Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)
Vol. XX (2020), 415-435

Sparse bounds for maximally truncated oscillatory singular integrals

BEN KRAUSE AND MICHAEL T. LACEY

Abstract. For a polynomial $P(x, y)$, and any Calderón-Zygmund kernel K, the operator below satisfies a $(1, r)$ sparse bound, for $1 < r \leq 2$:

$$\sup_{\epsilon>0} \left| \int_{|y|>\epsilon} f(x-y) e^{2\pi i P(x,y)} K(y)\, dy \right|.$$

The implied bound depends upon $P(x, y)$ only through the degree of P. We derive from this a range of weighted inequalities, including weak type inequalities on $L^1(w)$, which are new, even in the unweighted case. The unweighted weak-type estimate, without maximal truncations, is due to Chanillo and Christ (1987).

Mathematics Subject Classification (2010): 42B20.

1. Introduction

The Ricci-Stein [20, 21] theory of oscillatory singular integrals concern operators of the form

$$T_P f(x) = \int e(P(x, y)) K(y) f(x-y)\, dy, \qquad e(t) := e^{2\pi i t}$$

where $K(y)$ is a Calderón-Zygmund kernel on $\mathbb{R}^n$, and $P : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a polynomial of two variables. These operators are bounded on all L^p, with bounds that only depend upon the degree of the polynomial, the dimension, and the kernel K, an important point in the motivations for this theory.

Theorem A. *Under the assumptions above, there holds for a finite constant*

$$\sup_{\deg(P)=d} \| T_P : L^p \mapsto L^p \| < \infty, \qquad 1 < p < \infty.$$

Research supported in part by an NSF Postdoctoral Research Fellowship.
Research supported in part by grant NSF-DMS-1600693.
Received June 18, 2017; accepted in revised form May 18, 2018.
Published online May 2020.

The L^1 theory is more delicate, with the dominant result being that of Chanillo and Christ [2] proving that the operators T_P indeed map L^1 into weak L^1, with again the bound depending only on the degree of P, the dimension, and the kernel K. Their argument does not address maximal truncations.

Our main result proves sparse bounds for the maximal truncations

$$T_{P,*}f(x) = \sup_{\epsilon>0}\left|\int_{|y|>\epsilon} f(x-y)e(P(x,y))K(y)\,dy\right|.$$

The bound that we prove implies the weak L^1 bounds for the maximal truncations, as well as quantitative bounds in A_p, for $1 \le p < \infty$.

Call a collection of cubes $\mathcal{S}$ in $\mathbb{R}^n$ *sparse* if there are sets $\{E_S : S \in \mathcal{S}\}$ which are pairwise disjoint, $E_S \subset S$ and satisfy $|E_S| > \frac{1}{4}|S|$ for all $S \in \mathcal{S}$. For any cube I and $1 \le r < \infty$, set $\langle f\rangle_{I,r}^r = |I|^{-1}\int_I |f|^r\,dx$. Then the (r,s)-sparse form $\Lambda_{\mathcal{S},r,s} = \Lambda_{r,s}$, indexed by the sparse collection $\mathcal{S}$ is

$$\Lambda_{S,r,s}(f,g) = \sum_{I\in\mathcal{S}} |I|\langle f\rangle_{I,r}\langle g\rangle_{I,s}.$$

Given a sublinear operator T, and $1 \le r, s < \infty$, we set $\|T : (r,s)\|$ to be the infimum over constants C so that for all all bounded compactly supported functions f, g,

$$|\langle Tf, g\rangle| \le C \sup \Lambda_{r,s}(f,g),$$

where the supremum is over all sparse forms. It is essential that the sparse form be allowed to depend upon f and g. But the point is that the sparse form itself varies over a class of operators with very nice properties.

For singular integrals without oscillatory terms we have:

Theorem B ([6,16]). *Let K be a Calderón-Zygmund kernel on $\mathbb{R}^n$ as above. Then, the operator $Tf =$ p.v. $K * f(x)$ satisfies $\|T : (1,1)\| < \infty$.*

Below, we obtain a quantitative version of a conjecture from [17].

Theorem 1.1. *For all integers d, and $1 < r < 2$, there holds*

$$\sup_{\deg(P)\le d} \|T_{P,*} : (1,r)\| \lesssim \frac{1}{r-1}. \tag{1.1}$$

The implied constant depends upon the degree d, the dimension n, and the kernel K, but is otherwise absolute.

As a corollary, we have these quantitative weighted inequalities. The inequalities (1.2) are new even for Lebesgue measure. The case of no truncations has been addressed in [12,22], but without effective bounds in terms of the A_p characteristic.

Corollary 1.2. *For every $d \geq 2$ and weight $w \in A_1$ there holds*

$$\begin{aligned} \sup_{\deg(P)\leq d} \left\| T_{P,*} : L^1(w) \mapsto L^{1,\infty}(w) \right\| &\leq [w]_{A_1}^2 \log_+ [w]_{A_1}, \\ \sup_{\deg(P)\leq d} \left\| T_{P,*} : L^p(w) \mapsto L^p(w) \right\| &\lesssim [w]_{A_p}^{1+\max\{\frac{1}{p-1},1\}}, \quad 1 < p < \infty. \end{aligned} \tag{1.2}$$

Ricci-Stein theory has several interesting extensions. On the one hand, there are several variants on the main result of [25], which considers an estimate which is uniform over the polynomials P. See [18,19]. One also has weighted extensions of the inequalities in for instance [11].

Sparse bounds have recently been quite active research topic, impacting a range of operators. We point to the previously cited [6,16]. But also point to the range of operators addressed in [1,7,8,10,13,15].

Our quantitative sparse bound in Theorem 1.1 closely matches the bounds obtained for "rough" singular integrals by Conde, Culiuc, Di Plinio and Ou [5]. Their argument is a beautiful abstraction of the methods of Christ and others [3,4,23]. A large part of our argument can be seen as an extension of [5].

But, the oscillatory nature of the kernels present substantial difficulties, and additional new arguments are required to address maximal truncations. Our prior paper [14] proved the theorem above in the special case of $P(y) = y^d$ in one dimension, and the interested reader will find that argument has fewer complications than this one.

(1) The essential oscillatory nature of the question is captured in Lemma 2.3. It has two estimates, the first (2.2) being well-known, having its origins in the work of Ricci and Stein [20]. The second, (2.4) is the additional feature needed to understand the L^1 endpoint. It is proved with the aid of arguments that can be found in the work of Christ and Chanillo [2];
(2) The essential fact needed is the partial sparse bound of Lemma 3.2. This is a kind of 'L^1 improving' estimate. The initial steps in the proof of this lemma depend upon a standard Calderón-Zygmund decomposition, with additional tweaks of the argument to adapt to the oscillatory estimates;
(3) The further additional fact is Lemma 4.3. Crucially, a the Carleson measure estimate is proved, which allows one to control the number of scales that impact this lemma. The maximal truncations are then controlled by orthogonality considerations, and a general form of the Rademacher-Menshov theorem.

1.1. Notation

As mentioned previously, here and throughout we use $e(t) := e^{2\pi i t}$; M_{HL} denotes the Hardy-Littlewood maximal function. For cubes $I \subset \mathbb{R}^n$, we let $\ell(I) := |I|^{1/n}$ denote its side-length.

With $d \geq 1$, we fix throughout the constant

$$\epsilon_d := \frac{1}{2d}. \tag{1.3}$$

We use multi-index notation, $\alpha = (\alpha_1, \ldots, \alpha_n)$, so

$$x^\alpha := \prod_{i=1}^{n} x_i^{\alpha_i}.$$

We use $\alpha > \beta$ to mean that $\alpha_i \geq \beta_i$ for each $1 \leq i \leq n$ with at least one inequality being strict.

