Publ. Mat. **68** (2024), 251–265
DOI: 10.5565/PUBLMAT6812411

GREEDY APPROXIMATION ALGORITHMS FOR SPARSE COLLECTIONS

Guillermo Rey

Abstract: We describe a greedy algorithm that approximates the Carleson constant of a collection of general sets. The approximation has a logarithmic loss in a general setting, but is optimal up to a constant with only mild geometric assumptions. The constructive nature of the algorithm gives additional information about the almost disjoint structure of sparse collections.

As applications, we give three results for collections of axis-parallel rectangles in every dimension. The first is a constructive proof of the equivalence between Carleson and sparse collections, first shown by Hänninen. The second is a structure theorem proving that every finite collection $\mathcal{E}$ can be partitioned into $\mathcal{O}(N)$ sparse subfamilies, where N is the Carleson constant of $\mathcal{E}$. We also give examples showing that such a decomposition is impossible when the geometric assumptions are dropped. The third application is a characterization of the Carleson constant involving only $L^{1,\infty}$ estimates.

2020 Mathematics Subject Classification: 42B25.

Key words: sparse collection, maximal function, multiparameter, Carleson sequence.

1. Introduction

Consider a collection $\mathcal{E}$ of measurable sets in $\mathbb{R}^d$ with finite measure. We say that $\mathcal{E}$ is η-sparse if for every R in $\mathcal{E}$ there exists a subset $E(R) \subseteq R$ satisfying $|E(R)| \geq \eta|R|$ such that the family $\{E(R)\}$ is pairwise disjoint. The number η quantifies how much overlap exists in $\mathcal{E}$ in a scale-invariant way. In particular, the closer η is to 1 the closer $\mathcal{E}$ is to being pairwise disjoint. A closely related quantity is the Carleson constant of $\mathcal{E}$. For any collection $\mathcal{F}$ let $\operatorname{sh}(\mathcal{F}) = \bigcup_{R\in\mathcal{F}} R$ be its *shadow*; then we say that $\mathcal{E}$ satisfies the Carleson condition with constant C if

$$\sum_{R\in\mathcal{F}} |R| \leq C|\operatorname{sh}(\mathcal{F})|$$

for all subcollections $\mathcal{F} \subseteq \mathcal{E}$. The best constant in the inequality above is usually called the Carleson constant of $\mathcal{E}$. We have specialized the definition to the case of the Lebesgue measure, which we denote by $|\cdot|$, but these notions carry over to general measure spaces, as we will describe later in the article.

These notions have been used extensively in harmonic analysis, for example in connection with the boundedness of maximal functions (cf. [**3**], [**4**]). In recent years they have also gained a lot of attention for their applications to weighted inequalities; we direct the interested reader to [**10**] for a nice review in this direction.

It is very easy to see that η-sparse collections satisfy the Carleson condition with constant η^{-1}. With more work one can show that the converse is also true when $\mathcal{E}$ consists of dyadic intervals (or squares, cubes, etc.); see for example Lemma 6.3 in [**8**]. This can be done exploiting the strong nestedness property of dyadic intervals, and in fact this structural property allows one to explicitly find the sets $E(R)$ in the definition above. In particular, in the (one-parameter) dyadic setting the Carleson condition becomes *local*, being equivalent to

$$\sum_{\substack{S\in\mathcal{E}\\ S\subseteq R}} |S| \le C|R|$$

for all R in $\mathcal{E}$.

This locality is sadly lacking in general, failing even in the two-parameter setting where, instead of dyadic intervals, one works with collections consisting of axis-parallel dyadic rectangles. This was shown in [**2**] with what is now known as Carleson's counterexample; see [**11**] or [**9**].

The equivalence between Carleson and sparse collections of dyadic cubes was first shown by I. E. Verbitsky in [**12**]; see also [**8**] for a different proof. Then, T. S. Hänninen in [**7**] adapted some of the ideas from [**12**] and from L. E. Dor's article [**5**] to prove the existence of the sets $E(R)$ in the definition for general collections of sets (not necessarily dyadic cubes), thus proving the equivalence in general. The proof, which at its core uses a convexity argument together with the Hahn–Banach separation theorem, is strikingly clean but gives no clue about *how* the sets $\{E(R)\}$ can be found or about their structure. A more geometric proof was later found in [**1**], but this proof is also non-constructive.

The main purpose of this article is to describe a greedy algorithm that is able to construct the sets $\{E(R)\}$ for any Carleson collection $\mathcal{E}$. With no geometric assumptions on $\mathcal{E}$ the algorithm has a logarithmic loss, but if one imposes some geometric structure, then the algorithm provides sets that are optimal up to an absolute constant. The constructive nature of our methods allows us to prove a structural theorem about sparse collections with only mild geometric assumptions, for example valid for axis-parallel rectangles in every dimension.

Before stating the main results, let us begin with some definitions. We will frequently use the words *collection* and *family* to mean an unordered sequence, instead of the usual definition of *set*. In particular we allow repeated elements.

We can define the Carleson constant of $\mathcal{E}$ with respect to a measure μ as

$$\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} = \sup\left\{\frac{1}{\mu(\mathrm{sh}(\mathcal{F}))}\sum_{R\in\mathcal{F}}\mu(R) : \mathcal{F}\subseteq\mathcal{E}\right\}. \tag{1.1}$$

We will write just $\|\mathcal{E}\|_{\mathrm{Carleson}}$ when μ is the Lebesgue measure. For simplicity, all of our collections will be assumed to be finite. As a consequence, we can assume μ to be a general measure as long as all the elements of the collections have finite measure.

For general measures, the straightforward generalization of sparse collection is not equivalent to the Carleson condition above (one needs μ to have no point masses). This can be readily seen with the example

$$\mathcal{E} = \{\underbrace{\{1\},\dots,\{1\}}_{N\text{ times}}\} \quad \mu = \text{Counting measure.}$$

Instead, we can extend it as follows.

