Rev. Mat. Iberoam. **30** (2014), no. 2, 727–750
DOI 10.4171/RMI/798 

# Calderón commutators and the Cauchy integral on Lipschitz curves revisited I. First commutator and generalizations

Camil Muscalu

*Dedicated to Professor Nicolae Popa, on the occasion of his 70th birthday*

**Abstract.** This article is the first in a series of three papers, whose aim is to give new proofs to the well-known theorems of Calderón, Coifman, McIntosh and Meyer [2], [4] and [5]. Here we treat the case of the first commutator of Calderón and some of its generalizations.

## 1. Introduction

This is the first paper of three, whose aim is to give new proofs to the well-known theorems of Calderón, Coifman, McIntosh, and Meyer [2], [4], and [5], which established $L^p$ estimates for the so called Calderón commutators and the Cauchy integral on Lipschitz curves.

We refer the reader to the book [4] of Coifman and Meyer for a description of the history of these fundamental analytical objects, the role they play in analysis, and the various methods that have been further developed to understand these operators since the appearance of the original articles.

Other expository papers, where some of these results are described and connected with other parts of mathematics, are the proceedings of the plenary talks at the 1974 ICM in Vancouver and the 1978 ICM in Helsinki, given by Fefferman [6] and Calderón [3].

Our approach will also turn out to be sufficiently flexible and generic, to allow us to generalize these classical results in various new ways.

This first paper describes the case of the first commutator and its generalizations, the second one treats the case of the Cauchy integral on Lipschitz curves and its generalizations and, finally, the third will be devoted to the extension of all these results to the multiparameter setting of polydiscs of arbitrary dimension, solving completely along the way an open question of Coifman from the early eighties.

*Mathematics Subject Classification* (2010): Primary 42; Secondary 35.
*Keywords:* Calderón commutators, Littlewood–Paley projections, logarithmic estimates.

We naturally start with the first commutator.

Given a Lipschitz function $A$ on the real line (so $A' := a \in L^\infty(\mathbb{R})$) one formally defines the linear operator $C_1(f)$ by the formula

$$C_1(f)(x) = p.v. \int_{\mathbb{R}} \frac{A(x)-A(y)}{(x-y)^2} f(y)\,dy\,, \tag{1.1}$$

where the meaning of the principal value integral is

$$\lim_{\epsilon\to 0} \int_{\epsilon<|x-y|<1/\epsilon} \frac{A(x)-A(y)}{(x-y)^2} f(y)\,dy \tag{1.2}$$

whenever the limit exists. This is the so called first commutator of Calderón. Note that the simplest particular case is when $A(x) = x$, in which case $C_1(f)$ is the classical Hilbert transform.

Observe that when $a$ and $f$ are Schwartz functions, then (1.2) makes perfect sense. Indeed, for a fixed $\epsilon > 0$, one can rewrite the corresponding expression in (1.2) as

$$\begin{aligned} -\int_{\epsilon<|t|<1/\epsilon} \frac{A(x+t)-A(x)}{t^2} f(x+t)\,dt &= -\int_{\epsilon<|t|<1/\epsilon} \Big[\frac{A(x+t)-A(x)}{t}\Big] f(x+t)\frac{dt}{t} \\ &= -\int_{\epsilon<|t|<1/\epsilon} \Big[\int_0^1 a(x+\alpha t)\,d\alpha\Big] f(x+t)\frac{dt}{t}. \end{aligned} \tag{1.3}$$

Then, write $a$ and $f$ as

$$a(x+\alpha t) = \int_{\mathbb{R}} \widehat{a}(\xi_1)\, e^{2\pi i(x+\alpha t)\xi_1}\,d\xi_1 \quad \text{and} \quad f(x+t) = \int_{\mathbb{R}} \widehat{f}(\xi)\, e^{2\pi i(x+t)\xi}\,d\xi.$$

Using these formulas in (1.3), the expression becomes

$$-\int_{\mathbb{R}^2} m_\epsilon(\xi,\xi_1)\,\widehat{f}(\xi)\,\widehat{a}(\xi_1)\,e^{2\pi i x(\xi+\xi_1)}\,d\xi\,d\xi_1, \tag{1.4}$$

where

$$m_\epsilon(\xi,\xi_1) = \int_0^1 \int_{\epsilon<|t|<1/\epsilon} \frac{1}{t}\,e^{2\pi i t(\xi+\alpha\xi_1)}\,dt\,d\alpha,$$

which is known to converge pointwise to

$$-\int_0^1 \operatorname{sgn}(\xi+\alpha\xi_1)\,d\alpha.$$

In particular, the dominated convergence theorem implies that in (1.4) the limit as $\epsilon\to 0$ exists and it equals

$$\int_{\mathbb{R}^2} \Big[\int_0^1 \operatorname{sgn}(\xi+\alpha\xi_1)\,d\alpha\Big]\,\widehat{f}(\xi)\,\widehat{a}(\xi_1)\,e^{2\pi i x(\xi+\xi_1)}\,d\xi\,d\xi_1. \tag{1.5}$$

Because of (1.5), one can think of $C_1$ as being a bilinear operator in $f$ and $a$ and henceforth we will denote it by $C_1(f,a)$. The following theorem of Calderón is classical [2].

**Theorem 1.1.** *For every $A' = a \in L^\infty$ and every $1 < p < \infty$, the operator $C_1$ extends naturally to a bounded linear operator from $L^p(\mathbb{R})$ into $L^p(\mathbb{R})$, satisfying*

$$\|C_1(f,a)\|_p \lesssim \|a\|_\infty \cdot \|f\|_p. \tag{1.6}$$

The precise way in which the operator $C_1$ can be extended as claimed in the Theorem 1.1, will be described in detail a bit later.

At this point, one should observe that the symbol of (1.5), given by

$$(\xi, \xi_1) \to \int_0^1 \operatorname{sgn}(\xi + \alpha\xi_1)\, d\alpha \tag{1.7}$$

is not a Marcinkiewicz–Hörmander–Mihlin symbol (see [15]) and as a consequence, the Coifman–Meyer theorem on paraproducts (see [4]) cannot be applied. More precisely, one can see that away from the lines $\xi = 0$ and $\xi + \xi_1 = 0$, the symbol (1.7) is many times differentiable and behaves like a classical symbol, but along them it is only continuous. The observation on which our approach is based, is that in spite of this lack of differentiability, when one *smoothly* restricts (1.7) to an arbitrary Whitney square with respect to the origin[1], the Fourier coefficients of the corresponding function decay at least quadratically. This fact (which will be proved carefully in Lemma 2.4) will reduce the problem to one of proving estimates for the associated bilinear operators, which do not grow too fast with respect to the indices of the Fourier coefficients. We will see that these upper bounds can grow at most logarithmically, which will be more than enough to make the final power series convergent. This is, in just a few words, the strategy of our proof.

Before proceeding, let us also remark that if one permutes the two integrations in (1.5), one can rewrite that expression as

$$\int_0^1 BHT_\alpha(f,a)(x)\, d\alpha,$$

where $BHT_\alpha$ is the so called bilinear Hilbert transform of parameter $\alpha$. An alternative approach to the first commutator (suggested by Calderón), was to prove $L^p \times L^\infty \to L^p$ estimates for these operators, with implicit constants that are integrable or even uniform in $\alpha$. Estimates for the bilinear Hilbert transform have been first proved by Lacey and Thiele in [8] and [9], and uniform estimates have been later on obtained by Thiele [16], Grafakos and Li [7], and Li [10]. It is also interesting to remark that it is not yet known whether such an approach works for the second Calderón commutator [2] which this time can be written as

$$\int_{[0,1]^2} THT_{\alpha,\beta}(f,a,a)(x)\, d\alpha\, d\beta.$$

Recently, in [14], Palsson proved many estimates for the operator $\int_0^1 THT_{\alpha,\beta}\, d\alpha$ ($\beta$ is fixed now), but so far there are no $L^p$ estimates available (uniform or not)

[1] These are squares whose sides are parallel to the coordinate axes and whose distances to the origin are comparable to their sidelengths.

[2] The second commutator can similarly be seen as a trilinear operator with symbol $\int_{[0,1]^2} \operatorname{sgn}(\xi + \alpha\xi_1 + \beta\xi_2)\, d\alpha\, d\beta$.

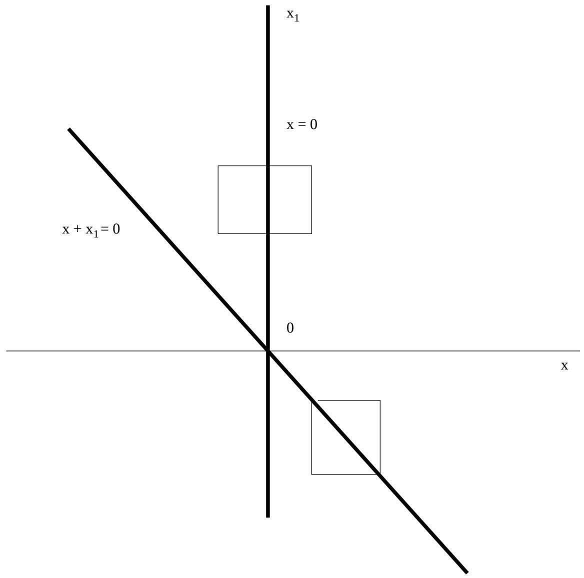


FIGURE 1. The singularities of the symbol of the first commutator

for the corresponding trilinear operator $THT_{\alpha,\beta}$ which has been called by several authors the trilinear Hilbert transform.

Now returning to (1.6), in order to describe the way in which $C_1(f,a)$ can be defined for any $a \in L^\infty$ and any $f \in L^p(\mathbb{R})$, we need to say a few words about adjoints of bilinear operators.

