



# Improved Range in the Return Times Theorem

Ciprian Demeter

*Abstract.* We prove that the Return Times Theorem holds true for pairs of  $L^p - L^q$  functions, whenever  $\frac{1}{p} + \frac{1}{q} < \frac{3}{2}$ .

## 1 Introduction

Let  $\mathbf{X} = (X, \Sigma, \mu, \tau)$  be a dynamical system, *i.e.*, a Lebesgue space  $(X, \Sigma, \mu)$  equipped with an invertible bimeasurable measure-preserving transformation  $\tau: X \rightarrow X$ . We recall that a complete probability space  $(X, \Sigma, \mu)$  is called a *Lebesgue space* if it is isomorphic with the ordinary Lebesgue measure space  $([0, 1], \mathcal{L}, m)$ , where  $\mathcal{L}$  and  $m$  denote the usual Lebesgue algebra and measure (see [10] for more on this topic). The system  $\mathbf{X}$  is called *ergodic* if  $A \in \Sigma$  and  $\mu(A \Delta \tau^{-1}A) = 0$  imply  $\mu(A) \in \{0, 1\}$ .

Bourgain proved the following result [4].

**Theorem 1.1** (Return Times Theorem) *For each function  $f \in L^\infty(X)$  there is a universal set  $X_0 \subseteq X$  with  $\mu(X_0) = 1$  such that for each second dynamical system  $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$ , each  $g \in L^\infty(Y)$ , and each  $x \in X_0$ , the averages*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(\tau^n x) g(\sigma^n y)$$

*converge  $\nu$ -almost everywhere.*

Subsequent proofs were given in [6, 18]. If in the above theorem  $f$  (or  $g$ ) is taken to be a constant function, one recovers the classical Birkhoff pointwise ergodic theorem; see [3]. However, Theorem 1.1 is much stronger, in that it shows that given  $f$  for almost every  $x$ , the sequence  $w_n = (f(\tau^n x))_{n \in \mathbb{N}}$  forms a system of universal weights for the pointwise ergodic theorem.

The difficulty in Theorem 1.1 lies in the fact that the weights provided by  $f$  work for *every* dynamical system  $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$ . If, on the other hand, the system  $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$  is fixed, then the result follows from an approximation argument combined with applications of Birkhoff's theorem to the functions  $f \otimes g_j$  in the product system  $\mathbf{X} \times \mathbf{Y}$ , where  $(g_j)_j$  is a dense class of functions in  $L^2(Y)$ .

A result by Assani, Buczolic, and Mauldin [1] shows that the return times theorem fails when  $p = q = 1$ .

---

Received by the editors August 27, 2009.

Published electronically June 14, 2011.

The author was supported by a Sloan Research Fellowship and by NSF Grant DMS-0556389

AMS subject classification: 42B25, 37A45.

Keywords: Return Times Theorem, maximal multiplier, maximal inequality.

**Theorem 1.2** ([1]) *Let  $\mathbf{X} = (X, \Sigma, \mu, \tau)$  be an ergodic dynamical system. There exist a function  $f \in L^1(X)$  and a subset  $X_0 \subseteq X$  of full measure with the following property: for each  $x_0 \in X_0$  and for each ergodic dynamical system  $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$ , there exists  $g \in L^1(Y)$  such that the averages*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(\tau^n x_0) g(\sigma^n y)$$

*diverge for almost every  $y$ .*

On the other hand, Hölder’s inequality and an elementary density argument show that Bourgain’s theorem holds for  $f \in L^p(X)$  and  $g \in L^q(Y)$ , whenever  $1 \leq p, q \leq \infty$ , and  $\frac{1}{p} + \frac{1}{q} \leq 1$ ; see [18] or [9, §4]. It is an interesting question to understand the precise range of  $p$  and  $q$  for which a positive result holds.

Significant progress on this issue appears in [9], where it was proved that the return times theorem remains valid when  $q \geq 2$  and  $p > 1$ . We build on the approach from [9] and prove the following result.

**Theorem 1.3** *Let  $1 < p, q \leq \infty$  be such that*

$$(1.1) \quad \frac{1}{p} + \frac{1}{q} < \frac{3}{2}.$$

*For each dynamical system  $(X, \Sigma, \mu, \tau)$  and each  $f \in L^p(X)$  there is a universal set  $X_0 \subseteq X$  with  $\mu(X_0) = 1$  such that for each second dynamical system  $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$ , each  $g \in L^q(Y)$ , and each  $x \in X_0$ , the averages*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(\tau^n x) g(\sigma^n y)$$

*converge  $\nu$ -almost everywhere.*

Given the result in [9] and the convergence for  $L^\infty$  functions  $f$  and  $g$ , an approximation argument will immediately prove Theorem 1.3, once we establish the following maximal inequality.

**Theorem 1.4** *Let  $1 < p < \infty$  and  $1 < q < 2$  satisfy (1.1). For each dynamical system  $\mathbf{X} = (X, \Sigma, \mu, \tau)$  and each  $f \in L^p(X)$*

$$\left\| \sup_{(Y, \mathcal{F}, \nu, \sigma)} \sup_{\|g\|_{L^q(Y)=1}} \left\| \sup_N \left| \frac{1}{N} \sum_{n=1}^N f(\tau^n x) g(\sigma^n y) \right| \right\|_{L^q(Y)} \right\|_{L^p(X)} \lesssim_{p,q} \|f\|_{L^p(X)},$$

*where the first supremum in the inequality above is taken over all dynamical systems  $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$ .*

Here and in the following, we have subscripted some of our  $L^p$  norms to clarify the variable of integration. As explained in [9], this theorem will follow by standard transfer, from the following real line version.

**Theorem 1.5** *Let  $1 < p < \infty$  and  $1 < q < 2$  satisfy (1.1). For each  $f \in L^p(\mathbf{R})$  we have*

$$\left\| \sup_{\|g\|_{L^q(\mathbf{R})}=1} \left\| \sup_{k \in \mathbf{Z}} \frac{1}{2^{k+1}} \int_{-2^k}^{2^k} |f(x+y)g(z+y)| dy \right\|_{L^q_2(\mathbf{R})} \right\|_{L^p_x(\mathbf{R})} \lesssim_{p,q} \|f\|_{L^p(\mathbf{R})}.$$

When  $1 \leq p, q \leq \infty$  are in the duality range, that is, when  $\frac{1}{p} + \frac{1}{q} \leq 1$ , Theorem 1.5 follows immediately from Hölder’s inequality. The case  $q = 2, 1 < p < \infty$  was proved in [9]. The approach from [9] consists of treating averages and singular integrals in a similar way: one performs Littlewood–Paley decompositions of each average, combined with Gabor frames expansions of  $f$ , to obtain a model sum. This discretized operator turns out to be a maximal truncation of the Carleson operator [7]

$$Cf(x, \theta) = p.v. \int_{\mathbf{R}} f(x+y)e^{iy\theta} \frac{dy}{y}.$$

The analysis in [9] is then driven by time-frequency techniques combined with an  $L^2$  maximal multiplier result of Bourgain. Most of the work in [9] is  $L^2$  based, and in particular, the fact that  $q = 2$  in Theorem 1.5 is heavily exploited.