Definition 1. We say that $\mathcal{E}$ is η-sparse with respect to the measure μ if one can find non-negative measurable functions $\varphi_R \ge 0$ for each R in $\mathcal{E}$ such that

$$\int_R \varphi_R\, d\mu \ge \eta\mu(R), \tag{1.2}$$

$$\sum_R \varphi_R \le 1. \tag{1.3}$$

We call the best constant η above (over all possible φ_R) the sparse constant of $\mathcal{E}$ with respect to μ, that is:

$$\|\mathcal{E}\|_{\mathrm{Sparse}(\mu)} = \sup\{\eta \geq 0 : \mathcal{E} \text{ is } \eta\text{-sparse with respect to } \mu\}.$$

This is only a slight generalization of the previous definition (one can just take $\varphi_R = \mathbb{1}_{E(R)}$ to recover the original). In fact, when the measure μ has no point masses, one can use a convexity argument like Lemma 2.3 from [**5**] to show that the two definitions of sparse collection are equivalent. With this notation the Carleson-sparse equivalence from [**7**] becomes

$$\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}\|\mathcal{E}\|_{\mathrm{Sparse}(\mu)} = 1.$$

We will not use this equivalence but only the easy direction alluded to earlier in the introduction, which we state here as

Lemma 1.1. *Let $\mathcal{E}$ be a finite collection of sets with finite μ-measure; then we have*

$$\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}\|\mathcal{E}\|_{\mathrm{Sparse}(\mu)} \leq 1.$$

Proof: Suppose that $\mathcal{E}$ is η-sparse with respect to the measure μ. We will show that $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \leq \eta^{-1}$. To this end, let $\mathcal{F}$ be any subcollection of $\mathcal{E}$ and let $\{\varphi_R\}$ be the functions from Definition 1. Then

$$\begin{aligned}\sum_{R\in\mathcal{F}} \mu(R) &\leq \eta^{-1} \sum_{R\in\mathcal{F}} \int_R \varphi_R \, d\mu \\ &= \eta^{-1} \int_{\mathrm{sh}(\mathcal{F})} \sum_{R\in\mathcal{F}} \varphi_R \, d\mu \leq \eta^{-1}\mu(\mathrm{sh}(\mathcal{F})),\end{aligned}$$

which is what we wanted to show. □

We are now ready to describe our algorithm and main results.

Suppose one were to compute $\|\mathcal{E}\|_{\mathrm{Carleson}}$ directly with (1.1). The definition involves computing a certain sum for each of the subcollections $\mathcal{F} \subseteq \mathcal{E}$, so the process quickly becomes intractable as the cardinality of $\mathcal{E}$ grows.

One could instead try to find functions φ_R as in Definition 1. However, this approach quickly runs into problems since the two conditions (1.2) and (1.3) are in direct opposition. Namely, (1.2) requires that the average of each φ_R be at least η, which together with (1.3) means that each φ_R cannot be much smaller than η on a large portion of R. But also, (1.3) implies that the functions cannot all be larger than η at the same place.

If $\mathcal{E}$ consists of only two elements $\{R_1, R_2\}$, then the problem becomes very easy: one can just set φ_{R_i} to be 1 on the symmetric difference of R_1 and R_2 and then equitably distribute the mass in $R_1 \cap R_2$ among the two in proportion to their masses. This could lead to an induction algorithm, but one easily sees that, even with simple examples, earlier choices of the functions φ_R can make the choice of the $(n+1)$-th function impossible, especially when the earlier choices do not take the global situation into account.

This suggests that we choose φ_R in a way that guarantees, independently of the choice of φ_S for $S \neq R$, that the sum of all the functions remains bounded by 1. In particular, we would like to find R so that arbitrarily solving the subproblem for $\mathcal{E} \setminus \{R\}$ still leaves space to choose φ_R appropriately.

We are not able to do this in general without at least some geometric information about $\mathcal{E}$. However, we *can* find R so that, if we choose the following functions in a special way (independent of the choice of φ_R), then there always exists a choice of φ_R that is valid up to a logarithmic factor. In particular, this strategy leads to functions φ_R satisfying

$$\int_R \varphi_R \, d\mu \gtrsim \frac{\eta}{\log \eta^{-1}} \mu(R),$$

where $\eta = \|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}^{-1}$.

In order to remove the logarithmic loss we can use the maximal operator associated to $\mathcal{E}$:

$$\mathcal{M}_{\mathcal{E}}^{\mu} f = \sup_{R\in\mathcal{E}} \mathbb{1}_R \frac{1}{\mu(R)} \int_R |f| \, d\mu.$$

The geometric condition alluded to previously is related to the restricted weak-type boundedness of $\mathcal{M}_{\mathcal{E}}^{\mu}$. In particular, if there exists an $0 < \eta < 1$ and $M > 0$ such that

$$\mu(\{x : \mathcal{M}_{\mathcal{E}}^{\mu}(\mathbb{1}_E)(x) > \eta\}) \leq M\mu(E) \tag{1.4}$$

for all measurable sets E, then the strategy described above leads to an algorithm that finds functions φ_R satisfying

$$\int_R \varphi_R \, d\mu \gtrsim \|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}^{-1} \mu(R),$$

where the implied constant depends only on M.

The inequality (1.4) would follow from the restricted weak-type $(1,1)$ of $\mathcal{M}_{\mathcal{E}}^{\mu}$, so for example it holds for the Lebesgue measure when $\mathcal{E}$ consists of axis-parallel rectangles, cubes, balls, etc. One can also consider other measures; for example, in the one-parameter dyadic case $\mathcal{M}^{\mu}$ is weak-type $(1,1)$-bounded for any measure μ, so (1.4) is true whenever $\mathcal{E}$ consists of dyadic intervals (or squares, cubes, etc.). In two or more parameters the weak type fails for general measures, but does hold when $d\mu(x) = w(x)\,dx$ and w is a strong A_∞ weight, as shown by R. Fefferman in [**6**].

Our algorithm finds the functions φ_R in Definition 1 assuming only that (1.4) holds, so it immediately gives a constructive proof of the equivalence between the sparse and Carleson conditions. Without (1.4) we can constructively prove the equivalence, but only up to a logarithmic factor.

As another application of the algorithm we can prove the following structural property of sparse collections.

