selected at the second step are collected in $\mathcal{I}_2^{(1)}$ and so on, obtaining the collections $(\mathcal{I}_{n_1}^{(1)})_{n_1}$. Clearly, there are only finitely many such steps, since our initial collection of intervals was finite.

After that, independently, we perform a similar stopping time, but one that involves the averages $\mathrm{ave}^1_{I_0}(\mathbf{1}_{E\cap\widetilde{F}})$ instead. We start by selecting those maximal intervals I_0 for which

$$\mathrm{ave}^1_{I_0}(\mathbf{1}_{E\cap\widetilde{F}})>\frac{1}{2}\,\mathrm{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}})$$

then, among those that are left (more specifically, those that are not inside any of the previously selected I_0) we again pick those maximal I_0 for which

$$\mathrm{ave}^1_{I_0}(\mathbf{1}_{E\cap\widetilde{F}})>\frac{1}{2^2}\,\mathrm{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}}) \tag{40}$$

and so on, exactly as before. In this way, one obtains a sequence of collections of maximal dyadic intervals I_0 denoted by $(\mathcal{I}_{n_2}^{(2)})_{n_2}$.

In the end, we combine them to be able to estimate (34). One can write

$$\left\|\Big(\sum_{I\in\mathcal{I},I\cap\Omega_0^c\neq\emptyset}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\right\|_s^s \tag{41}$$
$$\leq\sum_{n_1,n_2\geq 0}\sum_{I_1\in\mathcal{I}_{n_1}^{(1)},I_2\in\mathcal{I}_{n_2}^{(2)}}\left\|\Big(\sum_{I\in\mathcal{I}_{n_1}^{(1)}(I_1)\cap\mathcal{I}_{n_2}^{(2)}(I_2)}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\right\|_s^s,$$

where $\mathcal{I}_{n_1}^{(1)}(I_1)$ contains all the relevant dyadic intervals I with the property that $I\subseteq I_1$ but such that I is not contained in any of the previously selected intervals in $\mathcal{I}_l^{(1)}$ for $0\leq l\leq n_1-1$, and similarly for $\mathcal{I}_{n_2}^{(2)}(I_2)$. Clearly, any interval I participating in the summation (41) must satisfy $I\subseteq I_1\cap I_2$. Now, for every I_1, I_2 as before, the corresponding L^s quasi-norm in (41) can be estimated by

$$\left\|\Big(\sum_{I\subseteq I_1\cap I_2}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\right\|_1\cdot|E\cap\widetilde{F}\cap I_1\cap I_2|^{\frac{1-s}{s}} \tag{42}$$

by using Hölder, since $s\leq 1$. The L^1 norm in (42) can be dualized and estimated by

$$\sum_{I\subseteq I_1\cap I_2}\langle f,\phi_I^1\rangle\langle\mathbf{1}_{E\cap\widetilde{F}}\,g,\widetilde{\phi}_{I,0}^2\rangle$$

for some function g with the property $\|g\|_\infty=1$. Using Cauchy–Schwarz this can be further estimated by

$$\frac{1}{|I_1\cap I_2|^{1/2}}\left\|\Big(\sum_{I\subseteq I_1\cap I_2}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\right\|_2$$
$$\times\frac{1}{|I_1\cap I_2|^{1/2}}\left\|\Big(\sum_{I\subseteq I_1\cap I_2}\frac{|\langle\mathbf{1}_{E\cap\widetilde{F}}\,g,\widetilde{\phi}_{I,0}^2\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\right\|_2\cdot|I_1\cap I_2|.$$

Now using John–Nirenberg twice (see Theorem 2.10 in [**19**] for this robust, discrete variant of it) together with the standard local estimate of weak-L^1 averages (which can be found in Lemma 2.16 of [**19**], for instance), this can be further majorized by

$$\Big(\sup_{J_1\subseteq I_1\cap I_2}\frac{1}{|J_1|^{1/p}}\left\|\Big(\sum_{I\subseteq J_1}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\right\|_p\Big) \tag{43}$$
$$\times\big(\sup_{J_2\subseteq I_1\cap I_2}\operatorname{ave}^1_{J_2}(\mathbf{1}_{E\cap\widetilde{F}})\big)\cdot|I_1\cap I_2|.$$

If one raises these estimates to the power s, as required by (41), one can see that the corresponding expression there is smaller than

$$\begin{aligned}(44)\qquad &\Big(\sup_{J_1\subseteq I_1\cap I_2}\frac{1}{|J_1|^{1/p}}\Big\|\Big(\sum_{I\subseteq J_1}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_p\Big)^s\\ &\qquad\times\big(\sup_{J_2\subseteq I_1\cap I_2}\operatorname{ave}^1_{J_2}(\mathbf{1}_{E\cap\widetilde{F}})\big)^s\cdot|I_1\cap I_2|^s\\ &\qquad\times\big(\operatorname{ave}^1_{I_1\cap I_2}(\mathbf{1}_{E\cap\widetilde{F}})\big)^{1-s}\cdot|I_1\cap I_2|^{1-s},\end{aligned}$$

which is smaller still than

$$\begin{aligned}(45)\qquad &\Big(\sup_{J_1\subseteq I_1\cap I_2}\frac{1}{|J_1|^{1/p}}\Big\|\Big(\sum_{I\subseteq J_1}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_p\Big)^s\\ &\qquad\times\big(\sup_{J_2\subseteq I_1\cap I_2}\operatorname{ave}^1_{J_2}(\mathbf{1}_{E\cap\widetilde{F}})\big)\cdot|I_1\cap I_2|.\end{aligned}$$

Using these estimates in (41), the expression there can be estimated further by

$$(46)\qquad \sum_{n_1,n_2\geq 0}\ \sum_{I_1\in\mathcal{I}^{(1)}_{n_1},\,I_2\in\mathcal{I}^{(2)}_{n_2}}(2^{-n_1}\tilde{C}\|Sf\|_p)^s(2^{-n_2}\operatorname{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}}))|I_1\cap I_2|.$$

On the other hand the expression

$$\sum_{I_1\in\mathcal{I}^{(1)}_{n_1},\,I_2\in\mathcal{I}^{(2)}_{n_2}}|I_1\cap I_2|$$

is smaller than

$$\sum_{I_1\in\mathcal{I}^{(1)}_{n_1}}|I_1|\lesssim 2^{n_1p}$$

and also smaller than

$$\sum_{I_2\in\mathcal{I}^{(2)}_{n_2}}|I_2|\lesssim 2^{n_2}(\operatorname{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}}))^{-1},$$

given that $|\widetilde{F}|\sim 1$. This implies that

$$\sum_{I_1\in\mathcal{I}^{(1)}_{n_1},\,I_2\in\mathcal{I}^{(2)}_{n_2}}|I_1\cap I_2|\lesssim 2^{n_1p\theta_1}2^{n_2\theta_2}(\operatorname{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}}))^{-\theta_2}$$

for every $0\leq\theta_1,\theta_2\leq 1$ so that $\theta_1+\theta_2=1$. Using this in (46) one can majorize that expression by

$$(47)\qquad \|Sf\|_p^s(\operatorname{size}_{\mathcal{I}}(\mathbf{1}_{E\cap\widetilde{F}}))^{1-\theta_2}\sum_{n_1,n_2\geq 0}2^{-n_1(s-p\theta_1)}2^{-n_2(1-\theta_2)}.$$

The geometric series on the left-hand side above are convergent provided $\theta_1 < s/p$ and $\theta_2 < 1$, which is equivalent to

$$0 < \theta_1 < \frac{s}{p}.$$

If θ_1 is taken to be very close to s/p, this gives an upper bound of the type

$$\|Sf\|_p^s \cdot (\text{size}_{\mathcal{I}} \, \mathbf{1}_E)^{s/p-\epsilon},$$

as desired in (34).

To prove (33) for arbitrary $\ell > 0$ one proceeds similarly. The observation now is that since $\text{supp}\, \widetilde{\phi}^2_{I,\ell} \subseteq 2^\ell I$ one must have

$$2^\ell I \cap \Omega^c \neq \emptyset$$

and it is not difficult to see that this implies that

$$I \cap \Omega_\ell^c \neq \emptyset.$$

Indeed, if this were not true, then $I \subseteq \Omega_\ell$, which means that $2^\ell I \subseteq \widetilde{\Omega}_\ell \subseteq \Omega$, a contradiction. Here we can see the connection between the initially independent decomposition (31) of the functions ϕ_I^2 and the level sets Ω_k and $\tilde{\Omega}_k$ used in (30) to define the exceptional set Ω.

Now one simply repeats the previous argument. One difference is that the first L^p averages of the square function can be as large as $C2^{10\ell/p}\|Sf\|_p$, a bound which is responsible for the positive constant L in (33). Another difference is in the estimate (42), whose analogue now contains a factor of the type

$$|E \cap \widetilde{F} \cap 2^\ell (I_1 \cap I_2)|^{\frac{1-s}{s}}.$$

However, the small constant $2^{-\tilde{M}\ell/2}$ in the definition of $\widetilde{\phi}^2_{I,\ell}$ gets multiplied by it, and this allows one to write

$$2^{-\tilde{M}\ell/2}|E\cap\widetilde{F}\cap 2^\ell(I_1\cap I_2)|^{\frac{1-s}{s}} \lesssim \Big(\int_{\mathbb{R}} \mathbf{1}_{E\cap\widetilde{F}}(x)\Big(1+\frac{\text{dist}(x, I_1\cap I_2)}{|I_1\cap I_2|}\Big)^{-100} dx\Big)^{\frac{1-s}{s}}$$

and everything continues as before, if $\tilde{M}$ is large enough. This completes the proof of $\mathbb{P}(0)$.

4.2. The proof of $\mathbb{P}(n-1)$ implies $\mathbb{P}(n)$. Recall that what we need to prove now is the estimate

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\mathbf{1}_E\Big\|_{L^{p,\infty}(L^Q)} \lesssim \Big\|\Big(\sum_{I\in\mathcal{I}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p(L^Q)}\cdot(\mathrm{size}_{\mathcal{I}}\,\mathbf{1}_E)^{1/p-\epsilon} \tag{48}$$

for every $0<p<\infty$ and Q, an n-tuple of positive real numbers, assuming that even the stronger version of it, namely (24), holds true for $(n-1)$-tuples Q. Again, here we are implicitly assuming that the proof of the strong $L^p(L^Q)$ estimate in (48) will follow by standard interpolation arguments, which we will describe later on, as promised.

Define $q_{j_0} := \min_{1\le j\le n} q_j$ and let s be any positive real number such that $s\le \min(1,p,q_{j_0})$. Then, one can dualize the weak-L^p quasi-norm on the left-hand side of (48) through L^s, as explained in [**3**]. As before, this amounts to proving that for every $F\subseteq\mathbb{R}$ measurable set with $|F|=1$, there exists a subset $\widetilde{F}\subseteq F$ with $|\widetilde{F}|>1/2$ such that

$$\Big\|\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_Q\Big\|_s \lesssim \|Sf\|_{L^p(L^Q)}\cdot(\mathrm{size}_{\mathcal{I}}\,\mathbf{1}_E)^{1/p-\epsilon}. \tag{49}$$

To construct $\widetilde{F}$, one first constructs an exceptional set Ω, as in the scalar case, with the only difference that the corresponding Ω_k is now given by

$$\Omega_k := \{x : \|Sf(x)\|_Q > C\,2^{10k/p}\|\|Sf\|_Q\|_p\}.$$

After that, exactly as before, one defines $\widetilde{F} := F\setminus\Omega$, which is clearly a major subset of F, in the sense that it has a comparable measure. Then, again one uses the decomposition (31) to reduce matters to proving the analogue of (33), which is now given by

$$\Big\|\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,k}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_Q\Big\|_s^s \lesssim 2^{Lk}\|\|Sf\|_Q\|_p^s\cdot(\mathrm{size}_{\mathcal{I}}\,\mathbf{1}_E)^{(1/p-\epsilon)s}. \tag{50}$$

Recall from [**3**] that $s\le\min(1,p,q_{j_0})$ implies that the expression on the left-hand side of (50) is now subadditive. As before, we will describe the proof of (50) in the main case $k=0$, the changes in the general case being similar to the ones in the scalar case. We therefore want to show that

$$\Big\|\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_Q\Big\|_s^s \lesssim \|\|Sf\|_Q\|_p^s\cdot(\mathrm{size}_{\mathcal{I}}\,\mathbf{1}_E)^{(1/p-\epsilon)s}. \tag{51}$$

To estimate the left-hand side of (51) we combine, as before, two stopping times. The first one selects iteratively maximal dyadic intervals I_0 for which one has

$$\frac{1}{|I_0|^{1/p}}\Big\|\Big\|\Big(\sum_{I\subseteq I_0}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_Q\Big\|_{L^p} > \frac{\widetilde{C}}{2^{n_1}}\,\|Sf\|_{L^p(L^Q)} \tag{52}$$

for various $n_1 \geq 0$, while the second is identical to the one used in the scalar case – see (40) and its natural generalizations. This allows us to estimate the left-hand side of (51) by

$$\sum_{n_1,n_2}\sum_{I_1\in\mathcal{I}_{n_1}^{(1)},I_2\in\mathcal{I}_{n_2}^{(2)}}\Big\|\Big\|\Big(\sum_{I\in\mathcal{I}_{n_1}^{(1)}(I_1)\cap\mathcal{I}_{n_2}^{(2)}(I_2)}\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_Q\Big\|_s^s. \tag{53}$$

Now fix I_1 and I_2 and consider the corresponding term on the right-hand side of (53). Given variables $(a_1,\dots,a_n)\in A$ denote $\widetilde{a}:=(a_2,\dots,a_n)$ and given $Q=(q_1,\dots,q_n)$ denote $\widetilde{Q}:=(q_2,\dots,q_n)$. Using these notations, the expression becomes

$$\begin{aligned}
&\int_{\mathbb{R}}\Big\|\Big(\sum_I\langle f,\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_{I_1\cap I_2}\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_Q^s(x)\,dx\\
&\qquad=\int_{\mathbb{R}}\Big(\int_{A_1}\Big\|\Big(\sum_I\langle f_{(a_1,\widetilde{a})},\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_{L_{\widetilde{a}}^{\widetilde{Q}}}^{q_1}(x)d\,a_1\Big)^{s/q_1}dx\\
&\qquad=\int_{\mathbb{R}}\Big(\int_{A_1}\Big\|\Big(\sum_I\langle f_{(a_1,\widetilde{a})},\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_{L_{\widetilde{a}}^{\widetilde{Q}}}^{q_1}(x)d\,a_1\Big)^{s/q_1}\\
&\qquad\qquad\times\mathbf{1}_{I_1\cap I_2}(x)\mathbf{1}_E(x)\mathbf{1}_{\widetilde{F}}(x)\,dx.
\end{aligned}$$

