Acta Math., 189 (2002), 143–160

A characterization of product *BMO* by commutators

by

SARAH H. FERGUSON and MICHAEL T. LACEY

Wayne State University
Detroit, MI, U.S.A.

Georgia Institute of Technology
Atlanta, GA, U.S.A.

1. Introduction

In this paper we establish a commutator estimate which allows one to concretely identify the *product BMO* space, $BMO(\mathbf{R}^2_+\times\mathbf{R}^2_+)$, of A. Chang and R. Fefferman, as an operator space on $L^2(\mathbf{R}^2)$. The one-parameter analogue of this result is a well-known theorem of Nehari [8]. The novelty of this paper is that we discuss a situation governed by a two-parameter family of dilations, and so the spaces H^1 and BMO have a more complicated structure.

Here $\mathbf{R}^2_+$ denotes the upper half-plane and $BMO(\mathbf{R}^2_+\times\mathbf{R}^2_+)$ is defined to be the dual of the real-variable Hardy space H^1 on the product domain $\mathbf{R}^2_+\times\mathbf{R}^2_+$. There are several equivalent ways to define this latter space, and the reader is referred to [5] for the various characterizations. We will be more interested in the biholomorphic analogue of H^1, which can be defined in terms of the boundary values of biholomorphic functions on $\mathbf{R}^2_+\times\mathbf{R}^2_+$ and will be denoted throughout by $H^1(\mathbf{R}^2_+\times\mathbf{R}^2_+)$, cf. [10].

In one variable, the space $L^2(\mathbf{R})$ decomposes as the direct sum $H^2(\mathbf{R})\oplus\overline{H^2(\mathbf{R})}$, where $H^2(\mathbf{R})$ is defined as the boundary values of functions in $H^2(\mathbf{R}^2_+)$ and $\overline{H^2(\mathbf{R})}$ denotes the space of complex conjugate of functions in $H^2(\mathbf{R})$. The space $L^2(\mathbf{R}^2)$, therefore, decomposes as the direct sum of the four spaces $H^2(\mathbf{R})\otimes H^2(\mathbf{R})$, $\overline{H^2(\mathbf{R})}\otimes H^2(\mathbf{R})$, $H^2(\mathbf{R})\otimes\overline{H^2(\mathbf{R})}$ and $\overline{H^2(\mathbf{R})}\otimes\overline{H^2(\mathbf{R})}$, where the tensor products are the Hilbert space tensor products. Let $P_{\pm,\pm}$ denote the orthogonal projection of $L^2(\mathbf{R}^2)$ onto the holomorphic/anti-holomorphic subspaces, in the first and second variables, respectively, and let H_j denote the one-dimensional Hilbert transform in the jth variable, $j=1,2$. In terms of the projections $P_{\pm,\pm}$,

$$H_1=P_{+,+}+P_{+,-}-P_{-,+}-P_{-,-} \quad\text{and}\quad H_2=P_{+,+}+P_{-,+}-P_{+,-}-P_{-,-}.$$

The first author was supported by NSF Grant DMS-0071514, while the second author was supported by NSF Grant DMS-9706884.

The nested commutator determined by the function b is the operator $[[M_b, H_1], H_2]$ acting on $L^2(\mathbf{R}^2)$, where, for a function b on the plane, we define $M_b f := bf$.

In terms of the projections $P_{\pm,\pm}$, it takes the form

$$\tfrac{1}{4}[[M_b, H_1], H_2] = P_{+,+} M_b P_{-,-} - P_{+,-} M_b P_{-,+} - P_{-,+} M_b P_{+,-} + P_{-,-} M_b P_{+,+}. \quad (1.1)$$

Ferguson and Sadosky [4] established the inequality $\|[[M_b, H_1], H_2]\|_{L^2} \leqslant c\|b\|_{BMO}$. The main result is the converse inequality.

THEOREM 1.2. *There is a constant $c>0$ such that $\|b\|_{BMO} \leqslant c\|[[M_b, H_1], H_2]\|_{L^2\to L^2}$ for all functions b in $BMO(\mathbf{R}^2_+ \times \mathbf{R}^2_+)$.*

As A. Chang and R. Fefferman have established for us, the structure of the space BMO is more complicated in the two-parameter setting, requiring a more subtle approach to this theorem, despite the superficial similarity of the results to the one-parameter setting. The proof relies on three key ideas. The first is the dyadic characterization of the BMO norm given in [1]. The second is a variant of Journé's lemma, [6], (whose proof is included in the appendix). The third idea is that we have the estimates, the second of which was shown in [4],

$$\|b\|_{BMO(\mathrm{rec})} \leqslant c\|[[M_b, H_1], H_2]\|_{L^2\to L^2} \leqslant c'\|b\|_{BMO}.$$

An unpublished example of L. Carleson shows that the rectangular BMO norm is not comparable to the BMO norm, [3]. We may assume that the rectangular BMO norm of the function b is small. Indeed, this turns out to be an essential aspect of the argument.

From Theorem 1.2 we deduce a *weak* factorization for the (biholomorphic) space $H^1(\mathbf{R}^2_+ \times \mathbf{R}^2_+)$. The idea is that if the function b has biholomorphic extension to $\mathbf{R}^2_+ \times \mathbf{R}^2_+$ then for functions $f, g \in L^2(\mathbf{R}^2)$,

$$\tfrac{1}{4}\langle [[M_b, H_1], H_2] f, g\rangle = \langle b, \overline{P_{-,-} f} P_{+,+} g\rangle.$$

So in this case, the operator norm of the nested commutator $[[M_b, H_1], H_2]$ is comparable to the *dual* norm

$$\|b\|_* := \sup |\langle fg, b\rangle|,$$

where the supremum above is over all pairs f, g in the unit ball of $H^2(\mathbf{R}^2_+ \times \mathbf{R}^2_+)$. On the other hand, since $\|b\|_{BMO}$ and $\|[[M_b, H_1], H_2]\|_{L^2\to L^2}$ are comparable, the dual norm above satisfies

$$\|b\|_* \sim \sup |\langle h, b\rangle|,$$

where the supremum is over all functions h in the unit ball of $H^1(\mathbf{R}^2_+ \times \mathbf{R}^2_+)$.

COROLLARY 1.3. *Let h be in $H^1(\mathbf{R}^2_+\times\mathbf{R}^2_+)$ with $\|h\|_1=1$. Then there exist functions $(f_j),(g_j)\subseteq H^2(\mathbf{R}^2_+\times\mathbf{R}^2_+)$ such that $h=\sum_{j=1}^\infty f_jg_j$ and $\sum_{j=1}^\infty \|f_j\|_2\|g_j\|_2\leqslant c$.*

We remark that the weak factorization above implies the analogous factorization for H^1 of the bidisk. Indeed, for all $1\leqslant p<\infty$, the map u_p: $H^p(\mathbf{R}^2_+\times\mathbf{R}^2_+)\to H^p(\mathbf{D}^2)$ defined by

$$(u_pf)(z,w)=\pi^{2/p}\left(\frac{2i}{1-z}\right)^{2/p}\left(\frac{2i}{1-w}\right)^{2/p}f(\alpha(z),\alpha(w)),\quad \alpha(\lambda):=i\frac{1+\lambda}{1-\lambda},$$

is an isometry with isometric inverse

$$(u_p^{-1}g)(z,w)=\pi^{-2/p}\left(\frac{1}{z+i}\right)^{2/p}\left(\frac{1}{w+i}\right)^{2/p}g(\beta(z),\beta(w)),\quad \beta(\lambda):=\frac{\lambda-i}{\lambda+i}.$$

