

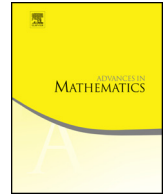


ELSEVIER

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim



Sparse domination and the strong maximal function



Alex Barron^a, José M. Conde-Alonso^a, Yumeng Ou^b,
Guillermo Rey^{c,*}

^a Department of Mathematics, Brown University, 151 Kassar House, Providence, RI, USA

^b Department of Mathematics, Box B6-230, One Bernard Baruch Way, New York, NY, USA

^c 127 Vincent Hall, 206 Church St. SE, Minneapolis, MN, USA

ARTICLE INFO

Article history:

Received 27 November 2018

Accepted 18 December 2018

Available online xxxx

Communicated by C. Fefferman

Keywords:

Biparameter

Multiparameter

Sparse domination

Maximal functions

ABSTRACT

We study the problem of dominating the dyadic strong maximal function by $(1, 1)$ -type sparse forms based on rectangles with sides parallel to the axes, and show that such domination is impossible. Our proof relies on an explicit construction of a pair of maximally separated point sets with respect to an appropriately defined notion of distance.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction

Recent years have seen a great deal of work around the concept of *sparse domination*. Perhaps the easiest domination result is the one for the dyadic maximal function

$$M_{\mathcal{D}}f(x) = \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |f|,$$

* Corresponding author.

E-mail addresses: abarron@math.brown.edu (A. Barron), jconde@math.brown.edu (J.M. Conde-Alonso), yumeng.ou@baruch.cuny.edu (Y. Ou), reyg@umn.edu (G. Rey).

where the supremum is taken over the family $\mathcal{D}$ of all dyadic cubes in $\mathbb{R}^d$. One can use the Calderón–Zygmund decomposition at varying heights to obtain (for each non-negative locally integrable function f) a collection $\mathcal{S}$ of dyadic cubes from the same filtration $\mathcal{D}$ such that

$$M_{\mathcal{D}}f \leq C \sum_{Q \in \mathcal{S}} \langle f \rangle_Q \mathbb{1}_Q,$$

where $\langle f \rangle_Q = \frac{1}{|Q|} \int_Q f$, and $\mathcal{S}$ is *sparse*:

Definition 1. A collection of sets $\mathcal{S}$ is η -sparse if for all $Q \in \mathcal{S}$ there exists a subset $E(Q) \subseteq Q$ such that $|E(Q)| \geq \eta|Q|$ and the collection $\{E(Q)\}$ is pairwise disjoint.¹

It is worth mentioning that this notion was already used by Sawyer [21] and by Christ–Fefferman [3] to prove certain weighted estimates, although the language that they used was different. Moreover, the result readily extends to more general martingale filtrations and a similar estimate allows one to sparsely dominate martingale transforms as well (see [15]).

The recent interest in sparse domination results was sparked by Lerner (see for example [18–20]) in connection with the famous A_2 conjecture as it gave a particularly simple proof of the – now – theorem (first settled by Hytönen in [13] after numerous efforts by the mathematical community). Lerner’s original result was a *dual* form of sparse domination:

$$\|Tf\|_{\mathbb{X}} \lesssim \sup_{\mathcal{S}} \left\| \sum_{Q \in \mathcal{S}} \langle |f| \rangle_Q \mathbb{1}_Q \right\|_{\mathbb{X}}, \quad (1.1)$$

where the supremum is taken over all sparse collections (with respect to a finite number of dyadic grids), and $\mathbb{X}$ is any *Banach function space*.

The number of sparse domination results is by now very large and it would be impossible to give a detailed survey of all of them in a reasonable amount of text. But let us note that (1.1) has been generalized in many ways, and the typical result is that for every pair of (sufficiently nice) functions f and g there exists a sparse collection of cubes $\mathcal{S}$ such that

$$|\langle Tf, g \rangle| \lesssim \sum_{Q \in \mathcal{S}} \langle |f| \rangle_Q \langle |g| \rangle_Q |Q| \quad (1.2)$$

or more generally

$$|\langle Tf, g \rangle| \lesssim \sum_{Q \in \mathcal{S}} \langle |f| \rangle_{Q,r} \langle |g| \rangle_{Q,s} |Q|, \quad (1.3)$$

¹ In what follows, we will usually omit the parameter η and just say “sparse” instead of “ η -sparse”.

where $\langle |f| \rangle_{Q,r} = \langle |f|^r \rangle_Q^{\frac{1}{r}}$ and $1 \leq r, s < \infty$. The (r, s) sparse forms of type (1.3) arise when studying operators that fall outside the scope of classical Calderón–Zygmund theory, and which may not satisfy the full range of strong and weak-type L^p estimates implied by (1.2). The operators that have been studied include but are not limited to rough singular integral operators, Bochner–Riesz multipliers, spherical maximal functions, singular integrals along manifolds, pseudodifferential operators, and the bilinear Hilbert transform. See for example [1,5,15,16,2,17,4,8,7]. Sparse domination provides a fine quantification of the mapping properties of operators that carries much more information than L^p bounds. In particular, it is a very effective way to obtain sharp quantitative weighted estimates and has also been used to study previously unknown endpoint behaviors (see for instance [14]).

In the present work we study the problem of sparsely dominating the bi-parameter analogue of the dyadic maximal function. This operator is the dyadic strong maximal function:

$$\mathcal{M}_S f(x) := \sup_{R \ni x} \langle |f| \rangle_R, \quad (1.4)$$

where the supremum is taken over all dyadic rectangles containing x with sides parallel to the axes. The strong maximal function is one of the most important operators in the theory of multi-parameter singular integrals, associated with which is an underlying non-isotropic dilation structure. This class of operators arise naturally in the theory of summation of multiple Fourier series, several complex variables and certain boundary value problems, and is a first step into the study of operators with more complicated dilation structure (e.g. Zygmund dilations). However, much less is known about these operators compared to the one-parameter case, for example the sharp weighted bound, thus it is natural to ask whether the sparse domination technique can be introduced into the multi-parameter setting to help the study. Unfortunately, this seems to be very difficulty due to the fact that one of the key ingredients in standard proofs of sparse domination, the stopping time argument, is missing in the multi-parameter setting.

We first note that in the study of multi-parameter sparse domination, the natural geometric objects to consider are axes-parallel rectangles instead of cubes, since $\mathcal{M}_S$ can be large on axis-parallel rectangles of arbitrary eccentricity. Then, one observes that the strong maximal function is indeed dominated by sparse forms (based on rectangles) when restricted to a single point mass. Indeed, one can make sense of (1.4) when applied to finite positive measures, and then it is easy to see that

$$\mathcal{M}_S(\delta_0)(x, y) \leq \frac{1}{|x||y|}. \quad (1.5)$$

And this function can actually be dominated by a sparse operator: if we define $I_m = [0, 2^m)$ and $J_m = [2^{m-1}, 2^m)$ then

$$\begin{aligned}
\frac{1}{|x||y|} &= \sum_{m,n \in \mathbb{Z}} \frac{1}{|x||y|} \mathbb{1}_{J_m \times J_n}(x, y) \\
&\leq \sum_{m,n \in \mathbb{Z}} 2^{2-m-n} \mathbb{1}_{J_m \times J_n}(x, y) \\
&= 4 \sum_{m,n \in \mathbb{Z}} \langle \delta_0 \rangle_{I_m \times I_n} \mathbb{1}_{I_m \times I_n}(x, y).
\end{aligned}$$

Note that the collection $\mathcal{S} = \{I_m \times I_n : m, n \in \mathbb{Z}\}$ is sparse since for every $R = I_m \times I_n$ we can define $E(R) = J_m \times J_n$ and this satisfies the conditions of Definition 1 with $\eta = 1/4$.

This example is relevant because (approximate) point masses are extremal examples for the weak-type behavior of $\mathcal{M}_S$ and the Hardy–Littlewood maximal functions near the L^1 endpoint. Moreover, we know from the one-parameter theory that there is a connection between sparse bounds and weak-type endpoint estimates; for example, a $(1, 1)$ sparse bound of type (1.2) implies that T maps L^1 into weak L^1 (see the appendix in [4]). Taking the above discussion into account, it is natural to ask whether or not $\mathcal{M}_S$ admits a $(1, 1)$ sparse bound of the type

$$|\langle \mathcal{M}_S f, g \rangle| \lesssim \sum_{R \in \mathcal{S}} \langle |f| \rangle_R \langle |g| \rangle_R |R|, \quad (1.6)$$

where the collection $\mathcal{S}$ consists of dyadic axis-parallel rectangles. We also note that there is no immediate contradiction implied by a bound of type (1.6); indeed, an estimate of the type

$$|\langle T f, g \rangle| \lesssim \sum_{R \in \mathcal{S}} \langle |f| \rangle_R \langle |g| \rangle_R |R|$$

implies that T maps $L \log L$ into weak L^1 , but does *not* imply that T maps L^1 into weak L^1 (the proof is similar to the argument in the appendix of [4]).

Our main result answers the question raised above in the negative: there can be no domination by positive sparse forms of the type (1.2) for the strong maximal function.

Theorem A. *For every $C > 0$ and $0 < \eta < 1$ there exist a pair of compactly supported integrable functions f and g such that*

$$|\langle \mathcal{M}_S f, g \rangle| \geq C \sum_{R \in \mathcal{S}} \langle |f| \rangle_R \langle |g| \rangle_R |R|,$$

for all η -sparse collections $\mathcal{S}$ of dyadic rectangles with sides parallel to the axes.

