SCHATTEN CLASS ESTIMATES FOR PARAPRODUCTS IN MULTI-PARAMETER SETTING

MICHAEL T. LACEY*, JI LI†, AND BRETT D. WICK‡

Abstract. Let Π_b be a bounded n parameter paraproduct with symbol b. We demonstrate that this operator is in the Schatten class $\mathbb{S}^p$, $0 < p < \infty$, if the symbol is in the n parameter Besov space $\mathbb{B}^p$. Our result covers both the dyadic and continuous version of the paraproducts in the multiparameter setting.

1. Introduction, Notation and Statement of Main Results

Paraproducts are an extremely useful tool in questions arising in harmonic analysis. They provide a nice class of singular integral operators, and when restricting to the dyadic case provide much insight into the mapping properties of Calderón–Zygmund operators. Paraproducts have natural connections with other important operators in analysis. In particular, it is possible to view paraproducts as either the commutator between a function and the Hilbert transform or equivalently a Hankel operator with a certain symbol. It turns out the properties of the symbol heavily influence the operator theoretic characteristics of the paraproduct. In this paper, we are interested in the property of the paraproduct (both continuous version and dyadic version) being in certain Schatten class, with applications to commutators in multi-parameter settings which link to the big Hankel operators.

We note that along the line of Calderón [2], Coifman–Rochberg–Weiss [6], Uchiyama [27], Janson–Wolff [11], Rochberg–Semmes [25], the theory of commutators (boundedness, compactness and Schatten class) plays an important role, which connects to the weak factorisation of Hardy space ([6]) and Hankel operators ([1, 22, 24]) in the complex analysis, compensated compactness in the PDEs [5], as well as the quantised derivative in non-commutative analysis and geometry [7, 17].

To state the main results of this paper, let us recall the relevant paraproducts. In so doing, we prefer a discrete formulation of these operators. As is well known, there are two distinct ways to formulate them, with the Haar basis playing a distinguished role. So, for the sake of definiteness, we first give the Haar paraproducts and the statements of the main results in this context.

Here we make some conventions on notation. Throughout the paper, the letter C denotes (possibly different) constants that are independent of the essential variables. If $A \leq CB$, we write $A \lesssim B$ or $B \gtrsim A$; and if $A \lesssim B \lesssim A$, we write $A \simeq B$.

One Parameter Paraproducts. We take the dyadic intervals to be

$$\mathcal{D} \stackrel{\text{def}}{=} \left\{ [j2^k, (j+1)2^k) \,:\, j, k \in \mathbb{Z} \right\}.$$

2020 *Mathematics Subject Classification.* 47B10, 42B20, 42B35.

Key words and phrases. Paraproducts, Besov Space, Schatten Class, multi-parameter analysis.

Each dyadic interval I is a union of its left and right halves, denoted I_- and I_+ respectively. The Haar function h_I adapted to I is

$$h_I^0 \stackrel{\text{def}}{=} |I|^{-1/2}\left(-\mathbf{1}_{I_-}+\mathbf{1}_{I+}\right). \tag{1.1}$$

The other function of importance is

$$h_I^1 = |I|^{-1/2}\mathbf{1}_I. \tag{1.2}$$

Thus, h_I^0 has cancellation, i.e. $\int_{\mathbb{R}} h_I^0(x)dx = 0$, while h_I^1 is a multiple of an indicator function. The function h_I^1 is, in wavelet nomenclature, the 'father' wavelet.

The one parameter Haar paraproducts are

$$\mathrm{B}_{\mathrm{Haar}}(f_1,f_2) \stackrel{\text{def}}{=} \sum_{I\in\mathcal{D}} |I|^{-1/2} h_I^1 \prod_{j=1}^{2} \langle f_j, h_I^0\rangle_{L^2(\mathbb{R})}. \tag{1.3}$$

A classic result in dyadic harmonic analysis is that (see for example [3, Section 2.1.])

$$\|\mathrm{B}_{\mathrm{Haar}}(f_1,\cdot)\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \simeq \|f_1\|_{\mathrm{BMO}_{\mathrm{dyadic}}\,(\mathbb{R})}. \tag{1.4}$$

In this last display, $\mathrm{BMO}_{\mathrm{dyadic}}\,(\mathbb{R})$ is the dyadic BMO space.

To define more general paraproducts, we will appeal to *wavelets*. We make this more precise now. For an interval I, we say that φ **is *adapted to*** I if and only if $\|\varphi\|_2 = 1$ and

$$\Big|\frac{d^\alpha}{dx^\alpha}\varphi(x)\Big| \lesssim |I|^{-\alpha-\frac12}\left(1+\frac{|x-c(I)|}{|I|}\right)^{-N}, \qquad \alpha = 0,1,2. \tag{1.5}$$

Here, $c(I)$ denotes the center of I, and N is a large fixed integer, whose exact value need not concern us. By $\{\varphi_I : I\in\mathcal{D}\}$ ***are adapted to*** $\mathcal{D}$ we mean that for all dyadic intervals I, φ_I are adapted to I. We shall consistently work with functions which are normalized in $L^2(\mathbb{R})$. Some of these functions we will also insist to be normalized to **have cancellation, i.e.**

$$\int_{\mathbb{R}} \varphi_I(x)dx = 0.$$

A collection $\{\varphi_I : I\in\mathcal{D}\}$ is ***uniformly adapted*** to $\mathcal{D}$ if for each $I\in\mathcal{D}$, φ_I is adapted to I, and each φ_I is obtained from a single fixed function $\varphi\in C^\infty(\mathbb{R})$ as follows

$$\varphi_I(x) = \frac{1}{\sqrt{|I|}}\varphi\left(\frac{x-c(I)}{|I|}\right), \tag{1.6}$$

where $c(I)$ is the center of I and φ satisfies

$$\int_0^\infty \left|\hat\varphi(t\xi)\right|^2\frac{dt}{t} = 1, \qquad \text{for all } \xi\neq 0.$$

Most typically, the notation φ_I will be used for a function adapted to I.

A collection of functions $\{w_I : I\in\mathcal{D}\}$ **is called a *wavelet basis*** if the collection is uniformly adapted to $\mathcal{D}$, and it is an orthonormal basis for $L^2(\mathbb{R})$. It is very easy to see that necessarily, w has cancellation. The examples of wavelet bases that will be important for us are $L^2(\mathbb{R})$ normalized functions adapted to an interval.

Paraproduct operators are constructed from rank one operators $f\mapsto\langle f,\varphi\rangle\phi$. A paraproduct is, in its simplest manifestation, of the form

$$\mathrm{B}(f_1,f_2) \stackrel{\text{def}}{=} \sum_{I\in\mathcal{D}} |I|^{-1/2}\varphi_{3,I}\prod_{j=1}^{2}\langle f_j,\varphi_{j,I}\rangle_{L^2(\mathbb{R})}.$$

Here, the functions $\varphi_{j,I}$, $j = 1, 2, 3$, are adapted to I, and $\langle f, g\rangle_{L^2(\mathbb{R})} = \int_{\mathbb{R}} f(x)g(x)dx$. Two of these three functions are assumed to have cancellation, and in particular we will always assume that $\varphi_{1,I}$ has cancellation. We are concerned with extensions of the classical conditions for this operator to be bounded on $L^2(\mathbb{R})$. For these more general paraproducts there is the following well-known extension of (1.4). See [14, 20] for proofs.

1.7. **Theorem** ([14, 20]). *If $\{\varphi_{1,I}\}$ and at least one collection of $\{\varphi_{j,I}\}$ $j = 2, 3$ have cancellation and are adapted to $\mathcal{D}$, then*

$$\|\mathrm{B}(f_1, \cdot)\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})} \lesssim \|f_1\|_{\mathrm{BMO}(\mathbb{R})}. \tag{1.8}$$

If all three collections $\{\varphi_{j,I}\}$ are uniformly adapted to $\mathcal{D}$, then the reverse inequality holds.

The BMO norm is explicitly given by

$$\|f\|_{\mathrm{BMO}(\mathbb{R})} \stackrel{\mathrm{def}}{=} \sup_U \left[|U|^{-1} \sum_{I\subset U} |\langle f, \varphi_I\rangle_{L^2(\mathbb{R})}|^2\right]^{1/2}. \tag{1.9}$$

Here, the supremum is formed over all intervals U and $\{\varphi_I\}$ is uniformly adapted to $\mathcal{D}$. We note that the definition is independent of the choice of $\{\varphi_I\}$ and the dyadic grid $\mathcal{D}$.

In this paper we are principally concerned with the Schatten norms of paraproducts $\mathrm{B}(f_1, f_2)$ and their multiparameter versions. Recall that the Schatten norm of an operator is given by a ℓ^p sum of its singular values, see Section 2 for the precise definition and more information about these norms. Much like the case of boundedness, membership of the paraproduct in a Schatten class can be characterized in terms of the function f_1.

We first recall some previous closely related results on Schatten class of related operators. In the case of continuous paraproducts, Janson and Peetre showed in [10] that membership in a Schatten class is equivalent to the symbol belonging to the Besov space. Their method of proof was very much Fourier analytic by viewing the continuous paraproduct as a certain multiplier on the Fourier side and then decomposing the operator in an appropriate manner. See also Pott and Smith, [24] for the dyadic version. Chao and Peng [3] showed that the one parameter paraproducts arising from (d-dyadic) martingale transforms are bounded if and only if the symbol belongs to the dyadic Besov space, whose definition is given below. The proof is very computational, and takes advantage of the notion of "nearly weakly orthonormal sequences" introduced by Rochberg and Semmes, [25]. In fact, in both [3, 10] it is shown that more generally the commutators with singular integral operators (or martingale transforms) belong to a certain Schatten class if and only if the symbol belongs to the appropriate Besov space.

Membership of the paraproduct in the Schatten class is related to smoothness on the symbol f_1, and this is governed by the symbol belonging to a certain Besov space. We can define the dyadic Besov spaces $\mathbb{B}^p_{\mathrm{dyadic}}(\mathbb{R})$ as the set of $f \in L^1_{loc}(\mathbb{R})$ such that $\|f\|_{\mathbb{B}^p_{\mathrm{dyadic}}(\mathbb{R})} < \infty$, where

$$\|f\|_{\mathbb{B}^p_{\mathrm{dyadic}}(\mathbb{R})} \stackrel{\mathrm{def}}{=} \left[\sum_{I\in\mathcal{D}} \left[|I|^{-1/2} |\langle f, h^0_I\rangle_{L^2(\mathbb{R})}|\right]^p\right]^{1/p}, \qquad 0 < p < \infty. \tag{1.10}$$

Also, the Besov space $\mathbb{B}^p(\mathbb{R})$ is the set of Schwartz distributions f such that $\|f\|_{\mathbb{B}^p(\mathbb{R})} < \infty$, where

$$\|f\|_{\mathbb{B}^p(\mathbb{R})} \stackrel{\mathrm{def}}{=} \left[\sum_{I\in\mathcal{D}} \left[|I|^{-1/2} |\langle f, \varphi_I\rangle_{L^2(\mathbb{R})}|\right]^p\right]^{1/p}, \qquad 0 < p < \infty \tag{1.11}$$

with $\{\varphi_I\}$ is uniformly adapted to $\mathcal{D}$. We note that the definition of $\mathbb{B}^p(\mathbb{R})$ is independent of the choice of $\{\varphi_I\}$ and the dyadic grid $\mathcal{D}$. That is, suppose there is another family $\{\psi_I\}$ that is uniformly adapted to dyadic system $\mathcal{D}'$. Then

$$\|f\|_{\mathbb{B}^p(\mathbb{R})} \simeq \left[\sum_{I\in\mathcal{D}'} \left[|I|^{-1/2}|\langle f,\psi_I\rangle_{L^2(\mathbb{R})}|\right]^p\right]^{1/p}, \qquad 0<p<\infty.$$

We refer the readers to Proposition 3.1 for a proof.

Our main result is then the following theorem, giving an extension and refinement of the boundedness of paraproducts in one parameter given in Theorem 1.7.

1.12. **Theorem** (Main Result 1)**.** *Assuming only the collection of functions $\{\varphi_{1,I} : I \in \mathcal{D}\}$ has cancellation, we have the estimate*

$$\|\mathrm{B}(f_1,\cdot)\|_{\mathbb{S}^p} \lesssim \|f_1\|_{\mathbb{B}^p(\mathbb{R})}, \qquad 0<p<\infty. \tag{1.13}$$

If each of the collections of functions $\{\varphi_{j,I}\}$, $j=1,2,3$, is uniformly adapted to $\mathcal{D}$, then the reverse inequality holds.

When taking the collections of functions $\{\varphi_{j,I}\}$ to be the Haar functions (1.1) and the normalized indicators (1.2), we have the following immediate corollary in the case of Haar paraproducts.

1.14. **Theorem.** *In the Haar case we have*

$$\left\|\sum_{I\in\mathcal{D}} \frac{\langle f_1,h_I^0\rangle_{L^2(\mathbb{R})}}{\sqrt{|I|}} h_I^\epsilon\otimes h_I^\delta\right\|_{\mathbb{S}^p} \simeq \|f_1\|_{\mathbb{B}^p_{\text{dyadic}}(\mathbb{R})}, \qquad 0<p<\infty,\ \{\epsilon,\delta\}\neq\{1,1\}. \tag{1.15}$$

We note that a variation of Theorem 1.14 for dyadic martingale transforms was previously studied by Chao and Peng [3]. Additionally, an alternate proof of Theorem 1.14 was given by Pott and Smith [24, Theorem 2.1]. We provide a different proof compared to [3, 24].