In this paper, we relax the restriction  $q = 2$ , and replace it with (1.1). There are two key new ingredients. The first is a simplification of the argument from [9], which consists in treating the Hardy–Littlewood kernel in a way distinct from the Hilbert kernel. In [9], the two kernels were treated on equal footing, as a byproduct of a unified approach for regular averages and signed averages. Here we treat each average as a single Littlewood–Paley piece. This decomposition simplifies the model sum to a significant extent, and is suited for analysis on spaces other than  $L^2$ . The main new ingredient we use here is the  $L^q$  version of Bourgain’s result on maximal multipliers that was proved in [8] in the frequency separated case, and in [17] in the general case (see Theorem 4.1 below).

It is interesting to remark that the range (1.1) that we establish is the same as the range where the bilinear Hilbert transform (see [12, 13])

$$BHT(f, g)(x) = p.v. \int_{\mathbf{R}} f(x+y)g(x-y) \frac{dy}{y}$$

and the bilinear maximal function (see [11])

$$BM(f, g)(x) = \sup_{t>0} \left| \frac{1}{2t} \int_{-t}^t f(x+y)g(x-y) dy \right|$$

are known to be bounded. This is perhaps not a coincidence, as the methods we use to prove Theorem 1.5 are related to those used in the proof of the bilinear Hilbert transform. Moreover, in both cases, the methods fail beyond the  $3/2$  threshold, essentially because of the same reason. Even the model sum that contains a single scale is unbounded if  $\frac{1}{p} + \frac{1}{q} \geq 3/2$ . Another interesting connection is that both the boundedness of the bilinear maximal function and the return times theorem fail for pairs of

$L^1$  functions, and they do so in quite a dramatic way. Even the (smaller) tail operators

$$T_1(f, g)(x) := \sup_{t>1} \left| \frac{1}{2t} \int_t^{t+1} f(x+y)g(x-y) dy \right|,$$

$$T_2(f, g)(x, y) := \sup_n \left| \frac{1}{n} f(\tau^n x)g(\sigma^n y) \right|,$$

fail to be bounded for pairs of  $L^1$  functions. See [1, 2].

## 2 Discretization

Let  $m_k: \mathbf{R} \rightarrow \mathbf{R}$  be a sequence of multipliers. For each  $1 \leq q \leq \infty$ , the maximal multiplier norm associated with them is defined as

$$\|(m_k)_{k \in \mathbf{Z}}\|_{M_q^*(\mathbf{R})} := \sup_{\|g\|_q=1} \left\| \sup_k \left| \int m_k(\theta) \widehat{g}(\theta) e^{2\pi i \theta z} d\theta \right| \right\|_{L_x^q(\mathbf{R})}.$$

Let  $K: \mathbf{R} \rightarrow [0, \infty)$  be a positive function with  $K(0) > 0$ , whose Fourier transform is supported in, say, the interval  $[-1, 1]$ . In particular, one can take  $K$  to be the inverse Fourier transform of  $\eta * \tilde{\eta}$ , where  $\eta: \mathbf{R} \rightarrow \mathbf{R}$  is supported in  $[-1/2, 1/2]$ ,  $\int \eta \neq 0$ , and  $\tilde{\eta}(\xi) = \eta(-\xi)$ . Of course, Theorem 1.5 will immediately follow if we can prove the same thing with

$$\sup_{\|g\|_{L^q(\mathbf{R})}=1} \left\| \sup_{k \in \mathbf{Z}} \frac{1}{2^{k+1}} \int_{-2^k}^{2^k} |f(x+y)g(z+y)| dy \right\|_{L_x^q(\mathbf{R})}$$

replaced by

$$Rf(x) := \sup_{\|g\|_{L^q(\mathbf{R})}=1} \left\| \sup_{k \in \mathbf{Z}} \frac{1}{2^k} \left| \int f(x+y)g(z+y)K\left(\frac{y}{2^k}\right) dy \right| \right\|_{L_x^q(\mathbf{R})}.$$

As remarked earlier, whenever  $p \geq \frac{q}{q-1}$ , we know that  $R$  maps  $L^p$  to  $L^p$ . By invoking restricted weak type interpolation, it thus suffices to prove that

$$(2.1) \quad m\{x : R1_F(x) > \lambda\} \lesssim_{p,q} \frac{|F|}{\lambda^p},$$

for each  $p < 2$ ,  $\lambda \leq 1$  and each finite measure set  $F \subset \mathbf{R}$ .

We next indicate how to discretize the operator  $R$ . Rather than going through the whole procedure in detail, we emphasize its key aspects. The interested reader is referred to [9, §6] for details. We note however that our approach here is a simplified version of the decomposition in [9], since we no longer perform Littlewood–Paley decompositions of a given average.

Let  $\varphi$  be a Schwartz function such that  $\widehat{\varphi}$  is supported in  $[0, 1]$  and such that

$$\sum_{l \in \mathbf{Z}} \left| \widehat{\varphi}\left(\xi - \frac{l}{2}\right) \right|^2 \equiv C.$$

If  $C$  is chosen appropriately, it will follow that for each  $F$  and each  $k \in \mathbf{Z}$ , one has the following Gabor basis expansion:

$$\sum_{m,l \in \mathbf{Z}} \langle 1_F, \varphi_{k,m,l/2} \rangle \varphi_{k,m,l/2} = 1_F.$$

Here  $\varphi_{k,m,l}(x) := 2^{-\frac{k}{2}} \varphi(2^{-k}x - m) e^{2\pi i 2^{-k}xl}$  is the  $L^2$  normalized wave packet that is quasi-localized in time frequency in the rectangle  $[m2^k, (m+1)2^k] \times [l2^{-k}, (l+1)2^{-k}]$ .

Given a scale  $2^k$ , one uses this expansion to get

$$\begin{aligned} R1_F(x) &= \sup_{\|g\|_{L^q(\mathbf{R})}=1} \left\| \sup_{k \in \mathbf{Z}} \left| \sum_{m,l \in \mathbf{Z}} \langle 1_F, \varphi_{k,m,l/2} \rangle \int \varphi_{k,m,l/2}(x+y) g(z+y) 2^{-k} K\left(\frac{y}{2^k}\right) dy \right| \right\|_{L^q(\mathbf{R})} \\ &= \left\| \left( \sum_{m,l \in \mathbf{Z}} \langle 1_F, \varphi_{k,m,l/2} \rangle \mathcal{F}[\varphi_{k,m,l/2}(x+\cdot) 2^{-k} K\left(\frac{\cdot}{2^k}\right)](\theta) \right)_{k \in \mathbf{Z}} \right\|_{M_{q,\theta}^*(\theta)}. \end{aligned}$$

The key observation is that the function

$$\phi_{k,m,l/2}(x, \theta) = \mathcal{F} \left[ \varphi_{k,m,l/2}(x+\cdot) 2^{-k} K\left(\frac{\cdot}{2^k}\right) \right] (\theta)$$

has the same decay (in  $x$ ) as  $\varphi_{k,m,l/2}$  and behaves like the function

$$\varphi_{k,m,l/2}(x) 1_{[l2^{-k}, (l+1)2^{-k}]}(\theta).$$

Note that in reality, the support in  $\theta$  of  $\phi_{k,m,l/2}(x, \theta)$  is slightly larger than

$$[l2^{-k}, (l+1)2^{-k}].$$

More precisely, it is a subset of  $[l2^{-k}, (l+1)2^{-k}] + [-2^{-k}, 2^{-k}]$ . This will force upon us the use of shifted dyadic grids. But, as explained in [9], for simplicity of notation (but not of the argument), we can really assume that we are working with the standard dyadic grid.