If $m(\xi_1, \xi_2)$ is a bounded symbol, denote by $T_m(f_1, f_2)$ the bilinear operator given by

$$T_m(f_1, f_2)(x) = \int_{\mathbb{R}^2} m(\xi_1, \xi_2)\, \widehat{f_1}(\xi_1)\, \widehat{f_2}(\xi_2)\, e^{2\pi i x(\xi_1+\xi_2)}\, d\xi_1\, d\xi_2, \tag{1.8}$$

for Schwartz functions $f_1$ and $f_2$. Associated with it is the trilinear form $\Lambda(f_1, f_2, f_3)$ defined by

$$\Lambda(f_1, f_2, f_3) = \int_{\mathbb{R}} T_m(f_1, f_2)(x)\, f_3(x)\, dx,$$

again for Schwartz functions $f_1$, $f_2$, and $f_3$.

There are two adjoint operators $T_m^{*1}$ and $T_m^{*2}$ naturally defined by the equalities

$$\int_{\mathbb{R}} T_m^{*1}(f_2, f_3)(x)\, f_1(x)\, dx = \Lambda(f_1, f_2, f_3)$$

and

$$\int_{\mathbb{R}} T_m^{*2}(f_1, f_3)(x)\, f_2(x)\, dx = \Lambda(f_1, f_2, f_3),$$

respectively. It is very easy to observe that both of them are also bilinear multipliers whose symbols are $m(-\xi_1 - \xi_2, \xi_2)$ and $m(\xi_1, -\xi_1 - \xi_2)$ respectively.

Now, if $a$ and $f$ are Schwartz functions, the inequality (1.6) is equivalent to

$$\left| \int_{\mathbb{R}} C_1(f,a)(x)\, g(x)\, dx \right| \lesssim \|a\|_\infty \cdot \|f\|_p \cdot \|g\|_{p'} \tag{1.9}$$

for any Schwartz function $g$, where $p'$ is the dual index of $p$ (so $1/p + 1/p' = 1$). We also know from the proceeding that

$$\int_{\mathbb{R}} C_1(f,a)(x)\, g(x)\, dx = \int_{\mathbb{R}} C_1^{*2}(f,g)(x)\, a(x)\, dx. \tag{1.10}$$

We are going to prove in the rest of the paper that

$$\|C_1^{*2}(f,g)\|_1 \lesssim \|f\|_p \cdot \|g\|_{p'} \tag{1.11}$$

for any Schwartz functions $f$ and $g$, and this shows that $C_1^{*2}$ can be extended by density to the whole $L^p \times L^{p'}$. However, this then means that the right hand side of (1.10) makes sense for any $a \in L^\infty$, not only for bounded Schwartz functions, and this suggests extending $C_1(f,a)$ by duality. More specifically, for $f \in L^p$ and $a \in L^\infty$, one can define $C_1(f,a)$ to be the unique $L^p$ function satisfying (1.10) for any $g \in L^{p'}$.

This discussion also proves that to demonstrate Theorem 1.1, we only need to prove (1.11). The idea now is to discretize $C_1^{*2}$ and reduce (1.11) to a discrete finite model.

**Acknowledgment.** I wish to thank Eli Stein who after a talk I gave in Pisa, kindly pointed out to me that the maximal Theorem 4.1 that will enter the picture later on, was actually known and can be found in Chapter II of [15].

## 2. Reduction to a finite localized model

We start with some standard notation and definitions. An interval $I$ of the real line $\mathbb{R}$ is called dyadic if it is of the form $I = [2^k n, 2^k(n+1)]$ for some $k, n \in \mathbb{Z}$. We will denote by $\mathcal{D}$ the set of all such dyadic intervals.

If $I \in \mathcal{D}$, we say that a smooth function $\Phi_I$ is a bump adapted to $I$ if and only if

$$|\partial^\alpha(\Phi_I)(x)| \leq C_{\alpha,N} \cdot \frac{1}{|I|^\alpha} \cdot \frac{1}{\big(1 + \mathrm{dist}(x,I)/|I|\big)^N}$$

for every integer $N$ and sufficiently many derivatives $\alpha$, where $|I|$ is the length of $I$. The intuition here is that the function $\Phi_I$ and many of its derivatives, are essentially supported on the interval $I$, in the sense that they decay very rapidly away from this interval. For example, if $\Phi$ is a fixed Schwartz function, then the function defined by $\Phi_I(x) := \Phi((x - c_I)/|I|)$ is clearly a bump function adapted to the interval $I$ (here, $c_I$ stands for the center of $I$).

Then, if $\Phi_I$ is a bump adapted to $I$, we say that $|I|^{-1/p}\, \Phi_I$ is an $L^p$-normalized bump adapted to $I$, for $1 \leq p \leq \infty$. Also, if $I \in \mathcal{D}$ and $\mathbf{n} \in \mathbb{Z}$ we denote by $I_{\mathbf{n}}$ the new dyadic interval $[2^k(n - \mathbf{n}), 2^k(n + 1 - \mathbf{n})]$ sitting $\mathbf{n}$ units of length $|I|$ away from $I$.

**Definition 2.1.** A sequence of $L^2$-normalized bumps $(\Phi_I)_I$ adapted to dyadic intervals $I$ is said to be of $\phi$ type if and only if for each $I$ there exists an interval $\omega_I$ $(=\omega_{|I|})$, symmetric with respect to the origin, such that $\operatorname{supp}\widehat{\Phi}_I \subseteq \omega_I$ and $|\omega_I| \sim |I|^{-1}$.

**Definition 2.2.** A sequence of $L^2$-normalized bumps $(\Phi_I)_I$ adapted to dyadic intervals $I$ is said to be of $\psi$ type if and only if for each $I$ there exists an interval $\omega_I$ $(=\omega_{|I|})$ such that $\operatorname{supp}\widehat{\Phi}_I \subseteq \omega_I$ and $|\omega_I| \sim |I|^{-1} \sim \operatorname{dist}(0, \omega_I)$.

Fix now two integers $\mathbf{n}_1$ and $\mathbf{n}_2$ and a finite arbitrary collection of dyadic intervals $\mathcal{I} \subseteq \mathcal{D}$. Consider also three sequences of $L^2$-normalized bumps $(\Phi^1_{I_{\mathbf{n}_1}})_{I\in\mathcal{I}}$, $(\Phi^2_{I_{\mathbf{n}_2}})_{I\in\mathcal{I}}$, $(\Phi^3_I)_{I\in\mathcal{I}}$ adapted to $I_{\mathbf{n}_1}$, $I_{\mathbf{n}_2}$, and I, respectively, such that at least two of them are of $\psi$ type. The following theorem holds.

**Theorem 2.3.** *The bilinear operator defined by*

$$T_{\mathcal{I}}(f,g) := \sum_{I\in\mathcal{I}} \frac{1}{|I|^{1/2}} \langle f, \Phi^1_{I_{\mathbf{n}_1}}\rangle \langle g, \Phi^2_{I_{\mathbf{n}_2}}\rangle \, \Phi^3_I \tag{2.1}$$

*is bounded from $L^p \times L^q \to L^r$ for any $1 < p, q < \infty$ and $0 < r < \infty$ so that $1/p + 1/q = 1/r$, with a bound of type*

$$O\big(\log <\mathbf{n}_1> \log <\mathbf{n}_2>\big)$$

*depending also implicitly on $p$ and $q$ but independent of the cardinality of $\mathcal{I}$ and of the families of bumps considered (here $<\mathbf{n}>$ simply means $2 + |\mathbf{n}|$).*

As we will see, Theorem 2.3 lies at the heart of our estimates. In the rest of the section we will show how it implies the desired inequality (1.11). The idea is to discretize $C_1^{*2}$ and show that it can be reduced to operators of type (2.1). Equivalently, since it has the same trilinear form, we will discretize $C_1$ instead. We start with two Littlewood–Paley decompositions and write

$$1(\xi) = \sum_{k_1} \widehat{\Psi}_{k_1}(\xi) \quad \text{and} \quad 1(\xi_1) = \sum_{k_2} \widehat{\Psi}_{k_2}(\xi_1)$$

where as usual, $\widehat{\Psi}_{k_1}(\xi)$ and $\widehat{\Psi}_{k_2}(\xi_1)$ are supported in the regions $|\xi| \sim 2^{k_1}$ and $|\xi_1| \sim 2^{k_2}$ respectively. In particular, we get

$$1(\xi, \xi_1) = \sum_{k_1,k_2} \widehat{\Psi}_{k_1}(\xi)\, \widehat{\Psi}_{k_2}(\xi_1). \tag{2.2}$$

By splitting (2.2) over the regions where $k_1 \ll k_2$, $k_2 \ll k_1$ and $k_1 \sim k_2$ we obtain the final decomposition

$$1(\xi, \xi_1) = \sum_k \widehat{\Phi}_k(\xi)\, \widehat{\Psi}_k(\xi_1) \tag{2.3}$$

$$+ \sum_k \widehat{\Psi}_k(\xi)\, \widehat{\Phi}_k(\xi_1) \tag{2.4}$$

$$+ \sum_{k_1\sim k_2} \widehat{\Psi}_{k_1}(\xi)\, \widehat{\Psi}_{k_2}(\xi_1). \tag{2.5}$$

By inserting this into (1.5), $C_1(f,a)$ splits as a sum of three different expressions. It is easy to see that the symbol of that corresponding to (2.4) is a classical symbol and for this part the inequality (1.11) follows from the Coifman–Meyer theorem on paraproducts [4]. We are thus left with understanding the other two terms. Notice that the first (corresponding to (2.3)) interacts with the line $\xi = 0$, while the third (corresponding to (2.5)) interacts with the line $\xi + \xi_1 = 0$ along which the original symbol

$$\int_0^1 \operatorname{sgn}(\xi + \alpha\xi_1)\, d\alpha$$

is only continuous. Also, for simplicity, henceforth we will replace $\int_0^1 \operatorname{sgn}(\xi + \alpha\xi_1)\, d\alpha$ with $\int_0^1 1_{\mathbb{R}_+}(\xi+\alpha\xi_1)\, d\alpha$ since the difference of the corresponding operators is just the product of $a$ and $f$ which clearly satisfies the original Hölder type inequalities.