We will now use these ideas to extend the results from one-parameter to the multi-parameter setting.

Multi-parameter paraproducts. We use many of the ideas from the previous section to form the tensor product basis in $L^2(\mathbb{R}^n) = L^2(\mathbb{R}\times\cdots\times\mathbb{R})$. Let $\mathcal{R} \stackrel{\text{def}}{=} \mathcal{D}\times\cdots\times\mathcal{D}$ be the collection of dyadic rectangles in $\mathbb{R}^n$.

• Dyadic version of paraproduct:

For a rectangle $R = R_1\times\cdots\times R_n \in \mathcal{R}$ and for a choice of $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)\in\{0,1\}^n$,

$$h_R^\varepsilon(x_1,\ldots,x_n) \stackrel{\text{def}}{=} \prod_{j=1}^n h_{R_j}^{\varepsilon_j}(x_j), \tag{1.16}$$

where $h_{R_j}^{\varepsilon_j}$ is given by either (1.1) or (1.2) depending on if $\varepsilon_j = 0$ or $\varepsilon_j = 1$. For $\varepsilon = \vec{0} = (0,\ldots,0)$, we denote $h_R^{\vec{0}}(x_1,\ldots,x_n)$ by $h_R(x_1,\ldots,x_n)$ for simplicity.

Define $E_n \stackrel{\text{def}}{=} \{0,1\}^n\setminus\vec{1}$, where $\vec{1} = (1,\ldots,1)$. Note that the cardinality of the set E_n is 2^n-1. We then have that $\{h_R^\varepsilon : R\in\mathcal{R}, \varepsilon\in E_n\}$ is the product Haar basis for $L^2(\mathbb{R}^n)$.

The n parameter Haar paraproducts are

$$\mathrm{B}_{\text{Haar}}(f_1,f_2) \stackrel{\text{def}}{=} \sum_{R\in\mathcal{R}} \frac{\langle f_1,h_R\rangle_{L^2(\mathbb{R}^n)}}{\sqrt{|R|}} \langle f_2,h_R^\varepsilon\rangle_{L^2(\mathbb{R}^n)}\, h_R^\delta. \tag{1.17}$$

The essential restriction to place on the two choices of $\varepsilon, \delta \in \{0,1\}^n$ is that

in no coordinate j do we have $\varepsilon_j = \delta_j = 1$.

Observe that this condition permits a wide variety of paraproducts, with most having no proper analog as compared to the one dimensional case.

• Continuous version of paraproduct:

Let us say that a collection $\{\varphi_R : R \in \mathcal{R}\}$ *is adapted to a rectangle* R if and only if $\varphi_R(x_1, \dots, x_n) = \prod_{j=1}^n \varphi_{j,R_j}(x_j)$, with each $\{\varphi_{j,R_j}\}$ adapted to $\mathcal{D}$. We say that $\{\varphi_R\}$ has *cancellation in the jth coordinate* if and only if $\{\varphi_{j,R_j}\}$ has cancellation. The collection $\{\varphi_R\}$ is *uniformly adapted to* $\mathcal{R}$ if and only if each $\{\varphi_{j,R_j}\}$, $j = 1, \dots, n$, is uniformly adapted to $\mathcal{D}$.

Paraproducts are then defined as follows:

$$\mathrm{B}(f_1, f_2) \stackrel{\text{def}}{=} \sum_{R \in \mathcal{R}} \frac{\varphi_R^{(3)}}{\sqrt{|R|}} \prod_{j=1}^{2} \langle f_j, \varphi_R^{(j)} \rangle_{L^2(\mathbb{R}^n)}, \tag{1.18}$$

where the functions $\varphi_R^{(j)}$ are adapted to R for $j = 1, 2, 3$. The construction of a smooth wavelet basis in $L^2(\mathbb{R}^n)$ is similar and standard. For the details we omit here.

The boundedness of the multi-parameter paraproducts was first studied by Journé when considering the $T(1)$ Theorem in the product setting, [9]. These results were later studied further by Muscalu, Pipher, Tao and Thiele in the following papers [18, 19] which showed the richness of the paraproduct structures in the multiparameter setting. One should also see the article by Lacey and Metcalfe, [14]. The following general theorem on the boundedness in $L^2(\mathbb{R}^n)$ of the Haar paraproducts and more general paraproducts from a wavelet basis is then given by:

1.19. **Theorem** ([9, 14, 18, 19]). *For the multi-parameter Haar paraproducts, we have the following estimate:*

$$\|\mathrm{B}_{\mathrm{Haar}}(f_1, \cdot)\|_{L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)} \lesssim \|f_1\|_{\mathrm{BMO}_{\mathrm{dyadic}}(\mathbb{R} \times \cdots \times \mathbb{R})}\,. \tag{1.20}$$

More generally, assume that for both coordinates $j = 1, 2$ there is a choice of $k \in \{2, 3\}$ for which $\varphi_R^{(1)}$ and $\varphi_R^{(k)}$ have cancellation in the jth coordinate. Then, for the paraproducts as in (1.18)*, we have the inequality:*

$$\|\mathrm{B}(f_1, \cdot)\|_{L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)} \lesssim \|f_1\|_{\mathrm{BMO}(\mathbb{R} \times \cdots \times \mathbb{R})}\,. \tag{1.21}$$

There are two points to make about this last inequality. The first is that the space $\mathrm{BMO}(\mathbb{R} \times \cdots \times \mathbb{R})$ is the product BMO space studied by S.-Y. A. Chang and R. Fefferman [4], and the BMO norm is given explicitly by

$$\|f\|_{\mathrm{BMO}(\mathbb{R} \times \cdots \times \mathbb{R})} \stackrel{\text{def}}{=} \sup_U \left[|U|^{-1} \sum_{R \subset U} |\langle f, w_R \rangle_{L^2(\mathbb{R}^n)}|^2 \right]^{1/2}. \tag{1.22}$$

Here, the supremum is formed over *open* sets U and $\{w_R\}$ is a product wavelet basis. Replacing the wavelet basis by the Haar basis, we have dyadic Chang–Fefferman BMO, [4]. The second is that we are not asserting the equivalence of norms. Indeed, for a 'degenerate' n parameter paraproduct, the equivalence of norms is not so clear. There are essentially two distinct cases. The first case, with the greatest similarity to the one parameter case, is where we have for example, $\{\varphi_R^{(2)}\}$ has cancellation in all coordinates. The second case with no proper analog in the one variable setting is, for instance, $\{\varphi_R^{(2)}\}$ has cancellation in one set of coordinates while $\{\varphi_R^{(3)}\}$ has cancellation in a complimentary set of coordinates.

Similar to the one-parameter case, with the Haar basis we can define the dyadic (product) Besov spaces in $\mathbb{R}^n$ as

$$\|f\|_{\mathbb{B}^p_{\text{dyadic}}(\mathbb{R}\times\cdots\times\mathbb{R})} \stackrel{\text{def}}{=} \left[\sum_{R\in\mathcal{R}} \left[|R|^{-1/2}|\langle f, h_R\rangle_{L^2(\mathbb{R}^n)}|\right]^p\right]^{1/p}, \qquad 0<p<\infty. \tag{1.23}$$

For the tensor product wavelet basis, we define the (product) Besov spaces in $\mathbb{R}^n$ as

$$\|f\|_{\mathbb{B}^p(\mathbb{R}\times\cdots\times\mathbb{R})} \stackrel{\text{def}}{=} \left[\sum_{R\in\mathcal{R}} \left[|R|^{-1/2}|\langle f, \varphi_R\rangle_{L^2(\mathbb{R}^n)}|\right]^p\right]^{1/p}, \qquad 0<p<\infty. \tag{1.24}$$

Again, we will see that this definition does not depend upon the choice of wavelet basis.

Our principal estimate, giving an extension of Theorem 1.19 to the multi-parameter setting is given next, especially for the continuous version. For simplicity, this theorem is stated in the case of two parameters. The correct statement of the general multi-parameter version can be obtained from Theorem 5.17 below.

1.25. **Theorem** (Main Result 2)**.** *Assume that $\{\phi_{1,R}\}$ has cancellation for both coordinates $j=1,2$, while $\{\phi_{2,R}\}$ and $\{\phi_{3,R}\}$ have the property that: if $\{\phi_{2,R}\}$ has cancellation in coordinate j then $\{\phi_{3,R}\}$ does not and vice versa, for $j=1,2$. Then,*

$$\|\mathrm{B}(f_1,\cdot)\|_{\mathbb{S}^p} \lesssim \|f_1\|_{\mathbb{B}^p(\mathbb{R}\times\mathbb{R})}, \qquad 0<p<\infty. \tag{1.26}$$

If all the collections $\{\phi_{j,R}\}$ are uniformly adapted to $\mathcal{R}$, then the reverse inequality holds.

Notice that when neither collection of functions have cancellation, the corresponding operator is *not* bounded for general functions in BMO. This is indicative of the well known fact that the result on Schatten norms is not as delicate as the criteria for being bounded.

Parallel to the above result, a corollary of Theorem 1.25 in the dyadic version is as follows, which is an extension of Theorem 1.14.

1.27. **Theorem.** *For any bounded n parameter dyadic paraproduct we have the equivalence*

$$\|\mathrm{B}_{\text{Haar}}(f_1,\cdot)\|_{\mathbb{S}^p} \simeq \|f_1\|_{\mathbb{B}^p_{\text{dyadic}}(\mathbb{R}\times\cdots\times\mathbb{R})}, \qquad 0<p<\infty. \tag{1.28}$$

Note that a variation of Theorem 1.27 for little Hankel operators was given in Pott and Smith [24, Theorem 5.1]. For the sake of notational simplicity, we state and prove the result in the bi-parameter setting. The argument for multi-parameter setting follows similarly.

In Section 2 we collect some properties of Schatten norms. In Section 3 we show that the Besov norms are independent of the choice of wavelet basis and establish the dyadic structure for the Besov spaces. In Section 4 we give a proof of Theorems 1.12 and 1.14. We first will handle the case of Haar paraproducts since that will turn out to be a model for the more general case of paraproducts built from wavelet bases. In Section 5 we give the proofs of Theorems 1.25 and 1.27, which are based upon the ideas appearing in Section 4 but will be complicated by additional notation necessary to handle the multi-parameter paraproducts.

2. Basic Properties of Schatten Norms

Let $\mathcal{H}$ be a separable Hilbert space. Recall that for elements $\varphi,\phi\in\mathcal{H}$ the operator denoted by $\varphi\otimes\phi$ takes an $f\in\mathcal{H}$ to $\phi\langle f,\varphi\rangle_{\mathcal{H}}$.

A compact operator $\mathrm{T} : \mathcal{H} \to \mathcal{H}$ has a decomposition

$$\mathrm{T} = \sum_n \lambda_n \, \mathrm{e}_n \otimes \mathrm{f}_n \tag{2.1}$$

in which $\lambda_n \in \mathbb{R}$, and $\{\mathrm{e}_n\}$ and $\{\mathrm{f}_n\}$ are orthonormal sequences in $\mathcal{H}$ (compactness implies that $|\lambda_n| \to 0$). The Schatten norm is then

$$\|\mathrm{T}\|_{\mathbb{S}^p} \stackrel{\text{def}}{=} \Bigl[\sum_n |\lambda_n|^p\Bigr]^{1/p}, \qquad 0 < p < \infty. \tag{2.2}$$

This is an actual norm for $1 \le p < \infty$, while for $0 < p < 1$ it is not. The trace class operators are the class $\mathbb{S}^1$ and the Hilbert–Schmidt operators are the class $\mathbb{S}^2$. Part of the interest in these classes are that the class $\mathbb{S}^1$ is in natural duality with $\mathcal{L}(\mathcal{H})$, the space of bounded operators on $\mathcal{H}$ and the class $\mathbb{S}^2$ has a simple way to compute the norm using any orthonormal basis. It is clear that $\|\mathrm{T}\|_{\mathbb{S}^p} = \|\mathrm{T}^*\|_{\mathbb{S}^p}$.

Define a collection of positive numbers

$$\mathcal{T} \stackrel{\text{def}}{=} \left\{ \Bigl[\sum_n \|\mathrm{T}\,\mathrm{e}_n\|_{\mathcal{H}}^p\Bigr]^{1/p} \;:\; \{\mathrm{e}_n\} \text{ is an orthonormal basis in } \mathcal{H} \right\}.$$

Then it is the case that

$$\|\mathrm{T}\|_{\mathbb{S}^p} = \inf \mathcal{T}, \qquad 0 < p \le 2, \tag{2.3}$$

$$\|\mathrm{T}\|_{\mathbb{S}^p} = \sup \mathcal{T}, \qquad 2 \le p < \infty. \tag{2.4}$$

For $1 \le p < \infty$, the Schatten norm obeys the triangle inequality:

$$\|\mathrm{S} + \mathrm{T}\|_{\mathbb{S}^p} \le \|\mathrm{S}\|_{\mathbb{S}^p} + \|\mathrm{T}\|_{\mathbb{S}^p}. \tag{2.5}$$

For $0 < p < 1$ this is no longer the case. There is the following quasi-triangle inequality, linked to the subadditivity of $x \mapsto x^p$.

$$\|\mathrm{S} + \mathrm{T}\|_{\mathbb{S}^p}^p \le \|\mathrm{S}\|_{\mathbb{S}^p}^p + \|\mathrm{T}\|_{\mathbb{S}^p}^p. \tag{2.6}$$

In the converse direction, there is a proposition below.