The dual formulation of weak factorization for $H^1(\mathbf{D}^2)$ is a Nehari theorem for the bidisk. Specifically, if $b\in H^2(\mathbf{D}^2)$ then the *little* Hankel operator with symbol b is densely defined on $H^2(\mathbf{D}^2)$ by the formula

$$\Gamma_bf=P_{-,-}(\bar{b}f).$$

By (1.1), $\|\Gamma_b\|=\|[[M_{\bar b},H_1],H_2]\|_{L^2\to L^2}$ and thus, by Theorem 1.2, $\|\Gamma_b\|$ is comparable to $\|b\|_{BMO}$, which, by definition, is just the norm of b acting on $H^1(\mathbf{D}^2)$. So the boundedness of the Hankel operator Γ_b implies that there is a function $\phi\in L^\infty(\mathbf{T}^2)$ such that $P_{+,+}\phi=b$.

Several variations and complements on these themes in the one-parameter setting have been obtained by Coifman, Rochberg and Weiss [2].

The paper is organized as follows. §2 gives the one-dimensional preliminaries for the proof of Theorem 1.2, and §3 is devoted to the proof of Theorem 1.2. The appendix contains the variant of Journé's lemma that we use in our proof in §3. One final remark about notation. $A\lesssim B$ means that there is an absolute constant C for which $A\leqslant CB$. $A\approx B$ means that $A\lesssim B$ and $B\lesssim A$.

We are indebted to the anonymous referee and Andreas Seeger.

2. Remarks on the one-dimensional case

Several factors conspire to make the one-dimensional case easier than the higher-dimensional case. Before proceeding with the higher-dimensional case, we make several comments about the one-dimensional case, comments that extend and will be useful in the subsequent section.

Let H denote the Hilbert transform in one variable, $P_+=\frac{1}{2}(I+H)$ be the projection of $L^2(\mathbf{R})$ onto the positive frequencies, and P_- is $\frac{1}{2}(I-H)$, the projection onto the negative frequencies. We shall in particular rely upon the following basic computation:

$$\tfrac{1}{2}[M_b,H]\bar{b}=P_-|P_-b|^2-P_+|P_+b|^2. \tag{2.1}$$

The frequency distribution of $|P_-b|^2$ is symmetric since it is real-valued. Thus,

$$\begin{aligned}\|b\|_4^2 &\lesssim \|P_-|P_-b|^2 - P_+|P_+b|^2\|_2 \\ &\leqslant \|[M_b, H]\|_{2\to 2}\|b\|_2.\end{aligned}$$

Moreover, if b is supported on an interval I, we see that

$$\|b\|_2 \leqslant |I|^{1/4}\|b\|_4 \lesssim |I|^{1/4}\|[M_b, H]\|_{2\to 2}^{1/2}\|b\|_2^{1/2},$$

which is the BMO estimate on I. We seek an extension of this estimate in the two-parameter setting.

We use a wavelet proof of Theorem 1.2, and specifically use a wavelet with compact frequency support constructed by Y. Meyer [7]. There is a Schwartz function w with these properties:

- $\|w\|_2 = 1$.
- $\widehat{w}(\xi)$ is supported on $\left[\frac{2}{3}, \frac{8}{3}\right]$ together with the symmetric interval about 0.
- $P_\pm w$ is a Schwartz function. More particularly, we have

$$|w(x)|, |P_\pm w(x)| \lesssim (1+|x|)^{-n}, \quad n \geqslant 1.$$

Let $\mathcal{D}$ denote a collection of dyadic intervals on $\mathbf{R}$. For any interval I, let $c(I)$ denote its center, and define

$$w_I(x) := \frac{1}{\sqrt{|I|}}\, w\left(\frac{x - c(I)}{|I|}\right).$$

Set $w_I^\pm := P_\pm w_I$. The central facts that we need about the functions $\{w_I : I \in \mathcal{D}\}$ are these:

First, that these functions are an orthonormal basis on $L^2(\mathbf{R})$. Second, that we have the Littlewood–Paley inequalities, valid on all L^p, though $p=4$ will be of special significance for us. These inequalities are

$$\|f\|_p \approx \left\| \left[\sum_{I\in\mathcal{D}} \frac{|\langle f, w_I\rangle|^2}{|I|} \mathbf{1}_I \right]^{1/2} \right\|_p, \quad 1 < p < \infty. \tag{2.2}$$

Third, that the functions w_I have good localization properties in the spatial variables. That is,

$$|w_I(x)|, |w_I^\pm(x)| \lesssim |I|^{-1/2}\chi_I(x)^n, \quad n \geqslant 1, \tag{2.3}$$

where $\chi_I(x) := (1 + \operatorname{dist}(x, I)/|I|)^{-1}$. We find the compact localization of the wavelets in frequency to be very useful. The price we pay for this utility below is the careful accounting of "Schwartz tails" we shall make in the main argument. Fourth, we have the identity below for the commutator of one w_I with a w_J. Observe that since P_+ is one half of $I+H$, it suffices to replace H by P_+ in the definition of the commutator.

$$\begin{aligned} w_{I,J} &:= [w_I, P_+]\overline{w_J} = w_I\overline{w_J^-} - P_+ w_I \overline{w_J} \\ &= P_- w_I \overline{w_J^-} - P_+ w_I \overline{w_J^+} = P_- w_I^- \overline{w_J^-} - P_+ w_I^+ \overline{w_J^+} \\ &= \begin{cases} 0 & \text{if } |I| \geqslant 4|J|, \\ P_-|w_I^-|^2 - P_+|w_I^+|^2 & \text{if } I = J, \\ w_I^- \overline{w_J^-} - w_I^+ \overline{w_J^+} & \text{if } |J| \geqslant 4|I|. \end{cases} \end{aligned} \tag{2.4}$$

From this we see a useful point concerning orthogonality. For intervals I, I', J and J', assume $|J| \geqslant 8|I|$, and likewise for I' and J'. Then

$$\operatorname{supp}(\widehat{w_{I,J}}) \cap \operatorname{supp}(\widehat{w_{I',J'}}) = \varnothing, \quad |I'| \geqslant 8|I|. \tag{2.5}$$

Indeed, this follows from a direct calculation. The positive frequency support of $w_I^+ \overline{w_J^+}$ is contained in the interval $[(3|I|)^{-1}, 8(3|I|)^{-1}]$. Under the conditions on I and I', the frequency supports are disjoint.

3. Proof of the main theorem

$BMO(\mathbf{R}_+^2 \times \mathbf{R}_+^2)$ will denote the BMO of two parameters (or *product BMO*) defined as the dual of (*real*) $H^1(\mathbf{R}_+^2 \times \mathbf{R}_+^2)$. The following characterization of the space $BMO(\mathbf{R}_+^2 \times \mathbf{R}_+^2)$ is due to A. Chang and R. Fefferman [1].