The proof of Theorem A is based on the construction of pairs of extremal functions for which the sparse bound cannot hold. These extremal examples take advantage of

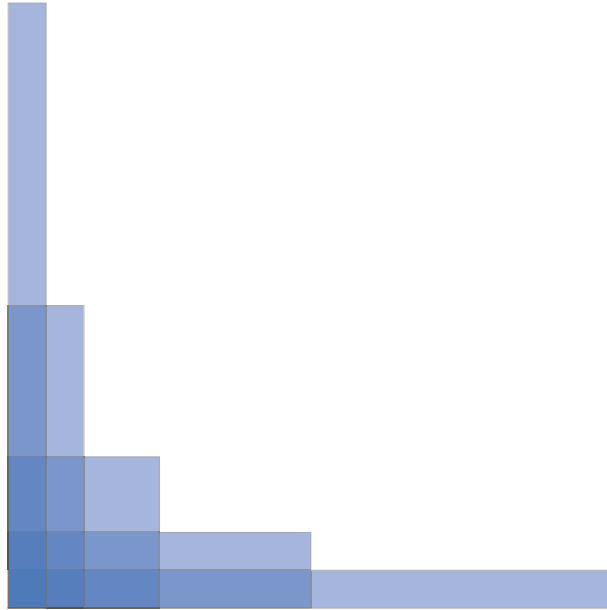


Fig. 1. Dyadic stairs.

the behavior of $\mathcal{M}_S$ when applied to sums of several point masses. Indeed, when we apply $\mathcal{M}_S$ to a point mass we get level sets that look like dyadic stairs, as displayed in Fig. 1. These stairs are sparse, in fact the measure of their union is proportional to the number of rectangles that form them. We shall describe a way to place many point masses sufficiently far from each other in such a way that these stairs are all necessary to dominate $\mathcal{M}_S$, but are packed too tightly to be sparse.

In particular, we shall construct a set of points $\mathcal{P}$ which are maximally separated with respect to a quantity that captures the biparametric structure of the problem. Our first extremal function is then the sum of point masses at the points in $\mathcal{P}$. As we said above, this will guarantee that $\mathcal{M}_S$ applied to it is uniformly large on the unit square. We complement the construction of the set $\mathcal{P}$ with another set, $\mathcal{Z}$, with the following structure: for each point $p \in \mathcal{P}$ there exist a large amount of points $z \in \mathcal{Z}$ that are both near p and far away from all the other points in $\mathcal{P}$. Moreover, the family of points in $\mathcal{Z}$ associated with a given $p \in \mathcal{P}$ are not clustered together, but are instead spread out over the boundary of a hyperbolic ball centered at p . Our second extremal function is then the normalized sum of point masses at the points in $\mathcal{Z}$. We will show that the placement of the points in $\mathcal{Z}$ implies that if a form given by a family $\mathcal{S}$ dominates the pair $\langle \mathcal{M}_S f, g \rangle$ then the rectangles in $\mathcal{S}$ need to cover the portion of the stair centered at p that contains each z , for all $z \in \mathcal{Z}$. But then the resulting collection of rectangles cannot be sparse.

We remark that our result actually extends to the situation in which we take slightly larger averaged norms of f . In particular our methods allow us to push Theorem A to

the case of Orlicz $\phi(L)$ -norms with $\phi(x) = x \log(x)^\alpha$ for $\alpha < 1/2$. We carry out this extension in Theorem 4.6.

We have learned that our construction of $\mathcal{P}$ is closely connected to the fields of discrepancy, combinatorics, and computational learning. In particular it is a special construction of a *low-discrepancy* sequence similar to those of Hammersley, see for example Definition 3.44 of [9]. These low-discrepancy sequences are also closely connected to the notion of ε -nets, though our point of view is in a sense opposite to the one usually taken in the theory of ε -nets: we are interested in *lower* bounds on the cardinality of the intersection of these sets with rectangles. The notion of (t, m, s) -nets is closer to our purposes, see [9]. Our results show that these kinds of sets are particularly well-suited to the study of the strong maximal function. However, the theory of discrepancy and ε -nets is much more general so it would be very interesting to find further connections between other types of maximal operators and the fields above. In particular Zygmund-type maximal operators (defined like the strong maximal function, but with rectangles with fewer “parameters” than the ambient dimension, see [6,22,11,23]) could be amenable to this connection.

The structure of the rest of the paper is the following. In Section 2, we prove lower bounds for $\mathcal{M}_S$ and, in the opposite direction, we prove an upper bound for sparse forms when acting on the same extremal functions. We assume the existence of the key sets of points $\mathcal{P}$ and $\mathcal{Z}$, which are needed to construct the aforementioned extremal functions. The sets $\mathcal{P}$ and $\mathcal{Z}$ are constructed in detail in Section 3, completing the proof of Theorem A. Finally, in Section 4 we show how our results extend to higher dimensions by a tensor product argument. We also show that the strong maximal function is in a sense supercritical for L^1 -sparse forms, which allows us to slightly strengthen our result.

Acknowledgments

The authors would like to thank Jill Pipher and Dmitriy Bilyk for useful discussions regarding the content of this paper. The third named author is supported by NSF-DMS #1764454.

2. Lower bounds for $\mathcal{M}_S$ and upper bounds for sparse forms

We start by introducing a quantity which is intimately related with the geometry of the bi-parameter setting: for every pair of points p, q in $[0, 1]^2$ we set

$$\text{dist}_{\mathbb{H}}(p, q) = \inf\{|R|^{1/2} : R \in \mathcal{D} \times \mathcal{D} \text{ is a dyadic rectangle containing } p \text{ and } q\}.$$

One can also define a distance between two (dyadic) rectangles:

$$\text{dist}_{\mathbb{H}}(R_1, R_2) := \inf_{p \in R_1, q \in R_2} \text{dist}_{\mathbb{H}}(p, q).$$

Note that $\text{dist}_{\mathbb{H}}$ is not a distance function, and not even a quasidistance, as the triangle inequality is completely false in general. We will, however, refer to it as *the distance*

between two points. Using this language, the strong maximal function of a point mass (1.5) can now be written as

$$\mathcal{M}_S(\delta_0)(p) = \frac{1}{\text{dist}_{\mathbb{H}}(p, 0)^2}.$$

We are going to study the precise behavior of $\mathcal{M}_S$ when applied to uniform probability measures concentrated on certain point sets: given a finite set of points $\mathcal{F}$ let

$$\mu = \mu_{\mathcal{F}} = \frac{1}{\#(\mathcal{F})} \sum_{p \in \mathcal{F}} \delta_p. \quad (2.1)$$

In general evaluating the strong maximal function of a sum of point masses is a computationally difficult problem: for every point p one has to consider all possible subsets of $\mathcal{F}$ whose minimal enclosing rectangle contains p . However, if we know a priori that $\mathcal{F}$ is ε -separated (with respect to $\text{dist}_{\mathbb{H}}$), and that it contains sufficiently many points, then this forces a certain uniform distribution of μ at large scales, reducing the problem to treating only small rectangles. In particular, let $\varepsilon > 0$, suppose that $\mathcal{F} = \mathcal{P}$ is such that $\#(\mathcal{P}) \geq C\varepsilon^{-2}$, and that it is ε -separated:

$$\text{dist}_{\mathbb{H}}(p, q) \geq \varepsilon \quad \forall p \neq q \text{ in } \mathcal{P}.$$

Then for any rectangle with $|R| \geq \varepsilon^2$ we have

$$\langle \mu \rangle_R = \frac{\mu(R)}{|R|} \leq \frac{\#(R \cap \mathcal{P})}{C|R|\varepsilon^{-2}} \lesssim \frac{|R|\varepsilon^{-2}}{C|R|\varepsilon^{-2}} = \frac{1}{C}.$$

If instead $|R| < \varepsilon^2$ then R can contain at most one point from $\mathcal{P}$ and so for any p for which $\text{dist}_{\mathbb{H}}(p, \mathcal{P}) \lesssim 1$ we have

$$M_S(\mu)(p) \sim \frac{1}{\#(\mathcal{P})} \sup_{q \in \mathcal{P}} \frac{1}{\text{dist}_{\mathbb{H}}(p, q)^2}. \quad (2.2)$$

As we mentioned in the introduction, if one can find another set of points $\mathcal{Z}$ each of which is very close to exactly one point p from $\mathcal{P}$ —and which are not very close to one another, so that they do not form large clusters—, then we can force a contradiction that makes sparse domination fail. The rest of this section is devoted to giving lower bounds for $\mathcal{M}_S$ assuming the existence of such sets of points $\mathcal{P}$ and $\mathcal{Z}$. In Section 2.2 we exploit the intuition given above and use it to give upper bounds for sparse forms that, together with the lower bounds from this section, will ultimately yield a contradiction.

The following two theorems give the full description of $\mathcal{P}$ and $\mathcal{Z}$. We postpone their proofs to Section 3.