2.7. **Proposition.** *Suppose that* T *is an operator from* $\mathcal{H}$ *to itself, and that* P *is a contraction, then*

$$\|\mathrm{T}\,\mathrm{P}\|_{\mathbb{S}^p}, \|\mathrm{P}\,\mathrm{T}\|_{\mathbb{S}^p} \le \|\mathrm{T}\|_{\mathbb{S}^p}, \qquad 0 < p < \infty.$$

Proof. For $\|\mathrm{P}\,\mathrm{T}\|_{\mathbb{S}^p}$, this follows from the characterization of the Schatten norm in terms of either an infimum or supremum, see (2.3) and (2.4). Combining this observation with the equivalence of the Schatten norms for dual operators proves the proposition. □

We also need an inequality for the Schatten norms of a $m \times n$ matrix $\mathrm{A} = (a_{i,j})$.

2.8. **Proposition.** *We have the inequality*

$$\|\mathrm{A}\|_{\mathbb{S}^p} \le (mn)^{\delta(p)} \Bigl[\sum_{i,j=1}^{m,n} |a_{i,j}|^p\Bigr]^{1/p}, \qquad 0 < p < \infty. \tag{2.9}$$

Here, $\delta(p) = \max\left(0, \tfrac{1}{2} - \tfrac{1}{p}\right)$.

Proof. The case of $0 < p < 2$ is clear. Appealing to (2.3), we use the standard basis e_k $1 \le k \le m$, so that

$$\|\mathrm{A}\|_{\mathbb{S}^p}^p \le \sum_{k=1}^m \|\mathrm{A}\,\mathrm{e}_k\|^p = \sum_{k=1}^m \left[\sum_{i=1}^n |a_{i,k}|^2\right]^{p/2} \le \sum_{k=1}^m \sum_{i=1}^n |a_{i,k}|^p.$$

This case is finished.

To conclude the proof for $2 < p < \infty$, observe that the norms decrease in p, hence

$$\|\mathrm{A}\|_{\mathbb{S}^p} \le \|\mathrm{A}\|_{\mathbb{S}^2} = \left[\sum_{i=1}^n \sum_{j=1}^m |a_{i,j}|^2\right]^{1/2} \le (mn)^{\delta(p)} \left[\sum_{i=1}^n \sum_{j=1}^m |a_{i,j}|^p\right]^{1/p}.$$

The proof is complete. □

More comments about the Schatten norms and nearly weakly orthogonal (NWO) functions are made in Section 4.

3. Besov space and its dyadic structure

3.1. **One-parameter.** We are interested in those results that relate the Schatten norms to Besov spaces of corresponding symbols. The functions $\{\varphi_I : I \in \mathcal{D}\}$ will be a wavelet basis for $L^2(\mathbb{R})$ with the function φ being continuous and rapidly decreasing. In the first definition, (1.11), one may be concerned that the definition depends upon the choice of function φ. There is a straight forward lemma which shows this is not the case.

3.1. **Proposition.** *Let φ and ϕ be two distinct wavelets, generating wavelet bases $\{\varphi_I\}$ and $\{\phi_I\}$, respectively. We have the equivalence*

$$\sum_{I\in\mathcal{D}} \left[|I|^{-1/2} |\langle f, \varphi_I\rangle_{L^2(\mathbb{R})}|\right]^p \simeq \sum_{I\in\mathcal{D}} \left[|I|^{-1/2} |\langle f, \phi_I\rangle_{L^2(\mathbb{R})}|\right]^p.$$

This is valid for any function f for which either side is finite, and implied constants depend only on the choice of $0 < p < \infty$.

Proof. This is a standard argument for wavelet characterisation of Besov spaces. We also note that the two wavelet bases need not be associated with the same dyadic grid $\mathcal{D}$, which could be different grids. The key steps are to use the wavelet expansion, almost orthogonality estimates and then the Plancherel–Pólya type inequality. See for example the standard argument in [8]. □

We have the standard characterization of the Besov space $\mathbb{B}^p(\mathbb{R})$, $1 \le p < \infty$, as follows (see for example [26, p. 242]), which also reflect that the definition of $\mathbb{B}^p(\mathbb{R})$ is independent of the choice of the dyadic grids, and the associated wavelet basis.

3.2. **Proposition.** *Let $b \in L^1_{\mathrm{loc}}(\mathbb{R})$ and $0 < p < \infty$. Then we say that b belongs to the Besov space $B^p(\mathbb{R})$ if*

$$\|b\|_{B^p(\mathbb{R})} = \left(\int_{\mathbb{R}}\int_{\mathbb{R}} \frac{|b(x) - b(y)|^p}{|x-y|^2} dydx\right)^{\frac{1}{p}} < \infty.$$

Further, if $1 \le p < \infty$, then we have $B^p(\mathbb{R}) = \mathbb{B}^p(\mathbb{R})$ with equivalence of norms.

In the dyadic case the same definition applies, defining the Besov space $\mathbb{B}^p_{\text{dyadic}}(\mathbb{R})$ as follows, with the functions φ_I replaced by Haar functions.

$$\|f\|_{\mathbb{B}^p_{\text{dyadic}}(\mathbb{R})} \stackrel{\text{def}}{=} \left[\sum_{I\in\mathcal{D}} \left[|I|^{-1/2}|\langle f, h_I\rangle_{L^2(\mathbb{R})}|\right]^p\right]^{1/p}, \qquad 0<p<\infty. \tag{3.3}$$

Based on our recent work [12, 13], it is direct to see the following dyadic structure of the Besov space.

3.4. **Proposition.** *Suppose $1 \le p < \infty$. There are two choices of dyadic grids $\mathcal{D}^0$ and $\mathcal{D}^1$ for which we have $\mathbb{B}^p_{\text{dyadic},0}(\mathbb{R}) \cap \mathbb{B}^p_{\text{dyadic},1}(\mathbb{R}) = \mathbb{B}^p(\mathbb{R})$, where $\mathbb{B}^p_{\text{dyadic},i}(\mathbb{R})$ is the dyadic Besov space associated with the dyadic grid $\mathcal{D}^i$, $i = 0, 1$. Moreover, we have*

$$\|b\|_{\mathbb{B}^p(\mathbb{R})} \simeq \|b\|_{\mathbb{B}^p_{\text{dyadic},0}(\mathbb{R})} + \|b\|_{\mathbb{B}^p_{\text{dyadic},1}(\mathbb{R})}.$$

3.2. **Product setting.** Recall that the (product) Besov space on $\mathbb{R}^n = \mathbb{R}\times\cdots\times\mathbb{R}$ is defined by

$$\|f\|_{\mathbb{B}^p(\mathbb{R}\times\cdots\times\mathbb{R})} \stackrel{\text{def}}{=} \left[\sum_{R\in\mathcal{R}} \left[|R|^{-1/2}|\langle f, \varphi_R\rangle_{L^2(\mathbb{R}^n)}|\right]^p\right]^{1/p}, \qquad 0<p<\infty. \tag{3.5}$$

3.6. **Proposition.** *Let φ and ϕ be two distinct wavelets, generating wavelet bases $\{\varphi_R\}$ and $\{\phi_R\}$, respectively. We have the equivalence*

$$\sum_{R\in\mathcal{R}} \left[|R|^{-1/2}|\langle f, \varphi_R\rangle_{L^2(\mathbb{R}^n)}|\right]^p \simeq \sum_{R\in\mathcal{R}} \left[|R|^{-1/2}|\langle f, \phi_R\rangle_{L^2(\mathbb{R}^n)}|\right]^p.$$

This is valid for any function f for which either side is finite, and implied constants depend only on the choice of $0 < p < \infty$. The two wavelet bases need not be associated with the same dyadic grid $\mathcal{R}$.

Proof. Again, the key step is to use the wavelet expansion, almost orthogonality estimates, and then the Plancherel–Pólya type inequality. See for example the standard argument for the product Hardy and BMO spaces in [8, Section 4], which can be easily adapted to the product Besov space setting. □

We also introduce the following definition of product Besov space via difference. For notational simplicity, for $j = 1, \ldots, n$, we let

$$\triangle^{(j)}_{y_j} b(x_1, \ldots, x_n) = b(x_1, \ldots, x_n) - b(x_1, \ldots, x_{j-1}, y_j, x_{j+1}, \ldots x_n).$$

To begin with, we first introduce the following definition.

3.7. **Definition.** *Let $b \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $0 < p < \infty$. Then we say that b belongs to the product Besov space $B^p(\mathbb{R}\times\cdots\times\mathbb{R})$ if*

$$\|b\|_{B^p(\mathbb{R}\times\cdots\times\mathbb{R})} = \left(\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} \frac{\left|\triangle^{(1)}_{y_1}\cdots\triangle^{(n)}_{y_n} b(x_1,\ldots,x_n)\right|^p}{\prod_{j=1}^n |x_j-y_j|^2} dy_1dx_1\cdots dy_ndx_n\right)^{\frac{1}{p}} < \infty.$$

Next, we first point out that parallel to the classical setting, we have the equivalence of $B^p(\mathbb{R}\times\cdots\times\mathbb{R})$ and $\mathbb{B}^p(\mathbb{R}\times\cdots\times\mathbb{R})$ when $p \ge 1$. That is,

3.8. **Proposition.** *Suppose $1 \le p < \infty$. We have $B^p(\mathbb{R}\times\cdots\times\mathbb{R}) = \mathbb{B}^p(\mathbb{R}\times\cdots\times\mathbb{R})$ with equivalence of norms.*

Proof. By using the reproducing formula and the almost orthogonality estimate in the tensor product setting (see for example [8, Section 4]) and Proposition 3.1, one obtains the above proposition. Details are omitted. □

To establish the dyadic structure of the product Besov space, we only provide details for $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. The n-parameter setting follows with appropriate modifications.

We establish the following natural containment of the Besov space and the dyadic Besov space.

3.9. **Proposition.** *Suppose $1 \le p < \infty$. We have $B^p(\mathbb{R}\times\mathbb{R}) \subset \mathbb{B}^p_{\rm dyadic}(\mathbb{R}\times\mathbb{R})$, with estimate*

$$\|b\|_{\mathbb{B}^p_{\rm dyadic}(\mathbb{R}\times\mathbb{R})} \lesssim \|b\|_{B^p(\mathbb{R}\times\mathbb{R})}.$$

Proof. The key inequalities here are specific to a choice of dyadic rectangle R. Let $R = I \times J$, where both I and J are dyadic intervals. Below, let $I' = I + 2|I|$ and $J' = J + 2|J|$. Then it is clear that $R' = I' \times J'$ is another dyadic rectangle with volume comparable to that of R. Hence, by the cancellation condition of h_R, we have

$$\begin{aligned}
&\left|\int_R b(x_1,x_2)h_R(x_1,x_2)dx_1dx_2\right| |R|^{-\frac12}\\
&\lesssim \inf_{(y_1,y_2)\in R'}\left|\int_R \Big(b(x_1,x_2)-b(x_1,y_2)-b(y_1,x_2)+b(y_1,y_2)\Big)h_R(x_1,x_2)dx_1dx_2\right| |R|^{-\frac12}\\
&\lesssim |R|^{-\frac32}\int_{R'}\left|\int_R \triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\ h_R(x_1,x_2)dx_1dx_2\right| dy_1dy_2\\
&\lesssim |R|^{-\frac32}\bigg(\int_{R'}\int_R \left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\ \right|^p dy_1dy_2dx_1dx_2\bigg)^{\frac1p}\bigg(\int_{R'}\int_R |h_R(x_1,x_2)|^{p'}dx_1dx_2dy_1dy_2\bigg)^{\frac1{p'}}\\
&\lesssim \bigg(\int_{R'}\int_R \left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\ \right|^p dy_1dy_2dx_1dx_2\ \frac{1}{|R|^2}\bigg)^{\frac1p}\\
&\lesssim \bigg(\int_{R'}\int_R \frac{\left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\ \right|^p}{|x_1-y_1|^2|x_2-y_2|^2} dy_1dy_2dx_1dx_2\bigg)^{\frac1p},
\end{aligned}$$

where in the third inequality we use Hölder's inequality and this is where we need $1 \le p < \infty$. Note also that if $p = 1$, then $p' = \infty$ and the second factor in the right-hand side of the third inequality will become $\|h_R\|_{L^\infty}$.

Take the power p on both sides, and sum over all $R, R' \in \mathcal{D}\times\mathcal{D}$, to get an expression dominated by $\|b\|^p_{B^p(\mathbb{R}\times\mathbb{R})}$. □

Moreover, we also have a weaker version of the reverse containment. That is, we will need another set of dyadic intervals. We take two dyadic grids $\mathcal{D}^0$ and $\mathcal{D}^1$ so that for all intervals I there is a $Q \in \mathcal{D}^0 \cup \mathcal{D}^1$ with

$$I \subset Q \subset 4I. \tag{3.10}$$

One option is that $\mathcal{D}^0$ is the standard dyadic system in $\mathbb{R}$ and $\mathcal{D}^1$ is the 'one-third shift' of $\mathcal{D}^0$, see for example [16].