The relevant class of rectangles is $\mathcal{R} = \mathcal{D} \times \mathcal{D}$, all rectangles which are products of dyadic intervals. These are indexed by $R \in \mathcal{R}$. For such a rectangle, write it as a product $R_1 \times R_2$ and then define

$$v_R(x_1, x_2) = w_{R_1}(x_1) w_{R_2}(x_2).$$

A function $f \in BMO(\mathbf{R}_+^2 \times \mathbf{R}_+^2)$ if and only if

$$\sup_U \left[|U|^{-1} \sum_{R \subset U} |\langle f, v_R \rangle|^2 \right]^{1/2} < \infty.$$

Here, the sum extends over those rectangles $R \in \mathcal{R}$, and the supremum is over all open sets in the plane of finite measure. Note that the supremum is taken over a much broader class of sets than merely rectangles in the plane. We denote this supremum as $\|f\|_{BMO}$.

In this definition, if the supremum over U is restricted to just rectangles, this defines the "rectangular BMO" space, and we denote this restricted supremum as $\|f\|_{BMO(\mathrm{rec})}$.

Let us make a further comment on the BMO condition. Suppose that for $R \in \mathcal{R}$, we have non-negative constants a_R for which

$$\sum_{R \subset U} a_R \leqslant |U|,$$

for all open sets U in the plane of finite measure. Then, we have the John–Nirenberg inequality

$$\Big\| \sum_{R \subset U} |R|^{-1} a_R \mathbf{1}_R \Big\|_p \lesssim |U|^{1/p}, \quad 1 < p < \infty.$$

See [1]. This, with the Littlewood–Paley inequalities, will be used several times below, and referred to as the John–Nirenberg inequalities.

3.1. The principal points in the argument

We begin the principle line of the argument. The function b may be taken to be of Schwartz class. By multiplying b by a constant, we can assume that the BMO norm of b is one. Set $B_{2\to 2}$ to be the operator norm of $[[M_b, H_1], H_2]$. Our purpose is to provide a lower bound for $B_{2\to 2}$. Let U be an open set of finite measure for which we have the equality

$$\sum_{R \subset U} |\langle b, v_R \rangle|^2 = |U|.$$

As b is of Schwartz class, such a set exists. By invariance under dilations by a factor of two, we can assume that $\frac{1}{2} \leqslant |U| \leqslant 1$. In several estimates below, the measure of U enters in, a fact which we need not keep track of.

An essential point is that we may assume that the rectangular BMO norm of b is at most ε. The reason for this is that we have the estimate $\|b\|_{BMO(\mathrm{rec})} \lesssim B_{2\to 2}$. See [4]. Therefore, for a small constant ε to be chosen below, we can assume that $\|b\|_{BMO(\mathrm{rec})} \lesssim \varepsilon$, for otherwise we have a lower bound on $B_{2\to 2}$.

Associated to the set U is a set V which contains U and has the properties specified in Lemma A.1. It is critical that the measure of V be only slightly larger than the measure of U, or more exactly, $|V| < (1+\delta)|U|$, for a choice of $0 < \delta < 1$ to be specified. Define

$$\mu(R) := \sup\{\mu : \mu R \subset V\}, \quad R \subset U.$$

The quantity $\mu(R)$ measures how deeply a rectangle R is inside V. This quantity enters into the essential Journé's lemma, see [6] or the variant we prove in the Appendix.

In the argument below, we will be projecting b onto subspaces spanned by collections of wavelets. These wavelets are in turn indexed by collections of rectangles. Thus, for a collection $\mathcal{A}\subseteq\mathcal{R}$, let us denote

$$b^{\mathcal{A}} := \sum_{R\in\mathcal{A}} \langle b, v_R\rangle v_R.$$

The relevant collections of rectangles are defined as

$$\begin{aligned} \mathcal{U} &:= \{R\in\mathcal{R} : R\subset U\}, \\ \mathcal{V} &= \{R\in\mathcal{R}-\mathcal{U} : R\subset V\}, \\ \mathcal{W} &= \mathcal{R}-\mathcal{U}-\mathcal{V}. \end{aligned}$$

For functions f and g, we set $\{f,g\} := [[M_f, H_1], H_2]\bar{g}$.

We will demonstrate that for all $\delta,\varepsilon>0$ there is a constant $K_\delta>0$ so that

(i) $\|\{b^{\mathcal{V}}, b^{\mathcal{U}}\}\|_2 \lesssim \delta^{1/4}$,

(ii) $\|\{b^{\mathcal{W}}, b^{\mathcal{U}}\}\|_2 \leqslant K_\delta \varepsilon^{1/3}$.

Furthermore, we will show that $1\lesssim\|\{b^{\mathcal{U}}, b^{\mathcal{U}}\}\|_2$. Since $b=b^{\mathcal{U}}+b^{\mathcal{V}}+b^{\mathcal{W}}$, $\|b^{\mathcal{U}}\|_2\lesssim 1$ and $\delta,\varepsilon>0$ are arbitrary, a lower bound on $B_{2\to 2}$ will follow from an appropriate choice of δ and ε. To be specific, one concludes the argument by estimating

$$\begin{aligned} 1 &\lesssim \|\{b^{\mathcal{U}}, b^{\mathcal{U}}\}\|_2 \\ &\lesssim \|\{b^{\mathcal{U}}+b^{\mathcal{V}}, b^{\mathcal{U}}\}\|_2 + \delta^{1/4} \\ &\lesssim \|\{b^{\mathcal{U}}+b^{\mathcal{V}}+b^{\mathcal{W}}, b^{\mathcal{U}}\}\|_2 + \delta^{1/4} + K_\delta \varepsilon^{1/3} \\ &\lesssim B_{2\to 2} + \delta^{1/4} + K_\delta \varepsilon^{1/3}. \end{aligned}$$

Implied constants are absolute. Choosing δ first and then ε appropriately small supplies a lower bound on $B_{2\to 2}$.

The estimate $1\lesssim\|\{b^{\mathcal{U}}, b^{\mathcal{U}}\}\|_2$ relies on the John–Nirenberg inequality and the two-parameter version of (2.1), namely

$$\tfrac{1}{4}[[M_b, H_1], H_2]\bar{b} = P_{+,+}|P_{+,+}b|^2 - P_{+,-}|P_{+,-}b|^2 - P_{-,+}|P_{-,+}b|^2 + P_{-,-}|P_{-,-}b|^2.$$

This identity easily follows from the one-variable identities. Here $P_{\pm,\pm}$ denotes the projection onto the positive/negative frequencies in the first and second variables. These projections are orthogonal and moreover, since $|P_{\pm,\pm}b|^2$ is real-valued we have that $\|P_{\pm,\pm}|P_{\pm,\pm}b|^2\|_2 \geqslant \frac{1}{4}\||P_{\pm,\pm}b|^2\|_2$. Therefore, $\|b^{\mathcal{U}}\|_4^2 \lesssim \|\{b^{\mathcal{U}}, b^{\mathcal{U}}\}\|_2$. It follows that

$$\begin{aligned} 1 \lesssim \|b^{\mathcal{U}}\|_2 &= \Bigl[\sum_{R\in\mathcal{U}} |\langle b, v_R\rangle|^2\Bigr]^{1/2} \lesssim \Bigl\|\Bigl[\sum_{R\in\mathcal{U}} \frac{|\langle b, v_R\rangle|^2}{|R|}\mathbf{1}_R\Bigr]^{1/2}\Bigr\|_4 \\ &\lesssim \|b^{\mathcal{U}}\|_4 \lesssim \|\{b^{\mathcal{U}}, b^{\mathcal{U}}\}\|_2^{1/2}. \end{aligned}$$