Theorem A. *Let $\mathcal{E}$ be a finite collection of sets and suppose* (1.4) *holds. Then there exists a partition into $\mathcal{O}(\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)})$ subcollections $\{\mathcal{E}_i\}$ satisfying*

$$\|\mathcal{E}_i\|_{\mathrm{Carleson}(\mu)} \lesssim 1.$$

This result is proved as a special case of Theorem 4.1, which is a more precise version where we track all the constants. As an application of Theorem 4.1 we can draw a connection with the notion of (P_1) sequence introduced in [**4**]. In particular, we can show that, under the same geometric hypothesis as Theorem A, every Carleson collection can be split into a finite number of (P_1) sequences. We will defer the definition of (P_1) sequence until Section 4, where this connection is explained in Remark 4.2.

We also show that there are situations where such a splitting is impossible in the absence of an estimate on $\mathcal{M}_{\mathcal{E}}^{\mu}$. In particular we have, if μ is the Lebesgue measure on $\mathbb{R}$:

Theorem B. *For every $\Lambda \geq 2$ and every integer $N \geq 1$ there exists a collection $\mathcal{E}$ of subsets of $\mathbb{R}$ with $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \leq \Lambda$ such that for any partition*

$$\mathcal{E} = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_N$$

there exists at least one $i \in \{1,\dots,N\}$ for which $\|\mathcal{E}_i\|_{\mathrm{Carleson}(\mu)} \gtrsim \Lambda$.

Even if we restrict $\mathcal{E}$ to consist of only dyadic rectangles, we can produce an example similar to this last one, but for a specially designed measure μ for which (1.4) does not hold.

Another application of the algorithms described in this article is that we can weaken the definition of the Carleson condition to require only weak-type instead of strong L^1 estimates.

Theorem C. *Suppose* (1.4) *holds and let* $M<\infty$ *be the constant in the inequality. Then*

$$\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \lesssim \sup\left\{\frac{1}{\mu(\mathcal{F})}\left\|\sum_{R\in\mathcal{F}} \mathbb{1}_R\right\|_{L^{1,\infty}(\mu)} : \mathcal{F}\subseteq\mathcal{E}\right\},$$

where the implied constant depends only on M.

In Section 2 we describe a general algorithm to approximate the Carleson constant of a collection following the strategy described here.

In Section 3 we show how to modify the algorithm from the previous section to remove the logarithmic loss, conditional on inequality (1.4). Theorem C is proved at the end of this section.

Finally, in Section 4 we prove Theorem A by inductively constructing said partition, and prove Theorem B as well as the dyadic version with explicit examples.

2. An algorithm for general collections

In this section μ will always denote a fixed positive measure. All sets will also be assumed to be of finite μ-measure. For any collection $\mathcal{E}$ of sets define its *height function*

$$h_{\mathcal{E}} = \sum_{R\in\mathcal{E}} \mathbb{1}_R.$$

Carleson's condition asserts a uniform bound on the average height of all subcollections. Indeed, if we denote the average height by

$$\Lambda_\mu(\mathcal{E}) = \frac{1}{\mu(\mathrm{sh}(\mathcal{E}))}\int h_{\mathcal{E}}\,d\mu,$$

then Carleson's condition becomes $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} = \sup\{\Lambda_\mu(\mathcal{F}) : \mathcal{F}\subseteq\mathcal{E}\}$.

The next lemma is the main iteration step in our algorithm.

Lemma 2.1. *Let* $\mathcal{E}$ *be a collection of sets and suppose* $\Lambda := \Lambda_\mu(\mathcal{E}) < \infty$. *Define*[1] *for every* R *in* $\mathcal{E}$

$$g_R := \frac{\mathbb{1}_R}{h_{\mathcal{E}}}\mathbb{1}_{\{x : h_{\mathcal{E}}(x)\leq 2\Lambda\}}.$$

Then there exists at least one R *in* $\mathcal{E}$ *such that*

$$\int_R g_R\,d\mu \geq \frac{1}{2\Lambda}\mu(R). \tag{2.1}$$

Proof: By definition we have

$$\frac{1}{\Lambda}\int h_{\mathcal{E}}\,d\mu = \mu(\mathrm{sh}(\mathcal{E})). \tag{2.2}$$

Define the set $G=\{x\in \mathrm{sh}(\mathcal{E}) : h_{\mathcal{E}}(x)\leq 2\Lambda\}$ and note that with this notation we have:

$$g_R = \mathbb{1}_G\frac{\mathbb{1}_R}{h_{\mathcal{E}}} \quad\text{and}\quad \sum_{R\in\mathcal{E}} g_R = \mathbb{1}_G.$$

[1] Here and throughout we will take the convention that $\frac{0}{0}=0$.

If $\mu(G) = \mu(\mathrm{sh}(\mathcal{E}))$, then we are done, since then if (2.1) failed for all R, we would have

$$\mu(\mathrm{sh}(\mathcal{E})) = \mu(G) = \sum_{R\in\mathcal{E}} \int g_R \, d\mu < \frac{1}{2\Lambda} \sum_{R\in\mathcal{E}} \mu(R) = \frac{1}{2}\mu(\mathrm{sh}(\mathcal{E})),$$

which is a contradiction. So we can assume that $\mu(G) < \mu(\mathrm{sh}(\mathcal{E}))$.

Then, from Markov's inequality and (2.2) we can estimate

$$\begin{aligned} \mu(G) &= \mu(\mathrm{sh}(\mathcal{E})) - \mu(\{x : h_{\mathcal{E}} > 2\Lambda\}) \\ &> \mu(\mathrm{sh}(\mathcal{E})) - \frac{1}{2\Lambda}\int h_{\mathcal{E}} \, d\mu \\ &= \frac{1}{2}\mu(\mathrm{sh}(\mathcal{E})), \end{aligned}$$

and hence $\mu(G) > \frac{1}{2}\mu(\mathrm{sh}(\mathcal{E}))$.

Suppose by way of contradiction that (2.1) fails for all R. That is, for all R in $\mathcal{E}$,

$$\int_R g_R \, d\mu < \frac{1}{2\Lambda}\mu(R).$$

Then

$$\begin{aligned} \mu(G) = \int \mathbb{1}_G \, d\mu &= \int \sum_{R\in\mathcal{E}} g_R \, d\mu \\ &< \sum_{R\in\mathcal{E}} \frac{1}{2\Lambda}\mu(R) = \frac{1}{2\Lambda}\int h_{\mathcal{E}} \, d\mu = \frac{1}{2}\mu(\mathrm{sh}(\mathcal{E})). \end{aligned}$$

This means $\mu(G) < \frac{1}{2}\mu(\mathrm{sh}(\mathcal{E}))$, which is a contradiction. □

If we iterate this lemma, we obtain the algorithm described in the introduction.