Since $s/q_1\leq 1$ one can apply Hölder and estimate the above expression by

$$\begin{aligned}
&\Big(\int_{\mathbb{R}}\int_{A_1}\Big\|\Big(\sum_I\langle f_{(a_1,\widetilde{a})},\phi_I^1\rangle\widetilde{\phi}_{I,0}^2\Big)\mathbf{1}_E\mathbf{1}_{\widetilde{F}}\Big\|_{L_{\widetilde{a}}^{\widetilde{Q}}}^{q_1}(x)d\,a_1\,dx\Big)^{s/q_1}\\
&\qquad\times|E\cap\widetilde{F}\cap I_1\cap I_2|^{(1-s/q_1)}
\end{aligned}\tag{54}$$

also using the fact that all the intervals I are now inside $I_1\cap I_2$. Then, one can use Fubini and integrate first with respect to the x variable in (54). This allows one to use the induction hypothesis locally (i.e. with respect to the collection $\mathcal{I}_{n_1}^{(1)}(I_1)\cap\mathcal{I}_{n_2}^{(2)}(I_2)$) in the case $p=q_1$, and estimate (54) by

$$\begin{aligned}
&\Big(\int_{\mathbb{R}}\int_{A_1}\|Sf(x)\|_{\widetilde{Q}}^{q_1}d\,a_1\,dx\Big)^{\frac{s}{q_1}}\cdot\big(\text{size}_{\mathcal{I}_{n_1}^{(1)}(I_1)\cap\mathcal{I}_{n_2}^{(2)}(I_2)}(\mathbf{1}_{E\cap\widetilde{F}})\big)^{(\frac{s}{q_1}-\epsilon)}\\
&\qquad\times|E\cap\widetilde{F}\cap I_1\cap I_2|^{(1-\frac{s}{q_1})}.
\end{aligned}\tag{55}$$

We emphasize that in (55)) the implicit sum in the definition of the square function $Sf(x)$ runs over the intervals I inside the local collection $\mathcal{I}_{n_1}^{(1)}(I_1) \cap \mathcal{I}_{n_2}^{(2)}(I_2)$.

It is then not difficult to see that the last expression in (55) can be rewritten and majorized by

$$(56) \quad \Big(\frac{1}{|I_1 \cap I_2|^{1/q_1}} \Big\| \Big\| \Big(\sum_{I \subseteq I_1 \cap I_2} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_Q \Big\|_{L^{q_1}} \Big)^s \times (\mathrm{size}_{\mathcal{I}_{n_1}^{(1)}(I_1) \cap \mathcal{I}_{n_2}^{(2)}(I_2)} (\mathbf{1}_{E \cap \widetilde{F}}))^{1-\epsilon} \cdot |I_1 \cap I_2|.$$

Using once again the John–Nirenberg inequality from [**19**] (which works equally well in our multiple vector-valued setting), we find that (56) is smaller than

$$(57) \quad \sup_{J \subseteq I_1 \cap I_2} \Big(\frac{1}{|J|^{1/p}} \Big\| \Big\| \Big(\sum_{I \subseteq J} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_Q \Big\|_{L^p} \Big)^s \times (\mathrm{size}_{\mathcal{I}_{n_1}^{(1)}(I_1) \cap \mathcal{I}_{n_2}^{(2)}(I_2)} (\mathbf{1}_{E \cap \widetilde{F}}))^{1-\epsilon} \cdot |I_1 \cap I_2|.$$

Using these, we can go back to (53) and majorize that expression by

$$\sum_{n_1, n_2 \geq 0} (2^{-n_1} \|Sf\|_{L^p(L^Q)})^s (2^{-n_2} \,\mathrm{size}_{\mathcal{I}} \, \mathbf{1}_{E \cap \widetilde{F}})^{(1-\epsilon)} \sum_{I_1 \in \mathcal{I}_{n_1}^{(1)}, I_2 \in \mathcal{I}_{n_2}^{(2)}} |I_1 \cap I_2|.$$

As before, one can estimate

$$\sum_{I_1 \in \mathcal{I}_{n_1}^{(1)}, I_2 \in \mathcal{I}_{n_2}^{(2)}} |I_1 \cap I_2|$$

in two distinct ways, by taking advantage of the stopping time decompositions performed earlier.

First, we can estimate it by $2^{n_1 p}$ and secondly, by $2^{n_2} (\mathrm{size}_{\mathcal{I}} \, \mathbf{1}_{E \cap \widetilde{F}})^{-1}$ given that $|\widetilde{F}| \sim 1$. In particular, this allows one to estimate the whole expression by

$$\|Sf\|_{L^p(L^Q)}^s (\mathrm{size}_{\mathcal{I}} \, \mathbf{1}_{E \cap \widetilde{F}})^{(1-\theta_2-\epsilon)} \sum_{n_1, n_2 \geq 0} 2^{-n_1(s - p\theta_1)} 2^{-n_2(1-\epsilon-\theta_2)}$$

as in the scalar case, for every $0 \leq \theta_1, \theta_2 \leq 1$ with $\theta_1 + \theta_2 = 1$. Then, if one chooses $\theta_1 < s/p$ but very close to it, this double sum becomes smaller than

$$\|Sf\|_{L^p(L^Q)}^s \cdot (\mathrm{size}_{\mathcal{I}} \, \mathbf{1}_E)^{\frac{s}{p} - \epsilon},$$

as desired. And this completes our proof.

The only thing left is the interpolation argument that we used implicitly several times.

4.3. Interpolation. Our interpolation result is somewhat unusual, in the sense that the collection $\mathcal{I}$ of dyadic intervals is as important as the operator it defines, the square function associated to it from (29). The result and its proof generalize straight away to collections of cubes in $\mathbb{R}^d$, and to arbitrary measures.

Proposition 4.1. *Consider $0 < p_1 < p < p_2 < \infty$ and let $\tilde{\mathcal{I}}$ be a collection of dyadic intervals. Assume that, for any subcollection $\mathcal{I} \subseteq \tilde{\mathcal{I}}$ of dyadic intervals and any L^Q-valued Schwartz function f on $\mathbb{R}$, we have for $j = 1, 2$*

$$\Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \mathbf{1}_E \Big\|_{L^{p_j,\infty}(L^Q)} \leq K_j \Big\| \Big(\sum_{I \in \mathcal{I}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_{L^{p_j}(L^Q)} \tag{58}$$

with the constants K_j independent of $\mathcal{I}$. Then for any $\mathcal{I} \subseteq \tilde{\mathcal{I}}$ we have the strong bound

$$\Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \mathbf{1}_E \Big\|_{L^{p}(L^Q)} \leq K \Big\| \Big(\sum_{I \in \mathcal{I}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{1/2} \Big\|_{L^{p}(L^Q)}, \tag{59}$$

where $K \lesssim (K_1^{p_1} + K_2^{p_2})^{\frac{1}{p}}$.

Observation 4.2. As mentioned before, the interpolation result in Proposition 4.1 can be stated in a more general setting, as the interested reader can verify. Our choice of presentation is motivated by the fact that in the present paper we need precisely the form presented above. The constants K_j in (58) do not depend on any of the subcollections $\mathcal{I}$ of intervals, but they could (and in most applications they do) depend on the general collection $\tilde{\mathcal{I}}$ and on the set E appearing on the left-hand side of (58); as a consequence, K is not dependent upon any of the subcollections $\mathcal{I}$, but could depend on $\tilde{\mathcal{I}}$ and on the set E.

We use the interpolation result above in order to deduce (24) from (26); notice that in that case

$$K_j = (\text{size}_{\tilde{\mathcal{I}}}\, \mathbf{1}_E)^{\frac{1}{p_j} - \epsilon} \quad \text{for } j = 1, 2.$$

Hence (24) follows immediately from (26) after interpolating carefully in a small neighborhood of the desired index $0 < p < \infty$.

On the other hand, in the proof of the interpolation result we will assume that E is the entire real line since it plays no role in the interpolation argument.

Proof of Proposition 4.1: Let $\mathcal{I} \subseteq \tilde{\mathcal{I}}$ be a subcollection of dyadic intervals, and denote by $F(x)$ the L^Q-valued function

$$F(x) := \sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2(x).$$

Our goal is to control $\|F\|_{L^p(L^Q)}$ by $\|S_{\mathcal{I}} f\|_{L^p(L^Q)}$, where $S_{\mathcal{I}} f$ is the associated square function:

$$S_{\mathcal{I}} f(x) := \Big(\sum_{I \in \mathcal{I}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I(x) \Big)^{1/2}.$$

The proof that we are about to provide will involve a partitioning of the collection $\mathcal{I}$ according to level sets of the "global" square function $S_{\mathcal{I}}$. First, for any $\alpha > 0$ and any $k \geq 0$, we define

$$S(k,\alpha) := \Big\{ x : \|S_{\mathcal{I}} f(x)\|_{L^Q} > \frac{\alpha}{C^k} \Big\}, \tag{60}$$

where C is a constant that will be determined later. Notice that the sets $S(k,\alpha)$ are nested:

$$S(0,\alpha) \subseteq S(1,\alpha) \subseteq \cdots S(k,\alpha) \subseteq \cdots$$

Each of the sets $S(k,\alpha)$ can be written as a disjoint union of maximal dyadic intervals:

$$S(k,\alpha) := \bigcup_{I_{\max}^k \in \mathcal{M}_k} I_{\max}^k, \quad \forall\, k \geq 0.$$

These will be used for the formerly mentioned partition:

- the collection $\mathcal{I}_0$ will consist of all intervals $I \in \mathcal{I}$ that are contained inside some maximal interval $I_{\max}^0$:

$$\mathcal{I}_0 := \{ I \in \mathcal{I} : \text{ there exists some } I_{\max}^0 \in \mathcal{M}_0 \text{ with } I \subseteq I_{\max}^0 \subset S(0,\alpha) \};$$

- for any $k \geq 1$, $\mathcal{I}_k$ is defined as

$$\mathcal{I}_k := \{ I \in \mathcal{I} : \exists I_{\max}^k \in \mathcal{M}_k \text{ with } I \subseteq I_{\max}^k \subset S(k,\alpha) \\ \text{and } I \nsubseteq I_{\max}^\ell \text{ for all } 0 \leq \ell < k \}.$$

  That is, $\mathcal{I}_k$ consists of all the intervals in $\mathcal{I}$ contained in some $I_{\max}^k \in \mathcal{M}_k$, which were not previously selected in any other $\mathcal{I}_\ell$ with $0 \leq \ell \leq k-1$.

Then we have $\mathcal{I} = \bigcup_{k\geq 0} \mathcal{I}_k$ and if F_k denotes the L^Q-valued function

$$F_k(x) := \sum_{I\in\mathcal{I}_k} \langle f, \phi_I^1\rangle \phi_I^2(x),$$

we have the decomposition $F(x) = \sum_{k\geq 0} F_k(x)$.

Notice that for all $k \geq 0$

$$\operatorname{supp}(S_{\mathcal{I}_k} f) \subseteq \bigcup_{I\in\mathcal{I}_k} I \subseteq \bigcup_{I^k_{\max}\in\mathcal{M}_k} I^k_{\max} = S(k,\alpha).$$

For any $k \geq 1$, if $I \in \mathcal{I}_k$ and $I^{k-1}_{\max} \in \mathcal{M}_{k-1}$ are such that $I \cap I^{k-1}_{\max} \neq \emptyset$, then necessarily $I \supsetneq I^{k-1}_{\max}$. Given the maximality condition based on which $I^{k-1}_{\max}$ was selected in $\mathcal{M}_{k-1}$, all intervals $I \in \mathcal{I}_k$ intersect $S(k-1,\alpha)^c$ and

$$\|S_{\mathcal{I}_k} f(x)\|_{L^Q} \leq \|S_{\mathcal{I}} f(x)\|_{L^Q} \cdot \mathbf{1}_{S(k-1,\alpha)^c}(x) \leq \frac{\alpha}{C^{k-1}}, \tag{61}$$

a feature that will be exploited later on.

Since Q is an arbitrary n-tuple of positive real numbers, there is no certainty that $\|\cdot\|_{L^Q}$ satisfies the triangle inequality; however, for s small enough (the condition that $s \leq \min(1, \min_{1\leq j\leq n} q_j)$ suffices), $\|\cdot\|^s_{L^Q}$ becomes subadditive. As a result,

$$\Big\{x : \Big\|\sum_{k\geq 0} F_k(x)\Big\|^s_{L^Q} > \alpha^s\Big\} \subseteq \bigcup_{k\geq 0}\Big\{x : \|F_k(x)\|^s_{L^Q} > \frac{\alpha^s}{2^{k+1}}\Big\}.$$

Moreover,

$$\Big|\Big\{x : \Big\|\sum_{k\geq 0} F_k(x)\Big\|_{L^Q} > \alpha\Big\}\Big| = \Big|\Big\{x : \Big\|\sum_{k\geq 0} F_k(x)\Big\|^s_{L^Q} > \alpha^s\Big\}\Big|$$
$$\leq \sum_{k\geq 0}\Big|\Big\{x : \|F_k(x)\|^s_{L^Q} > \frac{\alpha^s}{2^{k+1}}\Big\}\Big| = \sum_{k\geq 0}\Big|\Big\{x : \|F_k(x)\|_{L^Q} > \frac{\alpha}{2^{(k+1)/s}}\Big\}\Big|.$$

Such an inequality is important because it allows us to estimate $\|F\|_{L^p(L^Q)}$:

$$\begin{aligned}\|F\|^p_{L^p(L^Q)} &= p\int_0^\infty \alpha^{p-1}\Big|\Big\{x : \Big\|\sum_{k\geq 0} F_k(x)\Big\|_{L^Q} > \alpha\Big\}\Big|\, d\alpha \\ &\leq \sum_{k\geq 0} p\int_0^\infty \alpha^{p-1}\Big|\Big\{x : \|F_k(x)\|_{L^Q} > \frac{\alpha}{2^{(k+1)/s}}\Big\}\Big|\, d\alpha.\end{aligned} \tag{62}$$

We note that the functions F_k above depend in fact on the variable α (the collections of intervals $\mathcal{I}_k$ are determined by the level sets $S(k,\alpha)$);

this is the main difficulty in proving the interpolation result, and which differentiates Proposition 4.1 from standard interpolation results.