We will denote by  $\mathbf{S}_{\text{univ}}$  the collection of all tiles  $s = I_s \times \omega_s$  with area 1, where both  $I_s$  and  $\omega_s$  are dyadic intervals. We will refer to  $I_s, \omega_s$  as the time and frequency components of  $s$ .

**Definition 2.1** A collection  $\mathbf{S} \subset \mathbf{S}_{\text{univ}}$  of tiles will be referred to as convex, if whenever  $s, s'' \in \mathbf{S}$  and  $s' \in \mathbf{S}_{\text{univ}}, \omega_{s''} \subseteq \omega_{s'} \subseteq \omega_s$  and  $I_s \subseteq I_{s'} \subseteq I_{s''}$  will imply that  $s' \in \mathbf{S}$ .

The fact that we choose to work with convex collections of tiles is of a technical nature. It will allow us to use some results like Proposition 3.7 which are known to hold under the convexity assumption.

As explained in [9], (2.1) now follows from the following theorem.

**Theorem 2.2** *Let  $1 < q < 2$ . Let  $\mathbf{S}$  be an arbitrary convex finite collection of tiles. Consider also two collections,  $\{\phi_s, s \in \mathbf{S}\}$  and  $\{\varphi_s, s \in \mathbf{S}\}$ , of Schwartz functions. We assume the functions  $\phi_s: \mathbf{R}^2 \rightarrow \mathbf{R}$  satisfy*

$$(2.2) \quad \text{supp}_\theta(\phi_s(x, \theta)) \subseteq \omega_s \quad \text{for each } x,$$

$$(2.3) \quad \text{supp}_\xi(\mathcal{F}_x(\phi_s(x, \theta))(\xi)) \subseteq \omega_s \quad \text{for each } \theta,$$

$$(2.4) \quad \sup_{c \in \omega_s} \left\| \frac{\partial^n}{\partial \theta^n} \frac{\partial^m}{\partial x^m} [\phi_s(x, \theta) e^{-2\pi i c x}] \right\|_{L^\infty(\mathbf{R})} \lesssim_{n,m,M} |I_s|^{(n-m-1/2)} \chi_{I_s}^M(x) \quad \forall n, m, M \geq 0,$$

uniformly in  $s$ . We also assume that the functions  $\varphi_s: \mathbf{R} \rightarrow \mathbf{R}$  satisfy  $\text{supp}(\widehat{\varphi}_s) \subseteq \omega_s$  and

$$(2.5) \quad \sup_{c \in \omega_s} \left| \frac{\partial^n}{\partial x^n} [\varphi_s(x) e^{-2\pi i c x}] \right| \lesssim_{n,M} |I_s|^{-n-\frac{1}{2}} \chi_{I_s}^M(x) \quad \forall n, M \geq 0,$$

uniformly in  $s$ .

Then the following inequality holds for each measurable  $F \subset \mathbf{R}$  with finite measure, each  $0 < \lambda \leq 1$ , and each  $1 < p < 2$  such that  $\frac{1}{p} + \frac{1}{q} < \frac{3}{2}$ :

$$m \left\{ x : \left\| \left( \sum_{\substack{s \in \mathbf{S} \\ |I_s|=2^k}} \langle 1_F, \varphi_s \rangle \phi_s(x, \theta) \right)_{k \in \mathbf{Z}} \right\|_{M_{q,\theta}^*(\mathbf{R})} > \lambda \right\} \lesssim \frac{|F|}{\lambda^p}.$$

The implicit constant depends only on  $p, q$  and on the implicit constants in (2.4) and (2.5) (in particular, it is independent of  $\mathbf{S}, F$  and  $\lambda$ ).

The rest of the paper is devoted to proving this theorem. We fix the collection  $\mathbf{S}$  throughout the rest of the paper.

### 3 Some Results on Trees

We now recall some facts about trees. We refer the reader to [12, 16, 19] for more details.

**Definition 3.1** (Tile order) For two tiles  $s$  and  $s'$  we write  $s \leq s'$  if  $I_s \subseteq I_{s'}$  and  $\omega_{s'} \subseteq \omega_s$ .

**Definition 3.2** (Trees) A tree with top  $(I_T, \xi_T)$ , where  $I_T$  is an arbitrary (not necessarily dyadic) interval and  $\xi_T \in \mathbf{R}$ , is a convex collection of tiles  $\mathbf{T} \subseteq \mathbf{S}$  such that  $I_s \subset I_T$  and  $\xi_T \in \omega_s$  for each  $s \in \mathbf{T}$ .

We will say that the tree has top tile  $T \in \mathbf{T}$  if  $s \leq T$  for each  $s \in \mathbf{T}$ .

**Remark 3.3** Not all trees have a top tile, but, of course, each tree can be (uniquely) decomposed into a disjoint union of trees with top tiles, such that these top tiles are pairwise disjoint.

Note also each tree  $\mathbf{T}$  with top tile  $T$  can be regarded as a tree with top  $(I, \xi)$ , for each interval  $I_T \subseteq I$  and each  $\xi \in \omega_T$ . If this is the case, we will adopt the convention that  $I_{\mathbf{T}} := I_T$ .

It is important to emphasize the following technical point. In [9] there is a clear distinction between the so-called 1-trees (overlapping) and 2-trees (lacunary) that is completely ignored in the current argument. This simplification is due to the fact that we only investigate maximal functions, rather than singular integrals. But in many ways, the trees in the current setting behave like overlapping trees.

We now recall a few definitions and results from [8]. We will denote by  $T_m$  the Fourier quasi-projection associated with the multiplier  $m$ :

$$T_m f(x) := \int \widehat{f}(\xi) m(\xi) e^{2\pi i \xi x} d\xi.$$

We will use the notation

$$\tilde{\chi}_I(x) = \left( 1 + \frac{|x - c(I)|}{|I|} \right)^{-1}.$$

**Definition 3.4** Let  $f$  be an  $L^2$  function and let  $\mathbf{S}' \subset \mathbf{S}$ . We define the *size* of  $\mathbf{S}'$  relative<sup>1</sup> to  $f$  as

$$\text{size}(\mathbf{S}') := \sup_{s \in \mathbf{S}'} \sup_{m_s} \frac{1}{|I_s|^{1/2}} \|\tilde{\chi}_{I_s}^{10}(x) T_{m_s} f(x)\|_{L^2_s},$$

where  $m_s$  ranges over all functions adapted to  $10\omega_s$ .