Theorem 2.1. For every $m \geq 0$ there exists a collection of points $\mathcal{P} \subset [0, 1]^2$ such that

$$\text{dist}_{\mathbb{H}}(p, q) \geq 2^{-m}, \quad \forall p \neq q \text{ in } \mathcal{P}, \quad (\text{P.1})$$

and

$$\#(\mathcal{P}) = 2^{2m+1}. \quad (\text{P.2})$$

Theorem 2.2. For every $k \ll m$ and every 2^{-m} -separated set $\mathcal{P} \subset [0, 1]^2$ there exists a set $\mathcal{Z} \subset [0, 1]^2$ satisfying the following properties:

$$(\text{Z.1}) \quad \#(\mathcal{Z}) \gtrsim m2^{2m}.$$

(Z.2) For every $z \in \mathcal{Z}$ there exists exactly one point $p(z) \in \mathcal{P}$ such that

$$\text{dist}_{\mathbb{H}}(p(z), z)^2 < 2^{-2m-1}.$$

(Z.3) For every $z \in \mathcal{Z}$,

$$\text{dist}_{\mathbb{H}}(p(z), z)^2 \sim 2^{-2m-k}.$$

(Z.4) For every dyadic rectangle R with $|R| \leq 2^{-2m-2}$ that intersects $\mathcal{P}$ we have

$$\#(R \cap \mathcal{Z}) \lesssim k.$$

(Z.5) For every dyadic rectangle R with $|R| \geq 2^{-2m-1}$ we have

$$\#(R \cap \mathcal{Z}) \lesssim 2^{2m}mk|R|.$$

The implied constants are independent of k and m .

Remark 2.3. A simple pigeonholing argument shows that if $\mathcal{P}$ is the collection from Theorem 2.1 and R is a dyadic rectangle with $|R| \geq 2^{-2m-1}$, then $\#(R \cap \mathcal{P}) = 2^{2m+1}|R|$.

Remark 2.4. Above, we deliberately omitted the precise dependence of m on k . It turns out that any $m \geq k + C$ is admissible as one can check from the proofs in Section 3. This justifies our choices of m and k in the next subsections.

2.1. Lower bound for $\mathcal{M}_S$

For the rest of this section $k \ll m$ will be some fixed large numbers, and $(\mathcal{P}, \mathcal{Z})$ will be the sets given by Theorems 2.1 and 2.2. Also, $\mu = \mu_{\mathcal{P}}$ and $\nu = \nu_{\mathcal{Z}}$ will always denote the associated uniform probability measures introduced in (2.1). The proofs of Theorems 2.1 and 2.2 in the next section show that μ and ν can be arbitrarily well-approximated by

L^1 functions, in particular because we will be able to choose the points from $\mathcal{P}$ and $\mathcal{Z}$ in small cubes. Because of this we will prove Theorem A with f and g replaced by the measures μ and ν , respectively, and the full result can then be recovered by a simple limiting argument.

Proposition 2.5. *Under these conditions we have*

$$\langle \mathcal{M}_S(\mu), \nu \rangle \gtrsim 2^k. \quad (2.3)$$

Proof. The positivity of $\mathcal{M}_S$ makes the proof almost trivial since we do not need to care about possible interactions among the points in $\mathcal{P}$. In particular for all $z \in \mathcal{Z}$ let $p(z)$ be the point guaranteed by (Z.2) of Theorem 2.2. Then, by (2.2) and (Z.3),

$$\mathcal{M}_S(\mu)(z) \geq \frac{1}{\#(\mathcal{P}) \operatorname{dist}_{\mathbb{H}}(p(z), z)^2} \sim \frac{2^{2m+k}}{\#(\mathcal{P})}.$$

Now by (P.2) of Theorem 2.1,

$$M_S(\mu)(z) \gtrsim 2^k$$

and the claim follows from the fact that ν is a uniform probability measure over the set $\mathcal{Z}$. $\square$

2.2. Upper bound for sparse forms

We now prove upper bounds for sparse forms when acting on μ and ν . Our results here are written in the language of Carleson sequences:

Definition. Let α be a non-negative function defined on all rectangles that is zero for all but a finite collection of rectangles. We say that α is Λ -Carleson if for all open sets Ω we have

$$\sum_{R \subseteq \Omega} \alpha_R |R| \leq \Lambda |\Omega|,$$

where the sum is taken over all rectangles contained in Ω .

We call a collection of dyadic rectangles $\mathcal{S}$ a Λ -Carleson collection if the sequence

$$\alpha_R = \begin{cases} 1 & \text{if } R \in \mathcal{S}, \\ 0 & \text{otherwise} \end{cases}$$

is a Λ -Carleson sequence. The notions of sparse and Carleson collections are equivalent: a Λ -Carleson collection is Λ^{-1} -sparse and vice versa, as was shown in [12] (see also [10]).

Proposition 2.6. *For every Λ -Carleson collection $\mathcal{S}$ and all k and m we have*

$$\sum_{R \in \mathcal{S}} \langle \mu \rangle_R \langle \nu \rangle_R |R| \lesssim \Lambda k \left(1 + \frac{2^{2k}}{m}\right). \quad (2.4)$$

Proof. For any rectangle R let $\overline{R}$ be the intersection of R with the unit square $[0, 1]^2$. We will show that

$$\langle \mu \rangle_R \langle \nu \rangle_R \lesssim k \left(1 + \frac{2^{2k}}{m}\right) \left(\frac{|\overline{R}|}{|R|}\right)^2. \quad (2.5)$$

Assume first that R is *large*, i.e.: $|\overline{R}| \geq 2^{-2m-1}$. Then by the 2^{-m} -separation of $\mathcal{P}$

$$\langle \mu \rangle_R = \frac{1}{2^{2m+1}} \frac{\#(\mathcal{P} \cap \overline{R})}{|R|} = \frac{|\overline{R}|}{|R|}.$$

Similarly, for ν we have by (Z.5):

$$\langle \nu \rangle_R \lesssim k \frac{|\overline{R}|}{|R|}.$$

Therefore, for large rectangles we have

$$\langle \mu \rangle_R \langle \nu \rangle_R \lesssim k \left(\frac{|\overline{R}|}{|R|}\right)^2.$$

If instead R is *small*, i.e.: $|\overline{R}| < 2^{-2m-1}$, then we have to be more careful. If

$$\langle \mu \rangle_R \langle \nu \rangle_R \neq 0$$

then $\overline{R}$ must contain at least one point $p \in \mathcal{P}$. In fact, since $\overline{R}$ is sufficiently small, there exists exactly one $p \in \mathcal{P} \cap \overline{R}$. Hence

$$\langle \mu \rangle_{\overline{R}} \lesssim \frac{1}{2^{2m} |\overline{R}|}.$$

Since $\langle \nu \rangle_R \neq 0$ the rectangle $\overline{R}$ must contain at least one point from $\mathcal{Z}$. This implies a lower bound on the size of $\overline{R}$. To see this observe that if z is any point in $R \cap \mathcal{Z}$, then by the smallness of R and (Z.2) we must have $p = p(z)$, and hence $\text{dist}_{\mathbb{H}}(p, z) \sim 2^{-2m-k}$ by (Z.3), therefore

$$|\overline{R}| \gtrsim 2^{-2m-k}.$$

Now, by (Z.4) we have

$$\#(\overline{R} \cap \mathcal{Z}) \lesssim k,$$

so

$$\langle \mu \rangle_{\overline{R}} \lesssim 2^k \quad \text{and} \quad \langle \nu \rangle_{\overline{R}} \lesssim \frac{k2^k}{m},$$

from which inequality (2.5) follows.

Now we can finish the proof by splitting $\mathcal{S}$ into families $\mathcal{S}_j$ as follows:

$$\mathcal{S}_j = \{R \in \mathcal{S} : 2^{-j-1}|R| \leq |\overline{R}| < 2^{-j}|R|\}.$$

Note that the rectangles in $\mathcal{S}_j$ are contained in

$$\Omega_j = \{p \in \mathbb{R}^2 : M_S(\mathbb{1}_{[0,1)^2}) \gtrsim 2^{-j}\},$$

which, by the weak-type boundedness of the strong maximal function, satisfies

$$|\Omega_j| \lesssim j2^j. \quad (2.6)$$

Then by estimate (2.5), the Carleson condition, and (2.6),

$$\begin{aligned} \sum_{R \in \mathcal{S}} \langle \mu \rangle_R \langle \nu \rangle_R |R| &= \sum_{j=0}^{\infty} \sum_{R \in \mathcal{S}_j} \langle \mu \rangle_R \langle \nu \rangle_R |R| \\ &\lesssim k \left(1 + \frac{2^{2k}}{m}\right) \sum_{j=0}^{\infty} 2^{-2j} \sum_{R \in \mathcal{S}_j} |R| \\ &\leq k \left(1 + \frac{2^{2k}}{m}\right) \sum_{j=0}^{\infty} 2^{-2j} \Lambda |\Omega_j| \\ &\lesssim \Lambda k \left(1 + \frac{2^{2k}}{m}\right) \sum_{j=0}^{\infty} j 2^{-j} \\ &\lesssim \Lambda k \left(1 + \frac{2^{2k}}{m}\right). \quad \square \end{aligned}$$

Remark 2.7. An examination of the proof above shows that we do not use the full power of the Carleson condition. In particular, the collection $\mathcal{S}$ can be assumed to satisfy the Λ -Carleson packing condition with respect to the unit cube and the level sets Ω_j , but it does not need to be Λ -Carleson with respect to any other open sets (and in particular need not be Λ -Carleson at small scales).

We can now formally conclude the proof of Theorem A conditionally on Theorems 2.1 and 2.2.

Proof of Theorem A. Choose $m = k2^{2k}$. Proposition 2.5 implies that

$$\langle M_S(\mu), \nu \rangle \gtrsim 2^k,$$

so Proposition 2.6 forces $\Lambda \rightarrow \infty$ if we make $k \rightarrow \infty$, which leads to a contradiction. $\square$

3. Construction of the extremal sets

In this section, we will sometimes need to work with the projections of a rectangle R onto the axes. To that end, if $R = I \times J$ we denote

$$\pi_1(R) := I, \quad \pi_2(R) := J.$$

For dyadic intervals I we let $\widehat{I}$ denote the dyadic parent of I . We also let $I^{(j)}$ denote the j -fold dyadic dilation of I , so that $I^{(j)}$ is the dyadic interval containing I with $|I| = 2^j |I|$.