Recall that a Calderón-Zygmund kernel K on $\mathbb{R}$, satisfies the following properties:

- K is a tempered distribution which agrees with a C^1 function $K(x)$ for $x \neq 0$;
- $\hat{K}$, the Fourier transform of K, is an L^∞ function;
- $|\partial^\alpha K(x)| \lesssim |x|^{-n-|\alpha|}$ for each multi-index $0 \leq |\alpha| \leq 1$. (We recall multi-index notation in the subsection on notation below.)

The key property of such kernels K that we shall use is that we may decompose

$$K(x) = \sum_{j=-\infty}^{\infty} 2^{-nj} \psi_j(2^{-j}x), \ x \neq 0 \tag{1.4}$$

where ψ_j are each C^1 functions supported in $\{\frac{1}{4} < |x| \leq 1\}$, which satisfy

$$|\partial^\alpha \psi_j| \leq C \ \text{ for each multi-index } 0 \leq |\alpha| \leq 1 \text{ uniformly in } j$$

and have zero mean, $\int \psi_j(x) = 0$. This decomposition is presented in [24, Chapter 13].

We will make use of the modified Vinogradov notation. We use $X \lesssim Y$, or $Y \gtrsim X$ to denote the estimate $X \leq CY$ for an absolute constant C. If we need C to depend on a parameter, we shall indicate this by subscripts, thus for instance $X \lesssim_p Y$ denotes the estimate $X \leq C_p Y$ for some C_p depending on p. We use $X \approx Y$ as shorthand for $Y \lesssim X \lesssim Y$.

ACKNOWLEDGEMENTS. We thank the referee for a careful reading.

2. Lemmas

There are two categories of facts collected here, (a) those which reflect the oscillatory nature of the problem, (b) a variant of the Rademacher-Menshov theorem.

2.1. Oscillatory Estimates

This is a variant of the van der Corput lemma.

Lemma 2.1 ([25, Proposition 2.1]). *Suppose $\Omega \subset \{|x| \leq 1\}$ is a convex set, and $P(t) := \sum_{|\alpha| \leq d} \lambda_\alpha t^\alpha$ is a real polynomial, equipped with the coefficient norm, $\|P\| := \sum_{1 \leq |\alpha| \leq d} |\lambda_\alpha|$. Then for any C^1 function ϕ,*

$$\left| \int_\Omega e(P(t))\phi(t)\, dt \right| \lesssim \|P\|^{-1/d} \left(\sup_{|x| \leq 1} |\phi(x)| + \sup_{|x| \leq 1} |\nabla \phi(x)| \right).$$

Next, a sublevel set estimate.

Lemma 2.2 ([25, Proposition 2.2]). *We have the estimate below.*

$$|\{|x| \leq 1 : |P(x)| < \epsilon\}| \lesssim \epsilon^{1/d} \|P\|^{-1/d}.$$

Using a simple change of variables we have for any cube I, any convex set $\Omega \subset I$, and any C^1 function φ,

$$\begin{aligned} &\left| \int_\Omega e(P(x))\varphi(x)\, dx \right| \\ &\lesssim |I| \left(\sup_{x \in I} |\varphi(x)| + \ell(I) \sup_{x \in I} |\nabla \varphi(x)| \right) \left[\sum_\alpha \ell(I)^{|\alpha|} |\lambda_\alpha| \right]^{-1/d}, \qquad (2.1) \\ &|\{x \in I : |P(x)| < \epsilon\}| \lesssim |I| \epsilon^{1/d} \left[\sum_\alpha \ell(I)^{|\alpha|} |\lambda_\alpha| \right]^{-1/d}. \end{aligned}$$

Here, we let $\ell(I)$ denote the side length of the cube I.

We will be concerned with operators that have kernels

$$\phi_k(x, y) = 2^{-nk} e(P(x, y)) \psi_k(2^{-k} y), \qquad k \in \mathbb{Z}.$$

Above, the ψ_k are as in (1.4). This next lemma is the essential oscillatory fact, concerning a T^*T estimate for convolution with respect to ϕ_k. Note that it holds for polynomials with no constant or linear term.

Lemma 2.3. *Assume that*

(a) *$d \geq 2$;*
(b) *The polynomial P does not have constant or linear terms, and is not solely a function of x;*
(c) *$\|P\| \geq 1$, and $k \geq t_n$, for a dimensional constant t_n.*

For each cube K with $\ell(K) = 2^k$, there is a set $Z_K \subset K \times K$ so that these three conditions hold.

(1) *For all* $t_n \le j \le k$, *we have*

$$\begin{aligned} &\mathbf{1}_{K\times K}(x,y)\left|\int \phi_k(x,z)\overline{\phi_j(z,y)}\,dz\right| \\ &\le C_0\left\{2^{-nk}\mathbf{1}_{Z_K}(x,y)+2^{-(n+\epsilon_d)k}\mathbf{1}_{K\times K}(x,y)\right\}. \end{aligned} \tag{2.2}$$

Above, $\tilde{\phi}(x,y)=\overline{\phi}(-y)$, *and* ϵ_d *is as in* (1.3)*;*

(2) *The sets* Z_K *have the following "small-neighborhoods" property: for any* $1 \le 2^s \le 2^k$,

$$|Z_K+\{|(x,y)|\le 2^s\}| \lesssim |K|^2\big(2^{-\epsilon_d k}+2^{s-k}\big); \tag{2.3}$$

here we are taking the Minkowski sum of Z_K *and the* 2^s *ball;*

(3) *In fact, the sets* Z_K *satisfy the fiber-wise estimate*

$$\sup_{x\in K}|\pi_x Z_K+\{(0,y):|y|\le 2^s\}| \lesssim 2^{nk}\big(2^{-\epsilon_d k}+2^{s-k}\big), \quad 1\le 2^s\le 2^k, \tag{2.4}$$

where $\pi_x Z=\{y:(x,y)\in Z\}$, *is the* x*-fiber of* Z, *and in* (2.4) *we are taking the Minkowski sum of* Z_K *and a ball in* $\mathbb{R}^n$, *and measure is taken in* $\mathbb{R}^n$.

The estimate (2.2) is uniform in $1 \le j \le k$, and the right side has two terms. The second term is the one in which we have additional decay in the convolution, beyond what we naively expect. The first term involving Z_K is that term for which we do not claim any additional decay in the convolution. Thus, additional information about the set Z_K is needed, which is the content of (2.4). Part of this information is well-known: Z_K has small measure (2.3). The more refined information in (2.4) is that *on each fiber*, the measure of a neighborhood of the set is small. This condition is not formulated by Chanillo and Christ [2], but follows from their techniques, which we present below.