Now fix a parameter $k \in \mathbb{Z}$ and consider the corresponding expressions (also, since $k_1 \sim k_2$ we assume that they are equal, for simplicity). Their trilinear forms are given by

$$\int_{\xi+\xi_1+\xi_2=0} \Big[\int_0^1 1_{\mathbb{R}_+}(\xi+\alpha\xi_1)\, d\alpha\Big] \widehat{\Phi}_k(\xi)\, \widehat{\Psi}_k(\xi_1)\, \widehat{\Psi}_k(\xi_2)\, \widehat{f}(\xi)\, \widehat{g}(\xi_1)\, \widehat{h}(\xi_2)\, d\xi\, d\xi_1\, d\xi_2$$

and

$$\int_{\xi+\xi_1+\xi_2=0} \Big[\int_0^1 1_{\mathbb{R}_+}(\xi+\alpha\xi_1)\, d\alpha\Big] \widehat{\Psi}_k(\xi)\, \widehat{\Psi}_k(\xi_1)\, \widehat{\Phi}_k(\xi_2)\, \widehat{f}(\xi)\, \widehat{g}(\xi_1)\, \widehat{h}(\xi_2)\, d\xi\, d\xi_1\, d\xi_2.$$

Clearly, the functions $\widehat{\Psi}_k(\xi_2)$ and $\widehat{\Phi}_k(\xi_2)$ have been inserted naturally into the above expressions (the first are supported away from zero while the support of the second contains the origin).

Now, on the support of $\widehat{\Phi}_k(\xi)\, \widehat{\Psi}_k(\xi_1)$, the function $\int_0^1 1_{\mathbb{R}_+}(\xi+\alpha\xi_1)\, d\alpha$ can be written as a double Fourier series of type

$$\sum_{n,n_1} C^k_{n,n_1}\, e^{2\pi i \frac{n}{2^k}\xi}\, e^{2\pi i \frac{n_1}{2^k}\xi_1}. \tag{2.6}$$

Similarly, on the support of $\widehat{\Psi}_k(\xi)\, \widehat{\Psi}_k(\xi_1)$ the same function can also be written as

$$\sum_{n,n_1} \widetilde{C}^k_{n,n_1}\, e^{2\pi i \frac{n}{2^k}\xi}\, e^{2\pi i \frac{n_1}{2^k}\xi_1}. \tag{2.7}$$

The following lemma will be crucial and gives upper bounds for these Fourier coefficients.

**Lemma 2.4.** *One has*

$$|C^k_{n,n_1}| \lesssim \frac{1}{<n>^2} \cdot \frac{1}{<n_1>^M}$$

*and also*

$$|\widetilde{C}^k_{n,n_1}| \lesssim \frac{1}{<n>^2} \cdot \frac{1}{<n-n_1>^M} + \frac{1}{<n>^M} \cdot \frac{1}{<n_1>^M}$$

*for a fixed large integer $M$, uniformly in $k$.*

We prove Lemma 2.4 at the end of this section. Roughly speaking, it shows that all the Fourier coefficients decay at least quadratically.

Now, (2.6) yields expressions of the form

$$\int_{\xi+\xi_1+\xi_2=0} \left[\widehat{\Phi}_k(\xi)\, e^{2\pi i \frac{n}{2^k}\xi}\right] \left[\widehat{\Psi}_k(\xi_1)\, e^{2\pi i \frac{n_1}{2^k}\xi_1}\right] \widehat{\Psi}_k(\xi_2)\, \widehat{f}(\xi)\, \widehat{g}(\xi_1)\, \widehat{h}(\xi_2)\, d\xi\, d\xi_1\, d\xi_2,$$

which can be rewritten as

$$\int_{\mathbb{R}} (f * \Phi_k^{1,n})(x)(g * \Psi_k^{2,n_1})(x)(h * \Psi_k^3)(x)\, dx,$$

and this can be further discretized by standard arguments (as explained in [13] for instance) as an average of expressions of type

$$\sum_{|I|=2^{-k}} \frac{1}{|I|^{1/2}} \langle f, \Phi^1_{I_n}\rangle \langle g, \Phi^2_{I_{n_1}}\rangle \Phi^3_I, \tag{2.8}$$

where the functions $\Phi^2_{I_{n_1}}$ and $\Phi^3_I$ are of $\psi$ type.

Similarly, (2.7) yields expressions of the form

$$\int_{\xi+\xi_1+\xi_2=0} \left[\widehat{\Psi}_k(\xi)\, e^{2\pi i \frac{n}{2^k}\xi}\right] \left[\widehat{\Psi}_k(\xi_1)\, e^{2\pi i \frac{n_1}{2^k}\xi_1}\right] \widehat{\Phi}_k(\xi_2)\, \widehat{f}(\xi)\, \widehat{g}(\xi_1)\, \widehat{h}(\xi_2)\, d\xi\, d\xi_1\, d\xi_2$$

and as we have seen, these can be further rewritten and discretized again in the form (2.8), where this time $\Phi^1_{I_n}$ and $\Phi^2_{I_{n_1}}$ are of $\psi$ type. The connection with (2.1) should be clear by now. If one adds all the expressions of the form (2.8) for all the scales $k \in \mathbb{Z}$, one obtains a discrete trilinear form corresponding to the part of $C_1$ related to (2.3) (and of course, as we mentioned, there is a similar trilinear form related to (2.5)). In particular, since we are interested in estimating $C_1^{*2}$, its bilinear model is of the form

$$\sum_{I\in\mathcal{I}} \frac{1}{|I|^{1/2}} \langle f, \Phi^1_{I_n}\rangle \langle h, \Phi^2_I\rangle \Phi^3_{I_{n_1}}$$

which should be rewritten as

$$\sum_{I\in\mathcal{I}} \frac{1}{|I|^{1/2}} \langle f, \Phi^1_{I_{n-n_1}}\rangle \langle h, \Phi^2_{I_{-n_1}}\rangle \Phi^3_I$$

to be able better able to compare it with (2.1).

Now, using the fact that $C_1^{*2}(f,g)$ makes perfect sense for Schwartz functions (in fact, it can be written as an expression similar to (1.5)) and by the triangle inequality, Fatou's lemma and Theorem 2.3, it follows that

$$\|C_1^{*2}(f,g)\|_1 \lesssim \sum_{n,n_1} \sup_k \left(|C^k_{n,n_1}|, |\widetilde{C}^k_{n,n_1}|\right) \cdot \log <n-n_1> \cdot \log <n_1> \cdot \|f\|_p \cdot \|g\|_{p'},$$

which is clearly bounded by $\|f\|_p \cdot \|g\|_{p'}$ as a consequence of the quadratic decay in Lemma 2.4. This completes the proof of (1.11).

We now describe the proof of Lemma 2.5. We first record the following.

**Lemma 2.5.** *One has the following identities:*

(a) $\displaystyle \partial_\xi^2 \Big( \int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big) = \delta_0(\xi+\xi_1) - \delta_0(\xi).$

(b) $\displaystyle \partial_{\xi_1}^2 \Big( \int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big) = \delta_0(\xi+\xi_1).$

(c) $\displaystyle \partial_\xi \partial_{\xi_1} \Big( \int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big) = \partial_{\xi_1}\partial_\xi \Big( \int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big) = \delta_0(\xi+\xi_1),$

*where $\delta_0$ is the Dirac distribution with respect to the origin.*

*Proof.* This is straightforward. For example, we show (a). One has

$$\begin{aligned} \partial_\xi^2 \Big( \int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big) &= \partial_\xi \Big( \int_0^{\xi_1} \delta_0(\xi+\alpha)\, d\alpha \Big) \\ &= \partial_\xi \Big( \int_\xi^{\xi+\xi_1} \delta_0(\alpha)\, d\alpha \Big) = \delta_0(\xi+\xi_1) - \delta_0(\xi). \qquad \square \end{aligned}$$

To prove now the estimates in Lemma 2.4, we rewrite (for instance) $\widetilde{C}^k_{n,n_1}$ as

$$\begin{aligned} &\frac{1}{2^k}\frac{1}{2^k}\int_{\mathbb{R}^2} \Big[ \int_0^1 1_{\mathbb{R}_+}(\xi+\alpha\xi_1)\, d\alpha \Big] \widehat{\widetilde{\Psi}}_k(\xi)\, \widehat{\widetilde{\Psi}}_k(\xi_1)\, e^{-2\pi i \frac{n}{2^k}\xi}\, e^{-2\pi i \frac{n_1}{2^k}\xi_1}\, d\xi\, d\xi_1 \\ &\qquad = \int_{\mathbb{R}^2} \Big[ \int_0^1 1_{\mathbb{R}_+}(\xi+\alpha\xi_1)\, d\alpha \Big] \widehat{\widetilde{\Psi}}(\xi)\, \widehat{\widetilde{\Psi}}(\xi_1)\, e^{-2\pi i n\xi}\, e^{-2\pi i n_1\xi_1}\, d\xi\, d\xi_1 \\ &\qquad = \int_{\mathbb{R}^2} \Big[ \frac{1}{\xi_1}\int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big] \widehat{\widetilde{\Psi}}(\xi)\, \widehat{\widetilde{\Psi}}(\xi_1)\, e^{-2\pi i n\xi}\, e^{-2\pi i n_1\xi_1}\, d\xi\, d\xi_1 \\ &\qquad := \int_{\mathbb{R}^2} \Big[ \int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\, d\alpha \Big] \widehat{\widetilde{\Psi}}(\xi)\, \widehat{\widetilde{\widetilde{\Psi}}}(\xi_1)\, e^{-2\pi i n\xi}\, e^{-2\pi i n_1\xi_1}\, d\xi\, d\xi_1, \end{aligned} \tag{2.9}$$

where $\widehat{\widetilde{\Psi}}(\xi)$, $\widehat{\widetilde{\Psi}}(\xi_1)$, and $\widehat{\widetilde{\widetilde{\Psi}}}(\xi_1)$ are supported away from the origin and are adapted to scale 1.