Each tree defines a region in the time-frequency plane. A good heuristic for the size of the tree is to think of it as being comparable to the  $L^\infty$  norm of the restriction of  $f$  to this region. This heuristic is made precise by means of the phase space projections. We refer the reader to [8, 15] for more details.

We recall two important results regarding the size. The first one is immediate.

**Proposition 3.5** For each  $\mathbf{S}' \subset \mathbf{S}$  and each  $f \in L^1(\mathbf{R})$  we have

$$\text{size}(\mathbf{S}') \lesssim \sup_{s \in \mathbf{S}'} \inf_{x \in I_s} Mf(x),$$

where the size is understood with respect to  $f$ .

The following Bessel type inequality from [14] will be useful in organizing collections of tiles into trees. See also [8, Lemma 4.11] for a proof.

<sup>1</sup>The function with respect to which the size is computed will change throughout the paper; however, it will always be clear from the context

**Proposition 3.6** Let  $S' \subseteq S$  be a convex collection of tiles and define

$$\Delta := \lceil -\log_2(\text{size}(S')) \rceil,$$

where the size is understood with respect to some function  $f \in L^2(\mathbf{R})$ . Then  $S'$  can be written as a disjoint union  $S' = \bigcup_{n \geq \Delta} \mathcal{P}_n$ , where  $\text{size}(\mathcal{P}_n) \leq 2^{-n}$  and each  $\mathcal{P}_n$  is convex and consists of a family  $\mathcal{F}_{\mathcal{P}_n}$  of pairwise disjoint trees (that is, distinct trees do not share tiles)  $\mathbf{T}$  with top tiles  $T$  satisfying

$$\sum_{\mathbf{T} \in \mathcal{F}_{\mathcal{P}_n}} |I_{\mathbf{T}}| \lesssim 2^{2n} \|f\|_2^2,$$

with bounds independent of  $S'$ ,  $n$ , and  $f$ .

We next recall an important decomposition from [9]. Let  $\mathbf{T}$  be a tree with top  $(I_{\mathbf{T}}, \xi_{\mathbf{T}})$ . For each  $s \in \mathbf{T}$  and scale  $l \geq 0$  we split  $\phi_s(x, \theta)$  as

$$\phi_s(x, \theta) = \tilde{\phi}_{s, \mathbf{T}}^{(l)}(x, \theta) + \phi_{s, \mathbf{T}}^{(l)}(x, \theta).$$

For convenience, we set  $\phi_{s, \mathbf{T}}^{(0)} := \phi_s$  for each  $s \in \mathbf{T}$ . For  $l \geq 1$  we define the first piece to be localized in time:

$$\text{supp } \tilde{\phi}_{s, \mathbf{T}}^{(l)}(\cdot, \theta) \subseteq 2^{l-1}I_s, \quad \text{for each } \theta \in \mathbf{R}.$$

For the second piece we need some degree of frequency localization, but obviously full localization, as in the case of  $\phi_s$  is impossible. We will content ourselves with preserving the mean zero property with respect to the top of the tree. The advantage of  $\phi_{s, \mathbf{T}}^{(l)}$  over  $\phi_s$  is that it gains extra decay in  $x$ . More precisely, we have for each  $s \in \mathbf{T}$  and each  $M \geq 0$ ,

$$(3.1) \quad \phi_{s, \mathbf{T}}^{(l)}(x, \theta)e^{-2\pi i \xi_{\mathbf{T}} x} \text{ has mean zero, } \theta \in \mathbf{R},$$

$$(3.2) \quad \phi_{s, \mathbf{T}}^{(l)}(x, \theta)e^{-2\pi i \xi_{\mathbf{T}} x} \text{ is } c(M)2^{-Ml}\text{-adapted to } I_s \text{ for some constant } c(M), \theta \in \mathbf{R},$$

$$(3.3) \quad \text{supp } \phi_{s, \mathbf{T}}^{(l)}(x, \cdot) \subset \omega_{s, 2}, \text{ for each } x \in \mathbf{R},$$

$$(3.4) \quad \left| \frac{d}{d\theta} \phi_{s, \mathbf{T}}^{(l)}(x, \theta) \right| \lesssim 2^{-Ml} |I_s|^{\frac{1}{2}} \chi_{I_s}^M(x), \text{ uniformly in } x, \theta \in \mathbf{R}.$$

We achieve this decomposition by first choosing a smooth function  $\eta$  such that  $\text{supp}(\eta) \subset [-1/2, 1/2]$  and  $\eta = 1$  on  $[-1/4, 1/4]$ . We then define

$$\tilde{\phi}_{s, \mathbf{T}}^{(l)}(\theta; x) := \phi_s(\theta; x) \text{Dil}_{2^l I_s}^\infty \eta(x) - \frac{e^{2\pi i \xi_{\mathbf{T}} x} \text{Dil}_{2^l I_s}^\infty \eta(x)}{\int_{\mathbf{R}} \text{Dil}_{2^l I_s}^\infty \eta(x) dx} \int_{\mathbf{R}} \phi_s(\theta; x) e^{-2\pi i \xi_{\mathbf{T}} x} \text{Dil}_{2^l I_s}^\infty \eta(x) dx$$

and

$$\begin{aligned} \phi_{s, \mathbf{T}}^{(l)}(\theta; x) := & \\ & \frac{e^{2\pi i \xi_{\mathbf{T}} x} \text{Dil}_{2^l I_s}^\infty \eta(x)}{\int_{\mathbf{R}} \text{Dil}_{2^l I_s}^\infty \eta(x) dx} \int_{\mathbf{R}} \phi_s(\theta; x) e^{-2\pi i \xi_{\mathbf{T}} x} \text{Dil}_{2^l I_s}^\infty \eta(x) dx + \phi_s(\theta; x)(1 - \text{Dil}_{2^l I_s}^\infty \eta(x)). \end{aligned}$$



Properties (3.1) through (3.4) are now easy consequences of (2.2), (2.3) and (2.4).

The following result is essentially Proposition 4.9 from [8]. It can also be regarded as the “overlapping” counterpart of the “lacunary” result in Theorem 9.4 from [9].

**Proposition 3.7** *For each tree  $\mathbf{T}$  with top  $(I_{\mathbf{T}}, \xi_{\mathbf{T}})$ , and each  $l, M \geq 0, r > 2$  and  $1 < t < \infty$ ,*

$$\left\| \left\| \sum_{\substack{s \in \mathbf{T} \\ |I_s|=2^k}} \langle f, \varphi_s \rangle \phi_{s, \mathbf{T}}^{(l)}(x, \xi_{\mathbf{T}}) \right\|_{V_k^r} \right\|_{L_x^t(\mathbf{R})} \lesssim 2^{-Ml} \text{size}(\mathbf{T}) |I_{\mathbf{T}}|^{1/t},$$

with the implicit constants depending only on  $r, t$  and  $M$ .