Proof. Write the polynomial $P(x,y)$ as

$$P(x,y)=\sum_{\alpha,\beta:|\alpha|+|\beta|\ge 2,\ \beta\ne 0}\lambda_{\alpha,\beta}x^\alpha y^\beta,$$

where $\|\lambda\|=\sum_{\alpha,\beta}|\lambda_{\alpha,\beta}|=1$. We will use the van der Corput estimate (2.1) to estimate the integral in z in (2.2). The integral in (2.2) is explicitly

$$2^{-jn-kn}\int e(P(x,z)-P(y,z))\psi_j(2^{-j}(z-y))\psi_k(2^{-k}(x-z))\,dy.$$

Write $P(x,y)=\sum_{\alpha,\beta}\lambda_{\alpha,\beta}x^\alpha y^\beta$ the phase function above as

$$P(x,z)-P(y,z)=\sum_{\beta:|\beta|>0}\big[R_\beta(x)-R_\beta(y)\big]z^\beta,$$

$$\text{where}\quad R_\beta(x)=\sum_{\alpha:|\alpha|+|\beta|>1}\lambda_{\alpha,\beta}x^\alpha.$$

Above we have $|\alpha| \geq 1$. Then, by (2.1),

$$\text{LHS of (2.2)} \lesssim |I|^{-1} \Bigg[\sum_{\beta:|\beta|\geq 1} (\ell I)^{|\beta|} \big| R_\beta(x) - R_\beta(y) \big| \Bigg]^{-1/d}.$$

Therefore, we take the set Z_I to be

$$Z_I = \Bigg\{ (x, y) \in I \times I : \sum_\beta (\ell I)^{|\beta|} \big| R_\beta(x) - R_\beta(y) \big| < 2^{k/2} \Bigg\}. \tag{2.5}$$

We see that that (2.2) holds. This completes the first part of the conclusion of the lemma.

For the second part, the set Z_I in (2.5) is contained in the set $\bigcup_\sigma Z(\sigma)$, where

$$Z(\sigma) := \Bigg\{ (x, y) \in I \times I : \Bigg| \sum_\beta \sigma(\beta) \ell(I)^{|\beta|} \big[R_\beta(x) - R_\beta(y) \big] \Bigg| < 2^{k/2} \Bigg\}.$$

The union is over all choices of signs $\sigma : \{\beta\} \mapsto \{\pm 1\}$.There are $O(2^{nd}) = O(1)$ such choices of σ. Fixing σ, the polynomial of two variables $\sum_\beta \sigma(\beta)\ell(I)^{|\beta|}[R_\beta(x) - R_\beta(y)]$ has norm at least one, since $\|P\| \geq 1$; similarly, if we fix x as well, the polynomial of one variable $\sum_\beta \sigma(\beta)\ell(I)^{|\beta|} R_\beta(y)$ has norm at least one. It follows from Lemma 2.4 below, applied in dimensions $2n$ and n, that the set $Z(\sigma)$ satisfies the estimates (2.3) and (2.4). That completes the proof. □

This is the main point that remains to be addressed.

Lemma 2.4. *Let $P = P(x)$ be a polynomial on $\mathbb{R}^n$ with $\|P\| \geq 1$ and degree d. Let I be a cube of side length $\ell(I) = 2^k \geq 2^{t_n}$, and $2^{t_n} \leq 2^s \leq 2^k$. We have the estimate*

$$|Z_I + \{x : |x| \leq 2^s\}| \lesssim |I| \big\{ 2^{-\epsilon_d k/2} + 2^{s-k} \big\}, \tag{2.6}$$

where $Z_I = \{x \in I : |P(x)| < \ell(I)^{1/2}\}$.

The case of dimension $n = 1$ is easy. The set Z_I has small measure by the van der Corput estimate (2.1). But, it is the pre-image of an interval under a degree d polynomial P. Hence it has $O(d) = O(1)$ components. From this, (2.6) is immediate.

The higher dimensional case requires some additional insights, because level sets in two and higher dimensions are, in general, unbounded algebraic varieties. We need the following lemma, drawn from Chanillo and Christ [2]. By a k-strip we mean the set

$$S = \bigcup_{j \in \mathbb{Z}} Q + 2^k(0, \ldots, 0, j), \qquad Q \text{ is a cube.}$$

By a k-interval we mean a (possibly infinite) subset of S given by

$$I = \bigcup_{j_0 < j < j_1} Q + 2^k(0, \ldots, 0, j), \quad \text{for } j_0, j_1 \in \{\pm\infty\} \cup \mathbb{Z}.$$

Lemma 2.5 ([2], Lemma 4.2). *For any dimension n and degree d, there is a $C \lesssim_{d,n} 1$ so that for any $A > 0$, and any polynomial P of degree d, and any k-strip S, the subset of S given by*

$$\bigcup \left\{ Q \in \mathcal{D}_k^{\vec{\omega}} : Q \subset S, \, Q \cap \{|P(x)| < A\} \neq \emptyset \right\}$$

is a union of at most C k-intervals.

The lemma above is proven for $k = 0$, but as the result in [2] holds for polynomials of arbitrary norm, the statement therein implies the one above. It likewise holds for any rotation of a strip, which we will reference shortly.

We will also recall the following behavior of the coefficient norm $\|P\|$ under the action of the orthogonal group, $\mathcal{O}(n)$.

Lemma 2.6. *For any degree d, and any $n \geq 2$, if P has degree d, then*

$$\|P\| \approx_{d,n} \|P \circ \theta\|$$

for any $\theta \in \mathcal{O}(n)$. Moreover, for $d \geq 2$, there exists some $\theta = \theta(P) \in \mathcal{O}(n)$ so that for any choice of $1 \leq j < k \leq n$, there holds

$$\|\partial_j (P \circ \theta)\| \gtrsim_{d,n} \|P\|. \tag{2.7}$$

Proof. We argue by compactness and contradiction. If the conclusion does not hold for some choice of dimension n and degree d, for all integers j, we can select P_{λ_j} and θ_j so that $\|P_{\lambda_j}\| = 1$ and $\|P_{\lambda_j} \circ \theta_j\| < 1/j$. For some subsequence, we must have $\{\lambda_{\alpha,j} : |\alpha| \leq d\} \to \{\lambda_\alpha : |\alpha| \leq d\}$, and $\theta_j \to \theta$. We conclude that $\|P_\lambda\| = 1$ and $\|P_\lambda \circ \theta\| = 0$, which is a contradiction.

Turning to the second claim, let $\mathcal{D}$ be the collection of differential operators

$$\partial_j \quad 1 \leq j \leq n.$$

We argue again, by contradiction and compactness. For some choice of dimension n and degree d, there is a polynomial P with $\|P\| = 1$ so that for all $\theta \in \mathcal{O}(n)$, there is a choice of $D \in \mathcal{D}$ so that $D(P \circ \theta) = 0$. The map $\theta \to D(P \circ \theta)$ is continuous, so that the set $\Theta_D = \{\theta : D(P \circ \theta) = 0\}$ is closed.

We also have $\bigcup_{D \in \mathcal{D}} \Theta_D = \mathcal{O}(n)$. The Baire Category theorem implies that for some D, there is a θ_0 in the interior of Θ_D. Hence, for all θ sufficiently close to θ_0, we have $DP \circ \theta = 0$. But the degree of P is at least 2, so this is can only happen if P itself is zero, which is a contradiction. $\square$

Proof of Lemma 2.4. For $s \in \mathbb{Z}$, and subsets $A \subset \mathbb{R}^n$, set $A^s := A + \{|x| \leq 2^s\}$.

We prove: For each P of degree at most $d \geq 1$, $\|P\| \gtrsim_{d,n} 1$, and any $k \geq 0$, and any cube I with $\ell(I) = 2^k$, for the set $Z = \{x \in I : |P(x)| \leq 2^{k/2}\}$, we have

$$|Z^s| \lesssim_{d,n} 2^{nk} \big(2^{-\epsilon_d k} + 2^{s-k}\big), \qquad 1 \leq s \leq k. \tag{2.8}$$

In view of our definition of Z_I in (2.5), and the condition $\|Q_{\beta_0}\| \gtrsim 1$, this proves (2.6).

We will induct on the degree of the polynomial, so let us first assume that $d = 1$, and that P is linear. But then, the set Z is a of the form $\{x \in I : |\langle \xi, x\rangle| < 2^{k/2}\}$, for some choice of $\|\xi\| \approx 1$. It is clear that (2.8) holds.