Algorithm 1: ApproximateCarleson($\mathcal{E}$)

1 **begin**
2   Set $A = 1$.
3 **end**
4 **while** $\mathcal{E} \neq \emptyset$ **do**
5   Set $A = \max(A, \Lambda_\mu(\mathcal{E}))$.
6   Construct the functions g_R as in Lemma 2.1.
7   **for** $R \in \mathcal{E}$ **do**
8   **if** $\int g_R \, d\mu \geq \frac{\mu(R)}{2\Lambda_\mu(\mathcal{E})}$, **then**
9   Assign $f_R := g_R$ and $\Lambda_R := \Lambda_\mu(\mathcal{E})$.
10   Remove R from $\mathcal{E}$.
11   Go to line 5.
12   **end**
13   **end**
14 **end**

Result: The constant A, the sequence $\{\Lambda_R\}$, and the functions $\{f_R\}$.

The purpose of lines 7 through 13 is to find R so that

$$\int g_R \, d\mu \geq \frac{\mu(R)}{2\Lambda_\mu(\mathcal{E})}.$$

Lemma 2.1 shows that such an R always exists, so the algorithm removes one element from $\mathcal{E}$ at a time and thus always terminates. We note here that the functions f_R found by the algorithm are *not* the functions φ_R in the definition of sparse collection; those will be constructed later in Theorem 2.3.

Remark 2.2. We note that the symbols A, f_R, Λ_R, and $\mathcal{E}$ all change during the execution of the algorithm. One can rewrite the algorithm in a more traditional way by setting $\mathcal{E}_0 = \mathcal{E}$ and $A_0 = 1$ at the beginning. Then $\mathcal{E}_{n+1}$ is obtained from $\mathcal{E}_n$ by removing an R that satisfies the condition in line 8 and updating A_{n+1} to $\max(A_n, \Lambda_\mu(\mathcal{E}_{n+1}))$.

The following theorem shows that the approximation of $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$, namely A, is correct up to a logarithm.

Theorem 2.3. *Let A be the constant obtained as the result of running the algorithm on a collection $\mathcal{E}$. Then we have*

$$A \le \|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \le \|\mathcal{E}\|^{-1}_{\mathrm{Sparse}(\mu)} \lesssim A \log(e + A).$$

Proof: The inequality $A \le \|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$ is trivial since A is always one of the possible elements in the supremum of the definition of $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$. The second inequality is just Lemma 1.1, so we will focus only on showing the last inequality.

Suppose we could show

$$\sum_{R\in\mathcal{E}} f_R \le C \log(e + A). \tag{2.3}$$

Then we could set for each R in $\mathcal{E}$

$$\varphi_R := \frac{f_R}{C \log(e + A)}.$$

Now these functions obviously satisfy $\sum_{R\in\mathcal{E}} \varphi_R \le 1$ and

$$\begin{aligned} \int_R \varphi_R \, d\mu &= \frac{1}{C \log(e + A)} \int f_R \, d\mu \\ &\ge \frac{1}{C \log(e + A)} \frac{\mu(R)}{2A} \end{aligned}$$

for all R in $\mathcal{E}$. According to Definition 1, this would make $\mathcal{E}$ an η-sparse collection with

$$\eta = \frac{1}{2CA \log(e + A)}$$

and hence $\|\mathcal{E}\|^{-1}_{\mathrm{Sparse}(\mu)} \le 2CA \log(e + A)$.

We now proceed to prove (2.3). For any two $R, S \in \mathcal{E}$ set $R \prec S$ if and only if R was removed from $\mathcal{E}$ before S (in line 10). Define $\mathcal{E}_{\prec R} = \{S \in \mathcal{E} : S \prec R\}$ and $\mathcal{E}_{\succeq R} = \{S \in \mathcal{E} : S \succeq R\}$. Since $\prec$ is a total order, we have that $\mathcal{E} = \mathcal{E}_{\prec R} \sqcup \mathcal{E}_{\succeq R}$ for each R in $\mathcal{E}$. For $x \in \mathrm{sh}(\mathcal{E})$ set

$$\mathcal{B} = \{R \in \mathcal{E} : x \in \operatorname{supp} f_R\}$$

and let $(R_1, R_2, R_3, \dots)$ be the elements of $\mathcal{B}$ sorted in increasing order by $\prec$. Observe that the cardinality N of $\mathcal{B}$ satisfies on the one hand

$$N \le h_{\mathcal{E}_{\succeq R_1}}(x).$$

On the other hand, if x is in supp f_{R_1}, then $h_{\mathcal{E}_{\succeq R_1}}(x) \leq 2\Lambda_{R_1}$ and thus $N \leq 2\Lambda_{R_1} \leq 2A$. So

$$\sum_{R\in\mathcal{E}} f_R(x) = \sum_{n=1}^{N} f_{R_n}(x).$$

Note that by construction we have $f_{R_n}(x) \leq \frac{1}{1+N-n}$, therefore

$$\begin{aligned}\sum_{R\in\mathcal{E}} f_R(x) &\leq \sum_{n=1}^{N} \frac{1}{1+N-n}\\ &\lesssim \log(e+A)\end{aligned}$$

and we are done. □

3. An improvement with the maximal function

Recall the maximal operator associated to the family $\mathcal{E}$ and the measure μ from the introduction:

$$\mathcal{M}_{\mathcal{E}}^{\mu} f = \sup_{R\in\mathcal{E}} \frac{\mathbb{1}_R}{\mu(R)} \int_R |f|\, d\mu.$$

The measure μ will be fixed throughout this section, so we will abbreviate $\mathcal{M}_{\mathcal{E}} f := \mathcal{M}_{\mathcal{E}}^{\mu} f$.

We will show how Algorithm 1 can be slightly modified to give an essentially optimal approximation of $\|\mathcal{E}\|_{\mathrm{Carleson}}$ whenever $\mathcal{M}$ satisfies the condition in (1.4), which, we recall, was that for fixed $0<\eta<1$ and $M>0$

$$\mu(\{x : \mathcal{M}_{\mathcal{E}}(\mathbb{1}_E)(x) > \eta\}) \leq M\mu(E)$$

uniformly over all measurable sets E. We will denote by $M_\eta(\mathcal{E})$ the best constant in this inequality (again, dropping the dependence on μ for simplicity).