The idea is of course to integrate by parts as much as we can in (2.9) and keep track of the upper bounds that one gets in this way. We begin integrating by parts in $\xi$ as much as we can. Since both $\int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\,d\alpha$ and $\widehat{\widehat{\Psi}}(\xi)$ depend on $\xi$, the $\xi$ derivatives can hit either of the terms. If the derivative hits twice the term $\int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\,d\alpha$ then, because of Lemma 2.5, the $\xi$ variable disappears and becomes $-\xi_1$ (notice that $\xi$ cannot be zero in this case) at which point (2.9) simplifies to an expression of type

$$\int_{\mathbb{R}} \widehat{\widehat{\Psi}}(-\xi_1)\,\widehat{\widehat{\widehat{\Psi}}}(\xi_1)\,e^{-2\pi i\xi_1(n-n_1)}\,d\xi_1.$$

However this term can be integrated by parts as many times as we wish and this explains the appearance of the first upper bound for $|\widetilde{C}^k_{n,n_1}|$. If on the contrary, the $\xi$ derivative did not hit the term $\int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\,d\alpha$ two times, even after many integrations by parts, this means that we already gained a factor of type $1/{<}n{>}^M$ at which moment we stop integrating in $\xi$ and start integrating by parts in $\xi_1$. As before, if the $\xi_1$ derivatives hit the term $\int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\,d\alpha$ until one reaches $\delta_0(\xi+\xi_1)$ then $\xi_1$ becomes $-\xi$ and after that one integrates by parts a smooth function obtaining an upper bound of type $1/{<}n{>}^M\cdot 1/{<}n-n_1{>}^M$ which is smaller than the previously discussed one.

If finally, the $\xi_1$ derivative does not hit $\int_0^{\xi_1} 1_{\mathbb{R}_+}(\xi+\alpha)\,d\alpha$ until it becomes $\delta_0(\xi+\xi_1)$, then this means that it keeps hitting the smooth function of $\xi_1$, in which case we obtain an upper bound of type $1/{<}n{>}^M\cdot 1/{<}n_1{>}^M$ as desired.

The second term $C^k_{n,n_1}$ can be treated similarly. One should just remark that in this case the equality $\xi_1=-\xi$ is impossible and only $\delta_0(\xi)$ remains after integrating by parts, which explains the slight difference between the two upper bounds.

As a consequence, we are left with proving Theorem 2.3.

## 3. Proof of Theorem 2.3

The proof is based on the method introduced in [11] and [12].

We assume without loss of generality that the families $(\Phi^2_{I_{\mathbf{n}_2}})_I$ and $(\Phi^3_I)_I$ are of $\psi$ type (since all the other possible cases can be treated in a similar way). Fix also $1<p,q<\infty$ and $0<r<\infty$ so that $1/p+1/q=1/r$. We will prove that $T_{\mathcal{I}}$ maps $L^p\times L^q\to L^{r,\infty}$ since then (2.1) follows easily by standard interpolation arguments.

As usual (more specifically, as a consequence of scaling invariance and of the duality Lemma 5.4 from [1]), it is enough to show that given a measurable set $E\subseteq\mathbb{R}$ with $|E|=1$, one can find $E'\subseteq E$ with $|E'|\sim 1$ and so that

$$\sum_{I\in\mathcal{I}}\frac{1}{|I|^{1/2}}\,\big|\langle f,\Phi^1_{I_{\mathbf{n}_1}}\rangle\big|\,\big|\langle g,\Phi^2_{I_{\mathbf{n}_2}}\rangle\big|\,\big|\langle h,\Phi^3_I\rangle\big|\lesssim \log{<}\mathbf{n}_1{>}\cdot\log{<}\mathbf{n}_2{>}, \tag{3.1}$$

where $h:=\chi_{E'}$.

Define now the shifted maximal operator $M^{\mathbf{n}_1}$ and the shifted square function $S^{\mathbf{n}_2}$ by

$$M^{\mathbf{n}_1} f(x) := \sup_{x\in I} \frac{1}{|I|}\int_{\mathbb{R}} |f(y)|\, \widetilde{\chi_{I_{\mathbf{n}_1}}}(y)\, dy$$

where $\widetilde{\chi_{I_{\mathbf{n}_1}}}(y)$ denotes the function

$$\widetilde{\chi_{I_{\mathbf{n}_1}}}(y) = \Big(1 + \frac{\operatorname{dist}(y, I_{\mathbf{n}_1})}{|I_{\mathbf{n}_1}|}\Big)^{-100},$$

while $S^{\mathbf{n}_2}$ is given by

$$S^{\mathbf{n}_2} g(x) := \Big(\sum_I \frac{\big|\langle g, \Phi^2_{I_{\mathbf{n}_2}}\rangle\big|^2}{|I|}\, 1_I(x)\Big)^{1/2}.$$

As we will see later, both these are bounded on every $L^p$ space for $1<p<\infty$, with bounds of types $O(\log <\mathbf{n}_1>)$ and $O(\log <\mathbf{n}_2>)$, respectively. These important estimates will be proved in detail at the end of the paper (see Sections 4 and 5).

Using these two facts we define an exceptional set as follows.

First, define the set $\Omega_0'$ by

$$\Omega_0' := \big\{x : M^{\mathbf{n}_1} f(x) > C\log <\mathbf{n}_1> \big\} \bigcup \big\{x : S^{\mathbf{n}_2} f(x) > C\log <\mathbf{n}_2> \big\}.$$

Now let $d$ a positive integer and $\#$ be an integer so that $2^d < |\#| \le 2^{d+1}$. Define the set $\Omega^d_{\#}$ by

$$\Omega^d_{\#} := \big\{x : M^{\mathbf{n}_1 - \#} f(x) > C\log <\mathbf{n}_1 - \#>\ 2^{5d}\big\}$$

and then define also the set $\Omega_0''$ by

$$\Omega_0'' := \bigcup_{d\ge 0}\ \bigcup_{2^d < |\#| \le 2^{d+1}} \Omega^d_{\#}.$$

Define the set $\Omega_0'''$ in a similar way to $\Omega_0''$ but using the function $g$ and the corresponding index $\mathbf{n}_2$ instead. Then, define $\Omega_0$ to be

$$\Omega_0 := \Omega_0' \cup \Omega_0'' \cup \Omega_0''' \tag{3.2}$$

and finally, the exceptional set

$$\Omega := \Big\{x : M(1_{\Omega_0})(x) > \frac{1}{100}\Big\}.$$

Observe that $|\Omega| << 1$ if $C$ is chosen large enough and this allows us to define the set $E'$ by $E' := E\backslash\Omega$ and to observe that $|E'| \sim 1$, as desired. To estimate (3.1) properly, we split is into two parts as follows:

$$\sum_{I\cap\Omega^c \ne \emptyset} + \sum_{I\cap\Omega^c = \emptyset} := \mathrm{I} + \mathrm{II}. \tag{3.3}$$

**Estimates for I**

First, we observe that since $I \cap \Omega^c \neq \emptyset$ one has $|I \cap \Omega_0|/|I| < 1/100$ which means that $|I \cap \Omega_0^c| > \frac{99}{100}|I|$.

We now perform three independent stopping time type arguments for the functions $f, g$, and $h$ which will be combined carefully later.

Define first

$$\Omega_1 = \left\{x : M^{\mathbf{n}_1}(f)(x) > \frac{C \log < \mathbf{n}_1 >}{2^1}\right\}$$

and set

$$\mathcal{I}_1 = \left\{I \in \mathcal{I} : |I \cap \Omega_1| > \frac{1}{100}|I|\right\},$$

then define

$$\Omega_2 = \left\{x : M^{\mathbf{n}_1}(f)(x) > \frac{C \log < \mathbf{n}_1 >}{2^2}\right\}$$

and set

$$\mathcal{I}_2 = \left\{I \in \mathcal{I} \setminus \mathcal{I}_1 : |I \cap \Omega_2| > \frac{1}{100}|I|\right\},$$

and so on. The constant $C > 0$ here is the one in the definition of the set $E'$ before. Clearly, since $\mathcal{I}$ is finite, we will run out of dyadic intervals after a while, thus producing the sets $(\{\Omega_n\})_n$ and $(\{\mathcal{I}_n\})_n$.

Independently, define

$$\Omega'_1 = \left\{x : S^{\mathbf{n}_2}(g)(x) > \frac{C \log < \mathbf{n}_2 >}{2^1}\right\}$$

and set

$$\mathcal{I}'_1 = \left\{I \in \mathcal{I} : |I \cap \Omega'_1| > \frac{1}{100}|I|\right\},$$

then as before define

$$\Omega'_2 = \left\{x : S^{\mathbf{n}_2}(g)(x) > \frac{C \log < \mathbf{n}_2 >}{2^2}\right\}$$

and set

$$\mathbf{T}'_2 = \left\{I \in \mathcal{I} \setminus \mathcal{I}'_1 : |I \cap \Omega'_2| > \frac{1}{100}|I|\right\},$$

and so on, producing the finitely many sets $(\{\Omega'_n\})_n$ and $(\{\mathcal{I}'_n\})_n$. Of course, we would like to have such a decomposition available for $h$ as well. To do this, we first need to construct the analogue for it of the set $\Omega_0$. To do this, first choose an integer $N > 0$ large enough such that for every $I \in \mathcal{I}$ we have $|I \cap \Omega''^{c}_{-N}| > \frac{99}{100}|I|$ where we defined

$$\Omega''_{-N} = \left\{x : S(h)(x) > C\, 2^N\right\}.$$

Then, similarly to the previous algorithms, we define

$$\Omega''_{-N+1} = \Big\{x : S(h)(x) > \frac{C\,2^N}{2^1}\Big\}$$

and set

$$\mathcal{I}''_{-N+1} = \Big\{I \in \mathcal{I} : |I \cap \Omega''_{-N+1}| > \frac{1}{100}|I|\Big\},$$

and then define

$$\Omega''_{-N+2} = \Big\{x : S(h)(x) > \frac{C\,2^N}{2^2}\Big\}$$

and set

$$\mathbf{T}''_{-N+2} = \Big\{I \in \mathcal{I} \setminus \mathcal{I}''_{-N+1} : |I \cap \Omega''_{-N+2}| > \frac{1}{100}|I|\Big\},$$

and so on, constructing the finitely many sets $(\{\Omega''_n\})_n$ and $(\{\mathbf{T}''_n\})_n$.