First, we deal with the case corresponding to $k = 0$, by invoking the weak-type hypothesis (58) for the collection $\mathcal{I}_0$:

$$\int_0^\infty \alpha^{p-1}\Big|\Big\{x : \|F_0(x)\|_{L^Q} > \frac{\alpha}{2^{1/s}}\Big\}\Big|\,d\alpha$$

$$\leq \int_0^\infty \alpha^{p-1}\Big(\frac{\alpha}{2^{1/s}}\Big)^{-p_1} K_1^{p_1}\Big\|\Big(\sum_{I\in\mathcal{I}_0}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac{1}{2}}\Big\|_{L^{p_1}(L^Q)}^{p_1}\,d\alpha$$

$$\leq 2^{p_1/s}K_1^{p_1}\int_0^\infty \alpha^{p-p_1-1}\int_{S(0,\alpha)}\|S_{\mathcal{I}_0}f(x)\|_{L^Q}^{p_1}\,dx\,d\alpha.$$

The term above can be bounded by an expression involving only the "global" square function $S_{\mathcal{I}}$, which depends on neither the subcollection $\mathcal{I}_0$ nor on α, given by

$$2^{p_1/s}K_1^{p_1}\int_0^\infty \alpha^{p-p_1-1}\int_{\{\|S_{\mathcal{I}}f\|_{L^Q}>\alpha\}}\|S_{\mathcal{I}}f(x)\|_{L^Q}^{p_1}\,dx\,d\alpha.$$

Now we apply the usual trick which consists in changing the order of integration, obtaining in this way

$$\int_0^\infty \alpha^{p-1}\Big|\Big\{x :\|F_0(x)\|_{L^Q} > \frac{\alpha}{2^{1/s}}\Big\}\Big|\,d\alpha$$

$$\leq 2^{\frac{p_1}{s}}K_1^{p_1}\int_{\mathbb{R}}\|S_{\mathcal{I}}f(x)\|_{L^Q}^{p_1}\int_0^{\|S_{\mathcal{I}}f(x)\|_{L^Q}}\alpha^{p-p_1-1}\,d\alpha\,dx$$

$$= \frac{2^{\frac{p_1}{s}}K_1^{p_1}}{p-p_1}\|S_{\mathcal{I}}f\|_{L^p(L^Q)}^p.$$

Next we deal with a generic term involving F_k for some $k \geq 1$; we use the assumption (58) applied to the collection $\mathcal{I}_k$:

$$\int_0^\infty \alpha^{p-1}\Big|\Big\{x : \|F_k(x)\|_{L^Q} > \frac{\alpha}{2^{(k+1)/s}}\Big\}\Big|\,d\alpha$$

$$\leq \int_0^\infty \alpha^{p-1}\Big(\frac{\alpha}{2^{(k+1)/s}}\Big)^{-p_2} K_2^{p_2}\Big\|\Big(\sum_{I\in\mathcal{I}_k}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac{1}{2}}\Big\|_{L^{p_2}(L^Q)}^{p_2}\,d\alpha$$

$$= 2^{p_2\frac{k+1}{s}}K_2^{p_2}\int_0^\infty \alpha^{p-p_2-1}\int_{S(k,\alpha)}\|S_{\mathcal{I}_k}f(x)\|_{L^Q}^{p_2}\,dx\,d\alpha.$$

Changing the order of integration will not be helpful in this case because the collections on intervals $\mathcal{I}_0, \mathcal{I}_1, \ldots$ depend on the variable α, and lower[4] bounds for $S_{\mathcal{I}_k}$ independent of α are not available. Instead, we use the pointwise inequality $\|S_{\mathcal{I}_k} f(x)\|_{L^Q} \leq \frac{\alpha}{C^{k-1}}$ from (61). Recalling also the definition of $S(k,\alpha)$, we have

$$\begin{aligned}
&2^{p_2 \frac{k+1}{s}} K_2^{p_2} \int_0^\infty \alpha^{p-p_2-1} \int_{S(k,\alpha)} \|S_{\mathcal{I}_k} f(x)\|_{L^Q}^{p_2}\, dx\, d\alpha \\
&\quad\leq 2^{p_2 \frac{k+1}{s}} K_2^{p_2} \int_0^\infty \alpha^{p-p_2-1} |S(k,\alpha)| \Big(\frac{\alpha}{C^{k-1}}\Big)^{p_2} d\alpha \\
&\quad\leq 2^{p_2 \frac{k+1}{s}} K_2^{p_2} C^{-(k-1)p_2} \int_0^\infty \alpha^{p-1} \Big|\Big\{x : \|S_{\mathcal{I}} f(x)\|_{L^Q} > \frac{\alpha}{C^k}\Big\}\Big|\, d\alpha.
\end{aligned}$$

Making a change of variable we obtain

$$\begin{aligned}
&\int_0^\infty \alpha^{p-1} \Big|\Big\{x : \|F_k(x)\|_{L^Q} > \frac{\alpha}{2^{(k+1)/s}}\Big\}\Big|\, d\alpha \\
&\quad\leq 2^{p_2 \frac{k+1}{s}} K_2^{p_2} C^{-(k-1)p_2} C^{kp} \int_0^\infty \lambda^{p-1} \Big|\{x : \|S_{\mathcal{I}} f(x)\|_{L^Q} > \lambda\}\Big|\, d\lambda \\
&\quad\leq \frac{1}{p}\, 2^{p_2 \frac{k+1}{s}} C^{p_2} K_2^{p_2} C^{-k(p_2-p)} \|S_{\mathcal{I}} f\|_{L^p(L^Q)}^p.
\end{aligned}$$

Now it remains to put everything together and to sum in $k \geq 0$: due to (62),

$$\begin{aligned}
&\Big\|\sum_{I\in\mathcal{I}} \langle f, \phi_I^1\rangle \phi_I^2\Big\|_{L^p(L^Q)}^p \\
&\leq \Big(\frac{p}{p-p_1} 2^{\frac{p_1}{s}} K_1^{p_1} + 2^{p_2/s} K_2^{p_2} C^{p_2} \sum_{k\geq 1} \big(2^{p_2/s} C^{-(p_2-p)}\big)^k\Big) \|S_{\mathcal{I}} f\|_{L^p(L^Q)}^p.
\end{aligned}$$

Since $p_2 - p > 0$, if C is large enough so that $2^{p_2/s} C^{-(p_2-p)} < 1$ (which is equivalent to $C > 2^{p_2/s(p_2-p)}$), the series above is finite. We obtain in this way (59) with

$$K^p \lesssim_{p_1,p_2,s} K_1^{p_1} + K_2^{p_2}. \qquad \square$$

Observation 4.3. In the statement of Proposition 4.1, we could allow K_1 and K_2 to depend on the collection $\mathcal{I}$, which will yield an upper bound for K that also depends on $\mathcal{I}$. Thus, assuming that

$$\Big\|\Big(\sum_{I\in\mathcal{I}} \langle f, \phi_I^1\rangle \phi_I^2\Big) \mathbf{1}_E\Big\|_{L^{p_j,\infty}(L^Q)} \leq K_j(\mathcal{I}) \Big\|\Big(\sum_{I\in\mathcal{I}} \frac{|\langle f, \phi_I^1\rangle|^2}{|I|} \mathbf{1}_I\Big)^{1/2}\Big\|_{L^{p_j}(L^Q)} \tag{63}$$

[4]Classically the strong-type estimates are obtained from the weak-type ones via an identity similar to (62) by splitting the function into two pieces, one where the function is small and another one where the function is large (relative to the parameter α).

holds for all collections $\mathcal{I}$ of dyadic intervals, for $0 < p_1 < p < p_2 < \infty$, we deduce the strong bound

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\mathbf{1}_E\Big\|_{L^p(L^Q)} \leq K(\mathcal{I})\Big\|\Big(\sum_{I\in\mathcal{I}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\mathbf{1}_I\Big)^{1/2}\Big\|_{L^p(L^Q)}, \tag{64}$$

where this time $K(\mathcal{I}) \lesssim (\sup_{\mathcal{I}'\subseteq\mathcal{I}} K_1(\mathcal{I}')^{p_1} + \sup_{\mathcal{I}'\subseteq\mathcal{I}} K_2(\mathcal{I}')^{p_2})^{\frac{1}{p}}$.

5. Proof of Theorem 1.1 in the general case

Recall that our goal is to prove that

$$\|f\|_{L^P(L^Q)} \lesssim \|Sf\|_{L^P(L^Q)}, \tag{65}$$

where the d-tuple $P = (p_1,\dots,p_d)$ and the n-tuple $Q = (q_1,\dots,q_n)$ satisfy $0 < P, Q < \infty$ componentwise. Recall also that the N-parameter square function S is defined by

$$S := S_{d_1}\otimes\cdots\otimes S_{d_N}$$

while $d_1+\cdots+d_N = d$. So far we have proved this in the particular situation when $d_1=\cdots=d_N=1$. The goal of this section is to explain that similar ideas can handle the general case as well. First of all, let us observe that using a similar inductive argument to that in Section 2, it is enough to prove the particular case when $N=1$. In other words, from now on, our square function Sf is a one-parameter square function in $\mathbb{R}^d$ and the task is to prove multiple vector-valued, mixed-norm estimates for it, in the form of

$$\|f\|_{L^P(L^Q)} \lesssim \|S_d f\|_{L^P(L^Q)}. \tag{66}$$

It is now important to observe that when $p_1=\cdots=p_d=p$, then (66) becomes a multiple vector-valued $L^p(\mathbb{R}^d)$ estimate, which can be proved exactly as in the one-dimensional case $d=1$ treated before. This is because all of our previous arguments have natural higher-dimensional analogues. Instead of doing analysis with dyadic intervals, one does analysis with dyadic cubes of the corresponding dimension, in precisely the same way.

It will be more convenient to modify the notation a bit, in order to obtain a statement more suitable for the upcoming inductive argument. We will think of the Euclidean space $\mathbb{R}^d$ as being decomposed into

$$\mathbb{R}^d = \mathbb{R}^{n_1}\times\cdots\times\mathbb{R}^{n_\mathfrak{m}} \tag{67}$$

and consequently the mixed-norm space $L^P(\mathbb{R}^d)$ being unfolded as

$$L^P(\mathbb{R}^d) = L^{p_1}(\mathbb{R}^{n_1})(L^{p_2}(\mathbb{R}^{n_2})(\dots(L^{p_\mathfrak{m}}(\mathbb{R}^{n_\mathfrak{m}}))\dots)). \tag{68}$$

In other words, we implicitly assume that the first n_1 indices of the d-tuple P are all equal to p_1, the next n_2 indices of P are all equal to p_2, and so on, until the last $n_{\mathfrak{m}}$ set of indices of P, which are all equal to $p_{\mathfrak{m}}$.

The plan is to prove the corresponding (66) by induction with respect to the parameter $\mathfrak{m}$. As pointed out before, (66) is already known when $\mathfrak{m} = 1$ and we aim to show that it is also true for $\mathfrak{m} = d$, when all the entries of P are possibly different from each other.

As in the one-dimensional case, it is not difficult to see that things can be reduced to proving a discrete analogue of (66), namely

$$\Big\| \sum_{R \in \mathcal{R}} \langle f, \phi_R^1 \rangle \phi_R^2 \Big\|_{L^P(L^Q)} \lesssim \Big\| \Big(\sum_{R \in \mathcal{R}} \frac{|\langle f, \phi_R^1 \rangle|^2}{|R|} \mathbf{1}_R \Big)^{1/2} \Big\|_{L^P(L^Q)}. \tag{69}$$

This is because the methods in Section 3 have natural equivalents in higher dimensions. The families $(\phi_R^1)_R$ and $(\phi_R^2)_R$ in (69) are two lacunary families, L^2 normalized, indexed by a finite collection $\mathcal{R}$ of dyadic cubes in $\mathbb{R}^d$. And also as in the one-dimensional case, the statement of Observation 3.2 remains valid, in the sense that the two families of functions may depend on the implicit variables $(a_1, \ldots, a_n)$ of the space L^Q.

Using a higher-dimensional analogue of (31) we decompose each ϕ_R^2 as

$$\phi_R^2 = \sum_{\ell=0}^{\infty} 2^{-\#\ell} \phi_{R,\ell}^2 =: \sum_{\ell=0}^{\infty} 2^{-(\frac{\#}{2})\ell} \widetilde{\phi}_{R,\ell}^2, \tag{70}$$

where

$$\operatorname{supp}(\phi_{R,\ell}^2) \subseteq 2^\ell R$$

as before and where $\#$ is arbitrarily large. Using this in (69), it will be enough to show

$$\Big\| \sum_{R \in \mathcal{R}} \langle f, \phi_R^1 \rangle \widetilde{\phi}_{R,\ell}^2 \Big\|_{L^P(L^Q)} \lesssim 2^{L\ell} \Big\| \Big(\sum_{R \in \mathcal{R}} \frac{|\langle f, \phi_R^1 \rangle|^2}{|R|} \mathbf{1}_R \Big)^{1/2} \Big\|_{L^P(L^Q)} \tag{71}$$

for some large but fixed number L. The main case is when $\ell = 0$ and we will concentrate on it from now on (by this we mean that the general case follows by standard modifications as in the one-dimensional situation – see the proof of Theorem 3.1). Then (71) reads as

$$\Big\| \sum_{R \in \mathcal{R}} \langle f, \phi_R^1 \rangle \widetilde{\phi}_{R,0}^2 \Big\|_{L^P(L^Q)} \lesssim \Big\| \Big(\sum_{R \in \mathcal{R}} \frac{|\langle f, \phi_R^1 \rangle|^2}{|R|} \mathbf{1}_R \Big)^{1/2} \Big\|_{L^P(L^Q)}. \tag{72}$$

We think of the dyadic cubes R as being of the form

$$R = R_1 \times \cdots \times R_{\mathfrak{m}}$$

to match the decomposition (67), where each R_j is a dyadic cube in $\mathbb{R}^{n_j}$ of the same side length as R itself for $1 \leq j \leq \mathfrak{m}$.

Following the same earlier strategy for the estimate (72), one needs in fact to prove a more localized variant of it given by

$$(73)\quad \Big\| \sum_{R\in\mathcal{R}} \langle f, \phi_R^1\rangle \widetilde{\phi}^2_{R,0} \mathbf{1}_E \Big\|_{L^P(L^Q)} \lesssim \Big\| \Big(\sum_{R\in\mathcal{R}} \frac{|\langle f,\phi_R^1\rangle|^2}{|R|} \mathbf{1}_R \Big)^{1/2} \Big\|_{L^P(L^Q)} \cdot (\mathrm{size}_{\mathcal{R}_1}\, \mathbf{1}_E)^{1/p_1-\epsilon},$$

where

$$\mathcal{R}_1 := \{R_1 : R = R_1\times\cdots\times R_{\mathfrak{m}} \in \mathcal{R}\}$$

and $\mathrm{size}_{\mathcal{R}_1}\, \mathbf{1}_E$ is the corresponding n_1-dimensional *size* generalizing naturally the one-dimensional (25). In (73) the set E is an arbitrary measurable subset of $\mathbb{R}^{n_1}$.

