



Variational estimates for paraproducts

Yen Do, Camil Muscalu and Christoph Thiele

Abstract. We generalize a family of variation norm estimates (Theorem 1.1) of Lépingle with endpoint estimates of Bourgain and Pisier–Xu to a family of variational estimates for paraproducts (Theorem 1.2), both in the discrete and the continuous setting. This expands on work of Friz and Victoir, our focus being on the continuous case and an expanded range of variation exponents.

1. Introduction and main theorem

In this paper, a band limited function on \mathbb{R}^n , $n \geq 1$, with band width N has Fourier transform supported in $\{2^{-N} < |\xi| < 2^N\}$. A collection of functions $(f_i)_{i \in \mathbb{Z}}$ such that the dilates $f_i(2^{-i}x)$ are band limited with uniform band width is called a continuous Littlewood–Paley family. The collection is called a discrete Littlewood–Paley family if f_i is spanned by the Haar functions associated to dyadic cubes of side-length 2^{1-i} . It is then essentially a martingale. The following theorem is known:

Theorem 1.1. *Given $1 < p < \infty$, $1 \leq r \leq 2$, $r \leq t < \infty$, and $N > 0$, $n \geq 1$, there is a constant C such that, for any collection $(N_k)_{k \in \mathbb{N}}$ of measurable functions on \mathbb{R}^n and any Littlewood–Paley data f_i on \mathbb{R}^n (continuous with bandwidth N or discrete), the following holds:*

If $r < 2$ or $t > r$, then

$$(1.1) \quad \left\| \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} f_i \right|^t \right)^{1/t} \right\|_p \leq C \|(f_i)\|_{p,r} .$$

If $r = 2$ and $t = r$, then for every $\lambda > 0$

$$(1.2) \quad \left\| \lambda \left(\# \left\{ k : \left| \sum_{N_{k-1} < i \leq N_k} f_i \right| > \lambda \right\} \right)^{1/t} \right\|_p \leq C \|(f_i)\|_{p,r} .$$

Mathematics Subject Classification (2010): 42B20.

Keywords: Littlewood–Paley theory, paraproduct, variation norm.

Here we have set

$$(1.3) \quad \|(f_i)\|_{p,r} := \left\| \left(\sum_i |f_i|^r \right)^{1/r} \right\|_p.$$

In the discrete case, these inequalities are special cases of known inequalities for martingales: (1.1) is Lépingle’s inequality [6] if $t > r$ and Pisier and Xu’s inequality [12] if $t = r$. Inequality (1.2) was used by Bourgain [1] to give a proof of Lépingle’s inequality. Inequality (1.2) for $r < 2$ or $t > r$ follows from (1.1) by Chebysheff’s inequality.

For $r = 2$, the continuous case of Theorem 1.1 has been observed by several authors (see [1], [4] and [5]) in the following setting. Define for some fixed function f

$$(1.4) \quad f_i := \phi_i * f, \quad \phi_i(x) := 2^{ni} \phi(2^i x),$$

where ϕ is some band limited test function. Then (1.3) for $r = 2$ is the norm of a Littlewood–Paley square function and dominated by $\|f\|_p$, and the domination is an equivalence if ϕ_i is chosen appropriately. We present a proof of the continuous case $r < 2$ of Theorem 1.1 in Section 2 as model for the unfortunately somewhat technical Section 6.

We prove the following bilinear variant of Theorem 1.1:

Theorem 1.2. *Given $1 < p, q < \infty$, $1 \leq r, s \leq 2$, $1/t \leq 1/r + 1/s$, $N > 0$, and $n \geq 1$, there is a constant C such that, for any collection (N_k) of measurable functions on \mathbb{R}^n and any Littlewood–Paley data f_i and g_j (each continuous with bandwidth N or discrete), the following hold:*

If $\max(r, s) < 2$ or $t > rs/(r + s)$, then

$$(1.5) \quad \left\| \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right|^t \right)^{1/t} \right\|_{\frac{pq}{p+q}} \leq C \|(f_i)\|_{p,r} \|(g_j)\|_{q,s}.$$

If $\max(r, s) = 2$ and $t = rs/(r + s)$, then

$$(1.6) \quad \left\| \lambda \left(\#\left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right| > \lambda \right\} \right)^{1/t} \right\|_{\frac{pq}{p+q}} \leq C \|(f_i)\|_{p,r} \|(g_j)\|_{q,s}.$$

For $r = s = 2$ and $p = q$, $f_i = g_i$, the discrete case of (1.5) is a special case of a martingale inequality of Friz and Victoir [2]. All other instances of this theorem appear to be new. Inequality (1.6) for $\max(r, s) < 2$ or $t > \frac{rs}{r+s}$ follows from (1.5) by Chebysheff. Note that Theorem 1.2 has a continuous, a discrete, and a mixed continuous and discrete case, though the last is maybe less natural.

Theorems 1.1 and 1.2 have alternative formulations, where f_i and g_j are not required to be band limited. Instead, in the continuous case we replace f_i and g_j on the lefthand side by $\phi_i * f_i$ and $\phi_j * g_j$ with ϕ_i as in (1.4), while in the discrete case we use Haar projections on the lefthand side. In particular, we have

Theorem 1.3. *Let ϕ_i be as in (1.4). Given $1 < p, q < \infty$, $1 \leq r, s \leq 2$, $1/t \leq 1/r + 1/s$, and $n \geq 1$, there is a constant C such that, for any collection (N_k) of measurable functions on \mathbb{R}^n and any sequences of functions $f = (f_i)$, $g = (g_j)$ with the normalization*

$$(1.7) \quad \|(f_i)\|_{p,r} = \|(g_j)\|_{q,s} = 1,$$

the following hold:

If $\max(r, s) < 2$ or $t > rs/(r + s)$, then

$$(1.8) \quad \left\| \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right|^t \right)^{1/t} \right\|_{\frac{pq}{p+q}} \leq C.$$

If $\max(r, s) = 2$ and $t = rs/(r + s)$, then

$$(1.9) \quad \left\| \lambda \left(\#\left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right| > \lambda \right\} \right)^{1/t} \right\|_{\frac{pq}{p+q}} \leq C.$$

In the rest of this paper, including in the proofs of the main Theorems 1.2 and 1.3, we will restrict attention to the one-dimensional case, $n = 1$. Generalization to $n \geq 1$ is not difficult and requires merely notational changes.

Theorem 1.2 implies Theorem 1.3 by an application of the former to the Littlewood–Paley family $(\phi_i * f_i)$ and by the Fefferman–Stein inequality [13] for the Hardy–Littlewood maximal operator M :

$$\|(\phi_i * f_i)\|_{p,r} \leq C \|(Mf_i)\|_{p,r} \leq C \|(f_i)\|_{p,r}.$$

Conversely, the continuous case of Theorem 1.2 follows by specializing Theorem 1.3 to a continuous Littlewood–Paley family (f_i) and choosing ϕ with large band width such that $\phi_i * f_i = f_i$. Although we will use both formulations, Theorem 1.2 has simpler notation while Theorem 1.3 works better with truncations of f and g and interpolation.

Another alternative formulation arises from setting

$$f_i := \sum_{|I|=2^{1-i}} a_I \phi_I,$$

where a_I are coefficients and for each dyadic interval I the function ϕ_I is an L^∞ normalized band limited bump function adapted to I in the sense of [13]. In particular it satisfies, for $\epsilon > 0$,

$$|\phi_I(x)| \leq C (1 + (x - c(I))/|I|)^{-(1+\epsilon)},$$

which allows estimating (1.3) by

$$C \left\| \left(M \sum_{|I|=2^{1-i}} a_I 1_I \right) \right\|_{p,r} \leq C \left\| \left(\sum_{x \in I} |a_I|^r \right)^{1/r} \right\|_{L^p(x)}.$$

The L^∞ normalization of ϕ_I makes a_I have the same normalization as the values of the corresponding f_i .