The estimate (i) relies on the estimate $|V|<(1+\delta)|U|$. Now, if $R\in\mathcal{V}$, then $R\subset V$ and since b has BMO norm one, it follows that

$$|U|+\|b^{\mathcal{V}}\|_2^2=\sum_{R\in\mathcal{U}\cup\mathcal{V}}|\langle b,v_R\rangle|^2\leqslant(1+\delta)|U|.$$

Hence $\|b^{\mathcal{V}}\|_2\lesssim\delta^{1/2}$. Yet the BMO norm of $b^{\mathcal{V}}$ can be no more than that of b, which is to say one. Interpolating norms we see that $\|b^{\mathcal{V}}\|_4\lesssim\delta^{1/4}$, and so

$$\|\{b^{\mathcal{V}},b^{\mathcal{U}}\}\|_2\lesssim\|b^{\mathcal{V}}\|_4\|b^{\mathcal{U}}\|_4\lesssim\delta^{1/4}.$$

3.2. Verifying the estimate (ii)

We now turn to the estimate (ii). Roughly speaking $b^{\mathcal{U}}$ and $b^{\mathcal{W}}$ live on disjoint sets. But in this argument we are trading off precise Fourier support of the wavelets for imprecise spatial localization, that is the "Schwartz tails" problem. Accounting for this requires a careful analysis, invoking several subcases.

A property of the commutator that we will rely upon is that it controls the geometry of R and R'. Namely, $\{v_{R'},v_R\}\neq0$ if and only if writing $R=R_1\times R_2$ and likewise for R', we have for both $j=1,2$, $|R'_j|\leqslant4|R_j|$. This follows immediately from our one-dimensional calculations, in particular (2.4). We abbreviate this condition on R and R' as $R'\lesssim R$ and restrict our attention to this case.

Orthogonality also enters into the argument. Observe the following. For rectangles $R^k,\widetilde{R}^k$, $k=1,2$, with $\widetilde{R}^k\lesssim R^k$, and for $j=1$ or $j=2$,

$$\text{if } 8|\widetilde{R}^1_j|\leqslant|R^1_j| \text{ and } 8|\widetilde{R}^2_j|<|R^2_j|, \quad \text{then } \langle v_{\widetilde{R}^1}\overline{v_{R^1}},v_{\widetilde{R}^2}\overline{v_{R^2}}\rangle=0. \tag{3.1}$$

This follows from applying (2.5) in the jth coordinate.

Therefore, there are different partial orders on rectangles that are relevant to our argument. They are:

- $R'<R$ if and only if $8|R'_j|\leqslant|R_j|$ for $j=1$ and $j=2$.
- For $j=1$ or $j=2$, define $R'<_jR$ if and only if $R'\lesssim R$ and $8|R'_j|\leqslant|R_j|$ but $R'\not<R$.
- $R'\simeq R$ if and only if $\frac14|R_j|\leqslant|R'_j|\leqslant|R_j|$ for $j=1$ and $j=2$.

These four partial orders divide the collection $\{(R',R):R'\in\mathcal{W},\,R\in\mathcal{U},\,R'\lesssim R\}$ into four subclasses which require different arguments.

In each of these four arguments, we have recourse to this definition. Set $\mathcal{U}_k$, for $k=0,1,2,\ldots$, to be those rectangles in $\mathcal{U}$ with $2^{-k-1}<\mu(R)\leqslant2^k$, $R\in\mathcal{U}_k$.

Journé's lemma enters into the considerations. Let $\mathcal{U}'\subset\mathcal{U}_k$ be a collection of rectangles which are pairwise incomparable with respect to inclusion. For this latter collection, we have the inequality

$$\sum_{R\in\mathcal{U}'} |R| \leqslant K_\delta 2^{k/100} \Big| \bigcup_{R\in\mathcal{U}'} R \Big|. \tag{3.2}$$

See Journé [6], also see the Appendix. This together with the assumption that b has small rectangular *BMO* norm gives us

$$\|b^{\mathcal{U}_k}\|_{BMO} \leqslant K_\delta 2^{k/100} \varepsilon. \tag{3.3}$$

This interplay between the small rectangular *BMO* norm and Journé's lemma is a decisive feature of the argument.

Essentially, the decomposition into the collections $\mathcal{U}_k$ is a spatial decomposition of the collection $\mathcal{U}$. A corresponding decomposition of $\mathcal{W}$ enters in. Yet the definition of this class differs slightly depending on the partial order we are considering.

For $R'\in\mathcal{W}$ and $R\in\mathcal{U}$ the term $\{v^{R'}, v^R\}$ is a linear combination of

$$v_{R'} H_2 H_1 \overline{v_R}, \quad H_2(v_{R'} H_1 \overline{v_R}), \quad (H_1 v_{R'})(H_2 \overline{v_R}), \quad H_1 H_2 (v_{R'} \overline{v_R}).$$

Consider the last term. As we are to estimate an L^2-norm, the leading operators $H_1 H_2$ can be ignored. Moreover, the essential properties of wavelets used below still hold for the conjugates and Hilbert transforms of the same. These properties are Fourier localization and spatial localization. Similar comments apply to the other three terms, and so the arguments below applies to each type of term above.

3.2.1. *The partial order* '$<$'. We consider the case of $R'<R$ for $R'\in\mathcal{W}$ and $R\in\mathcal{U}$. The sums we consider are related to the following definition. Set

$$b^{\mathcal{U}_k}_{\mathrm{trun}}(x) := \sup_{R'} \Big| \sum_{\substack{R\in\mathcal{U}_k \\ R'<R}} \langle b, v_R\rangle v_R(x) \Big|.$$

Note that we consider the maximal truncation of the sum over all choices of dimensions of the rectangles in the sum. Thus, this sum is closely related to the strong maximal function M applied to $b^{\mathcal{U}_k}$, so that in particular we have the estimate below, which relies upon (3.3):

$$\|b^{\mathcal{U}_k}_{\mathrm{trun}}\|_p \lesssim \varepsilon 2^{k/100}, \quad 1<p<\infty.$$

(By a suitable definition of the strong maximal function M, one can deduce this inequality from the L^p-bounds for M.) We apply this inequality far away from the set U. For the set $W_\lambda = \mathbf{R}^2 - \bigcup_{R\in\mathcal{U}_k} \lambda R$, $\lambda>1$, we have the inequality

$$\|b^{\mathcal{U}_k}_{\mathrm{trun}}\|_{L^p(W)} \lesssim \varepsilon 2^{k/100} \lambda^{-100}, \quad 1<p<\infty. \tag{3.4}$$

We shall need a refined decomposition of the collection $\mathcal{W}$, the motivation for which is the following calculation. Let $\mathcal{W}'\subset\mathcal{W}$. For $n=(n_1,n_2)\in\mathbf{Z}^2$, set

$$\mathcal{W}'(n):=\{R'\in\mathcal{W}':|R'_j|=2^{n_j},\, j=1,2\}.$$

In addition, let

$$B(\mathcal{W}',n):=\sum_{R'\in\mathcal{W}'(n)}\sum_{\substack{R\in\mathcal{U}_k\\ R'<R}}\langle b,v_{R'}\rangle\overline{\langle b,v_R\rangle}\,v_{R'}\overline{v_R}.$$

And set $B(\mathcal{W}')=\sum_{n\in\mathbf{Z}^2}B(\mathcal{W}',n)$.