The plan is to prove (73) by induction with respect to the parameter $\mathfrak{m}$. Notice that when $\mathfrak{m} = 1$, then $\mathcal{R}_1 = \mathcal{R}$ and the corresponding (73) is known, as we pointed out before (its proof is identical to the one in the one-dimensional case). In particular, all one has to do is to prove that the case $\mathfrak{m}-1$ implies the case $\mathfrak{m}$, for every $\mathfrak{m}\geq 2$. We claim that this can be done by an argument similar to the one used earlier in the proof of "$\mathbb{P}(n-1)$ implies $\mathbb{P}(n)$" (see Subsection 4.2).

First of all, we like to see the left-hand side of (73) as being

$$(74)\quad \Big\| \sum_{R\in\mathcal{R}} \langle f, \phi_R^1\rangle \widetilde{\phi}^2_{R,0} \mathbf{1}_E \Big\|_{L^{p_1}(L^{\tilde{P}}(L^Q))},$$

where for $P = (p_1,\dots,p_{\mathfrak{m}})$ we define $\widetilde{P} := (p_2,\dots,p_{\mathfrak{m}})$. As before, by interpolation it would be enough to estimate the weaker analogue of it, namely

$$(75)\quad \Big\| \sum_{R\in\mathcal{R}} \langle f, \phi_R^1\rangle \widetilde{\phi}^2_{R,0} \mathbf{1}_E \Big\|_{L^{p_1,\infty}(L^{\tilde{P}}(L^Q))}$$

by the same right-hand side of (73). As explained previously, we dualize the $L^{p_1,\infty}$ quasi-norm through L^s, where s is a positive real number smaller than all the entries of P, of Q, and also smaller than 1. By scale invariance (in the ambient space $\mathbb{R}^d$) this amounts to proving that for every subset $F\subset\mathbb{R}^{n_1}$ with $|F|=1$ there exists a major subset $\widetilde{F}\subseteq F$ with $|\widetilde{F}|\geq 1/2$ such that

$$(76)\quad \Big\| \sum_{R\in\mathcal{R}} \langle f, \phi_R^1\rangle \widetilde{\phi}^2_{R,0} \mathbf{1}_E \mathbf{1}_{\widetilde{F}} \Big\|_{L^s(L^{\tilde{P}}(L^Q))} \lesssim RHS(73).$$

The subset $\widetilde{F}$ is defined as usual by $\widetilde{F} := F \setminus \Omega$ for a certain exceptional set $\Omega \in \mathbb{R}^{n_1}$. This exceptional set is constructed as before with the only difference that the corresponding Ω_k are now given by

$$\Omega_k := \{x_1 \in \mathbb{R}^{n_1} : \|Sf(x_1)\|_{L^{\tilde{P}}(L^Q)} > C 2^{10k/p_1} \|Sf\|_{L^P(L^Q)}\}. \tag{77}$$

In the above (77), by Sf one denotes the discrete square function given by the inner expression in the right-hand side of (69). Also, we now think of a generic variable in $\mathbb{R}^d$ as being of the form $(x_1, \dots, x_{\mathfrak{m}})$ with $x_j \in \mathbb{R}^{n_j}$ for $1 \leq j \leq \mathfrak{m}$. In particular, $Sf(x_1)$ can be thought of as a function depending on the rest of the variables $(x_2, \dots, x_{\mathfrak{m}})$ in an obvious way:

$$Sf(x_1)(x_2, \dots, x_{\mathfrak{m}}) := Sf(x_1, x_2, \dots, x_{\mathfrak{m}}).$$

To estimate (76) one needs to perform (again) two carefully designed stopping times. The second one involves *averages* over dyadic cubes, and it is essentially a higher-dimensional analogue of the one before. The first one, on the other hand, selects maximal dyadic cubes R_1^0 in $\mathbb{R}^{n_1}$ for which the corresponding averages

$$\frac{1}{|R_1^0|^{1/p_1}} \Big\| \Big(\sum_{R \in \mathcal{R} : R_1 \subseteq R_1^0} \frac{|\langle f, \phi_R^1 \rangle|^2}{|R|} \mathbf{1}_R \Big)^{1/2} \Big\|_{L^{p_1}(L^{\tilde{P}}(L^Q))} \tag{78}$$

are large (larger than $C\, 2^{-\kappa_1} \|Sf\|_{L^P(L^Q)}$, for κ_1, some positive integer), also as in the one-dimensional multiple vector-valued case. The way one uses these two together is similar to the way explained in the earlier "$\mathbb{P}(n-1)$ implies $\mathbb{P}(n)$" situation. At some point, exactly as before, one uses Hölder locally, to be able to rely on the induction hypothesis (as in the previous (54)) in the particular case when $p_1 = p_2$. More precisely, this amounts to estimating expressions of the type

$$\Big\| \sum_R \langle f, \phi_R^1 \rangle \widetilde{\phi}_{R,0}^2 \mathbf{1}_E \Big\|_{L^{p_2}(L^{\tilde{P}}(L^Q))}$$

locally, and here the induction hypothesis can be applied since the new P tuple is now $P = (p_2, p_2, \dots, p_{\mathfrak{m}})$, and in particular, one can think of $\mathbb{R}^d$ as being split as $\mathbb{R}^d = \mathbb{R}^{n_1+n_2} \times \dots \times \mathbb{R}^{n_{\mathfrak{m}}}$ and this now contains only $\mathfrak{m} - 1$ factors. There are only two observations that one needs to make in order to realize that the earlier argument goes through smoothly in our case as well.

The first is that the John–Nirenberg inequality is still available in this context. More explicitly, this means that the supremum over R_1^0 of averages of the type

$$\frac{1}{|R_1^0|^{1/p_2}} \Big\| \Big(\sum_{R \in \mathcal{R} : R_1 \subseteq R_1^0} \frac{|\langle f, \phi_R^1 \rangle|^2}{|R|} \mathbf{1}_R \Big)^{1/2} \Big\|_{L^{p_2}(L^{\tilde{P}}(L^Q))},$$

which appear naturally after one applies the induction, is controlled by the corresponding supremum of averages of the type

$$\frac{1}{|R_1^0|^{1/p_1}}\Big\|\Big(\sum_{R\in\mathcal{R}:R_1\subseteq R_1^0}\frac{|\langle f,\phi_R^1\rangle|^2}{|R|}\mathbf{1}_R\Big)^{1/2}\Big\|_{L^{p_1}(L^{\tilde{P}}(L^Q))},$$

which are the ones needed to capitalize on the stopping time procedure. To prove this, one just has to observe that the above inner expressions can also be seen as

$$\sum_{R\in\mathcal{R}:R_1\subseteq R_1^0}\frac{|\langle f,\phi_R^1\rangle|^2}{|R|}\mathbf{1}_R(x_1,x_2,\dots,x_{\mathfrak{m}})=\sum_{R_1\subseteq R_1^0}\frac{|a_{R_1}(x_2,\dots,x_{\mathfrak{m}})|^2}{|R_1|}\mathbf{1}_{R_1}(x_1),$$

where in general

$$a_C(x_2,\dots,x_{\mathfrak{m}}):=\Big(\sum_{R:R_1=C}\frac{|\langle f,\phi_R^1\rangle|^2}{|R_2\times\cdots\times R_{\mathfrak{m}}|}\mathbf{1}_{R_2\times\cdots\times R_{\mathfrak{m}}}(x_2,\dots,x_{\mathfrak{m}})\Big)^{\frac12},$$

and after that to realize that *BMO* expressions of the type

$$\sup_{C_0}\frac{1}{|C_0|^{1/q}}\Big\|\Big(\sum_{C\subseteq C_0}\frac{|a_C|^2}{|C|}\mathbf{1}_C\Big)^{\frac12}\Big\|_{L^q(B)} \tag{79}$$

are all equivalent to each other for every $0<q<\infty$ even when B is a quasi-Banach lattice.

And the second observation is that

$$\text{size}_{\mathcal{R}_{1\times 2}}\,\mathbf{1}_E\lesssim\text{size}_{\mathcal{R}_1}\,\mathbf{1}_E,$$

as one can easily check. By $\mathcal{R}_{1\times 2}$ one means

$$\mathcal{R}_{1\times 2}:=\{R_1\times R_2:R=(R_1,R_2,\dots,R_{\mathfrak{m}})\in\mathbb{R}\}$$

and they appear naturally after the application of the induction hypothesis in $\mathbb{R}^d=\mathbb{R}^{n_1+n_2}\times\cdots\times\mathbb{R}^{n_{\mathfrak{m}}}$. This concludes our proof of the weaker estimate (76).

After that the induction argument works exactly as before, allowing one to complete the proof of the desired discrete estimate (73).

6. Connections to weighted theory and extrapolation

In the present section we discuss a certain weighted version of inequality (6), which eventually yields an alternative proof of Theorem 1.1, upon adapting existing extrapolation results. Assuming such a weighted estimate, in Subsection 6.1, we detail this proof by extrapolation. In the second part, Subsection 6.2, we review the weighted estimates (which are indispensable for extrapolation) and provide a proof for them based on a sparse domination result implied by the helicoidal method.

Although the weighted estimates (either in the one-parameter or the multi-parameter setting) are not difficult to deduce in all their generality from certain particular cases via extrapolation, they do not seem to have appeared previously in the literature. For this reason, we will provide a self-contained proof of the weighted estimates, which relies on localization and stopping times.

A weighted, scalar version of (6) can be formulated in the following way: if f is a Schwartz function and w is "regular enough",

$$\|f\|_{L^p(w)} \lesssim \|Sf\|_{L^p(w)}. \tag{80}$$

For $0 < p \leq 1$, this inequality is related to the theory of weighted Hardy spaces and it was stated in [**9**]. There, the authors study the boundedness of singular integrals on such spaces, which was known previously under more stringent conditions on the weights (they were assumed to be A_1 weights). In [**9**], a theory of weighted Hardy spaces and boundedness of singular integrals is developed for A_∞ weights. Central to their theory is the inequality (80), which is stated for A_∞ weights. Starting from this and using a certain type of extrapolation (regarding collections of pairs of functions, rather than operators, and A_∞ weights), we recover the multiple vector-valued results of Theorem 1.1; the mixed-normed estimates are obtained through a generalization of a result of Kurtz [**15**].

On the other hand, we will see once again that a local estimate similar to (24) and a change in the direction of the stopping time will yield a (multiple vector-valued) sparse[5] estimate, and in consequence, also (multiple vector-valued) weighted estimates, in the one-parameter case. The weighted estimates obtained in this way are similar to (80) and to those of [**9**], and hence they are interconnected to weighted Hardy spaces.

Before proceeding, we briefly recall a few definitions and results about weights. For $1 < p < \infty$, the classes $A_p(\mathbb{R}^m)$ consist of measurable functions $w\colon \mathbb{R}^m \to [0, \infty]$ for which

$$[w]_{A_p} := \sup_{\substack{Q \subset \mathbb{R}^m \\ Q \text{ cube}}} \Big(\fint_Q w(x)\, dx \Big) \Big(\fint_Q w^{1-p'}(x)\, dx \Big)^{p-1} < +\infty.$$

If $p = 1$, then $w \in A_1(\mathbb{R}^m)$ provided there exists a constant C such that $Mw(x) \leq C\, w(x)$ for almost every $x \in \mathbb{R}^m$. Then $A_\infty(\mathbb{R}^m)$ is defined as

$$A_\infty(\mathbb{R}^m) := \bigcup_{1 \leq p < \infty} A_p(\mathbb{R}^m).$$

[5]While this is defined more precisely in Definition 6.7, one should think of a sparse estimate as one in which the information is concentrated on a "thin" collection.

For the classes $A_{p,\mathrm{Rectangle}}(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N})$, the collection of cubes is replaced by the collection of rectangles with sides parallel to the coordinate axes, and in the case $p = 1$, the Hardy–Littlewood maximal function is replaced by the *strong maximal function* M_S. For $p > 1$, it is well known that $w(x_1, \ldots, x_N) \in A_{p,\mathrm{Rectangle}}(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N})$ if and only if

$$w(\cdot, x_2, \ldots, x_N) \in A_p(\mathbb{R}^{d_1}), \ldots, w(x_1, \ldots, x_{N-1}, \cdot) \in A_p(\mathbb{R}^{d_N})$$

uniformly with respect to the fixed variables.

6.1. Weighted Hardy spaces and extrapolation. Let $0 < p < \infty$. If $w \in A_\infty(\mathbb{R}^m)$, then the *weighted Hardy space* H^p_w consists of

$$H^p_w := \{f : \mathbb{R}^m \to \mathbb{C} : S_m f \in L^p_w(\mathbb{R}^m)\}. \tag{81}$$

Setting $\|f\|_{H^p_w} := \|S_m f\|_{L^p_w}$, H^p_w becomes a quasi-Banach space, for which we have, whenever $s \le \min(p, 1)$,

$$\|f + g\|^s_{H^p_w} \le \|f\|^s_{H^p_w} + \|g\|^s_{H^p_w}.$$

By making use of a certain discrete Calderón reproducing formula, it was shown in [**9**, Theorem 3.5] that, for any $w \in A_\infty(\mathbb{R}^m)$ and any $0 < p \le 1$,

$$\|f\|_{L^p_w(\mathbb{R}^m)} \le C\|f\|_{H^p_w(\mathbb{R}^m)} = C\,\|S_m f\|_{L^p_w(\mathbb{R}^m)}. \tag{82}$$

The method of the proof does not immediately generalize to the case $p > 1$. Instead, in this situation the $L^p_w(\mathbb{R}^m)$ boundedness (which requires the stronger condition that $w \in A_p$) of the square function S_m is invoked to deduce, by means of duality, an estimate similar to (82). Hence, for $p > 1$, Ding et al. [**9**] state the inequality (82) only for weights $w \in A_p$.