3.1. Construction of $\mathcal{P}$ and the proof of Theorem 2.1

The following observation will be useful in the construction:

Lemma 3.1. *Let $R_1 = I_1 \times J_1$, $R_2 = I_2 \times J_2$ be two dyadic rectangles such that $I_1 \cap I_2 = J_1 \cap J_2 = \emptyset$. Then*

$$\text{dist}_{\mathbb{H}}(p_1, p_2) = \text{dist}_{\mathbb{H}}(q_1, q_2), \quad \forall p_1, q_1 \in R_1, p_2, q_2 \in R_2.$$

Proof. Given any $p_1 \in R_1$, $p_2 \in R_2$, it suffices to show that any dyadic rectangle $R = I \times J$ containing both p_1 and p_2 needs to contain R_1 and R_2 . Since $I \cap I_i \neq \emptyset$, $i = 1, 2$, and $I_1 \cap I_2 = \emptyset$, one has $I \supset I_i$, $i = 1, 2$. The same holds for J : $J \supset J_i$ for $i = 1, 2$. $\square$

Proof of Theorem 2.1. We are going to show the following claim by induction: for every non-negative integer m there exist 2^{2m+1} dyadic squares $Q_1^m, \dots, Q_{2^{2m+1}}^m$ in $[0, 1)^2$ satisfying

- (1) $\ell(Q_i^m) = 2^{-2m-1}$ for all i ,
- (2) $\text{dist}_{\mathbb{H}}(Q_i^m, Q_j^m) \geq 2^{-m}$ for all $i \neq j$.

Assuming the claim, it is enough to take

$$\mathcal{P} = \{p_Q : Q \in \{Q_1^m, \dots, Q_{2^{2m+1}}^m\}\},$$

where p_Q denotes the center of the cube Q —in fact, any point of Q would work. Then, by Lemma 3.1 we immediately get (P.1) and (P.2) and we end the proof.

We turn to the proof of the claim. The case $m = 0$ is easy: it suffices to take $Q_1^0 = [0, 1/2)^2$ and $Q_2^0 = [1/2, 1)^2$. Assume now by induction that the theorem is true for $m - 1$. Scale the family obtained in that step by $1/2$ in both dimensions and place a translated copy of the result in each square $Q \in \mathcal{D}_1([0, 1)^2)$. Denote the cubes so constructed by $P_1^Q, \dots, P_{2^{2m-1}}^Q$. By the dilation invariance of $\text{dist}_{\mathbb{H}}$ we have

$$\ell(P_i^Q) = 2^{-1} \cdot 2^{-2(m-1)-1} = 2^{-2m}.$$

Therefore, it is enough to choose a first-generation child from each of these squares $\{P_i^Q\}_{i,Q}$ in such a way that (2) is satisfied. Write $\mathcal{D}_1([0, 1)^2) = \{Q_0^0, Q_0^1, Q_1^0, Q_1^1\}$, where the children are listed in the order of upper left, upper right, lower left, and lower right. We first consider Q_0^0 and Q_1^1 . For each P_i^Q with $Q \in \{Q_0^0, Q_1^1\}$ we choose an arbitrary first-generation child and call it Q_i^Q . By induction we have

$$\text{dist}_{\mathbb{H}}(Q_i^Q, Q_j^Q) \geq \text{dist}_{\mathbb{H}}(P_i^Q, P_j^Q) \geq 2^{-1} \cdot 2^{-(m-1)} = 2^{-m} \quad \text{for all } i \neq j,$$

while the distance between any square in Q_0^0 and any other square in Q_1^1 must be exactly 1 according to Lemma 3.1.

Now we must choose the children from the squares in the other diagonal. This choice is more delicate since the squares from the first diagonal could be much closer. By the pigeonhole principle, for each $P = P_i^Q$, $Q \in \{Q_0^1, Q_1^0\}$, there exist exactly two other squares, $P_j^{Q_0^0}$ and $P_k^{Q_1^1}$, which are at distance at most 2^{-m} . Indeed, according to Lemma 3.1, the distance condition forces $P_j^{Q_0^0}, P_k^{Q_1^1}$ to be in the same row (or column) as P_i^Q , while each row (or column) of Q_0^0 (or Q_1^1) contains exactly one square by induction hypothesis. In fact, the distance between P_i^Q and $P_j^{Q_0^0}$ is precisely equal to 0, and the same holds for $P_k^{Q_1^1}$.

We now project $Q_j^{Q_0^0}$ and $Q_k^{Q_1^1}$, the children chosen from $P_j^{Q_0^0}$ and $P_k^{Q_1^1}$ onto P . Note that their projection leaves exactly one first-generation child of P untouched, which we select as Q_i^Q . It suffices to show that the distance from Q_i^Q to any square in Q_0^0 is larger than 2^{-m} , since the case of Q_1^1 is symmetric and any other combination can be dealt with in the same way as above. To see this, if the square in Q_0^0 does not come from the $P_j^{Q_0^0}$ chosen above, the distance has to be larger than 2^{-m} by Lemma 3.1. Otherwise, the square is exactly $Q_j^{Q_0^0}$. By the construction above, both its horizontal projection and its vertical projection and the respective ones of Q_i^Q are disjoint, which allows one to apply Lemma 3.1 once more to conclude the proof. $\square$

3.2. Construction of $\mathcal{Z}$

In the construction, there are two aspects that will require special care: each point in $\mathcal{Z}$ has to be close to exactly one point from $\mathcal{P}$ (but not too close), while two different

points in $\mathcal{Z}$ cannot be too close to each other. We begin by constructing certain special rectangles inside which we shall place the points forming $\mathcal{Z}$.

Definition. We say that a rectangle R is a standard rectangle if it belongs to the collection

$$\mathcal{E} = \{R \subseteq [0, 1)^2 : R \text{ dyadic, } |R| = 2^{-2m-2}, \text{ and } R \cap \mathcal{P} \neq \emptyset\}.$$

If $R \in \mathcal{E}$ we let p_R denote the unique point in $R \cap \mathcal{P}$. Also let $\mathcal{E}_p$ denote the collection of standard rectangles containing $p \in \mathcal{P}$.

One can obtain $\mathcal{E}_p$ by area-preserving dilations of one fixed $R \in \mathcal{E}_p$. Indeed, for any dyadic $R = I \times J$ and any $j \in \mathbb{Z}$ define

$$\text{Dil}_j R = I' \times J',$$

where $I' \times J'$ is the unique dyadic rectangle in $\mathcal{E}_p$ of dimensions $2^j|I| \times 2^{-j}|J|$. Then it is easy to see that for each standard rectangle R , $\mathcal{E}_p = \{\text{Dil}_j R \subseteq [0, 1)^2\}$. In particular, for any $p \in \mathcal{P}$ there are approximately m distinct standard rectangles in $\mathcal{E}_p$. Given $x, y \in \mathbb{R}$, let

$$\delta(x, y) = \inf\{|Q| : Q \in \mathcal{D} \text{ is a dyadic interval containing } x \text{ and } y\}.$$

Also, for dyadic intervals K define

$$\delta(x, K) := \inf_{y \in K} \delta(x, y).$$

Now suppose $R = I \times J \in \mathcal{E}$, with $p_R = p = (x, y)$. Given $k \geq 1$, we define $I_{(k)}$ to be the largest dyadic interval I' such that

$$\delta(x, I') = 2^{-k+1}|I|,$$

and define $J_{(k)}$ similarly. Note that $I_{(1)}$ is the child of I that does not contain x , and for all $k \geq 1$, $I_{(k)}$ is the unique dyadic offspring of I of generation k satisfying $x \notin I_{(k)}$ and $x \in \widehat{I_{(k)}}$ (the dyadic parent of $I_{(k)}$).

The following result is the main technical lemma needed for our construction.

Lemma 3.2. Fix an integer $k \geq 1$ with $k \ll m$. For any dyadic rectangle $R = I \times J \in \mathcal{E}_p$ there exists a sub-rectangle $R^* \subset I_{(1)} \times J_{(k)}$ such that for all $q \neq p$ in $\mathcal{P}$:

$$\text{dist}_{\mathbb{H}}(q, R^*)^2 \geq 2^{-2m-1}.$$

We will apply this lemma to construct the points $\mathcal{Z}$ used in the proof of the main theorem as follows: given $R = I \times J \in \mathcal{E}$, we choose one point z_R in $R^* \subset I_{(1)} \times J_{(k)}$

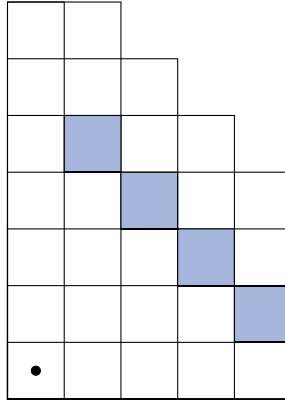


Fig. 2. Log-log sketch of some standard rectangles. In blue are the subrectangles $I_{(1)} \times J_{(k)}$ with $k = 3$. (For interpretation of the colors in the figure, the reader is referred to the web version of this article.)

and let $\mathcal{Z} = \bigcup_{R \in \mathcal{E}} z_R$. Fig. 2 gives some intuition about the situation. Note that this immediately guarantees properties (Z.1), (Z.2) and (Z.3) of Theorem 2.2. This placing also allows us to prove the remaining two properties. We start by proving directly (Z.4):

Lemma 3.3. *For every $R \in \mathcal{E}$ we have*

$$\#(R \cap \mathcal{Z}) \lesssim k.$$

Proof. First observe that if $z_{R'} \in R$ with $R, R' \in \mathcal{E}$ then by Lemma 3.2 we must have $p_{R'} = p_R = p$. So let $\mathcal{Z}_R$ be defined by

$$\mathcal{Z}_R = \{T \in \mathcal{E}_p : z_T \in R\}.$$

It suffices to show that $\#(\mathcal{Z}_R) \lesssim k$.