Then, in view of (3.1), we see that $B(\mathcal{W}',n)$ and $B(\mathcal{W}',n')$ are orthogonal if n and n' differ by at least three in either coordinate. Thus,

$$\Big\|\sum_{n\in\mathbf{Z}^2}B(\mathcal{W}',n)\Big\|_2^2\leqslant 3\sum_{n\in\mathbf{Z}^2}\|B(\mathcal{W}',n)\|_2^2.$$

The rectangles $R'\in\mathcal{W}(n)$ are all translates of one another. Thus, taking advantage of the rapid spatial decay of the wavelets, we can estimate

$$\|B(\mathcal{W}',n)\|_2^2\lesssim\sum_{R'\in\mathcal{W}(n)}\int\left|\frac{|\langle b,v_{R'}\rangle|}{\sqrt{|R'|}}(\chi_{R'}*\mathbf{1}_{R'})\,b^{\mathcal{U}_k}_{\mathrm{trun}}\right|^2dx.$$

In this display, we let $\chi(x_1,x_2)=(1+x_1^2+x_2^2)^{-10}$ and for rectangles R, $\chi_R(x_1,x_2)=\chi(x_1|R_1|^{-1},x_2|R_2|^{-1})$. Note that χ_R depends only on the dimensions of R and not its location.

Continuing, note the trivial inequality $\int(\chi_R*f)^2g\,dx\lesssim\int|f|^2\chi_R*g\,dx$. We can estimate

$$\begin{aligned}\|B(\mathcal{W}')\|_2^2&\lesssim\sum_{R'\in\mathcal{W}'}|\langle b,v_{R'}\rangle|^2\Big\{|R'|^{-1}\int_{R'}M(|b^{\mathcal{U}_k}_{\mathrm{trun}}|^2)\,dx\Big\}\\&\lesssim\Big|\bigcup_{R'\in\mathcal{W}'}R'\Big|\sup_{R'\in\mathcal{W}'}\operatorname{avg}(R').\end{aligned}\tag{3.5}$$

Here we take $\operatorname{avg}(R'):=|R'|^{-1}\int_{R'}M(|b^{\mathcal{U}_k}_{\mathrm{trun}}|^2)$.

The terms $\operatorname{avg}(R')$ are essentially of the order of magnitude ε^2 times the scaled distance between R' and the open set U. To make this precise requires a decomposition of the collection $\mathcal{W}$.

For integers $l>k$ and $m\geqslant0$, set $\mathcal{W}(l,m)$ to be those $R'\in\mathcal{W}$ which satisfy these three conditions:

- First, $\operatorname{avg}(R')\leqslant\varepsilon^2 2^{-4l}$ if $m=0$ and $\varepsilon^2 2^{-4l+m-1}<\operatorname{avg}(R')\leqslant\varepsilon^2 2^{-4l+m}$ if $m>0$.
- Second, there is an $R\in\mathcal{U}_k$ with $R'<R$ and $R'\subset2^{l+1}R$.
- Third, for every $R\in\mathcal{U}_k$ with $R'<R$, we have $R'\not\subset2^{l+1}R$. Certainly, this collection of rectangles is empty if $l\leqslant k$.

We see that

$$\Bigl|\bigcup_{R'\in\mathcal{W}(l,m)} R'\Bigr| \lesssim \min(2^{2lp}, 2^{-mp/2}), \quad 1<p<\infty.$$

The first estimate follows since the rectangles $R'\in\mathcal{W}(l,m)$ are contained in the set $\{M\mathbf{1}_U \geqslant 2^{-2l-1}\}$. The second estimate follows from (3.4).

But then from (3.5) we see that for $m>0$,

$$\|B(\mathcal{W}(l,m))\|_2^2 \lesssim \varepsilon^2 2^{-4l+m} \min(2^{2lp}, 2^{-mp/2}) \lesssim \varepsilon^2 2^{-(m+l)/10}.$$

In the case that $m=0$, we have the bound 2^{2lp}. This is obtained by taking the minimum to be 2^{2lp} for $p=\frac{5}{4}$ and $0<m<\frac{11}{8}l$. For $m\geqslant\frac{11}{8}l$ take the minimum to be $2^{-mp/2}$ with $p=4$.

This last estimate is summable over $0<k<l$ and $0<m$ to at most $\lesssim\varepsilon$, and so completes this case.

3.2.2. *The partial orders '$<_j$', $j=1,2$.* We treat the case of $R'<_1 R$, while the case of $R'<_2 R$ is the same by symmetry. The structure of this partial order provides some orthogonality in the first variable, leaving none in the second variable. Bounds for the expressions from the second variable are derived from a cognate of a Carleson measure estimate.

There is a basic calculation that we perform for a subset $\mathcal{W}'\subset\mathcal{W}$. For an integer $n'\in\mathbf{Z}$ define $\mathcal{W}'(n'):=\{R'\in\mathcal{W}' : |R_1'|=2^{n'}\}$ and

$$B(\mathcal{W}', n') := \sum_{R'\in\mathcal{W}'(n')} \sum_{\substack{R\in\mathcal{U}_k \\ R'<_1 R}} \langle b, v_{R'}\rangle \overline{\langle b, v_R\rangle}\, v_{R'} \overline{v_R}.$$

Recalling (3.1), if n' and n'' differ by more than 3, then $B(\mathcal{W}', n')$ and $B(\mathcal{W}', n'')$ are orthogonal.

Observe that for R' and R as in the sum defining $B(\mathcal{W}', n)$, we have the estimate

$$|v_{R'}(x)\overline{v_R}(x)| \lesssim (|R|\,|R'|)^{-1/2} \operatorname{dist}(R',R)^{1000} \chi_{R'} * \mathbf{1}_{R'}(x), \quad x\in\mathbf{R}^2. \tag{3.6}$$

In this display, we are using the same notation as before, $\chi(x_1,x_2)=(1+x_1^2+x_2^2)^{-10}$ and for rectangles R, $\chi_R(x_1,x_2)=\chi(x_1|R_1|^{-1}, x_2|R_2|^{-1})$. In addition, $\operatorname{dist}(R',R):= M\mathbf{1}_R(c(R'))$, with $c(R')$ being the center of R'. (This "distance" is more properly the inverse of a distance that takes into account the scale of the rectangle R.)

Now define

$$\beta(R') := \sum_{\substack{R\in\mathcal{U} \\ R'<_1 R}} |R|^{-1/2} |\langle b, v_R\rangle| \operatorname{dist}(R',R)^{1000}. \tag{3.7}$$

The main point of these observations and definitions is this. For the function $B(\mathcal{W}'):=\sum_{n'\in\mathbf{Z}} B(\mathcal{W}',n')$, we have

$$\begin{aligned}\|B(\mathcal{W}')\|_2^2 &\lesssim \sum_{n'\in\mathbf{Z}} \|B(\mathcal{W}',n')\|_2^2 \\ &\lesssim \sum_{n'\in\mathbf{Z}} \int_{\mathbf{R}^2} \Bigg[\sum_{R'\in\mathcal{W}'(n')} |\langle b, v_{R'}\rangle|\,\beta(R')\,|R'|^{-1/2}\chi_{R'}*\mathbf{1}_{R'}\Bigg]^2 dx \\ &\lesssim \sum_{n'\in\mathbf{Z}} \int_{\mathbf{R}^2} \Bigg[\sum_{R'\in\mathcal{W}'(n')} |\langle b, v_{R'}\rangle|\,\beta(R')\,|R'|^{-1/2}\mathbf{1}_{R'}\Bigg]^2 dx.\end{aligned}$$

At this point, it occurs to one to appeal to the Carleson measure property associated to the coefficients $|\langle b, v_{R'}\rangle|\,|R'|^{-1/2}$. This necessitates that one proves that the coefficients $\beta(R')$ satisfy a similar condition, which doesn't seem to be true in general. A slightly weaker condition is however true.