In the proof of Theorem 2.3 we showed how the logarithmic loss appears with Algorithm 1. In particular, dividing by $h_{\mathcal{E}}$ was needed in order to get a reasonably large value of $\int g_R$, which is where the logarithm appears as we end up having to sum the harmonic series. Dividing by a larger function would make the integral too small, while a smaller one makes bounding $\sum_R g_R$ harder.

Here we take a different approach. The idea is that, if $M_\eta(\mathcal{E})$ is finite, there must be a set R in $\mathcal{E}$ that intersects the high level set of $h_{\mathcal{E}}$ in only a small portion relative to itself. The next lemma is the main iteration step of the improved algorithm and is in the same spirit as Lemma 2.1.

Lemma 3.1. *Suppose that $M_\eta(\mathcal{E}) < \infty$ and*

$$\sup_{\lambda>0} \lambda\mu(\{x : h_{\mathcal{E}}(x) > \lambda\}) \leq \Lambda\mu(\mathrm{sh}(\mathcal{E})). \tag{3.1}$$

Then there must exist at least one R in $\mathcal{E}$ satisfying

$$\mu(\{x\in R : h_{\mathcal{E}}(x) \leq 2\Lambda M_\eta(\mathcal{E})\}) \geq (1-\eta)\mu(R). \tag{3.2}$$

Proof: To simplify the notation we will abbreviate $M := M_\eta(\mathcal{E})$. Suppose (3.2) does not hold for any R, that is: for every R in $\mathcal{E}$

$$\mu(\{x\in R : h_{\mathcal{E}}(x) \leq 2M\Lambda\}) < (1-\eta)\mu(R).$$

Then

$$\mu(\{x\in R : h_{\mathcal{E}}(x) > 2M\Lambda\}) = \mu(R) - \mu(\{x\in R : h_{\mathcal{E}}(x) \leq 2M\Lambda\}) > \eta\mu(R).$$

Set $B = \{x \in \mathrm{sh}(\mathcal{E}) : h_{\mathcal{E}}(x) > 2M\Lambda\}$. This estimate implies

$$\mu(B \cap R) > \eta\mu(R) \implies R \subseteq \{\mathcal{M}_{\mathcal{E}}(\mathbb{1}_B) > \eta\}.$$

Since (3.2) does not hold for any R we in fact have $\mathrm{sh}(\mathcal{E}) \subseteq \{\mathcal{M}_{\mathcal{E}}(\mathbb{1}_B) > \eta\}$.

By (3.1) we can estimate $\mu(B)$ from above as follows:

$$\mu(B) \le \frac{1}{2M\Lambda}\Lambda\mu(\mathrm{sh}(\mathcal{E})) = \frac{\mu(\mathrm{sh}(\mathcal{E}))}{2M}.$$

Thus, by the finiteness of M:

$$\mu(\mathrm{sh}(\mathcal{E})) \le \mu(\{\mathcal{M}_{\mathcal{E}}(\mathbb{1}_B) > \eta\}) \overset{(1.4)}{\le} M\mu(B) \le \frac{1}{2}\mu(\mathrm{sh}(\mathcal{E})),$$

which is a contradiction. □

Note that by Markov's inequality

$$\mu(\{x : h_{\mathcal{E}}(x) > \lambda\}) \le \lambda^{-1} \int h_{\mathcal{E}}\, d\mu = \lambda^{-1}\Lambda_\mu(\mathcal{E})\mu(\mathrm{sh}(\mathcal{E})).$$

So in particular condition (3.1) holds with $\Lambda \le \Lambda_\mu(\mathcal{E})$.

This lemma shows that one can find a set R in $\mathcal{E}$ with a large subset in which R is guaranteed to have bounded overlap with all the other sets in $\mathcal{E}$, in particular, if we set

$$F(R) = \{x \in R : h_{\mathcal{E}}(x) \le 2M_\eta(\mathcal{E})\Lambda_\mu^{1,\infty}(\mathcal{E})\},$$

where $\Lambda_\mu^{1,\infty}(\mathcal{E})$ is the best constant in (3.1). Then there must exist at least one R such that $\mu(F(R)) \ge (1-\eta)\mu(R)$.

We can now give the improved version of Algorithm 1:

Algorithm 2: ApproximateCarleson($\mathcal{E}$) - improved

begin
  Set $A = 1$.
end
while $\mathcal{E} \neq \emptyset$ **do**
  Set $A = \max(A, \Lambda_\mu^{1,\infty}(\mathcal{E}))$.
  for $R \in \mathcal{E}$ **do**
    if $\mu(F(R)) \ge (1-\eta)\mu(R)$, **then**
      Assign $E(R) := F(R)$ and $\Lambda_R := \Lambda_\mu^{1,\infty}(\mathcal{E})$.
      Remove R from $\mathcal{E}$.
    end
  end
end
Result: The constant A, the sequence $\{\Lambda_R\}$, and the sets $\{E(R)\}$.

As in the proof of Theorem 2.3, the order in which elements are removed from $\mathcal{E}$ is important. Set $R \prec S$ if and only if R was removed before S by Algorithm 2. Set also

$$\mathcal{E}_{\succeq R} = \{S \in \mathcal{E} : S \succeq R\}$$

with the natural definition of $\succeq$ in terms of $\prec$. The important property given by this order is the following inequality for the level sets of $h_{\mathcal{E}}$:

$$\mu(\{x \in R : h_{\mathcal{E}_{\succeq R}}(x) \le 2M\Lambda_\mu^{1,\infty}(\mathcal{E}_{\succeq R})\}) \ge (1-\eta)\mu(R), \tag{3.3}$$

where we have abbreviated $M = M_\eta(\mathcal{E})$. Note that $\Lambda_\mu^{1,\infty}(\mathcal{E}_{\succeq R}) \le A \le \|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$.

The next theorem shows that estimates like these imply upper bounds on the Carleson constant of $\mathcal{E}$.