Using all these decompositions, we can decompose the term $I$ further as

$$\sum_{l_1,l_2>0,l_3>-N}\ \sum_{I\in\mathcal{I}_{l_1,l_2,l_3}} \frac{1}{|I|^{3/2}}\big|\langle f,\Phi^1_{I_{\mathbf{n}_1}}\rangle\big|\,\big|\langle g,\Phi^2_{I_{\mathbf{n}_2}}\rangle\big|\,\big|\langle h,\Phi^3_I\rangle\big|\,\big|I\big|, \tag{3.4}$$

where

$$\mathcal{I}_{l_1,l_2,l_3} := \mathcal{I}_{l_1} \cap \mathcal{I}'_{l_2} \cap \mathcal{I}''_{l_3}.$$

Then, observe that, since $I$ belongs to $\mathcal{I}_{l_1,l_2,l_3}$, it has not been selected at any of the previous $l_1-1$, $l_2-1$ and $l_3-1$ steps respectively, which means that all of $|I\cap\Omega_{l_1-1}|$, $|I\cap\Omega'_{l_2-1}|$, and $|I\cap\Omega''_{l_3-1}|$ are smaller than $\frac{1}{100}|I|$. Equivalently, one has

$$|I\cap\Omega^c_{l_1-1}| > \frac{99}{100}|I|,\quad |I\cap\Omega'^c_{l_2-1}| > \frac{99}{100}|I|,\quad \text{and}\quad |I\cap\Omega''^c_{l_3-1}| > \frac{99}{100}|I|,$$

which implies that

$$\big|I\cap\Omega^c_{l_1-1}\cap\Omega'^c_{l_2-1}\cap\Omega''^c_{l_3-1}\big| > \frac{97}{100}|I|. \tag{3.5}$$

Using this in (3.4) one can estimate that expression by

$$\begin{aligned}
&\sum_{\substack{l_1,l_2>0,\\ l_3>-N}}\ \sum_{I\in\mathcal{I}_{l_1,l_2,l_3}} \frac{1}{|I|^{3/2}}\,|\langle f,\Phi^1_{I_{\mathbf{n}_1}}\rangle|\,|\langle g,\Phi^2_{I_{\mathbf{n}_2}}\rangle|\,|\langle h,\Phi^3_I\rangle|\,\big|I\cap\Omega^c_{l_1-1}\cap\Omega'^c_{l_2-1}\cap\Omega''^c_{l_3-1}\big| \\
&\quad= \sum_{\substack{l_1,l_2>0,\\ l_3>-N}} \int_{\Omega^c_{l_1-1}\cap\Omega'^c_{l_2-1}\cap\Omega''^c_{l_3-1}}\ \sum_{I\in\mathcal{I}_{l_1,l_2,l_3}} \frac{|\langle f,\Phi^1_{I_{\mathbf{n}_1}}\rangle|}{|I|^{1/2}}\,\frac{|\langle g,\Phi^2_{I_{\mathbf{n}_2}}\rangle|}{|I|^{1/2}}\,\frac{|\langle h,\Phi^3_I\rangle|}{|I|^{1/2}}\,\chi_I(x)\,dx \\
&\quad\lesssim \sum_{\substack{l_1,l_2>0,\\ l_3>-N}} \int_{\Omega^c_{l_1-1}\cap\Omega'^c_{l_2-1}\cap\Omega''^c_{l_3-1}\cap\Omega_{\mathcal{I}_{l_1,l_2,l_3}}} M^{\mathbf{n}_1}(f)(x)\,S^{\mathbf{n}_2}(g)(x)\,S(h)(x)\,dx \\
&\quad\lesssim \sum_{\substack{l_1,l_2>0,\\ l_3>-N}} \log<\mathbf{n}_1>\log<\mathbf{n}_2>\ 2^{-l_1}\,2^{-l_2}\,2^{-l_3}\,|\Omega_{\mathcal{I}_{l_1,l_2,l_3}}|,
\end{aligned} \tag{3.6}$$

where

$$\Omega_{\mathcal{I}_{l_1,l_2,l_3}} := \bigcup_{I \in \mathcal{I}_{l_1,l_2,l_3}} I.$$

On the other hand, we also have

$$\begin{aligned} |\,\Omega_{\mathcal{I}_{l_1,l_2,l_3}}| \leq |\,\Omega_{\mathcal{I}_{l_1}}| &\leq \left|\left\{x : M(\chi_{\Omega_{l_1}})(x) > \frac{1}{100}\right\}\right| \\ &\lesssim |\,\Omega_{l_1}| = \left|\left\{x : M^{\mathbf{n}_1}(f)(x) > \frac{C \log <\mathbf{n}_1>}{2^{l_1}}\right\}\right| \lesssim 2^{l_1 p}. \end{aligned}$$

Similarly, we have

$$|\,\Omega_{\mathcal{I}_{l_1,l_2,l_3}}| \lesssim 2^{l_2 q}$$

and also

$$|\,\Omega_{\mathcal{I}_{l_1,l_2,l_3}}| \lesssim 2^{l_3 \alpha},$$

for every $\alpha > 1$. Here we used the facts that the operators $M^{\mathbf{n}_1}$, $S^{\mathbf{n}_2}$ and $S$ are bounded on $L^s$ as long as $1 < s < \infty$ and that $|E_3'| \sim 1$. In particular, this implies that

$$|\Omega_{\mathcal{I}_{l_1,l_2,l_3}}| \lesssim 2^{l_1 p \theta_1}\, 2^{l_2 q \theta_2}\, 2^{l_3 \alpha \theta_3} \tag{3.7}$$

for any $0 \leq \theta_1, \theta_2, \theta_3 < 1$ such that $\theta_1 + \theta_2 + \theta_3 = 1$.

On the other hand, (3.6) can be split into

$$\begin{aligned} &\log <\mathbf{n}_1> \log <\mathbf{n}_2> \\ &\cdot \Big( \sum_{l_1,l_2>0, l_3>0} 2^{-l_1}\, 2^{-l_2}\, 2^{-l_3} |\Omega_{\mathcal{I}_{l_1,l_2,l_3}}| + \sum_{l_1,l_2>0, 0>l_3>-N} 2^{-l_1}\, 2^{-l_2}\, 2^{-l_3} |\Omega_{\mathcal{I}_{l_1,l_2,l_3}}| \Big). \end{aligned} \tag{3.8}$$

To estimate the first expression in (3.8) we use the inequality (3.7) for $\theta_1, \theta_2$, and $\theta_3$ such that $1 - p\theta_1 > 0$, $1 - q\theta_2 > 0$, and $1 - \alpha\theta_3 > 0$, while to estimate the second term we use (3.7) for $\theta_1, \theta_2$, and $\theta_3$ such that $1 - p\theta_1 > 0$, $1 - q\theta_2 > 0$, and $1 - \alpha\theta_3 < 0$. With these choices, the sum in (3.8) is indeed is $O(\log <\mathbf{n}_1> \log <\mathbf{n}_2>)$, as desired. This ends the discussion of $I$.

### Estimates for II

This term is simpler to estimate, now that we have defined the exceptional set carefully. Notice that the intervals of interest are those inside $\Omega$. One can split them as $\bigcup_{d \geq 0} \mathcal{I}_d$ where

$$\mathcal{I}_d := \Big\{ I \in \mathcal{I} : I \subseteq \Omega \quad \text{and} \quad 2^d \leq \frac{\operatorname{dist}(I, \Omega^c)}{|I|} < 2^{d+1} \Big\}.$$

Observe that for any $d \geq 0$ one has

$$\sum_{I \in \mathcal{I}_d} |I| \lesssim |\Omega| \lesssim 1.$$

Also, for every $I \in \mathcal{I}_d$ one has that $2^d I \cap \Omega^c = \emptyset$ and there exists $\widetilde{I}$ dyadic and of the same length, which lies # steps of length $|I|$ away from $I$ (with $2^d \leq |\#| \leq 2^{d+1}$), and having the property that $\widetilde{I} \cap \Omega^c \neq \emptyset$. In particular, this means that $I_{\mathbf{n}_1}$ and $I_{\mathbf{n}_2}$ are $\mathbf{n}_1 - \#$ and $\mathbf{n}_2 - \#$ steps of length $|I|$ away from $\widetilde{I}$. Using all these facts, one can estimate this term II by

$$\sum_{d \geq 0} \sum_{I \in \mathcal{I}_d} \frac{|\langle f, \Phi^1_{I_{\mathbf{n}_1}} \rangle|}{|I|^{1/2}} \frac{|\langle g, \Phi^2_{I_{\mathbf{n}_2}} \rangle|}{|I|^{1/2}} \frac{|\langle h, \Phi^3_I \rangle|}{|I|^{1/2}} |I|$$

$$\lesssim \sum_{d \geq 0} \sum_{2^d \leq |\#| \leq 2^{d+1}} \sum_{I \in \mathcal{I}_d} \big(\log < \mathbf{n}_1 - \# > \big) 2^{5d} \big(\log < \mathbf{n}_2 - \# > \big) 2^{5d} 2^{-Md} |I|$$

$$\lesssim (\log < \mathbf{n}_1 >) (\log < \mathbf{n}_2 >) \tag{3.9}$$

by using the trivial fact that, for $j = 1, 2$,

$$\log < \mathbf{n}_j - \# > \leq \log < \mathbf{n}_j > \cdot < \# > .$$

The proof is now complete.