3.11. **Proposition.** *Suppose $1 \le p < \infty$. There are two choices of grids $\mathcal{D}^0$ and $\mathcal{D}^1$ for which we have*

$$B^p(\mathbb{R}\times\mathbb{R}) = \mathbb{B}^{p,(0,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R}) \cap \mathbb{B}^{p,(0,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R}) \cap \mathbb{B}^{p,(1,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R}) \cap \mathbb{B}^{p,(1,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R}),$$

where $\mathbb{B}^{p,(i,j)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})$ *is the dyadic Besov space associated with* $\mathcal{D}^i\times\mathcal{D}^j$ *for* $i,j\in\{0,1\}$. *Moreover, we have*

$$\|b\|_{B^p(\mathbb{R}\times\mathbb{R})}\simeq\|b\|_{\mathbb{B}^{p,(0,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})}+\|b\|_{\mathbb{B}^{p,(0,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})}+\|b\|_{\mathbb{B}^{p,(1,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})}+\|b\|_{\mathbb{B}^{p,(1,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})}.$$

Proof. To begin with, we first note that

$$\begin{aligned}
&\|b\|^p_{B^p(\mathbb{R}\times\mathbb{R})}\\
&=\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\frac{\left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\right|^p}{|x_1-y_1|^2|x_2-y_2|^2}dy_1dy_2dx_1dx_2\\
&\lesssim\sum_{\substack{I\in\mathcal{D}\\J\in\mathcal{D}}}\int_{I\times J}\int_{\substack{\{y_1\in\mathbb{R}:2^a|I|<|x_1-y_1|\leq 2^{a+1}|I|\}\\ \{y_2\in\mathbb{R}:2^a|J|<|x_2-y_2|\leq 2^{a+1}|J|\}}}\frac{\left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\right|^p}{|x_1-y_1|^2|x_2-y_2|^2}dy_1dy_2dx_1dx_2\\
&\lesssim\sum_{\substack{I\in\mathcal{D}\\J\in\mathcal{D}}}\frac{1}{|R|^2}\int_{I\times J}\int_{\substack{\{y_1\in\mathbb{R}:2^a|I|<|x_1-y_1|\leq 2^{a+1}|I|\}\\ \{y_2\in\mathbb{R}:2^a|J|<|x_2-y_2|\leq 2^{a+1}|J|\}}}\left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\right|^pdy_1dy_2dx_1dx_2\\
&\lesssim\sum_{\substack{I\in\mathcal{D}\\J\in\mathcal{D}}}\frac{1}{|R|^2}\sum_{m_1=n_1}^{2n_1-1}\sum_{m_2=n_2}^{2n_2-1}\\
&\qquad\int_{I\times J}\int_{\substack{\{y_1\in\mathbb{R}:2^a\frac{m_1}{n_1}|I|<|x_1-y_1|\leq 2^a\frac{m_1+1}{n_1}|I|\}\\ \{y_2\in\mathbb{R}:2^a\frac{m_2}{n_2}|J|<|x_2-y_2|\leq 2^a\frac{m_2+1}{n_2}|J|\}}}\left|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)\right|^pdy_1dy_2dx_1dx_2.
\end{aligned}\tag{3.12}$$

The second integral above is over a symmetric interval. Consider the two intervals

$$I,\qquad I+2^a|I|[m_1/n_1,(m_1+1)/n_1],\qquad n_1\leq m_1<2n_1.\tag{3.13}$$

Now, we choose $a=5$, and $n_1=1000$, so the second interval is smaller in length, but still comparable to I in length. And, they are separated by a distance approximately $2^a|I|$. By (3.10), we can choose a so that there is a dyadic $I'\in\mathcal{D}^0\cup\mathcal{D}^1$ which contains both intervals above, and moreover I is contained in the left half of I', and $I+2^a|I|[m_1/n_1,(m_1+1)/n_1]$ the right half. We can argue similarly for $I-2^a|I|[m_1/n_1,(m_1+1)/n_1]$, as well as for the dyadic interval J and the parameters m_2 and n_2. Below we continue with $R=I\times J$ and $R'=I'\times J'$.

Next, observe the following identity:

$$\begin{aligned}
\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2}b(x_1,x_2)&=\big(b(x_1,x_2)-E^{(1,0)}b(x_2)-E^{(0,1)}b(x_1)+E^{(1,1)}b\big)\\
&\quad-\big(b(y_1,x_2)-E^{(1,0)}b(x_2)-E^{(0,1)}b(y_1)+E^{(1,1)}b\big)\\
&\quad-\big(b(x_1,y_2)-E^{(1,0)}b(y_2)-E^{(0,1)}b(x_1)+E^{(1,1)}b\big)\\
&\quad+\big(b(y_1,y_2)-E^{(1,0)}b(y_2)-E^{(0,1)}b(y_1)+E^{(1,1)}b\big)\\
&=:B_1(x_1,x_2)+B_2(y_1,x_2)+B_3(x_1,y_2)+B_4(y_1,y_2),
\end{aligned}$$

where

$$E^{(1,0)}b(\cdot)=\frac{1}{|I'|}\int_{I'}b(z_1,\cdot)dz_1,\qquad E^{(0,1)}b(\cdot)=\frac{1}{|J'|}\int_{J'}b(\cdot,z_2)dz_2,$$
$$E^{(1,1)}b=\frac{1}{|R'|}\int_{R'}b(z_1,z_2)dz_1dz_2.$$

In particular, for the main term in (3.12), with fixed $n_1 \le m_1 < 2n_1$ and $n_2 \le m_2 < 2n_2$, we have

$$\begin{aligned}(3.14)\quad &\frac{1}{|R|^2}\int_{I\times J}\int_{\substack{\{y_1\in\mathbb{R}:2^{a\frac{m_1}{n_1}}|I|<|x_1-y_1|\le 2^{a\frac{m_1+1}{n_1}}|I|\}\\ \{y_2\in\mathbb{R}:2^{a\frac{m_2}{n_2}}|J|<|x_2-y_2|\le 2^{a\frac{m_2+1}{n_2}}|J|\}}} \big|\triangle^{(1)}_{y_1}\triangle^{(2)}_{y_2} b(x_1,x_2)\big|^p dy_1dy_2dx_1dx_2\\
&\lesssim \frac{1}{|R|^2}\bigg\{\int_{R'}\int_{R'}\big|B_1(x_1,x_2)\big|^p dy_1dy_2dx_1dx_2 + \int_{R'}\int_{R'}\big|B_2(y_1,x_2)\big|^p dy_1dy_2dx_1dx_2\\
&\qquad + \int_{R'}\int_{R'}\big|B_3(x_1,y_2)\big|^p dy_1dy_2dx_1dx_2 + \int_{R'}\int_{R'}\big|B_4(y_1,y_2)\big|^p dy_1dy_2dx_1dx_2\bigg\}.\end{aligned}$$

It follows that the norm $\|b\|^p_{B^p(\mathbb{R}\times\mathbb{R})}$ is dominated by several terms, one of which is

$$(3.15)\qquad \sum_{R\in\mathcal{D}^0\times\mathcal{D}^0}|R|^{-2}\int_R\int_R\big|b(x_1,x_2)-E^{(1,0)}b(x_2)-E^{(0,1)}b(x_1)+E^{(1,1)}b\big|^p dy_1dy_2dx_1dx_2.$$

The other terms are obtained by varying the role of m_1 in (3.13) and the similar index m_2, considering the negative of the intervals in (3.13), exchanging the role of the dyadic grid, and the roles of x_i and y_i, $i = 1, 2$. All cases are similar, so we continue with the one above. In (3.15), the point is that

$$(3.16)\qquad \mathbf{1}_R\big(b(x_1,x_2)-E^{(1,0)}b(x_2)-E^{(0,1)}b(x_1)+E^{(1,1)}b\big) = \sum_{\substack{\tilde R\in\mathcal{D}^0\times\mathcal{D}^0\\ \tilde R\subset R}}\langle b, h_{\tilde R}\rangle h_{\tilde R}.$$

That is, only the smaller scales contribute. But then, it is straight forward to see that we can make a pure sum on scales.

$$(3.17)\qquad (3.15) \lesssim \sum_{R\in\mathcal{D}^0\times\mathcal{D}^0}|R|^{-2}\int_R\int_R\big|\langle b,h_R\rangle h_R\big|^p dydx \lesssim \|b\|^p_{\mathbb{B}^{p,(0,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})}.$$

This completes the proof. $\square$

Thus, based on Propositions 3.8 and 3.11, we obtain that

3.18. **Proposition.** *Suppose $1 \le p < \infty$. There are two choices of grids $\mathcal{D}^0$ and $\mathcal{D}^1$ for which we have*

$$\mathbb{B}^p(\mathbb{R}\times\mathbb{R}) = \mathbb{B}^{p,(0,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})\cap\mathbb{B}^{p,(0,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})\cap\mathbb{B}^{p,(1,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})\cap\mathbb{B}^{p,(1,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R}),$$

where $\mathbb{B}^{p,(i,j)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})$ is the dyadic Besov space associated with $\mathcal{D}^i\times\mathcal{D}^j$ for $i, j \in \{0, 1\}$. Moreover, we have

$$\|b\|_{B^p(\mathbb{R}\times\mathbb{R})} \simeq \|b\|_{\mathbb{B}^{p,(0,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})} + \|b\|_{\mathbb{B}^{p,(0,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})} + \|b\|_{\mathbb{B}^{p,(1,0)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})} + \|b\|_{\mathbb{B}^{p,(1,1)}_{\rm dyadic}(\mathbb{R}\times\mathbb{R})}.$$

4. The Proof of the One Parameter Result

The proof in fact has very little to do with the function theory of Besov spaces. The main results in Theorem 1.12 and Theorem 1.14 can be rephrased this way.

4.1. **Theorem.** *Suppose that $\{\varphi_I : I \in \mathcal{D}\}$ and $\{\phi_I : I \in \mathcal{D}\}$ are adapted to the dyadic intervals, and at least one collection of functions has cancellation. Then we have the estimate*

$$\Big\| \sum_{I\in\mathcal{D}} \alpha(I)\,\varphi_I \otimes \phi_I \Big\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p}, \qquad 0<p<\infty. \tag{4.2}$$

If both collections of functions are uniformly adapted to $\mathcal{D}$, then the reverse inequality holds.

In the setting of the Haar functions we have the estimate

$$\Big\| \sum_{I\in\mathcal{D}} \alpha(I)\,h_I^{\epsilon} \otimes h_I^{\delta} \Big\|_{\mathbb{S}^p} \simeq \|\alpha(\cdot)\|_{\ell^p}, \qquad 0<p<\infty, \tag{4.3}$$

with $\{\epsilon,\delta\} \neq \{1,1\}$.

We first point out that the conditions of φ_I and ϕ_I "to be adapted to the dyadic intervals" are close to the notion of "nearly weakly orthogonal sequences" introduced by Rochberg and Semmes, [25]. However, we do not require the compact support condition, but require suitable decay and regularity instead. Thus, it has the advantage in dealing with the continuous setting via wavelet functions or more general functions, which was missing before.

4.1. **The Proof of Theorem 4.1: The Haar Case.** We begin with the obvious estimate when $\epsilon=0$ and $\delta=0$; that is,

$$\Big\| \sum_{I\in\mathcal{D}} \alpha(I)\,h_I^{0} \otimes h_I^{0} \Big\|_{\mathbb{S}^p} \simeq \|\alpha(\cdot)\|_{\ell^p}.$$

To prove the full argument in (4.3), a certain extension of the above fact is needed. We will prove (4.3) for $\epsilon=1$ and $\delta=0$; that is, it suffices to prove

$$\Big\| \sum_{I\in\mathcal{D}} \alpha(I)\,h_I^{1} \otimes h_I^{0} \Big\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p} \tag{4.4}$$

and

$$\Big\| \sum_{I\in\mathcal{D}} \alpha(I)\,h_I^{1} \otimes h_I^{0} \Big\|_{\mathbb{S}^p} \gtrsim \|\alpha(\cdot)\|_{\ell^p}. \tag{4.5}$$

The case when $\epsilon=0$ and $\delta=1$ can be done symmetrically.

We now first prove inequality (4.4). The main point is the explicit representation

$$h_I^1 = \sum_{J:\ I\subsetneq J} \sqrt{|I|}\,h_J(c(I))\,h_J^0 = \sum_{J:\ I\subsetneq J} \nu_{I,J}\sqrt{\frac{|I|}{|J|}}\,h_J^0, \tag{4.6}$$

where the $\nu_{I,J} \in \{\pm 1\}$ are determined by I being in the left or the right half of J.

For an integer $m>0$ and for $J\in\mathcal{D}$, we define the m-fold children of J to be

$$\mathcal{D}(m,J) \stackrel{\text{def}}{=} \{I\in\mathcal{D} : I\subset J, 2^m|I| = |J|\}. \tag{4.7}$$

Now, consider the operators

$$H_{m,J} \stackrel{\mathrm{def}}{=} \sum_{I\in\mathcal{D}(m,J)} \nu_{I,J}\,\alpha(I)\,h_I^0 \tag{4.8}$$

and

$$\mathrm{S}_m \stackrel{\mathrm{def}}{=} \sum_{J\in\mathcal{D}} h_J^0 \otimes H_{m,J}. \tag{4.9}$$

The choices of signs $\nu_{I,J}$ are determined as in (4.6). Our observation is that this operator S_m also has an effective estimate of its norm as follows.

$$\|\mathrm{S}_m\|_{\mathbb{S}^p} \lesssim 2^{\delta(p)m}\,\|\alpha(\cdot)\|_{\ell^p}\,, \qquad 0<p<\infty, \tag{4.10}$$

where $\delta(p)=\max\left(0,\frac{1}{2}-\frac{1}{p}\right)$. Indeed, the functions $H_{m,J}$ are orthogonal in J, and we have the estimate

$$\|H_{m,J}\|_{L^2(\mathbb{R})} = \left[\sum_{I\in\mathcal{D}(m,J)} |\alpha(I)|^2\right]^{1/2} \le 2^{\delta(p)m}\left[\sum_{I\in\mathcal{D}(m,J)} |\alpha(I)|^p\right]^{1/p}. \tag{4.11}$$

And this clearly implies our observation in (4.10).