Henceforth, we will assume that $d \geq 2$. Now, since (2.8) is invariant under replacement of x by θx, for $\theta \in \mathcal{O}(n)$, we may assume by Lemma 2.6 that the condition (2.7) holds. Thus, the induction hypothesis applies to each polynomial $\partial_r P$, $1 \leq r \leq n$. As a consequence, we have this.

$$\begin{aligned}&\left|E^{s+c_n}\right| \lesssim_{d,n} 2^{nk}\big(2^{-\epsilon_{d-1}k} + 2^{s-k}\big) \leq 2^{nk}\big(2^{-\epsilon_d k} + 2^{s-k}\big), \qquad 1 \leq s \leq k,\\ &\text{where} \quad E := \bigcup_{r=1}^{n} \left\{|x| \leq 2^{k+1} : |\partial_r P(x)| \leq 2^{k/2}\right\}.\end{aligned}$$

Above, c_n is a dimensional constant. Observe that each of the sets that we form a union over, when restricted to a strip, can be covered by $C_{n,d}$ intervals, by Lemma 2.5. Thus, the same conclusion holds for the union.

The set we need to estimate is the set Z^s, but off of the set E^{s+c_n}, which we denote by $X := Z^s \setminus E^{s+c_n}$. For each $\sigma : \{2, \dots, n\} \to \{\pm 1\}$, let $\theta_\sigma \in \mathcal{O}_n$ be such that

$$\theta_\sigma^{-1}\vec{e_n} = \frac{1}{\sqrt{n}}(\sigma(1), \dots, \sigma(n)).$$

Here, $\vec{e_n} := (0, \dots, 0, 1)$ is the nth basis vector.

With this in mind, write

$$X = \bigcup_\sigma Y_\sigma,$$

where

$$\begin{aligned} Y_\sigma &:= \Big\{|x| \lesssim 2^k : d(x, E) > 2^{s+c_n}, |P(x)| \leq 2^{k/2},\\ &\qquad \sigma(r)\partial_r P(x) > 2^{k/2} \text{ for all } 1 \leq r \leq n\Big\}\\ &= \Big\{|x| \lesssim 2^k : d(x, \theta_\sigma^{-1}E) > 2^{s+c_n}, |P(\theta_\sigma x)| \leq 2^{k/2},\\ &\qquad \sigma(r)(\partial_r P)(\theta_\sigma x) > 2^{k/2} \text{ for all } 1 \leq r \leq n\Big\}.\end{aligned}$$

We will now favorably estimate $|Y_\sigma| \lesssim 2^{nk}(2^{-\epsilon_d k} + 2^{s-k})$ for each choice of σ.

Since any cube of dyadic side length is contained in a dyadic cube of six times its length shifted by some element $\vec{\omega} \in \{0, 1/3\}^n$, it suffices to show that for any grid shifted by any $\vec{\omega} \in \{0, 1/3\}^n$, and any $s + 3$ strip, $S = S_{\vec{\omega}}$, situated in that grid, S meets Y_σ in at most a bounded number of cubes. Since finite unions and complements of $s + 3$ intervals are expressible as finite unions of $s + 3$ intervals, it suffices to prove this result for $s + 3$ intervals. In particular, it suffices to show that any interval $I \subset S$ that meets Y_σ does so in at most 2 cubes.

So, suppose now that $Q \in I$, and $Q \cap Y_\sigma \neq \emptyset$. Since we have excised a 2^{s+c_n} neighborhood of E, this implies that for every $x \in Q$, $x \notin \theta_\sigma^{-1} E$, and thus

$$|\partial_r P(\theta_\sigma x)| > 2^{k/2} \text{ for all } 1 \leq r \leq n.$$

But, we know that there exists some point $y_Q \in Q \cap Y_\sigma$; for this y_Q, we have

$$\sigma(r)(\partial_r P)(\theta_\sigma y_Q) > 2^{k/2} \text{ for all } 1 \leq r \leq n,$$

so by connectedness and continuity, it follows that

$$\sigma(r)(\partial_r P)(\theta_\sigma x) > 2^{k/2} \text{ for all } 1 \leq r \leq n$$

for each $x \in Q$. But now we see that

$$\begin{aligned} \partial_n(P(\theta_\sigma x)) &= \vec{e_n} \cdot \theta_\sigma(\nabla P)(\theta_\sigma x) = \theta_\sigma^{-1}(\vec{e_n}) \cdot (\nabla P)(\theta_\sigma x) \\ &= \frac{1}{\sqrt{n}} \sum_{r=1}^{n} \sigma(r)(\partial_r P)(\theta_\sigma x) > \sqrt{n} 2^{k/2}. \end{aligned}$$

This strong monotonicity yields the result. $\square$

2.2. Rademacher-Menshov theorem

There is a general principle, a variant of the Rademacher-Menshov inequality that we will reference to control maximal truncations. This has been observed many times, for an explicit formulation and proof, see [9, Theorem 10.6].

Lemma 2.7. *Let (X, μ) be a measure space, and $\{\phi_j : 1 \leq j \leq N\}$ a sequence of functions which satisfy the Bessel type inequality below, for all sequences of coefficients $c_j \in \{0, \pm 1\}$,*

$$\left\| \sum_{j=1}^{N} c_j \phi_j \right\|_{L^2(X)} \leq A. \tag{2.9}$$

Then, there holds

$$\left\| \sup_{1<n\leq N} \left| \sum_{j=1}^{n} \phi_j \right| \right\|_{L^2(X)} \lesssim A \log(2+N). \tag{2.10}$$

3. The main lemma

The polynomials $P(x, y)$ in Theorem 1.1 are general polynomials. But, we can without loss of generality assume that P does not contain (a) constants, (b) terms that are purely powers of x, nor (c) terms that are linear in y. That is, we can write

$$P(x, y) = \sum_{\substack{\alpha,\beta:2\le|\alpha|+|\beta|\le d\\ |\beta|\neq 0}} \lambda_{\alpha,\beta} x^{\alpha} y^{\beta}.$$

Define $\|P\| = \sum_{\alpha,\beta} |\lambda_{\alpha,\beta}|$. Observe: Any dilate of a Calderón-Zygmund kernel is again a Calderón-Zygmund kernel. Therefore, in proving our sparse bounds, it suffices to do so for a polynomials satisfying $\|P\| = 1$. We will do so using induction on degree. Notice that the induction hypothesis implies that the sparse bounds hold without restriction on $\|P\|$. These remarks are important to the proof.

The essential step is to show that the sparse bounds of Theorem 1.1 holds in these cases.

Lemma 3.1. *The operators $T_{P,*}$ satisfy the sparse bounds* (1.1) *under either of these assumptions:*

(1) *The polynomial $P(x, y) = P(y)$ is only a function of y;*
(2) *The polynomial P satisfies $\|P\| = 1$, and the kernel $K(y)$ of the operator T is supported on $|y| \ge 2^{t_n}$, where $t_n \lesssim 1$ is a dimensional constant.*

We take up the proof of the lemma, returning to the conclusion of the proof our main Theorem 1.1 at the end of this section. Now, in the case of P being only a polynomial of y, the conclusion is invariant under dilations, so that we are free to assume that in this case $\|P\| = 1$. We need only concern ourselves with 'large scales.' For any finite constant t_n, the operator defined below, using the notation (1.4),

$$f \mapsto \int e(P(y)) f(x - y) \sum_{j<t_n} 2^{-jn} \psi_j(2^{-j} y)\, dy$$

is a Calderón-Zygmund operator, hence its maximal truncations are bounded on L^1 to weak L^1, with a norm bound that only depends upon the polynomial P through its degree.