## 4. Appendix 1 to Section 3: Logarithmic estimates for the shifted maximal function

The goal of this section is to prove the following theorem that has been used before. This result can be found in Stein [15], but we decided to give a selfcontained proof of it here (which we (re)discovered independently), not only for the reader's convenience, but also because some notation will be introduced that will be used later.

**Theorem 4.1** ([15]). *For any $\mathbf{n} \in \mathbb{Z}$, and for every $1 < p < \infty$, the shifted maximal function $M^{\mathbf{n}}$ is bounded on every $L^p$ space, with a bound of type $O(\log < \mathbf{n} >)$.*

*Proof.* First, we observe that in order to prove the desired estimates, it is enough to prove them for the corresponding *sharp* maximal function $\widetilde{M}^{\mathbf{n}}$ defined by

$$\widetilde{M}^{\mathbf{n}} f(x) := \sup_{x \in I} \frac{1}{|I_{\mathbf{n}}|} \int_{I_{\mathbf{n}}} |f(y)| \, dy \, , \tag{4.1}$$

where the suppremum is taken only over dyadic intervals.

To see this, fix $x$ and $I$ so that $x \in I$. One can write

$$\frac{1}{|I_{\mathbf{n}}|} \int_{I_{\mathbf{n}}} |f(y)| \, dy \lesssim \sum_{\# \in \mathbb{Z}} \Big[ \frac{1}{|I^{\#}_{\mathbf{n}}|} \int_{I^{\#}_{\mathbf{n}}} |f(y)| \, dy \Big] \frac{1}{< \# >^{100}},$$

where $I^{\#}_{\mathbf{n}}$ is the dyadic interval of the same length as $I_{\mathbf{n}}$ and lying # steps of length $|I_{\mathbf{n}}|$ away from it. In particular, using the proceeding and assuming that the theorem holds for $\widetilde{M}^{\mathbf{n}}$, one has

$$\begin{aligned}\|M^{\mathbf{n}}f\|_p &\lesssim \sum_{\#\in\mathbb{Z}} \frac{1}{<\#>^{100}} \big\|\widetilde{M}^{\mathbf{n}+\#} f\big\|_p \lesssim \sum_{\#\in\mathbb{Z}} \frac{1}{<\#>^{100}} \big(\log <\mathbf{n}+\#>\big)\|f\|_p \\ &\lesssim \sum_{\#\in\mathbb{Z}} \frac{1}{<\#>^{100}} \big(\log(<\mathbf{n}><\#>)\big) \lesssim \log <\mathbf{n}> \|f\|_p,\end{aligned}$$

as desired. We are then left with proving the theorem for $\widetilde{M}^{\mathbf{n}}$.

Now let $\lambda > 0$. We claim that one has the inequality

$$\big|\{x : \widetilde{M}^{\mathbf{n}} f(x) > \lambda\}\big| \lesssim \big(\log <\mathbf{n}>\big) \big|\{x : Mf(x) > \lambda\}\big| \tag{4.2}$$

where $M$ is the classical Hardy–Littlewood maximal operator. Assuming (4.2), the theorem for $\widetilde{M}^{\mathbf{n}}$ follows from the Hardy–Littlewood theorem by interpolation with the trivial $L^\infty$ estimate.

Finally, to prove (4.2) denote by $\mathcal{I}_{\mathbf{n}}^{\lambda}$ the collection, of all dyadic intervals $I_{\mathbf{n}}$, maximal with respect to inclusion, for which

$$\frac{1}{|I_{\mathbf{n}}|}\int_{I_{\mathbf{n}}} |f(y)|\,dy > \lambda.$$

Note that all of them are disjoint and one also has

$$\bigcup_{I_{\mathbf{n}}\in\mathcal{I}_{\mathbf{n}}^{\lambda}} I_{\mathbf{n}} = \{x : Mf(x) > \lambda\}.$$

Then, for every such maximal dyadic interval $I_{\mathbf{n}}$, consider its dyadic subintervals of lengths $|I_{\mathbf{n}}|$, $|I_{\mathbf{n}}|/2$, $|I_{\mathbf{n}}|/2^2$, etc. Observe that there exist only $[\log <\mathbf{n}>]$ disjoint dyadic intervals $I_{\mathbf{n}}^1, I_{\mathbf{n}}^2, \dots, I_{\mathbf{n}}^{[\log<\mathbf{n}>]}$ of the same length as $|I_{\mathbf{n}}|$, so that the translate with $-\mathbf{n}$ corresponding units of any such smaller dyadic subinterval of $I_{\mathbf{n}}$ becomes a subinterval of one of these $I_{\mathbf{n}}^1, I_{\mathbf{n}}^2, \dots, I_{\mathbf{n}}^{[\log<\mathbf{n}>]}$. The claim is now that

$$\{x : \widetilde{M}^{\mathbf{n}} f(x) > \lambda\} \subseteq \bigcup_{I_{\mathbf{n}}\in\mathcal{I}_{\mathbf{n}}^{\lambda}} \big(I_n \cup I_{\mathbf{n}}^1 \cup \dots \cup I_{\mathbf{n}}^{[\log<\mathbf{n}>]}\big).$$

To see this, pick $x_0$ so that $M^{\mathbf{n}} f(x_0) > \lambda$. This then means, in particular, that there exists a dyadic interval $J$ containing $x_0$, so that $\frac{1}{|J_{\mathbf{n}}|}\int_{J_{\mathbf{n}}} |f(y)|\,dy > \lambda$. Because of the previous construction, one can find selected maximal interval of type $I_{\mathbf{n}}$, so that $J_{\mathbf{n}} \subseteq I_{\mathbf{n}}$. But then, this means in particular that $J$ itself will be a subset of either $I_{\mathbf{n}}$ or $I_{\mathbf{n}}^1$ or…or $I_{\mathbf{n}}^{[\log<\mathbf{n}>]}$, which implies the claim.

It is now easy to see that together this claim and the disjointness of the maximal intervals $I_{\mathbf{n}}$, imply (4.2). The proof is then complete[3]. $\square$

[3]Of course, since the trivial $L^\infty$ estimate comes with an $O(1)$ bound, by interpolation the $L^p$ bound of $M^{\mathbf{n}}$ will be even $O((\log <\mathbf{n}>)^{1/p})$. However, for simplicity, we used the $O(\log <\mathbf{n}>)$ bound all the time.

## 5. Appendix 2 to Section 3: Logarithmic estimates for the shifted square function

The goal of this last section is to prove the following theorem which played an important role earlier in the argument[4].

**Theorem 5.1.** *For any $\mathbf{n} \in \mathbb{Z}$, and every $1 < p < \infty$, the shifted square function $S^{\mathbf{n}}$ is bounded on every $L^p$ space, with a bound of type $O(\log <\mathbf{n}>)$.*

*Proof.* Besides the observations of the previous section, the proof is based on a classical decomposition of Calderón and Zygmund [15].

First, observe that $S^{\mathbf{n}}$ is bounded on $L^2$ with a bound independent of $\mathbf{n}$. Indeed, one can see that

$$\|S^{\mathbf{n}} f\|_2 = \Big(\sum_I \langle f, \Phi_{I_{\mathbf{n}}}\rangle^2\Big)^{1/2}$$

which is clearly comparable to the $L^2$ norm of the classical Littlewood–Paley square function, which is known to be bounded on $L^2$.

Next, we show that

$$\|S^{\mathbf{n}} f\|_{1,\infty} \lesssim (\log <\mathbf{n}>)\, \|f\|_1, \tag{5.1}$$

or, more precisely, that

$$\big|\{x \in \mathbb{R} : S^{\mathbf{n}} f(x) > \lambda\}\big| \lesssim \log <\mathbf{n}> \frac{1}{\lambda}\|f\|_1. \tag{5.2}$$

Fix such a $\lambda > 0$ and construct a Calderón–Zygmund decomposition of the function $f$ *at level* $\lambda$. Choose maximal dyadic intervals one-by-one so that

$$\frac{1}{|J|}\int_J |f(y)|\, dy > \lambda.$$

Observe that these intervals are by construction pairwise disjoint and denote their union by $\Omega$. One has

$$|\Omega| = \sum_J |J| < \frac{1}{\lambda}\sum_J \int_J |f(y)|\, dy \le \frac{1}{\lambda}\|f\|_1. \tag{5.3}$$

Now split the function $f$ as

$$f = g + b,$$

where

$$g := f\chi_{\Omega^c} + \sum_J \Big[\frac{1}{|J|}\int_J f(y)\, dy\Big]\chi_J,$$

[4]It may very well be that this result has been observed before (as was the case with the previous shifted maximal function) but since we did not find it in the literature, we have included a self-contained proof of it in what follows.

and

$$b := f - g := \sum_J b_J, \quad \text{where} \quad b_J := \Big[ f - \frac{1}{|J|}\int_J f(y)\,dy \Big]\chi_J.$$

Clearly, $\mathrm{supp} b_J \subseteq J$. Observe also that one has

$$|f(x)| \le \lambda$$

for every $x \in \Omega^c$ and, as a consequence,

$$\|g\|_\infty \lesssim \lambda$$

since one also observes that

$$\Big| \frac{1}{|J|}\int_J f(y)\,dy \Big| \le \frac{1}{|J|}\int_J |f(y)|\,dy \le \frac{2}{|\widetilde{J}|}\int_{\widetilde{J}} |f(y)|\,dy \le 2\lambda,$$

where $\widetilde{J}$ is the unique dyadic interval containing $J$ and twice as long as $J$. It is also important to observe that

$$\int_{\mathbb{R}} b_J(y)\,dy = 0,$$

by definition, and also that

$$\|b_J\|_1 = \int_J |b_J(y)|\,dy \le \int_J |f(y)|\,dy + \Big( \frac{1}{|J|}\int_J |f(y)|\,dy \Big)|J| \lesssim \int_J |f(y)|\,dy \lesssim \lambda |J|,$$

as we have seen before.