Observe that any $T \in \mathcal{Z}_R$ must have $\pi_1(T) \subseteq \pi_1(R)$. Indeed, if $\pi_1(T) \supsetneq \pi_1(R)$, by definition $z_T \in \pi_1(T)_{(1)} \times \pi_2(T)_{(k)}$ and one has $\pi_1(R) \subset \pi_1(T)_{(1)}$. But this is impossible since $\pi_1(T)_{(1)} \subseteq \pi_1(T) \setminus \pi_1(R)$. In addition, there must hold

$$|\pi_2(T)| \leq 2^{k-1} |\pi_2(R)|,$$

as the smallest interval containing both $\pi_2(p_T)$ and $\pi_2(z_T)$ is $\widehat{\pi_2(T)_{(k)}}$. The desired estimate then follows from the fact that there are $O(k)$ such standard rectangles. $\square$

3.3. Proof of Lemma 3.2 and (Z.5)

In this subsection, given two dyadic rectangles R_1, R_2 that are not contained in one another but which do intersect at some nontrivial set, we say that R_2 intersects R_1 *horizontally* if

$$\pi_2(R_2) \subsetneq \pi_2(R_1),$$

and similarly R_2 intersects R_1 *vertically* if

$$\pi_1(R_2) \subsetneq \pi_1(R_1).$$

One can see that these two options are mutually exclusive for R_1 and R_2 . Also, one can see that if R_2 intersects R_1 horizontally then $\pi_1(R_1) \subsetneq \pi_1(R_2)$, and if R_2 intersects R_1 vertically then $\pi_2(R_1) \subsetneq \pi_2(R_2)$. We start by recording the following information that will be used later:

Lemma 3.4. *Let $R = I \times J \in \mathcal{E}$ be such that $I \times J^{(\ell+1)} \subseteq [0, 1]^2$, $\ell \geq 1$. Then*

$$I_{(\ell)} \times J^{(\ell+1)}$$

contains exactly one $q \in \mathcal{P}$, $q \neq p_R$, and furthermore

$$q \in I_{(\ell)} \times (J^{(\ell+1)} \setminus J^{(\ell)}).$$

Proof. We have that $|\widehat{I_{(\ell)}} \times J^{(\ell+1)}| = 2^{-\ell+1}2^{\ell+1}|I||J| = 2^{-2m}$, so there are exactly two points from $\mathcal{P}$ in $\widehat{I_{(\ell)}} \times J^{(\ell+1)}$. One of these points must be p_R , so the other one is q . To prove the second part of the assertion, note that we cannot have $q \in (\widehat{I_{(\ell)}} \setminus I_{(\ell)}) \times J^{(\ell+1)}$ or $q \in I_{(\ell)} \times J^{(\ell)}$, since in each case we would have p_R and q contained in a dyadic rectangle of area smaller or equal than 2^{-2m-1} , violating (P.1). $\square$

Lemma 3.5. *Let $R = I \times J \in \mathcal{E}$. For any $\ell \geq 1$ with $2^{\ell+1}|J| \leq 1$, there exists exactly one $T = W \times H \in \mathcal{E}$ intersecting R vertically such that $p_T \neq p_R$, $W \subset I_{(\ell)}$, and $|W| = 2^{-1}|I_{(\ell)}|$.*

Proof. By Lemma 3.4 the rectangle $I_{(\ell)} \times J^{(\ell+1)}$ contains exactly one $q \in \mathcal{P}$ and $q \neq p_R$. Moreover, we know that $q \in I_{(\ell)} \times (J^{(\ell+1)} \setminus J^{(\ell)})$. We define W as the child of $I_{(\ell)}$ that contains $\pi_1(q)$, and choose H so that $\pi_2(q) \in H$ and $T = W \times H$ has area 2^{-2m-2} . In particular, we have $H = J^{(\ell+1)}$. By construction, $T \in \mathcal{E}$.

We now claim that T is the unique standard rectangle intersecting R vertically with $p_T \neq p_R$, $W \subset I_{(\ell)}$, and $|W| = \frac{1}{2}|I_{(\ell)}|$. By contradiction, suppose there existed another $T' = W' \times H'$ satisfying the same properties. Then since $W, W' \subset I_{(\ell)}$ with $|W| = |W'|$ we know that T and T' are disjoint and therefore $p_T \neq p_{T'}$. Since T' is a standard rectangle and T, T' both intersect R vertically, we also know that $H' = H = J^{(\ell+1)}$. But this contradicts Lemma 3.4: that there can only be one $q \in \mathcal{P}$ in $I_{(\ell)} \times J^{(\ell+1)}$ with $q \neq p_R$, so $p_T = p_{T'}$, which is a contradiction. $\square$

Lemma 3.5 holds for rectangles intersecting R horizontally with exactly the same proof, which we omit for brevity. Assume now that the rectangle $R = I \times J$ has been fixed. Let

$$\mathcal{E}_R^{\ell,v} = \{T \in \mathcal{E} : T \text{ intersects } R \text{ vertically and } \pi_1(T) \subset I_{(\ell)}\},$$

and

$$\mathcal{E}_R^{\ell,h} = \{T \in \mathcal{E} : T \text{ intersects } R \text{ horizontally and } \pi_2(T) \subset J_{(\ell)}\}.$$

We will analyze the vertical rectangles $\mathcal{E}_R^{\ell,v}$; essentially the same arguments apply to the horizontal collection $\mathcal{E}_R^{\ell,h}$.

If S and T are in $\mathcal{E}_R^{\ell,v}$, we say that $S \preceq T$ if $\pi_1(S) \subset \pi_1(T)$ (an analogous order can be defined in $\mathcal{E}_R^{\ell,v,*}$). Denote by $\mathcal{E}_R^{\ell,v,*} = \{T_1, T_2, \dots\}$ the set of maximal elements with respect to the ordering $\preceq$. Assume they are ordered so that the sequence $a_i := |\pi_1(T_i)|$ is non-increasing.

Lemma 3.6. *The sequence a_i satisfies*

$$a_i = 2^{-i}|I_{(\ell)}|, 1 \leq i \leq m - \ell.$$

Proof. By Lemma 3.5 we find that T_1 satisfies $\pi_1(T_1) \subset I_{(\ell)}$ and $a_1 = 2^{-1}|I_{(\ell)}|$. Moreover, we know that there can be no other T_j with $a_j = a_1$, and so in particular $a_2 \leq 2^{-2}|I_{(\ell)}|$.

We assume inductively that the desired result holds for $T_1, \dots, T_{i-1}$, and aim to show that $a_i = 2^{-i}|I_{(\ell)}|$. Note that $\{\pi_1(T_1), \dots, \pi_1(T_{i-1})\}$ are pairwise disjoint. Then by induction

$$\pi_1(T_i) \subset I_{(\ell)} \setminus \bigcup_{j=1}^{i-1} \pi_1(T_j),$$

and we also have that $a_j = 2^{-1}a_{j-1}$ and $a_1 = 2^{-1}|I_{(\ell)}|$. Therefore, $\pi_1(T_i)$ must be contained in some interval K with $|K| = 2^{-i+1}|I_{(\ell)}|$. Moreover, by induction K and $\pi_1(T_{i-1})$ must have the same dyadic parent.

We first show that $a_i < |K|$. To see this, suppose by contradiction that $\pi_1(T_i) = K$. Then $a_i = a_{i-1}$ and

$$\widehat{\pi_1(T_i)} = \widehat{\pi_1(T_{i-1})}.$$

Since both T_{i-1} and T_i must also have the same height, their union is a dyadic rectangle of area 2^{-2m-1} which contains both $p_{T_{i-1}}$ and p_{T_i} , which is impossible by (P.1).