Taking (4.6) and (4.10) into account, it is clear that we can write

$$\sum_{I\in\mathcal{D}} \alpha(I)\,h_I^1\otimes h_I^0 = \sum_{m=1}^{\infty} 2^{-m/2}\,\mathrm{S}_m\,.$$

For $1\le p<\infty$, the Schatten norm obeys the triangle inequality, hence, together with (4.10), we can estimate

$$\begin{aligned}\left\|\sum_{I\in\mathcal{D}} \alpha(I)h_I^1\otimes h_I^0\right\|_{\mathbb{S}^p} &\le \sum_{m=1}^{\infty} 2^{-m/2}\,\|\mathrm{S}_m\|_{\mathbb{S}^p}\\ &\le \|\alpha(\cdot)\|_{\ell^p}\sum_{m=1}^{\infty} 2^{-m\min\left(\frac{1}{2},\frac{1}{p}\right)}\\ &\lesssim \|\alpha(\cdot)\|_{\ell^p}\,.\end{aligned} \tag{4.12}$$

In the case that $0<p<1$, we can rely upon the subadditivity as in (2.6), and a very similar argument finishes the proof of the upper bound in (4.4).

Our inequality (4.4) then follows.

We prove the lower bound in (4.5). Fix $\alpha(\cdot)$ so that $\|\alpha(\cdot)\|_{\ell^p}=1$, We want to show that for

$$\mathrm{T}_\alpha \stackrel{\mathrm{def}}{=} \sum_{I\in\mathcal{D}} \alpha(I)\,h_I^1\otimes h_I^0,$$

we have $\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \gtrsim 1$.

A collection of dyadic intervals $\mathcal{D}_\ell$ will *have scales separated by* ℓ if it satisfies the conditions $I\in\mathcal{D}_\ell$ implies $I\pm 2|I|\in\mathcal{D}_\ell$, but $I\pm|I|\notin\mathcal{D}_\ell$ and $\{\log_2|I| \,:\, I\in\mathcal{D}_\ell\}=a+\ell\mathbb{Z}$ for some choice of integer a. All dyadic intervals $\mathcal{D}$ are a union of 2ℓ subcollections with scales separated by ℓ. Of course passing to such a subcollection will suggest a loss of order ℓ^{-1}, but from other aspects of the argument below, we will be able to pick up an exponential decay in ℓ.

For an integer ℓ to be chosen, we can separate the scales in the dyadic intervals, choosing one particular $\mathcal{D}_\ell$ so that

$$\sum_{I\in\mathcal{D}_\ell} |\alpha(I)|^p \geq (2\ell)^{-p}. \tag{4.13}$$

Let P_ℓ be the projection onto the span of the functions $\{h_J : J \in \mathcal{D}_\ell\}$, i.e.,

$$\mathrm{P}_\ell(f) = \sum_{J\in D_\ell} \langle f, h_J^0\rangle_{L^2(\mathbb{R})} h_J^0.$$

Let $\mathcal{D}_\ell'$ be the collection of *parents* of those dyadic intervals in $\mathcal{D}_\ell$ (the parent of I is the next dyadic larger interval) and let P_ℓ' be the corresponding projection. Appealing to Proposition 2.7 we have the estimate

$$\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \geq \|\mathrm{P}_\ell' \, \mathrm{T}_\alpha \, \mathrm{P}_\ell\|_{\mathbb{S}^p} .$$

We shall show that $\|\mathrm{P}_\ell' \, \mathrm{T}_\alpha \, \mathrm{P}_\ell\|_{\mathbb{S}^p}$ is at least $(8\ell)^{-1}$ for ℓ sufficiently large depending only on p.

Define

$$\mathrm{T}_\alpha^1 \stackrel{\mathrm{def}}{=} \sum_{I\in\mathcal{D}_\ell} v_{I,I'} \; \alpha(I) \, h_{I'}^0 \otimes h_I^0,$$

where I' is the parent of I and again the $\nu_{I,I'} \in \{\pm 1\}$ are determined by I being in the left or the right half of I'. Then the $\mathbb{S}^p$ norm can be calculated exactly from (2.2) as follows

$$\left\|\mathrm{T}_\alpha^1\right\|_{\mathbb{S}^p}^p = \sum_{I\in\mathcal{D}_\ell} \alpha(I)^p \geq (2\ell)^{-p},$$

where the above inequality follows from (4.13). This is the main term in providing a lower bound on the Schatten norm for T_α.

We appeal to some of the estimates used in the proof of the upper bound. Define

$$\mathrm{T}_\alpha^m \stackrel{\mathrm{def}}{=} \sum_{I\in\mathcal{D}_\ell'} h_I \otimes \widetilde{H}_{m\ell+1,I},$$

where $\widetilde{H}_{m\ell+1,I}$ is similar as in the definition from (4.9), given by

$$\widetilde{H}_{m\ell+1,I} \stackrel{\mathrm{def}}{=} \sum_{\substack{K\in\mathcal{D}_\ell \\ K\in\mathcal{D}(m\ell+1,I)}} \nu_{K,I} \, \alpha(K) \, h_K^0. \tag{4.14}$$

Note that

$$\mathrm{P}_\ell' \, \mathrm{T}_\alpha \, \mathrm{P}_\ell = 2^{-\frac{1}{2}} \, \mathrm{T}_\alpha^1 + \sum_{m=2}^{\infty} 2^{-\frac{1}{2}(m\ell+1)} \, \mathrm{T}_\alpha^m \, .$$

Repeating the estimates as in (4.10) and (4.12), we have the estimate

$$\left\| \sum_{m=2}^{\infty} 2^{-\frac{1}{2}(m\ell+1)} \, \mathrm{T}_\alpha^m \right\|_{\mathbb{S}^p} \lesssim 2^{-\min\left(\frac{1}{2},\frac{1}{p}\right)\ell}.$$

The implied constant depends upon $0 < p < \infty$. Therefore, for an absolute choice of ℓ, we will have the estimate $\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \geq (8\ell)^{-1}$. This completes the proof of (4.5).

Thus, combining (4.4) and (4.5), we obtain that (4.3) holds. The proof of the Haar Case in Theorem 4.1 is complete. □

4.2. **The Proof of Theorem 4.1: The Wavelet Case.** Assume that both collections $\{\varphi_I\}$ and $\{\phi_I\}$ are wavelet bases, then the operator in (4.2) is already given in singular value form, and the theorem is trivial. The Theorem follows from the lemma below:

4.15. **Lemma.** *We have the inequality*

$$\left\| \sum_{I\in\mathcal{D}} \alpha(I)\varphi_I \otimes \phi_I \right\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p}\,, \qquad 0<p<\infty, \tag{4.16}$$

assuming only that $\{\phi_I\}$ are adapted to $\mathcal{D}$. If this collection of functions is uniformly adapted to $\mathcal{D}$, then the reverse inequality holds.

In the case of Haar functions, the main point is the explicit expansion of (4.6). In the current setting, of course we have the expansion

$$\varphi_I = \sum_{J\in\mathcal{D}} \langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})}\, w_J,$$

as $\{w_J\}$ is an orthogonal basis as defined in Page 2 in the introduction. But the expansion is not quite so clean. Nevertheless, we have the following general almost orthogonality estimate.

4.17. **Lemma.** *Denoting $2^m|I| = |J|$, we have the inequality*

$$|\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})}| \lesssim 2^{-\Delta(m)} \left(1+\frac{\operatorname{dist}(I,J)}{|I|+|J|}\right)^{-\eta}, \qquad m\in\mathbb{Z}. \tag{4.18}$$

Here, $\Delta(m)=|m|$ if $m\le 0$, and $\Delta(m)=\frac{1}{2}m$ if $m>0$, $\eta>0$ is a large positive constant, namely $N-1$, where N appears in (1.5).

Proof. This is elementary. On the one hand, by using (1.5), we have

$$|\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})}| \lesssim 2^{-\frac{1}{2}|m|} \left(1+\frac{\operatorname{dist}(I,J)}{|I|+|J|}\right)^{-N+1},$$

where N is as in (1.5). This treats the case $m>0$.

The case $m<0$ does not occur in the Haar setting. While it does occur here, there is an extra decay coming from the fact that the wavelet has mean zero, and are adapted to an interval of smaller length than J. Thus, we essentially gain a derivative in this case

$$|\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})}| \lesssim 2^{-\frac{3}{2}|m|}.$$

Taking the geometric mean of these two estimates proves the estimate in this last case. □

We use this lemma to prove another technical lemma, more specifically adapted to our purposes. Let P_{2^m} be the wavelet projection of functions onto the span of $\{w_J : |J| = 2^m\}$; that is

$$\mathrm{P}_{2^m} \overset{\text{def}}{=} \sum_{J\in\mathcal{D}:\ |J|=2^m} w_J\otimes w_J,$$

so that

$$\mathrm{P}_{2^m}(f) = \sum_{J\in\mathcal{D}:\ |J|=2^m} \langle f, w_J\rangle w_J.$$

4.19. **Lemma.** *We have the estimate*

$$\|\mathrm{B}_m\|_{\mathbb{S}^p} \lesssim 2^{-\min(\frac{1}{2},\frac{1}{p})|m|} \|\alpha(\cdot)\|_{\ell^p}, \qquad m \in \mathbb{Z},$$

$$\textit{where} \quad \mathrm{B}_m \stackrel{\text{def}}{=} \sum_{I \in \mathcal{D}} \alpha(I)[\mathrm{P}_{2^m|I|}\,\varphi_I] \otimes \phi_I.$$

In particular note that the scale of the wavelet projection being used depends upon the scale of I.

Proof. Consider first the case of $m < 0$. Now, for an integer $\ell \geq 0$, set

$$\mathcal{D}_0(I) \stackrel{\text{def}}{=} \{J \in \mathcal{D} \,:\, |J| = 2^m|I|,\ J \subset I\},$$
$$\mathcal{D}_\ell(I) \stackrel{\text{def}}{=} \{J \in \mathcal{D} \,:\, |J| = 2^m|I|,\ J \subset 2^\ell I,\ J \not\subset 2^{\ell-1}I\}.$$

Also consider the functions

$$W_{I,\ell} \stackrel{\text{def}}{=} \sum_{J \in \mathcal{D}_\ell(I)} \alpha(I)\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})} w_J.$$

For fixed ℓ, the functions $\{W_{I,\ell} \,:\, I \in \mathcal{D}\}$ are orthogonal. Hence the operator

$$\mathrm{T}_\ell \stackrel{\text{def}}{=} \sum_{I \in \mathcal{D}} W_{I,\ell} \otimes \phi_I$$

is in singular value form. Moreover, observe that from Lemma 4.17,

$$\|W_{I,\ell}\|_{L^2(\mathbb{R})} \leq \Bigg[\sum_{J \in \mathcal{D}_\ell(I)} |\alpha(I)\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})}|^2\Bigg]^{1/2} \lesssim 2^{-|m|-\eta'\ell} \Bigg[\sum_{J \in \mathcal{D}_\ell(I)} |\alpha(I)|^2\Bigg]^{1/2}.$$

An elementary estimate gives

$$\Bigg[\sum_{J \in \mathcal{D}_\ell(I)} |\alpha(I)|^2\Bigg]^{1/2} \lesssim \operatorname{card}(\mathcal{D}_\ell(I))^{\delta(p)} \Bigg[\sum_{J \in \mathcal{D}_\ell(I)} |\alpha(I)|^p\Bigg]^{1/p}. \tag{4.20}$$

The term in the exponent is $\delta(p) = \max\left(0, \frac{1}{2} - \frac{1}{p}\right)$ as before.

Hence, we can easily estimate the Schatten norm of T_ℓ.

$$\|\mathrm{T}_\ell\|_{\mathbb{S}^p}^p \leq \sum_{I \in \mathcal{D}} \|W_{I,\ell}\|_{L^2(\mathbb{R})}^p \lesssim 2^{-p(1-\delta(p))|m|-\eta''\ell} \|\alpha(\cdot)\|_{\ell^p}^p.$$

Because we are free to take η'' as large as needed, this completes the proof in this case.