Alternatively, one can use the A_∞ extrapolation developed in [**7**] (similarly, see [**8**, Corollary 3.15]) applied to the pairs of functions $(f, S_m(f))$. This will imply that (82) is valid for any $0 < p < \infty$, and for any $w \in A_\infty(\mathbb{R}^m)$. The same extrapolation result yields multiple vector-valued weighted inequalities: for any $0 < p < \infty$, any n-tuple Q, and any weight $w \in A_\infty(\mathbb{R}^m)$,

$$\|f\|_{L^p(L^Q)(w)} \le C\|S_m(f)\|_{L^p(L^Q)(w)}. \tag{83}$$

Theorem 3.5 in [**9**] remains valid in the context of multi-parameter Hardy spaces, and Theorem 2.1 in [**7**] holds for weights associated to Muckenhoupt bases. This simple observation extends the inequality in [**9**] and as a result, the multi-parameter multiple vector-valued inequality holds:

$$\|f\|_{L^p(L^Q)(w)} \le C\|S_{d_1} \otimes \cdots \otimes S_{d_N}(f)\|_{L^p(L^Q)(w)}, \tag{84}$$

where $0 < p < \infty$, $Q = (q_1, \dots, q_n)$ with $0 < q_j < \infty$ for all $1 \leq j \leq n$, and $w \in A_{\infty,\text{Rectangle}}(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N})$.

In order to obtain the full mixed-norm estimates of Theorem 1.1, we need an extrapolation result from [**15**] suited for mixed-norm spaces. The result extends without any important modification to pairs of functions, in which case the operator T is being disregarded. Once inequality (84) is deduced as above, the plan is to apply it to product weights and deduce the mixed-norm estimates from Theorem 6.1 below.

We recall the following reformulation of Kurtz's result, in a slightly more general setting, although the proof remains the same:

Theorem 6.1 (Similar to Theorem 2 of [**15**]). *Let $0 < s_0 < \infty$ and assume that there exists $s_0 < s < \infty$ such that*

$$\int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} |f(x,y)|^s w(x,y)\, dy\, dx \leq C \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} |g(x,y)|^s w(x,y)\, dy\, dx \tag{85}$$

for all pairs (f, g) in a certain collection of functions $\mathcal{F}$, and for all $w \in A_{\frac{s}{s_0},\text{Rectangle}}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$, with the constant C depending only on $[w]_{A_{\frac{s}{s_0},\text{Rectangle}}}$. Then for any $s_0 < p, q < \infty$, and any weights $w(x, y)$ of the type $w(x, y) = u(x)\, v(y)$ such that

$$u^{\frac{p}{q}} \in A_{\frac{p}{s_0}}(\mathbb{R}^{d_1}), \quad v \in A_{\frac{q}{s_0}}(\mathbb{R}^{d_2}),$$

we have for any pair $(f, g) \in \mathcal{F}$:

$$\int_{\mathbb{R}^{d_1}} \Big(\int_{\mathbb{R}^{d_2}} |f(x,y)|^q w(x,y)\, dy \Big)^{\frac{p}{q}} dx \\ \leq C([u]_{A_{\frac{p}{s_0}}}, [v]_{A_{\frac{q}{s_0}}}) \int_{\mathbb{R}^{d_1}} \Big(\int_{\mathbb{R}^{d_2}} |g(x,y)|^q w(x,y)\, dy \Big)^{\frac{p}{q}} dx.$$

In particular, if $w(x, y) \equiv 1$, mixed-norm estimates are implied by extrapolation, once the weighted result (85) *is known.*

Remark 6.2. In [**15**], one is in fact looking for a necessary and sufficient condition on weights $w(x, y)$ so that the strong maximal function M_S satisfies

$$\int_{\mathbb{R}^{d_1}} \Big(\int_{\mathbb{R}^{d_2}} |M_S f(x,y)|^q w(x,y)\, dy \Big)^{\frac{p}{q}} dx \\ \leq C \int_{\mathbb{R}^{d_1}} \Big(\int_{\mathbb{R}^{d_2}} |f(x,y)|^q w(x,y)\, dy \Big)^{\frac{p}{q}} dx.$$

While a necessary condition was found (the classes $A_p(A_q)$ from [**15**, Definition 2]), sufficiency is proved only in the particular case of product weights $w(x, y) = u(x)\, v(y)$. Since we are mainly interested in the unweighted, multiple vector-valued case, we do not elaborate on the properties of the classes of weights $A_p(A_q)$, but instead focus on the extrapolation result, which is also known to be true only for product weights.

We also disregard how the constants appearing in the inequalities above depend on the weights involved or on their characteristics.

Next, we generalize Theorem 6.1 to mixed-norm L^p spaces involving κ variables (with $\kappa \geq 2$), and A_∞ weights.

Theorem 6.3. *Assume there exists some $0 < s < \infty$ such that*

$$\begin{aligned}(86)\quad &\int_{\mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_\kappa}} |f(x_1,\dots,x_\kappa)|^s w(x_1,\dots,x_\kappa)\,dx_1\dots dx_\kappa \\ &\qquad\leq C\int_{\mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_\kappa}} |g(x_1,\dots,x_\kappa)|^s w(x_1,\dots,x_\kappa)\,dx_1\dots dx_\kappa\end{aligned}$$

for all $w \in A_{\infty,\mathrm{Rectangle}}(\mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_\kappa})$ and for all pairs of functions (f,g) belonging to a certain collection $\mathcal{F}$. Then for any $0 < p_1,\dots,p_\kappa < \infty$ and for any weight $w(x_1,\dots,x_\kappa) = w_1(x_1)\cdot\ldots\cdot w_\kappa(x_\kappa)$ such that $w_l^{\frac{p_l}{p_\kappa}} \in A_\infty(\mathbb{R}^{d_l})$ for all $1 \leq l \leq \kappa$ and all $(f,g) \in \mathcal{F}$, we have

$$\begin{aligned}(87)\quad &\Big(\int_{\mathbb{R}^{d_1}}\cdots\Big(\int_{\mathbb{R}^{d_\kappa}} |f(x_1,\dots,x_\kappa)|^{p_\kappa} w(x_1,\dots,x_\kappa)\,dx_\kappa\Big)^{\frac{p_{\kappa-1}}{p_\kappa}}\dots dx_1\Big)^{\frac{1}{p_1}} \\ &\leq C\Big(\int_{\mathbb{R}^{d_1}}\cdots\Big(\int_{\mathbb{R}^{d_\kappa}} |g(x_1,\dots,x_\kappa)|^{p_\kappa} w(x_1,\dots,x_\kappa)\,dx_\kappa\Big)^{\frac{p_{\kappa-1}}{p_\kappa}}\dots dx_1\Big)^{\frac{1}{p_1}},\end{aligned}$$

for all $(f,g) \in \mathcal{F}$.

Proof: We present a proof by induction over κ. If $\kappa = 2$, the statement is a reformulation of Theorem 6.1: the assumption that $w_1^{\frac{p_1}{p_2}}, w_2 \in A_\infty$ will be rewritten so that $w_1^{\frac{p_1}{p_2}} \in A_{\frac{p_1}{s_0}}$, $w_2 \in A_{\frac{p_2}{s_0}}$, for a suitable $0 < s_0 < \infty$.

Since $0 < p_1, p_2 < \infty$ and $w_1^{\frac{p_1}{p_2}} \in A_\infty$, $w_2 \in A_\infty$, there exists $1 \leq s_1, s_2 < \infty$ such that

$$w_1^{\frac{p_1}{p_2}} \in A_{s_1}, \quad w_2 \in A_{s_2}.$$

We pick s_0 with $0 < s_0 \leq s$ with $s_1 \leq \frac{p_1}{s_0}$, $s_2 \leq \frac{p_2}{s_0}$ (these conditions reduce to $s_0 \leq \min\big(\frac{p_1}{s_1}, \frac{p_2}{s_2}, s\big)$). Because the weight classes are nested, we have in this situation $w_1^{\frac{p_1}{p_2}} \in A_{\frac{p_1}{s_0}}$, $w_2 \in A_{\frac{p_2}{s_0}}$.

The hypothesis (86) holds for all weights $w \in A_{\infty,\text{Rectangle}}$, and in particular also for $w \in A_{\frac{s}{s_0},\text{Rectangle}}$; the inequality in (87) then follows from Theorem 6.1.

Next, we assume that the result holds true when $\kappa - 1$ variables are involved and will prove it for κ variables as well. We fix a κ-tuple $(p_1, \dots, p_\kappa)$ and weights $w_1, \dots, w_\kappa$ satisfying $w_l^{\frac{p_l}{p_\kappa}} \in A_\infty(\mathbb{R}^{d_l})$ for all $1 \leq l \leq \kappa$. Denote

$$F(x_1, x_2) := \|f(x_1, \dots, x_\kappa)\|_{L^{p_3}_{x_3} \dots L^{p_\kappa}_{x_\kappa}(w_3 \cdot \ldots \cdot w_\kappa)},$$
$$G(x_1, x_2) := \|g(x_1, \dots, x_\kappa)\|_{L^{p_3}_{x_3} \dots L^{p_\kappa}_{x_\kappa}(w_3 \cdot \ldots \cdot w_\kappa)}.$$

We want to show that

$$\int_{\mathbb{R}^{d_1}} \Big(\int_{\mathbb{R}^{d_2}} |F(x_1, x_2)|^{p_2} w_1^{\frac{p_2}{p_\kappa}}(x_1) w_2^{\frac{p_2}{p_\kappa}}(x_2)\, dx_2 \Big)^{\frac{p_1}{p_2}} dx_1$$
$$\leq C \int_{\mathbb{R}^{d_1}} \Big(\int_{\mathbb{R}^{d_2}} |G(x_1, x_2)|^{p_2} w_1^{\frac{p_2}{p_\kappa}}(x_1) w_2^{\frac{p_2}{p_\kappa}}(x_2)\, dx_2 \Big)^{\frac{p_1}{p_2}} dx_1,$$

given that $w_1^{\frac{p_1}{p_\kappa}} \in A_\infty$ and $w_2^{\frac{p_2}{p_\kappa}} \in A_\infty$.

If we denote

$$W(x_1, x_2) := w_1^{\frac{p_2}{p_\kappa}}(x_1)\, w_2^{\frac{p_2}{p_\kappa}}(x_2) := U(x_1)\, V(x_2),$$

we have $U^{\frac{p_1}{p_2}} \in A_\infty$ and $V \in A_\infty$. The problem is reduced to the case $\kappa = 2$, and it remains to check that the hypothesis (86) is satisfied. That is, we need to check that there exists $0 < s < \infty$ such that for all weights $W_0 \in A_{\infty,\text{Rectangle}}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$

$$\int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} |F(x_1, x_2)|^s\, W_0(x_1, x_2)\, dx_1\, dx_2 \leq C \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} |G(x_1, x_2)|^s\, W_0(x_1, x_2)\, dx_1\, dx_2. \tag{88}$$

The case of $(\kappa - 1)$ iterated Lebesgue spaces applied to the tuple $(\tilde{p}_2, p_3, \dots, p_\kappa)$ for some $0 < \tilde{p}_2 < \infty$ yields, for weights of the form $w(x_1, x_2, \dots, x_\kappa) = \tilde{w}_2(x_1, x_2) \cdot w_3(x_3) \cdot \ldots \cdot w_\kappa(x_\kappa)$ such that $w_l^{\frac{p_l}{p_\kappa}} \in A_\infty(\mathbb{R}^{d_l})$ for all $3 \leq l \leq \kappa$ and $\tilde{w}_2^{\frac{\tilde{p}_2}{p_\kappa}}(x_1, x_2) \in A_\infty(\mathbb{R}^{d_1 + d_2})$, the estimate

$$\Big(\int_{\mathbb{R}^{d_1+d_2}} \cdots \Big(\int_{\mathbb{R}^{d_\kappa}} |f(x_1, x_2, \dots, x_\kappa)|^{p_\kappa} w(x_1, \dots, x_\kappa)\, dx_\kappa \Big)^{\frac{p_{\kappa-1}}{p_\kappa}} \dots dx_1\, dx_2 \Big)^{\frac{1}{\tilde{p}_2}}$$
$$\leq C \Big(\int_{\mathbb{R}^{d_1+d_2}} \cdots \Big(\int_{\mathbb{R}^{d_\kappa}} |g(x_1, x_2, \dots, x_\kappa)|^{p_\kappa} w(x_1, \dots, x_\kappa)\, dx_\kappa \Big)^{\frac{p_{\kappa-1}}{p_\kappa}} \dots dx_1\, dx_2 \Big)^{\frac{1}{\tilde{p}_2}}.$$

If the functions f and g and the weights $w_3, \dots, w_\kappa$ are precisely those we started with, we obtain, for any weight $\tilde{w}_2$ such that $\tilde{w}_2(x_1, x_2)^{\frac{\tilde{p}_2}{p_\kappa}} \in A_\infty(\mathbb{R}^{d_1+d_2})$, the estimate

$$\int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} |F(x_1,x_2)|^{\tilde{p}_2}\tilde{w}_2(x_1,x_2)\,dx_1\,dx_2 \leq C\int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} |G(x_1,x_2)|^{\tilde{p}_2}\tilde{w}_2(x_1,x_2)\,dx_1\,dx_2. \tag{89}$$

We want (88) for some $0 < s < \infty$ and all weights $W_0(x_1, x_2) \in A_{\infty,\text{Rectangle}}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$. Instead, the $(k-1)$ induction case yields the similar estimate (89) for any $0 < \tilde{p}_2 < \infty$ and any weight $\tilde{w}_2$ such that $\tilde{w}_2(x_1, x_2)^{\frac{\tilde{p}_2}{p_\kappa}} \in A_\infty(\mathbb{R}^{d_1+d_2})$. We get the desired estimate by choosing $s = p_\kappa$ and by noting that the class of weights for which the supremum over rectangles is finite is a subcollection of the class of weights for which the supremum over cubes is finite:

$$A_{\infty,\text{Rectangle}}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}) \subset A_\infty(\mathbb{R}^{d_1+d_2}). \qquad \square$$

Proof of the main Theorem 1.1. Now we want to deduce the general inequality

$$\|f\|_{L^P(L^Q)} \lesssim \|S(f)\|_{L^P(L^Q)}.$$

By extrapolating the scalar result of [**9**], we obtain the multiple vector-valued estimate of (84). Then we apply Theorem 6.3 in the case of $d = d_1 + \cdots + d_N$ variables, to obtain the mixed-norm, multiple vector-valued result.

6.2. Obtaining the weighted result by using the helicoidal method. As previously mentioned, we can obtain the weighted result directly from a sparse domination estimate, which follows from a *local maximal inequality.* A similar strategy was used in [**5**].

6.2.1. The localization lemma. For the weighted result, it is more suitable to work with locally integrable functions than with characteristic functions, the reason being that the characteristic function cannot play the role of an A_∞ weight.

We recall a few notations, for convenience:

Notation. If $\mathcal{I}$ is a collection of cubes in $\mathbb{R}^d$ and $I_0 \subseteq \mathbb{R}^d$ is a fixed dyadic cube, then

$$\mathcal{I}(I_0) := \{I \in \mathcal{I} : I \subseteq I_0\} \quad \text{and} \quad \mathcal{I}^+(I_0) := \mathcal{I}(I_0) \cup \{I_0\}.$$

For any cube $I \subset \mathbb{R}^d$, $\tilde{\chi}_I(x)$ denotes a function that decays fast away from I:

$$\tilde{\chi}_I(x) := \Big(1 + \frac{\operatorname{dist}(x, I)}{|I|}\Big)^{-\tilde{M}}, \tag{90}$$

where $\tilde{M}$ can be as large as we wish.