The restriction $i < j$ in Theorem 1.2 is analogous to the paraproduct [8] of two functions f and g :

$$P(f, g) = \sum_{i < j} (\phi_i * f)(\phi_j * g).$$

The endpoint of Theorem 1.2 for $t = \infty$ is a classical maximal paraproduct inequality, for example in the convolutional case we have

Proposition 1.4. *For $1 < p, q < \infty$,*

$$(1.10) \quad \left\| \sup_{N_0, N_1} \left| \sum_{N_0 < i < j \leq N_1} (\phi_i * f)(\phi_j * g) \right| \right\|_{pq/(p+q)} \leq C_{p,q} \|f\|_p \|g\|_q.$$

Inequality (1.1) has a trivial endpoint for $r = t = 1$ by the triangle inequality. Similarly, the endpoint of (1.5) for $r = 1$ or $s = 1$ is easy and elaborated in detail in Proposition 4.1. This endpoint is proved by reduction to Theorem 1.1. and used as interpolation endpoint to prove part of Theorem 1.2. Note that while classical variation norms as in (1.1) are somewhat pointless for $t \leq 1$, because then the choice of consecutive points $N_{k+1} = N_k + 1$ is extremal, the bilinear variational expression in (1.5) is meaningful for $t > 1/2$.

We make two simple observations about Theorem 1.2. The variant of (1.5) without the paraproduct restriction $i < j$ follows from Theorem 1.1 by Hölder’s inequality. The high exponent case $t > 2$ of Theorem 1.2 has a simple proof that we sketch in the discrete case. Let Δ_m denote the projection onto the Haar functions associated to dyadic intervals of length 2^{1-m} . Then

$$\begin{aligned} \Delta_m \left(\sum_{i < j} f_i g_j \right) &= \sum_{i < m} f_i g_m, \\ \sum_{N_{k-1} < i < j \leq N_k} f_i g_j &= \sum_{N_{k-1} < m \leq N_k} \Delta_m \left[\sum_{i < j} f_i g_j \right] - \sum_{i \leq N_{k-1}} f_i \sum_{N_{k-1} < j \leq N_k} g_j. \end{aligned}$$

The first term on the right hand side can be estimated by (1.1) and (1.10), while the second term can be estimated by the Hardy–Littlewood maximal theorem and (1.1) applied to the terms involving f and g respectively.

We became interested in variational estimates for paraproducts while studying L^p estimates for a variational expression of the form

$$(1.11) \quad \left(\sum_k \left| \int_{N_{k-1} < \xi < \eta < N_k} \widehat{f}(\xi) \widehat{g}(\eta) e^{i\pi x(\xi + \eta)} d\xi d\eta \right|^r \right)^{1/r}.$$

This can be viewed as a bilinear analogue of the variation norm Carleson operator studied in [11] or as a variational variant of the bi-Carleson operator introduced in [10]; compare also with the bi-est operator introduced in [9]. It requires time-frequency analysis to be understood, and Theorem 1.2 provides the related lacunary estimates.

The general type of bilinear variational estimate as in Theorem 1.2 and (1.11) is motivated by Terry Lyons’ theory of ODE with rough driving terms, [7]. This theory bootstraps Theorem 1.2 to the diagonal case $p_1 = \dots = p_m$ of the following multilinear generalization of Theorem 1.2

Proposition 1.5. *For $2/M < r < \infty$, $1 < p_m < \infty$, and $f_m \in L^{p_m}(\mathbb{R}^n)$, for $m = 1, \dots, M$, we have, with $1/p = \sum_m 1/p_m$,*

$$\left\| \sup_{N_k} \left(\sum_k \left| \sum_{N_{k-1} < i_1 < \dots < i_M \leq N_k} \prod_m (\phi_{i_m} * f_m) \right|^r \right)^{1/r} \right\|_p \leq C \prod_m \|f_m\|_{p_m}.$$

Some non-diagonal estimates are likely to follow from a multilinear generalization of arguments as in this paper, though the exact elaboration of the details is beyond the scope of this paper.

Further consequences in the case of martingales are discussed in [2]. The authors would like to thank Terry Lyons for pointing out reference [2] after circulation of an earlier draft of this paper, and to thank the anonymous referee for several suggestions to improve the exposition.

2. Proof of Pisier and Xu’s inequality, continuous case

Recall that Pisier and Xu’s inequality is the case $t = r < 2$ of inequality (1.1). In the diagonal case, $p = t$, inequalities (1.1) and (1.2) can be written as strong and weak type L^p estimates (see [13] for this terminology) for mappings from functions on the measure space $\mathbb{R} \times \mathbb{Z}$ to functions on the measure space $\mathbb{R} \times \mathbb{N}$:

$$(2.1) \quad \left\| \sum_{N_{k-1}(x) < i \leq N_k(x)} \phi_i * f_i(x) \right\|_{L^p(x,k)} \leq C \|f_i(x)\|_{L^p(x,i)},$$

$$(2.2) \quad \left| \left\{ (x, k) : \left| \sum_{N_{k-1}(x) < i \leq N_k(x)} \phi_i * f_i(x) \right| > \lambda \right\} \right| \leq C \lambda^{-2} \|f_i(x)\|_{L^2(x,i)}^2.$$

Here we use a setup analogous to that in Theorem 1.3. For a proof of (2.2) we refer to [5]. The endpoint of (2.1) for $p = 1$ follows from the triangle inequality. Hence (2.1) follows by Marcinkiewicz interpolation from (2.2), which completes the discussion of the diagonal case.

Starting from this diagonal case, we shall lower and raise the exponent p by Calderón–Zygmund decomposition and sharp function techniques respectively. For the sub-diagonal case, $p < t$, fix r and employ Marcinkiewicz interpolation for l^r -valued functions. The interpolation endpoints are the diagonal case $p = r$ and the weak type bound

$$(2.3) \quad \left| \left\{ x : \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i(x) \right|^r \right)^{1/r} > \lambda \right\} \right| \leq \frac{C}{\lambda} \left\| \left(\sum_i |f_i|^r \right)^{1/r} \right\|_1$$

at L^1 . The latter follows via a Calderón–Zygmund decomposition of the vector function $f = (f_i)$ at level λ . Let

$$E = \left\{ x : M \left(\left(\sum_i |f_i|^r \right)^{1/r} \right) > \lambda \right\}$$

and let \mathbf{I} be the collection of maximal dyadic intervals contained in the set E . Split f as

$$f = g + b = g + \sum_I b_I,$$

where on each interval I the function g is constant equal to the average of f on I , and each b_I is $f - g$ restricted to I . It suffices to prove (2.3) separately with f replaced by g and b on the lefthand side. However,

$$\|\phi_i * b_{i,I}\|_{L^1((3I)^c)} \leq C \lambda |I| \min(2^i |I|, (2^i |I|)^{-\epsilon})$$

by smoothness ($|I| < 2^{-i}$) and decay ($|I| > 2^{-i}$) estimates for ϕ_i . Hence, by embedding l^1 into l^r ,

$$\begin{aligned} & \left\| \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * \sum_I b_{i,I} \right|^r \right)^{1/r} \right\|_{L^1(\cup 3I)^c} \\ & \leq C \sum_I \sum_i \|\phi_i * b_{i,I}\|_{L^1((3I)^c)} \leq C \lambda |E| \leq C \|f\|_{1,r}. \end{aligned}$$

Then (2.3) for b follows from Chebysheff's inequality. On the other hand,

$$\|g\|_{r,r}^r \leq C \lambda^{r-1} \|g\|_{1,r} \leq C \lambda^{r-1} \|f\|_{1,r}$$

and (2.3) for g follows from the known diagonal estimate and Chebysheff's inequality. This completes the proof of the sub-diagonal case.