To get around this difficulty, we make a further diagonalization of the terms $\beta(R')$ above. For integers $\nu\geqslant\nu_0$, $\mu\geqslant 1$ and a rectangle $R'\in\mathcal{W}$, consider rectangles $R\in\mathcal{U}_k$ such that

$$R'<_1 R, \quad 2^{-v}\leqslant \operatorname{dist}(R',R)\leqslant 2^{-v+1}, \quad 2^{\mu}|R'|=|R|.$$

(The quantity v_0 depends upon the particular subcollection $\mathcal{W}'$ we are considering.) We denote one of these rectangles as $\pi(R')$.

An important geometrical fact is this. We have $\pi(R')\subset 2^{v+\mu+10}R_1'\times 2^{v+10}R_2'$. And in particular, this last rectangle has measure $\lesssim 2^{2v+\mu}|R'|$.

Therefore, there are at most $O(2^{2v})$ possible choices for $\pi(R')$. (Small integral powers of 2^v are completely harmless because of the large power of $\operatorname{dist}(R',R)$ that appears in (3.7).)

Our purpose is to bound this next expression by a term which includes a power of ε, a small power of 2^v and a power of $2^{-\mu}$. Define

$$\begin{aligned}S(\mathcal{W}',\nu,\mu) &:= \sum_{n'\in\mathbf{Z}} \int_{\mathbf{R}^2} \Bigg[\sum_{R'\in\mathcal{W}'(n')} \frac{|\langle b, v_{R'}\rangle\langle b, v_{\pi(R')}\rangle|}{\sqrt{|R'|\,|\pi(R')|}}\,\chi_{R'}*\mathbf{1}_{R'}\Bigg]^2 dx \\ &\lesssim \sum_{n'\in\mathbf{Z}} \int_{\mathbf{R}^2} \Bigg[\sum_{R'\in\mathcal{W}'(n')} \frac{|\langle b, v_{R'}\rangle\langle b, v_{\pi(R')}\rangle|}{\sqrt{|R'|\,|\pi(R')|}}\,\mathbf{1}_{R'}\Bigg]^2 dx \\ &= \sum_{n'\in\mathbf{Z}} \sum_{R'\in\mathcal{W}'(n')} \frac{|\langle b, v_{R'}\rangle\langle b, v_{\pi(R')}\rangle|}{\sqrt{|R'|\,|\pi(R')|}} \sum_{\substack{R''\in\mathcal{W}'(n')\\ R''\subset R'}} \sqrt{\frac{|R''|}{|\pi(R'')|}}\,|\langle b, v_{R''}\rangle\langle b, v_{\pi(R'')}\rangle|.\end{aligned}$$

The innermost sum can be bounded this way. First $\|b\|_{BMO(\mathrm{rec})}\leqslant\varepsilon$, so that

$$\sum_{R''\subset R'}|\langle b, v_{R''}\rangle|^2\leqslant\varepsilon^2|R'|.$$

Second, by our geometrical observation about $\pi(R')$,

$$\sum_{R''\subset R'}\frac{|R''|}{|\pi(R'')|}|\langle b, v_{\pi(R'')}\rangle|^2\lesssim\varepsilon^2 2^{2v}|R'|.$$

In particular, the factor 2^u does not enter into this estimate.

This means that

$$\begin{aligned}S(\mathcal{W},\nu,\mu)&\lesssim\varepsilon^2 2^{2v}\sum_{R'\in\mathcal{W}'}\sqrt{\frac{|R'|}{|\pi(R')|}}\,|\langle b, v_{R'}\rangle\langle b, v_{\pi(R')}\rangle|\\&\lesssim\varepsilon^2 2^{2v-\mu/2}\Big[\sum_{R'\in\mathcal{W}'}|\langle b, v_{R'}\rangle|^2\sum_{R\in\mathcal{U}_k}|\langle b, v_R\rangle|^2\Big]^{1/2}\\&\lesssim\varepsilon^2 2^{2v-\mu/2}\Big|\bigcup_{R'\in\mathcal{W}'}R'\Big|^{1/2}.\end{aligned}$$

The point of these computations is that a further trivial application of the Cauchy–Schwarz inequality proves that

$$\|B(\mathcal{W}')\|_2\lesssim\varepsilon 2^{-100\nu_0}\Big|\bigcup_{R'\in\mathcal{W}'}R'\Big|^{1/4},$$

where ν_0 is the largest integer such that for all $R'\in\mathcal{W}'$ and $R\in\mathcal{U}_k$, we have $\operatorname{dist}(R',R)\leqslant 2^{-\nu_0}$.

We shall complete this section by decomposing $\mathcal{W}$ into subcollections for which this last estimate is summable to $\varepsilon 2^{-k}$. Indeed, take $\mathcal{W}_v$ to be those $R'\in\mathcal{W}$ with $R'\not\subset 2^vR$ for all $R\in\mathcal{U}_k$ with $R'<_1R$. And there is an $R\in\mathcal{U}_k$ with $R'\subset 2^{v+1}R$ and $R'<_1R$. Certainly, we need only consider $v\geqslant k$.

It is clear that this decomposition of $\mathcal{W}$ will conclude the treatment of this partial order.

3.2.3. *The partial order* '$\simeq$'. We now consider the case of $R'\simeq R$, which is less subtle as there is no orthogonality to exploit and the Carleson measure estimates are more directly applicable. We prove the bound

$$\Big\|\sum_{R'\in\mathcal{W}}\sum_{\substack{R\in\mathcal{U}\\R'\simeq R}}\langle b, v_{R'}\rangle\overline{\langle b, v_R\rangle}\,v_{R'}\overline{v_R}\Big\|_2\lesssim K_\delta\varepsilon^{1/3}.$$

The diagonalization in space takes two different forms. For $\lambda \geqslant 2^k$ and $R \in \mathcal{U}_k$ set $\sigma(\lambda, R)$ to be a choice of $R' \in \mathcal{W}$ with $R' \simeq R$ and $R' \subset 2\lambda R$. (The definition is vacuous for $\lambda < 2^k$.) It is clear that we need only consider $\simeq \lambda^2$ choices of these functions $\sigma(\lambda, \cdot): \mathcal{U}_k \to \mathcal{W}$. There is an L^1-estimate which allows one to take advantage of the spatial separation between R and $\sigma(\lambda, R)$:

$$\begin{aligned}\Big\| \sum_{R \in \mathcal{U}_k} \langle b, v_{\sigma(\lambda,R)} \rangle \overline{\langle b, v_R \rangle}\, v_{\sigma(\lambda,R)} \overline{v_R} \Big\|_1 &\lesssim \lambda^{-100} \sum_{R \in \mathcal{U}_k} |\langle b, v_{\sigma(\lambda,R)} \rangle \overline{\langle b, v_R \rangle}| \\ &\lesssim \lambda^{-100} \Big[\sum_{R \in \mathcal{U}_k} |\langle b, v_{\sigma(\lambda,R)} \rangle|^2 \sum_{R \in \mathcal{U}_k} |\langle b, v_R \rangle|^2 \Big]^{1/2} \\ &\leqslant K_\delta \varepsilon \lambda^{-90}.\end{aligned}$$

This estimate uses (3.3) and is a very small estimate.