Remark 6.4. For statements involving a weight $w \in A_\infty$, the decaying factor $\tilde{M}$ in the definition (90) might depend on w. More exactly, if $w \in A_\infty$, then we know that $w \in A_{q_w}$ for some $q_w > 1$; we will need, in certain situations, to make sure that $d\, q_w < \tilde{M}$.

Lemma 6.5. *Let $0 < p \le 1$. Let $\mathcal{I}$ be a finite collection of dyadic squares in $\mathbb{R}^d$, I_0 a fixed dyadic square, $f : \mathbb{R}^d \to \mathbb{C}$ a Schwartz function, and w a positive locally integrable function. Then for any $0 < p_1 < \infty$,*

$$\begin{aligned} \Big\| \sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle\, \phi_I^2 \Big\|_{L^p(w)}^p & \\ \lesssim \Big(\sup_{J_1 \in \mathcal{I}(I_0)} \frac{1}{|J_1|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I}(I_0) \\ I \subseteq J_1}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big)^p & \\ \times \Big(\sup_{J_2 \in \mathcal{I}^+(I_0)} \frac{1}{|J_2|} \int_{\mathbb{R}} w \cdot \tilde{\chi}_{J_2}\, dx \Big) \cdot |I_0|, & \end{aligned} \tag{91}$$

with an implicit constant independent of the collection $\mathcal{I}$ and of the functions f and w.

Proof: If $0 < p < 1$, then $\| \cdot \|_p^p$ is subadditive. In this case, we have for some $0 < \tau < \infty$

$$\frac{1}{p} = 1 + \frac{1}{\tau}.$$

First, we note that $\big\| \sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle\, \phi_I^2 \big\|_{L^p(w)} = \big\| \big(\sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle\, \phi_I^2 \big) \cdot w^{\frac{1}{p}} \big\|_{L^p}$. We let $v_1 := w$ and $v_2 := w^{\frac{1}{\tau}}$, so that

$$w^{\frac{1}{p}} = v_1 \cdot v_2 \quad \text{and} \quad \Big\| \sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \phi_I^2 \Big\|_{L^p(w)} = \Big\| \Big(\sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) v_1 \cdot v_2 \Big\|_{L^p}.$$

We also use the previous decomposition (31) $\phi_I^2(x) := \sum_{\ell \geq 0} 2^{-\frac{\ell \tilde{M}}{2}} \tilde{\phi}_{I,\ell}^2(x)$, so that it suffices to show instead of (91) the similar inequality, for every $\ell \geq 0$:

$$\begin{aligned} &\Big\| \Big(\sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \tilde{\phi}_{I,\ell}^2 \Big) v_1 \cdot v_2 \Big\|_{L^p}^p \\ &\qquad \lesssim 2^{10\ell d p} \Big(\sup_{J_1 \in \mathcal{I}(I_0)} \frac{1}{|J_1|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I}(I_0) \\ I \subseteq J_1}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big)^p \\ &\qquad\qquad \times \Big(\sup_{J_2 \in \mathcal{I}^+(I_0)} \frac{1}{|J_2|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_{J_2} \, dx \Big)^p \cdot |I_0|. \end{aligned} \tag{92}$$

We recall that the families $\big(\tilde{\phi}_{I,\ell}^2\big)_{I \in \mathcal{I}}$ are all lacunary, L^2-normalized, and $\operatorname{supp} \tilde{\phi}_{I,\ell}^2 \subseteq 2^\ell I$ for all $I \in \mathcal{I}$. As before, we only present the case $\ell = 0$, since the general case follows from almost identical arguments.

By Hölder's inequality and the fact that all the functions $\tilde{\phi}_{I,0}^2$ are supported on $I \subseteq I_0$, we have

$$\begin{aligned} &\Big\| \Big(\sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \tilde{\phi}_{I,0}^2 \Big) v_1 \cdot v_2 \Big\|_{L^p} \\ &\qquad \lesssim \Big\| \Big(\sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \tilde{\phi}_{I,0}^2 \Big) v_1 \Big\|_{L^1} \cdot \| v_2 \cdot \mathbf{1}_{I_0} \|_{\tau}. \end{aligned} \tag{93}$$

The first expression can be rewritten as

$$\int_{\mathbb{R}^d} \Big(\sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \tilde{\phi}_{I,0}^2(x) \Big) v_1(x) \cdot \overline{g(x)} \, dx = \sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \overline{\langle \overline{v_1} \cdot g, \tilde{\phi}_{I,0}^2 \rangle},$$

for a certain function $g \in L^\infty$ satisfying $\|g\|_\infty = 1$. Next, we apply the Cauchy–Schwarz and John–Nirenberg inequalities to get

$$\begin{aligned} &\Big| \sum_{I \in \mathcal{I}(I_0)} \langle f, \phi_I^1 \rangle \langle v_1 \cdot g, \tilde{\phi}_{I,0}^2 \rangle \Big| \\ &\qquad \lesssim \Big(\sup_{J_1 \in \mathcal{I}(I_0)} \frac{1}{|J_1|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I}(I_0) \\ I \subseteq J_1}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big) \\ &\qquad\quad \times \Big(\sup_{J_2 \in \mathcal{I}^+(I_0)} \frac{1}{|J_2|^{\frac{1}{p_2}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I}(I_0) \\ I \subseteq J_2}} \frac{|\langle v_1 \cdot g, \tilde{\phi}_{I,0}^2 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_2,\infty} \Big) |I_0|, \end{aligned}$$

or any $0 < p_1, p_2 < \infty$.

Setting $p_2 = 1$ and using the $L^1 \to L^{1,\infty}$ boundedness of the square function (see also [**19**, Lemma 2.13]) we obtain

$$\Big\|\Big(\sum_{I\in\mathcal{I}(I_0)}\langle f,\phi_I^1\rangle\tilde{\phi}_{I,\ell}^2\Big)v_1\Big\|_{L^1} \lesssim \Big(\sup_{J_1\in\mathcal{I}(I_0)}\frac{1}{|J_1|^{\frac{1}{p_1}}}\Big\|\Big(\sum_{\substack{I\in\mathcal{I}(I_0)\\ I\subseteq J_1}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac12}\Big\|_{p_1}\Big)$$
$$\times\Big(\sup_{J_2\in\mathcal{I}^+(I_0)}\frac{1}{|J_2|}\int_{\mathbb{R}^d}v_1\cdot\tilde{\chi}_{J_2}\,dx\Big)\cdot|I_0|.$$

Recalling that $v_1 = w$ and $\|v_2\cdot\mathbf{1}_{I_0}\|_\tau = \big(\frac{1}{|I_0|}\|w\cdot\mathbf{1}_{I_0}\|_1\big)^{\frac{1}{\tau}}\cdot|I_0|^{\frac1\tau}$, the above estimate and (93) imply that

$$\Big\|\Big(\sum_{I\in\mathcal{I}(I_0)}\langle f,\phi_I^1\rangle\tilde{\phi}_{I,0}^2\Big)v_1\cdot v_2\Big\|_{L^p}$$
$$\lesssim\Big(\sup_{J_1\in\mathcal{I}(I_0)}\frac{1}{|J_1|^{\frac{1}{p_1}}}\Big\|\Big(\sum_{\substack{I\in\mathcal{I}(I_0)\\ I\subseteq J_1}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac12}\Big\|_{p_1}\Big)$$
$$\times\Big(\sup_{J_2\in\mathcal{I}^+(I_0)}\frac{1}{|J_2|}\int_{\mathbb{R}^d}w\cdot\tilde{\chi}_{J_2}\,dx\Big)\cdot|I_0|\cdot\Big(\frac{1}{|I_0|}\|w\cdot\mathbf{1}_{I_0}\|_1\Big)^{\frac1\tau}\cdot|I_0|^{\frac1\tau}.$$

Raising the inequality to the power p we obtain exactly (92) in the case $\ell = 0$.

For $\ell \geq 1$, the difference will consist in replacing (93) by

$$\Big\|\Big(\sum_{I\in\mathcal{I}(I_0)}\langle f,\phi_I^1\rangle\tilde{\phi}_{I,\ell}^2\Big)v_1\cdot v_2\Big\|_{L^p}\lesssim\Big\|\Big(\sum_{I\in\mathcal{I}(I_0)}\langle f,\phi_I^1\rangle\tilde{\phi}_{I,\ell}^2\Big)v_1\Big\|_{L^1}$$
$$\times 2^{\frac{\ell\tilde{M}_1}{\tau}}\Big(\frac{1}{|I_0|}\int_{\mathbb{R}^d}v_2^\tau\cdot\tilde{\chi}_{I_0}\,dx\Big)^{\frac1\tau}$$

and using the $L^1 \to L^{1,\infty}$ boundedness of the modified square function

$$g\mapsto\Big(\sum_{\substack{I\in\mathcal{I}(I_0)\\ I\subseteq J_2}}\frac{|\langle g,\tilde{\phi}_{I,\ell}^2\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac12},$$

which satisfies the same L^p estimates as the classical discretized square function of [**19**], uniformly in $\ell \geq 0$.

The inequality remains true if $p = 1$; in that case $\tau = \infty$ and there will be no second term on the right-hand side of (93). □

Remark 6.6. The local estimate in Lemma 6.5 should be compared to the maximal inequality in Theorem 19 of [**5**].

6.2.2. The stopping time. Further, Theorem 12 in [**5**] explains how to deduce sparse estimates from a local estimate such as (91) of Lemma 6.5. The procedure in [**5**] is stated for averages of functions, but the same is true when averages of square functions are concerned. A similar algorithm, based on the helicoidal method, was used in [**1**] to deduce a result on sparse domination by averages of localized square functions.

Definition 6.7. A collection of dyadic cubes $\mathcal{S}$ is said to be *sparse* if there exists $0<\eta<1$ such that for each $Q\in\mathcal{S}$ we have

$$\sum_{P\in ch_{\mathcal{S}}(Q)} |P| \leq (1-\eta)|Q|,$$

where $ch_{\mathcal{S}}(Q)$ is the collection of direct descendants of Q in $\mathcal{S}$ – the maximal elements of $\mathcal{S}$ that are strictly contained in Q

$$ch_{\mathcal{S}}(Q)=\{P\subsetneq Q : P\in\mathcal{S} \text{ and if } P'\in\mathcal{S},\, P\subset P'\subsetneq Q, \text{ then } P'=P\}.$$

Our result, making use of sparse collections as defined above, reads as follows:

Theorem 6.8. *Let $\mathcal{I}$ be a collection of dyadic squares, $0<p,p_1<\infty$, and w a positive locally integrable function. Then, for any $\epsilon_p>0$ and any Schwartz function f, there exists a sparse collection $\mathcal{S}$ of cubes (which depends on the functions f, w, the exponent p) such that*

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\cdot w^{\frac{1}{p}}\Big\|_p^p \lesssim \sum_{Q\in\mathcal{S}}\Big(\frac{1}{|Q|^{\frac{1}{p_1}}}\Big\|\Big(\sum_{\substack{I\in\mathcal{I}\\ I\subseteq Q}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac{1}{2}}\Big\|_{p_1}\Big)^p \times\Big(\frac{1}{|Q|}\int_{\mathbb{R}^d} w^{1+\epsilon_p}\,\tilde{\chi}_Q\,dx\Big)^{\frac{1}{1+\epsilon_p}}\cdot|Q|. \tag{94}$$

However, if $0<p\leq 1$, the above inequality is true for $\epsilon_p=0$.

Proof: We briefly sketch the proof for completeness, which consists of a rather standard argument, first in the case $0<p\leq 1$. As per usual, the collection $\mathcal{S}:=\bigcup_{k\geq 0}\mathcal{S}_k$, where the cubes in the subcollection $\mathcal{S}_{k+1}$ are to be understood as the "descendants" of the dyadic cubes in the previous generation $\mathcal{S}_k$:

$$\mathcal{S}_{k+1}:=\bigcup_{Q\in\mathcal{S}_k} ch_{\mathcal{S}}(Q).$$

To every $Q\in\mathcal{S}$, we also associate a subcollection $\mathcal{I}_Q\subseteq\mathcal{I}$ of cubes so that

$$\mathcal{I}:=\bigcup_{Q\in\mathcal{S}}\mathcal{I}_Q$$

represents a partition of the initial collection $\mathcal{I}$.

The bottom-most collection $\mathcal{S}_0$ will consist of the maximal dyadic cubes of the collection $\mathcal{I}$:

$$\mathcal{S}_0 := \{Q \in \mathcal{I} : Q \text{ maximal with respect to inclusion}\}.$$

Next, we assume that $\mathcal{S}_0$, $\mathcal{S}_1$ up to $\mathcal{S}_k$ are known and we will show how to construct $\mathcal{S}_{k+1}$, and for every $Q_0 \in \mathcal{S}_k$, the collections $\mathcal{I}_{Q_0}$.

If $Q_0 \in \mathcal{S}_k$, then we define

$$\begin{aligned}\Omega_{Q_0} := \Big\{ x \in Q_0 : \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q_0}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I(x) \Big)^{\frac{1}{2}} \\ > C \frac{1}{|Q_0|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q_0}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big\} \qquad (95) \\ \cup \Big\{ x \in Q_0 : M(w \cdot \tilde{\chi}_{I_0})(x) > C \frac{1}{|Q_0|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_{Q_0}(y)\, dy \Big\},\end{aligned}$$

and we set $E_{Q_0} := Q_0 \backslash \Omega_{Q_0}$. It is not difficult to see that, if we choose $C > 0$ large enough, $|E_{Q_0}| > \frac{|Q_0|}{2}$. Then $ch_{\mathcal{S}}(Q_0)$ will consist of a maximal covering of Ω_{Q_0} by dyadic cubes:

$$ch_{\mathcal{S}}(Q_0) := \{Q \text{ dyadic cube: } Q \subseteq \Omega_{Q_0}, \text{ maximal with respect to inclusion}\}$$

and also, as already stated, $\mathcal{S}_{k+1} := \bigcup_{Q_0 \in \mathcal{S}_k} ch_{\mathcal{S}}(Q_0)$. We have that

$$|ch_{\mathcal{S}}(Q_0)| = |\Omega_{Q_0}| \leq \frac{1}{2} |Q_0|,$$

which guarantees that the collection $\mathcal{S}$ is sparse. Also, it will be useful later to notice that the sets $\{E_Q\}_{Q \in \mathcal{S}}$ are all mutually disjoint.