For the super-diagonal case, $p > r$, consider the sharp function

$$\begin{aligned} (2.4) \quad (Tf)^\sharp(x) &= \sup_{x \in I} \inf_c \frac{1}{|I|} \int_I |Tf(y) - c| dy, \\ Tf(x) &= \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i(x) \right|^r \right)^{1/r}. \end{aligned}$$

Here the supremum is over all sequences (N_k) . The L^p norms of Tf and $(Tf)^\sharp$ are comparable, [13], hence it suffices to estimate pointwise

$$(2.5) \quad (Tf)^\sharp(x) \leq CM_r \left(\left(\sum_i |f_i|^r \right)^{1/r} \right)(x),$$

where

$$M_r h(x) := \left(\sup_{x \in I} \frac{1}{|I|} \int_I |h(y)|^r dy \right)^{1/r}.$$

Fix an interval I and let \tilde{f}_i be a constant function on \mathbb{R} whose value equals the average of $\phi_i * f_i$ on I if $2^{-i} > |I|$ and equals zero if $2^{-i} \leq |I|$. Instead of taking the infimum in c , we evaluate the definition of the sharp function $(Tf)^\sharp$ with

$$c = \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \tilde{f}_i \right|^r \right)^{1/r}.$$

By the general norm inequality $\| \|a\| - \|b\| \| \leq \|a - b\|$ we have

$$\begin{aligned} & \left| \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i(y) \right|^r \right)^{1/r} - c \right| \\ & \leq \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i(y) - \tilde{f}_i(y) \right|^r \right)^{1/r}. \end{aligned}$$

We write

$$\phi_i * f_i - \tilde{f}_i = h_i^{[1]} + h_i^{[2]} + h_i^{[3]}.$$

where

$$h_i^{[2]} = h_i^{[3]} = 0$$

for $2^{-i} > |I|$, while for $2^{-i} \leq |I|$ we have $h_i^{[1]} = 0$ and

$$h_i^{[2]} = \phi_i * (f_i 1_{3I}) \quad , \quad h_i^{[3]} = \phi_i * (f_i 1_{(3I)^c}).$$

Estimating the three summands separately, we have for $h^{[1]}$, using the embedding of l^1 into l^r , and the smoothness of $\phi_i * f_i$,

$$\begin{aligned} & \frac{1}{|I|} \int_I \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} h_i^{[1]}(y) \right|^r \right)^{1/r} dy \\ & \leq \frac{1}{|I|} \int_I \sum_{2^{-i} > |I|} |h_i^{[1]}(y)| dy \leq C \sum_{2^{-i} > |I|} (2^i |I|) M f_i(x). \end{aligned}$$

This is dominated by the right hand side of (2.5). By the diagonal estimate, we have

$$\frac{1}{|I|} \int_I \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} h_i^{[2]}(y) \right|^r \right)^{1/r} dy \leq C |I|^{-1/r} \left\| \left(\sum_i |f_i 1_{3I}|^r \right)^{1/r} \right\|_r,$$

which is again bounded by the right hand side of (2.5). Finally, we have, by the decay estimates for ϕ_i and the embedding of l^1 into l^r ,

$$\begin{aligned} & \frac{1}{|I|} \int_I \sup_{(N_k)} \left(\sum_{N_{k-1} < i \leq N_k} |h_i^{[3]}(y)|^r \right)^{1/r} dy \\ & \leq \frac{1}{|I|} \int_I \sum_{2^{-i} \leq |I|} |h_i^{[3]}(y)| dy \leq C \sum_{2^{-i} \leq |I|} (2^i |I|)^{-\epsilon} M f_i(x), \end{aligned}$$

which is again bounded by the right hand side of (2.5). This concludes the proof of the super-diagonal case.

3. Reduction of Theorem 1.2 to the case $t = rs/(r + s)$

If $\max(r, s) < 2$ we simply observe that (1.5) is monotone in t to reduce to the case $t = rs/(r + s)$. If $\max(r, s) = 2$ we need to show that the relatively weaker estimate (1.6) for some t implies the relatively stronger estimate (1.5) for $t_0 > t$. This will be a bilinear variant of the argument used by Bourgain [1].

We apply Marcinkiewicz interpolation in the following form:

Proposition 3.1. *Let f, g denote measurable functions on measure spaces X, Y with values in Banach spaces F and G respectively. Suppose we have a bi-quasi-sublinear operation $(f, g) \rightarrow T(f, g)$ producing some measurable function $T(f, g)$ on a measure space Z . Here we mean quasi-sublinear in each argument, that is,*

$$T(f + \tilde{f}, g) \leq C(T(f, g) + T(\tilde{f}, g)),$$

$$T(f, g + \tilde{g}) \leq C(T(f, g) + T(f, \tilde{g})).$$

Assume we have for all $\lambda > 0$ the weak type estimates

$$\lambda \{z : |T(f, g)(z)| > \lambda\}^{1/p+1/q} \leq C \|f\|_p \|g\|_q$$

for all corners (p, q) of an axis parallel rectangle whose interior contains the point (p_0, q_0) , then we have the strong type estimate

$$\|T(f, g)\|_{p_0q_0/(p_0+q_0)} \leq \tilde{C} \|f\|_{p_0} \|g\|_{q_0}.$$

The proof of this proposition follows the standard Marcinkiewicz argument, in the multilinear setting described in Janson [3]. One splits both functions f and g according to small and large values (according to some level λ that is later integrated on) and estimates the four resulting terms of $T(f, g)$ using the assumed estimates. While Janson requires sublinearity in each argument, the adaption to quasi-sublinearity is not difficult since we split $T(f, g)$ only into four terms for each level λ .

Fix p, q, r, s, t as in the theorem and let $t_0 > t$. We shall work in the setting of Theorem 1.3 and deduce (1.8) for t_0 from (1.9) for t . By Proposition (3.1) with $F = l^r(\mathbb{Z})$ and $G = l^s(\mathbb{Z})$ it suffices to show that for any $\lambda > 0$ and for any $f = (f_i), g = (g_j)$ with normalization (1.7) we have

$$(3.1) \quad \left| \left\{ x : \left(\sum_k F(k)^{t_0} \right)^{1/t_0} > \lambda \right\} \right| \leq C \lambda^{-\frac{pq}{p+q}},$$

where

$$F(k) = \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right|.$$

Indeed, since the problem is invariant under dilation of the x axis by powers of 2, it suffices to prove (3.1) for $\lambda = 1$. Let

$$E = \left\{ x : \sup_k F(k) > 1 \right\}.$$

Then, by (1.9) for the exponent t and with $u = pq/(p + q)$, we have

$$|E| \leq \int |\{k : F(k) > 1\}|^{\frac{u}{t}} \leq C.$$

Considering level sets of $F(k)$, for $x \notin E$ we have

$$\sum_k F(k)^{t_0} \leq C \sum_{n \leq 0} 2^{nt_0} |\{k : F(k) > 2^n\}|.$$

By Hölder together with a geometric sum if $u/t > 1$ and by monotonicity of $l^v(\mathbb{Z})$ in v if $u/t \leq 1$ we obtain

$$\left(\sum_k F(k)^{t_0}\right)^{u/t} \leq C \sum_{n \leq 0} 2^{n(1-\epsilon)ut_0/t} |\{k : F(k) > 2^n\}|^{u/t}$$

for some small $\epsilon > 0$. Using Chebysheff we have

$$\begin{aligned} \left|\{x \notin E : \sum_k F(k)^{t_0} > 1\}\right| &\leq \int_{\mathbb{R} \setminus E} \left(\sum_k F(k)^{t_0}\right)^{u/t} \\ &\leq C \sum_{n \leq 0} 2^{n(1-\epsilon)ut_0/t} \int |\{k : F(k) > 2^n\}|^{u/t} \leq C \sum_{n \leq 0} 2^{n(1-\epsilon)ut_0/t} 2^{-nu}. \end{aligned}$$

In the last inequality we have applied (1.9). The righthand side is summable for sufficiently small ϵ , since $t_0 > t$. This proves (3.1) and completes the reduction to the case $t = rs/(r + s)$ for Theorem 1.3. Adaptations to the discrete and mixed cases are not difficult. We shall assume $t = rs/(r + s)$ throughout the rest of this paper.