We now consider the case of $m > 0$. Keeping the same notation $\mathcal{D}_\ell(J)$, we redefine

$$W_{J,\ell} \stackrel{\text{def}}{=} \sum_{I \in \mathcal{D}_\ell(J)} \alpha(I)\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})} w_J,$$
$$\mathrm{T}_\ell \stackrel{\text{def}}{=} \sum_{J \in \mathcal{D}} w_J \otimes W_{J,\ell}.$$

Again, this is an operator in singular value form. In particular

$$\|W_{J,\ell}\|_{L^2(\mathbb{R})}^2 = \sum_{I \in \mathcal{D}_\ell(J)} |\alpha(I)\langle \varphi_I, w_J\rangle_{L^2(\mathbb{R})}|^2$$

$$\lesssim 2^{-m-\eta\ell} \sum_{I \in \mathcal{D}(J)} |\alpha(I)|^2$$

$$\lesssim 2^{-m(1-2\delta(p))-\eta\ell} \Bigg[\sum_{I \in \mathcal{D}(J)} |\alpha(I)|^p \Bigg]^{2/p} .$$

Therefore, it is the case

$$\|\mathrm{T}_\ell\|_{\mathbb{S}^p}^p \lesssim 2^{-mp \min\left(\frac{1}{2},\frac{1}{p}\right)-\eta\ell} \, \|\alpha(\cdot)\|_{\ell^p}^p .$$

Due to the fact that η can be taken very large, for all $0 < p < \infty$, this estimate can be summed over ℓ, to prove the lemma in this case. The proof of Lemma 4.19 is complete. ☐

Proof of Lemma 4.15. We assume that $\{\phi_I\}$ are adapted to $\mathcal{D}$. Then, we have

$$\begin{aligned}
\sum_{I \in \mathcal{D}} \alpha(I)\, \varphi_I \otimes \phi_I &= \sum_{I \in \mathcal{D}} \alpha(I) \sum_{J \in \mathcal{D}} \langle \varphi_I, w_J \rangle_{L^2(\mathbb{R})}\, w_J \otimes \phi_I \\
&= \sum_{m \in \mathbb{Z}} \sum_{I \in \mathcal{D}} \alpha(I) \sum_{J \in \mathcal{D}\colon\ |J| = 2^m} \langle \varphi_I, w_J \rangle_{L^2(\mathbb{R})}\, w_J \otimes \phi_I \\
&= \sum_{m \in \mathbb{Z}} \sum_{I \in \mathcal{D}} \alpha(I) \big[\mathrm{P}_{2^m|I|}\, \varphi_I \big] \otimes \phi_I \\
&= \sum_{m \in \mathbb{Z}} \mathrm{B}_m .
\end{aligned}$$

With the estimates on the operators B_m provided by Lemma 4.19, we obtain that

$$\Bigg\| \sum_{I \in \mathcal{D}} \alpha(I) \varphi_I \otimes \phi_I \Bigg\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p}$$

for $0 < p < \infty$, which shows that (4.16) holds.

We turn to the proof of the reverse inequality, assuming that $\{\phi_I\}$ are uniformly adapted to $\mathcal{D}$. We aim to prove that

$$\Bigg\| \sum_{I \in \mathcal{D}} \alpha(I) \varphi_I \otimes \phi_I \Bigg\|_{\mathbb{S}^p} \gtrsim \|\alpha(\cdot)\|_{\ell^p} , \qquad 0 < p < \infty. \tag{4.21}$$

This argument is modeled on the proof of the lower bound in the Haar setting. Recall that this means that (1.6) is in force. Fix a dyadic interval I_0 with length 1, so that

$$|\langle \phi_{[0,1]}, w_{I_0} \rangle_{L^2(\mathbb{R})}|$$

is maximal.

To continue, we define

$$I < J \tag{4.22}$$

if the (orientation preserving) linear transformation that carries J to $[0,1]$ also carries I to I_0. (In the Haar case, we take I_0 to be the parent of $[0,1]$).

In Figure 1 below, we illustrate two examples of orientation-preserving linear transformations applied to dyadic intervals. Let I_0 be a dyadic interval of length 1. The interval I_0 may lie to the right of $[0,1]$, or conversely, $[0,1]$ may lie to the right of I_0. These examples demonstrate how the transformation maps intervals while preserving orientation and scale (up to dyadic structure), possibly shifting their position along the real line.

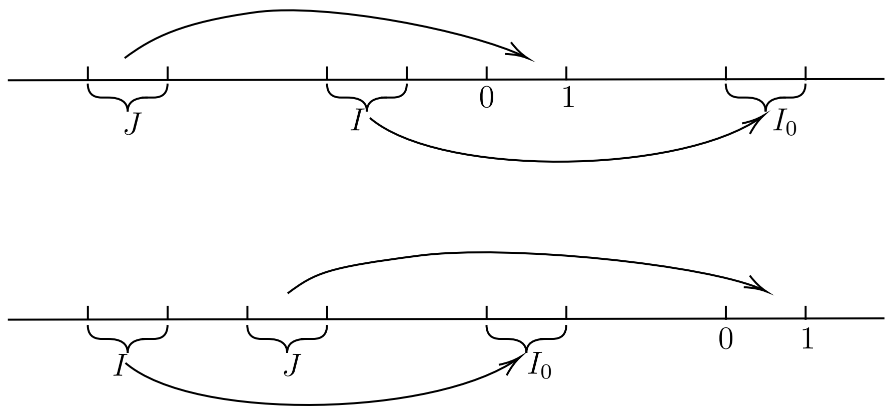

Figure 1. (orientation preserving) linear transform

The notion of *scales separated by* ℓ is modified slightly. A collection of dyadic intervals $\mathcal{D}_\ell$ will *have scales separated by* ℓ if it satisfies the conditions: "$I \in \mathcal{D}_\ell$ implies $I \pm \ell|I| \in \mathcal{D}_\ell$, but $I \pm j|I| \notin \mathcal{D}_\ell$ for $|j| < \ell$ and $\{\log_2|I| \,:\, I \in \mathcal{D}_\ell\} = a + \ell\mathbb{Z}$ for some choice of integer a".

All dyadic intervals $\mathcal{D}$ are a union of ℓ^2 subcollections with scales separated by ℓ. We expect a loss of ℓ^{-2} by passing to a subcollection with scales separated by ℓ. We will be able to pick up rapid (but not exponential) decay from other parts of the argument.

Fix $\alpha(\cdot)$ such that $\|\alpha(\cdot)\|_p = 1$. We want to provide a lower bound on

$$\mathrm{T}_\alpha \overset{\text{def}}{=} \sum_{I \in \mathcal{D}} \alpha(I)\, \phi_I \otimes w_I.$$

For a choice of ℓ to be specified, we can choose $\mathcal{D}_\ell$ with scales separated by ℓ such that

$$\Big(\sum_{I \in \mathcal{D}_\ell} |\alpha(I)|^p \Big)^{\frac{1}{p}} \geq \ell^{-2}.$$

Let P_ℓ be the projection onto the span of $\{w_I \,:\, I \in \mathcal{D}_\ell\}$, that is

$$\mathrm{P}_\ell \overset{\text{def}}{=} \sum_{I \in \mathcal{D}_\ell} w_I \otimes w_I.$$

Let $\mathcal{D}_{\ell,<}$ be those dyadic intervals I which satisfy $I < J$ for some $J \in \mathcal{D}_\ell$ (as defined in (4.22)). We will pick ℓ so large that there is a unique such J. Let $\mathrm{P}_{<,\ell}$ be the corresponding wavelet projection onto the span of $\{w_{I_<} \,:\, I_< \in \mathcal{D}_{\ell,<}\}$, defined similarly to P_ℓ as above.

By Proposition 2.7, we have the estimate

$$\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \geq \|\mathrm{P}_{<,\ell}\, \mathrm{T}_\alpha\, \mathrm{P}_\ell\|_{\mathbb{S}^p}\,.$$

We estimate the norm of the latter quantity. By definition, we have

$$\mathrm{P}_{<,\ell}\, \mathrm{T}_\alpha\, \mathrm{P}_\ell = \sum_{I \in \mathcal{D}_{\ell,<}} \sum_{I' \in \mathcal{D}_\ell} \alpha(I') \langle \phi_{I'}, w_{I_<} \rangle_{L^2(\mathbb{R})}\, w_{I_<} \otimes w_{I'}.$$

Set

$$\mathrm{T}^1_\alpha \overset{\text{def}}{=} \sum_{I \in \mathcal{D}_\ell} \alpha(I) \langle \phi_I, w_{I_<} \rangle_{L^2(\mathbb{R})}\, w_{I_<} \otimes w_I.$$

Here, by $I_<$ we mean that unique element of $\mathcal{D}_{\ell,<}$ for which $I_< < I$. Observe that $\langle \phi_I, w_{I_<}\rangle_{L^2(\mathbb{R})}$ is in fact independent of I, and so we have the estimate

$$\left\|\mathrm{T}_\alpha^1\right\|_{\mathbb{S}^p} \gtrsim \ell^{-2}, \tag{4.23}$$

with the implied constant depending only on the specific choice of ϕ and wavelet basis $\{w_I\}$.

The remainder of the argument consists of showing that

$$\left\|\mathrm{P}_{<,\ell}\,\mathrm{T}_\alpha\,\mathrm{P}_\ell - \mathrm{T}_\alpha^1\right\|_{\mathbb{S}^p} \lesssim \ell^{-4}, \tag{4.24}$$

where the implied constant depends on $0 < p < \infty$, and on the specific constants that enter into the inequality (1.5) (in particular, for the case of $0 < p < 1$, we will need to require that $Np > 5$).

To see (4.24), note that the remainder $\mathrm{P}_{<,\ell}\,\mathrm{T}_\alpha\,\mathrm{P}_\ell - \mathrm{T}_\alpha^1$ can be represented and split as follows. Set

$$\mathcal{D}_{\ell,j}(I') \stackrel{\text{def}}{=} \{I \in \mathcal{D}_{\ell,<} \,:\, |I| \subset 2^j I',\ I \not\subset 2^{j-1}I'\}, \quad j > 1.$$

And we consider the functions

$$W_{I',j} \stackrel{\text{def}}{=} \sum_{I\in\mathcal{D}_{\ell,j}(I')} \alpha(I')\langle \phi_{I'}, w_{I_<}\rangle_{L^2(\mathbb{R})}\, w_{I_<}.$$

Then we have

$$\begin{aligned}
\mathrm{P}_{<,\ell}\,\mathrm{T}_\alpha\,\mathrm{P}_\ell - \mathrm{T}_\alpha^1 &= \sum_{I'\in\mathcal{D}_\ell}\ \sum_{I\in\mathcal{D}_{\ell,<}, I\neq I'} \alpha(I')\langle \phi_{I'}, w_{I_<}\rangle_{L^2(\mathbb{R})}\, w_{I_<} \otimes w_{I'} \\
&= \sum_{j=2}^{\infty}\sum_{I'\in\mathcal{D}_\ell} W_{I',j} \otimes w_{I'} \\
&\stackrel{\text{def}}{=} \sum_{j=2}^{\infty} T_{\ell,j}.
\end{aligned}$$

Note that for each j, there are only $2j$ cases of $I \in \mathcal{D}_{\ell,j}(I')$. Hence, by using the almost orthogonality estimate in Lemma 4.17 for $\langle \phi_{I'}, w_{I_<}\rangle_{L^2(\mathbb{R})}$, and following a similar step in the proof of Lemma 4.19 for $\|W_{I',j}\|_2$, we obtain that $\|T_{\ell,j}\|_{\mathbb{S}^p} \lesssim 2^{-\eta j}2^{-\eta\ell}$, where η is a large positive number as in Lemma 4.17. Thus, summing over all $j \geq 2$, we obtain that $\left\|\mathrm{P}_{<,\ell}\,\mathrm{T}_\alpha\,\mathrm{P}_\ell - \mathrm{T}_\alpha^1\right\|_{\mathbb{S}^p} \lesssim 2^{-\eta\ell}$, that is, (4.24) holds.

Now, it is clear that we can choose ℓ sufficiently large, and then combine (4.23) with (4.24) to obtain that (4.21) holds. The proof of the Wavelet Case in Theorem 4.1 is complete. □

5. The Proof of the Multi-Parameter Result

5.1. **Two-Parameter Paraproducts.** We now consider paraproducts formed over sums of dyadic rectangles in the plane. The class of paraproducts is then invariant under a two parameter family of dilations, a situation that we refer to as one of "two parameters". This case contains all the essential difficulties for the higher parameter setting and is good for focusing ideas. Again, there is very little function theory in the argument and Theorems 1.25 and 1.27 in the two parameter setting can be rephrased as

5.1. **Theorem.** *Let $\{\phi_R : R \in \mathcal{R}\}$ and $\{\varphi_R : R \in \mathcal{R}\}$ be collections of functions adapted to $\mathcal{R}$, with at least one collection having cancellation in the jth coordinate for $j = 1, 2$. Then we have the inequality*

$$\Bigg\| \sum_{R \in \mathcal{R}} \alpha(R)\, \varphi_R \otimes \phi_R \Bigg\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p}. \tag{5.2}$$

In the Haar case, we have for $\epsilon, \delta \in \{0,1\}^2$, with the same assumption on cancellation, that the equivalence holds

$$\Bigg\| \sum_{R \in \mathcal{R}} \alpha(R)\, h_R^{\epsilon} \otimes h_R^{\delta} \Bigg\|_{\mathbb{S}^p} \simeq \|\alpha(\cdot)\|_{\ell^p}. \tag{5.3}$$

5.1.1. *The Proof of Theorem 5.1: The Haar Case.* Recall that as in (1.16), for a dyadic rectangle $R = R_1 \times R_2$, we set $h_R(x_1, x_2) = h^0_{R_1}(x_1) h^0_{R_2}(x_2)$.

We of course immediately have the inequality

$$\Bigg\| \sum_{R \in \mathcal{R}} \alpha(R)\, h_R \otimes h_R \Bigg\|_{\mathbb{S}^p} \simeq \|\alpha(\cdot)\|_{\ell^p}.$$

And, keeping in mind the proof in one parameter, we need a certain extension of this inequality.

To set some notation to capture the role of zeros, or their absence, we set

$$h_R^{(\epsilon_1,\epsilon_2)}(x_1, x_2) = h^{\epsilon_1}_{R_1}(x_1) h^{\epsilon_2}_{R_2}(x_2).$$

We discuss the equivalence (5.3). There are three possible forms of the paraproduct, after taking duality and permutation of coordinates into account. Of these, the first and simplest case is

$$\sum_{R \in \mathcal{R}} \alpha(R)\, h_R^{(1,0)} \otimes h_R.$$

This is very clearly a simple variant of the one parameter version, and we do not discuss it.