Moreover, for every $Q_0 \in \mathcal{S}_k$, we define

$$\mathcal{I}_{Q_0} := \{I \in \mathcal{I} : I \subseteq Q_0, I \nsubseteq \Omega_{Q_0}\}.$$

In consequence, every $I \in \mathcal{I}_{Q_0}$ has the property that either it is disjoint from the intervals in $ch_{\mathcal{S}}(Q_0)$ or, if $Q \in ch_{\mathcal{S}}(Q_0)$ and $I \cap Q \neq \emptyset$, then necessarily $Q \subsetneq I$. This implies in particular that the localized square function

$$S_{\mathcal{I}_{Q_0}} f(x) := \Big(\sum_{\substack{I \in \mathcal{I}_{Q_0} \\ I \subseteq Q_0}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I(x) \Big)^{\frac{1}{2}} \qquad (96)$$

is constant on each $Q \in ch_{\mathcal{S}}(Q_0)$ and moreover, for every $x \in \Omega_{Q_0}$,

$$S_{\mathcal{I}_{Q_0}} f(x) \lesssim \frac{1}{|Q_0|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q_0}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1}.$$

The same inequality remains true on E_{Q_0}, by definition (95). So that we have, for every $J_1 \in \mathcal{I}_{Q_0}$,

$$\frac{1}{|J_1|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I}_{Q_0} \\ I \subseteq J_1}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \lesssim \frac{1}{|Q_0|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q_0}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1}.$$

Also, all $J_2 \in \mathcal{I}_{Q_0}$ intersect $\big\{ x \in Q_0 : M(w \cdot \tilde{\chi}_{I_0})(x) > C \frac{1}{|Q_0|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_{Q_0}(y)\, dy \big\}^c$, which implies

$$\sup_{J_2 \in \mathcal{I}_{Q_0}^+} \frac{1}{|J_2|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_{J_2}\, dx \lesssim \frac{1}{|Q_0|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_{Q_0}\, dx.$$

Using the subadditivity of $\| \cdot \|_p^p$ and the result in Lemma 6.5, we have

$$\begin{aligned}
\Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \cdot w^{\frac{1}{p}} \Big\|_p^p &\lesssim \sum_{Q \in \mathcal{S}} \Big\| \Big(\sum_{I \in \mathcal{I}_Q} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \cdot w^{\frac{1}{p}} \Big\|_p^p \\
&\lesssim \sum_{Q \in \mathcal{S}} \Big(\sup_{J_1 \in \mathcal{I}_Q} \frac{1}{|J_1|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I}_Q \\ I \subseteq J_1}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big)^p \\
&\qquad\qquad \times \Big(\sup_{J_2 \in \mathcal{I}_Q^+} \frac{1}{|J_2|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_{J_2}\, dx \Big) \cdot |Q| \\
&\lesssim \sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big)^p \cdot \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} w \cdot \tilde{\chi}_Q\, dx \Big) \cdot |Q|.
\end{aligned}$$

If $p > 1$, we invoke a procedure that has already appeared in Proposition 20 of our previous [**5**]. In this situation, we can use duality:

$$\Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \cdot w^{\frac{1}{p}} \Big\|_p = \Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \cdot w^{\frac{1}{p}}\, u \Big\|_1,$$

for some function $u \in L^{p'}$ with $\|u\|_{L^{p'}} = 1$. Now we can apply the result of Theorem 6.8 for $p = 1$ to deduce the existence of a sparse collection $\mathcal{S}$ so that

$$\begin{aligned}
\Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \cdot w^{\frac{1}{p}}\, u \Big\|_1 &\lesssim \sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big) \\
&\qquad \times \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} w^{\frac{1}{p}}\, u \cdot \tilde{\chi}_Q\, dx \Big) \cdot |Q|.
\end{aligned}$$

Hölder's inequality, first with respect to the measure $\tilde{\chi}_Q\, dx$ and with exponents $p+\epsilon$ and $(p+\epsilon)'$, yields

$$\Big\| \Big(\sum_{I \in \mathcal{I}} \langle f, \phi_I^1 \rangle \phi_I^2 \Big) \cdot w^{\frac{1}{p}}\, u \Big\|_1$$
$$\lesssim \sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big)$$
$$\times \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} w^{\frac{p+\epsilon}{p}} \cdot \tilde{\chi}_Q\, dx \Big)^{\frac{1}{p+\epsilon}} \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} u^{(p+\epsilon)'} \cdot \tilde{\chi}_Q\, dx \Big)^{\frac{1}{(p+\epsilon)'}} \cdot |Q|.$$

Then we again use Hölder's inequality with respect to the discrete measure $\ell^p(\mathcal{S})$ to estimate the above expression by

$$\Big(\sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|^{\frac{1}{p_1}}} \Big\| \Big(\sum_{\substack{I \in \mathcal{I} \\ I \subseteq Q}} \frac{|\langle f, \phi_I^1 \rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{1}{2}} \Big\|_{p_1} \Big)^p \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} w^{\frac{p+\epsilon}{p}} \cdot \tilde{\chi}_Q\, dx \Big)^{\frac{p}{p+\epsilon}} \cdot |Q| \Big)^{\frac{1}{p}}$$
$$\times \Big(\sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} u^{(p+\epsilon)'} \cdot \tilde{\chi}_Q\, dx \Big)^{\frac{p'}{(p+\epsilon)'}} \cdot |Q| \Big)^{\frac{1}{p'}}.$$

For the last term, we take advantage of the sparseness property, or more exactly, we use the disjointness of the sets $\{E(Q)\}_{Q \in \mathcal{S}}$ and the fact that $|E(Q)| > |Q|/2$:

$$\Big(\sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} u^{(p+\epsilon)'} \cdot \tilde{\chi}_Q\, dx \Big)^{\frac{p'}{(p+\epsilon)'}} \cdot |Q| \Big)^{\frac{1}{p'}}$$
$$\lesssim \Big(\sum_{Q \in \mathcal{S}} \Big(\frac{1}{|Q|} \int_{\mathbb{R}^d} u^{(p+\epsilon)'} \cdot \tilde{\chi}_Q\, dx \Big)^{\frac{p'}{(p+\epsilon)'}} \cdot |E(Q)| \Big)^{\frac{1}{p'}}$$
$$\lesssim \|M_{(p+\epsilon)'} u\|_{p'} \lesssim \|u\|_{p'} = 1.$$

We are losing an ϵ (as small as we wish) in making sure that the maximal operator $M_{(p+\epsilon)'}$ is bounded on $L^{p'}$. We can choose ϵ such that $\epsilon_p = \frac{p+\epsilon}{p}$. □

Such a sparse estimate allows us to recover the weighted estimates from [**9**], in the one-parameter case.

Proposition 6.9. *Let $0 < p < \infty$, $w \in A_\infty$, and f a Schwartz function on $\mathbb{R}^d$; then*

$$\|f\|_{L^p(w)} \lesssim \|Sf\|_{L^p(w)}. \tag{97}$$

Proof: The weighted estimate follows easily once we prove a strengthening of the sparse estimate (94), with $p_1 > 0$ to be chosen later: there exists a sparse collection of dyadic cubes $\mathcal{S}$ such that

$$(98)\quad \Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\cdot w^{\frac{1}{p}}\Big\|_p^p \lesssim \sum_{Q\in\mathcal{S}}\Big(\frac{1}{|Q|^{\frac{1}{p_1}}}\Big\|\Big(\sum_{\substack{I\in\mathcal{I}\\ I\subseteq Q}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac{1}{2}}\Big\|_{p_1}\Big)^p w(E(Q)).$$

If such an estimate were true, we could deduce that

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\Big\|_{L^p(w)}^p \lesssim \sum_{Q\in\mathcal{S}}\Big(\inf_{y\in Q} M(|S\,f|^{p_1})(y)\Big)^{\frac{p}{p_1}}\cdot w(E(Q)),$$

and in consequence,

$$\Big\|\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big\|_{L^p(w)}^p \lesssim \int_{\mathbb{R}^d} M(|S\,f|^{p_1})(x))^{\frac{p}{p_1}}\,w(x)\,dx.$$

So far, no information has been required on p_1; it suffices to choose $p_1 < p$ and such that $w \in A_{\frac{p}{p_1}}$ (this will assure that M is bounded on $L^{\frac{p}{p_1}}(w)$) to obtain that

$$\Big\|\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big\|_{L^p(w)} \lesssim \|Sf\|_{L^p(w)}.$$

This is possible since $w \in A_\infty = \bigcup_{q\geq 1} A_q$. The final inequality (97) is deduced thanks to the formula (14).

It remains to show how (94) implies (98). We recall that

$$\begin{aligned}\Big(\frac{1}{|Q|}\int_{\mathbb{R}^d} w^{1+\epsilon_p}\,\tilde{\chi}_Q\,dx\Big)^{\frac{1}{1+\epsilon_p}} &\leq \sum_{\ell\geq 0} 2^{-\ell\tilde{M}}\Big(\frac{1}{|Q|}\int_{2^\ell Q} w^{1+\epsilon_p}\,dx\Big)^{\frac{1}{1+\epsilon_p}}\\ &\leq \sum_{\ell\geq 0} 2^{-\ell\tilde{M}}2^{\frac{\ell d}{1+\epsilon_p}}\Big(\frac{1}{|2^\ell Q|}\int_{2^\ell Q} w^{1+\epsilon_p}\,dx\Big)^{\frac{1}{1+\epsilon_p}}.\end{aligned}$$

Now we use the *Reverse Hölder* property of the weight w: there exists ϵ_w such that

$$\Big(\frac{1}{|2^\ell Q|}\int_{2^\ell Q} w^{1+\epsilon_w}\,dx\Big)^{\frac{1}{1+\epsilon_w}} \lesssim \frac{1}{|2^\ell Q|}\int_{2^\ell Q} w\,dx.$$

If we pick $\epsilon_p < \epsilon_w$, then the $L^{1+\epsilon_p}$ average in (94) can be replaced by an L^1 average (note that, for $0 < p \leq 1$, we have from the start $\epsilon_p = 0$).

Hence, we have

$$\Big\|\Big(\sum_{I\in\mathcal{I}}\langle f,\phi_I^1\rangle\phi_I^2\Big)\cdot w^{\frac{1}{p}}\Big\|_p^p$$

$$\lesssim \sum_{\ell\geq 0} 2^{-\ell\tilde{M}} 2^{\frac{\ell d}{1+\epsilon_p}} \sum_{Q\in\mathcal{S}}\Big(\frac{1}{|Q|^{\frac{1}{p_1}}}\Big\|\Big(\sum_{\substack{I\in\mathcal{I}\\ I\subseteq Q}}\frac{|\langle f,\phi_I^1\rangle|^2}{|I|}\cdot\mathbf{1}_I\Big)^{\frac{1}{2}}\Big\|_{p_1}\Big)^p 2^{-\ell d} w(2^\ell Q).$$

All we need to do is compare $w(2^\ell Q)$ and $w(E(Q))$. We know that $|Q| < 2\,|E(Q)|$ and $w\in A_\infty$. Then $w\in A_{q_w}$ for some $q_w>1$ and in consequence (see inequality (7.2) of [**10**])

$$w(2^\ell Q)\Big(\frac{|E(Q)|}{|2^\ell Q|}\Big)\lesssim w(E(Q)) \Longleftrightarrow w(2^\ell Q)\lesssim 2^{\ell d q_w} w(E(Q)).$$

If $\tilde{M}$, the decaying exponent of the auxiliary weights $\tilde{\chi}_Q$ (see Definition 90), satisfies $d\,q_w<\tilde{M}$, then we can sum in $\ell\geq 0$ and we are done. Since $\tilde{M}$ can be as large as we wish, we can arrange for this condition to be satisfied. □

We note that the sparse domination result (94) of Theorem 6.8 implies, for any collection $\mathcal{I}$ of dyadic squares and any fixed dyadic square I_0:

$$\Big\|\Big(\sum_{I\in\mathcal{I}(I_0)}\langle f,\phi_I^1\rangle\phi_I^2\Big)\Big\|_{L^p(w)}^p \lesssim \Big(\sup_{J_2\in\mathcal{I}^+(I_0)}\frac{1}{|J_2|}\int_{\mathbb{R}^d} w(x)\,\tilde{\chi}_{J_2}\,dx\Big)\|\mathcal{S}_{\mathcal{I}(I_0)} f\|_p^p. \tag{99}$$

This observation will be useful shortly, as we will show that it is possible to prove a multiple vector-valued weighted result without making use of extrapolation.

Proposition 6.10. *Let $0<p<\infty$, $0<Q<\infty$, and $w\in A_\infty$; then for any L^Q-valued Schwartz function f on $\mathbb{R}^d$, we have*

$$\|f\|_{L^p(L^Q;d\,w)}\lesssim\|S\,f\|_{L^p(L^Q;d\,w)}.$$

The proof combines all the previous techniques used to deduce multiple vector-valued estimates in Section 4 and weighted estimates. We sketch the proof of the crucial maximal inequality (the equivalent of (91) of Lemma 6.5) in the case of ℓ^q-valued functions, where $q<1$. The case $q\geq 1$ is in fact easier, since duality is available. The general multiple vector-valued case, corresponding to a general n-tuple Q, follows by induction over n.

Lemma 6.11. *Let $0 < q < 1$ and $0 < p \le q$; let $\mathcal{I}$ be a finite collection of dyadic squares in $\mathbb{R}^d$, I_0 a fixed dyadic square, $(f_k)_{k\in\mathbb{Z}}$ a sequence of Schwartz functions, and w a positive locally integrable function. Then for any $0 < p_1 < \infty$,*

$$\Big\| \Big(\sum_k \Big| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle\, \phi_I^2 \Big|^q \Big)^{\frac{1}{q}} \Big\|_{L^p(w)}$$

$$\lesssim \Big(\sup_{J_1\in\mathcal{I}(I_0)} \frac{1}{|J_1|^{\frac{1}{p_1}}} \Big\| \Big(\sum_k \Big(\sum_{\substack{I\in\mathcal{I}(I_0)\\ I\subseteq J_1}} \frac{|\langle f_k, \phi_I^1\rangle|^2}{|I|} \cdot \mathbf{1}_I \Big)^{\frac{q}{2}} \Big)^{\frac{1}{q}} \Big\|_{p_1} \Big) \tag{100}$$

$$\times \Big(\sup_{J_2\in\mathcal{I}^+(I_0)} \frac{1}{|J_2|} \int_{\mathbb{R}} w\cdot \tilde{\chi}_{J_2}\, dx \Big)^{\frac{1}{p}} \cdot |I_0|^{\frac{1}{p}},$$

with the implicit constant independent of the collection $\mathcal{I}$ and of the functions f and w.