Having removed that part of the kernel close to the origin, both cases in Lemma 3.1 fall under the assumptions of case 2. The maximal truncations are at most

$$\sup_{k_0 \ge C_d} \left| \sum_{k=k_0}^{\infty} \int e(P(x, y)) 2^{-kn} \psi_k\big(2^{-k} y\big) f(x - y)\, dy \right| + M_{\mathrm{HL}} f =: \tilde{T}_* f + M_{\mathrm{HL}} f.$$

On the right, the maximal function admits a sparse bound of type (1,1), so we show the sparse bound $(1, r)$ operator $\tilde{T}_* f$.

We make a familiar dyadic reduction, using *shifted dyadic grids*. For each $\vec{\omega} \in \{0, 1/3, 2/3\}^n$, let $\mathcal{D}^{\vec{\omega}}$ be the cubes in

$$\left\{2^k \left([0,1]^n + \vec{m} + (-1)^k \vec{\omega}\right) : \vec{m} \in \mathbb{Z}^n, k \in \mathbb{Z}\right\},$$

that is the dyadic grid $\mathcal{D}$ shifted by $\vec{\omega}$. It is well known that for any cube, I, there exists some $\vec{\omega} = \vec{\omega}(I) \in \{0, 1/3, 2/3\}^n$ and some $P = P(I) \in \mathcal{D}^{\vec{\omega}}$, we have the containment $I \subset P$, and $\ell(P) \leq 6\ell(I)$. Moreover, fixing the side length of a cube, we can resolve the identity function by

$$\sum_{\vec{\omega} \in \{0,1/3,2/3\}^n} \sum_{I \in \mathcal{D}^{\vec{\omega}} : \ell(I) = 2^k} \mathbf{1}_{\frac{1}{3}I} \equiv 1, \qquad k \in \mathbb{Z}.$$

Observe that we then have

$$\int e(P(x,y)) 2^{-kn} \psi_k\big(2^{-k}(x-y)\big) f(y)\, dy = \sum_{\vec{\omega} \in \{0,1/3,2/3\}^n} \sum_{\substack{I \in \mathcal{D}^{\vec{\omega}} \\ \ell(I) = 2^{k+t_n}}} T_I f(x),$$

where $$T_I g(x) = \int e(P(x,y)) 2^{-kn} \psi_k\big(2^{-k}(x-y)\big) \Big(g \mathbf{1}_{\frac{1}{3}I}\Big)(y)\, dy,$$

$$T_{*,\vec{\omega}} f = \sup_{\epsilon \geq 2^{C_d}} \left| \sum_{I \in \mathcal{D}^{\vec{\omega}} : \ell(I) \geq \epsilon} T_I f \right|. \tag{3.1}$$

The role of the grid $\mathcal{D}^{\vec{\omega}}$ in the remaining argument is of a standard nature, and so we suppress $\vec{\omega}$ in the notation in the argument to follow.

We will freely decompose the collection of cubes $\mathcal{D}_+ = \{I \in \mathcal{D}^{\vec{\omega}} : \ell(I) \geq 2^{C_d}\}$. Thus, extend the notation (3.1) to

$$T_{*,\mathcal{I}} f = \sup_{\epsilon > 0} \left| \sum_{I \in \mathcal{I} : \ell(I) \geq \epsilon} T_I f \right|, \tag{3.2}$$

where $\mathcal{I} \subset \mathcal{D}_+$. Our second main lemma is as below. As it forms the core of the proof, we place its proof in the next section.

Lemma 3.2. *Suppose that f, g are supported on cube I_0, and and $\mathcal{I}$ is a collection of subcubes of I_0 for which*

$$\sup_{I \in \mathcal{I}} \langle f \rangle_I < A \langle f \rangle_{I_0}, \qquad \sup_{I \in \mathcal{I}} \langle g \rangle_I < A \langle g \rangle_{I_0},$$

where $A \lesssim 1$ is a constant. Then,

$$\langle T_{*,\mathcal{I}} f, g \rangle \lesssim \frac{1}{r-1} |I_0| \langle f \rangle_{I_0} \langle g \rangle_{I_0, r}, \qquad 1 < r \leq 2. \tag{3.3}$$

The implied constant depends upon A, and $P(x, y)$ only through the degree of P.

Proof of Lemma 3.1. Recall that we are to prove the sparse bound for $T_{*,\mathcal{D}_+}$, as defined in (3.2). It suffices to consider bounded functions f, g supported on a fixed cube $I_0 \in \mathcal{D}_+$. Now, it is easy to see that

$$\sum_{J:J\supset I_0} |T_J f| \mathbf{1}_{I_0} \lesssim \langle f\rangle_{I_0}.$$

It suffices to prove the sparse bound for $T_{*,\mathcal{I}_0}$, where $\mathcal{I}_0$ consists of the dyadic cubes strictly contained in I_0.

Add the cube I_0 to the sparse collection $\mathcal{S}$. Take $\mathcal{E}$ to be the maximal dyadic cubes $P \subset I_0$ so that at least one of the following two inequalities hold:

$$\langle f\rangle_P > 100\langle f\rangle_{I_0} \quad \text{or} \quad \langle g\rangle_P > 100\langle g\rangle_{I_0}.$$

Let $E = \bigcup\{P : P \in \mathcal{E}\}$, so that $|E| \le \frac{1}{50}|I_0|$. And let $\mathcal{I} = \{I \in \mathcal{I}_0 : I \not\subset E\}$. It follows that

$$\langle T_{*,\mathcal{I}_0} f, g\rangle \le \langle T_{*,\mathcal{I}} f, g\rangle + \sum_P \langle T_{*,\mathcal{I}_0(P)} f, g\rangle,$$

where $\mathcal{I}_0(P) = \{I \in \mathcal{I}_0 : I \subset P\}$.

The first term on the right is controlled by (3.3). And, we add the collection $\mathcal{E}$ to the sparse collection $\mathcal{S}$, and recurse on the second group of terms above. This completes the proof of the sparse bound. □

To complete the proof of the Theorem 1.1, we need to consider the case *not covered* by Lemma 3.1, namely:

Lemma 3.3. *The operator $T_{P,*}$ satisfies the sparse bound inequalities* (1.1) *under the assumptions that polynomial P satisfies $\|P\| = 1$ and the kernel $K(y)$ of the operator T is supported on $|y| \le 2^{t_n}$.*

Proof. We induct on the degree of the polynomial $P(x, y)$ in the x-coordinate, call it d_x. The case of $d_x = 0$ is contained in the first case of Lemma 3.1, which we use as the base case.

We pass to the inductive case of $d_x > 0$. Note that the induction hypothesis implies that we have the full strength of our main theorem for polynomials of degree $d_x - 1$ in the x-coordinate. Now, the the kernel K is supported on the cube $I_0 = [-2^{-t_n}, 2^{t_n}]^n$, hence it suffices to prove the sparse bounds for functions f supported on a cube $m + I_0$, uniformly over $m \in \mathbb{Z}^d$. Equivalently, it is the same to prove the inequality functions supported on I_0, uniformly over polynomials $P(m + x, y)$, where P is a fixed polynomial of degree d_x in the x-coordinate, and $m \in 2^{t_n+1}\mathbb{Z}^n$. Write

$$R_m(x, y) = P(m + x, y) - P(x, y).$$

This is a polynomial with degree in x at most $d_x - 1$. (In fact, for $m = 0$, it is the zero polynomial.) Hence, $T_{R_m,*}$ satisfies the sparse bounds, uniformly in $m \in \mathbb{Z}^n$.