To complete this case we need to provide an estimate in L^4. Here, we can be quite inefficient. By Cauchy–Schwarz and the Littlewood–Paley inequalities,

$$\begin{aligned}&\Big\| \sum_{R \in \mathcal{U}} \langle b, v_{\sigma(\lambda,R)} \rangle \overline{\langle b, v_R \rangle}\, v_{\sigma(\lambda,R)} \overline{v_R} \Big\|_4 \\ &\qquad \lesssim \Big\| \Big[\sum_{R \in \mathcal{U}} |\langle b, v_{\sigma(\lambda,R)} \rangle v_{\sigma(\lambda,R)}|^2 \Big]^{1/2} \Big\|_4 \Big\| \Big[\sum_{R \in \mathcal{U}} |\langle b, v_R \rangle \overline{v_R}|^2 \Big]^{1/2} \Big\|_4 \lesssim \lambda.\end{aligned}$$

This follows directly from the *BMO* assumption on b. Our proof is complete.

Appendix: A remark on Journé's lemma

Let U be an open set of finite measure in the plane. Let $\mathcal{R}(U)$ be all dyadic rectangles in $\mathcal{R}$ that are contained in U. For each $R \in \mathcal{R}(U)$ and open set $V \supset U$, set

$$\mu(V; R) = \sup\{\mu > 0 : \mu R \subset V\}.$$

The form of Journé's lemma we need is

LEMMA A.1. *For each $0<\delta<1$ and open set U of finite measure in the plane, there is a set $V \supset U$ for which $|V| < (1+\delta)|U|$, and for all $0<\varepsilon<1$, there is a constant $K_{\delta,\varepsilon}$ so that for any subset $\mathcal{R}' \subset \mathcal{R}(U)$ such that $R \not\subset R'$ for any two rectangles $R \neq R' \in \mathcal{R}'$, we have the inequality*

$$\sum_{R \in \mathcal{R}'} \mu(V; R)^{-\varepsilon} |R| \leqslant K_{\delta,\varepsilon} \Big| \bigcup_{R \in \mathcal{R}'} R \Big|. \tag{A.2}$$

Journé's lemma is the central tool in verifying the Carleson measure condition, and points to the central problem in two dimensions: that there can be many rectangles close to the boundary of an open set.

Among the references we could find in the literature [6], [9], the form of Journé's lemma cited and proved take the set V to be $\{M\mathbf{1}_U>\frac{1}{2}\}$, which only satisfies $|V|<K|U|$.

Proof of Lemma A.1. There are two stages of the proof, with the first stage being the specification of the set V. This must be done with some care, and in a manner that depends upon $\delta>0$. Let us illustrate the difficulty.

At first guess, one would take $V:=\{M\mathbf{1}_U>1-\delta\}$, with M being the strong maximal function. But the problem is that the strong maximal function is not bounded on $L^1(\mathbf{R}^2)$, so it can't possibly satisfy the desired inequality on its measure.

It is then tempting to define V as some variant of the one-dimensional maximal function. While this maximal function is bounded on $L^1(\mathbf{R})$, as a map into $L^{1,\infty}(\mathbf{R})$, the norm is known to exceed one.

The dyadic maximal function, however, maps L^1 into $L^{1,\infty}$ with norm one. This well-known fact we shall utilize in a slightly more general form. Define a *grid* to be a collection $\mathcal{I}$ of intervals in the real line for which for all $I,I'\in\mathcal{I}$, $I\cap I'\in\{\varnothing,I,I'\}$. For a collection of intervals $\mathcal{I}$, not necessarily a grid, set

$$M^{\mathcal{I}}f(x):=\sup_{I\in\mathcal{I}}\mathbf{1}_I(x)|I|^{-1}\int_I f(y)\,dy.$$

Then, for any grid $\mathcal{I}$, $M^{\mathcal{I}}$ maps $L^1(\mathbf{R})$ into $L^{1,\infty}(\mathbf{R})$ with norm one. This is in particular true for the dyadic grid $\mathcal{D}$.

Now, let us take $0<\delta<1$, and in particular take $\delta=(2^d+1)^{-1}$ for integer d. We define shifted dyadic grids, modifying an observation due to M. Christ. For integers $0\leqslant b<d$, and $\alpha\in\{\pm(2^d+1)^{-1}\}$, let

$$\mathcal{D}_{d,b,\alpha}:=\{2^{kd+b}((0,1)+j+(-1)^k\alpha):k\in\mathbf{Z},\,j\in\mathbf{Z}\}.$$

One checks that this is a grid. Indeed, it suffices to assume $\alpha=(2^d+1)^{-1}$, and that $b=0$. Checking the grid structure can be done by induction. And it suffices to check that the intervals in $\mathcal{D}_{d,0,\alpha}$ of length one are a union of intervals in $\mathcal{D}_{d,0,\alpha}$ of length 2^{-d}. One need only check this for the interval $(0,1)+\alpha$. But certainly

$$\begin{aligned}(0,1)+\frac{1}{2^d+1}&=\bigcup_{j=1}^{2^d}(0,2^{-d})+j2^{-d}-\frac{1}{2^d(2^d+1)}\\&=\bigcup_{j=1}^{2^d}\left(\frac{1}{2^d(2^d+1)},\frac{1}{2^d+1}\right)+\frac{j}{2^d}.\end{aligned}$$

What is more important concerns the collections $\mathcal{D}_d := \bigcup_\alpha \bigcup_{b=0}^{d-1} \mathcal{D}_{d,b,\alpha}$. For each dyadic interval $I \in \mathcal{D}$, $I \pm \delta|I| \in \mathcal{D}_d$. (The problem we are avoiding here is that the dyadic grid distinguishes dyadic rational points. At the point 0, for instance, observe that for all integers k, $(1+\delta)(0,1) \not\subset (0,2^k)$, regardless of how big k is.) Moreover, the maximal function $M^{\mathcal{D}_d}$ maps L^1 into $L^{1,\infty}$ with norm at most $2d \simeq \log \delta$.

We may define V. For a collection of intervals $\mathcal{I}$ and $j=1,2$, set $M_j^{\mathcal{I}}$ to be the maximal function associated to $\mathcal{I}$, computed in the coordinate j. Initially, we use only the dyadic grids, setting

$$V_0 = \bigcup_{i \neq j} \{M_i^{\mathcal{D}} \mathbf{1}\{M_j \mathbf{1}_U > 1-\delta\} > 1-\delta\}.$$

It is clear that $|V_0| < (1+K\delta)|U|$. Invoking the collections $\mathcal{D}_d$, set

$$V = \bigcup_{i \neq j} \{M_i^{\mathcal{D}_d} \mathbf{1}\{M_j^{\mathcal{D}_d} \mathbf{1}_{V_0} > 1-\delta\} > 1-\delta\}.$$

Then $|V| < (1+K\delta \log \delta^{-1})|U|$, and we will work with this choice of V.