Finally, we have to show that $a_i \geq 2^{-1}a_{i-1}$ or, equivalently, that there exists $T \in \mathcal{E}_R^{\ell,v,*}$ with $|\pi_1(T)| = 2^{-1}a_{i-1}$. To that end, let $\tilde{T}_{i-1} \in \mathcal{E}_{p_{T_{i-1}}}$ be the unique rectangle with $\pi_1(\tilde{T}_{i-1}) = \widehat{K}$. If $2^2|\pi_2(\tilde{T}_{i-1})| \leq 1$, one can apply Lemma 3.5 with $\ell = 1$ to $\tilde{T}_{i-1}$ to conclude that there is exactly one $T \in \mathcal{E}$ intersecting $\tilde{T}_{i-1}$ vertically with $p_T \neq p_{T_{i-1}}$ and

$$|\pi_1(T)| = 2^{-1}|\widehat{K}_{(1)}| = 2^{-i}|I_{(\ell)}|.$$

$T_i = T$ by maximality, and this closes the induction. If on the other hand $2^2|\pi_2(\tilde{T}_{i-1})| > 1$, then we claim that $i > m - \ell$ and we have finished. Indeed, if T_i exists in this case, one has

$$|\pi_1(T_i)| < |K| = 2^{-1}|\pi_1(\tilde{T}_{i-1})|.$$

Therefore,

$$|\pi_2(T_i)| > 2|\pi_2(\tilde{T}_{i-1})| \geq 2^2|\pi_2(\tilde{T}_{i-1})| > 1,$$

which means it is impossible for T_i to be contained in $[0, 1)^2$. $\square$

Proof of Lemma 3.2. It is enough to show the following slightly stronger claim: for any pair of integers $k, \ell \geq 1$, any $R = I \times J \in \mathcal{E}$ contains a non-empty rectangle $E \times F$ such that

$$E \times F \subseteq I_{(\ell)} \times J_{(k)} \setminus \left(\bigcup_{\substack{R' \in \mathcal{E} \\ p_{R'} \neq p_R}} R' \right).$$

Any $R' \in \mathcal{E}$ which intersects $I_{(\ell)} \times J_{(k)}$ must be in one of the two collections $\mathcal{E}_R^{\ell, v}, \mathcal{E}_R^{k, h}$, when $p_R \neq p_{R'}$.

We start by constructing E , so now we only need to take rectangles belonging to $\mathcal{E}_R^{\ell, v, *}$ into account. From Lemma 3.6 we know that $a_i = |\pi_1(T_i)|$ satisfies $a_i = 2^{-i}|I_{(\ell)}|$. It thus follows that $|\pi_2(T_i)|$ must increase in size exponentially, and therefore the total number of $T_i \in \mathcal{E}_R^{\ell, v, *}$ is bounded by some constant that depends on R (since all rectangles are contained in $[0, 1)^2$). Since there are only finitely many $\pi_1(T_i) \subset I_{(\ell)}$, the estimate on their sizes from Lemma 3.6 implies that there must be some dyadic subinterval $E \subset I_{(\ell)}$ with

$$\pi_1(T_i) \cap E = \emptyset \text{ for all } i,$$

as desired. The same procedure yields an interval $F \subset J_{(k)}$ which no rectangles $T' \in \mathcal{E}_R^{k, h, *}$ can intersect. $\square$

Finally, we turn to the proof of (Z.5), which is the content of the next lemma.

Lemma 3.7. *For every dyadic rectangle $R \subset [0, 1)^2$ with $|R| \geq 2^{-2m-1}$ we have*

$$\#(R \cap \mathcal{Z}) \lesssim 2^{2m}mk|R|.$$

Proof. First, we may assume that $|R| = 2^{-2m-1}$ and show

$$\#(R \cap \mathcal{Z}) \lesssim mk.$$

Indeed, if it is larger we can just write it as a disjoint union of dyadic rectangles of area 2^{-2m-1} and use the estimate for rectangles of that size. Since $|R| = 2^{-2m-1}$ we know that it contains exactly one point $p_0 \in \mathcal{P}$, and from Lemma 3.3 we know R contains about k points $z \in \mathcal{Z}$ such that $p(z) = p_0$.

For every $z \in R \cap \mathcal{Z}$ such that $p(z) \neq p_0$, there exists $T \in \mathcal{E}$ for which $z_T = z$, which we denote by T_z . Since both R and T_z are dyadic rectangles and $|T_z| < |R|$ we must have

$$\mathcal{Z} \cap R = \mathcal{V} \sqcup \mathcal{H} \sqcup \mathcal{O},$$

where $\mathcal{V}$ consists of those points $z \in \mathcal{Z} \cap R$ such that T_z intersects R vertically, $\mathcal{H}$ consists of those points $z \in \mathcal{Z} \cap R$ such that T_z intersects R horizontally, and $\mathcal{O}$ is the collection of points $z \in \mathcal{Z}$ whose associated point $p(z) = p_0$. We will only estimate the size of $\mathcal{V}$, since the argument for $\mathcal{H}$ is similar.

By the construction of $\mathcal{Z}$ and the previous arguments, for each $z \in \mathcal{V}$ there exists $T \in \mathcal{E}_R^{\ell, v, *}$ containing z . By Lemma 3.3 it suffices to show that there are no more than some constant times m such rectangles.

Let $\tilde{R} = \tilde{I} \times \tilde{J}$ be the unique rectangle in $\mathcal{E}$ satisfying

$$p_R \in \tilde{R} \subset R \text{ and } \pi_1(\tilde{R}) = \pi_1(R).$$

Observe that any rectangle $T \in \mathcal{E}_R^{\ell, v, *}$ intersecting R vertically and whose z_T is inside R must intersect $(\tilde{I})_{(1)} \times \tilde{J}$ vertically (since otherwise z_T would be too close to p_0). According to Lemma 3.5 with $\ell = 1$, the maximal rectangle with shortest height must have height at least $4|\tilde{J}| = |\pi_2(R)|$ and the heights of these maximal rectangles increase exponentially. Therefore the number of maximal rectangles going through $(\tilde{I})_{(1)}$ is at most

$$\log_2(2^{2m}|\pi_2(R)|) \leq 2m + \log_2(|\pi_2(R)|) \leq 2m. \quad \square$$

The proof of Theorem 2.2 is complete.

4. Extensions and open questions

4.1. Failure of sparse bound for bi-parameter martingale transform

In this subsection, we extend our main theorem to the bi-parameter martingale transform, which is a 0-complexity dyadic shift that resembles the behavior of the bi-parameter Hilbert transform. In general, given a sequence $\sigma = \{\sigma_R\}_{R \in \mathcal{D} \times \mathcal{D}}$ satisfying $|\sigma_R| \leq 1$, the corresponding martingale transform is defined as

$$T_\sigma(f) := \sum_{R \in \mathcal{D}} \sigma_R \langle f, h_R \rangle h_R.$$

We have the following lower bound result.

Theorem 4.1. *Let measures μ, ν be the same as above. Then there exists a martingale transform T_σ such that*

$$|\langle T_\sigma(\mu), \nu \rangle| \gtrsim 2^k.$$

Recalling the upper estimate of sparse forms obtained in Subsection 2.2, one immediately gets the following corollary:

Corollary 4.2. *There cannot be any $(1, 1)$ sparse domination for all bi-parameter martingale transforms.*

Proof of Theorem 4.1. We first assume that $\sigma_R = 0$ unless $R \subset [0, 1]^2$. In general, for any $R = I \times J \in \mathcal{E}$ that gives rise to some point $z \in \mathcal{Z}$, denote $T_R := \widehat{I_{(\ell)}} \times \widehat{J_{(k)}}$, which is the smallest rectangle that contains both z and $p(z) \in \mathcal{P}$. Also, denote the standard rectangle that gives rise to $z \in \mathcal{Z}$ by R_z . Now for any fixed $z \in \mathcal{Z}$,

$$T_\sigma(\mu)(z) = \frac{1}{\#\mathcal{P}} \sum_{p \in \mathcal{P}} \sum_{R \subset [0, 1]^2} \sigma_R h_R(p) h_R(z).$$

Set $\sigma_T = 0$ unless $T = T_R$ for some $R \in \mathcal{E}$. In particular, $\{T_R\}_{R \in \mathcal{E}}$ and $\{R\}_{R \in \mathcal{E}}$ are one-to-one, and each T_R contains exactly one $p_R \in \mathcal{P}$. Hence,

$$T_\sigma(\mu)(z) \sim 2^{-2m} \sum_{R \in \mathcal{E}} \sigma_{T_R} h_{T_R}(p_R) h_{T_R}(z).$$

Note that z cannot be contained in T_R unless $R = R_z$ according to Lemma 4.3 below, therefore,

$$T_\sigma(\mu)(z) \sim 2^{-2m} \sigma_{T_{R_z}} h_{T_{R_z}}(p_{R_z}) h_{T_{R_z}}(z) = 2^{-2m} \sigma_{T_{R_z}} |T_{R_z}|^{-1} \xi_{T_{R_z}},$$

where $\xi_{T_{R_z}} = \pm 1$. Choosing $\sigma_{T_{R_z}} = \xi_{T_{R_z}}$, one has

$$T_\sigma(\mu)(z) \sim 2^{-2m} |T_{R_z}|^{-1} = 2^{-2m} \cdot (2^{-k+1} |R_z|)^{-1} \sim 2^k,$$

and the desired estimate follows immediately. $\square$

In the proof above, we have used the key observation that even though $R \in \mathcal{E}$ can contain many points $z \in \mathcal{Z}$ with $p_R = p_{R_z}$, T_R can only contain one $z_R \in \mathcal{Z}$. This is justified by the following result.

Lemma 4.3. *For any integers $k, \ell \geq 1$ and $R = I \times J \in \mathcal{E}$, let $z_R \in \mathcal{Z}$ be the point chosen in R as in Subsection 3.2, and $T_R = \widehat{I_{(\ell)}} \times \widehat{J_{(k)}}$. Then*

$$z_T \notin T_R, \quad \forall R, T \in \mathcal{E}, R \neq T.$$

Proof. Fix $R = I \times J \in \mathcal{E}$ and denote $z = z_R$. Given any other $R' = I' \times J' \in \mathcal{E}$, our goal is to show $z_{R'} \notin T_R$. Obviously, if $p_{R'} \neq p_R$, then by Lemma 3.2 $z_{R'}$ is not even contained in R . It thus suffices to assume $p_{R'} = p_R = p$. Suppose $|I| > |I'|$, i.e. $I = I'^{(j)}$ for some $j \geq 1$, then $|J| = 2^{-j}|J'|$ and $J \subset J'$. Since both $\widehat{J_{(k)}}$, $\widehat{J'_{(k)}}$ contain $\pi_2(p)$, one has $\widehat{J'_{(k)}} \supsetneq \widehat{J_{(k)}}$.