Using all these properties, one can write

$$\begin{aligned} &\big| \{x \in \mathbb{R} : S^{\mathbf{n}} f(x) > \lambda\} \big| \\ &\qquad \le \big| \{x \in \mathbb{R} : S^{\mathbf{n}} g(x) > \lambda/2\} \big| + \big| \{x \in \mathbb{R} : S^{\mathbf{n}} b(x) > \lambda/2\} \big|. \end{aligned} \tag{5.4}$$

To estimate the first term in (5.4), we use the $L^2$ boundedness of $S^{\mathbf{n}}$ and we write

$$\begin{aligned} \big| \{x \in \mathbb{R} : S^{\mathbf{n}} g(x) > \lambda/2\} \big| &\lesssim \frac{1}{\lambda^2}\|S^{\mathbf{n}} g\|_2^2 \lesssim \frac{1}{\lambda^2}\|g\|_2^2 = \frac{1}{\lambda^2}\int_{\mathbb{R}} |g(x)|^2\,dx \\ &\lesssim \frac{1}{\lambda^2}\lambda \int_{\mathbb{R}} |g(x)|\,dx = \frac{1}{\lambda}\|g\|_1 \\ &\lesssim \frac{1}{\lambda}\Big( \int_{\Omega^c} |f(x)|\,dx + \sum_J \int_J |f(x)|\,dx \Big) \lesssim \frac{1}{\lambda}\|f\|_1, \end{aligned}$$

as desired.

To estimate the second term in (5.4), we proceed as follows. First, for any interval $J$, consider the associated $J^1, J^2, \ldots, J^{[\log <\mathbf{n}>]}$ as defined in the previous section and define the set $\Omega_J$ by

$$\Omega_J := 5J \bigcup 5J^1 \bigcup 5J^2 \bigcup \ \ldots \ \bigcup 5J^{[\log <\mathbf{n}>]}.$$

Then, one has

$$\text{(5.5)}\qquad \left|\left\{x\in\mathbb{R}: S^{\mathbf{n}}b(x)>\lambda/2\right\}\right| \le \left|\left\{x\in\bigcup_J \Omega_J : S^{\mathbf{n}}b(x)>\lambda/2\right\}\right| + \left|\left\{x\in\Big(\bigcup_J\Omega_J\Big)^c : S^{\mathbf{n}}b(x)>\lambda/2\right\}\right|.$$

The first expression is easy to estimate since one can write

$$\left|\left\{x\in\bigcup_J \Omega_J : S^{\mathbf{n}}b(x)>\lambda/2\right\}\right| \le \left|\bigcup_J\Omega_J\right| \lesssim \left(\log <\mathbf{n}>\right)\sum_J |J| \lesssim \left(\log <\mathbf{n}>\right)\frac{1}{\lambda}\|f\|_1,$$

as we have seen before. The second expression in (5.5) can be majorized by

$$\frac{1}{\lambda}\int_{\left(\bigcup_J\Omega_J\right)^c} S^{\mathbf{n}}\, b(x)\,dx \le \frac{1}{\lambda}\sum_J\int_{\left(\bigcup_J\Omega_J\right)^c} S^{\mathbf{n}}\, b_J(x)\,dx \le \frac{1}{\lambda}\sum_J\int_{(\Omega_J)^c} S^{\mathbf{n}}\, b_J(x)\,dx$$

and we claim now that for any $J$ one has

$$\text{(5.6)}\qquad \int_{(\Omega_J)^c} S^{\mathbf{n}}\, b_J(x)\,dx \lesssim \lambda|J|.$$

Assuming (5.6), one can continue the previous inequality and further majorize it by

$$\frac{1}{\lambda}\lambda\sum_J |J| \lesssim |\Omega| \lesssim \frac{1}{\lambda}\|f\|_1$$

as desired.

We are then left with proving our claim (5.6). First, we majorize the left-hand side of it by

$$\begin{aligned}\int_{(\Omega_J)^c}\Big(\sum_I \frac{|\langle b_J,\Phi_{I_{\mathbf{n}}}\rangle|}{|I|^{1/2}}1_I(x)\Big)dx &= \sum_I\int_{(\Omega_J)^c}\frac{|\langle b_J,\Phi_{I_{\mathbf{n}}}\rangle|}{|I|^{1/2}}1_I(x)\,dx\\ &= \sum_{|I|\le|J|}\int_{(\Omega_J)^c}\frac{|\langle b_J,\Phi_{I_{\mathbf{n}}}\rangle|}{|I|^{1/2}}1_I(x)\,dx + \sum_{|I|>|J|}\int_{(\Omega_J)^c}\frac{|\langle b_J,\Phi_{I_{\mathbf{n}}}\rangle|}{|I|^{1/2}}1_I(x)\,dx\\ &:= A+B.\end{aligned}$$

### Estimating A

The main observation here is to realize that since $|I|\le|J|$ and $I\cap(\Omega_J)^c\neq\emptyset$, one must in particular have $I_{\mathbf{n}}\cap 3J=\emptyset$. This allows one to estimate $A$ by

$$\sum_{|I_{\mathbf{n}}|\le|J|}\Big(1+\frac{\operatorname{dist}(I_{\mathbf{n}},J)}{|I_{\mathbf{n}}|}\Big)^{-10}\int_{\mathbb{R}}|b_J(y)|\,dy \lesssim \lambda|J|\sum_{|I_{\mathbf{n}}|\le|J|}\Big(1+\frac{\operatorname{dist}(I_{\mathbf{n}},J)}{|I_{\mathbf{n}}|}\Big)^{-10} \lesssim \lambda|J|,$$

as required by (5.6).

**Estimating B**

This time, one has to take into account the fact that

$$\int_{\mathbb{R}} b_J(y)\,dy = 0. \tag{5.7}$$

As before, one can estimate $B$ by

$$\sum_{|I_{\mathbf{n}}|>|J|} |\langle b_J, \Phi^{\infty}_{I_{\mathbf{n}}}\rangle|,$$

where this time $\Phi^{\infty}_{I_{\mathbf{n}}} := |I_{\mathbf{n}}|^{1/2}\,\Phi_{I_{\mathbf{n}}}$ is an $L^\infty$ normalized bump. In order to emphasize that the dependence of $\mathbf{n}$ is now irrelevant now we rewrite the above expression as

$$\sum_{|K|>|J|} |\langle b_J, \Phi^{\infty}_{K}\rangle|\,,$$

where the sum is over dyadic intervals $K$.

Fix $K$ such that $|K|>|J|$ and observe that

$$|\langle b_J, \Phi^{\infty}_{K}\rangle| = \Big|\int_{\mathbb{R}} b_J(z)\,\overline{\Phi^{\infty}_{K}}(z)\,dz\Big| = \Big|\int_J b_J(z)\big(\overline{\Phi^{\infty}_{K}}(z) - \overline{\Phi^{\infty}_{K}}(c_J)\big)\,dz\Big|,$$

where $c_J$ denotes the midpoint of the interval $J$.

Then, observe that, for $z\in J$, one has

$$\big|\,\overline{\Phi^{\infty}_{K}}(z) - \overline{\Phi^{\infty}_{K}}(c_J)\big| \lesssim |J\,|\frac{1}{|K|}\Big(1+\frac{\mathrm{dist}(K,J)}{|K|}\Big)^{-10},$$

which can be further estimated by

$$|J\,|\,\frac{1}{|K|}\,\Big(1+\frac{\mathrm{dist}(K,J)}{|K|}\Big)^{-10}\int_J |b_J(y)|\,dy \lesssim |J\,|\,\frac{1}{|K|}\,\Big(1+\frac{\mathrm{dist}(K,J)}{|K|}\Big)^{-10}\lambda\,|J\,|.$$

Finally, the corresponding (5.6) follows from the straightforward observation that

$$\sum_{|K|>|J\,|}\frac{|J|}{|K|}\,\Big(1+\frac{\mathrm{dist}(K,J)}{|K|}\Big)^{-10}\lesssim 1.$$

By interpolating between $L^2$ and weak-$L^1$ we obtain the theorem for any $1<p\le 2$. To prove the rest of the estimates we proceed as usual, by duality. Fix $2<p<\infty$. By using Khinchin's inequality, one can write

$$\begin{aligned}\|S^{\mathbf{n}}f\|_p^p &= \int_{\mathbb{R}}\Big(\sum_I \frac{|\langle f,\Phi_{I_{\mathbf{n}}}\rangle|^2}{|I|}\,\chi_I(x)\Big)^{p/2} dx \lesssim \int_{\mathbb{R}}\int_0^1 \Big|\sum_I r_I(t)\langle f,\Phi_{I_{\mathbf{n}}}\rangle\,h_I(x)\Big|^p dx\,dt\\ &= \int_0^1 \Big\|\sum_I r_I(t)\langle f,\Phi_{I_{\mathbf{n}}}\rangle\,h_I\Big\|_p^p dt,\end{aligned} \tag{5.8}$$

where $(r_I)_I$ are the Rademacher functions and $(h_I)_I$ the $L^2$-normalized Haar functions.