### 4 A Result on Maximal Multipliers

Consider a finite set  $\Lambda = \{\lambda_1, \dots, \lambda_N\} \subset \mathbf{R}$ . For each  $k \in \mathbf{Z}$  define  $R_k$  to be the collection of all dyadic intervals of length  $2^k$  containing an element from  $\Lambda$ .

For each  $1 \leq r < \infty$  and each sequence  $(x_k)_{k \in \mathbf{Z}} \in \mathbf{C}$ , define the  $r$ -variational norm of  $(x_k)_{k \in \mathbf{Z}}$  to be

$$\|x_k\|_{V_k^r} := \sup_k |x_k| + \|x_k\|_{\tilde{V}_k^r},$$

where

$$\|x_k\|_{\tilde{V}_k^r} := \sup_{M, k_0 < k_1 < \dots < k_M} \left( \sum_{m=1}^M |x_{k_m} - x_{k_{m-1}}|^r \right)^{1/r}.$$

For each interval  $\omega \in R_k$ , let  $m_\omega$  be a complex valued Schwartz function  $C$ -adapted to  $\omega$ , that is, supported on  $\omega$  and satisfying

$$\|\partial^\alpha m_\omega\|_\infty \leq C |\omega|^{-\alpha}, \quad \alpha \in \{0, 1, \dots, M\},$$

for some  $M$  large enough, whose value is not important. Define

$$\Delta_k f(x) := \sum_{\omega \in R_k} \int m_\omega(\xi) \widehat{f}(\xi) e^{2\pi i \xi x} d\xi,$$

and also

$$\|m_\omega\|_{V^{r,*}} := \max_{1 \leq n \leq N} \|\{m_{\omega_k}(\lambda_n) : \lambda_n \in \omega_k \in R_k\}\|_{V_k^r}.$$

The following result was proved in [8] for the case when frequencies in  $\Lambda$  are separated. The general case was proved in [17].

**Theorem 4.1** *Let  $1 < q < 2$  and  $r > 2$ . There exists  $\epsilon(r) \rightarrow 0$  as  $r \rightarrow 2$ , such that for each  $f \in L^q(\mathbf{R})$  we have the inequality*

$$\left\| \sup_k |\Delta_k f(x)| \right\|_{L_x^q(\mathbf{R})} \lesssim N^{1/q-1/2+\epsilon(r)} (C + \|m_\omega\|_{V^{r,*}}) \|f\|_q,$$

with the implicit constant depending only on  $r$  and  $q$ .

### 5 Pointwise Estimates Outside Exceptional Sets

Let  $\mathcal{P}$  be a finite convex collection of tiles which can be written as a disjoint union of trees  $\mathbf{T}$  with tops  $T \mathcal{P} = \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}$ . To better quantify the contribution coming from individual tiles, we need to reorganize the collection  $\mathcal{F}$  in a more suitable way. For each  $\mathbf{T} \in \mathcal{F}$  define its saturation  $G(\mathbf{T}) := \{s \in \mathcal{P} : \omega_T \subseteq \omega_s\}$ . For the purpose of organizing  $G(\mathbf{T})$  as a collection of disjoint and better spatially localized trees we define for each  $l \geq 0$  and  $m \in \mathbf{Z}$  the tree  $\mathbf{T}_{l,m}$  to include all tiles  $s \in G(\mathbf{T})$  satisfying the following requirement:

- $I_s \cap 2^l I_T \neq \emptyset$ , if  $m = 0$ .
- $I_s \cap (2^l I_T + 2^l m |I_T|) \neq \emptyset$  and  $I_s \cap (2^l I_T + 2^l(m - 1)|I_T|) = \emptyset$ , if  $m \geq 1$ .
- $I_s \cap (2^l I_T + 2^l m |I_T|) \neq \emptyset$  and  $I_s \cap (2^l I_T + 2^l(m + 1)|I_T|) = \emptyset$ , if  $m \leq -1$ .

We remark that since  $|I_s| \leq |I_T|$  for each  $s \in G(\mathbf{T})$ , for a fixed  $l \geq 0$ , each  $I_s$  can intersect at most two intervals  $2^l I_T + 2^l m |I_T|$  (and they must be adjacent). Obviously, for each  $l \geq 0$  the collection consisting of  $(\mathbf{T}_{l,m})_{m \in \mathbf{Z}}$  forms a partition of  $G(\mathbf{T})$  into trees. The top of  $\mathbf{T}_{l,m}$  is formally assigned to be the pair  $(I_{\mathbf{T}_{l,m}}, \xi_{\mathbf{T}})$ , where  $I_{\mathbf{T}_{l,m}}$  is the interval  $(2^l + 2)m |I_T|$ , while  $\xi_{\mathbf{T}}$  is the frequency component of the top  $(I_T, \xi_T)$  of the tree  $\mathbf{T}$ .

Denote by  $\mathcal{F}_{l,m}$  the collection of all the trees  $\mathbf{T}_{l,m}$ . Consider  $\sigma, \gamma > 0, \beta \geq 1, r > 2$ , and the complex numbers  $a_s, s \in \mathcal{P}$ .

**Theorem 5.1** *Let  $1 < q < 2$ . Assume we are in the settings from above and also that the following additional requirement is satisfied:*

$$\sup_{s \in \mathcal{P}} \frac{|a_s|}{|I_s|^{1/2}} \leq \sigma.$$

Define the exceptional sets

$$E^{(1)} := \bigcup_{l \geq 0} \left\{ x : \sum_{\mathbf{T} \in \mathcal{F}} 1_{2^l I_T}(x) > \beta 2^{2l} \right\},$$

$$E^{(2)} := \bigcup_{l, m \geq 0} \bigcup_{\mathbf{T} \in \mathcal{F}_{l,m}} \left\{ x : \left\| \sum_{\substack{s \in \mathbf{T} \\ |I_s|=2^j}} a_s \phi_{s, \mathbf{T}}^{(\alpha(l,m))}(x, \xi_{\mathbf{T}}) \right\|_{V_j^r(\mathbf{Z})} > \gamma 2^{-10l} (|m| + 1)^{-2} \right\},$$

where the  $\alpha(l, m)$  equals  $l$  if  $m \in \{-1, 0, 1\}$  and  $l + \lceil \log_2 |m| \rceil$  otherwise.

Then for each  $0 < \epsilon < 1$  (say) and for each  $x \notin E^{(1)} \cup E^{(2)}$  we have

$$(5.1) \quad \left\| \left( \sum_{\substack{s \in \mathcal{P} \\ |I_s|=2^k}} a_s \phi_s(x, \theta) \right)_{k \in \mathbf{Z}} \right\|_{M_{q,\theta}^*(\mathbf{R})} \lesssim \beta^{1/q-1/r+\epsilon} (\gamma + \sigma),$$

with the implicit constants depending only on  $r, \epsilon$ , and  $q$ .