We claim that $\pi_2(z_{R'}) \notin \widehat{J_{(k)}}$, hence $z_{R'} \notin T_R$. Indeed, $\pi_2(z_{R'}) \in J'_{(k)}$, and $\widehat{J_{(k)}}$ is either contained in $J'_{(k)}$ or its dyadic sibling. But it is impossible for $\widehat{J_{(k)}}$ to be contained in $J'_{(k)}$ as it would imply $\pi_2(p) \in J'_{(k)}$, which is an obvious contradiction. Therefore, $\pi_2(z_{R'}) \notin \widehat{J_{(k)}}$ and the proof for the case $|I| > |I'|$ is complete.

The case $|I| < |I'|$ can be treated symmetrically, as one has $|J| < |J'|$ and can show in the same way as above that $\pi_1(z_{R'}) \notin \widehat{I_{(\ell)}}$. $\square$

Remark 4.4. Using the same method, one can easily show that there exists a bi-parameter dyadic shift of any given complexity which cannot have a $(1, 1)$ sparse bound. We omit the details.

4.2. Failure of sparse bound in higher dimensions

For $n \geq 1$, define the n -parameter strong maximal function

$$\mathcal{M}_n(f)(x) := \sup_{R \ni x} \frac{1}{|R|} \int_R |f(y)| dy, \quad f: \mathbb{R}^n \rightarrow \mathbb{C},$$

where R is any n -dimensional dyadic rectangle. Clearly, $\mathcal{M}_S$ equals $\mathcal{M}_2$. We have the following theorem.

Theorem 4.5. *Let (D_n) denote the following statement: there exists $C > 0$ such that for all compactly supported integrable functions f, g on $\mathbb{R}^n$, there exists a sparse collection of dyadic rectangles $\mathcal{S}$ such that*

$$|\langle \mathcal{M}_n(f), g \rangle| \leq C \sum_{R \in \mathcal{S}} \langle |f| \rangle_R \langle |f| \rangle_R |R|.$$

Then for all $n \geq 2$, there holds

$$(D_n) \implies (D_{n-1}).$$

Proof. The desired result follows from a tensor product argument. For the sake of simplicity, we will only prove $(D_3) \implies (D_2)$. Our goal is to show (D_2) : for any given functions f, g on $\mathbb{R}^2$, we would like to find a sparse dominating form for $\langle \mathcal{M}_2(f), g \rangle$. Define $\tilde{f} = f \otimes \chi_{[0,1]}$ and $\tilde{g} = g \otimes \chi_{[0,1]}$, both are compactly supported integrable functions on $\mathbb{R}^3$. By the assumption (D_3) , there exists a sparse collection $\tilde{\mathcal{S}}$ of rectangles in $\mathbb{R}^3$ such that

$$|\langle M_3(\tilde{f}), \tilde{g} \rangle| \leq C \sum_{\tilde{R}=R \times J \in \tilde{\mathcal{S}}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}|,$$

where $R \in \mathbb{R}^2$, $J \in \mathbb{R}$.

By definition, for all $\tilde{x} = (x, x_3) \in \mathbb{R}^3$ with $x_3 \in [0, 1)$,

$$M_3(\tilde{f})(\tilde{x}) = \sup_{R \times J \ni \tilde{x}} \frac{1}{|R| \cdot |J|} \int_{R \times J} |\tilde{f}(\tilde{y})| d\tilde{y} = M_2(f)(x) M(\mathbb{1}_{[0,1)})(x_3) = M_2(f)(x).$$

Therefore,

$$\langle M_3(\tilde{f}), \tilde{g} \rangle_{\mathbb{R}^3} = \langle M_2(f), g \rangle_{\mathbb{R}^2}.$$

It thus suffices to show that the sparse form in $\mathbb{R}^3$ can be dominated by a sparse form in $\mathbb{R}^2$.

To see this, rewrite

$$\sum_{\tilde{R}=R \times J \in \tilde{\mathcal{S}}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}| = \sum_{\tilde{R}=R \times J \subset \mathbb{R}^3} \alpha_{\tilde{R}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}|$$

where $\alpha_{\tilde{R}} = 1$ if $\tilde{R} \in \tilde{\mathcal{R}}$, and 0 otherwise. The sparsity of $\tilde{\mathcal{R}}$ thus means for some constant Λ there holds

$$\sum_{\tilde{R} \subset \tilde{\Omega}} \alpha_{\tilde{R}} |\tilde{R}| \leq \Lambda |\tilde{\Omega}|, \quad \forall \text{ open set } \tilde{\Omega} \subset \mathbb{R}^3.$$

One can further simplify the expression

$$\begin{aligned} & \sum_{\tilde{R}=R \times J \subset \mathbb{R}^3} \alpha_{\tilde{R}} \langle |\tilde{f}| \rangle_{\tilde{R}} \langle |\tilde{g}| \rangle_{\tilde{R}} |\tilde{R}| \\ &= \sum_{R \subset \mathbb{R}^2} \langle |f| \rangle_R \langle |g| \rangle_R |R| \sum_{J \subset \mathbb{R}} \alpha_{R \times J} |J| \cdot \left(\frac{|[0, 1) \cap J|}{|J|} \right)^2 \\ &=: \sum_{R \subset \mathbb{R}^2} \langle |f| \rangle_R \langle |g| \rangle_R |R| \beta_R. \end{aligned}$$

It suffices to show that $\{\beta_R\}_{R \subset \mathbb{R}^2}$ is a Carleson sequence.

Indeed, decomposing

$$\beta_R = \sum_{j=0}^{\infty} 2^{-2j} \sum_{\substack{J \subset \mathbb{R} \\ |J \cap [0, 1)| \sim 2^{-j} |J|}} \alpha_{R \times J} |J|,$$

one has for all open set $\Omega \subset \mathbb{R}^2$ that

$$\begin{aligned} \sum_{R \subset \Omega} \beta_R |R| &= \sum_{j=0}^{\infty} 2^{-2j} \sum_{R \subset \Omega} \sum_{\substack{J \subset \mathbb{R} \\ |J \cap [0,1]| \sim 2^{-j} |J|}} \alpha_{R \times J} |R \times J| \\ &\leq \sum_{j=0}^{\infty} 2^{-2j} \sum_{R \times J \subset \tilde{\Omega}} \alpha_{R \times J} |R \times J|, \end{aligned}$$

where $\tilde{\Omega} := \Omega \times \{x_3 \in \mathbb{R} : M(\chi_{[0,1]})(x_3) > C2^{-j}\}$. By the weak L^1 boundedness of M , one has $|\tilde{\Omega}| \lesssim 2^j |\Omega|$, hence

$$\sum_{R \subset \Omega} \beta_R |R| \lesssim \Lambda \sum_{j=0}^{\infty} 2^{-j} |\Omega| \lesssim \Lambda |\Omega|.$$

The proof is complete. $\square$

Similar results hold true for multi-parameter martingale transforms as well. One can define $\tilde{f} = f \otimes h_{[0,1]}$ and $\tilde{g} = g \otimes h_{[0,1]}$. Then one observes that $|\langle T_{\sigma}(h_{[0,1]}), h_{[0,1]} \rangle| \geq C$ for some martingale transform T_{σ} in the third variable and therefore can argue similarly as above. The details are left to the reader.

4.3. Other sparse forms

As mentioned in the introduction, it is also interesting to consider the possibility of domination of the bi-parameter operators by bigger sparse forms, for instance those of type $(\phi, 1)$:

$$\sum_{R \in \mathcal{S}} \langle |f| \rangle_{R, \phi} \langle |g| \rangle_R |R|,$$

where

$$\langle f \rangle_{R, \phi} = \inf \left\{ \lambda > 0 : \frac{1}{|R|} \int_R \phi \left(\frac{f(x)}{\lambda} \right) \leq 1 \right\} dx,$$

and interesting norms include for example $\phi(x) = x[\log(e+x)]^{\alpha}$, $\alpha > 0$, and $\phi(x) = x^p$, $p > 1$. We can actually extend our main theorem to the following:

Theorem 4.6. *Let $\phi(x) = x[\log(e+x)]^{\alpha}$, $0 < \alpha < \frac{1}{2}$. Then there cannot be any $(\phi, 1)$ -sparse domination for the strong maximal function or the martingale transform T_{σ} .*

Proof. Fix $k \ll m$. We need to first slightly modify the measure μ to get a function that lies in the correct normed space. Let ν be the same as before and define

$$f = \frac{1}{\#\mathcal{P}} \sum_{p \in \mathcal{P}} \frac{\mathbb{1}_{Q_p}}{|Q_p|},$$

where for each $p \in \mathcal{P}$, $Q_p \ni p$ is the little cube (of side-length $\sim 2^{-2m}$) that is constructed in the proof of Theorem 2.1. By the exact same argument in Section 2, it is easy to check that the same lower bound for the bilinear forms associated to $\mathcal{M}_S$ and the special martingale transform T_σ still hold true, i.e.

$$\langle M_S(f), \nu \rangle \gtrsim 2^k, \quad \langle T_\sigma(f), \nu \rangle \gtrsim 2^k.$$

We shall now focus on showing the upper bound. For every Λ -Carleson collection $\mathcal{S}$ we will show that

$$\sum_{R \in \mathcal{S}} \langle f \rangle_{R, \phi} \langle \nu \rangle_R |R| \lesssim \Lambda k m^\alpha \left(1 + \frac{2^{2k}}{m}\right). \quad (4.1)$$

If (4.1) holds true, then by taking $m = 2^{2k}$ and by noting that $\alpha < \frac{1}{2}$, one immediately sees that Λ blows up to infinity as $k \rightarrow \infty$, which completes the proof.