4. The endpoint at $r = 1$ or $s = 1$

If $r = s = 1$, hence $t = 1/2$, we have the following trivial observation for $0 < p, q \leq \infty$, which holds for arbitrary sequences (f_i) and (g_j) :

$$\begin{aligned} (4.1) \quad &\left\| \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right|^{1/2} \right)^2 \right\|_{pq/(p+q)} \\ &\leq \left\| \left(\sum_k \left(\sum_{N_{k-1} < i \leq N_k} |f_i| \right)^{1/2} \left(\sum_{N_{k-1} < j \leq N_k} |g_j| \right)^{1/2} \right)^2 \right\|_{pq/p+q} \\ &\leq \left\| \left(\sum_i |f_i| \right) \left(\sum_j |g_j| \right) \right\|_{pq/(p+q)} \leq \|(f_i)\|_{p,1} \|(g_j)\|_{q,1}. \end{aligned}$$

Consider just one of r, s equal to 1. By symmetry of the argument below we may assume $r = 1$. We then have the following result:

Proposition 4.1. *Let $0 < p \leq \infty$ and $1 < q < \infty$ and assume $t = s/(1 + s)$. Let N_k be measurable functions, (f_i) any sequence of measurable functions and let (g_j) be a Littlewood–Paley family.*

If $1 < s < 2$, then

$$(4.2) \quad \left\| \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right|^t \right)^{1/t} \right\|_{\frac{pq}{p+q}} \leq C \| (f_i) \|_{p,1} \| (g_j) \|_{q,s}.$$

If $s = 2$, then

$$(4.3) \quad \left\| \lambda \left(\# \left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right| > \lambda \right\} \right)^{1/t} \right\|_{\frac{pq}{p+q}} \leq C \| (f_i) \|_{p,1} \| (g_j) \|_{q,s}.$$

Inequality (4.2) follows quickly from Theorem 1.1. Note that

$$\left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right| \leq \left(\sum_{N_{k-1} < i \leq N_k} |f_i| \right) \left(\sup_{N_{k-1} < i \leq N_k} \left| \sum_{i < j \leq N_k} g_j \right| \right).$$

Then estimate the lefthand side of (4.2) with Hölder’s inequality by

$$\| (f_i) \|_{p,1} \left\| \left(\sum_k \sup_{N_{k-1} < i \leq N_k} \left| \sum_{i < j \leq N_k} g_j \right|^s \right)^{1/s} \right\|_q.$$

Introducing for each x a new sequence which consists of the sequence $N_k(x)$ interlaced with extremal choices of i where applicable, we can estimate the second factor by Theorem 1.1 and conclude (4.2).

It remains to prove (4.3). It will suffice to prove the analogue inequality with g_j replaced by $\phi_j * g_j$ on the lefthand side and g_j not necessarily bandlimited as in Theorem 1.3. We will use Proposition 3.1. Since the quantity

$$(4.4) \quad \lambda \left(\# \left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} (f_i) (\phi_j * g_j) \right| > \lambda \right\} \right)^{1/t}$$

is not quasi subadditive, we shall replace it by an equivalent quasi subadditive quantity. For $\lambda > 0$ define $\rho_\lambda(x) = \min(|x|, \lambda)$ and note that $\rho_\lambda(x - y)$ defines a metric distance between x and y . Hence the operation T_λ with

$$(4.5) \quad T_\lambda(f, g) = \lambda^{-1} \left(\sum_k \rho_\lambda \left(\sum_{N_{k-1} < i < j \leq N_k} (f_i) (\phi_j * g_j) \right)^2 \right)^{1/t}$$

is quasi subadditive. It dominates (4.4) and hence it suffices to show (here $r = 1$)

$$\| T_\lambda(f, g) \|_{\frac{pq}{p+q}} \leq C \| f \|_{p,r} \| g \|_{q,s}.$$

By Proposition 3.1, it suffices to prove

$$(4.6) \quad |\{x : T_\lambda(f, g) > \mu\}| \leq C \mu^{-\frac{pq}{p+q}} \| f \|_{p,r} \| g \|_{q,s}$$

with C independent of λ and μ . Switching back to expression (4.4) we will prove

$$(4.7) \quad \left| \left\{ \lambda \left| \left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} (f_i)(\phi_j * g_j) \right| > \lambda \right\} \right|^{1/t} > \mu \right\} \right| \leq C \mu^{-\frac{pq}{p+q}} (\|f\|_{p,r} \|g\|_{q,s})^{\frac{pq}{p+q}}.$$

We argue that this is sufficient to conclude (4.6). Consider for fixed x the level sets of the function

$$F(k) = \left| \sum_{N_{k-1} < i < j \leq N_k} (f_i)(\phi_j * g_j) \right|,$$

i.e., the sets of k for which $2^n \lambda \leq F(k) < 2^{(n+1)} \lambda$ for $n \leq 0$. Using geometric decay of $2^{n\epsilon}$ for some small $\epsilon > 0$, we obtain

$$T_\lambda(f, g) \leq C \sum_{n \leq 0} 2^{(2-\epsilon)n} \lambda |\{k : F(k) > 2^n \lambda\}|^{1/t}.$$

Then we can estimate

$$\begin{aligned} |\{x : T_\lambda(f, g) > \mu\}| &\leq \sum_{n \leq 0} \left| \left\{ x : 2^{(2-\epsilon)n} \lambda |\{k : F(k) > 2^n \lambda\}|^{1/t} > c 2^{n\epsilon} \mu \right\} \right| \\ &\leq C \sum_{n \leq 0} (2^{(2\epsilon-1)n} \mu)^{-\frac{pq}{p+q}} (\|f\|_{p,r} \|g\|_{q,s})^{\frac{pq}{p+q}}. \end{aligned}$$

Taking $\epsilon < 1/2$ we obtain (4.6).

Now we prove (4.7). By dilating the x -axis, we may assume that the righthand side of (4.7) is C . By multiplying f, g, λ , and μ by appropriate constants we may assume the normalization (1.7), and by the previous also $\mu = 1$. We write

$$\begin{aligned} \left| \left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} (f_i)(\phi_j * g_j) \right| > \lambda \right\} \right| &\leq \left| \left\{ k : \sum_{N_{k-1} < i < N_k} |f_i| > \lambda^t \right\} \right| \\ &\quad + \left| \left\{ k : \sup_{N_{k-1} < i < N_k} \left| \sum_{i < j \leq N_k} \phi_j * g_j \right| > \lambda^{t/s} \right\} \right|. \end{aligned}$$

Therefore the lefthand side of (4.7) has the upper bound

$$\begin{aligned} \left| \left\{ x : \lambda \left| \left\{ k : \sum_{N_{k-1} < i < N_k} |f_i| > \lambda^t \right\} \right|^{1/t} > c \right\} \right| \\ + \left| \left\{ x : \lambda \left| \left\{ k : \sup_{N_{k-1} < i < N_k} \left| \sum_{i < j \leq N_k} \phi_j * g_j \right| > \lambda^{t/s} \right\} \right|^{1/t} > c \right\} \right|. \end{aligned}$$

Using Chebysheff twice, the first term can be estimated by

$$\left| \left\{ x : \sum_i |f_i| > c^t \right\} \right| \leq C \|(f_i)\|_{p,1}^p.$$

The second term can be estimated by

$$C \left\| \lambda^{t/s} \left| \left\{ k : \sup_{N_{k-1} < i < N_k} \left| \sum_{i < j \leq N_k} \phi_j * g_j \right| > \lambda^{t/s} \right\} \right|^{1/s} \right\|_q^q \leq C \|g\|_{q,s}^q$$

by Theorem 1.1, applied for a sequence N_k interlaced with elements i . This proves (4.7) and completes the proof of Proposition 4.1.

5. The diagonal case of Theorem 1.2 and L^2 theory

The goal of this section is to prove Theorems 1.2 and 1.3 for one pair of exponents (p, q) and for any given triple (r, s, t) with $t = rs/(r + s)$. In Section 6, we will raise and lower (p, q) using a bilinear variant of the argument in Section 2, and thereby complete the proof of Theorem 1.2. As in the previous section, the desired estimates are technically easier under some diagonality assumptions on (p, q) . We will assume throughout this section that $p = r$ and $q = s$ and in particular $t = pq/(p + q)$, which allow us to turn (1.9) into a weak type formulation analogous to (2.2), namely the estimate

$$(5.1) \quad \int \#\left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right| > \lambda \right\} dx \leq C \lambda^{-t}$$

(the formulation is as in Theorem 1.3), with the normalization assumption (1.7).