**Proof** For each  $l \geq 0$  and each  $x \in \mathbf{R}$ , define inductively

$$\begin{aligned} \mathcal{F}_{0,x} &:= \{\mathbf{T} \in \mathcal{F}, x \in I_T\}, \\ \mathcal{F}_{l,x} &:= \{\mathbf{T} \in \mathcal{F}, x \in 2^l I_T \setminus 2^{l-1} I_T\}, \quad l \geq 1, \\ \mathcal{P}_{0,x} &:= \bigcup_{\mathbf{T} \in \mathcal{F}_{0,x}} G(\mathbf{T}) \\ \mathcal{P}_{l,x} &:= \bigcup_{\mathbf{T} \in \mathcal{F}_{l,x}} G(\mathbf{T}) \setminus \bigcup_{l' < l} \mathcal{P}_{l',x}, \quad l \geq 1, \\ \tilde{\mathcal{F}}_{l,x} &:= \{\mathbf{T} \in \mathcal{F}_{l,x} : G(\mathbf{T}) \setminus \bigcup_{l' < l} \mathcal{P}_{l',x} \neq \emptyset\}, \\ \Xi_{x,l} &:= \{c(\omega_T) : \mathbf{T} \in \tilde{\mathcal{F}}_{l,x}\}. \end{aligned}$$

Note that for each  $x \in \mathbf{R}$ ,  $\{\mathcal{P}_{l,x}\}_{l \geq 0}$  forms a partition of  $\mathcal{P}$ . Since  $x \notin E^{(1)}$ , it also follows that  $\#\Xi_{x,l} \leq \beta 2^{2l}$ .

Fix  $x \notin E^{(1)} \cup E^{(2)}$ . Next, we fix  $l$  and try to estimate

$$\left\| \left( \sum_{\substack{s \in \mathcal{P}_{l,x} \\ |I_s|=2^k}} a_s \phi_s(x, \theta) \right)_{k \in \mathbf{Z}} \right\|_{M_{q,\theta}^*(\mathbf{R})}.$$

Note that for each  $\lambda \in \Xi_{x,l}$ ,

$$(5.2) \quad \{s \in \mathcal{P}_{l,x} : \lambda \in \omega_s\} = G(\mathbf{T}') \cap \mathcal{P}_{l,x},$$

for some  $\mathbf{T}' \in \tilde{\mathcal{F}}_{l,x}$  (and, perhaps surprisingly, if  $\lambda = c(\omega_T)$ ,  $\mathbf{T}'$  is not necessarily the tree whose top tile is  $T$ ). Indeed, let  $\omega$  be the shortest frequency component of a tile  $s$  from  $\mathcal{P}_{l,x}$  such that  $\lambda \in \omega$ . In other words,  $\omega = \omega_s$ . This tile belongs to  $G(\mathbf{T}')$ , for some  $\mathbf{T}' \in \tilde{\mathcal{F}}_{l,x}$  (if there are more such  $\mathbf{T}'$ , select any of them). Note that its top tile  $T'$  must be in  $\mathcal{P}_{l,x}$  (otherwise, it must be that  $T' \in \bigcup_{l' < l} \mathcal{P}_{l',x}$ , hence  $T'$  was eliminated earlier, and thus the whole  $G(\mathbf{T})$  must have been eliminated at the same stage). Equation (5.2) is now immediate.

For each  $\mathbf{T} \in \tilde{\mathcal{F}}_{l,x}$  define  $\mathbf{T}_{l,m,x} := \mathbf{T}_{l,m} \cap \mathcal{P}_{l,x}$ , and note that  $(\mathbf{T}_{l,m,x})_m$  partition  $G(\mathbf{T}) \cap \mathcal{P}_{l,x}$ . An important observation is that for each  $k$ , the set  $\{s \in \mathbf{T}_{l,m,x} : |I_s| = 2^k\}$  either equals  $\{s \in \mathbf{T}_{l,m} : |I_s| = 2^k\}$ , or else it is empty. As a consequence,

$$(5.3) \quad \left\| \sum_{\substack{s \in \mathbf{T}_{l,m,x} \\ |I_s|=2^k}} a_s \phi_{s,\mathbf{T}}^{(\alpha(l,m))}(x, \xi_{\mathbf{T}}) \right\|_{V_k^r} \leq \left\| \sum_{\substack{s \in \mathbf{T}_{l,m} \\ |I_s|=2^k}} a_s \phi_{s,\mathbf{T}}^{(\alpha(l,m))}(x, \xi_{\mathbf{T}}) \right\|_{V_k^r}.$$

For each dyadic  $\omega$  denote by

$$m_\omega(\theta) := \sum_{s \in \mathcal{P}_{l,x} : \omega_s = \omega} a_s \phi_s(x, \theta).$$

The key observation is that if  $s \in \mathcal{P}_{l,x}$  and  $l \geq 1$ , then  $x \notin 2^{l-1}I_s$ , as can be easily checked. This together with property (2.4) easily implies that  $m_\omega$  is  $O(2^{-10l}\sigma)$ -adapted to  $\omega$ . Theorem 4.1 applied to  $\Lambda := \Xi_{l,x}$ , and (5.2) imply that

$$\begin{aligned} \left\| \left( \sum_{\substack{s \in \mathcal{P}_{l,x} \\ |I_s|=2^k}} a_s \phi_s(x, \theta) \right)_{k \in \mathbf{Z}} \right\|_{M_{q,\theta}^*(\mathbf{R})} \\ \lesssim 2^{4l} \beta^{1/q-1/r+\epsilon} \left( 2^{-10l} \sigma + \max_{\mathbf{T} \in \tilde{\mathcal{F}}_{l,x}} \left\| \sum_{\substack{s \in \mathcal{G}(\mathbf{T}) \cap \mathcal{P}_{l,x} \\ |I_s|=2^k}} a_s \phi_s(x, \xi_{\mathbf{T}} \right\|_{V_k^r} \right). \end{aligned}$$

It remains to show that for each  $\mathbf{T} \in \tilde{\mathcal{F}}_{l,x}$ ,

$$\left\| \sum_{\substack{s \in \mathcal{G}(\mathbf{T}) \cap \mathcal{P}_{l,x} \\ |I_s|=2^k}} a_s \phi_s(x, \xi_{\mathbf{T}} \right\|_{V_k^r} \lesssim 2^{-10l} \gamma.$$