But, note that for $x, y \in I_0$,

$$|e(P(m+x,y)) - e(R_m(x,y))| = |e(P(x,y)) - 1| \lesssim |y|,$$

since $\|P\| \leq 1$. Therefore, we have

$$\left|T_{P(m+\cdot,\cdot),*} - T_{R_m,*} f(x)\right| \lesssim Mf.$$

The maximal function also satisfies the (1, 1) sparse bound, so the proof is complete. □

4. Proof of Lemma 3.2

We begin with a Calderón-Zygmund decomposition. Let $\mathcal{B}$ be the maximal subcubes $J \subset I_0$ for which $\langle f \rangle_J \geq A \langle f \rangle_{I_0}$. Write $f = \gamma + b$ where

$$b = \sum_{J \in \mathcal{B}} f \mathbf{1}_J = \sum_{s=0}^{s_0} \sum_{J \in \mathcal{B}_s} f \mathbf{1}_J =: \sum_{s=0}^{s_0} b_s.$$

Above, we set $\mathcal{B}_0 = \{J \in \mathcal{B} : \ell(J) \leq 0\}$, and for $0 < s \leq s_0 = \log_2 \ell(I_0) - 1$, set $\mathcal{B}_s = \{J \in \mathcal{B} : \ell(J) = s\}$. No cancellation property of b is needed. And, as a matter of convenience, we set $B_s \equiv 0$ if $s < 0$.

The first step is to observe that by Lemma 4.4 below we have

$$\langle T_{*,\mathcal{I}} \gamma, g \rangle \lesssim \left\| T_{*,\mathcal{I}} : L^{r'} \to L^{r'} \right\| \cdot \|\gamma\|_{r'} \|g\|_r \lesssim \frac{1}{r-1} |I_0| \langle f \rangle_{I_0} \langle g \rangle_{I_0,r},$$

since $\|\gamma\|_\infty \lesssim \langle f \rangle_{I_0}$. Note in particular that we have written the norms on f and g in sparse form.

It remains to consider the bad function. Observe that if $J \in \mathcal{B}$ and $K \in \mathcal{I}$, we have either $J \cap K = \emptyset$ or $J \subsetneq K$. Therefore, we can write

$$\begin{aligned} T_K b &= \sum_{s : 1 \leq 2^s < \ell(K)} T_K b_s = \sum_{s=1}^{\infty} T_K b_{k-s} \\ &=: \sum_{s=1}^{\infty} T_{K,s} b, \qquad \ell(K) = 2^k. \end{aligned}$$

We will consistently assume that $\ell(K) = 2^k$ below.

For integers $0 \leq s \leq s_0$ we decompose $\mathcal{K} = \mathcal{S}_s \cup \mathcal{N}_s$ where $K \in \mathcal{S}_s$ if $2^s < \ell(K)$ and there holds

$$\|T_{K,s} b\|_2^2 < 100 C_0 |K|^{-(1+\epsilon_d/n)} \|b_{k-s} \mathbf{1}_K\|_1^2, \tag{4.1}$$

where C_0 is the constant in (2.2). We refer to $\mathcal{S}_s$ as the 'standard' collection for which the second, simpler, term in (2.2) is decisive.

Lemma 4.1. *We have these inequalities*

$$\left\| \sup_{\epsilon} \left| \sum_{s \geq 0:} \sum_{K \in \mathcal{S}_s : \ell(I) \geq \epsilon} T_{K,s} b \right| \right\|_q \lesssim q \langle f \rangle_{I_0} |I_0|^{1/q}.$$

Proof. We have a gain in the scale. Holding the side length of K fixed, it is clear that for any integer j

$$\left\| \sum_{s \geq 0} \sum_{K \in \mathcal{S}_s : \ell(K) = 2^j} T_{K,s} b \right\|_\infty \lesssim 1.$$

And, in L^2, we have

$$\begin{aligned} \left\| \sum_{s \geq 0:} \sum_{K \in \mathcal{S}_s : \ell(K) = 2^j} T_{K,s} b \right\|_2 &\leq j \sum_{s=0}^{j-1} \sum_{K \in \mathcal{S}_s : \ell(K) = 2^j} \| T_{K,s} b \|_2^2 \\ &\lesssim j \sum_{s=0}^{j-1} \sum_{K \in \mathcal{S}_s : \ell(K) = 2^j} |K|^{-1-\epsilon_d/n} \| b_{k-s} \mathbf{1}_K \|_1^2 \\ &\lesssim j 2^{-\epsilon_d j} \sum_s \| b_s \|_1 \lesssim 2^{-\frac{j \epsilon_d}{2}} |I_0|, \end{aligned}$$

where we have used (4.1). Interpolating these two estimates gives us

$$\left\| \sum_{s \geq 0:} \sum_{K \in \mathcal{S}_s : \ell(K) = 2^j} T_{K,s} b \right\|_q \lesssim 2^{-\epsilon_d j / 2q} |I_0|^{1/q}.$$

Summing this over $t_n \leq j \leq s_0$ completes the proof. □

It therefore remains to consider the complementary 'non-standard' collection $\mathcal{N}_s$. Note that for it, we have

$$\| T_{K,s} b \|_2^2 < \frac{2C_0}{|I|} \int_{Z_I} b_s(x) b_s(y) \, dx \, dy.$$

There is an elementary endpoint estimate.

Lemma 4.2. *Under the assumption that*

$$\sup_{K \subset I_0} \langle g \rangle_K \leq A \langle g \rangle_{I_0},$$

we have these inequalities below, uniformly in $s \geq 0$

$$\left\langle \sup_{\epsilon} \left| \sum_{K \in \mathcal{N}_s : \ell(K) \geq \epsilon} T_{K,s} b \right|, g \right\rangle \lesssim |I_0| \langle f \rangle_{I_0} \langle g \rangle_{I_0}. \tag{4.2}$$

Proof. We argue by duality. For measurable $\sigma : I_0 \mapsto (0, \infty)$, we set

$$\tilde{T}_K f = \mathbf{1}_{\ell(K) \geq \sigma(x)} T_K f,$$

so that for arbitrary choice of σ, we can estimate

$$\begin{aligned}\left\langle \sum_{K \in \mathcal{N}_s} \tilde{T}_K b_{k-s}, g \right\rangle &= \sum_{K \in \mathcal{N}_s} \langle b_{k-s}, \tilde{T}_K^* g \rangle \\ &\lesssim \sum_{K \in \mathcal{N}_s} \|b_{k-s} \mathbf{1}_K\|_1 \langle g \rangle_K \lesssim |I_0| \langle f \rangle_{I_0} \langle g \rangle_{I_0}. \qquad \square\end{aligned}$$

The essence of the argument is therefore the L^2 bound below.

Lemma 4.3. *We have these inequalities below, uniformly in $s \geq 0$, for a choice of $\eta = \eta(n, d) > 0$.*

$$\left\| \sup_\epsilon \left| \sum_{K \in \mathcal{N}_s : \ell(K) \geq \epsilon} T_{K,s} b \right| \right\|_2 \lesssim 2^{-\eta s} \langle f \rangle_{I_0} |I_0|^{1/2}. \tag{4.3}$$

Interpolating between (4.2) and (4.3), we have

$$\left\langle \sup_\epsilon \left| \sum_{K \in \mathcal{N}_s : \ell(K) \geq \epsilon} T_{K,s} b \right|, g \right\rangle \lesssim 2^{-s\eta/q} |_0|^{1/q} \langle f \rangle_{I_0} \langle g \rangle_{I_0, q'}, \qquad s \geq 0.$$

Summing this over $s \geq 0$ completes the proof, with a single power of $q \simeq \frac{1}{r-1}$ as the leading coefficient in (3.3).