The additional important property that V has can be formulated this way. For all dyadic rectangles $R = R_1 \times R_2 \subset V_0$, the four rectangles

$$(R_1 \pm \delta |R_1|) \times (R_2 \pm \delta |R_2|) \subset V. \tag{A.3}$$

This follows immediately from the construction of the shifted dyadic grids. The first stage of the proof is complete.

In the second stage, we verify (A.2). A typical proof of Journé's lemma shows that the rectangles in $\mathcal{R}'$ have logarithmic overlap, measured in terms of $\log \mu(V;U)$. We adopt that method of proof. Fix a subset $\mathcal{R}' \subset \mathcal{R}(U)$ satisfying the incomparability condition of the lemma, and fix $\mu \geqslant 1$. Set $\mathcal{S}$ to be those rectangles in $\mathcal{R}'$ with $\mu \leqslant \mu(R) \leqslant 2\mu$. It suffices to show that

$$\sum_{R \in \mathcal{S}} |R| \lesssim (1+\log \mu)^2 \Big| \bigcup_{R \in \mathcal{S}} R \Big|.$$

For then this estimate is summed over $\mu \in \{2^k : k \in \mathbf{Z}\}$.

In showing this estimate, we can further assume for all $R, R' \in \mathcal{S}$, writing $R = R_1 \times R_2$ and likewise for R', that if for $j=1,2$, $|R_j| > |R'_j|$ then $|R_j| > 16\mu\delta^{-1}|R'_j|$. This is done by restricting $\log_2 |R_j|$ to be in an arithmetic progression of difference $\simeq \log \mu\delta^{-1}$. This necessitates the division of all rectangles into $\lesssim (1+\log \mu\delta^{-1})^2$ subclasses, and so we prove the bound above without the logarithmic term.

We define a "bad" class of rectangles $\mathcal{B} = \mathcal{B}(\mathcal{S})$ as follows. For $j=1,2$, let $\mathcal{B}_j(\mathcal{S})$ be those rectangles R for which there are rectangles

$$R^1, R^2, \ldots, R^K \in \mathcal{S} - \{R\},$$

so that for each $1 \leqslant k \leqslant K$, $|R_j^k| > |R_j|$ and

$$\Big|R \cap \bigcup_{k=1}^{K} R^k\Big| > \big(1-\tfrac{1}{10}\delta\big)\,|R|.$$

Thus $R \in \mathcal{B}_j$ if it is nearly completely covered by dyadic rectangles in the jth direction of the plane. Set $\mathcal{B}(\mathcal{S}) = \mathcal{B}_1(\mathcal{S}) \cup \mathcal{B}_2(\mathcal{S})$. It follows that if $R \notin \mathcal{B}(\mathcal{S})$, it is not covered in both the vertical and horizontal directions, hence

$$\Big|R \cap \bigcap_{R' \in \mathcal{S}-\{R\}} (R')^c\Big| \geqslant \frac{\delta^2}{100|R|}.$$

And so

$$\sum_{R \in \mathcal{S}-\mathcal{B}(\mathcal{S})} |R| \leqslant 100\delta^{-2}\Big|\bigcup_{R \in \mathcal{S}} R\Big|.$$

Thus, it remains to consider the set of rectangles $\mathcal{B}_1(\mathcal{S})$ and $\mathcal{B}_2(\mathcal{S})$. Observe that for any collection $\mathcal{S}'$, $\mathcal{B}_j(\mathcal{S}') \subset \mathcal{S}'$, as follows immediately from the definition. Hence $\mathcal{B}_1(\mathcal{B}_2(\mathcal{B}_1(\mathcal{S}))) \subset \mathcal{B}_1(\mathcal{B}_1(\mathcal{S}))$. And we argue that this last set is empty. As our definition of V and $\mu(V;R)$ is symmetric with respect to the coordinate axes, this is enough to finish the proof.

We argue that $\mathcal{B}_1(\mathcal{B}_1(\mathcal{S}))$ is empty by contradiction. Assume that R is in this collection. Consider those rectangles R' in $\mathcal{B}_1(\mathcal{S})$ for which (i) $|R_1'| > |R_1|$ and (ii) $R' \cap R \neq \varnothing$. Then

$$\Big|R \cap \bigcup_{R' \in \mathcal{B}_1(\mathcal{S})} R'\Big| \geqslant \big(1-\tfrac{1}{10}\delta\big)\,|R|.$$

Fix one of these rectangles R' with $|R_1'|$ being minimal. We then claim that $8\mu R' \subset V$, which contradicts the assumption that $\mu(V;R')$ is no more than 2μ.

Indeed, all the rectangles in $\mathcal{B}_1(\mathcal{S})$ are themselves covered by dyadic rectangles in the first coordinate axis. We see that the the set $\{M_2^{\mathcal{D}} \mathbf{1}_U > 1-\delta\}$ contains the dyadic rectangle $R_1'' \times R_2$, in which R_2 is the second coordinate interval for the rectangle R and R_1'' is the dyadic interval that contains R_1' and has measure $8\mu\delta^{-1}|R_1'| \leqslant |R_1''| < 16\mu\delta^{-1}|R_1'|$.

That is, $R_1'' \times R_2$ is contained in V_0. And the dimensions of this rectangle are very much bigger than those of R. Applying (A.3), the rectangles $(R_1'' \pm |R_1''|) \times R_2 \pm \delta|R_2|$ are contained in V. And since $8\mu R'$ is contained in one of these last four rectangles, we have contradicted the assumption that $\mu(V;R') < 2\mu$. □

References

[1] CHANG, S.-Y. A. & FEFFERMAN, R., Some recent developments in Fourier analysis and H^p-theory on product domains. *Bull. Amer. Math. Soc. (N.S.)*, 12 (1985), 1–43.

[2] COIFMAN, R. R., ROCHBERG, R. & WEISS, G., Factorization theorems for Hardy spaces in several variables. *Ann. of Math. (2)*, 103 (1976), 611–635.

[3] FEFFERMAN, R., Bounded mean oscillation on the polydisk. *Ann. of Math. (2)*, 110 (1979), 395–406.

[4] FERGUSON, S. H. & SADOSKY, C., Characterizations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures. *J. Anal. Math.*, 81 (2000), 239–267.

[5] GUNDY, R. F. & STEIN, E. M., H^p theory for the poly-disc. *Proc. Nat. Acad. Sci. U.S.A.*, 76 (1979), 1026–1029.

[6] JOURNÉ, J.-L., A covering lemma for product spaces. *Proc. Amer. Math. Soc.*, 96 (1986), 593–598.

[7] MEYER, Y., *Wavelets and Operators.* Translated from the 1990 French original by D. H. Salinger. Cambridge Stud. Adv. Math., 37. Cambridge Univ. Press, Cambridge, 1992.

[8] NEHARI, Z., On bounded bilinear forms. *Ann. of Math. (2)*, 65 (1957), 153–162.

[9] PIPHER, J., Journé's covering lemma and its extension to higher dimensions. *Duke Math. J.*, 53 (1986), 683–690.

[10] STEIN, E. M. & WEISS, G., *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Math. Ser., 32. Princeton Univ. Press, Princeton, NJ, 1971.

SARAH H. FERGUSON
Department of Mathematics
Wayne State University
Detroit, MI 48202
U.S.A.
sarah@math.wayne.edu

MICHAEL T. LACEY
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
U.S.A.
lacey@math.gatech.edu

Received April 19, 2001