Another key observation is that if  $s \in \mathbf{T}_{l,m,x}$  and  $l \geq 1$ , then  $x \notin 2^{\alpha(l,m)-1}I_s$ . It follows that for each  $l \geq 0$  and each  $m \in \mathbf{Z}$ ,  $\phi_s(x, \theta) = \phi_{s,\mathbf{T}}^{(\alpha(l,m))}(x, \theta)$ . Using this and (5.3) we get that

$$\begin{aligned} \left\| \sum_{\substack{s \in \mathcal{G}(\mathbf{T}) \cap \mathcal{P}_{l,x} \\ |I_s|=2^k}} a_s \phi_s(x, \xi_{\mathbf{T}} \right\|_{V_k^r} &\leq \sum_m \left\| \sum_{\substack{s \in \mathbf{T}_{l,m,x} \\ |I_s|=2^k}} a_s \phi_s(x, \xi_{\mathbf{T}} \right\|_{V_k^r} \\ &= \sum_m \left\| \sum_{\substack{s \in \mathbf{T}_{l,m,x} \\ |I_s|=2^k}} a_s \phi_{s,\mathbf{T}}^{(\alpha(m,l))}(x, \xi_{\mathbf{T}} \right\|_{V_k^r} \\ &\leq \sum_m \left\| \sum_{\substack{s \in \mathbf{T}_{l,m} \\ |I_s|=2^k}} a_s \phi_{s,\mathbf{T}}^{(\alpha(l,m))}(x, \xi_{\mathbf{T}} \right\|_{V_k^r}. \end{aligned}$$

Finally, since  $x \notin E^{(2)}$ , the last sum is  $O(2^{-10l}\gamma)$ , as desired. Now, (5.1) follows from the triangle inequality. ■

### 6 Proof of Theorem 2.2

For each collection of tiles  $\mathbf{S}' \subseteq \mathbf{S}$  define the following operator

$$V_{\mathbf{S}'} f(x) := \left\| \left( \sum_{\substack{s \in \mathbf{S}' \\ |I_s|=2^k}} \langle f, \varphi_s \rangle \phi_s(x, \theta) \right)_{k \in \mathbf{Z}} \right\|_{M_{q,\theta}^*(\mathbf{R})}.$$

Note that for each  $\mathbf{S}'$  the operator  $V_{\mathbf{S}'}$  is sublinear as a function of  $f$ . Also, for each  $f$  and  $x$  the mapping  $\mathbf{S}' \rightarrow V_{\mathbf{S}'} f(x)$  is sublinear as a function of the tile set  $\mathbf{S}'$ .

Let  $1 < p < 2$  be fixed such that  $\frac{1}{p} + \frac{1}{q} < \frac{3}{2}$ . We will prove in the following that for each  $\delta > 0$  and each  $0 < \lambda < 1$ ,

$$m\{x : V_{\mathbf{S}}1_F(x) \gtrsim \lambda^{1-\delta}\} \lesssim_{\delta,p,q} \frac{|F|}{\lambda^p}.$$

Since the range of  $p$  is open, this will immediately imply Theorem 2.2. Now fix  $\delta > 0$ . Let  $\epsilon > 0$  be sufficiently small, depending on  $\delta$ . Its value will not be specified, but it will be clear from the argument below that such an  $\epsilon$  exists. Define  $Q = \frac{1}{q} - \frac{1}{2} + \epsilon$ . We can arrange that  $Q < 1 - \frac{1}{p}$ , and define  $b := \frac{1-pQ}{1-2Q}$ . It will follow that  $0 < b < p$ . We can also arrange that  $\epsilon + (2 + \epsilon)Q < 1$ .

Define the first exceptional set  $E := \{x : M1_F(x) \geq \lambda^b\}$ . Note that

$$(6.1) \quad |E| \lesssim \frac{|F|}{\lambda^b}.$$

Split  $\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2$  where

$$\mathbf{S}_1 := \{s \in \mathbf{S} : I_s \cap E^c \neq \emptyset\}, \quad \mathbf{S}_2 := \{s \in \mathbf{S} : I_s \cap E^c = \emptyset\}.$$

We first argue that

$$(6.2) \quad m\{x \in \mathbf{R} : V_{\mathbf{S}_1}1_F(x) \gtrsim \lambda^{1-\delta}\} \lesssim \frac{|F|}{\lambda^p}.$$

Proposition 3.5 guarantees that  $\text{size}(\mathbf{S}_1) \lesssim \lambda^b$ , where the size is understood here with respect to the function  $1_F$ . Define  $\Delta := \lceil -\log_2(\text{size}(\mathbf{S}_1)) \rceil$ . Use the result of Proposition 3.6 to split  $\mathbf{S}_1$  as a disjoint union  $\mathbf{S}_1 = \bigcup_{n \geq \Delta} \mathcal{P}_n$ , where  $\text{size}(\mathcal{P}_n) \leq 2^{-n}$  and each  $\mathcal{P}_n$  consists of a family  $\mathcal{F}_{\mathcal{P}_n}$  of trees satisfying

$$(6.3) \quad \sum_{\mathbf{T} \in \mathcal{F}_{\mathcal{P}_n}} |I_{\mathbf{T}}| \lesssim 2^{2n}|F|.$$

For each  $n \geq \Delta$  define  $\sigma = \sigma_n := 2^{-n}$ ,  $\beta = \beta_n := 2^{(2+\epsilon)n}\lambda^p$ ,  $\gamma = \gamma_n := 2^{-n[(2+\epsilon)Q+\epsilon]}\lambda^{1-Qp-3\epsilon}$ . Define  $a_s := \langle 1_F, \varphi_s \rangle$  for each  $s \in \mathcal{P}_n$  and note that the collection  $\mathcal{P}_n$  together with the coefficients  $(a_s)_{s \in \mathcal{P}_n}$  satisfy the requirements of Theorem 5.1. Let  $\mathcal{F}_{\mathcal{P}_n,l,m}$  be the collection of all the trees  $\mathbf{T}_{l,m}$  obtained from all the trees  $\mathbf{T} \in \mathcal{F}_{\mathcal{P}_n}$  by the procedure described in the beginning of the previous section. Let  $r > 2$  be any number such that  $\frac{1}{q} - \frac{1}{r} < Q$ . Define the corresponding exceptional sets

$$E_n^{(1)} := \bigcup_{l \geq 0} \left\{ x : \sum_{\mathbf{T} \in \mathcal{F}_{\mathcal{P}_n}} 1_{2^l I_{\mathbf{T}}}(x) > \beta_n 2^{2l} \right\},$$

$$E_n^{(2)} := \bigcup_{l,m \geq 0} \bigcup_{\mathbf{T} \in \mathcal{F}_{\mathcal{P}_n,l,m}} \left\{ x : \left\| \sum_{\substack{s \in \mathbf{T} \\ |I_s| < 2^l}} a_s \phi_{s,\mathbf{T}}^{(\alpha(l,m))}(x, \xi_{\mathbf{T}}) \right\|_{V_j(\mathbf{Z})} > \gamma_n 2^{-10l} (|m| + 1)^{-2} \right\}.$$

By (6.3) we get  $|E_n^{(1)}| \lesssim 2^{-n\epsilon} \lambda^{-p} |F|$ . By Theorem 3.7, for each  $1 < s < \infty$  we get

$$|E_n^{(2)}| \lesssim \gamma_n^{-s} \sigma_n^{s-2} |F| \lesssim 2^{-n[-2-s(-1+(2+\epsilon)Q+\epsilon)]} \lambda^{-s(1-Qp-3\epsilon)} |F|.$$

Define  $E^* := \bigcup_{n \geq \Delta} (E_n^{(1)} \cup E_n^{(2)})$ . Trivial computations show that since  $\lambda \leq 1$  and  $2^{-\Delta} \lesssim \lambda^b$ , we have  $|E^*| \lesssim \lambda^{-p} |F|$  (work with a sufficiently large  $s$ , depending only on  $p, q, Q, \epsilon$ ).