Theorem 3.2. *Let $\mathcal{E}$ be a finite collection totally ordered by some binary relation $\prec$. Suppose that for some $\Lambda > 0$ and $0 < \eta \leq 1$ we have*

$$\mu(\{x \in R : h_{\mathcal{E}_{\succeq R}}(x) \leq \Lambda\}) \geq \eta\mu(R) \tag{3.4}$$

for all R in $\mathcal{E}$. Then $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \leq \|\mathcal{E}\|^{-1}_{\mathrm{Sparse}(\mu)} \leq \Lambda\eta^{-1}$.

Proof: We will show that $\|\mathcal{E}\|_{\mathrm{Sparse}(\mu)} \geq \eta\Lambda^{-1}$, which will provide the second inequality in the theorem; the first one is proved in Lemma 1.1.

For each R in $\mathcal{E}$ let $E(R) = \{x \in R : h_{\mathcal{E}_{\succeq R}}(x) \leq \Lambda\}$ and define the functions

$$\varphi_R = \frac{\mathbb{1}_{E(R)}}{\Lambda}.$$

By (3.4) we have

$$\int_R \varphi_R \, d\mu \geq \frac{\eta}{\Lambda}\mu(R),$$

which is (1.2) in Definition 1.

For any point x in $\mathrm{sh}(\mathcal{E})$ let $\mathcal{B}(x) = \{S \in \mathcal{E} : x \in E(S)\}$; then

$$\sum_{R \in \mathcal{E}} \varphi_R = \frac{\#(\mathcal{B}(x))}{\Lambda}.$$

So it suffices to show that $\mathcal{B}(x)$ has at most Λ elements as this will imply (1.3).

Let $N = \#(\mathcal{B}(x))$, and let R be the minimal element of $\mathcal{B}(x)$ with respect to $\prec$. Then obviously $h_{\mathcal{E}_{\succeq R}}(x) \geq N$. And since $x \in E(R)$, we must have $N \leq \Lambda$.

Thus, the functions $\{\varphi_R\}$ satisfy the conditions of Definition 1 and we are done. $\square$

Corollary 3.3. *If $M_\eta < \infty$ and A is the output constant of Algorithm 2, then*

$$A \leq \|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \leq \|\mathcal{E}\|^{-1}_{\mathrm{Sparse}(\mu)} \leq 2(1-\eta)^{-1}M_\eta A.$$

Proof: As in the proof of Theorem 2.3, the proof of the first inequality is trivial, and the second is just Lemma 1.1. Observe that $\Lambda^{1,\infty}_\mu(\mathcal{E}_{\succeq R}) \leq A$ for all R so (3.3) implies

$$\mu(\{x \in R : h_{\mathcal{E}_{\succeq R}}(x) \leq 2M_\eta A\}) \geq (1-\eta)\mu(R)$$

for all $R \in \mathcal{E}$. Thus, the last inequality follows from applying Theorem 3.2. $\square$

The fact that we only really needed the weak-type bound in (3.1) allows us to prove Theorem C:

Proof: Suppose

$$\|h_{\mathcal{F}}\|_{L^{1,\infty}(\mu)} \leq C_0\mu(\mathrm{sh}(\mathcal{F}))$$

for all $\mathcal{F} \subseteq \mathcal{E}$. This means that, at each iteration in the algorithm, (3.1) holds with $\Lambda \leq C_0$, thus the constant A output as a result is at most C_0 and the claim follows by Corollary 3.3. $\square$

4. Breaking up sparse collections

We are now ready to prove the structure theorem mentioned in the introduction.

Theorem 4.1. *Let $\mathcal{E}$ be a finite collection of sets with finite μ-measure, and suppose the maximal operator $\mathcal{M}^{\mu}_{\mathcal{E}}$ satisfies* (1.4) *with constant $M = M_\eta$.*

Then for any $0 < \gamma < 1 - \eta$ there exists a partition of $\mathcal{E}$ into at most

$$1 + \frac{2M(1-\eta)}{1-\eta-\gamma}\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$$

subcollections $\{\mathcal{E}_i\}$ satisfying

$$\|\mathcal{E}_i\|_{\mathrm{Sparse}(\mu)} \geq \gamma.$$

Proof: After applying Algorithm 2 and reversing the order, one obtains a total order $<$ on $\mathcal{E}$ such that

$$\mu(\{x \in R : h_{\mathcal{E}_{\leq R}}(x) > 2M\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}\}) \leq \eta\mu(R) \tag{4.1}$$

for all $R \in \mathcal{E}$.

Create N empty *buckets* $\{\mathcal{E}_1, \dots, \mathcal{E}_N\}$, where N is a large integer to be chosen later. These buckets will be constructed by iteratively inserting elements from $\mathcal{E}$.

We start with the smallest (with respect to $<$) element in $\mathcal{E}$, which we can insert into an arbitrary bucket, say $\mathcal{E}_1$. Let R be any set in $\mathcal{E}$ and assume that we have placed all the previous sets $S < R$ in the sets $\{\mathcal{E}_1, \dots, \mathcal{E}_N\}$. We will show that there must exist a bucket $\mathcal{E}_i$ such that

$$\mu(R \cap \mathrm{sh}(\mathcal{E}_i)) \leq (1-\gamma)\mu(R). \tag{4.2}$$

Indeed, suppose (4.2) fails for all the N buckets. Then, for $A_i = \mathrm{sh}(\mathcal{E}_i) \cap R$, we have

$$\mu(A_i) > (1-\gamma)\mu(R) \tag{4.3}$$

for all $i \in \{1, \dots, N\}$. Set

$$\alpha = \frac{1-\eta-\gamma}{1-\eta} \tag{4.4}$$

and let $U = \{x \in R : \sum_{i=1}^N \mathbb{1}_{A_i} \geq \alpha N\}$. We will show that

$$\mu(U) > \eta\mu(R).$$

This will contradict (4.1) if $N \geq \alpha^{-1}2M\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$, since we would have

$$\{x \in R : h_{\mathcal{E}_{\leq R}}(x) > 2M\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}\} \supseteq U.$$