Proof. There is an important calculation, relating the L^2 norm of $T_{K,s} b$ to s. The argument below is illustrated in Figure 4.1.

$$\begin{aligned}\|T_{K,s} b\|_2^2 &\lesssim |K|^{-1} \int_{Z_K} b_{k-s}(x) b_{k-s}(y)\, dx\, dy \qquad \big(\ell(K) = 2^k\big) \\ &\lesssim |K|^{-1} \sum_{\substack{I, J \in \mathcal{B}_{k-s} \\ I, J \subset K}} |I \times J| \mathbf{1}_{I \times J \cap Z_K \neq \emptyset} \qquad \left(b_{k-s} = \sum_{I \in \mathcal{B}_{k-s}} b \mathbf{1}_I \right);\end{aligned}$$

here, we have used (4.4) twice. Crucially, we have (2.4), and so we can continue

$$\begin{aligned}&\lesssim |K|^{-1} |Z_K + \{(x, y) : |(x, y)| \leq C 2^{k-s}\}| \\ &\lesssim (2^{-\epsilon_d k} + 2^{-s}) |K|.\end{aligned}$$

At the end of the argument, we will again need this argument, but done fiberwise, using the full strength of (2.4).

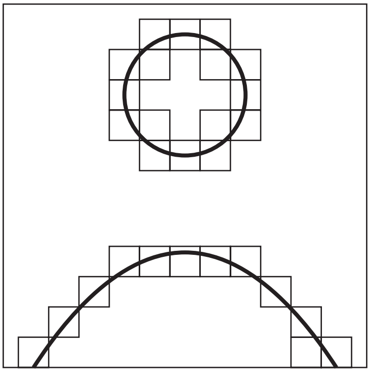

Figure 4.1. The large square is $K \times K$, with the set $\{(x, y) : x - y \in Z_K\}$ indicated by the two thick curves above. The set Z_K is then covered by rectangles of the form $I \times J$ for $I, J \in \mathcal{B}_{k,s}$, and the function $b \times b$ has integral $2^{-2t}|I| \cdot |J|$ on each rectangle $I \times J$.

Make this secondary division of $\mathcal{N}_s$. For integers $t \geq 0$ set $K \in \mathcal{N}_{s,t}$ if

$$K_0 2^{-t} \leq \langle b_{k-s} \rangle_K < K_0 2^{-t+1}, \qquad \ell(K) = 2^k. \tag{4.4}$$

The key additional property that we have is this Carleson measure estimate:

$$\sum_{K \in \mathcal{N}_{s,t} : K \subset J} |K| \lesssim 2^t |J| \tag{4.5}$$

for all dyadic $K \subset I_0$ with $K \in \mathcal{N}_{s,t}$. Indeed, we necessarily have $\sum_{t \geq 0 : 2^t \leq |J|} \langle b_t \rangle_J \lesssim 1$, so that

$$\sum_{K \in \mathcal{N}_{s,t} : K \subset J} |K| \lesssim 2^t \sum_{t \geq 0 : 2^t \leq |J|} \int_J b_t \lesssim 2^t |J|.$$

The first order of business is to discard the scales and locations where the bad functions have "large" averages, relative to the parameter s: we show that for each $t \leq \eta s$,

$$\left\| \sup_{\epsilon} \left| \sum_{\ell(K) \geq \epsilon, K \in \mathcal{N}_{s,t}} T_{K,s} b \right| \right\|_2 \lesssim 2^{-\eta' s} |I_0|^{1/2}. \tag{4.6}$$

The Carleson measure condition (4.5) implies that the set

$$E = \left\{ \sum_{I \in \mathcal{N}_{s,t}} \mathbf{1}_I > C 2^t \right\}$$

has measure at most $\frac{1}{4}|I_0|$, for appropriate constant C. This permits a further modification of (4.6), namely we restrict the sum to $\mathcal{N}^\sharp_{s,t} := \{I \in \mathcal{N}_{s,t} : I \not\subset E\}$, and show

$$\left\| \sup_\epsilon \left| \sum_{\ell(K)\geq\epsilon, K\in\mathcal{N}^\sharp_{s,t}} T_{K,s}b \right| \right\|_2 \lesssim 2^{-\eta' s}|I_0|^{1/2}. \tag{4.7}$$

This proves (4.6) upon a straight forward recursion inside the set E.

But now, taking into account that $t \leq \eta s$, (4.7) now follows from an application of Cauchy-Schwarz, and the observation that for each $K \in \mathcal{N}_s$, $\|T_{K,s}b\|_2^2 \lesssim 2^{-\epsilon_d s}|K|$, since each such K has $\ell(K) \geq 2^s$. The upshot is that we can proceed under the assumptions $t \geq \eta s$.

We now claim a strengthened version of (4.3), namely

$$\left\| \sup_\epsilon \left| \sum_{K\in\mathcal{N}^\sharp_{s,t}:\ell(I)\geq\epsilon} T_{K,s}b \right| \right\|_2 \lesssim 2^{-t/3}\langle f\rangle_{I_0}|I_0|^{1/2}, \qquad s,t \geq 0, \tag{4.8}$$

which leads to the same conclusion for the full set $\mathcal{N}_{s,t}$ as above.

The essence of this reduction is that it places the Rademacher-Menshov inequality (2.10) at our disposal. Namely, after proving an appropriate orthogonality condition, an instance of (2.9), we can conclude a result for maximal truncations from (2.10). Let $\mathcal{M}_1$ be the minimal elements of $\mathcal{N}^\sharp := \bigcup_{s=0}^{s_0} \mathcal{N}^\sharp_{s,t}$, and inductively set $\mathcal{M}_{u+1}$ to be the maximal elements of $\mathcal{N}^\sharp \setminus \bigcup_{v=1}^{u} \mathcal{M}_v$. Note that this set is empty for $u + 1 \geq u_0 = C2^t$. Then, set $\beta_u := \sum_{K\in\mathcal{M}_u} T_{K,s}b$. The required orthogonality statement is this. For any choice of constants $\varepsilon_t \in \{-1, 0, 1\}$, there holds

$$\left\| \sum_{u=1}^{u_0} \varepsilon_u \beta_u \right\|_2 \lesssim 2^{-t/3}|I_0|^{1/2}, \qquad t \geq \eta s. \tag{4.9}$$

This is the hypothesis (2.9) of the Rademacher-Menshov lemma, so we conclude that

$$\left\| \sup_v \left| \sum_{u=1}^{v} \varepsilon_u \beta_u \right| \right\|_2 \lesssim t2^{-t/3}|I_0|^{1/2}.$$

And this implies (4.8).

First, observe that

$$\begin{aligned} \sum_{u=1}^{C2^t} \|\beta_u\|_2^2 = \sum_{K\in\mathcal{N}_{s,t}} \|T_{K,s}b\|_2^2 &\lesssim \sum_{K\in\mathcal{N}_{s,t}} \langle b_{k-s}\rangle_K^2 |K| \\ &\lesssim 2^{-t} \sum_{K\in\mathcal{N}_{s,t}} \int_K b_{k-s}\, dx \lesssim 2^{-t}|I_0|. \end{aligned} \tag{4.10}$$

It remains to consider $u < v$, and the inner product

$$\langle \beta_u, \beta_v \rangle = \sum_{J \in \mathcal{M}_u} \sum_{\substack{K \in \mathcal{M}_v \\ J \subset K}} \langle b_{j-s}, T_J^* T_K b_{k-s} \rangle \qquad (\ell J = 2^j,\ \ell K = 2^k).$$

The kernel of $T_J^* T_K$ is controlled by (2.2). There are two terms on the right in (2.2), and for the second we have

$$\begin{aligned}\sum_{J \in \mathcal{M}_u} \sum_{\substack{K \in \mathcal{M}_v \\ J \subset K}} |K|^{-1-\epsilon_d/n} \int_K b_{k-s} \cdot \int_J b_{j-s} &\lesssim 2^{-2t} \sum_{J \in \mathcal{M}_u} \sum_{\substack{K \in \mathcal{M}_v \\ J \subset K}} |K|^{-1-\epsilon_d/n} |K| \cdot |J| \\ &\lesssim 2^{-2t-\epsilon_d v} |I_0|.\end{aligned}$$

The depends upon (4.4), and the fact that $K \in \mathcal{M}_v$ implies $\ell(K) \geq 2^v$.