Note that for $r = 1$ or $s = 1$ we have already proven (5.1) in the previous section for many values of (p, q) . By interpolation, it suffices to prove (5.1) for $r = s = 2$. Proposition 3.1, applied to the square with corners $(2, 2), (2, 1), (1, 2)$ and $(1, 1)$, will prove the diagonal case of (1.8), hence (5.1), in the open interior of the square. Note that we apply Proposition 3.1 with $F, G = \mathbb{C}$ and $X, Y = \mathbb{R} \times \mathbb{Z}$.

On the edges of the square we only need to prove the weak type estimate (5.1), which requires not Marcinkiewicz interpolation, rather simple interpolation by one time truncation of f or g , and application of the endpoint estimates.

The rest of this section will be concerned with proving the case $r = s = p = q = 2$ and $t = 1$ of (5.1).

5.1. Discrete case, stopping times

As the first step, we assume here that (N_k) is a stopping time, which means that if I is the dyadic interval of length $2^{-N_k(x)}$ containing x , then for all $y \in I$ we have $N_k(x) = N_k(y)$. Such an interval I is called a tree top and we can form the tree T (or stopping time region) consisting of all dyadic intervals contained in I but not contained in any smaller tree top.

For x in a tree top I of length $|I| = 2^{-N_{k-1}(x)}$ we can write

$$(5.2) \quad \sum_{N_{k-1} < i < j \leq N_k} f_i(x)g_j(x) = \sum_{i < j} \Delta_i \Pi_T f(x) \Delta_j \Pi_T g(x),$$

where $f = \sum_i f_i, g = \sum_j g_j$ and Π_T denotes the projection onto the space spanned by the Haar functions associated to dyadic intervals in the tree T . Here we have

used that $\Delta_i f(x)$ equals the evaluation of the projection of f onto a single Haar function depending on i and x and for a stopping time (N_k) this Haar function corresponds to an interval in the tree T if and only if $N_{k-1} < i \leq N_k$.

In the case of stopping times, we prove strong type (1.5) directly, which of course implies weak type (5.1). Inserting (5.2) on the lefthand side of (1.5) and interchanging the order of sum and integration we obtain

$$\sum_{T \in \mathbf{T}} \left\| \sum_{i < j} (\Delta_i \Pi_T f)(\Delta_j \Pi_T g) \right\|.$$

By the paraproduct estimate (1.10) this is bounded by a constant times

$$\sum_{T \in \mathbf{T}} \|\Pi_T f\|_2 \|\Pi_T g\|_2.$$

Applying Cauchy–Schwarz to the sum in T , and using orthogonality of the projections Π_T , this implies (1.5).

5.2. Discrete case, arbitrary (N_k)

We will prove the weak type estimate (5.1) for arbitrary (N_k) by comparing the set of λ jumps for (N_k) to the set of $\lambda/4$ jumps of an adapted stopping time (\tilde{N}_k) . Fix measurable functions N_k , discrete Littlewood–Paley families f_i and g_j , and $\lambda > 0$. For each x choose $\tilde{N}_0(x)$ sufficiently close to $-\infty$ that the maximal paraproduct restricted to $i < j < \tilde{N}_0(x)$ is pointwise bounded by $\lambda/4$, and the tree tops for \tilde{N}_0 partition the real axis. Define $\tilde{N}_k(x)$ recursively as the first time past $\tilde{N}_{k-1}(x)$ such that one of the following two conditions is satisfied:

$$\begin{aligned} & \left| \sum_{\tilde{N}_{k-1}(x) < i < j \leq \tilde{N}_k(x)} f_i(x)g_j(x) \right| \geq \lambda/4, \\ \sup_{\tilde{N}_{k-1}(x) < m < \tilde{N}_k(x)} & \left| \sum_{\tilde{N}_{k-1}(x) < i \leq m} f_i(x) \right| \left| \sum_{m < j \leq \tilde{N}_k(x)} g_j(x) \right| \geq \lambda/4. \end{aligned}$$

If no such $\tilde{N}_k(x)$ exists, we set $\tilde{N}_k(x) = \tilde{N}_{k-1}(x) + 1$. The established estimate for stopping times gives good control for the first kind of jumps. The integral over the number of the second kind of jumps is controlled by

$$\begin{aligned} & \int \#\left\{k : \sup_{N_{k-1}(x) < m < N_k(x)} \left| \sum_{N_{k-1}(x) < i \leq m} (\Delta_i f(x)) \right| \geq c\lambda^{1/2} \right\} dx \\ & + \int \#\left\{k : \sup_{N_{k-1}(x) < m < N_k(x)} \left| \sum_{m < j \leq N_k(x)} (\Delta_j g(x)) \right| \geq c\lambda^{1/2} \right\} dx, \end{aligned}$$

where each term can be controlled via (1.2) by $C\lambda$.

It remains to show that for every instance

$$\left| \sum_{N_{k-1}(x) < i < j \leq N_k(x)} f_i(x)g_j(x) \right| > \lambda$$

there is a \tilde{k} satisfying

$$(5.3) \quad N_{k-1}(x) < \tilde{N}_{\tilde{k}}(x) \leq N_k(x) .$$

Assume no such \tilde{k} exists. Then, by the choice of \tilde{N}_0 , one must have $\tilde{N}_0(x) \leq N_{k-1}(x)$. Let \tilde{k} be the largest index such that $\tilde{N}_{\tilde{k}-1}(x) \leq N_{k-1}(x)$. Then $\tilde{N}_{\tilde{k}}(x) > N_k(x)$ and we have

$$\begin{aligned} \left| \sum_{N_{k-1}(x) < i < j \leq N_k(x)} f_i(x)g_j(x) \right| &\leq \left| \sum_{\tilde{N}_{\tilde{k}-1}(x) < i < j \leq N_k(x)} f_i(x)g_j(x) \right| \\ &+ \left| \sum_{\tilde{N}_{\tilde{k}-1}(x) < i < j \leq N_{k-1}(x)} f_i(x)g_j(x) \right| \\ &+ \left| \sum_{\tilde{N}_{\tilde{k}-1}(x) < i \leq N_{k-1}(x)} f_i(x) \sum_{N_{k-1}(x) < j \leq N_k(x)} g_j(x) \right|. \end{aligned}$$

Each term on the right hand side is less than $\lambda/4$ by the choice of $\tilde{N}_{\tilde{k}}$, while the lefthand side is larger than λ . This contradiction proves (5.3).

5.3. Mixed and continuous case

We compare the discrete case with the mixed and continuous case of (5.1) using an inequality by Jones, Seeger and Wright ([5]), namely that the square function

$$Sf(x) = \left(\sum_i |\mathbb{E}_i f - \varphi_i * f|^2 \right)^{1/2}$$

is bounded in L^2 , where

$$\mathbb{E}_i = \sum_{j \leq i} \Delta_j,$$

and $\widehat{\varphi}_i$ is supported in $|\xi| \leq 2^N$ and is constant, equal to 1, on $|\xi| \leq 2^{N-1}$.