To estimate $\mu(U)$ it is easier to bound the measure of the complement $V = R \setminus U$. If x is in fewer than αN of the subsets $\{A_i\}$, then x is in at least $(1-\alpha)N$ of the subsets $\{R \setminus A_i\}$. Thus

$$\begin{aligned}
\mu(V) &= \mu\left(\left\{x \in R : \sum_{i=1}^N \mathbb{1}_{A_i^c} > (1-\alpha)N\right\}\right) \\
&\leq \frac{1}{(1-\alpha)N}\int_R \sum_{i=1}^N \mathbb{1}_{A_i^c}\,d\mu \\
&= \frac{1}{(1-\alpha)N}\sum_{i=1}^N \mu(R \setminus A_i).
\end{aligned}$$

By (4.3) we have $\mu(R \setminus A_i) < \gamma\mu(R)$, so

$$\mu(V) < \frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\gamma\mu(R)$$

$$= \frac{\gamma}{1-\alpha}\mu(R).$$

Thus, with our choice of α in (4.4):

$$\begin{aligned}\mu(U) &= \mu(R) - \mu(V)\\ &> \mu(R)\left(1 - \frac{\gamma}{1-\alpha}\right)\\ &= \mu(R)\left(\frac{1-\alpha-\gamma}{1-\alpha}\right)\\ &= \eta\mu(R),\end{aligned}$$

which is our contradiction.

Finally, it remains to choose N, but this is easy as the smallest integer $N \geq \alpha^{-1}2M\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)}$ will suffice. □

Remark 4.2. We would like to note here that Theorem 4.1 proves that every finite Carleson collection of axis-parallel rectangles (or sets for which the associated maximal function satisfies (1.4)) can be decomposed into finitely many collections of type (P_1) in the nomenclature of [**4**]. Recall that a sequence $\{R_1, R_2, \dots\}$ is of type (P_1) if for every $n \geq 1$

$$|R_{n+1} \setminus (R_1 \cup \cdots \cup R_n)| \geq \frac{1}{2}|R_{n+1}|.$$

Observe that, in the proof of Theorem 4.1, the buckets $\mathcal{E}_i$ satisfy (4.2), which is exactly the (P_1) condition when $\gamma = \frac{1}{2}$.

We now show that the structure theorem is not true in general.

Theorem 4.3. *For every $\Lambda \geq 2$ and every integer $N \geq 1$ there exists a collection $\mathcal{E}$ of subsets of $\mathbb{R}$ with $\|\mathcal{E}\|_{\mathrm{Carleson}} \leq \Lambda$ such that for any partition*

$$\mathcal{E} = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_N$$

there exists at least one $i \in \{1, \dots, N\}$ for which $\|\mathcal{E}_i\|_{\mathrm{Carleson}} \geq \frac{1}{2}\Lambda$.

Proof: Fix a large integer M to be chosen later. For any integer $m \geq 1$ define the sets

$$R_m = [0, 1) \cup [m, m + (\Lambda - 1)^{-1}).$$

If $\mathcal{F}$ is any non-empty subcollection of $\{R_1, R_2, \dots\}$, then

$$|\mathrm{sh}(\mathcal{F})| = 1 + \#(\mathcal{F})(\Lambda - 1)^{-1}.$$

For each $m \geq 1$ let $E(R_m) = [m, m + (\Lambda - 1)^{-1})$; then the collection $\{E(R_m)\}$ is pairwise disjoint and

$$\frac{|E(R_m)|}{|R_m|} = \frac{(\Lambda-1)^{-1}}{1 + (\Lambda-1)^{-1}} = \frac{1}{\Lambda}.$$

These two facts mean that the collection $\{R_0, \dots, R_{M-1}\}$ is Λ^{-1}-sparse, and hence $\|\mathcal{E}\|_{\mathrm{Carleson}} \leq \Lambda$ by Lemma 1.1.

Now let $\mathcal{E}_1 \cup \cdots \cup \mathcal{E}_N$ be any partition of $\mathcal{E}$. Since $\#\mathcal{E} = M$, there must exist an $i \in \{1, \dots, N\}$ such that $\#\mathcal{E}_i \geq \frac{M}{N}$. For this family we have

$$\|\mathcal{E}_i\|_{\mathrm{Carleson}} \geq \Lambda_\mu(\mathcal{E}_i) = \frac{\#\mathcal{E}_i(1 + (\Lambda-1)^{-1})}{1 + \#\mathcal{E}_i(\Lambda-1)^{-1}}.$$

When M is sufficiently large (depending only on N and Λ) we have

$$\begin{aligned}\frac{\#\mathcal{E}_i(1+(\Lambda-1)^{-1})}{1+\#\mathcal{E}_i(\Lambda-1)^{-1}} &\geq \frac{1+(\Lambda-1)^{-1}}{2(\Lambda-1)^{-1}}\\ &= \frac{\Lambda}{2},\end{aligned}$$

which is what we wanted. □

One may wonder whether one can improve matters by imposing additional geometry on the sets contained in $\mathcal{E}$. For example, when $\mathcal{E}$ consists of dyadic rectangles in $\mathbb{R}^d$ then Theorem 4.1 applies. However, if one is allowed to change the measure, then we can construct an example that behaves like the one in Theorem 4.3.

The construction, which has essentially the same behavior as that of Theorem 4.3, is similar to one used by R. Fefferman in [**6**].

Theorem 4.4. *There exists a measure μ on $\mathbb{R}^d$ such that for any integers $N \geq 1$ and $\Lambda \geq 2$ there exists a finite collection $\mathcal{E}$ of dyadic rectangles with $\|\mathcal{E}\|_{\mathrm{Carleson}(\mu)} \leq \Lambda$ such that any partition into N subfamilies has at least one with Carleson constant $\geq \frac{1}{2}\Lambda$.*

Proof: For integers m and j consider the dyadic rectangles

$$R_m^j = [0, 2^m) \times [j, j+2^{-m})$$

and let $\mathcal{S}^j = \{R_m^j : m \geq 0\}$. Define also

$$E(R_m^j) = [2^{m-1}, 2^m) \times [j + 2^{-m-1}, j + 2^{-m}).$$

Observe that the sets$\{E(R_m^j)\}$ are pairwise disjoint.