Proof: Since $\|\cdot\|^p_{L^p(\ell^q;d\,w)}$ is subadditive, and hence, using the decomposition (31),

$$\Big\| \Big(\sum_k \Big| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle \phi_I^2 \Big|^q \Big)^{\frac{1}{q}} \Big\|^p_{L^p(w)}$$

$$\lesssim \sum_{\ell\ge 0} 2^{-\ell p\tilde{M}} \Big\| \Big(\sum_k \Big| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle \tilde{\phi}_{I,\ell}^2 \Big|^q \Big)^{\frac{1}{q}} \Big\|^p_{L^p(w)}.$$

Since $p \le q$ and all the functions $\tilde{\phi}^2_{I,\ell}$ are supported inside $2^\ell I_0$:

$$\Big\| \Big(\sum_k \Big| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle \tilde{\phi}_{I,\ell}^2 \Big|^q \Big)^{\frac{1}{q}} \Big\|_{L^p(w)}$$

$$\lesssim \Big\| \Big(\sum_k \Big| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle\, \tilde{\phi}_{I,\ell}^2 \Big|^q \Big)^{\frac{1}{q}} \Big\|_{L^q(w)} \cdot \|\mathbf{1}_{2^\ell I_0}\|_{L^\tau(w)},$$

where $\frac{1}{p} = \frac{1}{q} + \frac{1}{\tau}$. For the first term on the right-hand side, we use Fubini and the known scalar version of Lemma 6.11 (more precisely, inequality (99) above):

$$\Big\| \Big(\sum_k \Big| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle \tilde{\phi}_{I,\ell}^2 \Big|^q \Big)^{\frac{1}{q}} \Big\|^q_{L^q(w)} = \sum_k \Big\| \sum_{I\in\mathcal{I}(I_0)} \langle f_k, \phi_I^1\rangle \tilde{\phi}_{I,\ell}^2 \Big\|^q_{L^q(w)}$$

$$\lesssim \sum_k \|\mathcal{S}_{\mathcal{I}(I_0)} f_k\|_q^q \Big(\sup_{J_2\in\mathcal{I}^+(I_0)} \frac{1}{|J_2|} \int_{\mathbb{R}^d} w(x)\, \tilde{\chi}_{J_2}\, dx \Big)$$

$$\lesssim \Big(\frac{1}{|I_0|^{\frac{1}{q}}} \Big\| \Big(\sum_k |\mathcal{S}_{\mathcal{I}(I_0)} f_k|^q \Big)^{\frac{1}{q}} \Big\|_q \Big)^q \cdot \Big(\sup_{J_2\in\mathcal{I}^+(I_0)} \frac{1}{|J_2|} \int_{\mathbb{R}^d} w(x)\, \tilde{\chi}_{J_2}\, dx \Big) \cdot |I_0|.$$

By a vector-valued version of the John–Nirenberg inequality, which was also used in proving the multiple vector-valued version of Theorem 1.1, the above can be estimated by

$$\Big(\sup_{J_1 \in \mathcal{I}(I_0)} \frac{1}{|I_0|^{\frac{1}{p_1}}} \Big\| \Big(\sum_k |\mathcal{S}_{\mathcal{I}(I_0)} f_k|^q \Big)^{\frac{1}{q}} \Big\|_{p_1} \Big)^q \times \Big(\sup_{J_2 \in \mathcal{I}^+(I_0)} \frac{1}{|J_2|} \int_{\mathbb{R}^d} w(x)\, \tilde{\chi}_{J_2}\, dx \Big) \cdot |I_0|,$$

where $0 < p_1 < \infty$ is any Lebesgue exponent.

On the other hand,

$$2^{-\ell p \tilde{M}/2} \| \mathbf{1}_{2^\ell I_0} \|^p_{L^\tau(w)} \lesssim \Big(\frac{1}{|I_0|} \int_{\mathbb{R}^d} w(x)\, \tilde{\chi}_{I_0}\, dx \Big)^{\frac{p}{\tau}} |I_0|^{\frac{p}{\tau}}.$$

After summing in $\ell \geq 0$, we get the inequality (100). □

Applying the usual stopping time, the maximal inequality of Lemma 6.11 will imply a vector-valued version of Theorem 6.8. We leave the details to the interested reader. Although Lemma 6.11 is stated for $p \leq q$, a vector-valued version of Theorem 6.8 is valid for any Lebesgue exponents, as we can pass from lower Lebesgue exponents to larger ones at the expense of losing an ϵ.

6.2.3. The multi-parameter case. The multi-parameter version of Proposition 6.9 follows easily from the properties of the weights $A_{\infty,\mathrm{Rectangle}}(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N})$. We will only illustrate the scalar bi-parameter case, but state the result in its generality.

Proposition 6.12. *Let $0 < p < \infty$, $0 < Q < \infty$; then for any $w \in A_{\infty,\mathrm{Rectangle}}(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N})$ and any L^Q-valued Schwartz function f,*

$$\|f\|_{L^p(L^Q)(w)} \leq C \| S_{d_1} \otimes \cdots \otimes S_{d_N}(f) \|_{L^p(L^Q)(w)}.$$

Proof: In fact, we will prove that

$$\|f\|_{L^p(w)} \leq C \| S_{d_1} \otimes S_{d_2}(f) \|_{L^p(w)}, \tag{101}$$

for any $w \in A_{\infty,\mathrm{Rectangle}}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$. An important property of the weights in the class $A_{\infty,\mathrm{Rectangle}}(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_N})$ is that if we fix one of the variables, we still obtain an A_∞ weight in the other variable and we can use the one-parameter result:

$$\begin{aligned} w_y(x) &= w(x,y) \in A_\infty(\mathbb{R}^{d_1}) \quad \text{for a.e. } y \in \mathbb{R}^{d_2}, \\ w_x(y) &= w(x,y) \in A_\infty(\mathbb{R}^{d_2}) \quad \text{for a.e. } x \in \mathbb{R}^{d_2}. \end{aligned} \tag{102}$$

We start by fixing the variable y; then $f_y(x) := f(x,y)$ is a function on $\mathbb{R}^{d_1}$. By Proposition 6.9,

$$\int_{\mathbb{R}^{d_1}} |f(x,y)|^p w(x,y)\,dx \lesssim \int_{\mathbb{R}^{d_1}} |S_{d_1} f_y(x)|\,dx$$
$$= \int_{\mathbb{R}^{d_1}} \Big(\sum_k |Q_k(f_y)(x)|^2\Big)^{\frac{p}{2}} w(x,y)\,dx.$$

Above,

$$Q_k(f_y)(x) := Q_k^1 f(x,y) := Q_k^x f(y) := \int_{\mathbb{R}^{d_1}} f(x-s,y)\psi_k(s)\,ds. \tag{103}$$

If we integrate with respect to y and use Fubini, we have

$$\int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_2}} |f(x,y)|^p w(x,y)\,dy\,dx$$
$$\lesssim \int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_2}} \Big(\sum_k |(Q_k^x f)(y)|^2\Big)^{\frac{p}{2}} w(x,y)\,dy\,dx.$$

Now we consider x fixed and we apply Proposition 6.9 (or more specifically an ℓ^2-valued extension which follows also from a well-known result of Marcinkiewicz and Zygmund [**17**]) to the sequence of functions $(Q_k^x f)_{k\in\mathbb{Z}}$:

$$\int_{\mathbb{R}^{d_2}} \Big(\sum_k |(Q_k^x f)(y)|^2\Big)^{\frac{p}{2}} w(x,y)\,dy$$
$$\lesssim \int_{\mathbb{R}^{d_2}} \Big(\sum_k |S_{d_2}(Q_k^x f)(y)|^2\Big)^{\frac{p}{2}} w(x,y)\,dy$$
$$\lesssim \int_{\mathbb{R}^{d_2}} \Big(\sum_k \Big|\sum_l Q_l(Q_k^x f)(y)\Big|^2\Big)^{\frac{p}{2}} w(x,y)\,dy.$$

Here it is useful that we can interchange the role played by the variables: if x is fixed, $w(x,\cdot)$ is still an A_∞ weight and vice-versa.

We need to understand the last expression; the explicit formula for $Q_l(Q_k^x f)(y)$ is

$$\int_{\mathbb{R}^{d_2}} (Q_k^x f)(y-t)\,\psi_l(t)\,dt = \int_{\mathbb{R}^{d_2}} \Big(\int_{\mathbb{R}^{d_1}} f(x-s,y-t)\psi_k(s)\,ds\Big)\psi_l(t)\,dt$$
$$= f * (\psi_k \otimes \psi_l)(x,y),$$

so that

$$\Big(\sum_k |S_{d_2}(Q_k^x f)(y)|^2\Big)^{\frac{p}{2}} = \Big(\sum_k \Big|\sum_l Q_l(Q_k^x f)(y)\Big|^2\Big)^{\frac{1}{2}} = S_{d_1}\otimes S_{d_2}(f)(x,y).$$

Integrating in x we obtain (101). Note that here it is important that we can use Fubini, fix one of the variables and perform the usual one-parameter analysis; in particular, the properties (102) are critical. For mixed-norm estimates most of the weighted results are known only for weights that tensorize: $w(x,y) = u(x)\,v(y)$, the reason being that Fubini and property (102) do not hold any longer. □

References

[1] C. Benea and F. Bernicot, Conservation de certaines propriétés à travers un contrôle épars d'un opérateur et applications au projecteur de Leray–Hopf, *Ann. Inst. Fourier (Grenoble)* **68(6)** (2018), 2329–2379. DOI: 10.5802/aif.3211.

[2] C. Benea and C. Muscalu, Multiple vector-valued inequalities via the helicoidal method, *Anal. PDE* **9(8)** (2016), 1931–1988. DOI: 10.2140/apde.2016.9.1931.

[3] C. Benea and C. Muscalu, Quasi-Banach valued inequalities via the helicoidal method, *J. Funct. Anal.* **273(4)** (2017), 1295–1353. DOI: 10.1016/j.jfa.2017.04.014.

[4] C. Benea and C. Muscalu, The helicoidal method, in: *"Operator Theory: Themes and Variations"*, Theta Ser. Adv. Math. **20**, Theta, Bucharest, 2018, pp. 45–96.

[5] C. Benea and C. Muscalu, Sparse domination via the helicoidal method, *Rev. Mat. Iberoam.* **37(6)** (2021), 2037–2118. DOI: 10.4171/rmi/1266.

[6] D. Cruz-Uribe and J. M. Martell, Limited range multilinear extrapolation with applications to the bilinear Hilbert transform, *Math. Ann.* **371(1-2)** (2018), 615–653. DOI: 10.1007/s00208-018-1640-9.

[7] D. Cruz-Uribe, J. M. Martell, and C. Pérez, Extrapolation from A_∞ weights and applications, *J. Funct. Anal.* **213(2)** (2004), 412–439. DOI: 10.1016/j.jfa.2003.09.002.

[8] D. Cruz-Uribe, J. M. Martell, and C. Pérez, *"Weights, Extrapolation and the Theory of Rubio de Francia"*, Operator Theory: Advances and Applications **215**, Birkhäuser/Springer Basel AG, Basel, 2011. DOI: 10.1007/978-3-0348-0072-3.

[9] Y. Ding, Y. Han, G. Lu, and X. Wu, Boundedness of singular integrals on multiparameter weighted Hardy spaces $H^p_w(\mathbb{R}^n \times \mathbb{R}^m)$, *Potential Anal.* **37(1)** (2012), 31–56. DOI: 10.1007/s11118-011-9244-y.

[10] J. Duoandikoetxea, *"Fourier Analysis"*, Translated and revised from the 1995 Spanish original by David Cruz-Uribe, Graduate Studies in Mathematics **29**, American Mathematical Society, Providence, RI, 2001. DOI: 10.1090/gsm/029.

[11] M. Frazier and B. Jawerth, A discrete transform and decompositions of distribution spaces, *J. Funct. Anal.* **93(1)** (1990), 34–170. DOI: 10.1016/0022-1236(90)90137-A.

[12] R. F. Gundy and E. M. Stein, H^p theory for the poly-disc, *Proc. Nat. Acad. Sci. U.S.A.* **76(3)** (1979), 1026–1029. DOI: 10.1073/pnas.76.3.1026.

[13] J. Hart, R. H. Torres, and X. Wu, Smoothing properties of bilinear operators and Leibniz-type rules in Lebesgue and mixed Lebesgue spaces, *Trans. Amer. Math. Soc.* **370(12)** (2018), 8581–8612. DOI: 10.1090/tran/7312.

[14] L. Huang, J. Liu, D. Yang, and W. Yuan, Atomic and Littlewood–Paley characterizations of anisotropic mixed-norm Hardy spaces and their applications, *J. Geom. Anal.* **29(3)** (2019), 1991–2067. DOI: 10.1007/s12220-018-0070-y.

[15] D. S. Kurtz, Classical operators on mixed-normed spaces with product weights, *Rocky Mountain J. Math.* **37(1)** (2007), 269–283. DOI: 10.1216/rmjm/1181069331.

[16] K. Li, J. M. Martell, and S. Ombrosi, Extrapolation for multilinear Muckenhoupt classes and applications, *Adv. Math.* **373** (2020), 107286, 43 pp. DOI: 10.1016/j.aim.2020.107286.

[17] J. Marcinkiewicz and A. Zygmund, Quelques inégalités pour les opérations linéaires, *Fund. Math.* **32** (1939), 115–121. DOI: 10.4064/fm-32-1-115-121.

[18] C. Muscalu, J. Pipher, T. Tao, and C. Thiele, Multi-parameter paraproducts, *Rev. Mat. Iberoam.* **22(3)** (2006), 963–976. DOI: 10.4171/RMI/480.

[19] C. Muscalu and W. Schlag, *"Classical and Multilinear Harmonic Analysis"*, Vol II, Cambridge Studies in Advanced Mathematics **138**, Cambridge University Press, Cambridge, 2013. DOI: 10.1017/CBO9781139410397.

[20] E. M. Stein, *"Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals"*, With the assistance of Timothy S. Murphy, Princeton Mathematical Series **43**, Monographs in Harmonic Analysis, III, Princeton University Press, Princeton, NJ, 1993. DOI: 10.1515/9781400883929.

Cristina Benea
Université de Nantes, Laboratoire Jean Leray, Nantes, 44311, France
E-mail address: cristina.benea@univ-nantes.fr

Camil Muscalu
Department of Mathematics, Cornell University, Ithaca, NY 14853, USA
E-mail address: camil@math.cornell.edu

Received on July 20, 2020.
Accepted on November 8, 2021.