We write the telescopic sum

$$\sum_{N_{k-1} < i < j \leq N_k} (\Delta_i f)(\Delta_j g) = \sum_{N_{k-1} < j \leq N_k} (\mathbb{E}_{j-1} f - \mathbb{E}_{N_{k-1}} f)(\Delta_j g)$$

and compare with

$$\sum_{N_{k-1} < j \leq N_k} (f * \varphi_{j-1} - f * \varphi_{N_{k-1}})(\Delta_j g).$$

The difference can be estimated using another telescopic sum by

$$(5.4) \quad \left| \sum_{N_{k-1} < j \leq N_k} (\mathbb{E}_{j-1} f - f * \varphi_{j-1})(\Delta_j g) \right| + \left| (\mathbb{E}_{N_{k-1}} f - f * \varphi_{N_{k-1}})(\mathbb{E}_{N_k} g - \mathbb{E}_{N_{k-1}} g) \right| .$$

Considering the first term, we estimate

$$\begin{aligned} & \lambda \#\left\{k : \left| \sum_{N_{k-1} < i < j \leq N_k} (\mathbb{E}_{j-1} f - f * \varphi_{j-1})(\Delta_j g) \right| > \lambda \right\} \\ & \leq \sum_j |(\mathbb{E}_{j-1} f - f * \varphi_{j-1})(\Delta_j g)| \leq \left(\sum_j |\mathbb{E}_j f - f * \varphi_j|^2 \right)^{1/2} \left(\sum_j |\Delta_j g|^2 \right)^{1/2} . \end{aligned}$$

The $L^{pq/(p+q)}$ norm of the righthand side can be estimated using Hölder’s inequality and the classical square function estimate together with the square function estimate of Jones, Seeger and Wright. Turning to the second term in (5.4) we have

$$\begin{aligned} & \lambda \#\left\{k : |(\mathbb{E}_{N_{k-1}} f - f * \varphi_{N_{k-1}})(\mathbb{E}_{N_k} g - \mathbb{E}_{N_{k-1}} g)| > \lambda \right\} \\ & \leq \lambda \#\left\{k : |\mathbb{E}_{N_{k-1}} f - f * \varphi_{N_{k-1}}| > \lambda^{1/2}\right\} + \lambda \#\left\{k : |\mathbb{E}_{N_k} g - \mathbb{E}_{N_{k-1}} g| > \lambda^{1/2}\right\} . \end{aligned}$$

Then we estimate the first term by the square function of Jones–Seeger–Wright and the second term using Theorem 1.1. This proves (5.1) for the mixed convolutional and discrete case.

Using a very similar argument, one can obtain (5.1) for the mixed discrete (in f) and convolutional (in g) cases, and by combination one can obtain the pure convolutional case.

6. Completion of the proof of Theorem 1.2

For each pair (r, s) and $t = rs/(r + s)$, the previous section proves Theorem 1.2 for the specific pair $(p, q) = (r, s)$. Beginning with this pair, we will lower and raise the exponents p and q by Calderón–Zygmund decomposition and sharp function techniques respectively, as in Section 2. This will complete the proof of the theorem.

We will consider operators $S(f, g)$ in two different cases:

1. The *quasi-bi-sublinear* case,

$$S(f, g) = \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right|^t \right)^{1/t}$$

for $t < 1$. This case is needed to prove (1.8) for $\max(r, s) < 2$.

2. The *equivalent-to-quasi-bi-sublinear* case

$$S(f, g) = \lambda \left(\#\left\{k : \left| \sum_{N_{k-1} < i < j \leq N_k} f_i g_j \right| > \lambda \right\} \right)^{1/t}$$

for arbitrary λ and $t \leq 1$. This case is needed to prove (1.9) when $\max(r, s) = 2$.

The desired estimate now becomes an $L^{pq/(p+q)}$ estimate for $S(f, g)$. We will discuss the more difficult second case in more detail; the other case is similar, but easier.

In this section we focus on the continuous case. The mixed or discrete case follows from essentially the same arguments, with possible simplifications at some places. We shall make no further comment on the mixed and discrete case.

Lowering exponents

Fix (r, s) and first consider case (2). We work in the setting of Theorem 1.3. Assuming that (1.9) holds for specific exponents (p_0, q_0) , we prove (1.9) for all p, q with $1 < p < p_0, 1 < q < q_0$. More precisely, passing from $S(f, g)$ to the quasi bi-sublinear operation

$$T_\lambda(f, g) = \lambda^{-1} \left(\sum_k \rho_\lambda \left(\sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right)^{2t} \right)^{1/t}$$

similarly as in Section 4 it suffices to prove

$$\|T_\lambda(f, g)\|_{\frac{pq}{p+q}} \leq C \|f\|_{p,r} \|g\|_{q,s} .$$

By Proposition 3.1 it suffices to prove for all $1 \leq p \leq p_0, 1 \leq q \leq q_0$ and all $\mu > 0$ the weak type bound

$$|\{x : T_\lambda(f, g) \geq \mu\}| \leq C \mu^{-pq/(p+q)} \|f\|_{p,r} \|g\|_{q,s} .$$

As in Section 4 it suffices to prove

$$\left| \{ \lambda \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right| > \lambda \} \right|^{1/t} > \mu \} \leq C \mu^{-pq/(p+q)} \|f\|_{p,r} \|g\|_{q,s} .$$

By the symmetry of the argument, it suffices to fix $q = q_0$ and vary p . By simple interpolation it suffices to consider the extremal cases. The endpoint $p = p_0$ is true by assumption, so we may assume $p = 1$. As in Section 4 we may assume (1.7) and $\mu = 1$.

Split $f = (f_i)$ into a Calderón–Zygmund decomposition at level 1 as in Section 2. The good function is in L^{p_0} and we can apply the known estimate and Chebysheff. The bad function satisfies, outside the expanded bad set, the estimate

$$(6.1) \quad \left\| \sum_i \phi_i * b_i \right\|_{L^1((\cup 3I)^c)} \leq C \|f\|_{1,r} .$$

Applying Proposition 4.1 to the sequences $((\phi_i * b_i)1_{(\cup 3I)^c})$ and (g_j) we get the desired estimate for the bad function. This completes the proof of the weak type estimate for $p = 1$.

In case (1) we proceed similarly, except we work directly with the quasi-sublinear expression

$$T_\lambda(f, g) = \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right|^t \right)^{1/t} .$$

Raising the exponents in case (2)

In conjunction with the previous subsection, it suffices to prove (1.9) for every $p > r$ and $q > s$ satisfying the “diagonal” condition

$$(6.2) \quad \frac{p}{r} = \frac{q}{s},$$

which includes pairs with arbitrarily large p and q . Consider the function

$$T_\lambda(f, g)(x) = \sup_{(N_k)} \left(\sum_k \rho_\lambda \left(\sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(x) (\phi_j * g_j)(x) \right)^{2t} \right)^{\frac{1}{2t}}.$$

Similarly as before, (1.9) follows from bounds

$$\|\lambda^{-1} T_\lambda(f, g)^2\|_{pq/(p+q)} \leq C$$

under the normalization assumption (1.7). These are $L^{2pq/(p+q)}$ bounds for $T_\lambda(f, g)$ which we can prove by the corresponding bound for the sharp function $T_\lambda(f, g)^\sharp$.

The desired estimate now follows from bounds

$$(6.3) \quad \begin{aligned} & \lambda^{-1} (T_\lambda(f, g)^\sharp)^2 \leq C M_r \left(\left(\sum_i |f_i|^r \right)^{1/r} \right) M_s \left(\left(\sum_j |g_j|^s \right)^{1/s} \right) \\ & + C \sup_{x \in I} \left(\frac{\lambda^{1/2}}{|I|} \int_I \sup_{N_k} \left| \left\{ k : \left| \sum_{N_{k-1} < i, j \leq N_k} (\phi_i * f_i)(\phi_j * g_j - \tilde{g}_{j,I}) \right| > \lambda \right\} \right|^{\frac{1}{2t}} \right)^2 \end{aligned}$$

pointwise at every x . Here we have defined the function $\tilde{g}_{j,I}$ with respect to the interval I as in Section 2.