The first term on the right in (2.2) is the essential oscillatory term. It involves the set Z_K, and our initial estimate is as below, fixing the interval K:

$$\begin{aligned}&\sum_{\substack{J \in \mathcal{M}_u \\ J \subset K}} \langle b_{j-s}, T_J^* T_K b_{k-s} \rangle \\ &\lesssim |K|^{-1} \sum_{\substack{J \in \mathcal{M}_u \\ J \subset K}} \int_K b_{k-s}(x) \int_J b_{j-s}(y) \mathbf{1}_{Z_K}(x, y)\, dy\, dx.\end{aligned} \tag{4.11}$$

Fix x above, and recall that $\pi_x Z$ is the fiber of Z over x. The integral in y is over the set $\pi_x Z_K$, and the integral is

$$\begin{aligned}\sum_{\substack{J \in \mathcal{M}_u \\ J \subset K}} \int_{J \cap \pi_x Z_K} b_{j-s}(y)\, dy &\lesssim 2^{-t} |\pi_x Z_k + \{|y| \lesssim 2^{k-v+u}\}| \\ &\lesssim 2^{-t} |K| (2^{-\epsilon_s v} + 2^{-v+u}).\end{aligned}$$

Above, we appeal to the fact that $|J| \leq 2^{-v+u}|K|$, and (4.4), and the condition (2.6), which is an estimate uniform over all fibers. It follows that we have

$$(4.11) \lesssim 2^{-2t} (2^{-\epsilon_d v} + 2^{-s-|u-v|}) |K|.$$

Sum this estimate over $K \in \mathcal{M}_v$, to conclude the bound

$$|\langle \beta_u, \beta_v \rangle| \lesssim 2^{-2t-\epsilon_d |u-v|} |I_0|.$$

Combine this with (4.10) easily prove (4.9), completing the proof of our lemma. □

Lemma 4.4. *For any collection $\mathcal{I} \subset \mathcal{D}_+$ we have*

$$\| T_{*,\mathcal{I}} : L^q \to L^q \| \lesssim q, \qquad 2 \leq q < \infty.$$

Proof. Observe that for fixed scales, we have

$$\|T_{\mathcal{I}(k)} : L^\infty \mapsto L^\infty\| \lesssim 1,$$
$$\|T_{\mathcal{I}(k)} : L^2 \mapsto L^2\| \lesssim 2^{-\eta k}, \qquad k \geq 0.$$

The first estimate is trivial, and the second is a consequence of the oscillatory estimate (2.2). Interpolating these estimates, and adding up gives the proof. □

References

[1] C. BENEA, F. BERNICOT and T. TERESA, *Sparse bilinear forms for Bochner Riesz multipliers and applications*, Trans. London Math. Soc. **4** (2017), 110–128.

[2] S. CHANILL and M. CHRIST, *Weak* (1, 1) *bounds for oscillatory singular integrals*, Duke Math. J. **55** (1987), 141–155.

[3] M. CHRIST, *Weak type* (1, 1) *bounds for rough operators*, Ann. of Math. (2) **128** (1988), 19–42.

[4] M. CHRIST and J. RUBIO DE FRANCIA, *Weak type* (1, 1) *bounds for rough operators. II*, Invent. Math. **93** (1988), 225–237.

[5] J. M. CONDE-ALONSO, A. CULIUC, F. DI PLINIO and O. YUMENG, *A sparse domination principle for rough singular integrals*, Anal. PDE **10** (2017), 1255–1284.

[6] J. CONDE-ALONSO and G. REY, *A pointwise estimate for positive dyadic shifts and some applications*, Math. Ann. **365** (2016), 1111–1135.

[7] A. CULIUC, F. DI PLINIO and Y. OU, *Domination of multilinear singular integrals by positive sparse forms*, J. Lond. Math. Soc. (2) **98** (2018), 269–392.

[8] A. CULIUC, R. KESLER and M. T. LACEY, *Sparse bounds for the discrete cubic Hilbert transform*, Anal. PDE **12** (2019), 1295–1272.

[9] D. CIPRIAN, T. TAO and C. THIELE, *Maximal multilinear operators*, Trans. Amer. Math. Soc. **360** (2008), 4989–5042.

[10] F. DI PLINIO Y. Q. DO and G. N. URALTSEV, *Positive sparse domination of variational Carleson operators*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **18** (2018), 1443–1458.

[11] F. DI PLINIO and A. K. LERNER, *On weighted norm inequalities for the Carleson and W alsh-Carleson operator*, J. Lond. Math. Soc. (2) **90** (2014), 654–674.

[12] Y. DING and L. HONGHAI, *Uniform weighted estimates for oscillatory singular integrals*, Forum Math. **24** (2012), 223–238.

[13] G. KARAGULYAN, *An abstract theory of singular operators*, arXiv:1611.03808.

[14] B. KRAUSE and M. T. LACEY, *Sparse bounds for maximal monomial oscillatory Hilbert transforms*, Studia Math. **242** (2018), 217–229.

[15] B. KRAUSE and M. T. LACEY, *Sparse bounds for random discrete Carleson theorems*, In: “50 Years with Hardy Spaces. Operator Theory: Advances and Applications”, Baranov A., Kisliakov S., Nikolski N. (eds.), Vol. 261, Birkhäuser, Cham, 317–332.

[16] M. T. LACEY, *An elementary proof of the A_2 bound*, Israel J. Math. **217** (2017), 181–195.

[17] M. T. LACEY and S. SPENCER, *Sparse bounds for oscillatory and random singular integrals*, New York J. Math. **23** (2017), 119–131.

[18] V. LIE, *The* (*weak-L^2*) *boundedness of the quadratic Carleson operator*, Geom. Funct. Anal. **19** (2009), 457–497.

[19] V. LIE, *The polynomial Carleson operator*, arXiv:1105.4504.

[20] F. RICCI and E. M. STEIN, *Oscillatory singular integrals and harmonic analysis on nilpotent groups*, Proc. Nat. Acad. Sci. U.S.A. **83** (1986), 1–3.

[21] F. RICCI and E. M. STEIN, *Harmonic analysis on nilpotent groups and singular integrals. I. Oscillatory integrals*, J. Funct. Anal. **73** (1987), 179–194.

[22] S. Sato, *Weighted weak type* (1, 1) *estimates for oscillatory singular integrals*, Studia Math. **141** (2000), 1–24.
[23] A. SEEGER, *Singular integral operators with rough convolution kernels*, J. Amer. Math. Soc. **9** (1996), 95–105.
[24] E. M. STEIN, "Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals", Princeton Mathematical Series, Vol. 43, Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.
[25] E. M. STEIN and S. WAINGER, *Oscillatory integrals related to Carleson's theorem*, Math. Res. Lett. **8** (2001), 789–800.

Department of Mathematics
The University of British Columbia
1984 Mathematics Road
Vancouver, B.C.
Canada V6T 1Z2
benkrause@math.ubc.ca

School of Mathematics
Georgia Institute of Technology
Atlanta GA 3034, USA
lacey@math.gatech.edu