Choose any set of points $\{x_m^j\}$ such that $x_m^j \in E(R_m^j)$ for every non-negative m and j. Then define the measure

$$\mu = \sum_{j=0}^{\infty}\left(\delta_{(0,j)} + \sum_{m=0}^{\infty}(1+j)^{-1}\delta_{x_m^j}\right).$$

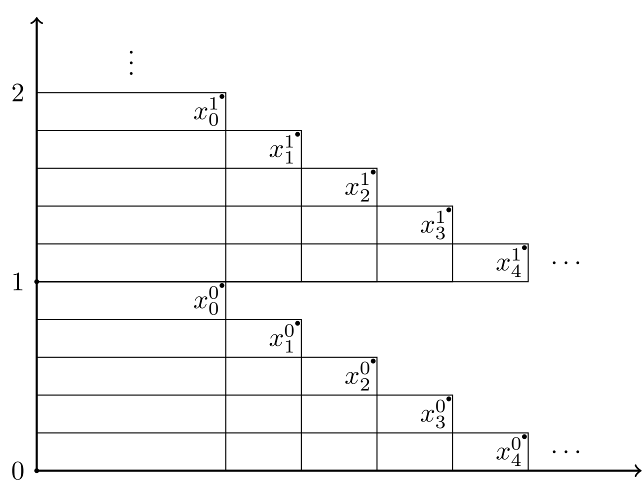

Figure 1. The first few rectangles $\{R_m^j\}$ and points $\{x_m^j\}$ (not to scale).

With this measure we have

$$\mu(R_m^j) = 1 + (1+j)^{-1} \quad \text{and} \quad \mu(E(R_m^j)) = (1+j)^{-1},$$

for all m and j. As in the proof of Theorem 4.3, for any finite collection $\mathcal{F} \subset \mathcal{S}^j$ we have

$$\|\mathcal{F}\|_{\mathrm{Carleson}(\mu)} \leq \frac{1+(1+j)^{-1}}{(1+j)^{-1}} = 2+j, \tag{4.5}$$

$$\|\mathcal{F}\|_{\mathrm{Carleson}(\mu)} \geq \frac{\#(\mathcal{F})(1+(1+j)^{-1})}{1+\#(\mathcal{F})(1+j)^{-1}} = \frac{\#(\mathcal{F})(2+j)}{1+j+\#(\mathcal{F})}. \tag{4.6}$$

Let M be a large integer and take any subset $\mathcal{E}$ from $\mathcal{S}^{\Lambda-2}$ with $\#(\mathcal{E}) = M$. By (4.5) we can bound the Carleson constant of $\mathcal{E}$ by Λ. Suppose $\{\mathcal{E}_i\}$ is any partition of $\mathcal{E}$ into N subfamilies. There must exist at least one subfamily, say $\mathcal{E}_i$, with $\#(\mathcal{E}_i) \geq M/N$. For this subfamily we have by (4.6)

$$\begin{aligned}\|\mathcal{E}_i\|_{\mathrm{Carleson}(\mu)} &\geq \frac{\frac{M}{N}\Lambda}{\Lambda - 1 + \frac{M}{N}} \\ &= \frac{M\Lambda}{N(\Lambda-1)+M}.\end{aligned}$$

The claim follows by taking M so large that

$$\frac{M\Lambda}{N(\Lambda-1)+M} \geq \frac{\Lambda}{2}.$$

$\square$

Acknowledgements

The author would like to thank the referees for their thoughtful comments and many helpful suggestions.

Supported by Grant MICIN/AEI/PID2019-105599GB-I00.

References

[1] A. Barron, Sparse bounds in harmonic analysis and semiperiodic estimates, Thesis (Ph.D.)-Brown University (2019). Available on `https://repository.library.brown.edu/studio/item/bdr:918918/PDF/`.

[2] L. Carleson, *A Counter Example for Neasures Bounded on H_p for the Bi-disc*, Report no. 7 - 1974, Institut Mittag-Leffler, 1974.

[3] A. Córdoba, Maximal functions, covering lemmas and Fourier multipliers, in: *Harmonic Analysis in Euclidean Spaces* (Proc. Sympos. Pure Math., Williams Coll., Williamstown, Mass., 1978), Part 1, Proc. Sympos. Pure Math. **35.1**, American Mathematical Society, Providence, RI, 1979, pp. 29–50. DOI: 10.1090/pspum/035.1.

[4] A. Córdoba and R. Fefferman, A geometric proof of the strong maximal theorem, *Ann. of Math. (2)* **102(1)** (1975), 95–100. DOI: 10.2307/1970976.

[5] L. E. Dor, On projections in L_1, *Ann. of Math. (2)* **102(3)** (1975), 463–474. DOI: 10.2307/1971039.

[6] R. Fefferman, Strong differentiation with respect to measures, *Amer. J. Math.* **103(1)** (1981), 33–40. DOI: 10.2307/2374188.

[7] T. S. Hänninen, Equivalence of sparse and Carleson coefficients for general sets, *Ark. Mat.* **56(2)** (2018), 333–339. DOI: 10.4310/ARKIV.2018.v56.n2.a8.

[8] A. K. Lerner and F. Nazarov, Intuitive dyadic calculus: the basics, *Expo. Math.* **37(3)** (2019), 225–265. DOI: 10.1016/j.exmath.2018.01.001.

[9] C. Muscalu and W. Schlag, *Classical and Multilinear Harmonic Analysis. Vol. II*, Cambridge Stud. Adv. Math. **138**, Cambridge University Press, Cambridge, 2013.

[10] M. C. Pereyra, Dyadic harmonic analysis and weighted inequalities: the sparse revolution, in: *New Trends in Applied Harmonic Analysis, Vol. 2. Harmonic Analysis, Geometric Measure Theory, and Applications*, Appl. Numer. Harmon. Anal., Birkhäuser/Springer, Cham, 2019, pp. 159–239. DOI: 10.1007/978-3-030-32353-0_7.

[11] T. Tao, Dyadic product H^1, BMO, and Carleson's counterexample, Preprint. http://www.math.ucla.edu/~tao/preprints/harmonic.html.

[12] I. E. Verbitsky, Imbedding and multiplier theorems for discrete Littlewood–Paley spaces, *Pacific J. Math.* **176(2)** (1996), 529–556. DOI: 10.2140/PJM.1996.176.529.

Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain
E-mail address: guillermo.rey@uam.es

Received on February 23, 2022.
Accepted on July 20, 2022.