We will first show how to estimate the righthand side of (6.3). The first term satisfies the desired $L^{pq/(p+q)}$ bound by Hölder and the maximal theorem. We estimate the second term pointwise by a constant times

$$\begin{aligned} & M_{1/2} \left(\sup_{N_k} \lambda \left| \left\{ k : \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i \right| > \lambda^{t/r} \right\} \right|^{1/t} \right) \\ & + \sup_I \left(\frac{1}{|I|} \int_I \sup_{N_k} \lambda^{1/2} \left| \left\{ k : \left| \sum_{N_{k-1} < j \leq N_k} \phi_j * g_j - \tilde{g}_{j,I} \right| > \lambda^{t/s} \right\} \right|^{1/(2t)} \right)^2. \end{aligned}$$

Since $pq/(p+q) > 1/2$, we can estimate the $L^{pq/(p+q)}$ norm of the first term using the maximal theorem adapted to $M_{1/2}$, by

$$C \left\| \sup_{N_k} \lambda \left| \left\{ k : \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i \right| > \lambda^{t/r} \right\} \right|^{1/t} \right\|_{\frac{pq}{p+q}}.$$

Observe that $\frac{r pq}{t(p+q)} = p$, thanks to (6.2). So by (1.2), with $\lambda^{t/r}$ in place of λ , we can estimate the last display by

$$C \left\| \left(\sum_i |f_i|^r \right)^{1/r} \right\|_p^{r/t} = C.$$

To estimate the g_j -term, we decompose as in Section 2,

$$\phi_j * g_j - \tilde{g}_{j,I} = k_j^{[1]} + k_j^{[2]} + k_j^{[3]}.$$

To estimate the contribution of $k_j^{[2]}$, we will use a weak type variant of the argument in Section 2. First, by adapting the sequence N_k , one can bound this contribution by

$$\sup_I \frac{1}{|I|^2} \left\| \sup_{N_k} \lambda^{t/s} \left\{ k : \left| \sum_{N_{k-1} < j \leq N_k} \phi_j * (g_j 1_{3I}) \right| > \lambda^{t/s} \right\} \right\|_{L^{s/(2t)}(I)}^{s/t},$$

and since $t \geq 1/2$ we can use Hölder to increase $L^{s/(2t)}(I)$ to $L^s(I)$,

$$\leq \sup_I \frac{1}{|I|^{1/t}} \left\| \sup_{N_k} \lambda^{t/s} \left\{ k : \left| \sum_{N_{k-1} < j \leq N_k} \phi_j * (g_j 1_{3I}) \right| > \lambda^{t/s} \right\} \right\|_{L^s(I)}^{s/t}.$$

The point is that now we can use the diagonal case of (1.2) (with $\lambda^{t/s}$ in place of λ). This gives the following estimate for the preceding expression:

$$\leq C \sup_I \frac{1}{|I|^{1/t}} \left\| \left(\sum_j |g_j 1_{3I}|^s \right)^{1/s} \right\|_{L^s}^{s/t} \leq C \left(M \left(\sum_j |g_j|^s \right) \right)^{1/t}.$$

Since $\frac{pq}{t(p+q)} = \frac{q}{s} > 1$, we can estimate the $L^{pq/(p+q)}$ norm of the last display using the maximal theorem, by

$$\leq C \left\| \sum_j |g_j|^s \right\|_{q/s}^{1/t} = C \|g\|_{q,s}^{s/t} = C.$$

The arguments for estimating the contributions of $k_j^{[3]}$ and $k_j^{[1]}$ will be similar to each other, and below we will only show details for the latter.

The contribution of $k_j^{[1]}$ is estimated using decay in j as in Section 2, by

$$\begin{aligned} &\leq \sup_I \left(\frac{1}{|I|} \int_I \left(\sup_{N_k} \sum_k \left| \sum_{N_{k-1} < j \leq N_k} k_j^{[1]} \right|^s \right)^{1/(2t)} \right)^2 \\ &\leq C \sup_I \left(\frac{1}{|I|} \int_I \left(\sum_j |k_j^{[1]}|^s \right)^{s/(2t)} \right)^2 \leq C \sup_I \left(\frac{1}{|I|} \int_I \left(\sup_j M g_j \right)^{s/(2t)} \right)^2 \\ &\leq C M_{1/2} \left(\left(\sup_j M g_j \right)^{s/t} \right)(x) \end{aligned}$$

which satisfies the desired $L^{pq/(p+q)}$ bound by the maximal theorem and (6.2). This concludes the bound of the righthand side of (6.3).

It remains to prove the bound (6.3) for the sharp function. Fix an interval I and define constant functions \tilde{f} and \tilde{g} as in Section 2. We will use the constant

$$c := \sup_{(N_k)} \left(\sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} \tilde{f}_i \tilde{g}_j \right|^{2t} \right)^{\frac{1}{2t}}$$

for the sharp function estimate. By metric properties we obtain

$$(6.4) \quad |T_\lambda(f, g) - c| \leq T_\lambda(f - \tilde{f}, g - \tilde{g}) + T_\lambda(f - \tilde{f}, \tilde{g}) + T_\lambda(\tilde{f}, g - \tilde{g}).$$

Focusing on the first term in (6.4), we estimate the average

$$\lambda^{-1} \left(\frac{1}{|I|} \int_I T_\lambda(f - \tilde{f}, g - \tilde{g})(x) dx \right)^2$$

using the equivalent estimate with

$$\frac{1}{|I|^2} \left\| \sup_{N_k} \lambda \left\{ k : \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i - \tilde{f}_i)(\phi_j * g_j - \tilde{g}_j) \right| > \lambda \right\} \right\|_{L^{1/2}(I)}^{1/t}.$$

We will show that the contribution of $T_\lambda(f - \tilde{f}, g - \tilde{g})$ is bounded by the product of two maximal functions on the righthand side of (6.3). We split as in Section 2

$$\begin{aligned} \phi_i * f_i - \tilde{f}_i &= h_i^{[1]} + h_i^{[2]} + h_i^{[3]}, \\ \phi_j * g_j - \tilde{g}_j &= k_j^{[1]} + k_j^{[2]} + k_j^{[3]}. \end{aligned}$$

The estimate for the term involving $h^{[2]}$ and $k^{[2]}$ follows from Hölder (to increase $L^{1/2}$ to L^t) and the known estimate for the diagonal case $(p, q) = (r, s)$. For all other estimates we can use the 1-variation bound of the error terms as in Section 2 in conjunction with the endpoint estimates of Proposition 4.1. The key idea is to incorporate the restriction to the interval I into the error terms *before* applying this Proposition, because we want to preserve locality for future estimates by maximal functions. For instance, to estimate the combination $(h^{[2]}, k^{[3]})$ we first equivalently replace $k_j^{[3]}$ by $k_j^{[3]} 1_I$, then increase $L^{1/2}$ to $L^{r/(r+1)}$ and decrease t to $\frac{r}{r+1}$, then estimate using Proposition 4.1, by

$$\leq C \frac{1}{|I|^{(r+1)/r}} \| (f_i 1_{3I}) \|_{r,r} \| (k_j^{[3]} 1_I) \|_{1,1} \leq C M_r \left(\left(\sum_i |f_i|^r \right)^{1/r} \sup_j M g_j \right).$$

Turning to the second term in (6.4), note that by support of \tilde{g}_j we may restrict attention to $j < -\log_2 |I|$. Thanks to the constraint $i < j$, only the term $h_i^{[1]}$ in the splitting of the function $\phi_i * f_i - \tilde{f}_i$ then appears in the summation. While \tilde{g}_j itself does not have geometric decay in j , the geometric decay of $h_i^{[1]}$ in i is sufficient to obtain geometric summability in both i and j thanks to the constraint $i < j$. Specifically, using $t \geq 1/2$ and Hölder,

$$\begin{aligned} \lambda^{-1} \frac{1}{|I|^2} \left(\int_I T_\lambda(h^{[1]}, \tilde{g}) \right)^2 &\leq \left(\frac{1}{|I|} \int_I \left(\sum_{i < j < -\log_2 |I|} |h_i^{[1]}| |\tilde{g}_j| \right)^{1/2} \right)^2 \\ &\leq C \left(\frac{1}{|I|} \int_I \left(\sum_{i < j < -\log_2 |I|} (2^i |I|) \inf_{y \in I} M f_i(y) \inf_{y \in I} M g_j(y) \right)^{1/2} \right)^2 \\ &\leq C \sup_i M f_i(x) \sup_j M g_j(x). \end{aligned}$$

This proves the desired estimate for the second term. Combining the first and second terms in (6.4) we also obtain bounds for the variant of the second term with \tilde{g} replaced by g .