The second case is that $\epsilon = (1,1)$ and $\delta = (0,0)$, i.e.,

$$\sum_{R \in \mathcal{R}} \alpha(R)\, h_R^{(1,1)} \otimes h_R.$$

It is clear that one function is a Haar function, and the other is a normalized indicator function. This case is the most natural analog of the one parameter case.

From (4.6) we see that for $j = 1, 2$,

$$h^1_{R_j} = \sum_{S_j :\ R_j \subsetneqq S_j} \nu_{R_j, S_j} \sqrt{\frac{|R_j|}{|S_j|}}\, h^0_{S_j}. \tag{5.4}$$

Define, for a multi-integer $m = (m_1, m_2) \in \mathbb{N}^2$, the collections

$$\mathcal{R}(m, S) = \{R \in \mathcal{R} : R \subset S,\ 2^{m_j}|R_j| = |S_j|,\ j = 1, 2\}.$$

There is a corresponding operator A_m defined as follows.

$$\mathrm{A}_m \stackrel{\text{def}}{=} \sum_{S \in \mathcal{R}} h_S \otimes H_{m,S}, \tag{5.5}$$

where

$$H_{m,S} \stackrel{\text{def}}{=} \sum_{R \in \mathcal{R}(m,S)} \nu_{R,S}\, \alpha(R) h_R. \tag{5.6}$$

Here, $\nu_{R,S} = \nu_{R_1,S_1} \cdot \nu_{R_2,S_2}$. Observe that, just as in (4.10), we have the estimate

$$\|\mathrm{A}_m\|_{\mathbb{S}^p} \lesssim 2^{\delta(p)(m_1+m_2)} \|\alpha(\cdot)\|_{\ell^p} . \tag{5.7}$$

Here, $\delta(p) = \max\left(0, \frac{1}{2} - \frac{1}{p}\right)$ as before. This follows just as before, namely observe that the functions $H_{m,S}$ are orthogonal in S, and that

$$\|H_{m,S}\|_2 = \left[\sum_{R \in \mathcal{R}(m,S)} |\alpha(R)|^2\right]^{1/2} \leq 2^{\delta(p)(m_1+m_2)} \left[\sum_{R \in \mathcal{R}(m,S)} |\alpha(R)|^p\right]^{1/p} .$$

In addition, by (5.4), we have

$$\sum_{R \in \mathcal{R}} \alpha(R)\, h_R^{(1,1)} \otimes h_R = \sum_{m \in \mathbb{N}^2} 2^{-\frac{1}{2}(m_1+m_2)}\, \mathrm{A}_m$$

for appropriate choices of signs in the definition of A_m. The remainder of the proof is just as in the one parameter case. For $1 \leq p < \infty$, the Schatten norm obeys the triangle inequality, hence, together with (5.7), we can estimate

$$\begin{aligned} \left\| \sum_{R \in \mathcal{R}} \alpha(R) h_R^{(1,1)} \otimes h_R \right\|_{\mathbb{S}^p} &\leq \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} 2^{-\frac{1}{2}(m_1+m_2)} \|\mathrm{A}_m\|_{\mathbb{S}^p} \\ &\leq \|\alpha(\cdot)\|_{\ell^p} \sum_{m_1=1}^{\infty} 2^{-m_1 \min\left(\frac{1}{2},\frac{1}{p}\right)} \sum_{m_2=1}^{\infty} 2^{-m_2 \min\left(\frac{1}{2},\frac{1}{p}\right)} \\ &\lesssim \|\alpha(\cdot)\|_{\ell^p} . \end{aligned} \tag{5.8}$$

In the case that $0 < p < 1$, we can rely upon the subadditivity as in (2.6), and a very similar argument finishes the proof of the upper bound as above.

We turn to the case with no proper analog in the one parameter setting, namely

$$\mathrm{T}_\alpha := \sum_{R \in \mathcal{R}} \alpha(R)\, h_R^{(1,0)} \otimes h_R^{(0,1)}. \tag{5.9}$$

Of course we want to apply (4.6) to the first function in the first coordinate, and the second function in the second coordinate. Doing so suggests these definitions. For $m \in \mathbb{N}^2$, and $S \in \mathcal{R}$,

$$\mathcal{R}(m,S) \stackrel{\text{def}}{=} \{R \in \mathcal{R} \,:\, R \subset S,\ 2^{m_j}|R_j| = |S_j|,\ j = 1,2\}, \tag{5.10}$$

$$\mathrm{A}_m \stackrel{\text{def}}{=} \sum_{S \in \mathcal{R}} \mathrm{A}_{m,S}, \tag{5.11}$$

where

$$\mathrm{A}_{m,S} \stackrel{\text{def}}{=} \sum_{R \in \mathcal{R}(m,S)} \nu_{R,S}\, \alpha(R)\, h^0_{S_1}(x_1) h^0_{R_2}(x_2) \cdot h^0_{R_1}(y_1) h^0_{S_2}(y_2). \tag{5.12}$$

Here, $\nu_{R,S}$ are appropriate choices of signs.

The principle observation is that Proposition 2.8 applies to the operators $\mathrm{A}_{m,S}$, giving the estimate

$$\|\mathrm{A}_{m,S}\|_2 \lesssim 2^{\delta(p)(m_1+m_2)} \left[\sum_{R\in\mathcal{R}(m,S)} |\alpha(R)|^p \right]^{1/p}.$$

It is easy to see that we then have

$$\|\mathrm{A}_m\|_{\mathbb{S}^p} \lesssim 2^{\delta(p)(m_1+m_2)} \|\alpha(\cdot)\|_p .$$

Finally, we have

$$\begin{aligned}
&\sum_{R\in\mathcal{R}} \alpha(R)\, h_R^{(1,0)}(x_1,x_2) \otimes h_R^{(0,1)}(y_1,y_2)\\
&= \sum_{\substack{R=R_1\times R_2\\ R_1,R_2\in\mathcal{D}}} \alpha(R) \Big[\sum_{S_1:\ R_1\subsetneqq S_1} \nu_{R_1,S_1} \sqrt{\frac{|R_1|}{|S_1|}}\, h^0_{S_1}(x_1)\Big] h^0_{R_2}(x_2)\, h^0_{R_1}(y_1)\\
&\qquad \times \Big[\sum_{S_2:\ R_2\subsetneqq S_2} \nu_{R_2,S_2} \sqrt{\frac{|R_2|}{|S_2|}}\, h^0_{S_2}(y_2)\Big]\\
&= \sum_{\substack{S=S_1\times S_2\\ S_1,S_2\in\mathcal{D}}} \sum_{\substack{R_1\subsetneqq S_1\\ R_2\subsetneqq S_2\\ R_1,R_2\in\mathcal{D}}} \alpha(R)\nu_{R,S}\ h^0_{R_1}(y_1)h^0_{R_2}(x_2)\ \left[\sqrt{\frac{|R_1|}{|S_1|}}\, h^0_{S_1}(x_1)\right]\left[\sqrt{\frac{|R_2|}{|S_2|}}\, h^0_{S_2}(y_2)\right]\\
&= \sum_{m\in\mathbb{N}^2} 2^{-\frac{1}{2}(m_1+m_2)}\, \mathrm{A}_m .
\end{aligned}$$

The conclusion of this case is just as that of (5.8). Thus, we obtain that

$$\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p}, \qquad 0<p<\infty,$$

where T_α is as defined in (5.9).

Combining all these cases above, we obtain that for $(\epsilon,\delta) = ((0,0),(0,0))$, $((1,1),(0,0))$, $((0,0),(1,1))$, $((1,0),(0,1))$ and $((0,1),(1,0))$

$$\left\| \sum_{R\in\mathcal{R}} \alpha(R)\, h_R^{\epsilon} \otimes h_R^{\delta} \right\|_{\mathbb{S}^p} \lesssim \|\alpha(\cdot)\|_{\ell^p}, \qquad 0<p<\infty.$$

We now prove the reverse direction. We only pick one case, $(\epsilon,\delta) = ((1,0),(0,1))$, to discuss the details, that is, we aim to prove that

$$\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \gtrsim \|\alpha(\cdot)\|_{\ell^p}, \qquad 0<p<\infty, \tag{5.13}$$

where T_α is as defined in (5.9).

To begin with, we fix $\alpha(\cdot)$ such that $\|\alpha(\cdot)\|_{\ell^p}=1$. We want to show that

$$\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \gtrsim 1.$$

All dyadic intervals $\mathcal{D}$ are a union of 2ℓ subcollections with scales separated by ℓ (for 'scales separated by ℓ', we refer to the definition provided in the proof of (4.5)). We denote by $\mathcal{D}_\ell^{(1)}$ such collection of intervals in the first variable, and by $\mathcal{D}_\ell^{(2)}$ such collection of intervals in the second variable. Let $\big(\mathcal{D}_\ell^{(1)}\big)'$ be the collection of *parents* of those dyadic intervals in $\mathcal{D}_\ell^{(1)}$ (the parent of I is the next dyadic larger interval) and let $\big(\mathcal{D}_\ell^{(2)}\big)'$ be the collection of *parents* of those dyadic intervals in $\mathcal{D}_\ell^{(2)}$.

For an integer ℓ to be chosen, we can separate the scales in the dyadic intervals in each variable, choosing $\mathcal{D}_\ell^{(1)}$ and $\mathcal{D}_\ell^{(2)}$ so that

$$\sum_{\substack{R=R_1\times R_2\in\mathcal{R},\\ R_1\in\mathcal{D}_\ell^{(1)},R_2\in\mathcal{D}_\ell^{(2)}}} |\alpha(R)|^p \geq (4\ell^2)^{-p}. \tag{5.14}$$

Let $\widetilde{\mathrm{P}}_\ell$ be the projection onto the span of the functions $\big\{h_S\,:\,S=S_1\times S_2\in\mathcal{D}_\ell^{(1)}\times\big(\mathcal{D}_\ell^{(2)}\big)'\big\}$, i.e.,

$$\widetilde{\mathrm{P}}_\ell(f)=\sum_{S\in\mathcal{D}_\ell^{(1)}\times\left(\mathcal{D}_\ell^{(2)}\right)'}\langle f,h_S^0\rangle_{L^2(\mathbb{R})}h_S^0.$$

Let $\widetilde{\mathrm{P}}'_\ell$ be the corresponding projection onto the span of the functions $\big\{h_S\,:\,S=S_1\times S_2\in\big(\mathcal{D}_\ell^{(1)}\big)'\times\mathcal{D}_\ell^{(2)}\big\}$, i.e.,

$$\widetilde{\mathrm{P}}'_\ell(f)=\sum_{S\in\left(\mathcal{D}_\ell^{(1)}\right)'\times\mathcal{D}_\ell^{(2)}}\langle f,h_S^0\rangle_{L^2(\mathbb{R})}h_S^0.$$

Appealing to Proposition 2.7 we have the estimate

$$\|\mathrm{T}_\alpha\|_{\mathbb{S}^p}\geq\left\|\widetilde{\mathrm{P}}'_\ell\,\mathrm{T}_\alpha\,\widetilde{\mathrm{P}}_\ell\right\|_{\mathbb{S}^p}.$$

We shall show that $\left\|\widetilde{\mathrm{P}}'_\ell\,\mathrm{T}_\alpha\,\widetilde{\mathrm{P}}_\ell\right\|_{\mathbb{S}^p}$ is at least $(16\ell^2)^{-1}$ for ℓ sufficiently large depending only on p.

Define

$$\mathrm{T}_\alpha^{(1,1)}\stackrel{\text{def}}{=}\sum_{\substack{R=R_1\times R_2\in\mathcal{R},\\ R_1\in\mathcal{D}_\ell^{(1)},R_2\in\mathcal{D}_\ell^{(2)}}}\nu_{R,R'}\;\alpha(R)\,h_{R'}(x_1,y_2)\,h_R(y_1,x_2),$$

where $R'=R_1'\times R_2'$ with R_i' the parent of R_i for $i=1,2$. Then the $\mathbb{S}^p$ norm can be calculated exactly from (2.2) as follows

$$\left\|\mathrm{T}_\alpha^{(1,1)}\right\|_{\mathbb{S}^p}^p=\sum_{\substack{R=R_1\times R_2\in\mathcal{R},\\ R_1\in\mathcal{D}_\ell^{(1)},R_2\in\mathcal{D}_\ell^{(2)}}}|\alpha(R)|^p\geq(4\ell^2)^{-p},$$

where the above inequality follows from (5.14). This is the main term in providing a lower bound on the Schatten norm for T_α.