For each  $x \notin E^*$ , Theorem 5.1 guarantees that

$$V_{S_1} 1_F(x) \leq \sum_{n \geq \Delta} V_{\mathcal{P}_n} 1_F(x) \lesssim \sum_{n \geq \Delta} \beta_n^Q (\gamma_n + \sigma_n).$$

The latter sum is easily seen to be  $O(\lambda^{1-\delta})$  if  $\epsilon$  is sufficiently small. This ends the proof of (6.2). We next prove that (and note that this is enough, due to (6.1))

$$m\{x \notin E : V_{S_2} 1_F(x) \gtrsim \lambda^{1-\delta}\} \lesssim \frac{|F|}{\lambda^p}.$$

To achieve this, we split  $S_2 := \bigcup_{\kappa > 0} S_{2,\kappa}$ , where

$$S_{2,\kappa} := \{s \in S_2 : 2^{\kappa-1} I_s \cap E^c = \emptyset, 2^\kappa I_s \cap E^c \neq \emptyset\},$$

and we prove that, uniformly over  $\kappa > 0$ ,

$$(6.4) \quad m\{x \notin E : V_{S_{2,\kappa}} 1_F(x) \gtrsim 2^{-\kappa} \lambda^{1-\delta}\} \lesssim 2^{-\kappa} \frac{|F|}{\lambda^p}.$$

Note further that if  $s \in S_{2,\kappa}$  then

$$\frac{|\langle 1_F, \varphi_s \rangle|}{|I_s|^{1/2}} \lesssim \inf_{x \in I_s} M 1_F(x) \lesssim 2^\kappa \inf_{x \in 2^\kappa I_s} M 1_F(x) \lesssim \lambda^b 2^\kappa,$$

and thus  $S_{2,\kappa}$  has size  $O(\lambda^b 2^\kappa)$ . Note also that  $S_{2,\kappa}$  remains convex. The proof of (6.4) now follows exactly the same way as the proof of (6.2). The fact that the size of  $S_{2,\kappa}$  is (potentially) greater than that of  $S_1$  by a factor of  $2^\kappa$  is compensated by the fact that for each  $x \notin E$  and each  $s \in S_{2,\kappa}$ ,  $x \notin 2^{\kappa-1} I_s$ . It follows that in the definition of the exceptional sets  $E_n^{(1)}$  and  $E_n^{(2)}$  for this case, we can restrict the union to  $l \geq \kappa - 1$ . We leave the details to the interested reader.

**Acknowledgements** This project is a continuation of the work in [9]. The author is indebted to M. Lacey, C. Thiele, and T. Tao for helpful conversations over the last few years. We are also indebted to R. Oberlin and C. Thiele for sharing a preprint of [17].

## References

- [1] I. Assani, Z. Buczolich, and R. D. Mauldin, *An  $L^1$  counting problem in ergodic theory* J. Anal. Math. **95**(2005), 221–241. <http://dx.doi.org/10.1007/BF02791503>
- [2] I. Assani, and Z. Buczolich, *The  $(L^1, L^1)$  bilinear Hardy-Littlewood function and Furstenberg average*. Rev. Mat. Iberoam. **26**(2010), no. 3, 861–890.
- [3] G. D. Birkhoff, *Proof of the ergodic theorem*. Proc. Nat. Acad. Sci. U.S.A. **17**(1931), 656–660.
- [4] J. Bourgain, *Temps de retour pour les systèmes dynamiques*. C. R. Acad. Sci. Paris Sér. I. Math. **306**, (1988), no. 12, 483–485.
- [5] ———, *Pointwise ergodic theorems for arithmetic sets*. Inst. Hautes Études Sci. Publ. Math. **69**(1989), 5–45.
- [6] J. Bourgain, H. Furstenberg, Y. Katznelson, and D. Ornstein, *Return times of dynamical systems*. (Appendix to [5]). Inst. Hautes Études Sci. Publ. Math. **69**(1989), 47–50.
- [7] L. Carleson, *On convergence and growth of partial sums of Fourier series*. Acta Math. **116**(1966), 135–157. <http://dx.doi.org/10.1007/BF02392815>
- [8] C. Demeter, *On some maximal multipliers in  $L^p$* . Rev. Mat. Iberoam. **26**(2010), no. 3, 947–964.
- [9] C. Demeter, M. Lacey, T. Tao, and C. Thiele, *Breaking the duality in the return times theorem*. Duke Math. J. **143**(2008), no. 2, 281–355 <http://dx.doi.org/10.1215/00127094-2008-020>
- [10] P. Halmos, *Lectures on Ergodic Theory*, Chelsea Publishing, New York, 1956.
- [11] M. Lacey, *The bilinear maximal functions map into  $L^p$  for  $2/3 < p \leq 1$* . Ann. of Math. **151**(2000), no. 1, 35–57. <http://dx.doi.org/10.2307/121111>
- [12] M. Lacey and C. Thiele,  *$L^p$  estimates on the bilinear Hilbert transform for  $2 < p < \infty$* . Ann. of Math. **146**(1997), no. 3, 693–724. <http://dx.doi.org/10.2307/2952458>
- [13] ———, *On Calderón’s conjecture*. Ann. of Math. **149**(1999), no. 2, 475–496. <http://dx.doi.org/10.2307/120971>
- [14] ———, *A proof of boundedness of the Carleson operator*. Math. Res. Lett. **7**(2000), no. 4, 361–370.
- [15] C. Muscalu, T. Tao, and C. Thiele, *Uniform estimates on multi-linear operators with modulation symmetry*. J. Anal. Math. **88**(2002), 255–307. <http://dx.doi.org/10.1007/BF02786579>
- [16] ———, *Multi-linear operators given by singular multipliers*. J. Amer. Math. Soc. **15**(2002), no. 2, 469–496. <http://dx.doi.org/10.1090/S0894-0347-01-00379-4>
- [17] F. Nazarov, R. Oberlin, and C. Thiele, *A Calderon Zygmund decomposition for multiple frequencies and an application to an extension of a lemma of Bourgain*. Math. Res. Lett. **17**(2010), no. 3, 529–545.
- [18] D. Rudolph, *A joinings proof of Bourgain’s return time theorem*. Ergodic Theory Dynam. Systems **14**(1994), no. 1, 197–203.
- [19] C. Thiele, *Wave packet Analysis*. CBMS Regional Conference Series in Mathematics 105. American Mathematical Society, Providence, RI, 2006.

Department of Mathematics, Indiana University, Bloomington, IN 47405, USA  
 e-mail: demeterc@indiana.edu