Turning to the last term in (6.4), we may also replace \tilde{f} with f . To apply the same argument as for the second term, we need to replace the constraint $i < j$ by $i \geq j$, which can be done via triangle inequality and estimates on the unconstrained expression

$$\frac{1}{|I|^2} \left\| \sup_{N_k} \lambda \left| \left\{ k : \left| \sum_{N_{k-1} < i, j \leq N_k} (\phi_i * f_i)(\phi_j * g_j - \tilde{g}_j) \right| > \lambda \right\} \right|^{1/t} \right\|_{L^{1/2}(I)}.$$

This term can be estimated by the second term on the righthand side of (6.3). This completes the proof of the pointwise bound for the sharp function.

Raising exponents in case (1)

The argument in this section is similar to and simpler than last section and we will only comment on the necessary changes. Consider a pair (p, q) such that $r < p < \infty, s < q < \infty$. To show the desired $L^{\frac{pq}{p+q}}$ bound for

$$S(f, g) = \left(\sup_{N_k} \sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} (\phi_i * f_i)(\phi_j * g_j) \right|^t \right)^{1/t}$$

we will show the corresponding bound for $T(f, g)^\sharp$, where $T(f, g) = S(f, g)^{1/2}$. Below we will show the pointwise bound:

$$(6.5) \quad T(f, g)^\sharp \leq C \left(M_r \left(\left(\sum_i |f_i|^r \right)^{1/r} \right) M_s \left(\left(\sum_j |g_j|^s \right)^{1/s} \right) \right)^{1/2} + C \left(M(\|\phi_i * f_i\|_{V^r}) M_s \left(\left(\sum_j |g_j|^s \right)^{1/s} \right) \right)^{1/2}.$$

Using this inequality, the desired $L^{\frac{2pq}{p+q}}$ bound for $T(f, g)^\sharp$ follows by Hölder’s inequality, the maximal theorem and Theorem 1.1.

Below we show (6.5). Fix any interval I and define \tilde{f} and \tilde{g} as in Section 2. We will use the following constant for the sharp function estimate

$$c = \left(\sup_{N_k} \sum_k \left| \sum_{N_{k-1} < i < j \leq N_k} \tilde{f}_i \tilde{g}_j \right|^t \right)^{1/t}.$$

Since $\frac{1}{2} \leq t < 1$, both $|x|^t$ and $|x|^{1/(2t)}$ define a norm, so $T(f, g)$ is sublinear. We actually do not need $t < 1$ for this to be true, because when $t \geq 1$ we could combine the $|x|^{1/2}$ and $\ell^t(\mathbb{Z})$ norms. We then have

$$|T(f, g) - c| \leq T(f - \tilde{f}, g - \tilde{g}) + T(f - \tilde{f}, \tilde{g}) + T(\tilde{f}, g - \tilde{g}).$$

The contributions of $T(f - \tilde{f}, g - \tilde{g})$ and of $T(f - \tilde{f}, \tilde{g})$ can be estimated as in the consideration of case (2) in the last subsection. To estimate the contribution of $T(\tilde{f}, g - \tilde{g})$, as before it suffices to estimate the unconstrained expression:

$$\frac{1}{|I|} \int_I \sup_{N_k} \left(\sum_k \left| \sum_{N_{k-1} < i, j \leq N_k} \tilde{f}_i(\phi_j * g_j - \tilde{g}_j) \right|^t \right)^{\frac{1}{2t}}.$$

By Hölder’s inequality we can estimate this by

$$\leq \sup_{N_k} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \tilde{f}_i |^r \right|^{\frac{1}{2r}} \frac{1}{|I|} \int_I \left(\sup_{N_k} \sum_k \left| \sum_{N_{k-1} < i \leq N_k} (\phi_j * g_j - \tilde{g}_j) \right|^s \right)^{\frac{1}{2s}} \right).$$

Using the estimates in Section 2 we can estimate the last display by

$$\leq C \left(\sup_{N_k} \left(\sum_k \left| \sum_{N_{k-1} < i \leq N_k} \tilde{f}_i |^r \right|^{\frac{1}{r}} \right)^{1/2} \left(M_s \left(\left(\sum_j |g_j|^s \right)^{1/s} \right) \right)^{1/2} \right).$$

Now, since

$$\tilde{f}_i = \frac{1}{|I|} \int_I \phi_i * f_i$$

for $i < -\log_2 |I|$ and $\tilde{f}_i = 0$ otherwise, by Minkowski’s inequality we can estimate the last display by

$$\begin{aligned} &\leq C \left(\frac{1}{|I|} \int_I \left(\sup_{N_k} \sum_k \left| \sum_{N_{k-1} < i \leq N_k} \phi_i * f_i |^r \right|^{\frac{1}{r}} \right)^{1/2} \left(M_s \left(\left(\sum_j |g_j|^s \right)^{1/s} \right) \right)^{1/2} \right) \\ &\leq C \left(M(\|\phi_i * f_i\|_{V^r}) M_s \left(\left(\sum_j |g_j|^s \right)^{1/s} \right) \right)^{1/2}. \end{aligned}$$

References

- [1] BOURGAIN, J.: Pointwise ergodic theorems for arithmetic sets. *Inst. Hautes Études Sci. Publ. Math.* **69** (1989), 5–45.
- [2] FRIZ, P. AND VICTOIR, N.: The Burkholder–Davis–Gundy inequality for enhanced martingales. In *Séminaire de probabilités XLI*, 421–438. Lecture Notes in Math., 1934. Springer, Berlin, 2008.
- [3] JANSON, S.: On interpolation of multilinear operators. In *Function spaces and applications (Lund, 1986)*, 290–302. Lecture Notes in Math., 1302. Springer, Berlin, 1988.
- [4] JONES, R. L., KAUFMAN, R., ROSENBLATT, J. M. AND WIERDL, M.: Oscillation in ergodic theory. *Ergod. Theory Dynam. Systems* **18** (1998), 889–935.
- [5] JONES, R. L., SEEGER, A. AND WRIGHT, J.: Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.* **360** (2008), no. 12, 6711–6742.
- [6] LÉPINGLE, D.: La variation d’ordre p des semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **36** (1976), 295–316.

- [7] LYONS, T.: Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14** (1998), no. 2, 215–310.
- [8] MEYER, Y. AND COIFMAN, R.: *Wavelets. Calderón–Zygmund and multilinear operators*. Cambridge Studies in Advanced Mathematics, 48. Cambridge University Press, Cambridge, 1997.
- [9] MUSCALU, C., TAO, T. AND THIELE, C.: L^p estimates for the biest. II. The Fourier case. *Math. Ann.* **329** (2004), no. 3, 427–461.
- [10] MUSCALU, C., TAO, T. AND THIELE, C.: The bi-Carleson operator. *Geom. Funct. Anal.* **16** (2006), no. 1, 230–277.
- [11] OBERLIN, R., SEEGER, A., TAO, T., THIELE, C. AND WRIGHT, J.: A variation norm Carleson theorem. *J. Eur. Math. Soc.* **14** (2012), no. 2, 421–464.
- [12] PISIER, G. AND XU, Q.: The strong p -variation of martingales and orthogonal series. *Probab. Theory Related Fields* **77** (1988), 497–514.
- [13] STEIN, E.: *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

Received September 15, 2010; revised April 8, 2011.

YEN DO: Department of Mathematics, Yale University, New Haven, CT 06520-8283, USA.

E-mail: yen.do@yale.edu

CAMIL MUSCALU: Department of Mathematics, Cornell University, Ithaca, NY 14852-4201, USA.

E-mail: camil@math.cornell.edu

CHRISTOPH THIELE: Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA.

E-mail: thiele@math.ucla.edu

This material is based upon work supported by the National Science Foundation under grant DMS-0635607002. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. C.M. partially supported by NSF grant DMS-0653519. C.Th. partially supported by NSF grant DMS 1001535.