We appeal to some of the estimates used in the proof of the upper bound above. Define

$$\mathrm{T}_\alpha^m\stackrel{\text{def}}{=}\sum_{S=S_1\times S_2,\;S_1\in(\mathcal{D}_\ell^{(1)})',\;S_2\in(\mathcal{D}_\ell^{(2)})'}\widetilde{\mathrm{A}}_{m\ell+1,S},\qquad m=(m_1,m_2)\in\mathbb{N}^2,$$

where

$$\widetilde{\mathrm{A}}_{m,S}\stackrel{\text{def}}{=}\sum_{R\in\widetilde{\mathcal{R}}(m,S)}\nu_{R,S}\,\alpha(R)\,h_{S_1}^0(x_1)h_{R_2}^0(x_2)\cdot h_{R_1}^0(y_1)h_{S_2}^0(y_2) \tag{5.15}$$

and

$$\widetilde{\mathcal{R}}(m,S)\stackrel{\text{def}}{=}\{R=R_1\times R_2\in\mathcal{R}\,:\,R_1\in\mathcal{D}_\ell^{(1)},R_2\in\mathcal{D}_\ell^{(2)},R\subset S,\;2^{m_j}|R_j|=|S_j|,\;j=1,2\}.$$

Note that

$$
\begin{aligned}
\widetilde{\mathrm{P}}'_\ell \, \mathrm{T}_\alpha \, \widetilde{\mathrm{P}}_\ell &= \sum_{\substack{S=S_1\times S_2 \\ S_1\in(\mathcal{D}_\ell^{(1)})' \\ S_2\in(\mathcal{D}_\ell^{(2)})'}} \sum_{\substack{R=R_1\times R_2 \\ R_1\in\mathcal{D}_\ell^{(1)},\ R_1\subsetneq S_1 \\ R_2\in\mathcal{D}_\ell^{(2)},\ R_2\subsetneq S_2}} \alpha(R)\, \nu_{R,S} \sqrt{\frac{|R|}{|S|}}\, h^0_{S_1}(x_1) h^0_{R_2}(x_2) h^0_{R_1}(y_1) h^0_{S_2}(y_2) \\
&= \frac{1}{2}\, \mathrm{T}_\alpha^{(1,1)} + \frac{1}{\sqrt{2}} \sum_{m_2\geq 2} 2^{-\frac{1}{2}(m_2\ell+1)}\, \mathrm{T}_\alpha^{(1,m_2)} + \frac{1}{\sqrt{2}} \sum_{m_1\geq 2} 2^{-\frac{1}{2}(m_1\ell+1)}\, \mathrm{T}_\alpha^{(m_1,1)} \\
&\quad + \sum_{\substack{m\in\mathbb{N}^2 \\ m_1,m_2\geq 2}} 2^{-\frac{1}{2}(m_1\ell+1+m_2\ell+1)}\, \mathrm{T}_\alpha^{m} \\
&=: Term_1 + Term_2 + Term_3 + Term_4.
\end{aligned}
$$

Then similar to the estimates in (5.8), we have the estimate

$$
\|Term_i\|_{\mathbb{S}^p} \lesssim 2^{-\min\left(\frac{1}{2},\frac{1}{p}\right)\ell}, \qquad i = 2,3,4.
$$

The implied constant depends upon $0 < p < \infty$. Therefore, for an absolute choice of ℓ, we will have the estimate $\|\mathrm{T}_\alpha\|_{\mathbb{S}^p} \geq (16\ell^2)^{-1}$. This completes the proof of (5.14).

5.1.2. *The Proof of Theorem 5.1: The Wavelet Case.* We can derive the same argument using similar steps as in the Haar setting in Section 5.1.1, combining with the wavelet basis and the almost orthogonality estimate as used in the one-parameter setting in Section 4.2.

5.2. **The Case of Multi-parameters.** In the general n parameter setting, the number of paraproducts increases dramatically. The only restriction in forming a paraproduct is that in each of the n coordinates, there must be zeros in one of the three classes of functions corresponding to the paraproduct.

Let us pass immediately to the Haar case, without mention of the Besov norms. Given $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \{0,1\}^n$, set

$$
h_R^\epsilon(x_1,\ldots,x_n) = \prod_{j=1}^n h_{R_j}^{\epsilon_j}(x_j). \tag{5.16}
$$

In this case, the proper statement of Theorem 5.1 is

5.17. **Theorem.** *Given $\epsilon, \delta \in \{0,1\}^n$, we assume that in each coordinate j, $\epsilon_j \cdot \delta_j = 0$. Then, we have the equivalence*

$$
\Big\| \sum_{R\in\mathcal{R}} \alpha(R)\, h_R^\epsilon \otimes h_R^\delta \Big\|_{\mathbb{S}^p} \simeq \|\alpha(\cdot)\|_{\ell^p}. \tag{5.18}
$$

The proof of Theorem 5.17 is again similar to what appeared above. Let $\mathbb{O}_j$ denote the coordinates in which ϵ equals j. These sets of coordinates are disjoint, and not necessarily all of $\{1,\ldots,n\}$.

The case that both $\mathbb{O}_j$, for $j = 0,1$, are empty, is trivial. The case that one of these two is the empty set is (essentially) the one parameter case, but this will fit into the discussion below. Assume that $\mathbb{O}_1$ is not empty. Let $\mathbb{O} \stackrel{\mathrm{def}}{=} \mathbb{O}_0 \cup \mathbb{O}_1$.

We define, for choices of $S \in \mathcal{R}$, and $m \in \mathbb{N}^{\mathbb{O}}$,

$$\mathcal{R}(m,S) \stackrel{\text{def}}{=} \{R \in \mathcal{R} \,:\, R_j \subset S_j,\ 2^{m_j}|R_j| = |S_j|,\ j \in \mathbb{O}\},$$

$$\mathrm{A}_m \stackrel{\text{def}}{=} \sum_{S\in\mathcal{R}} \mathrm{A}_{m,S},$$

$$\mathrm{A}_{m,S} \stackrel{\text{def}}{=} \sum_{R\in\mathcal{R}(m,S)} \nu_{R,S}\alpha(R)\, H_{0,m,R,S} \otimes H_{1,m,R,S},$$

$$H_{j,m,R,S} \stackrel{\text{def}}{=} \prod_{j\in\mathbb{O}_j} h_{S_j}(x_j) \prod_{k\notin\mathbb{O}_j} h_{R_k}(x_k), \qquad j = 0,1.$$

The main point is that Proposition 2.8 applies to the operators $\mathrm{A}_{m,S}$, giving the estimate

$$\|\mathrm{A}_{m,S}\|_2 \lesssim 2^{\delta(p)\sum_{\mathbb{O}} m_j} \Bigg[\sum_{R\in\mathcal{R}(m,S)} |\alpha(R)|^p\Bigg]^{1/p}.$$

And this, plus a simple argument, gives us the estimate

$$\|\mathrm{A}_m\|_{\mathbb{S}^p} \lesssim 2^{\delta(p)\sum_{\mathbb{O}} m_j} \|\alpha(\cdot)\|_{\ell^p}.$$

Finally, it is the case that for appropriate choices of signs $\nu_{R,S}$, we have

$$\sum_{R\in\mathcal{R}} \alpha(R)\, h_R^{\epsilon} \otimes h_R^{\delta} = \sum_{m\in\mathbb{N}^{\mathbb{O}}} 2^{-\frac{1}{2}\sum_{\mathbb{O}} m_j} \mathrm{A}_m.$$

This is clearly enough to prove the theorem for both the upper bound and lower bound of (5.18) following the arguments in Section 5.1.1.

6. Concluding Remarks

We close this paper by making several concluding remarks. First, we chose to work with the Haar/wavelet bases for $L^2(\mathbb{R})$, and then the appropriate generalization of these to the tensor product setting for $L^2(\mathbb{R}^n)$. It is possible to formulate and prove analogous results by starting with the Haar/wavelet basis for $L^2(\mathbb{R}^{n_j})$ and then forming the d-parameter tensor product bases for $L^2(\mathbb{R}^{\sum_{j=1}^d n_j})$. The only additional difficulty that this produces is more complicated notation. The interested reader can easily modify the results appearing in this paper to arrive at these more general theorems.

In this paper we have focused on continuous and dyadic paraproducts. As mentioned in the introduction, paraproducts are an equivalent way to view Hankel operators. Once estimates for the paraproducts are known, then using a nice result of S. Petermichl [23] it is possible to recover the commutator operator from an averaging of a dyadic shift operator; this commutator can then be understood as a sum of a Hankel operator and it adjoint. Using these results one can prove that a Hankel operator belongs to the Schatten class $\mathbb{S}^p$ if the symbol belongs to the Besov space $\mathbb{B}^p$. It is carried out very nicely in the paper [24] of Pott–Smith in their discussion of $\mathbb{S}^p$ class, $1 < p < \infty$, of one-parameter dyadic paraproducts, and the related Besov spaces.

Acknowledgement: The authors really appreciate the very helpful comments from the referees' reports, which made this paper more readable and accurate.

Lacey is supported in part by National Science Foundation DMS Grant. B. D. Wick's research is supported in part by National Science Foundation Grants DMS # 1800057, # 2054863, and # 2000510 and Australian Research Council DP 220100285. Li is supported by ARC DP 220100285.

References

[1] A Bonami and M. M. Peloso, *Hankel operators on Hardy spaces of two classical domains*, Various Publ. Ser. (Aarhus), vol. 45, Univ. Aarhus, Aarhus, 2000. ↑1

[2] A.P. Calderón, *Commutators of Singular Integral Operators*, Proc. Nat. Acad. Sci., USA **53** (1965), 1092–1099. ↑1

[3] J.-A. Chao and L. Peng, *Schatten classes and commutators on simple martingales*, Colloq. Math. **71** (1996), no. 1, 7–21. ↑2, 3, 4

[4] S.-Y. A. Chang and R. Fefferman, *A continuous version of duality of H^1 with BMO on the bidisc*, Ann. of Math. (2) **112** (1980), no. 1, 179–201. ↑5

[5] R. Coifman, P.L. Lions, Y. Meyer, and S. Semmes, *Compensated compactness and Hardy spaces*, J. Math. Pures Appl. **72** (1993), 247–286. ↑1

[6] R. R. Coifman, R. Rochberg, and G. Weiss, *Factorization theorems for Hardy spaces in several variables*, Ann. of Math. **103** (1976), 611–635. ↑1

[7] A. Connes, *Noncommutative geometry*, Academic Press, Inc., San Diego, CA, 1994. ↑1

[8] Y. Han, J. Li, and L. A. Ward, *Hardy space theory on spaces of homogeneous type via orthonormal wavelet bases*, Appl. Comput. Harmon. Anal. **45** (2018), no. 1, 120–169. ↑8, 9, 10

[9] J.-L. Journé, *Calderón-Zygmund operators on product spaces*, Rev. Mat. Iberoamericana **1** (1985), no. 3, 55–91. ↑5

[10] S. Janson and J. Peetre, *Paracommutators—boundedness and Schatten-von Neumann properties*, Trans. Amer. Math. Soc. **305** (1988), no. 2, 467–504. ↑3

[11] S. Janson and T. Wolff, *Schatten classes and commutators of singular integral operators*, Ark. Mat **20** (1982), 301–310. ↑1

[12] M. T. Lacey, J. Li, and B. D. Wick, *Schatten Classes and Commutators in the Two Weight Setting, I. Hilbert Transform*, Potential Anal. **60** (2024), no. 2, 875–894. ↑9

[13] M. T. Lacey, J. Li, B. D. Wick, and L. Wu, *Schatten classes and commutators of Riesz transforms in the two weight setting*, J. Funct. Anal. **289** (2025), no. 6, 111028. ↑9

[14] M. Lacey and J. Metcalfe, *Paraproducts in one and several parameters*, Forum Math. **19** (2007), no. 2, 325–351. ↑3, 5

[15] C. Li, *Boundedness of paracommutators on L^p-spaces*, Acta Math. Sinica **6** (1990), no. 2, 131–147. ↑

[16] J. Li, J. Pipher, and L. A. Ward, *Dyadic structure theorems for multiparameter function spaces*, Rev. Mat. Iberoam. **31** (2015), no. 3, 767–797. ↑10

[17] S. Lord, E. McDonald, F. Sukochev, and D. Zanin, *Quantum differentiability of essentially bounded functions on Euclidean space*, J. Funct. Anal. **273** (2017), no. 7, 2353–2387. ↑1

[18] C. Muscalu, J. Pipher, T. Tao, and C. Thiele, *Multi-parameter paraproducts*, Rev. Mat. Iberoam. **22** (2006), no. 3, 963–976. ↑5

[19] ———, *Bi-parameter paraproducts*, Acta Math. **193** (2004), no. 2, 269–296. ↑5

[20] F. Nazarov, G. Pisier, S. Treil, and A. Volberg, *Sharp estimates in vector Carleson imbedding theorem and for vector paraproducts*, J. Reine Angew. Math. **542** (2002), 147–171. ↑3

[21] M. Paluszyński, *Characterization of the Besov spaces via the commutator operator of Coifman, Rochberg and Weiss*, Indiana Univ. Math. J. **44** (1995), no. 1, 1–17. ↑

[22] V. V. Peller, *Vectorial Hankel operators, commutators and related operators of the Schatten-von Neumann class γ_p*, Integral Equations Operator Theory **5** (1982), no. 2, 244–272. ↑1

[23] S. Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*, C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), no. 6, 455–460 (English, with English and French summaries). ↑26
[24] S. Pott and M. P. Smith, *Paraproducts and Hankel operators of Schatten class via p-John-Nirenberg theorem*, J. Funct. Anal. **217** (2004), no. 1, 38–78. ↑1, 3, 4, 6, 26
[25] R. Rochberg and S. Semmes, *Nearly weakly orthonormal sequences, singular value estimates, and Calderón–Zygmund operators*, J. Funct. Anal. **86** (1989), no. 2, 237–306. ↑1, 3, 13
[26] H. Triebel, *Theory of function spaces*, Monographs in Mathematics, vol. 78, Birkhäuser Verlag, Basel, 1983. ↑8
[27] A. Uchiyama, *On the compactness of operators of Hankel type*, Tôhoku Math. J. **30** (1978), 163–171. ↑1

Michael T. Lacey, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 USA
Email address: `lacey@math.gatech.edu`

Ji Li, School of Mathematical and Physical Sciences, Macquarie University, NSW 2109, Australia
Email address: `ji.li@mq.edu.au`

Brett D. Wick, Department of Mathematics, Washington University - St. Louis, St. Louis, MO 63130-4899 USA
Email address: `bwick@wustl.edu`