We now prove (4.1). Compared to the estimate of $\langle \mu \rangle_R$ in the proof of Theorem 2.6, it suffices to show that there is at most an extra factor of m^α in the upper estimate of the new average form $\langle f \rangle_{R, \phi}$. Specifically, in the case $|\overline{R}| = |R \cap [0, 1)^2| \geq 2^{-2m-1}$, we would like to show that

$$\langle f \rangle_{R, \phi} \lesssim \frac{|\overline{R}|}{|R|} \cdot \max \left(m^\alpha, \left[\log \left(\frac{|R|}{|\overline{R}|} \right) \right]^\alpha \right). \quad (4.2)$$

Note that if this estimate holds true, then one can proceed as in the proof of Theorem 2.6 to conclude that contribution to $\sum_{R \in \mathcal{S}} \langle f \rangle_{R, \phi} \langle \nu \rangle_R |R|$ from rectangles of this type is controlled by $\Lambda k(1 + m^\alpha)$. Similarly, in the case $|\overline{R}| < 2^{-2m-1}$, we will show that

$$\langle f \rangle_{\overline{R}, \phi} \lesssim \frac{m^\alpha}{2^{2m} |\overline{R}|}, \quad (4.3)$$

which, combined with the estimates for large rectangles, completes the proof of (4.1).

To see why (4.2) holds, let $\lambda > 0$ and one has

$$\begin{aligned} \frac{1}{|R|} \int_R \phi \left(\frac{f(x)}{\lambda} \right) dx &= \frac{1}{|R|} \sum_{p \in \mathcal{P} \cap \overline{R}} \int_{R \cap Q_p} \phi \left(\frac{\sum_{q \in \mathcal{P}} \frac{\mathbb{1}_{Q_q}(x)}{|Q_q|}}{\lambda \cdot \#\mathcal{P}} \right) dx \\ &= \frac{1}{|R|} \cdot \#\{\mathcal{P} \cap \overline{R}\} \cdot |R \cap Q_p| \cdot \phi \left(\frac{1}{\lambda \cdot \#\mathcal{P} \cdot |Q_p|} \right) \\ &\leq \frac{|\overline{R}|}{|R|} \cdot 2^{-2m} \cdot \phi \left(\frac{2^{2m}}{\lambda} \right). \end{aligned}$$

Take $\lambda = \frac{|\overline{R}|}{|R|} \cdot \max \left(m^\alpha, \left[\log \left(\frac{|R|}{|\overline{R}|} \right) \right]^\alpha \right)$, it suffices to show that the expression above with this λ value is $\lesssim 1$. Indeed, note that $e \ll \frac{2^{2m}}{\lambda}$, so

$$\frac{|\overline{R}|}{|R|} \cdot 2^{-2m} \cdot \phi \left(\frac{2^{2m}}{\lambda} \right) \lesssim \frac{\left[2m + \log \left(\frac{|R|}{|\overline{R}|} \right) - \alpha \log \left(\max \left(m, \log \left(|R|/|\overline{R}| \right) \right) \right) \right]^\alpha}{\max \left(m^\alpha, \left[\log \left(\frac{|R|}{|\overline{R}|} \right) \right]^\alpha \right)} \lesssim 1.$$

To see (4.3), again let $\lambda > 0$ and one can calculate similarly as above to obtain

$$\frac{1}{|\overline{R}|} \int_{\overline{R}} \phi \left(\frac{f(x)}{\lambda} \right) dx \leq 2^{-2m} \phi \left(\frac{2^{2m}}{\lambda} \right) = \frac{1}{\lambda} \left[\log \left(e + \frac{2^{2m}}{\lambda} \right) \right]^\alpha.$$

Take $\lambda = \frac{m^\alpha}{2^{2m}|\overline{R}|}$, then it suffices to show that the expression above with this λ value is $\lesssim 1$. Indeed, the lower bound $|\overline{R}| \gtrsim 2^{-2m-k}$ guarantees that $\frac{2^{2m}}{\lambda} \gg e$, hence

$$\frac{1}{\lambda} \left[\log \left(e + \frac{2^{2m}}{\lambda} \right) \right]^\alpha \lesssim \frac{2^{2m}|\overline{R}|}{m^\alpha} [4m + \log |\overline{R}| - \alpha \log m]^\alpha \lesssim 2^{2m}|\overline{R}| \lesssim 1. \quad \square$$

The argument in the theorem above is not strong enough to produce a contradiction when $\alpha \geq \frac{1}{2}$. In other words, it could be possible that the bi-parameter operators that are considered have a $(\phi, 1)$ -type sparse bound with some $\alpha \geq \frac{1}{2}$. Moreover, it is easy to check that a very similar argument as in Theorem 4.6 barely fails for concluding a contradiction when $\phi(x) = x^p$, $p > 1$. We think that positive or negative results for those types of bounds would be very interesting.

Problem. Determine whether $(\phi, 1)$ -type sparse bounds for $\mathcal{M}_S$ are possible for $\phi(x) = x^p$, $p > 1$ or $\phi(x) = x[\log(e+x)]^\alpha$ ($\alpha \geq \frac{1}{2}$).

References

- [1] D. Beltran, L. Cladek, Sparse bounds for pseudodifferential operators, arXiv e-prints, Nov. 2017.
- [2] C. Benea, F. Bernicot, T. Luque, Sparse bilinear forms for Bochner Riesz multipliers and applications, *Trans. Lond. Math. Soc.* 4 (1) (2017) 110–128.
- [3] M. Christ, R. Fefferman, A note on weighted norm inequalities for the Hardy–Littlewood maximal operator, *Proc. Amer. Math. Soc.* 87 (3) (1983) 447–448.
- [4] J.M. Conde-Alonso, A. Culiuc, F. Di Plinio, Y. Ou, A sparse domination principle for rough singular integrals, *Anal. PDE* 10 (5) (2017) 1255–1284.
- [5] J.M. Conde-Alonso, G. Rey, A pointwise estimate for positive dyadic shifts and some applications, *Math. Ann.* 365 (3–4) (2016) 1111–1135.
- [6] A. Córdoba, Maximal functions: a problem of A. Zygmund, in: *Euclidean Harmonic Analysis*, Proc. Sem., Univ. Maryland, College Park, MD, 1979, in: *Lecture Notes in Math.*, vol. 779, Springer, Berlin, 1980, pp. 154–161.
- [7] A. Culiuc, F. Di Plinio, Y. Ou, Domination of multilinear singular integrals by positive sparse forms, *J. Lond. Math. Soc.* 98 (2) (2018) 369–392.
- [8] A. Culiuc, F. Di Plinio, Y. Ou, Uniform sparse domination of singular integrals via dyadic shifts, *Math. Res. Lett.* 2 (1) (2018).

- [9] J. Dick, F. Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, Cambridge, 2010.
- [10] L.E. Dor, On projections in L_1 , *Ann. of Math.* (2) 102 (3) (1975) 463–474.
- [11] R. Fefferman, J. Pipher, A covering lemma for rectangles in $\mathbb{R}^n$, *Proc. Amer. Math. Soc.* 133 (11) (2005) 3235–3241.
- [12] T.S. Hanninen, Equivalence of sparse and Carleson coefficients for general sets, arXiv e-prints, Sept. 2017.
- [13] T.P. Hytönen, The sharp weighted bound for general Calderón–Zygmund operators, *Ann. of Math.* (2) 175 (3) (2012) 1473–1506.
- [14] B. Krause, M.T. Lacey, Sparse bounds for maximally truncated oscillatory singular integrals, arXiv e-prints, Jan. 2017.
- [15] M.T. Lacey, An elementary proof of the A_2 bound, *Israel J. Math.* 217 (1) (2017) 181–195.
- [16] M.T. Lacey, Sparse bounds for spherical maximal functions, arXiv e-prints, Feb. 2017.
- [17] M.T. Lacey, D. Mena Arias, The sparse T_1 theorem, *Houston J. Math.* 43 (1) (2017) 111–127.
- [18] A.K. Lerner, A pointwise estimate for the local sharp maximal function with applications to singular integrals, *Bull. Lond. Math. Soc.* 42 (5) (2010) 843–856.
- [19] A.K. Lerner, On an estimate of Calderón–Zygmund operators by dyadic positive operators, *J. Anal. Math.* 121 (2013) 141–161.
- [20] A.K. Lerner, A simple proof of the A_2 conjecture, *Int. Math. Res. Not. IMRN* (14) (2013) 3159–3170.
- [21] E.T. Sawyer, A characterization of a two-weight norm inequality for maximal operators, *Studia Math.* 75 (1) (1982) 1–11.
- [22] F. Soria, Examples and counterexamples to a conjecture in the theory of differentiation of integrals, *Ann. of Math.* (2) 123 (1) (1986) 1–9.
- [23] A. Stokolos, Properties of the maximal operators associated with bases of rectangles in $\mathbb{R}^3$, *Proc. Edinb. Math. Soc.* (2) 51 (2) (2008) 489–494.