Now fix $t \in [0,1]$ and consider the linear operator

$$f \to \sum_I r_I(t)\langle f, \Phi_{I_{\mathbf{n}}}\rangle h_I.$$

Using the fact that $S^{\mathbf{n}}$ and the Littlewood–Paley square function associated to $(h_I)_I$ are bounded on $L^s$ for $1 < s \leq 2$, an argument identical to the one used to prove Theorem 2.3 shows that the above operator is also bounded on $L^s$ for $1 < s \leq 2$ and by duality, bounded on $L^s$ for $2 \leq s < \infty$ as well, with bounds independent of $t$ that grow logarithmically in $< \mathbf{n} >$. Using these observations in (5.8) completes the proof of the theorem. □

## 6. Generalizations

First observe that the first commutator $C_1 f$ can also be written as

$$C_1 f(x) = p.v. \int_{\mathbb{R}} \Big(\frac{\Delta_t}{t} A(x)\Big) f(x+t)\,\frac{dt}{t}\,, \tag{6.1}$$

where $\Delta_t$ is the *finite difference* operator at scale $t$ given by

$$\Delta_t g(x) := g(x+t) - g(x).$$

There is a very simple way to motivate the introduction of this operator. Start with the Leibnitz rule

$$(Af)' = A'f + Af'$$

and solve for $A'f$ to obtain

$$A'f = (Af)' - Af' = D(Af) - ADf = [D, A]f$$

where $D$ is the operator of taking one derivative and $A$ is viewed now as the operator of multiplication with the function $A(x)$. In particular, assuming that $A' \in L^\infty$, the commutator $[D, A]$ maps $L^p$ into itself boundedly, for every $1 < p < \infty$. One might ask: *Does this property hold for the operator* $[|D|, A]$ *as well?* A straightforward calculation shows that $[|D|, A]$ is precisely the first commutator of Calderón.

Given this, it is of course natural to ask: *What can be said about the double commutator* $[|D|, [|D|, A]]$*?*

A direct calculation shows that the expression $[|D|, [|D|, A]](f)(x)$ equals

$$p.v. \int_{\mathbb{R}^2} \Big(\frac{\Delta_t}{t} \circ \frac{\Delta_s}{s} A(x)\Big)\, f(x+t+s)\,\frac{dt}{t}\,\frac{ds}{s}, \tag{6.2}$$

a formula that can be naturally seen as a bilinear operator, this time depending on $f$ and $A''$. Its symbol can be again calculated easily and it is given by

$$\Big(\int_0^1 \mathrm{sgn}(\xi + \alpha\xi_1)\, d\alpha\Big)^2,$$

which is precisely the square of the symbol of the first commutator of Calderón.

**Theorem 6.1.** *Let $a \neq 0$ and $b \neq 0$ and consider the expression*

$$p.v. \int_{\mathbb{R}^2} \Big( \frac{\Delta_{at}}{t} \circ \frac{\Delta_{bs}}{s} A(x) \Big) f(x+t+s) \, \frac{dt}{t} \, \frac{ds}{s}.$$

*Viewed as a bilinear operator in $f$ and $A''$, it extends naturally as a bounded operator from $L^p \times L^q$ into $L^r$ for every $1 < p, q \leq \infty$ with $1/p + 1/q = 1/r$ and $1/2 < r < \infty$.*

To prove this theorem, one applies the same method described earlier for the first commutator. One just has to observe that the symbol of this operator is given by

$$\Big( \int_0^1 \operatorname{sgn}(\xi + \alpha a \xi_1) \, d\alpha \Big) \Big( \int_0^1 \operatorname{sgn}(\xi + \alpha b \xi_1) \, d\alpha \Big),$$

and after that to realize that *each factor* satisfies the same desired *quadratic estimates.* So this time one needs to decompose each factor as a double Fourier series as we did before. The fact that one can go all the way down to $1/2$ with the estimates is a simple consequence of the statement that series of type

$$\sum_{n_1, n_2 \in \mathbb{Z}} |C(n_1, n_1)|^r \log {<} n_1 {>} \log {<} n_2 {>}$$

are always convergent as long as the constants $C(n_1, n_2)$ decay at least quadratically in $n_1$ and $n_2$ and $r > 1/2$. The details are straightforward and are left to the reader. Clearly, one can generalize the above theorem even further, in the most obvious way. We will come back to this in the second paper of the sequel.

Another generalization we have in mind comes from the identity

$$A'B' = (AB)'' - (BA')' - (AB')' + A'B'. \tag{6.3}$$

As a consequence of this identity, the right hand side of (6.3) satisfies Hölder estimates of the form

$$\big\| (AB)'' - (BA')' - (AB')' + A'B' \big\|_r \lesssim \|A'\|_p \, \|B'\|_q$$

for indices $p, q$, and $r$ as before. *Does this inequality continue to hold if one replaces every derivative $D$ by its modulus $|D|$?* As before, a direct calculation shows that the new expression

$$|D|^2 (AB) - |D| \big( B \, |D| \, A \big) - |D| \big( A \, |D| \, B \big) + (|D| \, A)(|D| \, B)$$

can be rewritten as

$$p.v. \int_{\mathbb{R}^2} \Big( \frac{\Delta_t}{t} A(x+s) \Big) \Big( \frac{\Delta_s}{s} B(x+t) \Big) \, \frac{dt}{t} \, \frac{ds}{s}. \tag{6.4}$$

The right way to look at this formula is to view it as a bilinear operator in $A'$ and $B'$. Its symbol can be calculated quite easily and it is given by

$$\Big( \int_0^1 \operatorname{sgn}(\xi_1 + \alpha \xi_2) \, d\alpha \Big) \Big( \int_0^1 \operatorname{sgn}(\xi_2 + \beta \xi_1) \, d\beta \Big) \tag{6.5}$$

which is a symmetric function in the variables $\xi_1$ and $\xi_2$. Because of this symmetry we like to call expressions such as the ones in (6.4) *circular commutators.* We will return to them in the second paper of the series.

**Theorem 6.2.** *Let $a \neq 0$ and $b \neq 0$ and consider the expression*

$$p.v. \int_{\mathbb{R}^2} \Big(\frac{\Delta_{at}}{t}\, A(x+s)\Big)\Big(\frac{\Delta_{bs}}{s}\, B(x+t)\Big)\, \frac{dt}{t}\, \frac{ds}{s}.$$

*Viewed as a bilinear operator in $A'$ and $B'$, it extends naturally as a bounded operator from $L^p \times L^q$ into $L^r$ for every $1 < p, q \leq \infty$ with $1/p + 1/q = 1/r$ and $1/2 < r < \infty$.*

The proof uses the same method, since it is not difficult to see that the symbols of such bilinear operators are again products of symbols of the first commutator kind and they each satisfy the same *quadratic estimates.*

## References


[1] Auscher, P., Hofmann, S., Muscalu, C., Tao, T. and Thiele, C.: Carleson measures, trees, extrapolation and $Tb$ theorems. *Publ. Mat.* **46** (2002), no. 2, 257–325.

[2] Calderón, A.: Commutators of singular integral operators. *Proc. Nat. Acad. Sci. U.S.A.* **53** (1965), 1092–1099.

[3] Calderón, A.: Commutators, singular integrals on Lipschitz curves and applications. In *Proc. Int. Congress of Math. (Helsinki, 1978)*, 85–96. Acad. Sci. Fennica, Helsinki, 1980.

[4] Coifman, R. R. and Meyer, Y.: *Wavelets. Calderón–Zygmund and multilinear operators.* Cambridge Studies in Advanced Mathematics 48, Cambridge University Press, Cambridge, 1997.

[5] Coifman, R. R., McIntosh, A. and Meyer, Y.: L'intégrale de Cauchy définit un opérateur borné sur $L^2$ pour les courbes Lipschitziennes. *Ann. of Math. (2)* **116** (1982), no. 2, 361–387.

[6] Fefferman, C.: Recent progress in classical Fourier analysis. In *Proceedings of the ICM (Vancouver, 1974), Vol. 1,* 95–118. Canad. Math. Congress, Montreal, Que., 1975.

[7] Grafakos, L. and Li, X.: Uniform bounds for the bilinear Hilbert transform. I. *Ann. of Math. (2)* **159** (2004), no. 3, 889–933.

[8] Lacey, M. and Thiele, C.: $L^p$ estimates on the bilinear Hilbert transform for $2 < p < \infty$. *Ann. of Math. (2)* **146** (1997), no. 3, 693–724.

[9] Lacey, M. and Thiele, C.: On Calderón's conjecture. *Ann. of Math. (2)* **149** (1999), no. 2, 475–496.

[10] Li, X.: Uniform bounds for the bilinear Hilbert transform. II. *Rev. Mat. Iberoam.* **22** (2006), no. 3, 1069–1126.

[11] Muscalu, C., Pipher, J., Tao, T. and Thiele, C.: Bi-parameter paraproducts. *Acta Math.* **193** (2004), no. 2, 269–296.

[12] Muscalu, C., Pipher, J., Tao, T. and Thiele, C.: Multi-parameter paraproducts. *Rev. Mat. Iberoam.* **22** (2006), no. 3, 963–976.

[13] Muscalu, C., Tao, T. and Thiele, C.: $L^p$ estimates for the biest. II. The Fourier case. *Math. Ann.* **329** (2004), no. 3, 427–461.

[14] Palsson, E.: *$L^p$ estimates for a singular integral operator motivated by Calderón's second commutator*. PhD Thesis, Cornell University, 2011.

[15] Stein, E. M.: *Harmonic analysis: real-variable methods, orthogonality and oscillatory integrals*. Princeton Mathematical Series 43, Monographs in Harmonic Analysis III, Princeton University Press, Princeton, NJ, 1993.

[16] Thiele, C.: A uniform estimate. *Ann. of Math. (2)* **156** (2002), no. 2, 519–563.

Received November 11, 2012.

Camil Muscalu: Department of Mathematics, Cornell University, Ithaca, NY, 14850, USA.
E-mail: camil@math.cornell.edu

The author is a member of the Simion Stoilow Institute of Mathematics of the Romanian Academy. The present work has been partially supported by the NSF.